THE SPATIO-TEMPORAL

DISTRIBUTION OF HONEY BEES AND

FLORAL RESOURCES IN AUSTRALIA

Jonathan Arundel

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

November 2015

Supervisor: Stephan Winter

Melbourne School of Engineering

The University of Melbourne

Produced on archival quality paper
ABSTRACT

The Western honey bee, *Apis mellifera*, is the most economically important pollinator worldwide. In Australia today there are more than half a million hives managed by beekeepers around the country. In addition, there is a large population of feral *A. mellifera* providing free pollination services to native and disturbed environments.

The sustainability of pollinator populations globally, including both managed and feral *A. mellifera*, is under considerable pressure from four key threats; pests and pathogens, alien species, land-use intensification and climate change. Understanding what impact these threats will have on pollination in Australia requires better knowledge of the spatio-temporal distribution of honey bee colonies (both managed and feral) and the resources on which these colonies depend.

In Australia, the life of honey bees is inextricably linked to the flowering of eucalypts. The majority of honey in Australia is produced from eucalypts, and beekeepers frequently move their hives across distances of hundreds of kilometres in pursuit of flowering events. Little is known about how this erratic flowering effects the distribution and abundance of feral colonies.

This thesis demonstrates how agent-based models can be used in conjunction with genetics-based field survey methods to make inferences about the density of feral honey bee colonies in south eastern Australia. Colony densities vary with number of colonies observed in the sample in a log-linear relationship, rather than a linear relationship as previously thought. As a consequence, previous field surveys have overestimated
densities by an order of magnitude. The surveys described in this thesis have shown that densities of feral honey bees are remarkably uniform across different types of environments, with some evidence that densities are marginally higher in undisturbed environments. Regardless of the type of environment, densities of feral honey bees are not high enough to provide pollination of most horticultural or agricultural crops.

The spatial-temporal distribution of managed honey bees is determined by the aggregate behaviour of individual beekeepers. This thesis shows that in planning the movement of their hives from one set of flowering events to the next beekeepers need to solve complex optimisation problems. Where a beekeeper is able to use foresight to predict the location of future flowering events, the routes they choose can be up to 2-6% shorter than those obtained from using current flowering information only.

To better understand the spatio-temporal distribution of floral resources, this thesis outlines the development of a new software application that processes remote-sensing data and makes the output available to beekeepers and researchers. This application is used to highlight agreement between patterns of eucalypt growth, flowering, and honey production at landscape scale.

Further work is still needed to fully understand the relationships between climate, growth, eucalypt flowering, and the effects of flowering on feral colony densities and beekeeper movements. An agenda for future
research outlines how threats to pollination security in the future can be better managed.
DECLARATION

This is to certify that:

i. the thesis comprises only my original work towards the PhD except where indicated in the Preface,

ii. due acknowledgement has been made in the text to all other material used,

iii. the thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices

Signed: [Signature]

Date: 10/11/2015

Jonathan Arundel
## PREFACE

This thesis is based on the following published journal papers for which I was lead author:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Author</th>
<th>Planning</th>
<th>Execution</th>
<th>Contribution</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arundel, J., Oldroyd, B. P. and Winter, S. (2012) 'Modelling honey bee queen mating as a measure of feral colony density', <em>Ecological Modelling</em>, 247, 48-57.</td>
<td>Jonathan Arundel</td>
<td>Conceived and designed the experiment.</td>
<td>Developed, executed and analysed simulations.</td>
<td>Wrote initial draft, responded to co-author, editor and reviewer feedback.</td>
<td>70%</td>
</tr>
<tr>
<td></td>
<td>Benjamin P. Oldroyd</td>
<td></td>
<td></td>
<td>Contributed to manuscript.</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>Stephan Winter</td>
<td>Contributed to design.</td>
<td></td>
<td>Provided editorial advice.</td>
<td></td>
</tr>
<tr>
<td>Arundel, J., Oldroyd, B. P. and Winter, S. (2013) 'Modelling estimates of honey bee (<em>Apis</em> spp.) colony density from drones', <em>Ecological Modelling</em>, 267, 1-10.</td>
<td>Jonathan Arundel</td>
<td>Conceived and designed the experiment.</td>
<td>Developed, executed and analysed simulations.</td>
<td>Wrote initial draft, responded to co-author, editor and reviewer feedback.</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>Benjamin P. Oldroyd</td>
<td></td>
<td></td>
<td>Provided editorial advice.</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Stephan Winter</td>
<td></td>
<td></td>
<td>Provided editorial advice.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Peter R. Oxley</td>
<td></td>
<td>Genotyped samples.</td>
<td>Contributed to manuscript.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alen Faiz</td>
<td></td>
<td>Collected samples. Genotyped samples.</td>
<td></td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>John Crawford</td>
<td></td>
<td></td>
<td>Provided editorial advice.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stephan Winter</td>
<td></td>
<td></td>
<td>Provided editorial advice.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Benjamin P. Oldroyd</td>
<td>Co-conceived and designed the experiment.</td>
<td>Prepared and deployed test hives.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In addition, this thesis contains material from the following peer-reviewed conference abstract:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Author</th>
<th>Planning</th>
<th>Contribution</th>
<th>Publication</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stephan Winter</td>
<td></td>
<td></td>
<td>Provided editorial advice.</td>
<td>10%</td>
</tr>
</tbody>
</table>

This thesis is also based on the following paper submitted to an international journal for review:

<table>
<thead>
<tr>
<th>Publication</th>
<th>Author</th>
<th>Planning</th>
<th>Contribution</th>
<th>Publication</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stephan Winter</td>
<td>Contributed to the design of the application.</td>
<td></td>
<td>Provided editorial advice.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Guan Gui</td>
<td>Contributed to the design of the application.</td>
<td></td>
<td>Provided editorial advice.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marie R. Keatley</td>
<td>Contributed to the design of the application.</td>
<td></td>
<td>Provided editorial advice.</td>
<td></td>
</tr>
</tbody>
</table>
Finally, I also contributed to the following journal paper which is discussed in this thesis:


<table>
<thead>
<tr>
<th>Publication</th>
<th>Author</th>
<th>Planning</th>
<th>Execution</th>
<th>Publication</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hinson, E. M., Duncan, M., Lim, J., Arundel, J. and Oldroyd, B. P. (2015)</td>
<td>Other authors</td>
<td>Conceived and designed the experiment.</td>
<td>Prepared and deployed drone traps. Collected samples. Genotyped samples.</td>
<td>Wrote initial draft, responded to co-author, editor and reviewer feedback.</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>Jonathan Arundel</td>
<td></td>
<td>Executed and analysed simulations.</td>
<td>Contributed to manuscript.</td>
<td>10%</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would firstly like to acknowledge my supervisor, Stephan Winter, for his guidance and his patience over the last seven and a half years. It was a course I took under Stephan that inspired me to undertake a PhD, and Stephan has provided unfailing support even when I pursued avenues of investigation outside his areas of research interest. The end result may be more "science with geographic information" than "geographic information science" but throughout the journey I've always tried to be, like Stephan, an ambassador for GIScience.

I'd also like to thank my unofficial co-supervisor, Ben Oldroyd from the University of Sydney. The opportunity to work with Ben on surveying feral colony densities changed the direction of my research, and I consider myself very lucky to have had this experience. Ben has always been generous with his time and he mentored me through the writing of my first papers. Ben has also been very patient throughout our collaboration; not an easy task given our different academic backgrounds, my work commitments, and my enrolment at a different institution in a different city.

My Advisory Committee, Matt Duckham and Angus Campbell, have throughout my candidature helped me to sharpen my focus and avoid distractions and dead ends. Marie Keatley has taught me about eucalypt phenology, and Peter Oxley from the University of Sydney has given me a crash course on genetics. Michael Duncan and Alen Faiz assisted with fieldwork, and Peter and Alen undertook the lab work for the feral surveys. At Puckapunyal Military Area, Sean-Paul Smith provided the
valuable assistance needed to successfully conduct a scientific experiment in an area where army training exercises are conducted. Guan Gui did an outstanding job as the lead developer of BeeBox, and he has helped me keep BeeBox up and running since its launch in 2012.

I have also been fortunate to receive support from the beekeeping community. Castlemaine beekeepers Bob McDonald and his son Peter have taught me much about honey bees, and they also provided the hives for our fieldwork. Bob and Peter have both taken time out of running their large commercial beekeeping operation to assist with my research and provide a sounding board for new ideas. They also helped me set up my first hive. Capilano Honey, and in particular Bill Winner, have been generous in their support of my research through sharing their honey production data.

I also gratefully acknowledge the Rural Industries Research and Development Corporation (RIRDC), who have provided financial support for my research in the form of a top-up stipend for my PhD, and funding for the development of BeeBox under Project Number 004872.

Finally, I'd like to thank my partner Alice Olenich for making all of this possible. The sacrifices I've made in undertaking a PhD have been sacrifices that Alice has shared with me, and even as the years have stretched on her support has been unwavering.
# Table of Contents

1 Introduction ................................................................. 1

2 The Western Honey Bee .................................................. 5

2.1 Introduction ................................................................................................................................. 5

2.2 Honey Bees and Other Pollinators Globally ................................................................. 5

2.2.1 Honey bee origins ............................................................. 5

2.2.2 The economics of pollination and the honey bee industry ................................. 7

2.2.3 Pollinator shortfalls and pollinator declines ................................................ 10

2.3 Honey Bee Behaviour ........................................................... 15

2.4 Honey Bee Genetics ............................................................... 23

2.5 Managed Honey Bees in Australia ............................................ 26

2.5.1 Honey bees in Australia ...................................................... 26

2.5.2 The Australian honey bee industry .............................................. 26

2.5.3 Honey and pollen flora in Australia ............................................... 32

2.5.4 Crop pollination in Australia ..................................................... 34

2.5.5 Honey bee colony losses in Australia ........................................... 34

2.6 Feral Honey Bees ................................................................. 35

2.6.1 Feral honey bee impacts ............................................................ 36

2.6.2 Surveying feral colony densities .............................................. 38
5.2.1 Abstract .................................................................................................................... 147

5.2.2 Introduction ........................................................................................................... 148

5.2.3 Materials and Methods ...................................................................................... 151

5.2.4 Results ....................................................................................................................... 156

5.2.5 Discussion ................................................................................................................ 161

5.2.6 Acknowledgements ............................................................................................. 167

5.3 Estimating feral colony densities in South East Australia from captured drones .....................................................................................................................................................167

5.3.1 Site selection and land use .............................................................................. 168

5.3.2 Generating sampling distributions .............................................................. 173

5.3.3 Results ....................................................................................................................... 174

5.3.4 Discussion ................................................................................................................ 176

6 The spatio-temporal distribution of managed colonies ..................... 182

6.1 Introduction ........................................................................................................... 182

6.2 Motivation .................................................................................................................. 183

6.3 Model development ............................................................................................. 186

6.3.1 Simulation design ................................................................................................ 186

6.3.2 Formal specification ........................................................................................... 189

6.3.3 Computational complexity ................................................................................ 191
6.3.4 Heterogeneous attractiveness ................................................................. 192

6.4 EXPERIMENTS ................................................................................................................................. 193

6.4.1 Experiment 1: Effect of foresight on performance ......................... 194

6.4.2 Experiment 2: Effect of heterogeneous attractiveness on performance ....................................................................................................................... 197

6.5 DISCUSSION ..................................................................................................................................... 198

6.6 CONCLUSION ................................................................................................................................... 200

7 THE SPATIO-TEMPORAL DISTRIBUTION OF EUCALYPT FLOWERING ........ 202

7.1 EUCALYPTS AND REMOTE SENSING .................................................................................. 202

7.1.1 Introduction ................................................................................................................................. 202

7.1.2 The distribution of eucalypts ................................................................................................. 202

7.1.3 The phenology of eucalypts ................................................................................................. 209

7.1.4 Remote sensing of vegetation ................................................................................................. 218

7.1.5 Vegetation indices ..................................................................................................................... 224

7.1.6 Satellite platforms and sensors ................................................................................................. 230

7.1.7 Remote sensing for phenology ................................................................................................. 240

7.1.8 Remote sensing of eucalypts and eucalypt phenology ....................... 247

7.1.9 Summary ................................................................................................................................... 251

7.2 VISUALISING PATTERNS OF EUCALYPT GROWTH .................................................. 252

xv
LIST OF TABLES

TABLE 1: HONEY BEE SPECIES WITHIN THE GENUS *Apis* .............................................................. 6

TABLE 2: ESTIMATES OF THE VALUE OF POLLINATION SERVICES BY REGION IN $AUD BILLIONS (UNADJUSTED FOR INFLATION) .............................................................................. 8

TABLE 3: EXAMPLE OF TWO SEASONS’ ACTIVITIES FOR A VICTORIAN MIGRATORY BEEKEEPER (McDONALD, 2007) ....................................................................................... 30

TABLE 4: SUMMARY OF ALL STUDIES GLOBALLY INTO THE DENSITIES OF FERAL *Apis mellifera* colonies, sorted by estimated density............................................................. 45

TABLE 5: VEHICLE ROUTING PROBLEM (VRP) PARAMETERS RELEVANT TO MANAGED BEEKEEPER HIVE MOVEMENTS .............................................................................................. 73

TABLE 6: ELEMENTS OF THE ODD PROTOCOL (ADDITIONS FROM 2010 IN BOLD) .......... 83

TABLE 7: ENTITIES, STATE VARIABLES AND SCALES ................................................................... 94

TABLE 8: THE EFFECT OF THE SPATIAL DISTRIBUTION OF FERAL COLONIES ON THE NUMBER OF FERAL COLONIES CAPTURED IN TEST QUEEN PROGENY WITH A SINGLE TEST QUEEN .......................................................................................................................111

TABLE 9: THE EFFECT OF THE SPATIAL DISTRIBUTION OF FERAL COLONIES ON THE NUMBER OF FERAL COLONIES CAPTURED IN TEST QUEEN PROGENY WITH RANDOMLY DISTRIBUTED DCAS AND 10 TEST QUEENS ..........................................................113

TABLE 10: THE EFFECT OF TEST QUEEN NUMBER ON THE PROPORTION OF FERAL COLONIES ‘CAPTURED’ IN THE PROGENY OF TEST QUEENS ..................................................................115

TABLE 11: ENTITIES, STATE VARIABLES AND SCALES .............................................................. 124
TABLE 24: COMPARISON OF MODIS AND MERIS SPECTRAL BANDS ........................................237

TABLE 25: NOISE FILTERING TECHNIQUES USED IN LAND SURFACE PHENOLOGY STUDIES
.................................................................................................................................................. 243
LIST OF FIGURES

FIGURE 1: APIARY SITES IN VICTORIA SHOWN IN COMPARISON TO BIOREGIONS................... 28

FIGURE 2: ILLUSTRATIVE LOCATIONS AND NETWORK OF HIVE MOVEMENTS FOR TWO SEASONS' ACTIVITIES FOR A VICTORIAN MIGRATORY BEEKEEPER ................................. 31

FIGURE 3: PRE-INITIALISATION AND INITIALISATION PROCESS OVERVIEW OF THE SIMULATIONS............................................................................................................................. 98

FIGURE 4: SIMULATION PROCESS OVERVIEW ............................................................................. 100

FIGURE 5: GENERATING EQUAL DENSITY SPATIAL DISTRIBUTIONS ......................................... 105

FIGURE 6: RESULTS OF 100 SIMULATIONS FOR ONE TEST QUEEN. EACH HISTOGRAM SUMMARISES 100 SIMULATIONS FOR THE SPECIFIED COMBINATION OF DCA DENSITY AND FERAL COLONY DENSITY. ............................................................................................................................. 108

FIGURE 7: RESULTS OF 100 SIMULATIONS FOR 10 TEST QUEENS ........................................ 109

FIGURE 8: FLOWCHART FOR PRE-INITIALISATION AND INITIALISATION PROCESS .......... 127

FIGURE 9: FLOWCHART FOR SIMULATION PROCESS .................................................................... 129

FIGURE 10: RESULTS OF 100 SIMULATIONS FOR 96 DRONES AT DIFFERING DCA AND COLONY DENSITIES .................................................................................................................. 135

FIGURE 11: LOG-LINEAR RELATIONSHIP OF UNIQUE COLONIES OBSERVED IN A SAMPLE WITH COLONY DENSITY ........................................................................................................ 136

FIGURE 12: THE RELATIONSHIP BETWEEN FERAL COLONY DENSITY AND THE NUMBER OF COLONIES REPRESENTED IN TEST QUEEN BROODS ASSUMING (A) 3, (B) 4 AND (C) 5 TEST QUEENS BASED ON AGENT-BASED MODELLING (ARUNDEL ET AL., 2012) ...... 164
FIGURE 25: MODIS VI TILE COVERAGE FOR AUSTRALIA (GOOGLE EARTH, 2013) ........239

FIGURE 26: BeeBox Application Architecture.................................................................262

FIGURE 27: BeeBox Visualisation Comparing MODIS 16-day Images Before and
After the Victorian Black Saturday Bushfires...........................................................265

FIGURE 28: Eucalyptus tricarpa (green dots) and Eucalyptus sideroxylon (grey
dots) Herbaria Records.........................................................................................267

FIGURE 29: Comparison of EVI Growth Time Series (A) with Flowering Data (B)
and Honey Production (C) for Eucalyptus tricarpa.............................................269

FIGURE 30: BeeBox Region Analysis Wizard Showing Upload of Google Earth
KML File Against NDVI Background......................................................................271

FIGURE 31: Comparison of EVI Time Series Data with Seasonal Rainfall and
Monthly Maximum Daily Temperature Averages for Tunstalls Nature
Conservation Reserve............................................................................................278

FIGURE 32: Comparison of EVI Time Series Data with Seasonal Rainfall and
Monthly Maximum Daily Temperature Averages for Havelock Nature
Conservation Reserve............................................................................................279

FIGURE 33: Comparison of EVI Time Series Data with Seasonal Rainfall and
Monthly Maximum Daily Temperature Averages for Puckapunyal Army
Reserve ....................................................................................................................280

FIGURE 34: Comparison of EVI Time Series Data with Seasonal Rainfall and
Monthly Maximum Daily Temperature Averages for Rushworth State
Forest.......................................................................................................................281
FIGURE 35: COMPARISON OF EVI TIME SERIES DATA WITH SEASONAL RAINFALL AND MONTHLY MAXIMUM DAILY TEMPERATURE AVERAGES FOR WARBY OVENS NATIONAL PARK ........................................................................................................................................ 282

FIGURE 36: RAW AND FILTERED EVI TIME SERIES FOR BOX IRONBARK (EVC61) SITES IN VICTORIA................................................................................................................................. 284

FIGURE 37: SELECTED BOX IRONBARK (EVC61) SITES (YELLOW POLYGONS/GREEN LABELS), TEMPERATURE DATA SITES (RED LABELS), AND RAINFALL DATA SITES (BLUE LABELS) ACROSS VICTORIA. SITES WITH MATCHING EVI TIME SERIES ARE SHOWN LINKED WITH A GREEN LINE. ............................................................................................... 286

FIGURE 38: COMPARISON OF EVI TIME SERIES DATA, 90-DAY ROLLING DAILY AVERAGE RAINFALL AND MONTHLY MAXIMUM DAILY TEMPERATURE AVERAGES FOR THE RUSHWORTH STATE FOREST AND WARBY-OVENS NATIONAL PARK SITES. .......... 288

FIGURE 39: AGGREGATED SPATIAL POINT PATTERN WITH KERNEL DENSITY BANDWIDTHS OF (A) 0.28 KM, (B) 4 KM, AND (C) 15 KM ......................................................................................................................... 299

FIGURE 40: CAUSAL RELATIONSHIP CHAIN LINKING CLIMATE TO COLONY DENSITIES VIA VEGETATION GROWTH AND FLOWERING ........................................................................................................... 305

FIGURE 41: SIMULATION SPACE WITH RANDOMLY DISTRIBUTED DCAS (GREEN STARS) AND RANDOMLY DISTRIBUTED FERAL COLONIES (CIRCLES). WITHIN RANGE COLONIES SAMPLED ARE BLUE, AND NOT SAMPLED ARE RED .............................................................................................................. 415

FIGURE 42: SIMULATION SPACE WITH RANDOMLY DISTRIBUTED DCAS (GREEN STARS) AND AGGREGATED FERAL COLONIES (CIRCLES). WITHIN RANGE COLONIES SAMPLED ARE BLUE, AND NOT SAMPLED ARE RED .............................................................................................................. 416
FIGURE 43: SIMULATION SPACE WITH RANDOMLY DISTRIBUTED DCAS (GREEN STARS) AND OVERDISPERSED FERAL COLONIES (CIRCLES). WITHIN RANGE COLONIES SAMPLED ARE BLUE, AND NOT SAMPLED ARE RED. ................................................................. 417

FIGURE 44: RESULTS OF 100 SIMULATIONS FOR A SAMPLE SIZE OF 96 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .................. 418

FIGURE 45: RESULTS OF 100 SIMULATIONS FOR A SAMPLE SIZE OF 96 DRONES WITH HYPER-AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS ........ 419

FIGURE 46: RESULTS OF 100 SIMULATIONS FOR A SAMPLE SIZE OF 96 DRONES WITH OVERDISPERSED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .......... 420

FIGURE 47: RESULTS OF 100 SIMULATIONS FOR A SAMPLE SIZE OF 96 DRONES WITH RANDOMLY DISTRIBUTED FERAL COLONIES AND AGGREGATED DCAS .............. 421

FIGURE 48: RESULTS OF 100 SIMULATIONS FOR A SAMPLE SIZE OF 15 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .............. 422

FIGURE 49: RESULTS OF 100 SIMULATIONS FOR A SAMPLE SIZE OF 192 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .......... 423

FIGURE 50: DISTRIBUTION OF MEAN FLIGHT DISTANCES OF ALL DRONES PRESENT AT THE SAMPLED DCA WITH RANDOMLY DISTRIBUTED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS................................................................. 424

FIGURE 51: DISTRIBUTION OF FLIGHT DISTANCES FOR ALL DRONES PRESENT AT THE SAMPLED DCA ACROSS ALL SIMULATIONS FOR RANDOMLY GENERATED FERAL COLONIES OF DENSITY 0.1/km² AND RANDOMLY DISTRIBUTED DCAS OF DENSITY 0.1/km² ................................................................. 425
FIGURE 52: DISTRIBUTION OF THE PROPORTION OF COLONIES DETECTED IN THE SAMPLE
VERSUS WITHIN RANGE OF THE DCA FOR A SAMPLE SIZE OF 15 DRONES WITH
RANDOMLY DISTRIBUTED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS...426

FIGURE 53: DISTRIBUTION OF THE PROPORTION OF COLONIES DETECTED IN THE SAMPLE
VERSUS WITHIN RANGE OF THE DCA FOR A SAMPLE SIZE OF 96 DRONES WITH
RANDOMLY DISTRIBUTED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS...427

FIGURE 54: DISTRIBUTION OF THE PROPORTION OF COLONIES DETECTED IN THE SAMPLE
VERSUS WITHIN RANGE OF THE DCA FOR A SAMPLE SIZE OF 192 DRONES WITH
RANDOMLY DISTRIBUTED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS...428

FIGURE 55: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE
SIZE OF 14 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY
DISTRIBUTED DCAS..............................................................................................................429

FIGURE 56: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A
SAMPLE SIZE OF 14 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY
DISTRIBUTED DCAS..............................................................................................................430

FIGURE 57: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE
SIZE OF 29 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY
DISTRIBUTED DCAS..............................................................................................................431

FIGURE 58: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A
SAMPLE SIZE OF 29 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY
DISTRIBUTED DCAS..............................................................................................................432
FIGURE 59: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE SIZE OF 62 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .................................................................433

FIGURE 60: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A SAMPLE SIZE OF 62 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .................................................................434

FIGURE 61: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE SIZE OF 70 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .................................................................435

FIGURE 62: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A SAMPLE SIZE OF 70 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .................................................................436

FIGURE 63: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A SAMPLE SIZE OF 72 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .................................................................437

FIGURE 64: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE SIZE OF 74 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .................................................................438

FIGURE 65: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A SAMPLE SIZE OF 74 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS .................................................................439
FIGURE 66: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE SIZE OF 79 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS..............................................................................................................440

FIGURE 67: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A SAMPLE SIZE OF 79 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS..............................................................................................................441

FIGURE 68: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE SIZE OF 123 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS..............................................................................................................442

FIGURE 69: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A SAMPLE SIZE OF 123 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS..............................................................................................................443

FIGURE 70: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE SIZE OF 222 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS..............................................................................................................444

FIGURE 71: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A SAMPLE SIZE OF 222 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS..............................................................................................................445

FIGURE 72: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE SIZE OF 241 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY DISTRIBUTED DCAS..............................................................................................................446
FIGURE 73: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A
SAMPLE SIZE OF 241 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY
DISTRIBUTED DCAs.............................................................................................................. 447

FIGURE 74: RESULTS OF 100 SIMULATIONS AT VARYING DCA DENSITIES FOR A SAMPLE
SIZE OF 278 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY
DISTRIBUTED DCAs.............................................................................................................. 448

FIGURE 75: RESULTS OF 100 SIMULATIONS AT A DCA DENSITY OF 1 DCA/KM² FOR A
SAMPLE SIZE OF 278 DRONES WITH AGGREGATED FERAL COLONIES AND RANDOMLY
DISTRIBUTED DCAs.............................................................................................................. 449
Chapter 1: Introduction

1 INTRODUCTION

The Western honey bee, *Apis mellifera*, is the most economically important pollinator worldwide. In Australia, *A. mellifera* was introduced with the arrival of European settlers, and today there are more than half a million hives managed by beekeepers around the country. In addition to the sale of honey and beeswax, beekeepers earn income from providing pollination services to growers of pollination-dependent crops such as almonds. In areas where pollination occurs without the presence of managed hives, this is provided by the large population of feral *A. mellifera*.

In Australia, the life of honey bees is inextricably linked to the flowering of eucalypts. The majority of honey in Australia is produced from eucalypts, and beekeepers frequently move their hives across distances of hundreds of kilometres in pursuit of flowering events. Eucalypts flower erratically, and so beekeepers must become experts in observing the predictors of flowering and planning their seasonal movements accordingly. Little is known about how this erratic flowering effects the distribution and abundance of feral colonies.
The sustainability of pollinator populations globally, including both managed and feral *A. mellifera*, is under considerable pressure from four key threats. Pests and pathogens have caused significant losses of managed colonies in recent decades and have had an even more dramatic effect on feral populations. The introduction of alien species, even alien races of *A. mellifera* itself, lead to increase competition for resources and possible displacement of existing pollinators. Land-use intensification, including the removal of habitat and the use of pesticides, is a continuing pressure on pollinator populations. Finally, climate change may also be playing a role in pollinator decline; quantifying this impact is challenging.

The initial motivation of this PhD was to gain insight and understanding into the threat posed to pollination in Australia by the introduction of pests and pathogens, most specifically the parasitic honey bee mite *Varroa destructor*. While others have since attempted to address that question (Clifford *et al.*, 2011, Hafi *et al.*, 2012, Gordon *et al.*, 2014), none of these efforts have directly addressed the role played by feral *A. mellifera* colonies as pollinators, or indeed as vectors of pests or pathogens. Despite prevailing biosecurity policies calling for the elimination of feral colonies in response to an incursion (Animal Health Australia, 2010), remarkably little is known about the densities of feral colonies across different landscapes, and how these densities change over time with the effects of weather, climate change and fire. This thesis seeks to address these gaps in knowledge and thus provide a foundation on which the impacts of pollinator pressures can be better understood and predicted.
Chapter 1: Introduction

To understand the dynamics of the honey bee environment in Australia, this thesis has broken the research question into three parts;

(a) an assessment of the spatial distribution of feral colonies across different landscapes;
(b) an attempt to model behaviour of beekeepers moving their hives in response to changing resource availability; and
(c) a study into the temporal change of floral resources, as resource availability underpins both changes in the abundance of feral populations and patterns of beekeeper movement.

The remainder of this thesis is structured in an attempt to answer these research questions as follows:

- **Chapter 2** contains a general introduction to honey bees and the honey bee industry in Australia.
- **Chapter 3** discusses the relevance of spatially-explicit models in gaining insights into the dynamics of the honey bee environment.
- **Chapter 4** describes how simulations of the honey bee mating system can be used to generate synthetic sampling distributions for the application of genetics-based survey techniques of feral colonies, and contains two journal papers published in Ecological Modelling.
- **Chapter 5** applies the sampling distributions from Chapter 4 to two large-scale field surveys of feral colonies in Victoria, Australia and contains a journal paper from Austral Entomology.
• Chapter 6 presents a model of beekeepers moving their hives in response to resource availability.

• Chapter 7 provides an overview of the main floral resource of honey bees in Australia; eucalypts. It also highlights a relationship between remotely-sensed observations of eucalypt growth with flowering and honey production. A prototype web-based application has been developed for beekeepers to visualise these same patterns of growth, and is discussed in a journal paper currently under peer review.

• Chapter 8 concludes with a short summary of the major contributions of the research described in this thesis, provides a general discussion of progress made in analysing the spatio-temporal distributions of honey bees and floral resources in Australia, and outlines an agenda for further research.
2 THE WESTERN HONEY BEE

2.1 Introduction
Of all the honey bees, the Western honey bee, *Apis mellifera*, is the globally most important from an economic perspective. This chapter provides an introduction to the honey bee and the honey bee industry in Australia, and discusses threats to the honey bee and pollination generally. The current state of knowledge of the spatio-temporal distribution of both managed and feral honey bees is also discussed.

2.2 Honey bees and other pollinators globally

2.2.1 Honey bee origins
There are more than 16,000 species of bees worldwide, divided into seven families (Danforth *et al.*, 2006). Bees, wasps, ants and sawflies are members of the order of membranous-winged insects, the Hymenoptera (Oldroyd and Wongsiri, 2006). Bees co-evolved with angiosperms (flowering plants), favouring pollen and nectar as a source of nutrition for their nests, diverging from the preference to the predation exhibited by
wasps 120-130 million years ago (Winston, 1987, Oldroyd and Wongsiri, 2006). Oldroyd and Wongsiri (2006) recognise nine species of honey bees within the genus *Apis*, as shown in Table 1.

**Table 1: Honey bee species within the genus *Apis***

<table>
<thead>
<tr>
<th>Group</th>
<th>Species</th>
<th>Common name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwarf honey bees</td>
<td><em>Apis florea</em></td>
<td>Red dwarf honey bee</td>
</tr>
<tr>
<td></td>
<td>*Apis andreniformis</td>
<td>Black dwarf honey bee</td>
</tr>
<tr>
<td>Giant honey bees</td>
<td><em>Apis dorsata</em></td>
<td>Common giant honey bee</td>
</tr>
<tr>
<td></td>
<td><em>Apis laboriosa</em></td>
<td>Giant mountain honey bee</td>
</tr>
<tr>
<td>Cavity-nesting honey bees</td>
<td><em>Apis cerana</em></td>
<td>Eastern hive bee</td>
</tr>
<tr>
<td></td>
<td><em>Apis koschevnikovi</em></td>
<td>Red honey bee</td>
</tr>
<tr>
<td></td>
<td><em>Apis nulensis</em></td>
<td>Mountain honey bee</td>
</tr>
<tr>
<td></td>
<td><em>Apis nigrocincta</em></td>
<td>Sulawesian honey bee</td>
</tr>
<tr>
<td></td>
<td><em>Apis mellifera</em></td>
<td>Western honey bee</td>
</tr>
</tbody>
</table>

Commercial honey production globally is dominated by the Western honey bee, *Apis mellifera*. *A. mellifera* is native to Europe and Africa with a distribution extending from the Scandinavia and Russia in the north, through the Middle East and the entire African continent (Crane, 1999). Within its native range, 25 subspecies and races of *A. mellifera* have been identified differing in body size and behaviour (Ruttner, 1988, Ruttner, 1992, Crane, 1999). The dark bee of northern Europe, *A. m. mellifera*, was the race introduced into North America in 1616 (Crane, 1999), Australia in 1822 (Barrett, 1995) and New Zealand in 1839 (Crane, 1999). In the latter half of the nineteenth century *A. m. mellifera* was supplanted in commercial use by three other races with improved productivity and temperament; the Italian honey bee *A. m. ligustica*, the Carniolan honey bee *A. m. carnica*, and the Caucasian honey bee *A. m. caucasica* (Schiff and Sheppard, 1995, Crane, 1999, Warhurst and Goebel, 1999).
2.2.2 The economics of pollination and the honey bee industry

Pollination is an ecosystem service on which humans depend through its link to world food production (Hanley et al., 2015). While some plants depend on vertebrates such as birds and bats for pollination, the majority depend on invertebrates including bees, flies, beetles, wasps and thrips (Klein et al., 2007). Of all the invertebrate pollinators, A. mellifera is considered to be the most important pollinator in most regions of the world where flowers exist (Tautz, 2008).

For at least four decades, attempts to quantify the economic benefits of pollinators, and often more specifically Apis mellifera, have been made (Table 2). In the most recent estimate of the global total economic value of pollination services, Gallai et al. (2009) estimate this for 2005 as €153 billion, or 9.5% of world agricultural production used for human food. Where ranges of values provided in other estimates, these often arise from making assumptions about whether pollinator substitution could occur in the event that a primary pollinator is lost (Southwick and Southwick, 1992). In Australia, Gordon and Davis (2003) estimated the value of honey bee pollination at AUD$1.7 billion in 1999-2000, with a possible flow-on effect of another AUD$2 billion. This equates to ~45% of the total value of horticulture in the same period (AUD$3.8 billion).
Table 2: Estimates of the value of pollination services by region in $AUD billions (unadjusted for inflation)

<table>
<thead>
<tr>
<th>Region</th>
<th>Estimated value in $AUD billions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>$151.0</td>
<td>Costanza et al. (1998)</td>
</tr>
<tr>
<td></td>
<td>$258.0</td>
<td>Pimentel et al. (1997)</td>
</tr>
<tr>
<td></td>
<td>$211.1</td>
<td>Gallai et al. (2009)</td>
</tr>
<tr>
<td>Australia</td>
<td>$0.6-1.2</td>
<td>Gill (1991)</td>
</tr>
<tr>
<td></td>
<td>$1.7</td>
<td>Gordon and Davis (2003)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>$1.8-2.6</td>
<td>Matheson and Schrader (1987)</td>
</tr>
<tr>
<td>USA</td>
<td>$10.3</td>
<td>Martin (1975)</td>
</tr>
<tr>
<td></td>
<td>$24.5</td>
<td>Levin (1983)</td>
</tr>
<tr>
<td></td>
<td>$2.0-10.7</td>
<td>Southwick and Southwick (1992)</td>
</tr>
<tr>
<td></td>
<td>$12.0</td>
<td>Robinson et al. (1989)</td>
</tr>
<tr>
<td></td>
<td>$18.8</td>
<td>Morse and Calderone (2000)</td>
</tr>
<tr>
<td></td>
<td>$4.0</td>
<td>Losey and Vaughan (2006)</td>
</tr>
<tr>
<td></td>
<td>$19.5</td>
<td>Calderone (2012)</td>
</tr>
<tr>
<td>Canada</td>
<td>$0.5-1.3</td>
<td>Richards and Kevan (2002)</td>
</tr>
<tr>
<td>China</td>
<td>$67.3</td>
<td>An and Chen (2011)</td>
</tr>
<tr>
<td>Europe</td>
<td>$6.5</td>
<td>Borneck and Merle (1989)</td>
</tr>
<tr>
<td>France</td>
<td>$0.6</td>
<td>Borneck and Bricout (1984)</td>
</tr>
<tr>
<td>Hungary</td>
<td>Unknown</td>
<td>Benedek (1983)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Unknown</td>
<td>Fluri and Frick (2005)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>$0.4</td>
<td>Carreck and Williams (1998)</td>
</tr>
<tr>
<td>Egypt</td>
<td>$3.1</td>
<td>Brading et al. (2009)</td>
</tr>
</tbody>
</table>

Despite these decades of effort, there is still currently no reliable estimate of the value of the pollination ecosystem service at a global scale (Hein, 2009). Existing work has focussed on illustrative studies, with few studies useful for natural capital accounting or in cost-benefit analyses (Hanley et al., 2015). To progress the economic analysis of pollination services, Hanley et al. (2015) set the following five research priorities; (1) key pollinators for representative crops, (2) factors influencing substitution and synergies with pollinator communities, (3) monitoring schemes to establish linkages between habitats and pollinator populations, (4) econometric analysis linking insect pollinated crop yields
and prices, and (5) assessment of the non-market benefits of pollination services. Melathopoulos et al. (2015) have questioned the value of valuing pollination ecosystem services to agriculture, arguing that current estimates do not accurately estimate the contribution of wild pollinators, and are significantly based on estimates of pollination requirements for crops obtained from fewer than three field studies.

With *A. mellifera* as the primary pollinator of many crops, many beekeepers provide paid pollination services as a substantial source of revenue in their beekeeping operations. Champetier et al. (2015) estimate that in between 2006 and 2010, pollination revenues comprised about 49% of total revenues per hive for beekeepers in the US. Thus any comprehensive economic model of pollination services also needs to account for the economic behaviour of beekeepers. Economic models have been developed of beekeeping and honey production (Leonard and Long, 1992, Champetier et al., 2015). The differential equation model of Champetier et al. (2015) is based on profit maximisation by a single beekeeper who sells honey and provides pollination services. Constraints in the model include the availability of forage and the growth rate of bee and honey stocks. The authors conclude that contrary to prevailing beliefs, an increase in the price of honey may result in decreases in managed honey bee populations. They also reflect on the need for spatially explicit models of hive migration where multiple populations of bees are connected.
2.2.3 Pollinator shortfalls and pollinator declines

There has been an ~23% increase in the area cultivated for agriculture during the last five decades, and this expansion has mostly comprised crops that depend on pollinators (Aizen et al., 2008). Aizen and Harder (2009) used data from the United Nations Food and Agriculture Organisation to establish that the global population of managed honey bee hives has increased by ~45% in the last half century. However, over the same timeframe agricultural pollination demands have increased by more than 300%, highlighting potential future pollinator shortages. In the UK, insect pollinated crops accounted for 20% of cropland and 19% of farm gate value as of 2007. In the period since 1984, crop yields have risen by 54% yet the proportion of pollination capable of being supplied by honey bees has fallen from 70% to 34% (Breeze et al., 2011). In the UK at least, it appears that insect pollination is being provided primarily by insects other than managed honey bees. In a study conducted across 41 European countries, Breeze et al. (2014) found that the requirements for honey bee stocks to provide pollination services has risen 4.9 times as fast as the growth in actual honey bee numbers between 2005 and 2010. Such conclusions are drawn in a context where the understanding of the pollination requirements of different crops and the required stocking rates of honey bee colonies is still limited, as evidenced by wide ranges quoted for different crops in the literature (Breeze et al., 2014).

Aebi et al. (2012) point out that projections of future pollination shortfalls fail to take into account the changing dependence of crops on insect pollination, with the development of self-fertile true hybrids in
crops such as canola (*Brassica napus*). Ghazoul (2005) has questioned claims of a global pollination crisis, arguing that staple crops such as cereals (wheat, rice and maize) and tubers (potatoes, yams and cassava) are not dependent on animal pollinators.

Productivity of some monoecious (separate male and female flowers on the same individual) is either enhanced by pollinators or dependent on them. This is particularly the case for many economically-important members of the Rosaceae family such as almonds, apples, pears, plums, cherries, peaches, raspberries and strawberries (Ghazoul, 2005). Klein *et al.* (2007) determined that 87 of the leading global food crops are dependent on animal pollination accounting for 35% of global production volumes. While 28 do not require animal pollination, these crops account for 60% of global production volumes. Of 60 crops critical to the North American economy, seven crops are pollinated by wild insects (Buchmann and Nabhan, 1996). Veddeler *et al.* (2008) found that for coffee (*Coffea arabica*) a fourfold increase in bee density was associated with an 80% increase in yield and 800% increase in net revenues. Many legumes also depend on animal pollinators, and in turn are used as forage for animals. Of the legumes, lucerne (alfalfa), *Medicago sativa*, is considered the most economically important (Buchmann and Nabhan, 1996). While *A. mellifera* is not an efficient pollinator of some crops such as *M. sativa*, adding more colonies in a brute force approach to pollination ensures that sufficient pollination occurs (Morse, 1991).

In early 2007, media headlines around the world highlighted a mysterious loss of honey bee colonies in the US and Europe during the
winter and spring of 2006-2007 (Oldroyd, 2007). The media attention shifted public focus onto the threat posed by pollinator decline, typified by a (most-likely) apocryphal quote attributed to Albert Einstein; "If the bee disappeared off the surface of the globe then man would only have four years of life left." (snopes.com, 2007).

Nabhan et al. (1998) observed that as of 1998 managed pollinator numbers in North America were at their lowest level for 50 years. In the US, total overwintering losses of all honey bee colonies has fluctuated around 30% from 2006-2013, with the exception of one winter in 2011-2012 where total loss was 21.2% (Steinhauer et al., 2014). Unfortunately, beekeepers have only been surveyed for overwintering losses since the winter of 2006-2007, and thus no quantitative data is available to indicate how unusual the events of that winter were, but the data does show a consistent trend in the period since then. The commencement of the surveys comes eight years after a working group of more than 20 scientists from North America made recommendations to establish the geographic extent and magnitude of declines in invertebrate and vertebrate pollinators and the impacts to horticulture and agriculture in an appeal backed by 14 conservation and sustainable agriculture organisations (Nabhan et al., 1998).

Declines in beekeeper numbers and colony numbers have also been observed in Europe. In the period from 1965-2005, Simon Potts et al. (2010) have shown a decline in beekeeper numbers across all of Europe, and a decline in colony numbers in central European countries. This was in contrast to an increase in colony numbers in some Mediterranean
countries. Over the winter of 2012-2013, van der Zee et al. (2014) established colony losses of over 30% in the United Kingdom, Scandinavia and the Baltic in line with those observed in the US, but in other countries losses were relatively low compared with previous years.

In South Africa, colony losses were 30% in 2009-2010 and 46% in 2010-2011, with losses higher for migratory beekeepers than for stationary beekeepers (Pirk et al., 2014).

In a review of studies into the key pressures on pollinators, (Vanbergen and Insect Pollinators Initiative, 2013) grouped causes into four categories; land-use intensification (including habitat fragmentation and pesticide use), climate change, alien species, pests and pathogens. Alien species and pests and pathogens are further discussed in Section 2.7 in this thesis.

The area and connectivity of habitat fragments is particularly important for the conservation of habitat specialists such as bees (Steffan-Dewenter, 2003). Kremen et al. (2004) found that in California a strong dependency existed between the pollination services provided by native bees and the proportion of natural upland habitat within 1-2.5 km of the farm site. Klein et al. (2012) have shown that wild bee species visited almond flowers only in orchards with adjacent semi-natural habitat or vegetation strips. Contradicting these results, a study conducted in New Jersey and Pennsylvania, Winfree et al. (2007) found that that crop visitation by wild bees was not associated by natural habitat cover at either local or landscape scale. In a synthesis of 23 studies representing 16 crops on 5
continents, Ricketts et al. (2008) observed strong exponential declines in pollinator richness and visitation rates with increasing distances from natural habitat. Pollinator visitation rates dropped to half of their maximum at 0.6 km from natural habitat. To test whether isolation from florally diverse natural habitats reduces the spatial and temporal stability of flower-visitor richness and pollination services in field crops, Garibaldi et al. (2011) synthesised data from 29 field studies. Spatial stability, temporal stability, mean richness, visitation and fruit set all decreased at 1 km from adjacent natural areas. However, *A. mellifera* visitation showed no change with isolation which the authors attributed to *A. mellifera* being less effected by landscape composition because it has larger foraging ranges than many solitary bees. There is still a need to maintain native habitat to conserve wild insects as pollinators, as visitation by wild insects and honey bees promotes fruit set independently and thus pollination by *A. mellifera* supplements rather than substitutes pollination by wild insects (Garibaldi et al., 2013).

Recent pollinator losses are not confined to *A. mellifera*, although studies into the declines of other pollinators are scarce. Many bumble bee species have declined in western Europe and North America, and these declines are largely attributed to agricultural intensification (Goulson et al., 2008). In addition to habitat loss, and potentially climate change, Oldroyd and Nanork (2009) have found that some Asian honey bee species are threatened by overhunting. Most of the evidence for the effects of climate change on insect pollinators comes from studies into butterflies with a few studies on bees (S. G. Potts et al., 2010). Further
studies are required, as well as ongoing monitoring programs to detect changes in the spatio-temporal distribution of pollinators. To detect small (2-5%) annual declines in the number of bee species and total abundance, Lebuhn et al. (2013) estimate that a sampling program with 200-250 sampling locations each sampled twice over 5 years would be required, at an estimated cost of USD$2 million. This seems like a small price to pay given the huge potential impacts of pollinator loss on global crop production.

It is currently assumed that for most Australian agroecosystems there are sufficient numbers of native and introduced pollinators to provide adequate pollination (Cunningham et al., 2002). However a combination of pesticide use and land clearing, which reduces nest site availability of feral honey bees and floral resources for native pollinators, may mean that there are now insufficient pollinators available in some agricultural systems (Cunningham et al., 2002, Schellhorn et al., 2008).

2.3 Honey bee behaviour

Honey bee colonies consist of an average of 27,000-45,000 bees (Seeley and Morse, 1976) made up of predominantly female workers, a smaller number of male drones and a single queen. The population of feral colonies tends to be smaller than that of managed colonies (Seeley and Morse, 1976). In managed colonies beekeepers try to minimise drone populations for reasons of productivity and pest control (Calderone, 2005); in feral colonies the number of drones produced in the spring prior to swarming ranges between 660 to 3960, with a mean of 2400 (Winston, 1987).
The honey bee colony collectively functions as a superorganism (Tautz, 2008), exhibiting the traits of mammals such as a low rate of reproduction, nourishment of offspring by a substance produced in special glands, a precisely controlled development environment for offspring, thermoregulation, and significant cognitive and learning abilities.

The honey bee nest is structured from multiple vertical panels of wax comb. The comb itself consists of a regular lattice of hexagonal cells; worker cells have a depth 13 mm (Goetz and Koeniger, 1992) and race-dependent diameters ranging from 4.6-5.4 mm. As drones are larger than workers, so too are drone cells with a cell diameter ranging from 6.0–6.4 mm across the different races (Winston, 1987, Hepburn and Radloff, 1998). Cells are used to store nectar (as honey), pollen, and to raise brood and the comb also functions as a mechanism to store and transmit information (Tautz, 2008). The cell area, comb volume and colony size are correlated.

Managed colonies of *Apis mellifera* are kept in bee hives with movable frames, primarily based on a design first proposed by Reverend L. L. Langstroth in 1851 (Winston, 1987). The Langstroth hive consists of a bottom board, a lid, and one or more boxes each having the capacity to hold either 8 or 10 frames. When a box is given over entirely to the storage of honey, it is referred to as a super. Descriptions of the equipment, tools and practices generally used in beekeeping are given in many texts (Bailey, 1982, Warhurst and Goebel, 2005, Delaplane, 2007, Root et al, 2007, Sammataro and Avitabile, 2011). A full super of honey
will generally weigh around 30kg, including 10kg of box and frame weight (Somerville, 2005). Commercial beekeepers in Australia generally fix four hives to a pallet to allow equipment such as forklifts and skid steer loaders to be used to lift hives on and off their trucks for transportation. Depending on the size of truck used, up to 360 hives on pallets will be transported on a single truck (Benecke, 2007).

While managed colonies are kept in hives managed by beekeepers, feral colonies choose to nest in cavities or hollows, either manmade or natural. In urban areas feral colonies will frequently build nests in or around buildings, and this sometimes leads them to be regarded as pests and exterminated (Baum et al., 2008, Ryan, 2010). In natural areas feral colonies will typically occupy tree hollows, typically with a volume of around 45 L (Seeley and Morse, 1976). In the case of eucalypts, tree hollows do not form until trees are at least 120-150 years old (Gibbons and Lindenmayer, 2002). Using nest boxes as a substitute for naturally formed hollows, Lindenmayer et al. (2009) found that in older forest (> 70 years old) only 2% of nest boxes were infested by honey bees whereas for young forest (20-year regrowth) the proportion was 33%. In contrast with the results of Seeley and Morse (1976), Lindenmayer et al. (2009) found that swarming colonies had a preference to take up residence in nest boxes of 19 L volume in preference to nest boxes of 38 L volume.

For the general populace, their encounters with honey bees are generally encounters with foraging workers. Worker larvae hatch from eggs laid by the queen 3 days earlier (Winston, 1987). The larvae develop in uncapped cells for 5-6 days before the cell is then capped by workers and
The pupae develops. The total development time from egg laying to emergence ranges from 16-24 days for workers, with an average of 21 days. Worker lifespans vary between seasons; in summer workers will typically live for 15-38 days whereas in winter average lifespans are 140 days. Over their lifetime, workers engage in many tasks at different ages including cleaning, brood and queen tending, comb building, food handling, ventilation, guard duty and foraging. Foraging is the last task performed by workers before their death, commencing at around day 23 of their lives. The phenomenon of recruitment has been well studied in honey bees, and nearly half a century has passed since von Frisch (1967) first decoded the waggle dance used by foraging workers to signal the location of floral resources. Workers typically forage when temperatures reach around 12 °C (Horskins and Turner, 1999), and during daylight hours only. Seeley (1995) found that 98% of sites foraged by workers over a consecutive 6-day period were within 4 km of the hive and Visscher and Seeley (1982) found that 95% of foraging was within 6 km (median 1.7 km), although distances of up to 10 km have been recorded (Tautz, 2008). Assuming foraging occurs (equally in all directions) these figures equate to a typical colony foraging area of 50 km² and maximum foraging area of >300 km². Workers will typically fly 800 km before dying; generally this occurs in 4-5 days (Winston, 1987).

The single most important individual in each colony is the queen. Each colony typically has only a single queen. The queen's role in the colony is to lay eggs; a healthy queen will typically lay 1000-2000 eggs per day; about the equivalent of her own bodyweight (Winston, 1987). The queen
stores sperm obtained from mating flights with drones in her sperm bank, and uses this to selectively fertilise eggs. Unfertilised eggs develop as drones. Fertilised eggs develop as workers. While queens can live for as long as 3 years (Winston, 1987), their productivity can drop as they age. Should the workers decide that the queen is no longer fit, they will raise a queen from a worker egg placed in a specially created "queen cell" by feeding the larvae a diet of "royal jelly" (Winston, 1987). This mechanism is typically responsible for queen supersedure in feral colonies. For managed colonies, beekeepers intervene directly when they observe falling colony productivity by requeening the colony with a new queen obtained from a specialised queen breeder. The process for requeening is discussed in many beekeeping texts (Bailey, 1982, Warhurst and Goebel, 2005, Delaplane, 2007, Root et al., 2007, Sammataro and Avitabile, 2011). The only time a queen leaves the hive is to mate or when swarming.

Drones exist only to mate with the queen and perform no other useful role within the colony (Winston, 1987). Most drones will however die before having the opportunity to mate, and those drones who do mate with the queen die immediately after mating (Winston, 1987). A week after hatching a new drone will become sexually mature (Tautz, 2008). The drone will then leave the hive in the afternoon to fly to a drone congregation area (DCA); the area where matings with the queen typically occur (Loper et al., 1987, Loper et al., 1992). DCAs have been observed to form in areas of the landscape that contain clear visual features (Tautz, 2008). DCAs persist in the same location for year after year (Loper et al., 1987, Loper et al., 1992, Tautz, 2008). Ruttner (1974)
found that drones from a colony newly introduced into the flight area behave in a manner similar to existing colonies, even on the first day after being introduced. The DCA itself will range in size from 30-200 m across (Tautz, 2008), at an elevation of 12-50 m above ground (Loper et al., 1987, Loper et al., 1992).

Upon leaving the hive, drones have been observed to prefer nearer DCAs over DCAs further away (N. Koeniger et al., 2005). Taylor and Rowell (1988) found that the relationship between the number of drones and the distance from the hive can be approximated by a bivariate normal distribution with a standard deviation ranging from 800-1500 m. However, N. Koeniger et al. (2005) did not find any direct correlation between flight distances and the ratio of drones from each colony within range of the DCA. Ruttner and Ruttner (1972) observed that drones will fly as far as 7 km from the hive to reach a DCA, crossing altitudes of 800-1000 m. The average distance travelled by drones was estimated by Ruttner (1974) as 3 km, with drones at a DCA being drawn from an area of 50-100 km². This contrasts with the results of Rowell et al. (1992), who found that 80% of drones are recaptured at distances of less than 2 km from their hive. Nikolaus Koeniger et al. (2005) estimate the average number of drones present at a DCA as around 12,000. Using genetic techniques, Baudry et al. (1998) established that 238 different colonies were represented at a single DCA, and postulated that most colonies within range of the DCA contributed drones to the congregation with an equal probability.
Drones at DCAs can be captured in a variety of aerial traps. Designs are described by Taylor (1984) and Williams (1987). All work on similar principles of elevating a net several metres off the ground, and luring drones to the net by way of caged queens or pheromone impregnated lures suspended below the net. Williams (1987) found that traps consistently caught 150-300 drones per 3 minute trapping period.

When a virgin queen is about a week old, she will leave the hive on multiple mating flights to the DCA where she will mate with the waiting drones. The queen is typically escorted on these mating flights by a group of workers (Tautz, 2008). The queen manufactures pheromones in her mandibular gland of which the major component is (2E)-9-oxodecenoic acid (9-ODA) (Brockmann et al., 2006), which can be readily synthesised (Milte et al., 2012). The pheromone plume emanating from the queen in flight attracts drones within an area of approximately 100 m in diameter (Winston, 1987, Brockmann et al., 2006). The total number of matings by each queen has been estimated to range from 7-20 (Estoup et al., 1994, Palmer and Oldroyd, 2000, Jensen et al., 2005). In observing 2434 flights of 628 queens, Woyke (1964) observed that 63% took a second mating flight (38% mating) and 8% a third mating flight (6% mating). Similar results were obtained by Tarpy and Page (2000) with 8 out of 30 (27%) trying to take a second flight and Schlüns et al. (2005) with 10 out of 18 (56%) making the attempt. The number of flights taken is thought to be dependent on the number of matings achieved in previous nuptial flights (Schlüns et al., 2005). The evolution of this multiple mating behaviour in honey bees, and in members of Hymenoptera more generally, has been
extensively studied but no hypothesis can explain why more than 10 matings frequently occur (Schlüns *et al.*, 2005).

It has been speculated that the queens fly further than drones from the same colony to avoid inbreeding (Taylor and Rowell, 1988, Rowell *et al.*, 1992). Ruttner and Ruttner (1972) found that the average distance between the queen’s colony and the mating location was more than 2 km. In over a third of cases mating occurred between 5-7 km from the hive, and in several cases over 7 km. In an experiment where the maximum mating distance recordable was constrained by the design of the experiment to be 3.22 km, Taylor and Rowell (1988) found that the distribution of minimal queen flight distances was left-skewed, with a mean distance of 899 m. Mating was recorded at the maximum measureable mating distance. Jensen *et al.* (2005) found that 90% of matings occurred within a range of 7.5 km, and 50% within 2.5 km. Estimates of maximum mating distances range from 12 km (Ruttner and Ruttner, 1972) to 15 km (Jensen *et al.*, 2005).

Reproduction at the colony level is achieved through the phenomenon of swarming. Prior to swarming occurring, worker bees begin queen rearing in response to stimuli including resource abundance, colony size, brood nest congestion, worker age distribution and reduced transmission of queen pheromones (Winston, 1987). A few days before the new queen (or queens) emerge, the old queen will leave the nest with about 70% of the worker bees (Tautz, 2008). Swarm dispersal distances range from a few hundred metres to 10 km (Baum *et al.*, 2005). Swarms send out several hundred scouts to search for and evaluate potential nest sites.
More than a dozen nest sites are typically discovered, and the colony collectively chooses one of these via a mechanism described by Seeley, 2010) before collectively flying to the chosen nest location. A similar phenomenon is that of absconding, where the entire colony leaves its nest and moves to a new location; absconding is more prevalent amongst Africanised honey bees (Winston, 1987).

Aside from the act of mating, there are three other situations in which honey bees from different colonies may interact. The first is during foraging, where bees from multiple colonies will often be present on a single flowering tree or shrub. The second type of interaction arises usually when resources are scarce, and foraging workers are motivated to rob from other hives (Winston, 1987). The final situation occurs when foraging workers returning from the field become disoriented and enter a hive other than their own. This situation seldom arises in feral colonies due to their separation distances, but occurs frequently in managed apiaries (Fries and Camazine, 2001).

2.4 Honey bee genetics

As drones developed from unfertilised eggs, they have only a single set of chromosomes which is referred to as the haploid chromosome state. Workers and queens develop from fertilised eggs; they have two sets of chromosomes, referred to as the diploid chromosome state (Tautz, 2008).

Particular genes localised at the same place (locus) in particular organism and which influence the character are called alleles (Tautz, 2008). At the locus for the gene for sex determination, if an individual is heterozygous (having different alleles), the individual becomes females. If the
individual is homozygous (identical alleles) they become male. This is the case for all haploid individuals as they possess only a single allele at the locus but also gives rise to the possibility of a diploid drone. Diploid drones are not competitive with haploid drones and are eaten by workers as young larvae (Winston, 1987). Thus any mechanism increasing heterozygosity at this locus is favoured by natural selection, and this includes multiple matings (Winston, 1987).

The parentage of individuals can be inferred from a sample subjected to genetic analysis at selected marker loci. If loci are chosen to be highly polymorphic, with many potential alleles for each locus, then each queen in a population is likely to carry unique haplotypes (Shaibi et al., 2008). Additionally, if multiple such marker loci that are tightly linked are chosen, then recombination events between them will be rare or absent. Finally, by choosing markers linked to genes under different modes of selection, additional information can be obtained (Shaibi et al., 2008).

The sex locus and thelytoky locus satisfy all of the above criteria, and have therefore been used in several parentage inference studies (Moritz et al., 2008, Shaibi et al., 2008, Jaffé et al., 2009, Jaffé et al., 2010, Jaffé and Moritz, 2010, Oxley et al., 2010, Oldroyd et al., 2011, Yañez et al., 2012, Moritz et al., 2013). A queen heterozygous at two sets of linked microsatellites will produce haploid drones with one of four genotypes (i.e., the four possible combinations of her two pairs of haploid genotypes, or haplotypes). In a sample of drones from a DCA, these genotypes may occur between one and four times. Some colonies contributing drones to the DCA may not be represented in the sample at all (Baudry et al., 1998).
Neither the scenario where every colony is represented by exactly one genotype in the sample, or represented by all four genotypes is very likely. An estimate on the minimum number of colonies contributing drones can thus be made by taking the linkage group with the maximum number of drone haplotypes and dividing this number by two. For a brood sample, the paternal haplotype for each worker, derived from the drone with which the test queen mated, is first determined by subtraction of the maternal haplotype (Oldroyd et al., 1996, Oldroyd and Wongsiri, 2006).

An alternative method for parentage inference relies on using computational maximum likelihood methods. In the case of drone samples, alleles at each loci within a linkage group are concatenated as a pseudo-haplotype (Moritz et al., 2007, Jaffé et al., 2010, Moritz et al., 2013). With brood samples the queen genotype is deduced manually based on allelic frequencies, and the resulting complementary drone haplotypes are then concatenated as a pseudo-haplotype (Moritz et al., 2007, Jaffé et al., 2010). The pseudo-haplotypes are then analysed by inputting them as haplotypes into COLONY computer program (Jones and Wang, 2010). While this is a pragmatic solution using available tools, it is still not optimal. Errors or missing data at a single loci are unable to be corrected. When using brood samples, using COLONY optimally requires that identical loci are used for genotyping of brood from multiple queens in order to maximise the number of structurally similar pseudo-haplotypes. Finally, inference of the maternal genotype from the brood samples introduces a further (albeit small) source of potential error,
although this error could be corrected for by sampling the queen directly. It remains computationally challenging to incorporate linkage in the inference of sibship/parentage relationships (Sieberts et al., 2002), and no software currently exists for this specialised purpose.

2.5 Managed honey bees in Australia

2.5.1 Honey bees in Australia

Australia has at least 1647 species of bees, with an unusually high number of species in the family Colletidae but no native members of the genus *Apis* (Batley and Hogendoorn, 2009). As outlined in Section 2.2.1, *A. mellifera* was first introduced to Australia from England in April 1822 (Barrett, 1995). Early settlers had found previously that the crops they had brought with them from England were not being satisfactorily pollinated and sought to remedy this by importing *A. m. mellifera* from their homelands. Within the first spring these colonies are recorded to have swarmed and started to establish themselves in the Australian bush as a feral population (Barrett, 2006). Bumblebees, *Bombus terretris*, were also introduced into Tasmania in 1992 (Kingston and McQuillan, 1998, Kingston et al., 2002), but have not yet spread to the mainland.

2.5.2 The Australian honey bee industry

As of 2006-2007, the Australian honey bee industry comprised 10,000 registered beekeepers operating a total of 572,000 hives (Crooks, 2008). Many apiarist's are second or third generation beekeepers, with knowledge of flowering patterns passing from generation to generation (Birtchnell and Gibson, 2008). More than 90% of Australia's total honey
and related product production is attributable to the 1700 beekeepers operating more than 50 hives (Crooks, 2008). More than 90% of beekeepers are based in New South Wales, Queensland, Victoria or South Australia (Crooks, 2008) and beekeepers in any one of these states may move hives between these states. There are also at least 600 beekeepers keeping stingless bee (primarily *Tetragonula carbonaria* and *Austroplebeia australis*) hives in Australia but the industry as a whole is still underdeveloped (Halcroft *et al.*, 2013) and is not further discussed in this thesis.

The total estimated value of honey and beeswax was estimated at AUD$75 million during 2007-2008 (Crooks, 2008), and this number appears small in comparison to the estimated AUD$1.7 billion value placed on honey bee pollination of Australian crops (Gordon and Davis, 2003). In Australia, the dominant honey processor, or "honey packer" is Capilano with 74.1% of the market at the end of 2014; supermarket private label brands accounted for another 9.1%, and Beechworth Honey around 7% (Capilano Honey Limited, 2015). Australia also exports live bees to other countries, although exports to the largest market in the US were halted in 2010 due to concerns about the introduction of viruses from Australian bees (Grant, 2014).

Around two thirds of commercial beekeepers use public land (e.g., state forests) for honey production and in 2006-2007 an estimated 37% of the total national honey crop was produced from public land (Crooks, 2008). The sites on public land are leased for a small annual fee and are closely held by beekeepers; public land site leases typically form part of the sale
of a beekeeping enterprise (Benecke, 2007). The remainder of sites are held on land either owned by beekeeper, or more commonly on land held by farmers/graziers on which the beekeeper has permission to place his or her hives. In general, it is only government-leased apiary sites whose locations are known, and no further information is readily available on the history of usage of these sites. In Victoria, data on the location of the government leased apiary sites is available publicly (Department of Environment and Primary Industries, 2013a) and these sites are shown against the background of defined bioregions in Figure 1. For each load of bees (roughly between 100 and 300 hives), each commercial beekeeper must keep between 8 and 12 sites spread over a large geographic area permanently booked (Benecke, 2007).

Figure 1: Apiary sites in Victoria shown in comparison to bioregions
In Australia, commercial beekeepers practice migratory beekeeping, moving their hives on a regular basis to access flowering events. During 2008, a surveyed sample of 147 beekeepers moving their hives in eastern Australia visited 488 locations, of which 288 locations were joined in an extended network spreading from western Victoria to central Queensland (Gordon et al., 2014). Illustrative of the hive moments undertaken by a migratory beekeeper, McDonald published a chronology of hive movements over two seasons as part of the educational material distributed on the 2007 Apimondia Technical Tour (McDonald, 2007). The McDonalds operate approximately 2000 hives, and are based in Castlemaine, Victoria. The full seasonal summary is shown in Table 3 with sites marked on the map in Figure 2.
Table 3: Example of two seasons' activities for a Victorian migratory beekeeper (McDonald, 2007)

<table>
<thead>
<tr>
<th>ID</th>
<th>Time period</th>
<th>Location</th>
<th>Attractor</th>
<th>Yield (tonnes)</th>
<th>Distance (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Winter</td>
<td>Big Desert</td>
<td>Desert banksia (Banksia ornata)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>Late winter/early spring</td>
<td>Robinvale</td>
<td>Almonds (Prunus amygdalus)</td>
<td>N/A</td>
<td>205</td>
</tr>
<tr>
<td>3</td>
<td>Late August</td>
<td>Carwarp</td>
<td>White mallee (Eucalyptus gracilis)</td>
<td>30</td>
<td>132</td>
</tr>
<tr>
<td>4</td>
<td>Late October</td>
<td>Underbool</td>
<td>Giant angular mallee (Eucalyptus incrassata)</td>
<td>70</td>
<td>175</td>
</tr>
<tr>
<td>5</td>
<td>Mid December</td>
<td>Kyalite</td>
<td>River red gum (Eucalyptus camaldulensis)</td>
<td>65</td>
<td>220</td>
</tr>
<tr>
<td>6</td>
<td>Mid January</td>
<td>Castlemaine</td>
<td>Yellow box (Eucalyptus melliodora)</td>
<td>30</td>
<td>325</td>
</tr>
<tr>
<td></td>
<td>March</td>
<td>Castlemaine</td>
<td>Red stringybark (Eucalyptus macroryncha)</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Mid April</td>
<td>Northwest Mallee</td>
<td>Winter weeds including wild turnip (Brassica rapa var. oleifera)</td>
<td>N/A</td>
<td>416</td>
</tr>
<tr>
<td>8</td>
<td>Late winter/early spring</td>
<td>Robinvale</td>
<td>Almonds (Prunus amygdalus)</td>
<td>N/A</td>
<td>130</td>
</tr>
<tr>
<td>9</td>
<td>Late August</td>
<td>Hay</td>
<td>Canola (Brassica napus), Paterson's curse (Echium plantagineum)</td>
<td>35</td>
<td>320</td>
</tr>
<tr>
<td>10</td>
<td>Mid/late October</td>
<td>Corowa</td>
<td>Paterson's curse (Echium plantagineum)</td>
<td>30</td>
<td>284</td>
</tr>
<tr>
<td>11</td>
<td>Late December</td>
<td>Sale</td>
<td>Saw banksia (Banksia serrata)</td>
<td>60</td>
<td>502</td>
</tr>
<tr>
<td>12</td>
<td>Mid February</td>
<td>Castlemaine</td>
<td>Long-leaf box (Eucalyptus goniocalyx)</td>
<td>30</td>
<td>376</td>
</tr>
<tr>
<td>13</td>
<td>Early April</td>
<td>Wedderburn</td>
<td>Blue mallee (Eucalyptus polybractea)</td>
<td>20</td>
<td>101</td>
</tr>
<tr>
<td>14</td>
<td>Early May</td>
<td>Big Desert</td>
<td>Desert banksia (Banksia ornata)</td>
<td>N/A</td>
<td>386</td>
</tr>
</tbody>
</table>

Total: 1,473

Second season

<table>
<thead>
<tr>
<th>ID</th>
<th>Time period</th>
<th>Location</th>
<th>Attractor</th>
<th>Yield (tonnes)</th>
<th>Distance (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Late winter/early spring</td>
<td>Robinvale</td>
<td>Almonds (Prunus amygdalus)</td>
<td>N/A</td>
<td>130</td>
</tr>
<tr>
<td>9</td>
<td>Late August</td>
<td>Hay</td>
<td>Canola (Brassica napus), Paterson's curse (Echium plantagineum)</td>
<td>35</td>
<td>320</td>
</tr>
<tr>
<td>10</td>
<td>Mid/late October</td>
<td>Corowa</td>
<td>Paterson's curse (Echium plantagineum)</td>
<td>30</td>
<td>284</td>
</tr>
<tr>
<td>11</td>
<td>Late December</td>
<td>Sale</td>
<td>Saw banksia (Banksia serrata)</td>
<td>60</td>
<td>502</td>
</tr>
<tr>
<td>12</td>
<td>Mid February</td>
<td>Castlemaine</td>
<td>Long-leaf box (Eucalyptus goniocalyx)</td>
<td>30</td>
<td>376</td>
</tr>
<tr>
<td>13</td>
<td>Early April</td>
<td>Wedderburn</td>
<td>Blue mallee (Eucalyptus polybractea)</td>
<td>20</td>
<td>101</td>
</tr>
<tr>
<td>14</td>
<td>Early May</td>
<td>Big Desert</td>
<td>Desert banksia (Banksia ornata)</td>
<td>N/A</td>
<td>386</td>
</tr>
</tbody>
</table>

Total: 2,099
The stated distances for hive movements considerably understate the distance travelled by McDonald and his employees; each trip for them commences and concludes from their home base at Castlemaine. Additionally, the seasonal activities do not list the reconnaissance trips undertaken to plan future hive movements. The chronology shows considerable variation in flowering events between seasons, with only a few events such as almond pollination occurring in a consistent location at a consistent time each year.

While most state governments view the movement of hives as a movement of livestock, there is no mandatory reporting of hive movements. Movement of livestock such as cattle, sheep and goats is recorded through the National Livestock Identification System (Meat & Livestock Australia, 2015). No such system exists for beekeepers, and
while an industry code is being drafted that would require beekeepers to maintain records of hive movements (Australian Honey Bee Industry Council, 2015) there would be no auditing and mandatory reporting on these.

With commercial beekeepers in Australia practicing migratory beekeeping, it is unsurprising that along with labour and insurance costs the largest items of expenditure in their operations are fuel, oil and grease, and vehicle repairs (Crooks, 2008).

2.5.3 Honey and pollen flora in Australia

The large majority of honey produced in Australia comes from eucalypts (Somerville, 2005). There are approximately 800 species of eucalypts in Australia, with 138 species occurring naturally in Victoria (Nicolle, 2006). Almost all natural landscapes in Victoria are dominated by eucalypts (Nicolle, 2006). In the wetter parts of the Great Dividing Range, forests of *E. regnans* (mountain ash) occur (Nicolle, 2006). With specimens exceeding 90 m in height, and reliable records of a 114 m tall specimen (Mifsud, 2003), *E. regnans* is the tallest flowering plant in the world (Hickey *et al.*, 2000), the tallest hardwood (Mifsud, 2003) and one of the tallest plants. It is also a source of strongly flavoured, dark honey and abundantly yielded pollen for honey bees (Goodman, 1973). In contrast to the wet hilly forests of eastern Victoria, the north-west of Victoria is hot, dry and flat. Multi-stemmed, shrub like mallee gums dominate the landscape of low heath and sand dunes, and several species are important sources of nectar and pollen for honey bees (Goodman, 1973).
Based on production data for Victorian beekeepers provided by Capilano Honey, the top-producing eucalypt species (in order) for the period from July 2009 to March 2014 were *E. camaldulensis*, *E. microcarpa*, *E. leucoxylon*, *E. melliodora* (Winner, 2014). The same ordering of species was found in the survey of Victorian commercial apiarists with more than 30 years of experience conducted by Birtchnell and Gibson (2006, 2008). Both sources also agree that additional key eucalypt species include mallee species, *E. tricarpa*, *E. macrorhyncha*, and *E. obliqua*.

The most important non-eucalypt species are canola (*Brassica napus*), Paterson’s curse (*Echium plantagineum*) and cape weed (*Arctotheca calendula*) (Birtchnell and Gibson, 2008). While significant volumes of canola and Paterson’s curse honey are produced (Winner, 2014), the main value of cape weed to beekeepers is as a pollen source. Canola is grown as crop, with 24,800 km$^2$ planted in 2014/2015 nationally including 4000 km$^2$ in Victoria (AOF, 2014). Canola is grown differently in Australia compared with most other countries; it is generally sown in autumn and then harvested in late spring (Walton et al., 1999). For this reason, it is an important source of nectar for honey bees to help build strength heading in the traditional warmer weather production season. *E. plantagineum* is regarded as a significant environmental weed in most Australian states (The University of Queensland, 2011b), and *A. calendula* is regarded as an environmental weed in most states (The University of Queensland, 2011a).
2.5.4 Crop pollination in Australia

Paid pollination in Australia is currently very small compared with countries such as the US, with only 200,000 hives (35% of all managed hives and 28% of beekeepers) used for paid pollination (Crooks, 2008, Monck, 2008). Additionally, a further 80,000 to 100,000 hives are used for pollination on a mutually beneficial basis. In the US, 1,725,070 hives (71%) are used to pollinate almonds in California alone (vanEngelsdorp et al., 2008, Carman, 2011), noting that California does account for 80% of the world's almond production. Australian native bees are currently being assessed, and to a very small extent used for pollination services, especially for greenhouse crops such as tomatoes which require buzz pollination unable to be provided by A. mellifera. (Hogendoorn et al., 2006, Halcroft et al., 2013).

2.5.5 Honey bee colony losses in Australia

Unlike most other westernised beekeeping nations, Australia does not monitor annual managed honey bee colony losses. In part, this is because Australia has not experienced Colony Collapse Disorder. Another factor is that migratory beekeepers tend to move their hives to warmer inland areas where overwintering conditions are relatively mild. However, colony losses have been experienced recently in a series of poor honey seasons. The summer of 2011-2012 was reported as the worst season in 30 years Western Australia (Diss, 2011). The summer of 2012-2013 was reported as the worst season in 50 years in Victoria (Peake, 2013). The following year, summer 2013-2014 was the worst season in 20 years in Queensland (Foster, 2014). Most recently, summer 2014-2015 was the
worst season in 50 years in Tasmania (Humpries, 2015). Excluding the possibility of exaggeration in the media coverage, these claims raise serious concerns about why there appears to be an increased frequency of extraordinarily poor honey producing seasons, the effects this has on the viability of the beekeeping industry and the flow-on impacts on pollination as well as the impact on other species which depend on nectar and pollen for survival.

2.6 Feral honey bees

Feral honey bees are colonies of *A. mellifera* which have naturalised in an environment. Within the natural range of *A. mellifera* in Europe and in Africa, naturalised colonies may be referred to as wild colonies. Feral colonies of *A. mellifera* exist across all temperate areas of the world, and were typically established through the escape of swarms from managed colonies. This mechanism continues to replenish feral populations in many regions, although under more adverse climatic conditions modern races of managed honey bees have been shown to exhibit poor survival rates as feral populations (Oldroyd *et al.*, 1997). Feral honey bees colonies are present throughout Australia, and populations have been shown to be self-sustaining (Oldroyd *et al.*, 1997) without requiring new swarms from managed colonies. There is evidence that by the mid-1800s feral populations were already established throughout much of the forest in Australia, providing food, income and a source of apiary stock for European settlers (Barrett, 1999).
2.6.1 Feral honey bee impacts

In many natural environments, feral honey bees are thought to exert a range of harmful influences. Feral honey bees have been shown to compete with native wildlife for tree hollows, including birds (Oldroyd et al., 1994, Hansen et al., 2002) and mammals (Suckling and Goldstraw, 1989, Wood and Wallis, 1998a, Wood and Wallis, 1998b, Goldingay et al., 2007, Lindenmayer et al., 2009). Competition for nectar, pollen and in some cases water has also been observed (Schaffer et al., 1979, Schaffer et al., 1983, Roubik et al., 1986, Markwell et al., 1993, Kato et al., 1999, Gross, 2001, Dupont et al., 2004, Paini and Roberts, 2005) but negative effects on native pollinators do not always occur (Paton, 1999, Roubik and Wolda, 2001) and some studies into negative effects have been found to be poorly designed (Paini, 2004). Certain introduced weeds are thought to be pollinated by feral honey bees, leading to an increased spread (Kato et al., 1999, Barthell et al., 2001, Goulson and Derwent, 2004, Simpson et al., 2005). While finding that honey bees are important pollinators of some weeds, Butz Huryn and Moller (1995) concluded that they probably did not contribute substantially to weed problems. Conversely, some native plants are thought to be inadequately pollinated by feral honey bees, threatening their ability to propagate (Taylor and Whelan, 1988, Vaughton, 1996, Paton, 1997, Gross and Mackay, 1998).

From a biosecurity perspective, feral honey bees have been shown to perform a role in the transmission of pests and diseases, and to serve as reservoirs for reinfection of treated managed colonies (Kraus and Page, 1995). However, feral populations are also themselves impacted by the
introduction of pests and diseases. Evidence from other countries suggests that established feral colonies numbers are reduced by 75% to 90% following the introduction of *Varroa destructor*, an external parasitic mite of honey bees further discussed in Section 2.7 (Kraus and Page, 1995, Somerville, 2008, Thompson *et al.*, 2014). The possibility of eradicating feral populations as a biosecurity measure has been examined in several studies (Oldroyd, 1998, Taylor and Goodwin, 2001, Taylor *et al.*, 2007). A combination of manual extermination and deliberate introduction of varroa has been successfully used to eradicate all feral honey bee colonies on Santa Cruz Island off the coast of California (Wenner *et al.*, 2008).

Not all effects of feral honey bees are negative. As part of their recommendations for a nationally coordinated effort to address the issue of honey bee declines, Nabhan *et al.* (1998) recommended monitoring changes in the density of feral honey bee colonies at fixed sampling regions throughout North America. They also emphasised a need to better understand the pollination efficiency and value of (feral) Africanised honey bees as major crop pollinators for cultivated plants in North America. Feral colonies have been identified as providing essential pollination for many crops (Gordon and Davis, 2003). They also serve as important reservoir of genetic diversity (Daly *et al.*, 1991, Chapman *et al.*, 2008) and exhibit adaptations not seen in domestic colonies (Atmowidjojo *et al.*, 1997).
2.6.2 Surveying feral colony densities

It is impossible to gauge the effects of feral populations, both harmful and beneficial, without an adequate understanding of the distribution of these populations (Moritz et al., 2007). The distribution and abundance of all species varies across space and time in accordance with a range of abiotic and biotic factors (Krebs, 2009). As outlined in Section 2.3, A. mellifera possesses a number of adaptations that enable it to exist across a large natural range. Numerous studies have been conducted into the density and distribution of unmanaged A. mellifera colonies globally, and these are summarised in Table 4. Generally, a density of 100-200 colonies/km² is required for effective pollination of fruit crops (Free, 1970). Only the highest density of feral honey bees ever recorded globally – 150 colonies/km² (Oldroyd et al., 1997) comes close to this. More typically feral colonies have a density of $1 \times 10^{-5}$ to $1 \times 10^{-3}$ colonies/km² (reviewed in Ratnieks et al., 1991).

For several reasons it would be useful to have good estimates of the densities of feral honey bee colonies in different habitats in Australia. In agricultural settings, it is necessary to know if the density of feral bee colonies is sufficient to provide adequate pollination. In conservation areas it is valuable to know if the density of feral colonies is sufficient to be of concern. A third reason is that should an incursion of Varroa destructor or other exotic parasite or pathogen occur in Australia it will be necessary to determine whether location and destruction of feral nests is a feasible method of containing the outbreak (Stevenson et al., 2005, Taylor et al., 2007).
There are a number of methods by which the density of feral colonies can be determined. The most conceptually straightforward method is a direct search of a survey area for feral colonies. Colonies can be located through observation of activity near suitable nest sites or an assisted search using beelining techniques (Galton, 1971, Taber III, 1979, Visscher and Seeley, 1982, Schneider and Blyther, 1988, Wenner, 1989, Ratnieks et al., 1991, Oldroyd et al., 1994, McNally and Schneider, 1996, Paton, 1996, Oldroyd et al., 1997, Goodman and Hepworth, 2004, Baum et al., 2005, Seeley, 2007, Vaudo et al., 2012, Oleksa et al., 2013). The challenge with the direct search method is that it is extremely time consuming (and hence costly), can typically only be performed for small areas accessible by foot, and is prone to non-detection errors (false negatives) as colonies are cryptic and difficult to locate (Oldroyd et al., 1997). An advantage of the direct search method is that managed colonies in the area are unlikely to be incorrectly recorded as feral colonies (false positives). A number of other studies provide a variation on direct search of feral colonies through the use of data on colony removals and extermination (Boreham and Roubik, 1987, Baum et al., 2008).

The main alternative to methods of direct search are those that rely on utilising genetic techniques to infer parentage and thus colony densities within a survey area (Baudry et al., 1998, Moritz et al., 2007, Moritz et al., 2008, Jaffé and Moritz, 2010, Moritz et al., 2013, Arundel et al., 2014, Hinson et al., 2015). The genetic basis of these techniques, which sample either drones from a DCA, or offspring of newly mated queen, is discussed in Section 2.4.
As discussed in Section 2.3, drones can be lured into an aerial trap at a DCA by using a bait impregnated with pheromones. To construct a Williams trap (Williams, 1987), blackened cigarette filters impregnated with 9-ODA are placed within a net with an opening at the bottom, and this is then attached to a weather balloon and raised to an altitude of between 8-40 m (Koeniger et al., 1989, Baudry et al., 1998). Flying drones are difficult to trap outside DCAs (Taylor, 1984). Thus to successfully sample drones in large numbers it is necessary to locate the trap within a DCA. This can be achieved through either previous observation of the DCA locations, or a search of the environment using the trap as a detector (Brockmann et al., 2006, Jaffé et al., 2009). There are usually several thousand drones present at a DCA, and this method can capture several hundred drones in a few minutes. If the sampling is done during the breeding season (in Australia, September to March), then virtually all colonies will be producing males and thus be represented in the male pool of potential mates (Arundel et al., 2014). The method cannot sample weak colonies that are not producing drones, but these colonies are likely to die, and are therefore not ecologically significant. Similarly, the method will not sample nascent colonies which are yet to produce drones, and thus the estimates obtained from the data are conservative (Arundel et al., 2014). Typically, researchers genotype a sample of 96 drones from each site (96 being the number that conveniently fits in a standard plate used for PCR reactions) (Arundel et al., 2013). Surveys utilising drone capture to infer honey bee population densities are summarised in Table 4.
Rather than directly sampling the feral population through the capture of drones, it is instead possible to introduce one or more virgin queens into a survey area and allow the queens to mate with the indigenous drone population (Moritz et al., 2007, Moritz et al., 2008, Jaffé and Moritz, 2010, Arundel et al., 2014). Worker brood from each queen can then be sampled, the queen genotype inferred, and the parentage of the drones which mated with the queen inferred. Surveys utilising queen mating to infer honey bee population densities are summarised in Table 4.

Both techniques yield an estimate of the number of colonies contributing to the sample, and corrections may be applied to the estimate to adjust for non-detection errors based on the sample size. To interpret the count of colonies as a density typically requires assumptions to be made about the area from which the sample has been drawn. There have been several approaches used in previous studies. From the drone capture method, Jaffé and Moritz (2010), Moritz et al. (2007) and Moritz et al. (2008) all use an assumption derived from Taylor and Rowell (1988) that the average flight distance of drones is 900 m. This value was obtained in a study conducted on nearly treeless prairies in Kansas, USA. From this, they infer that the average area from which the sample is drawn is

\[ A = \pi r^2 = \pi (0.9)^2 = 2.5 \text{ km}^2 \]

and use this to calculate densities in a variety of landscape across Europe, Central Asia and Africa. Unfortunately this reflects a false degree of precision in the density estimate. The actual results of Taylor and Rowell (1988) are that the drone distribution can be approximated by a bivariate normal distribution with standard deviation of 800-1500 m. For a fixed colony
count, the density should therefore also take the form of a distribution. Different studies in different areas are likely to lead to differing conclusion on the distribution of drone flight distances to the DCA. Ruttner (1974) found an average drone flight distance of 3 km, equivalent to an area, \( A = \pi r^2 = \pi (3)^2 = 28.3 \text{ km}^2 \). Applying this value instead of that derived from Taylor and Rowell (1988) would lead to an order of magnitude smaller feral colony density estimate in the results of Jaffé and Moritz (2010), Moritz et al. (2007) and Moritz et al. (2008).

There are similar challenges in inferring the size of the area sampled in the queen mating methods. Moritz et al. (2007) make the assumption that queens fly as far as drones, and thus the sampling radius will be on average at least twice that used in the estimation of densities from drone samples \( A = \pi r^2 = \pi (2 \times 0.9)^2 = 10 \text{ km}^2 \). Unfortunately this assumption is also based on the flawed application of a result from a single survey as discussed earlier. In applying both the drone capture and queen mating techniques in a single survey area, Moritz et al. (2008) found that the number of colonies detected in the queen sample was 1.8 times greater than that in the drone sample. Jaffé et al. (2010) use the results of Moritz et al. (2008) to infer that the brood samples correspond to an area 1.8 times greater than that of the drone samples \( A = 1.8 \times \pi r^2 = 1.8 \times \pi (0.9)^2 = 4.5 \text{ km}^2 \). The inferred area is therefore based on a ratio observed one study and a drone flight distance from a single different study. On the basis that colonies further from the DCA contribute comparatively fewer drones and are therefore likely to be less represented in the sample, Jaffé et al. (2010) adjust their estimate of the
number of colonies to exclude the count of those colonies represented by less than a median number of drones. To apply such an adjustment correctly would also require modification of the drone flight distance distribution. Based on the results of (Ruttner and Ruttner, 1972), where over a third of matings occurred between 5-7 km from the hive, an alternative interpretation of the area being sampled by the mated queen is $A = \pi r^2 = \pi (7)^2 = 154 \text{ km}^2$, and using this alternative value would again lead to an order-of-magnitude decrease in the reported densities.

Irrespective of what value is used for estimating the area from which the sample is being drawn, another assumption inherent in this approach is that the number of colonies detected in the sample scales linearly with the colony density. At very low densities, it may be reasonable to expect that most, if not all, colonies present at the DCA are represented in the sample. As the density increases, this would require either an increasing number of drones to be sampled (in the case of the drone capture survey method) or a larger number of queens to be used (in the case of the queen mating survey method) in order to continue sampling a majority of colonies present. Guidelines for how many drones should be sampled or queens used are discussed in Section 4.

Two other forms of error arise in the application of genetics-based survey techniques. The uncertainty arising from the genetic analysis in the estimated count of colonies present in the sample should propagate through into the estimation of the density. Additionally, the large area from which the sample is drawn increases the likelihood of detecting managed colonies (false positives) and efforts need to be made to
establish beekeeper activity in the vicinity of the survey area (Jaffé and Moritz, 2010).

At a local scale, the distribution of colonies within an area has been shown to exhibit spatial aggregation (Oldroyd et al., 1995, McNally and Schneider, 1996, Oldroyd et al., 1997, Baum et al., 2005, Baum et al., 2011). The spatial point process and natural selection drivers leading to colony aggregations are not fully understood. Oldroyd et al. (1995) have proposed that existing colonies attract swarms and that new swarms do not travel far from their originating colony in environments providing suitable nest sites. Baum et al. (2005) suggest that the aggregations result from the distribution of resources, especially suitable nest sites. In a study of rural avenues, Oleksa et al. (2013) found no evidence of spatial aggregation.

While there are many studies addressing the spatial distribution and abundance of feral colonies, fewer studies focus on temporal variations. A small proportion of the overall studies from Table 4 involve studies conducted over multiple years or seasons (Taber III, 1979, Boreham and Roubik, 1987, Oldroyd et al., 1997, Baum et al., 2005, Baum et al., 2008). Some surveys do, however, revisit areas sampled in previous studies albeit often with differing survey techniques that make the direct comparison of results difficult (McNally and Schneider, 1996, Seeley, 2007, Moritz et al., 2013, Arundel et al., 2014, Hinson et al., 2015).
Table 4: Summary of all studies globally into the densities of feral *Apis mellifera* colonies, sorted by estimated density

<table>
<thead>
<tr>
<th>Region</th>
<th>Location</th>
<th>Climate</th>
<th>Habitat</th>
<th>Survey method</th>
<th>Survey area km²</th>
<th>Number of queens</th>
<th>Sample size</th>
<th>Number of colonies</th>
<th>Estimated density colonies km²</th>
<th>Multi-year survey</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia – Victoria</td>
<td>Whela State Forest</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Direct search</td>
<td>0.11</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
</tr>
<tr>
<td>Australia – Victoria</td>
<td>Black Mountain Flora Reserve</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Direct search</td>
<td>0.06</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
</tr>
<tr>
<td>Australia – Victoria</td>
<td>West of Murchison</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Direct search</td>
<td>0.06</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
</tr>
<tr>
<td>Australia – South Australia</td>
<td>Ngarkat Conservation Park</td>
<td>Temperate</td>
<td>Mallee heath</td>
<td>Direct search</td>
<td>36</td>
<td>N/A</td>
<td>N/A</td>
<td>4</td>
<td>0.1</td>
<td>No</td>
<td>Paton (1996)</td>
</tr>
<tr>
<td>Europe – Poland</td>
<td>Northern Poland</td>
<td>Temperate</td>
<td>Agricultural land</td>
<td>Direct search</td>
<td>10</td>
<td>N/A</td>
<td>N/A</td>
<td>45</td>
<td>0.1</td>
<td>No</td>
<td>Oleksa et al. (2013)</td>
</tr>
<tr>
<td>Australia – Victoria</td>
<td>Yaapeet</td>
<td>Semi-arid</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>0.15</td>
<td>N/A</td>
<td>14</td>
<td>6</td>
<td>0.15</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>Europe - Russia</td>
<td>Sergach, Nizhny Novgorod</td>
<td>Continental</td>
<td>Temperate forest</td>
<td>Direct search</td>
<td>72.5</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0.17</td>
<td>No</td>
<td>Galton (1971)</td>
</tr>
<tr>
<td>North America - California</td>
<td>Santa Cruz Island</td>
<td>Subtropical</td>
<td>Arid Island</td>
<td>Direct search</td>
<td>230</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0.25</td>
<td>No</td>
<td>Wenner (1989)</td>
</tr>
<tr>
<td>Australia – Victoria</td>
<td>Stratford Park</td>
<td>Semi-arid</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>0.25</td>
<td>N/A</td>
<td>123</td>
<td>14</td>
<td>0.25</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>Australia - NSW</td>
<td>Lake Albacutya</td>
<td>Temperate</td>
<td>Rural town/Agricultural land</td>
<td>Drone capture</td>
<td>0.25</td>
<td>N/A</td>
<td>29</td>
<td>9.5</td>
<td>0.25</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>Europe - Russia</td>
<td>Sergach, Nizhny Novgorod</td>
<td>Continental</td>
<td>Temperate forest</td>
<td>Direct search</td>
<td>10</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0.3</td>
<td>No</td>
<td>Galton (1971)</td>
</tr>
<tr>
<td>Australia – South Australia</td>
<td>Mt Rescue Conservation Park</td>
<td>Temperate</td>
<td>Mallee heath</td>
<td>Direct search</td>
<td>20</td>
<td>N/A</td>
<td>N/A</td>
<td>7</td>
<td>0.4</td>
<td>No</td>
<td>Paton (1996)</td>
</tr>
<tr>
<td>Australia - NSW</td>
<td>Gloucester Tops</td>
<td>Temperate</td>
<td>Dry sclerophyll forest</td>
<td>Drone capture</td>
<td>0.4</td>
<td>N/A</td>
<td>79</td>
<td>18</td>
<td>0.4</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>Australia - NSW</td>
<td>Tyagong</td>
<td>Temperate</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>0.45</td>
<td>N/A</td>
<td>74</td>
<td>18</td>
<td>0.45</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>North America - New York</td>
<td>Arnot Forest</td>
<td>Temperate</td>
<td>Mixed forest</td>
<td>Direct search</td>
<td>16.5</td>
<td>N/A</td>
<td>N/A</td>
<td>9</td>
<td>0.5</td>
<td>No</td>
<td>Visscher and Seeley (1982)</td>
</tr>
<tr>
<td>Australia - NSW</td>
<td>Ben Hall's Cave Camp</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Drone capture</td>
<td>0.55</td>
<td>N/A</td>
<td>62</td>
<td>19</td>
<td>0.55</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>Australia - NSW</td>
<td>Grenfell</td>
<td>Temperate</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>0.55</td>
<td>N/A</td>
<td>70</td>
<td>20.5</td>
<td>0.55</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>Europe - Russia</td>
<td>Sergach, Nizhny Novgorod</td>
<td>Continental</td>
<td>Temperate forest</td>
<td>Direct search</td>
<td>88</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0.57</td>
<td>No</td>
<td>Galton (1971)</td>
</tr>
<tr>
<td>Australia - NSW</td>
<td>Monkera</td>
<td>Temperate</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>0.6</td>
<td>N/A</td>
<td>70</td>
<td>21.5</td>
<td>0.6</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>Australia - NSW</td>
<td>Allsyn River</td>
<td>Temperate</td>
<td>Rainforest/wet sclerophyll</td>
<td>Drone capture</td>
<td>0.65</td>
<td>N/A</td>
<td>72</td>
<td>23.5</td>
<td>0.65</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
</tbody>
</table>

1 For genetics-based techniques, the survey area listed is the assumed area used for density calculations

2 Applicable for genetics-based techniques only, the sample size reflects either the number of drones (for the drone capture survey method) or worker brood (for the queen mating survey method) analysed

3 For genetics-based techniques this represents the estimated number of colonies inferred from the sample
<table>
<thead>
<tr>
<th>Region</th>
<th>Location</th>
<th>Climate</th>
<th>Habitat</th>
<th>Survey method</th>
<th>Survey area km$^2$</th>
<th>Number of queens</th>
<th>Sample size</th>
<th>Number of colonies</th>
<th>Estimated density colonies km$^2$</th>
<th>Multi-year survey</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe - Sweden</td>
<td>Gotland</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>2.5</td>
<td>23</td>
<td>2</td>
<td>0.8</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
<td></td>
</tr>
<tr>
<td>Australia - NSW</td>
<td>Holy Camp</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Drone capture</td>
<td>0.8</td>
<td>278</td>
<td>37.5</td>
<td>0.8</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
<td></td>
</tr>
<tr>
<td>Europe - Russia</td>
<td>Sergach, Nizhny Novgorod</td>
<td>Continental</td>
<td>Temperate forest</td>
<td>Direct search</td>
<td>10.36</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0.96</td>
<td>No</td>
<td>Gaiton (1971)</td>
</tr>
<tr>
<td>Africa – South Africa</td>
<td>Eastern Cape</td>
<td>Temperate</td>
<td>Thicket, savanna</td>
<td>Direct search, feeding station observations</td>
<td>Not specified</td>
<td>N/A</td>
<td>33</td>
<td>&lt; 0.1</td>
<td>No</td>
<td>Vaudo et al. (2012)</td>
<td></td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Wyperfeld National Park, Lake Brambruk</td>
<td>Temperate - Semi-arid</td>
<td>Riparian woodland</td>
<td>Drone capture</td>
<td>1</td>
<td>N/A</td>
<td>241</td>
<td>42</td>
<td>1</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>North America - Arizona, USA</td>
<td>Tucson</td>
<td>Arid</td>
<td>Urban/suburban/desert</td>
<td>Swarm removal data</td>
<td>897</td>
<td>N/A</td>
<td>323 - 1035</td>
<td>0.36 - 1.12</td>
<td>Yes</td>
<td>Baum et al. (2008)</td>
<td></td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Puckapunyal North and South</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Queen mating</td>
<td>N/A</td>
<td>3</td>
<td>144</td>
<td>12.34</td>
<td>0.1 - 1.5</td>
<td>No</td>
<td>Arundel et al. (2014)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Wyperfeld National Park, Black Flat</td>
<td>Temperate - Semi-arid</td>
<td>Riparian woodland</td>
<td>Drone capture</td>
<td>1.5</td>
<td>N/A</td>
<td>222</td>
<td>49</td>
<td>1.5</td>
<td>No</td>
<td>Hinson et al. (2015)</td>
</tr>
<tr>
<td>Europe - Germany</td>
<td>Schwarzenau</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>92</td>
<td>10</td>
<td>2.2</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe - Germany</td>
<td>Hochharz</td>
<td>Cool subarctic</td>
<td>Nature reserve</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>77</td>
<td>11</td>
<td>2.4</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe Germany</td>
<td>Hochharz</td>
<td>Cool subarctic</td>
<td>Nature reserve</td>
<td>Queen mating</td>
<td>10</td>
<td>7</td>
<td>Not specified</td>
<td>24</td>
<td>2.4</td>
<td>No</td>
<td>Moritz et al. (2007)</td>
</tr>
<tr>
<td>Central Asia - Tjurgen-Aka-Suu</td>
<td>Tjurgen-Ako-Su</td>
<td>Temperate semi-arid</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>62</td>
<td>12</td>
<td>2.6</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>North America - New York</td>
<td>Oswego</td>
<td>Temperate</td>
<td>Urban/suburban</td>
<td>Volunteered information</td>
<td>4.2</td>
<td>N/A</td>
<td>11</td>
<td>2.7</td>
<td>No</td>
<td>Morse et al. (1990)</td>
<td></td>
</tr>
<tr>
<td>Europe - Ireland</td>
<td>Caher, Republic of Ireland</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>7</td>
<td>2.8</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe Germany</td>
<td>Schwarzenau</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>10</td>
<td>5</td>
<td>Not specified</td>
<td>28</td>
<td>2.8</td>
<td>No</td>
<td>Moritz et al. (2007)</td>
</tr>
<tr>
<td>Central Asia - Kyrgyzstan</td>
<td>Dshuuka</td>
<td>Temperate semi-arid</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>67</td>
<td>14</td>
<td>3.1</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe Germany</td>
<td>Müritz</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>10</td>
<td>8</td>
<td>Not specified</td>
<td>32</td>
<td>3.2</td>
<td>No</td>
<td>Moritz et al. (2007)</td>
</tr>
<tr>
<td>Europe - Germany</td>
<td>Müritz</td>
<td>Wet warm oceanic</td>
<td>Nature reserve</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>101</td>
<td>16</td>
<td>3.5</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe - Ireland</td>
<td>Belfast, Northern Ireland</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>2.5</td>
<td>96</td>
<td>10</td>
<td>3.9</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
<td></td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Dookie</td>
<td>Temperate</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>N/A</td>
<td>192</td>
<td>14.51</td>
<td>0.1 - 4</td>
<td>No</td>
<td>Arundel et al. (2014)</td>
<td></td>
</tr>
<tr>
<td>Africa - Sudan</td>
<td>Khartoum</td>
<td>Hot desert</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>58</td>
<td>19</td>
<td>4.2</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Africa - Sudan</td>
<td>El-Faw</td>
<td>Hot desert</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>56</td>
<td>19</td>
<td>4.2</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe - Spain</td>
<td>Caravaca de la Cruz</td>
<td>Temperate semi-arid</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>233</td>
<td>22</td>
<td>4.9</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Region</td>
<td>Location</td>
<td>Climate</td>
<td>Habitat</td>
<td>Survey method</td>
<td>Survey area km&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Number of queens</td>
<td>Sample size</td>
<td>Number of colonies</td>
<td>Estimated density colonies km&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Multi-year survey</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>---------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-------------------------------------------</td>
<td>-------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Marysville</td>
<td>Temperate</td>
<td>Burnt woodland</td>
<td>Queen mating</td>
<td>N/A</td>
<td>4</td>
<td>192</td>
<td>13-52</td>
<td>0.1 - 5</td>
<td>No</td>
<td>Arundel et al. (2014)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Chiapas, Huixtla Rd</td>
<td>Tropical</td>
<td>Agricultural land</td>
<td>Volunteered information</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>5</td>
<td>No</td>
<td>Ratnieks et al. (1991)</td>
</tr>
<tr>
<td>North America - Arizona</td>
<td>Cottonwood</td>
<td>Arid</td>
<td>Semi-desert canyon</td>
<td>Direct search</td>
<td>3.1</td>
<td>N/A</td>
<td>N/A</td>
<td>Not known</td>
<td>2.9 - 5.1</td>
<td>Yes</td>
<td>Taber III (1979)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Tswalu</td>
<td>Hot desert</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>148</td>
<td>14.5</td>
<td>5.7</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe - Croatia</td>
<td>Pokupsko</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>15</td>
<td>5.9</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Chiaspas, San Rosque</td>
<td>Tropical</td>
<td>Agricultural land</td>
<td>Direct search and volunteered information</td>
<td>2.1</td>
<td>N/A</td>
<td>30</td>
<td>6.2</td>
<td>No</td>
<td>Ratnieks et al. (1991)</td>
<td></td>
</tr>
<tr>
<td>Africa - Libya</td>
<td>Al Kufrah</td>
<td>Hot desert</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>4.5</td>
<td>183</td>
<td>29</td>
<td>6.4</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
<td></td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Eisenburg</td>
<td>Hot Mediterranean</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>95</td>
<td>17</td>
<td>6.7</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Leeuwfontein</td>
<td>Dry winters-humid subtropical</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>17</td>
<td>6.7</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe - Germany</td>
<td>Halle</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>95</td>
<td>17</td>
<td>6.7</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe - Croatia</td>
<td>Svinnjucko</td>
<td>Wet warm oceanic</td>
<td>Agricultural land</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>17</td>
<td>6.7</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Chiapas, San Rosque</td>
<td>Tropical</td>
<td>Various</td>
<td>Swarm removal data</td>
<td>50</td>
<td>N/A</td>
<td>N/A</td>
<td>4.7 - 7.1</td>
<td>Yes</td>
<td>Boreham and Roubik (1987)</td>
<td></td>
</tr>
<tr>
<td>Europe - Italy</td>
<td>Internmesoli</td>
<td>Hot Mediterranean</td>
<td>Nature reserve</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>108</td>
<td>33</td>
<td>7.3</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Europe - Italy</td>
<td>Elba</td>
<td>Hot Mediterranean</td>
<td>Agricultural land</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>102</td>
<td>34.5</td>
<td>7.6</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Africa - Botswana</td>
<td>Okavango Delta</td>
<td>Subtropical</td>
<td>Desert/savannah</td>
<td>Direct search</td>
<td>~50</td>
<td>N/A</td>
<td>89</td>
<td>4.2 - 7.8</td>
<td>No</td>
<td>McNally and Schneider (1996)</td>
<td></td>
</tr>
<tr>
<td>Africa - Botswana</td>
<td>Okavango Delta</td>
<td>Subtropical</td>
<td>Desert/savannah</td>
<td>Honeyguide bird and direct survey</td>
<td>6</td>
<td>N/A</td>
<td>47</td>
<td>7.8</td>
<td>No</td>
<td>Schneider and Blyther (1988)</td>
<td></td>
</tr>
<tr>
<td>Europe - Italy</td>
<td>Farindola</td>
<td>Hot Mediterranean</td>
<td>Nature reserve</td>
<td>Queen mating</td>
<td>4.5</td>
<td>Not specified</td>
<td>110</td>
<td>37</td>
<td>8.2</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Wellsford State Forest</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Direct search</td>
<td>0.12</td>
<td>N/A</td>
<td>1</td>
<td>8.3</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
<td></td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Chiapas, Río Florido</td>
<td>Tropical</td>
<td>Agricultural land</td>
<td>Volunteered information</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>9</td>
<td>No</td>
<td>Ratnieks et al. (1991)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Suikerbosrand</td>
<td>Dry winters-warm oceanic</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>191</td>
<td>25</td>
<td>9</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Ezemvelo</td>
<td>Dry winters-humid subtropical</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>24</td>
<td>9.4</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Pietermaritzburg</td>
<td>Dry winters-warm oceanic</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>25</td>
<td>9.8</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Eildon</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Queen mating</td>
<td>4</td>
<td>192</td>
<td>15-59</td>
<td>0.1 - 10</td>
<td>No</td>
<td>Arundel et al. (2014)</td>
<td></td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Puckapunyal East</td>
<td>Temperate</td>
<td>Woodland</td>
<td>Queen mating</td>
<td>4</td>
<td>192</td>
<td>18-58</td>
<td>0.15 - 10</td>
<td>No</td>
<td>Arundel et al. (2014)</td>
<td></td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Yaapeet</td>
<td>Temperate-Semi-arid Agricultural land</td>
<td>Queen mating</td>
<td>5</td>
<td>240</td>
<td>25-68</td>
<td>0.15 - 10</td>
<td>No</td>
<td>Arundel et al. (2014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Wyperfeld</td>
<td>Temperate-Semi-arid Riparian woodland</td>
<td>Queen mating</td>
<td>5</td>
<td>240</td>
<td>29-73</td>
<td>0.2 - 10</td>
<td>No</td>
<td>Arundel et al. (2014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Jonkershoek</td>
<td>Hot Mediterranean</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>26</td>
<td>10.2</td>
<td>No</td>
<td>Jaffé and Moritz (2010)</td>
</tr>
<tr>
<td>Region</td>
<td>Location</td>
<td>Climate</td>
<td>Habitat</td>
<td>Survey method</td>
<td>Survey area km²</td>
<td>Number of queens</td>
<td>Sample size</td>
<td>Number of colonies</td>
<td>Estimated density colonies km⁻²</td>
<td>Multi-year survey</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------</td>
<td>--------------------------------</td>
<td>----------------------</td>
<td>-----------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>---------------------------------</td>
<td>-------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Gosa Lodge, Motze Lodge</td>
<td>Arid</td>
<td>Desert</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>49</td>
<td>26.5</td>
<td>10.6</td>
<td>No</td>
<td>Moritz et al. (2008)</td>
</tr>
<tr>
<td>Australia - South Australia</td>
<td>Scott Conservation Park</td>
<td>Temperate</td>
<td>Open woodland</td>
<td>Direct search</td>
<td>0.09</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>11.1</td>
<td>No</td>
<td>Paton (1996)</td>
</tr>
<tr>
<td>North America - Texas, USA</td>
<td>Welder Wildlife Reserve</td>
<td>Temperate - subtropical</td>
<td>Coastal prairie</td>
<td>Direct search</td>
<td>0.25</td>
<td>N/A</td>
<td>N/A</td>
<td>24-78</td>
<td>3.8 - 12.5</td>
<td>Yes</td>
<td>Baum et al. (2005)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Leeuwfontein</td>
<td>Dry winters-humid subtropical</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>32</td>
<td>12.8</td>
<td>No</td>
<td>Moritz et al. (2007)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Suikerbosrand</td>
<td>Dry winters-humid subtropical</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>127</td>
<td>43</td>
<td>17.2</td>
<td>No</td>
<td>Moritz et al. (2007)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Ezemvelo</td>
<td>Dry winters-warm oceanic</td>
<td>Nature reserve</td>
<td>Drone capture</td>
<td>2.5</td>
<td>N/A</td>
<td>96</td>
<td>44</td>
<td>17.6</td>
<td>No</td>
<td>Moritz et al. (2007)</td>
</tr>
<tr>
<td>Australia - South Australia</td>
<td>Flinders Chase National Park</td>
<td>Temperate</td>
<td>Mallee forest</td>
<td>Direct search</td>
<td>2</td>
<td>N/A</td>
<td>N/A</td>
<td>45-80</td>
<td>22.5 - 40</td>
<td>No</td>
<td>Paton (1996)</td>
</tr>
<tr>
<td>Australia - South Australia</td>
<td>Cromer Conservation Park</td>
<td>Temperate</td>
<td>Open woodland</td>
<td>Direct search</td>
<td>0.15</td>
<td>N/A</td>
<td>N/A</td>
<td>6</td>
<td>40</td>
<td>No</td>
<td>Paton (1996)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Wyperfeld National Park</td>
<td>Temperate - Semi-arid</td>
<td>Riparian woodland</td>
<td>Direct search</td>
<td>0.35</td>
<td>N/A</td>
<td>N/A</td>
<td>27</td>
<td>77.1</td>
<td>No</td>
<td>Odkroyd et al. (1994)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Cobram</td>
<td>Temperate</td>
<td>Riparian woodland</td>
<td>Direct search</td>
<td>~0.15</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>30-120</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Shepparton East</td>
<td>Temperate</td>
<td>Riparian woodland</td>
<td>Direct search</td>
<td>~0.18</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>30-130</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Wyperfeld National Park</td>
<td>Temperate - Semi-arid</td>
<td>Riparian woodland</td>
<td>Direct search</td>
<td>0.25</td>
<td>N/A</td>
<td>N/A</td>
<td>45-80</td>
<td>40-148</td>
<td>Yes</td>
<td>Odkroyd et al. (1997)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Kyabram/Merrigum</td>
<td>Temperate</td>
<td>Riparian woodland</td>
<td>Direct search</td>
<td>~0.15</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>20-260</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Invergordon</td>
<td>Temperate</td>
<td>Riparian woodland</td>
<td>Direct search</td>
<td>~0.15</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0.320</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
</tr>
<tr>
<td>Australia - Victoria</td>
<td>Tatura/Ardomina</td>
<td>Temperate</td>
<td>Riparian woodland</td>
<td>Direct search</td>
<td>~0.27</td>
<td>N/A</td>
<td>N/A</td>
<td>Not specified</td>
<td>0.4-400</td>
<td>No</td>
<td>Goodman and Hepworth (2004)</td>
</tr>
<tr>
<td>Europe - Germany</td>
<td>Oberursel (near Frankfurt)</td>
<td>Wet warm oceanic</td>
<td>Not specified</td>
<td>Williams trap and direct analysis from apiaries</td>
<td>Not specified</td>
<td>N/A</td>
<td>142</td>
<td>107</td>
<td>Not specified</td>
<td>No</td>
<td>Baardry et al. (1998)</td>
</tr>
<tr>
<td>North America - New York</td>
<td>Arnot Forest</td>
<td>Temperate</td>
<td>Mixed forest</td>
<td>Direct search</td>
<td>11</td>
<td>N/A</td>
<td>N/A</td>
<td>8</td>
<td>Not specified</td>
<td>No</td>
<td>Seeley (2007)</td>
</tr>
<tr>
<td>Africa - South Africa</td>
<td>Kalahari Desert, Tswalu Game Reserve</td>
<td>Arid</td>
<td>Desert</td>
<td>Queen mating</td>
<td>Not specified</td>
<td>3</td>
<td>60</td>
<td>47.2</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2008)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Tapachula</td>
<td>Tropical</td>
<td>Mango plantation</td>
<td>Drone capture</td>
<td>Not specified</td>
<td>N/A</td>
<td>&gt;90</td>
<td>34</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Tapachula</td>
<td>Tropical</td>
<td>Garden, urban landscape</td>
<td>Drone capture</td>
<td>Not specified</td>
<td>N/A</td>
<td>&gt;90</td>
<td>34</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Tapachula</td>
<td>Tropical</td>
<td>Mango plantation</td>
<td>Drone capture</td>
<td>Not specified</td>
<td>N/A</td>
<td>&gt;90</td>
<td>35</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Rosario</td>
<td>Tropical</td>
<td>Agriculture: maize, grassland</td>
<td>Drone capture</td>
<td>Not specified</td>
<td>N/A</td>
<td>&gt;90</td>
<td>38</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Region</td>
<td>Location</td>
<td>Climate</td>
<td>Habitat</td>
<td>Survey method</td>
<td>Survey area km²</td>
<td>Number of queens</td>
<td>Sample size</td>
<td>Number of colonies</td>
<td>Estimated density colonies km&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Multi-year survey</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td>--------------------------------------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>------------------------------------------</td>
<td>------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Cancún</td>
<td>Tropical</td>
<td>Extensive agriculture, river valley</td>
<td>Drone capture</td>
<td>N/A</td>
<td>&gt;90</td>
<td>52</td>
<td>Not specified</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Izamal</td>
<td>Tropical</td>
<td>Deciduous forest, thorny bushes and shrubs</td>
<td>Drone capture</td>
<td>N/A</td>
<td>&gt;90</td>
<td>37</td>
<td>Not specified</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Zacapulco</td>
<td>Tropical</td>
<td>Mangrove</td>
<td>Drone capture</td>
<td>N/A</td>
<td>&gt;90</td>
<td>37</td>
<td>Not specified</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Puerto Chiapas</td>
<td>Tropical</td>
<td>Mangrove, grassland</td>
<td>Drone capture</td>
<td>N/A</td>
<td>&gt;90</td>
<td>43</td>
<td>Not specified</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Toniná</td>
<td>Tropical</td>
<td>Patchy agriculture and secondary forest</td>
<td>Drone capture</td>
<td>N/A</td>
<td>&gt;90</td>
<td>39</td>
<td>Not specified</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>San Cristóbal</td>
<td>Tropical</td>
<td>Agriculture, pine forest</td>
<td>Drone capture</td>
<td>N/A</td>
<td>&gt;90</td>
<td>44</td>
<td>Not specified</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Pujiltic</td>
<td>Tropical</td>
<td>Sugarcane plantation</td>
<td>Drone capture</td>
<td>N/A</td>
<td>&gt;90</td>
<td>54</td>
<td>Not specified</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
<tr>
<td>Central America - Mexico</td>
<td>Dzoncauich</td>
<td>Tropical</td>
<td>Low deciduous forest</td>
<td>Drone capture</td>
<td>N/A</td>
<td>&gt;90</td>
<td>42</td>
<td>Not specified</td>
<td>Not specified</td>
<td>No</td>
<td>Moritz et al. (2013)</td>
</tr>
</tbody>
</table>
2.7 Biosecurity concerns and policy

2.7.1 Pest, pathogens and alien species

Key factors identified in the decline of pollinators worldwide, as discussed in Section 2.2.3, include pests, pathogens, and alien species. To understand the spatio-temporal distribution of managed and feral colonies requires an understanding of the pressure these factors place on populations. In the context of managed colonies, biosecurity controls also influence patterns of movement of migratory colonies. The epidemiology of pests, pathogens, and alien species is in turn influenced by the spatial-temporal distribution of managed and feral colonies, and the interaction that occurs between them. A comprehensive treatment of all the pests and diseases threatening honey bees in Australia is available in beekeeping texts (Warhurst and Goebel, 2005), and from the BeeAware website (BeeAware, 2015b), with selected pests and diseases discussed further below.

Amongst temperate regions around the world, Australia has the unique but likely unsustainable position of being free of one of the worst pests of *A. mellifera* globally; *Varroa destructor*. *V. destructor* is an external parasitic mite of honey bees whose natural host is the Asian honey bee, *Apis cerana*. Around 50-60 years ago it was reported to have jumped species, and had acquired the ability to reproduce in colonies of *A. mellifera* (Anderson, 2006). It has since spread from Asia to other continents, and most recently New Zealand (discussed extensively in Section 2.7.3). In California, surveyed feral populations decreased by 75% shortly after the arrival of varroa (Kraus and Page, 1995) and in
Chapter 2: The Western Honey Bee

New Zealand by 95% (Goodwin and Taylor, 2007). It is almost inevitable that varroa will arrive in Australia, most likely via an illegal importation of queen bees (Cunningham et al., 2002). When this happens the estimated losses to agriculture are $A 21-50 million per annum (Cook et al., 2007). There is also evidence from many studies that feral colonies adapt over time to the presence of varroa (Seeley, 2007, Rosenkranz et al., 2010, Locke and Fries, 2011). There are no studies undertaken which have explicitly examined the role played by feral colonies in the spread of varroa or demonstrated transmission between apiaries via feral populations. The spread of pathogens is aided by varroa, as varroa feeds on the bee equivalent of blood; haemolymph (Kevan et al., 2006, Chen and Siede, 2007). The combined pressure on colonies introduced by varroa and pathogens is thought to be a cause of Colony Collapse Disorder (CCD) (Chen and Siede, 2007, Francis et al., 2013).

There are other mites which threaten *A. mellifera* populations in Australia. Closely related to *V. destructor* is *Varroa jacobsoni*, whose natural host is also *A. cerana*. There is recent evidence that *V. jacobsoni* is now also able to reproduce in colonies of *Apis mellifera* based on colonies found in Papua New Guinea in 2008 (Roberts et al., 2015). Two other mites are on the World Organisation for Animal Health (OIE) list of notifiable diseases; the Tracheal mite *Acarapis woodi* and the Tropilaelaps mites *Tropilaelaps spp*. Australia currently remains free of both pests (Animal Health Australia, 2015a).

Australia has recent experience with the introduction of another pest of honey bees, the small hive beetle, *Aethina tumida*. *A. tumida* is native to
sub-Saharan Africa, and was first detected in Australia near Richmond, New South Wales in 2002 (BeeAware, 2015c). Efforts were undertaken to establish the extent of the incursion through sampling of colonies and movement tracing, and it was found that *A. tumida* had already spread to locations 170 km to the north and 190 km to the west and south of Richmond (Gillespie *et al.*, 2003). In New South Wales, of the 120 positive detections initially made, 12 of these were in feral colonies (Gillespie *et al.*, 2003). Currently *A. tumida* has spread throughout New South Wales into Queensland and Victoria. Since 2007, *A. tumida* has also been present in northern part of Western Australia (Animal Health Australia, 2015a).

Pathogens of honey bees are also a significant cause of colony losses in both managed and feral populations. Two diseases of brood are also endemic in most regions of Australia; American foulbrood (AFB) and European foulbrood (EFB) (Animal Health Australia, 2015a). AFB is caused by a spore-forming bacterium *Paenibacillus larvae* subsp. *larvae* which is resistant to heat, drying and chemicals (Animal Health Australia, 2015a). AFB is considered the most significant bee disease already present in Australia (Australian Honey Bee Industry Council, 2015). AFB outbreaks are notifiable, and treatment involves destruction or irradiation of hive material (Animal Health Australia, 2015a). EFB is caused by the bacterium *Melissococcus plutonius*, and since first reported in Australia in 1977 has spread to all states and territories except Western Australia (Animal Health Australia, 2015a).
Aside from pests and pathogens, the other pressure on both domesticated and feral colonies are the arrival of other alien honey bee species. The Asian honey bee, *Apis cerana*, was first detected in Australia in Cairns in 2007 (Koetz, 2013). *A. cerana* is a known carrier of exotic pests and diseases, many of which are not present in Australia and hence the arrival of *A. cerana* was therefore a serious biosecurity incident. Many of same concerns pertaining to possible negative impacts of *A. mellifera* in natural environments are also equally applicable to *A. cerana*. In the Solomon Island *A. cerana* has been found to be able to out-compete *A. mellifera* (Anderson, 2008) but no effect on *A. mellifera* populations in the Australian incursion has yet been observed (Koetz, 2013). Response measures included reporting of swarms and nests by the public, targeted floral observations, feeding stations and traps, and monitoring with the aid of Rainbow bee-eaters (*Merops ornatus*) (Koetz, 2013). However, from 2011 the Australian Government took the decision that eradication was not possible, and moved to a program of management (Animal Health Australia, 2015a). The distribution of *A. cerana* now stretches across an area 140 km long and 65 km wide (Animal Health Australia, 2015a).

In North America, the most significant honey bee invasion in the last century has been the spread of the so-called "killer bee"; the Africanised honey bee, *Apis mellifera scutellata*. *A. m. scutellata* was deliberately introduced from South Africa to South America in 1956, and by 1990 had spread through Central America to the southern United States (Pinto et al., 2005). The feral population in areas where *A. m. scutellata* is genetically a hybrid with around one quarter to one third of its genetic
material originating from European races (Pinto et al., 2005). *A. m. scutellata* is a notifiable pest in most states of Australia, with the major impacts being the increased aggressiveness of hybrid species leading to an increase in attacks on humans and animals (BeeAware, 2015a).

Although not a notifiable pest in Australia, another subspecies of *A. mellifera; A. m. capensis* is thought to be the main cause for colony losses in South Africa (Pirk et al., 2014). *A. m. capensis* has an unusual reproductive strategy, as workers have the ability to produce female offspring including daughter queens and can therefore parasitise other colonies (Beekman et al., 2011).

### 2.7.2 Australia's honey bee biosecurity policies

Biosecurity is defined as "the protection of livelihoods, lifestyles and the natural environment, all of which could be harmed by the introduction of new pests, or through the impact of pests already established" (Plant Health Australia, 2012a). In 2012 responsibilities for Australian honey bee biosecurity at a national level moved from animal biosecurity to plant biosecurity. While the industry is currently a member of both Animal Health Australia (AHA) and Plant Health Australia (PHA), the industry plans to resign from AHA during 2015. Both agencies maintain contractual arrangements between federal governments, state and territory governments and industries to collectively prepare for and respond to emergency incursions. In the case of AHA, this is the Emergency Animal Disease Response Agreement (EADRA) (Animal
EADRA approaches for each industry are outlined in a series of manuals known as the Australian Veterinary Emergency Plan (AUSVETPLAN) (Animal Health Australia, 2015b). The AUSVETPLAN disease strategy manual for the honey bee industry identifies a number of characteristics of apiculture in Australia that make the eradication of honey bee pests and diseases particularly difficult to eradicate. These are reproduced in their entirety below (Animal Health Australia, 2010, p.31):

- the commercial necessity for apiarists to move honey bees, hive equipment and apiary products from place to place over long distances, often interstate, to harvest flora resources, service pollination contracts and maintain the condition of their bees;
- the unpredictable flowering of key native melliferous species (eucalypts, other myrtaceous species and various proteaceous species);
- the time-critical nature of requirements for the provision of pollination services to important horticultural and agricultural crops;
- intrastate and interstate movement of queen bees and packaged bees via the post or other means of transport;
- the natural tendency of honey bee colonies to swarm and establish feral nests in nearby areas;
- inter-colony drifting of infested worker and drone bees;
• the activities of worker bees robbing weakened colonies;
• the ability of honey bee drones to travel up to 16 km and enter hives other than their own;
• the presence of feral honey bee nests, in which diseases and pests may establish and persist;
• the possibility that diseases and pests may not be immediately apparent after their introduction;
• the specialist nature of the skills required to manage and handle bees; and
• the occupational health and safety issues associated with handling bees.

Consequently, the ability to eradicate bee pests or diseases is highly dependent on early detection and immediate action (Animal Health Australia, 2010). AUSVETPLAN provides industry-specific control and eradication policies, and recommended movement and quarantine controls.

With the shift of responsibilities to PHA, biosecurity approaches will now be governed by the EPPRD and its equivalent policy manual PLANTPLAN (Plant Health Australia, 2014). PLANTPLAN does not provide any honey bee industry-specific policies currently.

The Australian Honey bee Industry Council (AHBIC) has circulated a draft Biosecurity Code of Practice (Australian Honey Bee Industry Council, 2015) for comment. AHBIC are working with state and territory governments to make compliance with Part B of the code mandatory for
all beekeepers, and Part C mandatory for beekeepers managing more than 50 hives. Under Part C, beekeepers must complete training such as Plant Health Australia’s Biosecurity Online Training (Plant Health Australia, 2012b); unfortunately the training course does not provide any specific details on identifying pests, diseases or pathogens nor does it describe in detail what happens in the event of an incursion. The industry code is designed to be complementary to state and territory legislation, which takes precedence. A limiting factor in the broader adoption of the code is that beekeeper registration is not yet compulsory in all states and territories in Australia. Under the draft industry code, beekeepers must keep records all hive movements including swarm catch boxes for a minimum of 3 years.

The main surveillance in place for honey bee pests and diseases in Australia is the National Bee Pest Surveillance Program (NBPSP) administered by PHA (National Bee Pest Surveillance Program, 2014). The NBPSP is the successor of the National Sentinel Hive Program administered by AHA. Under the NBPSP, 130 hives with known health status are monitored across 29 sea and airport locations. Hives are tested every two months for the presence of mites. Empty hives, some equipped with cameras, are also used to detect possible swarms of exotic species.

If surveillance detects an incident, the Consultative Committee on Emergency Plant Pests first makes a determination whether eradication or containment is likely to be feasible (Plant Health Australia, 2014). Typically a restricted area (RA) around the infested premises (IP) is declared. PLANTPLAN does not specify a minimum radius, but
AUSVETPLAN stipulates a distance of 25 km and required that all feral colonies within 10 km (or 6.5 km) should be destroyed (Animal Health Australia, 2015b). All managed apiaries identified within the RA are quarantined and a larger control area (CA), possibly as large as a state or territory, is designated around the RA as a buffer zone by the relevant state or territory's Chief Veterinary Officer (CVO). Movements within the CA permitted, but hives within the CA cannot be moved outside the CA without the CVO's approval. All-other apiaries owned by the beekeeper are quarantined and inspected and all dangerous contact premises (DCP), determined via movement tracing of any materials or honey bees that have come into contact with materials or honey bees from the IP, will be inspected and quarantined as necessary (Animal Health Australia, 2015b). Experience in both Australia and New Zealand has found that some apiarists have quickly moved hives from an RA, in some cases before the RA has been officially declared (Animal Health Australia, 2010).

2.7.3 The *Varroa destructor* incursion in New Zealand

*Varroa destructor* was first detected in the upper North Island of New Zealand in April 2000 (Sanson, 2007). Subsequent analysis indicated that despite a sampling program of 400 hives per annum in high risk areas, *varroa* may have been present for at least 5 years before detection (Controller and Auditor-General, 2002) and it was first detected by an urban hobbyist beekeeper rather than the surveillance program (Goodwin, 2004). Movement restrictions were immediately
implemented, and analysis was done to determine whether eradication was possible.

Within the first week of the response, 15 field teams had checked 3,356 hives on 318 properties and tracked 777 hive movements from one property to another (Controller and Auditor-General, 2002). The survey continued until early June 2000 by which time 60,479 hives from 3,106 apiaries had been checked (Benard et al., 2001). Tracing of 2,327 beekeepers revealed 394 movements from infested areas, included 44 movements from the North Island to the South Island (Benard et al., 2001). The cost of the delimiting survey was over AUD$2m (Controller and Auditor-General, 2002) and the government received criticised for taking so long to complete it (Somerville, 2008). With the exception of movement of hives for pollination, beekeeping in New Zealand is largely stationary (Somerville, 2008), and authorities were fortunate to have access to records of the locations of most managed hives (Goodwin, 2004).

The delimiting survey found at least 248 apiaries were infested with varroa, all in the North Island, but including an area that had several thousand colonies brought in for kiwifruit pollination the previous spring (Goodwin, 2004). Analysis undertaken established that the maximum rate of local spread of varroa was 10-15 km/year (Stevenson et al., 2005). The uncertainty over the extent of the spread coupled with the difficulties eradicating managed and feral colonies over a large area led to a decision not to attempt eradication, and a movement restriction line across the middle of the North Island was put in place (Goodwin, 2004).
September 2003 almost all 10,000 plus apiaries north of the line were infested, with around 100 apiaries south of the line also infested. The line was shifted southwards and eastwards, having slowed the spread of varroa by about two years (Goodwin, 2004).

In 2006, varroa was detected in the South Island, where the sampling program had identified the incursion early enough to attempt elimination (Sanson, 2007). However, a decision not to attempt eradication was made, and varroa spread further through the South Island. In September 2008, the Ministry of Agriculture and Forestry (MAF) revoked all movement controls, having determined that declaring the infestation had progressed beyond the point where eradication was possible and that the geography prevented the establishment of effective movement control lines (Hamblyn, 2008).

2.8 Summary
Honey bees are of immense importance globally because of the role they play in pollination. This role is under pressure in many environments due to pests and pathogens, alien species and land use intensification. The effects of these pressures on honey bee populations, and in turn pollination, are difficult to assess without knowledge of the distribution of feral honey bees. Genetic-based survey methods provide a mechanism to measure feral population densities across large areas. However, it is not sufficient to make simple assumptions about queen and drone flight distances in order to make density inferences from genetic data; better techniques are required.
3 SPATIO-TEMPORAL MODELLING AND SIMULATION

3.1 Introduction

To fully understand the dynamics of the honey bee environment requires and understanding of how each of the elements vary across space and across time:

- **managed colonies** - moved by beekeepers in response to floral events and pollination events

- **feral colonies** - densities vary across space and across time, and genetics-based survey techniques necessitate an understanding of how observations at the population level (genotypes) arise from interactions at an individual level (reproduction)

- **floral resources** - the spatio-temporal distribution of floral resources, especially eucalypts, influences the movements of managed colonies and the density of feral colonies
This chapter discusses the motivation for developing models, examine spatial process concepts, and outlines frameworks for simulating honey bee and beekeeper movements.

### 3.2 The role of models and simulation

According to Longley *et al.* (2005), the term model is one of the most overworked terms in the English language. In discussing models, this thesis focuses on conceptual models (in particular scientific models) and their realisation as computerised models (which are then simulated).

O'Sullivan and Perry (2013, p.3) define scientific models as "a simplified representation of a system under study, which can be used to explore, to understand better or to predict the behaviour of the system it represents". The conceptual model of this system consists of the components, state variables, processes that operate on the components and interactions between components (O'Sullivan and Perry, 2013).

From models, and with the aid of computers, we can execute simulations. Winsberg (2009, pp.835-836) describes a computer simulation as "an algorithm, run on a computer, which uses step-by-step methods to explore the approximate behaviour of a mathematical model, usually because the model contains equations that cannot be solved analytically", but also notes that computer simulation may also refer more broadly to the entire modelling process. O'Sullivan and Perry (2013, p.9) state that in a simulation model "a computer is programmed to iteratively recalculate the modelled system state as it changes over time in accordance with the relationships represented by the mathematical and other relationships that describe the system".
Two key purposes of modelling and simulation are prediction (where the key quality of the model is accuracy) and understanding (where the key quality of the model is transparency) (Keeling and Rohani, 2008). For some authors, it is the latter purpose that is more important. O’Sullivan and Perry (2013) view the purpose of models as exploratory or heuristic learning tools, to be used as an aid to clarify thinking and generate further questions for investigation. Grimm and Railsback (2005) state simply that the purpose of models is to solve problems or answer questions. A model must not be viewed merely as a representation of a system, but a purposeful representation where the inclusions in the model are determined by the purpose for which the model is to be used. Axelrod (1997) suggests that in addition to prediction and discovery (understanding), additional roles of models include proof, performance of tasks, education, training and even entertainment.

The use of models and simulations to make predictions should be undertaken with considerable caution. Silver (2012) identifies many reasons why predictive models fail, including chaotic systems, bias, overfitting, extrapolation and feedback loops. A chaotic system is a dynamic system in which small changes to the input conditions create exponentially magnified changes in the output dynamics, giving the appearance of randomness (Lorenz, 1995). This phenomenon was first observed by Lorenz in the study of weather systems during the 1970s. Bias is extensively discussed in Kahneman (2011); whose work involves the study of cognitive biases in the application of judgement and decision making. Overfitting arises when a model describes random noise, rather
than the underlying relationship and generally occurs as a result of utilising too many parameters in the fitting of the model to a particular set of sample data. Finally, feedback loops arise when the behaviours in the system under study change as a result of a change in state.

Models and simulations are frequently used in the study of complex adaptive systems (CAS). In a CAS, the components (often termed agents) interact and adapt or learn (Holland, 2006). CAS are broadly equivalent to multi-agent systems (MAS); CAS is more focussed on high-level properties of the system while MAS is more focussed on the low-level implementation (Fromm, 2004). Complex systems can be characterised by properties such as path dependence, positive feedback, self-organisation, and emergence (O’Sullivan and Perry, 2013). Emergent properties are those distinct from the properties of individual agents or the sum of those properties, and typically cannot be easily predicted from individual properties (Grimm and Railsback, 2005).

3.3 Spatial processes and patterns
A particular sub-class of models and simulation are those explicitly incorporating space. O’Sullivan and Perry (2013) define three broad categories of spatial model; aggregation, movement and spread. These models incorporate spatial processes which lead to the emergence of spatial patterns. Many spatial analyses start with observations of spatial patterns and seek to decode the process that gave rise to those patterns (Grimm and Railsback, 2005, O’Sullivan and Perry, 2013). This is a challenging task; changes in process intensity can create different patterns, and different processes can generate the same process signature.
Spatial patterns often comprise a combination of first order and second order effects. First order effects arise from variation in the mean value of the process in space leading to global or large scale trends (Bailey and Gatrell, 1995). Patterns with first-order structure are sometimes referred to as non-stationary or inhomogeneous (O’Sullivan and Perry, 2013). Second order effects arise from the spatial correlation structure or spatial dependence in the process (Bailey and Gatrell, 1995), implying that deviations in process values from the mean are more alike at neighbouring sampling locations (Fortin and Dale, 2005) and thus leading to local or small scale trends.

The choice of scale influences how patterns are perceived, how processes are represented, and the interactions between pattern and process (O’Sullivan and Perry, 2013). The term scale may refer to the spatial and temporal resolution of the data, or the extent of the processes generating the spatial patterns (Fortin and Dale, 2005). There is no single correct scale for model construction or data collection, and the scale chosen should primarily be determined by the purpose of the model (Levin, 1992, Urban, 2005, O’Sullivan and Perry, 2013).

Point patterns represent the simplest form of spatial data, where each point represents the location of an object. Point patterns are among the
most frequently collected and encountered forms of spatial information (O’Sullivan and Perry, 2013). The null model for point patterns arise in the case of a point pattern that lacks both first-order and second-order structure; this is often referred to as showing complete spatial randomness (CSR) and is more formally known as the homogeneous Poisson process (O’Sullivan and Perry, 2013).

Where second-order effects exist such that the presence of one point increases the probability of finding another nearby, the resulting pattern contains clusters or aggregations (Dale, 1999). An example of a spatial process giving rise to aggregations is the Matern cluster process; parent points are generated by a uniform Poisson process with intensity $\kappa$, then replaced by a cluster of evenly distributed offspring points within a circle of radius $r$. The number of points per offspring cluster being is Poisson distributed with mean $\mu$, and the overall intensity is $\kappa \times \mu$ (Baddeley and Turner, 2005).

Where the presence of one decreases the probability of finding another nearby, the distribution is referred to as either overdispersed, inhibited, uniform or regular (O’Sullivan and Unwin, 2003, Fortin and Dale, 2005). An example of a spatial process giving rise to an overdispersed distribution is simple sequential inhibition (SSI), whereby events are sequentially placed at randomly generated locations, but new events are constrained such that they cannot be placed within some inhibition distance of any already placed event (O’Sullivan and Perry, 2013).
The crude density or overall intensity of a point pattern can be calculated by determining the number of events in the study area (O'Sullivan and Unwin, 2003). For any case other than the CSR, the challenge with using density as a measure is the sensitivity to definition of the study area (O'Sullivan and Unwin, 2003). Rather than determining a single density value for the entire study area, it is possible to derive local density estimates at every location within the study area (not just at event locations) using a technique called kernel density estimation (O'Sullivan and Unwin, 2003). For each location in the study area, kernel density estimation counts the number of events within a circle of radius $r$ (the "kernel"), and divides by the area of the circle to determine a local intensity (O'Sullivan and Unwin, 2003). The local density estimate is thus sensitive to the choice of $r$, the kernel bandwidth; O'Sullivan and Unwin (2003) suggest choosing a kernel bandwidth that has some meaning in the context of the study.

Alternatives to density based point pattern measures are those based on the analysis of the distances between points. Measures include the mean nearest neighbour distance, $G$ function (cumulative frequency distribution of the nearest neighbour distances), $F$ function (similar to $G$, but using point locations anywhere in the study region selected at random) and Ripley's $K$ function (O'Sullivan and Unwin, 2003). Distance based measures are able to reveal second-order structure such as aggregation or overdispersion in the point pattern.
3.4 Emergent movement patterns

At its simplest level, movement can be viewed as a spatial process. Over time, the path traced by the movement of an individual creates a spatial pattern. Similarly, the aggregate movements of a large number of individuals can also give rise to patterns. While movement itself is continuous, most movements at the geographic scale are sampled with respect to some appropriate time interval for the model of movement being developed. The path then traced can be represented by a polyline where the vertices represent the individual’s location at a point in time and the edges represent steps or jumps interpolating the trajectory of movement between sampling times.

The simplest possible emergent pattern of movement is that of random movement, where both the direction and magnitude of each step is drawn from a random distribution. There is little evidence of random movement occurring in intelligent life, but it serves as a useful null model for comparing other observed patterns (O’Sullivan and Perry, 2013). Where the direction of movement in one step is correlated with the direction of movement in the previous step, this gives rise to a correlated random walk (CRW). CRWs better reflect the tendency of animals to move forwards, and thus create a local directional bias or persistence effect in the movement trajectory (Codling et al., 2008). A global directional bias arises when the probability of movement in a particular direction is greater, and this gives rise to biased random walks (BRW) (Codling et al., 2008). Random walks can be both correlated and biased (CBRW). If the random distribution from which the step lengths is drawn has finite
variance, then the root mean squared displacement grows proportionally with the square root of the walk duration in accordance with the Central Limit Theorem (O'Sullivan and Perry, 2013).

A particular class of movement trajectories known as Levy flights arise when the magnitude of step sizes is drawn not from a heavy-tailed distribution; that is, one with infinite variance (Codling et al., 2008). Levy flights thus describe movement trajectories in terms of a large number of small displacements coupled with a small number of large displacements; a biologically-inspired analogy would be the movement of migratory animals between seasonal home ranges. A large body of research has associated Levy flights with patterns of movements for many different animals including bees (Reynolds et al., 2007, Reynolds, 2009, Reynolds et al., 2009), but other authors have called into question the validity of many of these studies. A review of the evidence against Levy flights is provided by (Pyke, 2015).

In practice, most animal movements are goal-oriented (Cassini, 2013) and influenced by behavioural drivers. Movements occur within an environment, and the patterns observed are influenced strongly by the configuration of the environment; for stable environment configurations the movement trajectories have also been found to be remarkably stable (O'Sullivan and Perry, 2013). Movement may be directed by foraging or search for resources, and where those resources are spatially aggregated the search strategy will utilise this property. Behaviour arising from interactions with other individuals (e.g., predation, competition, reproduction) can also give rise to characteristic patterns of movement.
3.5 Computational movement patterns

In contrast with "bottom-up" emergent patterns of movement, movement may also occur in the form of "top-down" solutions to optimisation problems. Beekeepers planning the movement of their hives need seek to maximise their yield while minimising their travel costs and times. This optimisation problem bears close resemblance to an extensively studied problem in operations research; namely the travelling salesman problem (TSP).

The name given to the TSP originates from the idea of travelling salesman needing to visit each one of a fixed number of cities once (and only once) before returning to their starting location. Given the cost of travel between each pair of cities, typically based on the distance between these cities, the travelling salesman should choose a route (or more explicitly, a “tour”) that minimises the total cost. The Euclidean TSP is a variant of the TSP where the costs for travel between cities are based only on the Euclidean distances separating the points (Applegate et al., 2006).

It is uncertain where the name of the TSP originates; references to the problem in its original guise date back to 1832, although the study of the TSP within the operations research community began much later in the 1930s (Applegate et al., 2006). The applications of the TSP are diverse, and include logistics, genome sequencing, scan chains, printed circuit board drilling problems and data clustering (Applegate et al., 2006).

For any TSP instance with \( n \) cities, there are \( (n - 1)!/2 \) possible tours and thus an exhaustive solution of the search space is \( O(n!) \). The TSP has been shown to be NP-hard, and the problem has become a benchmark for
examining the quality and performance of combinatorial optimisation procedures due to its easy specification and extensive library of problem instances (de Smith et al., 2007).

Over the last 60 years, a variety of computationally exact and heuristic solutions have been developed for solving the TSP. As the performance of these techniques has improved, so too has the processing power of computers. With Dantzig’s creation of the simplex algorithm in 1947, an exact solution to the TSP using linear programming techniques based on cutting planes was developed by Dantzig, Fulkerson and Johnson (Applegate et al., 2006). This technique was used to solve the United States’ 49-city problem for the first time in 1954. Other techniques soon followed including branch-and-bound methods and dynamic programming (Applegate et al., 2006). The largest optimal tour found so far is for 85,900 cities (Applegate et al., 2006). The simplest heuristic procedure is the nearest-neighbour procedure. In the nearest-neighbour procedure a tour is constructed by adding to the tour the location nearest to the last added location. The performance of the nearest-neighbour heuristic is not particularly good, but it often serves to generate a starting tour on which further heuristic optimisation procedures are performed. A major contribution to heuristic solutions for the TSP was provided by Lin and Kernighan in 1973, with a tour improvement method that finds modifications to a given tour that result in a new tour of lower cost (Applegate et al., 2006). Greedy heuristics, interchange heuristics, and metaheuristics such as tabu search, cross-entropy methods and simulated annealing have also been used to solve the TSP (de Smith et al., 2007).
Evolutionary computing techniques based on biological inspiration such as ant-colony optimisation also provide solutions to the TSP (Dorigo and Gambardella, 1997).

As actual human tours were the inspiration for the travelling salesman problem, it is appropriate that there has been significant research into how humans find efficient tours without computational aid. Analysis shows that human performance decreases linearly with an increasing number of locations, yet the calculation complexity increases factorially (MacGregor et al., 1999). Human heuristic procedures consistently outperform simple construction algorithms by an order of magnitude in the case of 10-city and 20-city problems (MacGregor and Ormerod, 1996). The avoidance of crossings and/or convex hull procedures are also thought to play a role in human heuristic procedures (MacGregor et al., 2004). Convex hull procedures are of interest because it is an established property of optimal tours that the locations on the boundary of the point set (the convex hull) will be visited in order (MacGregor and Ormerod, 1996). Human performance on Euclidean TSPs has been shown to improve as spatial point patterns tend towards aggregated distributions (Dry et al., 2012). A comprehensive review of research into human performance on the TSP and related problems is provided by MacGregor and Yun (2011).

While human heuristic solutions to the TSP are relevant in the case of a beekeeper performing site reconnaissance, finding efficient solutions to the movement of hives themselves requires the solving of a slightly different optimisation problem. Among the many variations of the
general TSP, there are a class of Vehicle Routing Problems (VRP) inspired by finding efficient tours for vehicles required to pickup and/or deliver goods at each of the points in the tour. The parameters relevant to the development of an optimal tour for the pickup and delivery of hives from one apiary site to the next are as follows:

Table 5: Vehicle Routing Problem (VRP) parameters relevant to managed beekeeper hive movements

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Choice</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depot involved</td>
<td>Yes</td>
<td>Beekeepers commence and complete their journeys at a home base or depot. In some cases, hives are taken to/from a depot.</td>
</tr>
<tr>
<td>Origin/destination</td>
<td>many-to-many (M-M)</td>
<td>Hives from any pickup delivery site can be taken to any destination site</td>
</tr>
<tr>
<td>matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of vehicles</td>
<td>1</td>
<td>Most beekeepers operate a single large truck with the capacity to load and unload all hives at an apiary.</td>
</tr>
<tr>
<td>Vehicle capacity</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Profits</td>
<td>Yes</td>
<td>The profitability of an apiary site varies considerably.</td>
</tr>
<tr>
<td>Static/dynamic</td>
<td>Dynamic</td>
<td>Beekeepers update their decisions about hive movements based on observations taken when they arrive at an apiary, observations made in transit (e.g., whether flowering is happening as planned), and weather observations.</td>
</tr>
<tr>
<td>Time windows</td>
<td>Yes</td>
<td>To optimise their yields, beekeepers need to have their hives at a certain location by a certain point of time.</td>
</tr>
</tbody>
</table>

Given the highly specific nature of the above problem, no specific formulation or heuristics have been currently been developed to solve it. However, variations of the VRP with one or more of the individual parameters outlined above exist. The best match for a single-vehicle case is known as the One-commodity Pickup and Delivery Travelling Salesman Problem (1-PDTSP) (Hernández-Pérez and Salazar-González, 2004). The 1-PDTSP refers to a class of vehicle routing problems where the pickup and delivery locations are unpaired and a homogeneous good is
transported (Parragh et al., 2008). The starting (and finishing) point for routes in this context is referred to as the depot. The goal, as for the generic TSP, is to service all the necessary pickup and delivery sites in a minimum distance route. Computationally, the 1-PDTSP is harder to solve than the TSP, and this difficulty increases as the capacity of the vehicle decreases (Battarra et al., 2014). Many of the same techniques used to solve the TSP are equally applicable to the 1-PDTSP including branch-and-cut algorithms, greedy algorithms and genetic algorithms (Battarra et al., 2014). The 1-PDTSP does not directly provide a solution for vertices with profits, dynamic routing problems or time window constraints and these are therefore discussed in the context of other TSP variants below.

The generic TSP requires that all locations are visited, and does not attach any value to the servicing of the locations themselves. If locations are given a profit or attractiveness and the requirement to service all sites is relaxed, then it becomes possible to select which locations are visited based on both the distance cost and the location profit. In the case of a single vehicle, this class of TSP variants is known in the literature as the TSPs with profits (Feillet et al., 2005). There are conflicting objectives involved in such a solution; the traveller is compelled to travel to collect profit but also seeks to minimise the travel costs. Most researchers do not seek to solve the bicriteria problem directly, but combine the objectives in a linear or weighted manner and solve the resulting single criteria problem instead (Feillet et al., 2005).
In dynamic routing problems, data needed to make routing decisions is revealed over time and therefore planned routes may need to be updated based on new information (Berbeglia et al., 2010, Battarra et al., 2014). This contrasts with simpler static routing problems where the input data to the routing decision are known in full prior to computation of the route and do not change over time. The uncertainty in dynamic problems may be stochastic in nature (Battarra et al., 2014), and the planning horizon of a dynamic problem may be unbounded (Berbeglia et al., 2010).

In both static and dynamic routing problems, an additional constraint can be imposed by time windows, whereby the pickup and delivery must occur within a given time interval (Doerner and Salazar-González, 2014).

3.6 Models of species distribution and abundance

The geographic distribution of a species will be controlled by the environmental factor for which the organism has the narrowest range of tolerance; this ecological rule is known as Shelford's Law of Tolerance (Krebs, 2009). On a global scale, temperature and moisture are the main limiting factors and distributions also be impacted by barriers that block dispersal. On a local scale, light, fire, pH, and other physical and chemical factors can limit distributions, but few species are limited by a failure to disperse. Other organisms may also play a role in limiting the distribution of a given species, either as predators, parasites, pathogens or competitors. Competition arises not just from a scarcity of resources, but also from interference in gaining access to needed resources.
Within the distribution of a given species, population abundance is determined from the integration of natality, mortality, immigration, and emigration. Distribution and abundance are commonly positively related in that species are comparatively more abundant when their distributions are comparatively larger.

A species distribution model (SDM) is a "descriptive model that relates species occurrence to environmental (biotic and abiotic) factors to describe environmental conditions within which a species occurs" (Zurell et al., 2010, p.623). SDMs are built from presence-only data, and this creates challenges validating the predicted distribution and assessing model accuracy (van Proosdij et al., 2015). As knowledge of the real world is always incomplete, it is difficult to compare sampling schemes and models. An alternative approach, termed the "virtual ecologist" (VE) approach, proposes the development of a virtual world as a surrogate for reality (Zurell et al., 2010). This virtual environment allows simulation of the observation of the virtual ecosystem by a virtual ecologist mimicking the way that data would be collected by real ecologists in real ecosystems. In this manner, testing and comparison of sampling schemes and models can be achieved. Miller (2014) provides a progress report on the use of virtual species distribution models (vSDMs) within the context of the VE approach.

An emerging area of research relevant to the study of species distribution and abundance is landscape genetics. Landscape genetics amalgamates landscape ecology; the study of the interaction between spatial patterns and ecological processes, with molecular population genetics (Manel et
al., 2003). By combining complex and realistic life histories, behaviours, landscape features and genetic data, landscape genetics aims to relate spatial patterns to process (Epperson et al., 2010). Computer simulations can be used to show how demographic processes such as dispersal or reproduction interact with landscape features to affect probability of site occupancy, population size, and gene flow, which in turn determine spatial genetic structure (Epperson et al., 2010). The models described in Chapter 4 can thus be thought of as operating within the paradigm of landscape genetics. A general review of genetic techniques in ecology is provided in DeYoung and Honeycutt (2005) with a comprehensive review of software tools outlined in Hoban et al. (2012).

3.7 Spatio-temporal modelling techniques

Spatially explicit models are those where the relationship between distinct spatial elements in the model is important for a full understanding of the system. The two key requirements of a spatial model are that there is variation across the space being manipulated by the model and the results of modelling changes when the locations of objects change (Longley et al., 2005). Many of the spaces in these model exist at geographic scales, and the application of computational models to geographic problems is often termed geocomputation (Longley et al., 2005).

The representation of space can utilise a grid or lattice representation (analogous to a raster representation in a GIS) or continuous space (analogous to a vector representation) (O’Sullivan and Perry, 2013). The co-ordinate space may be finite, infinite, or wrapped; in the latter case the
boundaries of the two-dimensional space are wrapped to form what is nominally a toroidal-shape in three dimensions thereby avoiding edge effects (O’Sullivan and Perry, 2013). A key design consideration for spatially-explicit model is the resolution, which influences both the costs of acquiring the data and also the costs of running the model (Longley et al., 2005).

Many spatially explicit models seek to represent dynamic spatial processes, and therefore include a temporal dimension as well as spatial dimensions. Just as the choice of spatial resolution represents an important design decision, so too does the choice of temporal resolution. Another fundamental issue is whether or not all of the state changes in a model occur simultaneously, when they are termed synchronous, or if they happen one at a time to each affected component of the model, when they are termed asynchronous (O’Sullivan and Perry, 2013).

The simplest spatio-temporal models utilise a modelling paradigm known as cellular automata (CA). The origins of CA lie in work done by von Neumann and Ulam in the 1940s and 1950s (Wolfram, 2002). Two-dimensional CA were popularised in the 1970s when Conway introduced a model called the Game of Life, and one of the first to suggest using CA to model geographic processes was Tobler in 1979 (Castle et al., 2007). In spatially explicit models, the study area is typically decomposed into a lattice of regular (square, hexagonal or triangular) cells, but other geometries such as Voronoi polygons can be used (Castle et al., 2007). Each cellular automaton has a state; in the case of ecological models this may be the presence of a species, the amount of biomass or the size of a
local subpopulation (Grimm and Railsback, 2005). Around each cell a 
neighbourhood structure determines which adjacent cells will exert 
influence on the automaton (Castle et al., 2007). Transition rules 
consider the state of the cell itself and its neighbour cells, to determine 
the state of the cell in the next time step; in ecological models these rules 
are usually probabilistic (Grimm and Railsback, 2005).

The inability of an automaton to move within a CA frame led to the 
development of an alternative modelling paradigm based on autonomous 
agents moving in a continuous space referred to as agent-based modelling 
(ABM) (Castle et al., 2007). ABMs are synonymous with individual-based 
modelling (IBM), a term frequently used in the ecological modelling 
literature (Railsback and Grimm, 2011). In ABMs, multiple autonomous 
agents interact within a simulation environment (Castle et al., 2007), and 
thus ABMs often provide a mechanism to study complex adaptive 
systems.

While in general there is no consensus among researchers on the 
definition of an agent, there is some consensus in the geospatial 
literature. Both Castle et al. (2007) and Crooks and Heppenstall (2012) 
reproduce the text of an earlier paper by Castle and Crooks (2006) 
(without attribution!), in which a number of key characteristics of agents 
are defined:

- autonomy – agents can acquire, process and act on information 
  independently of each other or of any centralised control
The concept of an agent as a model of an actor in a real environment provides what O'Sullivan et al. (2012) refer to as a "intuitively satisfying representational approach", and as the ease with which ABMs can be created has increased so too has their adoption across many areas of science.

This is not to suggest, however, that ABMs should be used for geocomputation simply because they can be (O'Sullivan et al., 2012). There are particular situations for which ABMs provide an advantage over other modelling techniques, and these are reviewed by O'Sullivan
et al. (2012) and Bonabeau (2002). In summary, ABMs are best suited to models where any or all of the following apply:

- when space is critical, and the position of agents is not fixed
- the interactions between agents are complex, nonlinear, discontinuous or discrete, and the magnitude of the interaction effect is not weak or minor
- when either the population is heterogeneous, or heterogeneity of the decision-making context of agents leads to a heterogeneous topology of interactions
- the overall size and organization of the system is small enough to preclude mean-field approaches
- where the agents exhibit complex behaviour such as adaptation and learning

ABM is implemented in software, and in that respect there is a natural equivalency between agents and software objects. Several free and open source modelling environments and libraries are available for creating models. A recent review of these has been provided by North et al. (2013). Some of these frameworks have also been adapted for use in high-performance computing (HPC) environments, as the computation demands of ABMs may exceed the processing and storage capacity of a single computer (Parry and Bithell, 2012). A review of HPC ABM environments is provided by Rousset et al. (2014). Three recent texts (Railsback and Grimm, 2011, O’Sullivan and Perry, 2013, Wilensky andRand, 2015) are devoted to creation of ABMs using the NetLogo (Wilensky, 1999) framework. An alternative environment, and the one
used in this thesis, is Repast Simphony (North et al., 2013). The major advantage of Repast Simphony over NetLogo is speed of execution (Railsback et al., 2006).

3.8 Model description and evaluation
An historical criticism of ABMs is that results are irreproducible, largely because their published descriptions are ambiguous and/or incomplete (Grimm et al., 2006, Topping et al., 2010, Railsback and Grimm, 2011). Within the domain of ecological modelling, there has been a strong move towards the use of a standard protocol for model description called the ODD (Overview, Design concepts, Details) protocol. The ODD protocol was first defined in 2006 (Grimm et al., 2006), with an updated version published in 2010 based on an analysis of 54 publications that had used the protocol in the intervening period (Grimm et al., 2010). The number of citations of the original paper based on data from Web of Science has increased from 87 in December 2009 (Grimm et al., 2010) to 611 in August 2015 with 319 citations of the 2010 paper.

Models described using ODD adopt the following format, described in detail in Grimm et al. (2010) and Railsback and Grimm (2011):
Table 6: Elements of the ODD protocol (additions from 2010 in bold)

<table>
<thead>
<tr>
<th>ODD section</th>
<th>Element number (optional)</th>
<th>Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>1</td>
<td>Purpose</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td><strong>Entities, state variables and scales</strong></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Process overview and scheduling</td>
</tr>
<tr>
<td>Design Concepts</td>
<td>4</td>
<td>Basic principles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emergence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adaptation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Objectives</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Learning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prediction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stochasticity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collectives</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observation</td>
</tr>
<tr>
<td>Details</td>
<td>5</td>
<td>Initialisation</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Input <strong>data</strong></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Submodels</td>
</tr>
</tbody>
</table>

In addition to describing the model, it is important to also evaluate the performance of the model. At a minimum, the developers of a model should take steps to ensure that the model is realistic enough to meet its intended purpose (Rykiel, 1996). For the last 60 years, there has been debate in the ecological modelling community as to what steps should be taken, what tests of sufficiency exist, and finally what terminology to use with respect to these steps (Augusiak et al., 2014). Augusiak et al. (2014) outline a number of terms used in the literature either with variable definitions or as synonyms; validation, verification, evaluation, testing, corroboration and substantiation. Because of the confusion over terminology Augusiak et al. (2014) suggest the use of a new hybrid term; "evaludation". Evaludation is not a single activity that occurs after
development of the model is completed, but rather a series of six elements used iteratively as part of the "modelling cycle" (Grimm and Railsback, 2005). The six elements described by Augusiak et al. (2014) are:

i. **data evaluation**: the data used to parameterise the model may be both quantitative and qualitative in nature, where the latter is acquired from expert knowledge. Data evaluation requires that available data be assessed for appropriateness and accuracy. Particular care should be taken to understand what inferences and assumptions are inherent in the data (Oreskes et al., 1994).

ii. **conceptual model evaluation**: as the conceptual model of any system represents an abstraction of the system being studied, it is necessarily based on simplifying assumptions and choices in model design. These simplifying assumptions and choices should be reviewed as part of conceptual model evaluation.

iii. **implementation verification**: implementation verification is primarily focussed on whether a model is built as designed; that is, the intended technical functionality of the model has been tested for correct operation.

iv. **model output verification**: a model output is verified by comparing it with observations. In practice, this is often difficult; an inability to gather observational data is often a primary reason for the development of a model. Where observational data do exist and model output is shown to match observations well, two further
considerations apply. The first is to what extent calibration of the model has occurred to match observation; has the model been overfitted? The second is known as the equifinality problem (O'Sullivan and Perry, 2013); a large numbers of models, including the same model with different input data, could produce output consistent with observational data.

v. **model analysis:** in model analysis a sensitivity analysis of the model is undertaken to ensure that that the emergence of model output is well understood. The goal of a local sensitivity analysis (SA) is to isolate the parameters to which a model is most sensitive and rank them in terms of the model's sensitivity (O'Sullivan and Perry, 2013). Related to SA is the more general approach of uncertainty analysis (UA), which is concerned with estimating how uncertainty in multiple likely interacting parameters and their representation will affect a model (O'Sullivan and Perry, 2013).

vi. **model output corroboration:** in model output corroboration, the output of the model is corroborated with data not used for model development and parameterisation.

While model description and evaluation (or evaludation) are important, a final element needed to both build confidence and communicate the model is documentation. As ecological models tend to be developed iteratively, design decisions and testing undertaken at each iteration often go undocumented. For this reason, there is a need to broadly implement a standardised protocol to establish a culture of
comprehensive modelling documentation; one possible solution is the TRACE protocol (TRAnsparent and Comprehensive Ecological modelling documentation) (Schmolke et al., 2010, Grimm et al., 2014).

3.9 Honey bee models and simulation
As one of the most studied of all insects, it is perhaps unsurprising that most aspects of a honey bee's lifecycle and behaviour have been captured in models. (Becher et al., 2013) provide a comprehensive review of 31 different models of colony dynamics (8 models), honey bee/varroa/virus interactions (12 models) and foraging behaviour (11 models). They found that model analysis in most models was limited and focussed on sensitivities of the output state to one or a few parameters only. For the 11 foraging behaviour models, none simulated the spatiotemporal variability of forage availability and quality on foraging success.

Using a similar search strategy to Becher et al. (2013), additional and newer models of honey bees can be identified. Searching the Web of Science database using the same advanced search strings as Becher et al. (2013) identifies more recent models not included in their review. To identify models potentially related to honey bee mating behaviour, a search was performed using the advanced search string; TS = ((honeybee OR honey bee OR Apis) AND (queen OR drone) AND (movement OR flight OR mating OR trap OR pattern) AND (model* OR simulation)). As of September 2015, this yields 199 results. To identify models potentially related to the use of microsatellite markers to infer population densities an additional search string was used; TS = ((microsatellite OR genetic*) AND (density OR distribution OR abundance) AND (survey OR sampling...
OR trap) AND (agent-based OR ABM OR individual-based OR IBM)). As of September 2015, this yields 129 results. Many of the search results were not directly relevant to honey bee models and were thus excluded from further analysis.

The models collected through this search were based on a range of different modelling techniques. The majority used a formulation based upon mathematical equations. Examples include intra-colony dynamics (Khoury et al., 2013, Russell et al., 2013, Torres et al., 2015), epidemiology (Betti et al., 2014) and bioeconomics of the honey bee industry (Champetier et al., 2015). Other models made use of GIS-based analysis, such as for the estimation of honey production at apiary scale (Janssens et al., 2006, Sponsler and Johnson, 2015). Finally, several models made use of agent-based or individual-based modelling techniques. Examples include foraging (Schürch and Grüter, 2014), nest-site choice by swarms (List et al., 2009) and competition amongst species for pollen resources (Everaars and Dormann, 2014).

The most comprehensive model of honey bees yet developed is BEEHAVE (Becher et al., 2014). BEEHAVE provides an integrated model of intra-colony dynamics, honey bee/varroa/virus interactions and foraging behaviour in a spatially explicit landscape. BEEHAVE has been developed as an agent-based model in NETLOGO, and both source code and a user manual are available. The overall trend in intra-colony dynamics and foraging behaviour emerging from BEEHAVE has been shown to correlate well with empirical observations. Verification of the varroa model was performed through comparison with the models of Martin (1998, 2001).
Other models more specifically focussed on honey bee biosecurity issues in Australia and New Zealand have been developed, although these did not appear in the search results from the Web of Science. Clifford et al. (2011) developed a model based on cellular automata to determine optimum biosecurity surveillance parameters, and showed that increasing the number of surveillance hives along with improving the efficiency of the detection method yielded the largest improvement in detection times. Gordon et al. (2014) used survey data from beekeepers in conjunction with a cellular automata model to develop a connected network of sites based on beekeeper hive movements and demonstrate the potential for the rapid spread of pests and diseases in Australia. Hafi et al. (2012) combined a cellular automata model of varroa spread with mathematical equation models commodity and pollination markets to perform a cost-benefit analysis of responding to a varroa incursion in Australia. Sanson (2007) used a model to evaluating surveillance programs for the detection of varroa in New Zealand, and Stevenson et al. (2005) provide an analysis of varroa epidemiology in New Zealand based on a mathematical equation model.

Across all the honey bee models reviewed, inter-colony interactions are considerably underrepresented compared with intra-colony interactions. Most inter-colony models focus on varroa epidemiology (Stevenson et al., 2005, Sanson, 2007, Clifford et al., 2011, Hafi et al., 2012, Gordon et al., 2014). Additionally, the techniques of landscape genetics have been used in analysis of honey bee populations on the Iberian peninsula (Pinto et al., 2010, Cánovas et al., 2011) and in East Africa (Fuller et al., 2015). Very
few studies specifically model honey bee competition. Competition arises in many forms, including between managed and feral honey bees and between honey bees and other pollinators and hollow-dwellers. Competition for pollen amongst non-\textit{Apis} bees is modelled in Everaars and Dormann (2014), but no similar model has been developed for \textit{A. mellifera}. Inter-colony interactions also arise in honey bee mating, including drone congregation area formation and queen/drone multiple matings. Except for the papers reproduced in the next chapter (Arundel \textit{et al.}, 2012, Arundel \textit{et al.}, 2013), no other models of honey bee mating have been developed at the time of writing.

3.10 Summary

To infer the density of feral colonies from genetic survey data will require the development of a new model. Agent-based modelling provides a spatially-explicit modelling technique whereby behaviours at an individual level (e.g., mating between queens and drones) and emergent properties at an aggregate level (e.g., the number of genotypes in a sample) can be simulated in a virtual environment.

For beekeepers to efficiently plan their seasonal movements, they rely on human heuristics to solve complex optimisation problems. The form of the problem closely resembles variants of the Travelling Salesman Problem and Vehicle Routing Problems.
4 AGENT-BASED MODELS OF HONEY BEE BEHAVIOUR

4.1 Introduction

This chapter describes how simulations of the honey bee mating system can be used to generate synthetic sampling distributions for the application of genetics-based survey techniques of feral colonies. This chapter reproduces two journal papers published in Ecological Modelling, namely:


My contribution to these papers is outlined in the Preface. The literature review of each paper has been incorporated into Chapter 2 and Chapter 3, and thus removed from this chapter.

4.2 Modelling honey bee queen mating as a measure of feral colony density

4.2.1 Abstract

Robust estimates of feral and wild honey bee (*Apis mellifera*) colony densities are essential for the understanding of the role of honey bees in an ecosystem, and for planning responses to an incursion of an exotic honey bee disease or pest. New genetic methods make it possible to identify the number of feral colonies with which a test queen mates. We present an agent-based model of the spatial process of honey bee mating that describes the relationship between feral colony densities and the number of unique colonies with which a queen mates. This model incorporates random, aggregated, and overdispersed spatial distributions of feral colonies and drone congregation areas. We model different densities of feral colonies and drone congregation areas with various numbers of test queens and determine the range of counts of matings with drones from unique colonies. The model shows that the range of counts for a given set of parameters is consistent with a Poisson distribution, and that the most significant explanatory variable is density of the feral colonies. The results obtained reveal that ten or more test queens may be needed in field studies to resolve order of magnitude differences in feral colony densities. We conclude that the new genetic methods provide a powerful tool for making indirect inferences about
feral colony densities, so long as the field survey results are treated as outcomes of a stochastic spatial process.

4.2.2 Introduction

To determine the relationship between feral colony densities and the number of unique colonies determined by indirect sampling, we present an agent-based model (ABM) of the spatial process of honey bee mating. It is possible to construct a simple deterministic model of a single mating flight of a test queen, and derive a sampling distribution for the count of unique colonies with which that queen mated. However, to do so would be unrealistic because queens normally take several mating flights, to either the same or different DCAs. It is at this point that a simple deterministic model breaks down; such a model cannot take into consideration the effects of returning to a DCA nearby to one visited previously. Nearness is a spatial concept that can only be dealt with adequately in an agent-based model. The quantitative effect of visiting nearby DCAs is that the count of unique colonies with which the queen mated is smaller, as many of the same feral colonies will be represented at both DCAs. It is only in an agent-based model that this important effect can be quantified.

We use the model to quantify how the density and spatial distribution of feral colonies and drone congregation areas (DCAs) within the honey bee mating environment influences the number of unique colonies with which a queen will mate, and the proportion of colonies in the environment that are likely to be ‘captured’ (i.e., represented) in the test queen’s progeny.
Chapter 4: Agent-Based Models of Honey Bee Behaviour

We summarise results across multiple simulations and the results are analysed to determine which variables best explain the variation in the resulting counts of unique colony matings. In particular, we explore the hypothesis that the indirect sampling technique is primarily sensitive to variations in feral colony density rather than any other spatial characteristic of the honey bee mating environment. Finally, we make recommendations for the design of field experiments, and provide sampling distributions to interpret the results of those field experiments.

4.2.3 Materials and methods

The model description below follows the ODD (Overview, Design concepts, Details) protocol for describing individual- and agent-based models (Grimm et al., 2006, Grimm et al., 2010).

4.2.3.1 Purpose

The purpose of the agent based model is to understand how the spatial configuration of a honey bee mating environment influences the count of unique feral colonies with which a test queen mates. More specifically, we aim to determine the relationship between the underlying parameters used to generate various spatial configurations and the count itself. The mating environment consists of (i) feral colonies with a given density and characteristic spatial distribution and (ii) DCAs with a given density and characteristic spatial distribution.

4.2.3.2 Entities, state variables and scales

The model comprises three types of agents: queens, feral colonies and DCAs. Queens are characterised by an identifier, a current spatial
location, a target number of matings, a target number of mating flights, and a list of DCAs within their flight range. Queens are assigned a maximum flight range of 8 km, based on the maximum mating distance ever recorded of 15 km (Jensen et al., 2005), minus the 7 km maximum flight range of drones (Ruttner and Ruttner, 1972). Queens also hold a list of feral colonies with whose drones they have mated. At initialisation, this list is empty. Descriptions of the queen parameters and default parameter values are provided in Table 7.

Table 7: Entities, state variables and scales

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queen</td>
<td></td>
</tr>
<tr>
<td>Current spatial location</td>
<td>Initially central</td>
</tr>
<tr>
<td>Target number of matings</td>
<td>From a normal distribution with a mean of 14 and standard deviation of 4</td>
</tr>
<tr>
<td>Target number of mating flights</td>
<td>From a uniform distribution where probability of n flights, ( p(n) ) is:</td>
</tr>
<tr>
<td></td>
<td>( p(1) = 0.55 )</td>
</tr>
<tr>
<td></td>
<td>( p(2) = 0.40 )</td>
</tr>
<tr>
<td></td>
<td>( p(3) = 0.05 )</td>
</tr>
<tr>
<td>List of DCAs within flight range</td>
<td>Initially null</td>
</tr>
<tr>
<td>List of IDs of feral colonies mated with</td>
<td>Initially null</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Feral colonies</td>
<td></td>
</tr>
<tr>
<td>Spatial location</td>
<td>Taken from input file</td>
</tr>
<tr>
<td>Number of drones associated with feral colony</td>
<td>From a normal distribution with a mean of 2400 and standard deviation of 1000 (positive values only)</td>
</tr>
<tr>
<td>List of DCAs within flight range</td>
<td>Initially null</td>
</tr>
<tr>
<td>Population level characteristic spatial distribution</td>
<td>Aggregated, random, overdispersed</td>
</tr>
<tr>
<td>Population level density</td>
<td>0.1, 0.5, 1.0, 5.0, 10.0 colonies/km²</td>
</tr>
<tr>
<td>Drone congregation area (DCA)</td>
<td></td>
</tr>
<tr>
<td>Spatial location</td>
<td>Taken from input file</td>
</tr>
<tr>
<td>List of represent feral colonies with drone</td>
<td>Initially null</td>
</tr>
<tr>
<td>count from each</td>
<td>Aggregated, random, overdispersed</td>
</tr>
<tr>
<td>Population level characteristic spatial</td>
<td>0.1, 0.5, 1.0, 5.0, 10.0 colonies/km²</td>
</tr>
<tr>
<td>distribution</td>
<td></td>
</tr>
<tr>
<td>Population level density</td>
<td></td>
</tr>
</tbody>
</table>

A feral colony is characterised by an identifier, a spatial location, a list of DCAs within flight range, and a count of the number of sexually mature...
drones present in the colony (Table 7). Each feral colony carries a number of drones according to a normal distribution with mean of 2400 and standard deviation of 1000 (Winston, 1987). The maximum flight range of drones is set at 7 km (Ruttner and Ruttner, 1972). For each feral colony distances to each DCA within range are converted to a Z-score based on a normal distribution with mean of 0 km and standard deviation of 1.5 km. These Z-scores are then converted to probability densities. The probability densities are rescaled such that the total for all probability densities for DCAs within range adds to 1. These rescaled probability densities are then used to proportionally determine the drone allocation to each DCA. For instance, if there are only two within-range DCAs at distances of 1.5 km (1 standard deviation) and 3 km (2 standard deviations), then this gives probability densities of 0.161 and 0.036 respectively, and for a feral colony with 1000 drones this would proportionally distribute 818 drones to the closer DCA and 182 to the more distant DCA. In this respect, the model replicates the empirical observation that drones are more likely to fly short distances than long distances and thereby populate nearby DCAs preferentially to distant DCAs. The overall population of feral colonies is characterised by a particular spatial distribution and density. The characteristic spatial distribution of feral colonies is either aggregated, random or overdispersed (Table 7).

DCAs are characterised by an identifier, a spatial location, and a table of feral colonies that contribute to the DCA, a feral colony identifier and a count of the drones present from each feral colony. As a simplifying
assumption the DCAs do not have any area, and are considered instead as points. The overall population of DCAs is described by a characteristic spatial distribution and density. Similarly to the feral colonies, the characteristic spatial distribution of DCAs is either aggregated, random or overdispersed (Table 7).

The simulation executes for as long as it takes for the queen to complete her target number of mating flights.

4.2.3.2.1 Determining spatial size
In determining the size for the spaces used in the simulation, the goal is to make the space sufficiently large to avoid edge effects that would impact the model's results. These edge effects can be avoided if each feral colony capable of populating drone congregation areas within flight range of the queen does not have any simulation boundaries within flight range of the drones. This means being located a minimum of 7 km from the edge, and a maximum of 15 km (7 km drone flight range plus 8 km queen flight range) from the centre of the simulation space. Hence the simulation spaces needs to have a minimum radius of 22 km from the centre where the queen is initially located. To meet this requirement, the spatial dimensions chosen for our model are 50 km by 50 km.

4.2.3.3 Process overview and scheduling
The processes built into the model can be grouped into those associated with initialising the spatial configuration of the model, and those associated with running the mating simulation. Time is modelled using discrete time steps; at each time step a scheduled activity occurs. There is no concept of duration in the model; a scheduled activity is started and
completed as quickly as processing allows before moving onto the next time step.

4.2.3.3.1 Pre-initialisation and initialisation

In the pre-initialisation phase (Figure 1), simulation spaces with the desired characteristic spatial distribution and density are generated externally within R (Baddeley and Turner, 2005, R Core Team, 2014). Further detail on how these spaces are generated is available in Section 4.2.3.5.1.

During initialisation (Figure 3), feral colonies and DCAs are created in accordance with the locations specified in the R output files. Each feral colony determines which DCAs are within range and populates these DCAs with its drones according to a normal distribution.

A queen is created at a central location and is initialised with a target number of mating flights and matings. The queen determines which DCAs are within range of this central location.
Figure 3: Pre-initialisation and initialisation process overview of the simulations.
4.2.3.3.2 Simulation of the mating process

The process simulating queen matings is depicted in Figure 4. For each mating flight, the queen selects a DCA according to a uniform random distribution from her list of DCAs within range, moves to this DCA and then mates with the target number of drones for this mating flight. After mating, a drone dies insofar as the count of drones associated with the particular feral colony at the chosen DCA is decreased by one.
Figure 4: Simulation process overview
4.2.3.4 Design concepts

The level of genetic diversity in the offspring of a test queen, as measured by the number of distinct feral colonies with which she mates across all mating flights, emerges as a consequence of the spatial configuration of the mating environment. Through multiple simulations these counts form a distribution.

Our model makes the simplifying assumption that all males present in a test queen’s colony are sampled without error. This assumption is close to reality, provided that the number of worker bees genotyped from each test queen progeny is large (>200). The theory behind this kind of sampling error is well understood (Foster et al., 1999). The probability of not sampling a patriline (i.e., at least one worker daughter of each male in a mating) is given by \((1 - p)^n\) where \(p\) is the proportion of the progeny sired by a male, and \(n\) is the sample size. Thus if a male is the father of 5% of progeny, the probability of not sampling his progeny in a sample of 200 is 0.000003.

The other kind of non-sampling error arises from the possibility that two unrelated males share the same haplotypes by chance rather than because they are brothers. This probability is the frequency of each haplotype in the population. For these studies, geneticists use 4-5 linked microsatellite markers that are highly polyallelic. Thus the probability that two random males in a population would share the same haplotype by chance is vanishingly small and approaches \(\prod q_i\), where \(q_i\) is the frequency of the allele shared by the two random males at the \(i\)th locus.
For example Oldroyd et al. (2011) found that 151 random workers all had unique haplotypes.

For each simulation, the environment is fixed therefore no adaptation occurs. Queens mate with random drones at the mating location. Thus without adaptation, there is no objective sought by the agents and no learning behaviour is incorporated in the model. Furthermore, no predictions are made by the agents about future conditions.

There is an implied sensing behaviour in that drones from the feral colonies and queens can sense the presence of DCAs within flight range, and orientate themselves to these DCAs. This sensing ability is consistent with observable behaviour exhibited by drones and queens in vivo (Loper et al., 1992).

Interactions between agents in the model occur as a consequence of their initial spatial configuration, where feral colonies populate DCAs, queens move towards DCAs, and queens mate at DCAs with drones from the feral colonies. No communication is necessary between queen and drone agents for these interactions to occur.

Within the initialisation phase, stochasticity arises when, for each test queen, a target number of matings and target number of mating flights is chosen. The values are taken from distributions that have been empirically established in field studies (Schlüns et al., 2005). Within the simulation phase, the choice of DCA by the test queen is random, as is the choice of drones with which she mates. The random seed used is explicitly specified and this allows each simulation to be replicated.
Drones are modelled only at the collective level. At DCAs, the drones from a single feral colony are represented by the identity of the feral colony to which they belong, and a count of the number of drones from that colony present at the DCA. Although the queens mate with a single drone at a time, the individual drone is never modelled.

For each combination of inputs and parameters, including the random seed, the recorded observation is the number of unique colonies with which the queen has mated in that simulation run.

4.2.3.5 Initialisation

As the purpose of this model is to examine the relationship between the initial state of the model and the count of matings that occurs as a consequence of this initial state, care has been taken to accurately generate spaces of sufficient size and consistency for each simulation.

4.2.3.5.1 Generating equal density spatial distributions

It is necessary to ensure that for each density of feral colonies the densities at the location where the test queens are initially placed are the same in each simulation space generated. For each space, density is measured as entities/unit², where the entities may be either feral colonies or DCAs and the units are taken to mean kilometres. For the purposes of this experiment, density is defined as the number of entities within flight range (in this case, the flight range of the queen); each point in the space has an associated local density estimate. The local density estimates were computed using a technique called kernel density estimation. Each entity was represented by a point, and a Gaussian (multivariate normal) smoothing kernel of the point pattern was applied with the bandwidth for
the smoothing set to equal the flight range of the queen. The local density estimates vary across the simulation space to differing degrees, depending on the characteristic spatial distribution used to generate the space. To give one example, where the target central point density is 1 feral colony/km², the point densities were found to range from 0.96 to 1.08 feral colonies/km² in the overdispersed case, 0.75 to 1.05 feral colonies/km² in the random case, and 0.46 to 2.29 feral colonies/km² in the aggregated case. In all three cases, the overall number of colonies in the entire space was approximately the same, yet the effect of different characteristic spatial distributions can be seen in the variations of local density estimates. To generate multiple equal-density spatial distributions, a process is used whereby the initial space generated is 100 km by 100 km, and a 50 km by 50 km subregion within this space for which the central point matches the target kernel density is chosen systematically. This process is illustrated in Figure 5. The characteristic spatial distributions are generated using the spatstat library within R with (i) aggregated spaces generated using the rMatClust function, (ii) random spaces generated using the rpoispp function and (iii) overdispersed spaces generated using the SSI function. The range of densities chosen (0.1 colonies/km² – 10 colonies/km²) reflects the three orders of magnitude for feral colony densities most commonly found in field experiments (Morse et al., 1990, Ratnieks et al., 1991, Oldroyd et al., 1997, Goodman and Hepworth, 2004, Moritz et al., 2007, Jaffé et al., 2010). For each combination of entity type (feral colony or DCA), density, and characteristic spatial distribution, 100 unique spaces were generated.
Figure 5: Generating equal density spatial distributions
The locations of entities within the subregions generated according to the process illustrated in Figure 5 are stored in a text file where each entity is represented by an ID and a location. The procedure illustrated in Figure 5 is; (1) generate a characteristic spatial distribution with required density; in this case random with a density of 0.1 colonies/km², (2) calculate kernel density estimates based on number of colonies within 8 km radius, (3) partition the space by selecting areas where the density is between 0.099 and 0.101 colonies/km² (“TRUE”) and select an area of 50 km by 50 km centred on a TRUE grid cell, (4) create a subregion based on this newly selected area, and (5) export locations of all colonies in subregion to a flat text file. The generation of these files is referred to as the pre-initialisation phase in Figure 3. Each simulation is conducted using a different input file for the feral colony locations, and a different input file for the DCA locations.

4.2.3.5.2 Number of matings for each mating flight
Based on the results of Schlüns et al. (2005) for queens taking more than one mating flight, the first mating flight may account for between 40% and 84% of the total number of matings. The uncertainty is due to the small sample size, and no distinction is drawn between queens taking two mating flights and queens taking three mating flights. In our model, the target number of matings for each mating flight is determined from the chosen target number of matings, and chosen target number of mating flights, according to the following heuristic, reflecting the findings of Schlüns et al. (2005). If the target number of mating flights is two, then two thirds of the matings occur on the first flight, and one third on the
second flight. If the target number of mating flights is three, then half the matings occur on the first flight, three tenths on the second flight, and one fifth on the third flight. For the case where there is only one target mating flight, then all target matings occur on this flight.

4.2.3.6 Input data

The model does not use input data to represent time-varying processes.

4.2.3.7 Submodels

The submodel descriptions are included in the process overview described in Section 4.2.3.3.

4.2.4 Results

We present the results of our simulations grouped by the three spatial distributions (random, aggregated and overdispersed) and three densities (0.1/km², 1.0/km² and 10.0/km²) of feral colonies and DCAs (Figure 6 and Figure 7). The counts of unique colonies mated with by each test queen (i.e., the number of feral colonies that contributed at least one male to the mating) are presented for two situations: when the progeny of a single queen are analysed (Figure 6) and when the progeny of 10 queens are analysed (Figure 7).

For Figure 6, the histogram bars show for each recorded mated colony count the proportion of the total simulations that recorded that count; (a) randomly distributed feral colonies and randomly distributed DCAs, (b) aggregated feral colonies and randomly distributed DCAs, (c) overdispersed feral colonies and randomly distributed DCAs, and (d) randomly distributed feral colonies and overdispersed DCAs.
Figure 6: Results of 100 simulations for one test queen. Each histogram summarises 100 simulations for the specified combination of DCA density and feral colony density.

Figure 7 shows, for an experiment utilising 10 test queens, how the count of unique colonies with which these queens mated varies in relation to feral colony density. The shaded grey area represents the region within 2 standard deviations of the mean (thus accounting for 95% of observations) for each of the densities for which simulations were performed. This region can be used to determine the range of actual densities associated with a predicted maximum likelihood density. For
instance, in a field experiment where 10 test queens were used, and the
count of the unique colonies with which these queens mated was 100, the
intercept with interpolated line and shaded region indicates a predicted
feral colony density of 1.5 colonies/km² with actual densities ranging
from 0.7 colonies/km² to 4 colonies/km². The results are for randomly
distributed feral colonies and randomly distributed DCAs.

Figure 7: Results of 100 simulations for 10 test queens.

4.2.4.1 Distribution for samples taken with single queens

The number of colonies captured in the progeny of test queens when a
single queen was mated increased approximately linearly with the log of
the density of feral colonies in the environment (Table 8, Figure 6). The spatial distribution of feral colonies, the spatial distribution of DCAs and the density of DCAs had almost no effect on the mean or standard deviation of the number feral colonies captured in offspring of test queens (Table 8, Figure 6).
Table 8: The effect of the spatial distribution of feral colonies on the number of feral colonies captured in test queen progeny with a single test queen

<table>
<thead>
<tr>
<th>Feral colony density (colonies/km$^2$)</th>
<th>Random, random</th>
<th>Spatial distributions of feral colonies, DCAs</th>
<th>Overdispersed, random</th>
<th>Random, overdispersed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>1</td>
<td>10</td>
<td>0.1</td>
</tr>
<tr>
<td>DCA density (DCAs/km$^2$) = 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies mated with</td>
<td>3.8 (1.9)</td>
<td>10.5 (3.0)</td>
<td>13.1 (3.4)</td>
<td>3.3 (3.2)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.74</td>
<td>0.85</td>
<td>0.78</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>DCA density (DCAs/km$^2$) = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies mated with</td>
<td>4.5 (2.0)</td>
<td>11.4 (2.4)</td>
<td>14.1 (3.8)</td>
<td>4.2 (3.2)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.77</td>
<td>1.00</td>
<td>0.43</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>DCA density (DCAs/km$^2$) = 0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies mated with</td>
<td>4.9 (1.9)</td>
<td>11.9 (3.5)</td>
<td>13.5 (3.9)</td>
<td>4.6 (3.0)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.95</td>
<td>0.45</td>
<td>0.18</td>
<td>p &lt; 0.05</td>
</tr>
</tbody>
</table>
4.2.4.2 Distribution for samples taken with multiple queens

Given that the spatial distribution of feral colonies has no effect on the number of unique feral colonies mated with by test queens, we performed simulations for multiple queens using a random distribution of feral colonies and DCAs only. The simulations recorded a count of unique colonies with which each test queen mated, and generated an aggregated list of the unique colonies with which one or more queens mated. The count of unique colonies mated with by the ten test queens increased approximately linearly with the log of the density of feral colonies in the simulated space (Table 9, Figure 7).
### Table 9: The effect of the spatial distribution of feral colonies on the number of feral colonies captured in test queen progeny with randomly distributed DCAs and 10 test queens

<table>
<thead>
<tr>
<th>Feral colony density (colonies/km²)</th>
<th>0.1</th>
<th>0.5</th>
<th>1</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DCA density (DCAs/km²) = 10</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies mated with</td>
<td>21.7 (4.8)</td>
<td>66.8 (7.2)</td>
<td>89.9 (8.2)</td>
<td>126.3 (10.4)</td>
<td>134.1 (13.3)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.37</td>
<td>0.95</td>
<td>0.98</td>
<td>0.85</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td><strong>DCA density (DCAs/km²) = 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies mated with</td>
<td>22.8 (4.3)</td>
<td>68.1 (6.8)</td>
<td>91.2 (7.8)</td>
<td>128.3 (12.3)</td>
<td>132.1 (9.9)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.90</td>
<td>0.99</td>
<td>1.00</td>
<td>0.11</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>DCA density (DCAs/km²) = 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies mated with</td>
<td>24.0 (4.1)</td>
<td>68.9 (7.7)</td>
<td>91.5 (7.9)</td>
<td>127.6 (9.8)</td>
<td>133.9 (12.5)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.99</td>
<td>0.85</td>
<td>0.99</td>
<td>0.97</td>
<td>0.14</td>
</tr>
<tr>
<td><strong>DCA density (DCAs/km²) = 0.5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies mated with</td>
<td>23.9 (4.3)</td>
<td>68.7 (7.5)</td>
<td>92.6 (8.2)</td>
<td>126.9 (11.2)</td>
<td>133.3 (12.1)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.95</td>
<td>0.90</td>
<td>0.98</td>
<td>0.52</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>DCA density (DCAs/km²) = 0.1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies mated with</td>
<td>22.7 (4.4)</td>
<td>68.3 (7.8)</td>
<td>92.5 (9.5)</td>
<td>126.4 (12.5)</td>
<td>131.4 (10.3)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.86</td>
<td>0.79</td>
<td>0.56</td>
<td>0.06</td>
<td>0.93</td>
</tr>
</tbody>
</table>
4.2.4.3 Proportion of colonies sampled

In considering the proportion of feral colonies sampled in each simulation, we examined the mean count of unique colonies with which the queen mates in relation to the average number of colonies within flight range. Based on the combined flight range of queens and drones, the furthest distance at which a feral colony could be sampled through analysing test queen progeny is 15 km, from an area of 706.9 km². For densities of 0.1, 1 and 10 feral colonies/km², this gives an average number of in-range colonies of 71, 707 and 7069 respectively. The proportion of colonies typically sampled using 1 queen and 10 queens is shown in Table 10. Even though we only sampled a small percentage of the feral colonies at higher densities, we are still able to make sound inferences about order of magnitude differences in the feral colony densities (Table 10). For any given number of test queens there is a ceiling on the maximum number of unique colonies that could theoretically be recorded in the offspring. For 10 test queens, assuming each of the queens mates 25 times (already an unlikely scenario), each time with a drone from a different unique colony (also very unlikely), a maximum of 250 feral colonies could be captured in the progeny of the test queens. Even such an unlikely occurrence would still only sample a small proportion (3.5%) of the in-range feral colonies at colony densities of 10 feral colonies/km².
Table 10: The effect of test queen number on the proportion of feral colonies ‘captured’ in the progeny of test queens

<table>
<thead>
<tr>
<th>Colony density per km²</th>
<th>Average number of colonies within range</th>
<th>1 queen</th>
<th>10 queens</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Minimum average</td>
<td>Maximum average</td>
</tr>
<tr>
<td>0.1</td>
<td>71</td>
<td>3.3</td>
<td>5.2</td>
</tr>
<tr>
<td>1</td>
<td>707</td>
<td>10.5</td>
<td>11.9</td>
</tr>
<tr>
<td>10</td>
<td>7069</td>
<td>13.1</td>
<td>14.5</td>
</tr>
</tbody>
</table>
4.2.5 Discussion

The results show that the technique of using the progeny of test queens to survey feral colony densities can work well (Moritz et al., 2007, Jaffé et al., 2010), so long as it is understood that what is being sampled is the outcome of a spatial process rather than the underlying spatial distribution itself. The relationship between feral colony density, number of test queens and the resulting distribution of mated colony counts is complex, and for this reason the synthetic sampling distributions generated in this investigation should be used to interpret results obtained in field studies.

It is perhaps surprising that the only variable that has a significant impact on the distribution of mated colony counts is the main variable of interest; namely the underlying feral colony density. The DCA density and the underlying characteristic spatial distributions (random, aggregated, and overdispersed) of feral colonies and DCAs have no significant effect on the number of feral colonies observed in the progeny. As the feral colony count varies primarily with feral colony density, this means that the count is sensitive to the number of colonies within flight range. Honey bee mating behaviour, which involves both queens and drones flying to remote drone congregation areas (Loper et al., 1987, Loper et al., 1992) overcomes sensitivity to the particular locations of the colonies supplying queens and drones. The results of this model provide evidence of evolutionary adaptation insofar as the spatial process used by honey bees for mating appears to guarantee outbreeding. Because of the mechanism of sex determination (Page, 1980) and division of labour
(Oldroyd and Fewell, 2007), the fitness of a honey bee colony is highly sensitive to inbreeding. Our simulations suggest that the genetic diversity of males involved in matings is largely unaffected by the density or distribution of male-producing colonies in the environment.

A single test queen is not sufficient to detect order of magnitude differences in feral colony densities. For a single test queen, there is a strong difference in the distribution centre and spread between densities of 0.1 feral colonies/km² and the other two densities, and a weak difference between densities of 1.0 and 10.0 feral colonies/km². In analysing results from a field experiment, it would not be possible to distinguish between an order of magnitude difference in the feral colony density with a single test queen, as the distributions of feral colony counts overlap. For instance, an experimentally observed count of 8 unique matings can be explained by any of the densities used in this investigation. Therefore no inference about feral colony density can be made.

With multiple queens it becomes possible to distinguish between order of magnitude differences in feral colony densities. As shown in Figure 5, for ten test queens it becomes generally possible to distinguish between densities of 0.1, 1 and 10 feral colonies/km². For instance, an observed count of 90 unique feral colonies in the progeny of 10 test queens is consistent with a feral colony density of 1 colony/km², but inconsistent with 0.1 or 10 feral colonies/km². Additional test queens strengthen the ability to resolve order-of-magnitude differences in densities and/or allow more precise estimates of colony density. We therefore
recommend that the progeny of a minimum of 5 test queens be examined for field experiments, and that in areas where high (> 1 colony per km$^2$) feral colony densities are anticipated, there should be at least 10 test queens.

For simulations conducted with a single test queen, the lower bound for the count of unique feral colonies with which the queen has mated across all simulations is zero; that is, no mating occurs. This situation only occurred where the feral colony density was set to its lowest value of 0.1 feral colonies/km$^2$. For other densities mating always occurred. In field experiments, there may be a number of other environmental or biological reasons why mating does not occur even for these higher densities.

The upper bound observed for the number of unique colonies with which a single queen mated was 28. As the number of target matings is normally distributed with a mean of 14 matings and a standard deviation of 4 matings, it is very unlikely ($P = 0.02\%$) for any simulation that a queen would acquire more than 28 matings, and thus the instances in which a count of 28 matings has occurred are highly likely to represent cases where each mating was with a unique colony. A count of 28 matings only occurred in a case where the feral colony density was set to its highest value of 10 feral colonies/km$^2$.

The distribution of mated colony counts is generally consistent with the Poisson distribution. The Poisson distribution describes the expected probability distribution for a quadrat count description of a complete spatial randomness point pattern (O’Sullivan and Unwin, 2003). This
distribution was observed in our experimental results irrespective of whether the underlying point patterns for the feral colonies and DCAs were random, aggregated, or overdispersed. This implies that queen mating randomly samples the feral colonies within range irrespective of spatial configurations.

In the model, differences in the density of DCAs have a negligible effect on the count of unique feral colonies with which the test queen mates. In real environments it may be more difficult for a queen to locate a DCA where there are a low number of DCAs, but on the assumption that the test queen succeeds in locating a DCA, our model shows that the count itself is unlikely to be affected. Thus while the only interactions between feral colony drones and queens occur at DCAs, the density of these DCAs in the model environment has no effect on the number of colonies captured in test queen progeny.

The results of our modelling call into question the implied level of precision with which some of the results from previous fieldwork surveys have been reported (Moritz et al., 2007, Jaffé et al., 2010). As the maximum number of test queens used at any survey site was 10, our model shows that it is not appropriate to report densities to a single decimal point, but that this sampling effort is sufficient to make biologically-relevant estimates of colony densities in the field. Furthermore, the implied linear relationship between the count of unique feral colonies with which the test queens mate and the feral colony densities has been shown not to exist. Rather, the count varies near-linearly with the logarithm of the feral colony density. The implication of
this is that the count is much less sensitive to variations in underlying feral colony density than previously thought. Nonetheless, an order of magnitude estimate of the number of feral colonies is sufficient for most purposes, and we thoroughly endorse the use of this innovative technique.

4.2.6 Conclusions

Indirect sampling of honey bee feral colonies using genetic techniques provides a valuable survey method for determining feral colony densities. Our agent-based model generated sampling distributions which proved the validity of the indirect sampling method for determining feral colony densities. The model output also provides the quantitative link between observed unique mated colony counts and feral colony density that is needed to interpret the data from field surveys using this technique. The results obtained revealed that ten or more test queens may be needed in field studies to resolve order of magnitude differences in feral colony densities.

4.2.7 Acknowledgements

We acknowledge the financial support of the Rural Industries Research and Development Corporation, and a seed grant from the Sustainable Futures program of the University of Sydney.
4.3 Modelling estimates of honey bee (*Apis* spp.) colony density from drones

4.3.1 Abstract

Given reports of declines in populations of pollinators globally, it is increasingly important to develop efficient procedures to assess the density and distribution of honey bee colonies in both agricultural and natural landscapes. One such procedure utilizes the fact that drone honey bees from different colonies congregate in mating leks where they can be conveniently sampled. Genetic analysis of the captured drones can determine the number of colonies contributing to the sampled population. Here, through the use of sampling distributions derived from an agent-based model, we provide an improved procedure for estimating the density of colonies from the number of unique colonies identified from the sampled drones. We present simulations for different spatial environments and densities, and show that the number of unique colonies observed in a sample of drones collected at a drone trap covaries with the density of colonies in range of the sampled drone congregation area in a log-linear manner. As a consequence of this relationship, we find that colony densities from past surveys are likely to be lower than previously reported.

4.3.2 Introduction

As discussed in Section 2.6, in previous studies using the drone capture method (*Moritz et al.,* 2007, *Moritz et al.,* 2008, *Jaffé et al.,* 2010) the density has been determined by dividing the count of unique colonies by 2.5. This is derived from converting the average distance of a drone’s
flight (900 m; Taylor and Rowell, 1988) to an area of 2.5 km². The assumptions underpinning this calculation are that (i) all colonies represented at the DCA are detected in the genotyped drones, (ii) the drone flight range measured in a single experiment (Taylor and Rowell, 1988) is representative of drone flight ranges in different environments and insensitive to variations in underlying densities of colonies or drone congregation areas.

In this study, we examine the validity of these assumptions and make suggestions that will enhance the value of the drone trap survey method. Our approach extends the work of Arundel et al. (2012) who used a simulation of the spatial environment of honey bee colonies to aid the interpretation of survey results. Arundel et al. (2012) showed that for surveys using unmated queens the relationship between the number of colonies observed in the sample and the underlying colony density is log-linear. Here we apply the “virtual ecologist” approach (Zurell et al., 2010) and simulate the drone trap capture process in environments characterised by different spatial distributions and densities of colonies and DCAs. We generate distributions which can be used to determine the relationship between the count of unique drone haplotypes found in a sample and the underlying parameters used to generate the simulation environment. In particular, we explore the hypotheses that the count of unique haplotypes in a sample of drones captured using a Williams trap is primarily related to the underlying density of colonies, and that this relationship is non-linear. To assess the implications of our findings, we use this new method to estimate colony densities from field experiments.
Finally, we make recommendations for the design of field experiments, and provide sampling distributions that should be useful for the interpretation of data.

4.3.3 Materials and methods

The model description below follows the ODD (Overview, Design concepts, Details) protocol for describing individual- and agent-based models (Grimm et al., 2006, Grimm et al., 2010). The submodel descriptions are included in the process overview described in Section 4.3.3.3.

4.3.3.1 Purpose

The purpose of the agent based model is to understand how the spatial distribution of honey bee colonies in a landscape impacts the count of unique haplotypes observed in a sample of drones captured in a drone trap at a DCA. We determine the impacts of (i) the density and spatial distribution of colonies in the landscape and (ii) the density and spatial distribution of DCAs in the landscape.

4.3.3.2 Entities, state variables and scales

The model comprises three types of agents: the drone trap, colonies and DCAs. The drone trap is characterised by a spatial location and a target number of samples to be collected. The drone trap also holds a list of unique colonies represented in the sample. At initialisation, this list is empty. A description of the drone trap parameters and default parameter values is provided in Table 11.
Each colony is characterised by an identifier, a spatial location, a list of DCAs within flight range, and a count of the number of drones present in the colony (Table 11). The number of drones is normally distributed with a mean of 2400 and standard deviation of 1000 (Winston, 1987). The flight range of drones is set at 7 km (Ruttner and Ruttner, 1972). Drones are allocated to DCAs within range such that proximate DCAs are preferred to more distant DCAs (N. Koeniger et al., 2005). This is achieved by proportionally distributing the drones according to a normalised weighting of the distances to each of the DCAs within range (Arundel et al., 2012). As a population, colonies are characterised by a type of spatial distribution and a density. The possible types of spatial distributions of colonies are random, aggregated, hyper-aggregated, and

### Table 11: Entities, state variables and scales

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drone trap</strong></td>
<td></td>
</tr>
<tr>
<td>Current spatial location</td>
<td>Initially central</td>
</tr>
<tr>
<td>Target number of samples</td>
<td>50, 100, 200</td>
</tr>
<tr>
<td>List of IDs of captured feral colonies</td>
<td>Initially null</td>
</tr>
<tr>
<td><strong>Colonies</strong></td>
<td></td>
</tr>
<tr>
<td>Spatial location</td>
<td>Taken from input file</td>
</tr>
<tr>
<td>Number of drones associated with feral colony</td>
<td>From a normal distribution with a mean of 2400 and standard deviation of 1000 (positive values only)</td>
</tr>
<tr>
<td>List of DCAs within flight range</td>
<td>Initially null</td>
</tr>
<tr>
<td>Population level characteristic spatial distribution</td>
<td>Random, aggregated, hyper-aggregated, overdispersed</td>
</tr>
<tr>
<td>Population level density</td>
<td>0.1, 0.5, 1.0, 5.0, 10.0 colonies/km²</td>
</tr>
<tr>
<td><strong>Drone congregation area (DCA)</strong></td>
<td></td>
</tr>
<tr>
<td>Spatial location</td>
<td>Taken from input file</td>
</tr>
<tr>
<td>List of feral colonies within flight range</td>
<td>Initially null</td>
</tr>
<tr>
<td>List of represented feral colonies with drone count from each</td>
<td>Initially null</td>
</tr>
<tr>
<td>Population level characteristic spatial distribution</td>
<td>Random, aggregated, overdispersed</td>
</tr>
<tr>
<td>Population level density</td>
<td>0.1, 0.5, 1.0, 5.0, 10.0 colonies/km²</td>
</tr>
</tbody>
</table>
overdispersed, where overdispersed means that the colonies are more evenly spaced than would be expected by chance. (Table 11).

Each DCA is characterised by an identifier, a spatial location, a list of colonies within flight range, and a table listing the count of the drones present at the DCA from each colony within range. For simplicity, the DCAs are modelled as points rather than areas. As a population, DCAs are characterised by a type of spatial distribution and a density. Similarly to colonies, the possible types of spatial distributions are aggregated, random or overdispersed (Table 11).

4.3.3.2.1 Determining spatial size
The size of the spaces used in the simulation must ensure that edge effects are avoided at the location of the drone trap. The drone trap is located within 7 km of the central point of the simulated landscape. From the drone trap location at a particular DCA, the edge of the simulation space must be at least a further 14 km distant as this represents the diameter of the area covered by a colony only just within flight range (7 km) of the DCA. Hence the simulation spaces need a minimum radius of 21 km from the centre of the space. For consistency with Arundel et al. (2012), we have chosen to generate simulated landscapes of size 50 km by 50 km.

4.3.3.3 Process overview and scheduling
Processes are defined for initialising the spatial configuration, and for running the simulation of the drone trap capture itself. Time is modelled as discrete steps with scheduled activity occurring at each step.
4.3.3.3.1 Pre-initialisation
In the pre-initialisation phase (Figure 8), simulation spaces with the desired characteristic spatial distribution and density of colonies and DCAs are generated in R (Baddeley and Turner, 2005, R Core Team, 2014), according to the methodology of Arundel et al. (2012).

During initialisation (4.3.3.5), these R files are imported and feral colonies and DCAs are created in accordance with the locations specified therein. A drone trap agent is created and moved to a central location.
Figure 8: Flowchart for pre-initialisation and initialisation process
4.3.3.3.2 Simulation of the drone trap capture process

We developed an agent-based model simulating the drone capture process using the Repast Simphony modelling environment (North et al., 2013). The process simulating capture of drones in a drone trap is depicted in Figure 9. The drone trap is initially moved from its central location to the nearest DCA. This simulates the experimenter using the drone trap to locate a nearby DCA (Brockmann et al., 2006, Jaffé et al., 2009). Before capture can proceed, the DCA must be initialised with drones from the surrounding colonies. The DCA agent determines which colonies are within range. Each of these colonies then conversely determines which DCAs are within flight range, and populates the DCAs according to their proximity. Although more than one DCA is initialised in this process, the only DCA guaranteed its full share of drones is the DCA co-located with the drone trap. While this process introduces some small additional complexity to the simulation compared to fully initialising all DCAs, it provides the advantage of minimising the computation required. Once the DCA where the trap is located has been initialised, the specified sample size of drones is captured and their colony origins recorded.
Figure 9: Flowchart for simulation process
4.3.3.4 Design concepts

4.3.3.4.1 Basic principles

The basic principle underlying the model’s design is that the mix of drones present at any DCA emerges as a consequence of the particular spatial environment in which that DCA is located. Baudry et al. (1998) used a truncated Poisson distribution to estimate the number of colonies present at a DCA based on the number of unique colonies observed in a sample of a given sample size. While this approach is useful, it relies on the assumption that all colonies send, on average, the same number of drones to a DCA. Our model extends the analysis for spatial environments with known configurations. Specifically, the factors to be investigated are the density and characteristic spatial distributions (random, aggregated, hyper-aggregated, overdispersed) of both colonies and DCAs, and what effect these have on the number of unique colonies observed in a drone trap sample of a given size. By performing 100 simulations for each combination of factors, we can identify the factors to which the count is sensitive and interpret the counts of drone haplotypes observed in field experiments as colony densities. To illustrate the utility of this approach, we re-estimate the densities from selected field experiments as listed in Table 4.

4.3.3.4.2 Emergence

As well as investigating the emergence of the count of the density of colonies in a landscape from underlying factors, we investigate how the drone flight distances are distributed as a consequence of the spatial configurations. The distribution emerging from our simulation will be compared with observations from field experiments.
4.3.3.4.3 Adaptation, objectives, learning and predictions
As the simulation environment is fixed for the duration of the simulation, no adaptation occurs. Agents do not have any objective or goal set, and do not learn through the simulation or make predictions.

4.3.3.4.4 Sensing
Sensing of the spatial environment is implied for all agents. The drone trap senses the nearest DCA, simulating the process in the field where an experimenter uses the drone trap as a DCA detector. A colony senses the DCAs within range through a collective intelligence, and distributes drones to each DCA based on proximity. Reciprocally, the DCA itself is able to sense the colonies within flight range, although this is a convenience for the simulation rather than a model of the real world.

4.3.3.4.5 Interaction
_agents interact through their spatial proximity, but do not require any communication in order for this to happen.

4.3.3.4.6 Stochasticity
Stochasticity arises from the generation of spatial environments according to random processes, and from the random sampling of drones at the DCA by the drone trap. The particular spatial environments used in each simulation run are pre-specified.

4.3.3.4.7 Collectives
Individual drones are not modelled, so the drones associated with a colony and with a DCA are modelled at the collective level.
4.3.3.4.8 Observation
The observations recorded from each simulation are the number of
unique colonies in the drone sample, and the average flight distance of all
drones present at the DCA.

4.3.3.5 Initialisation
For each simulation from a series of 100 simulations, the model is
initialised with (i) a particular configuration of colonies with
characteristic spatial distribution and density, (ii) a particular
configuration of DCAs with characteristic spatial distribution and density,
and (iii) a target sample size representing the number of drones to be
captured at the drone trap. The individual agents are initialised
according to the description in Section 2.3.

4.3.3.5.1 Generating equal density spatial distributions
The spatial distributions of colonies for each density were generated as in
Arundel et al. (2012) using the spatstat (Baddeley and Turner, 2005)
library within R. Aggregated colony distributions were generated using
the rMatClust function, random distributions using the rpoispp function
and overdispersed distributions using the rSSI function. In addition, we
have simulated hyper-aggregated distributions, as are seen in the giant
honey bee *Apis dorsata* (Deodikar *et al.*, 1977, Oldroyd *et al.*, 2000,
generated hyper-aggregated colony distributions using the rMatClust
function with a mean number of 100 points per cluster, and a cluster
radius of 0.05 km. This additional simulation will allow interpretation of
any future study on the density of *A. dorsata* colonies. Such studies are
sorely needed because *A. dorsata* is heavily hunted, unsustainably in some areas (Oldroyd and Wongsiri, 2006, Oldroyd and Nanork, 2009).

4.3.3.6 Input data

The model does not use input data to represent time-varying processes.

4.3.3.7 Submodels

The submodel descriptions are included in the process overview described in Section 4.3.3.3.

4.3.4 Results

We simulated five densities (0.1/km², 0.5/km², 1.0/km², 5.0/km² and 10.0/km²) of colonies and DCAs with the four spatial distributions (random, aggregated, hyper-aggregated and overdispersed). For each combination of parameters, we executed 100 simulations and we present the results of these simulations as distributions.

4.3.4.1 Distribution of unique colonies present in the sample

The count of unique colonies present in each sample is approximated by a log-linear relationship with the density of colonies in the simulated landscape (Table 12, Figure 3, Figure 4, Figure 44 - Figure 47). The density and spatial distribution of DCAs had no appreciable effect on the number of colonies represented in the trap by their drones. The spatial distribution of colonies had a small effect on the spread of estimated colony densities, but not on the mean value (Table 12).
Table 12: The effect of the spatial distribution of feral colonies on the number of unique colonies observed in a sample of 96 captured drones

<table>
<thead>
<tr>
<th>Feral colony density (colonies/km²)</th>
<th>Random, random</th>
<th>Aggregated, random</th>
<th>Hyper-aggregated, random</th>
<th>Overdispersed, random</th>
<th>Random, aggregated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>1</td>
<td>10</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>DCA density (DCAs/km²) = 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies present in the sample</td>
<td>3.9 (2.0)</td>
<td>31.0 (3.7)</td>
<td>78.8 (3.2)</td>
<td>3.9 (4.5)</td>
<td>25.4 (15.3)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.15</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>DCA density (DCAs/km²) = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies present in the sample</td>
<td>6.3 (2.1)</td>
<td>33.4 (4.1)</td>
<td>80.2 (3.4)</td>
<td>5.9 (4.9)</td>
<td>31.6 (11.9)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.28</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>DCA density (DCAs/km²) = 0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD) of count of unique colonies present in the sample</td>
<td>6.3 (2.4)</td>
<td>36.8 (4.2)</td>
<td>82.5 (4.1)</td>
<td>7.6 (4.6)</td>
<td>35.0 (12.1)</td>
</tr>
<tr>
<td>Poisson dispersion test</td>
<td>0.32</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.05</td>
</tr>
</tbody>
</table>
Figure 10 shows the results of 100 simulations for a sample size of 96 drones with randomly distributed colonies and randomly distributed DCAs. Each histogram summarises 100 simulations for the specified combination of DCA density and colony density. The histogram bars show for each recorded count of unique colonies in the sample, what proportion of the total simulations recorded that count.
Figure 11: Log-linear relationship of unique colonies observed in a sample with colony density

Figure 11 shows that the number of unique colonies observed in a sample of drones collected at a drone trap varies in a log-linear relationship with colony density in range of the sampled drone congregation area. The shaded grey area represents the region within 2 standard deviations of the mean (thus accounting for 95% of observations) for each of the densities for which simulations were performed. This region can be used to determine the range of actual densities associated with a predicted
maximum likelihood density. The results are for aggregated colonies and randomly distributed DCAs.

The ability to distinguish between the underlying densities of colonies in the simulated landscape from the count of unique colonies represented by the presence of their drones in the trap increases with increasing sample size, as would be expected from statistical theory. With a sample size of 96 drones it is possible to distinguish order-of-magnitude differences in the density of colonies per square kilometre. In contrast, with a sample size of 15 drones, the minimum number drones with which an *A. mellifera* queen typically mates (Palmer and Oldroyd, 2000), it is not possible to distinguish between order-of-magnitude differences in densities when the actual number of colonies present in a landscape is greater than 0.5 colonies/km² (Figure 48 - Figure 49).

4.3.4.2 Distribution of mean drone flight distances

The distribution of the mean distance flown by drones present at the sampled DCA is presented in Figure 50 - Figure 51. The average of the means is 2.25 km, and this value is not affected by the density or distribution of DCAs. The variance is strongly dependent on the density of colonies, with the variance decreasing as the density increases.

4.3.4.3 Distribution of the proportion of colonies sampled

Colonies can be present in a sample, or present at the DCA but not in the sample (Figure 41 - Figure 43). The proportion of colonies sampled is dependent primarily on the density of colonies and the sample size, but is also affected by the underlying characteristic spatial distribution of
colonies. The proportion of colonies detected varies depending on the density and sample size (Figure 52 - Figure 54).

4.3.4.4 Reanalysis of field surveys conducted with a sample size of 96 drones

Researchers typically analyse 96 drones per sampling site because this is the number of PCR reactions that fits in a standard plate (Table 4). Using our model, and assuming a random spatial distribution of colonies and DCAs, we have simulated the expected density of colonies in the landscape based on the reported number of unique colonies represented in the sampled traps. Using the sampling distribution of these estimates, we present revised density estimates for each of the observed counts of unique colonies (Table 13). Because of our more realistic assumptions about the mating behaviour of drones, these revised estimates are an order-of-magnitude less than the published estimates.
### Table 13: Re-analysis of surveys with sample size of 96 drones

<table>
<thead>
<tr>
<th>Country</th>
<th>Surveyed location</th>
<th>Estimated unique colonies</th>
<th>Published density estimate (colonies/km²)</th>
<th>Revised density estimate (colonies/km²)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Croatia</td>
<td>Pokupsko</td>
<td>15</td>
<td>5.9</td>
<td>0.3</td>
<td>Jaffé et al. (2010)</td>
</tr>
<tr>
<td>Croatia</td>
<td>Svinjicko</td>
<td>17</td>
<td>6.7</td>
<td>0.4</td>
<td>Jaffé et al. (2010)</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>Belfast</td>
<td>10</td>
<td>3.9</td>
<td>0.2</td>
<td>Jaffé et al. (2010)</td>
</tr>
<tr>
<td>Republic of Ireland</td>
<td>Caher</td>
<td>7</td>
<td>2.8</td>
<td>0.1</td>
<td>Jaffé et al. (2010)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Ezemvelo</td>
<td>24</td>
<td>9.4</td>
<td>0.6</td>
<td>Jaffé et al. (2010)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Ezemvelo, March 2004</td>
<td>46.7</td>
<td>N/A</td>
<td>2</td>
<td>Jaffé et al. (2009)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Ezemvelo, November 2004</td>
<td>51.96</td>
<td>N/A</td>
<td>3</td>
<td>Jaffé et al. (2009)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Jonkershoek</td>
<td>26</td>
<td>10.2</td>
<td>0.6</td>
<td>Jaffé et al. (2010)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Leeuwfontein</td>
<td>34</td>
<td>13.6</td>
<td>0.9</td>
<td>Moritz et al. (2007)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Leeuwfontein</td>
<td>17</td>
<td>6.7</td>
<td>0.4</td>
<td>Jaffé et al. (2010)</td>
</tr>
<tr>
<td>South Africa</td>
<td>Pietermaritzburg</td>
<td>25</td>
<td>9.8</td>
<td>0.6</td>
<td>Jaffé et al. (2010)</td>
</tr>
</tbody>
</table>
4.3.5 Discussion

The count of unique haplotypes present in a sample of drones captured in a trap is strongly related to the underlying density of colonies in the environment. We find a log-linear relationship between the count of unique colonies present in a trap sample and the true density of colonies in the environment. Crucially, the most likely density that would result in the count of unique colonies observed from the sample must be estimated from a distribution specific to the sample size. Thus, the use of drone traps for determining colony density does not require prior knowledge about the likely location, distribution, or density of colonies.

Except in the case of hyper-aggregation, the spatial distribution of colonies in the landscape does not affect the distribution of mean estimates of colony density, but does impact the variance. The smallest variance in estimates occurs when colonies are overdispersed, and the largest variance occurs when colonies are aggregated. Honey bee colonies, especially Asian species, are often aggregated in a landscape (Deodikar et al., 1977, Oldroyd et al., 1995, McNally and Schneider, 1996, Oldroyd et al., 2000, Rinderer et al., 2002, Paar et al., 2004, Baum et al., 2005, Baum et al., 2008, Rattanawannee et al., 2013). Thus the likelihood of aggregation needs to be recognised when choosing the appropriate sample size to infer colony density. The extreme aggregation distributions are bi-modal as a consequence of an artefact of the model. All of the simulations at 0.1 colonies/km², and a significant proportion of those at 0.5 and 1 colony/km², failed to capture any drones as the DCA chosen was devoid of drones. This situation arises in the model because
DCA locations are randomly created independently of colony locations. After the adjustment to re-centre the spatial distributions is performed in the pre-initialisation stage (Arundel et al., 2012), the requirement to maintain the desired lower densities necessitates that the colony aggregations are placed far from the centre of the simulated space. Ignoring this artefact, the remaining results indicate that a sample from any DCA within range of an aggregation is likely to yield a large proportion of unique colonies irrespective of the underlying colony density. One consequence of this finding is that the technique of trapping drones at a DCA to estimate colony densities is unlikely to be directly useful for *A. dorsata* as the distributions at different densities exhibit substantial overlap.

Returning to *A. mellifera* and other species that do not have hyper-aggregated colonies, our results suggest that a sample size of 96 drones is sufficient to detect order-of-magnitude differences in colony density. This is not withstanding the fact that at high densities of colonies (greater than 5 colonies/km$^2$), the proportion of colonies sampled at a trap will be less than 15%. For more precise estimates, a larger sample size should be used. Based on our results (Table 12, Figure 3, Figure 44 - Figure 49) we suggest a sample size of 192 drones if it is suspected that the number of colonies in a landscape is > 5 colonies/km$^2$.

From a sample from a single DCA in Oberursel, Germany, Baudry et al. (1998) identified 107 unique colonies represented by the 142 sampled drones and estimated the presence of 238 colonies at the DCA based on the assumption that the number of unique colonies present in a sample is
Poisson distributed. Our simulations predict that 107 colonies would be
detected from a sample of 142 drones when the density approximates 10
colonies/km². For this scenario, the mean number of colonies
represented at the DCA predicted by the simulation is approximately
1550. This is much higher than the estimate of Baudry et al. (1998) and
arises because the model allocates drones from each colony to all DCAs
within the 7 km flight range. However, while every colony within range is
represented at the DCA in the simulations, representation from the
closest colonies is greatest.

The initial location of a trap is unlikely to be close enough to a DCA for the
9-ODA attractant to be effective, as the range of 9-ODA is less than 100 m.
This gives rise to the need to first locate a DCA and then sample from that
DCA. In our model we sample only from the chosen DCA, even though in
some circumstances there may other DCAs also within the 9-ODA range
threshold. Our justification for excluding samples from these other DCAs
is that any such DCA in our model would have a very similar population of
males, and can effectively be regarded as a single DCA.

Previously published estimates of densities from the capture of drones in
traps rely on the observations of Taylor and Rowell (1988) that drones fly
on average 900 m and have assumed on this basis that the flight range of
drones is 2.5 km². 900 m is actually the minimum queen flight distance
(Taylor and Rowell, 1988) and the average flight distance of males is
more likely to be ~1040 m based on the Taylor and Rowell estimates. We
argue that average area covered by a colony's drones cannot be derived
directly from the average flight distance, but must be derived from the
distribution of distances. In the case of Taylor and Rowell (1988), this area is closer to 5 km$^2$ than 2.5 km$^2$. Thus we suggest that many published estimates of colony density are inflated by a factor of 2, even without changing the analysis approach. The distribution of flight distances emerging from our model fits better the data of Taylor and Rowell (1988) when using a standard deviation of 1.1 km for the variability in flight range of individual drones rather than the 1.5 km used in our simulation. In our model, the standard deviation of drone flight distances can be interpreted as the extent to which drones prefer nearer rather than further DCAs. Irrespective of which value is used, the distribution of flight distances for drones present at the DCA arises not just from the extent to which drones prefer nearer DCAs, but also the proximity of those DCAs. In other words, just because drones prefer to fly shorter distances does not mean they always do so. Based on the persistence of DCA locations from year to year, irrespective of changes in population size, we made the assumption in our model that DCA locations are independent of colony locations and densities. Thus in our model, the distribution of flight distances arises as a consequence of the spatial environment.

Our re-analysis of previous survey results highlights that the new sampling distributions are able to be applied to previously published results and improve the accuracy of the density estimates. It also allows estimates of densities to be made for studies where this analysis was not previously performed (Baudry et al., 1998, Jaffé et al., 2009, Moritz et al., 2013). On the basis that the published densities are an order of
magnitude lower than originally indicated, the conclusions drawn about the decline in honey bee populations worldwide need to be reassessed. The largest analysis of feral colony densities where the location of the colonies was physically verified was undertaken by Baum et al. (2008) in a survey of the city of Tucson, Arizona with an area of ~900 km². Baum et al. (2008) report densities of 0.36 to 1.12 colonies/km², and the re-analysed density estimates applied using sampling distributions for drone traps gives densities more consistent with this range (0.1 to 3 colonies/km²).

As a final observation, an *A. mellifera* queen typically mates with 15 - 20 drones (Palmer and Oldroyd, 2000) and thus Figure 48 could also be considered to represent the number of unique ‘colonies’ that a queen will mate with. The mechanism of mating at DCAs seems to ensure that the queen samples the population with a similar level of genetic diversity across a two order of magnitude range of colony density. This effect is even more pronounced with the extreme colony aggregations typical of *A. dorsata*. A potential area for further research is an analytical comparison of the differently evolved mating model with other candidate models. Of particular interest is whether the evolved mating system optimises genetic diversity in a queen’s offspring in a range of different spatial environments.

### 4.3.6 Conclusion

The number of unique colonies observed in a sample of drones collected at a drone trap varies in a log-linear relationship with the density of
colonies in range of the sampled DCA. This finding validates the usefulness of the drone trap technique as an important survey tool, but highlights the challenges of interpreting field survey results. Typically, unless the sample size is large and the underlying density small, the proportion of colonies captured by the sample will be low. Our sampling distributions, generated from agent-based models, provide a significant improvement in the ability to estimate colony densities from drone samples. This is demonstrated through a re-analysis of published survey results using the new sampling distributions which show that colony densities are lower than previously thought, and more similar to those reported from direct field observation. This highlights the importance of further research into honey bee populations worldwide, and the ongoing development of genetic methods to support population surveys. Our recommendation for field surveys is that a sample size of at least 192 drones is used to estimate densities of up to 10.0 colonies/km².

4.3.7 Acknowledgements

We acknowledge the financial support of the Rural Industries Research and Development Corporation, and a seed grant from the Sustainable Futures program of the University of Sydney.
5 THE SPATIO-TEMPORAL DISTRIBUTION OF FERAL COLONIES

5.1 Introduction

This chapter applies the sampling distributions from Chapter 4 to two large-scale field surveys of feral colonies in Victoria, Australia. The first of these was conducted using the queen mating method, and the results published in the journal Austral Entomology. This chapter reproduces the paper from Austral Entomology:


My contribution to this paper is outlined in the Preface. The literature review of has been incorporated into Chapter 2 and Chapter 3, and thus removed from this chapter.
Sampling distributions for the drone capture method were also used to interpret field surveys, as published in the following paper:


A brief summary of my contribution to this paper as a co-author is featured in this chapter, along with a discussion of its key findings.

5.2 Remarkable uniformity in the densities of feral honey bee (*Apis mellifera*) colonies in South Eastern Australia

5.2.1 Abstract

It is often assumed that the density of feral honey bee colonies in Australia is sufficient to provide adequate pollination services to the many agricultural crops that require pollination. In contrast, there is concern that the density of feral colonies is sufficiently high to have inimical effects on Australian biota. For both these reasons it is desirable to have robust estimates of the density of feral honey bee colonies in Australian landscapes. In this study we mated 4-5 queens with wild drones at disturbed and undisturbed sites in three of the major ecosystems in Victoria Australia, and examined the paternities of worker offspring to estimate the density of feral colonies within mating range of the test queens. We show that the density of feral colonies differs little with land use (cleared or uncleared), and is similar across the state. Our data suggest that the density of feral colonies is probably insufficient to
provide adequate pollination of agricultural crops and that neither land use nor local climate variation is a major factor determining density. Finally, our data suggest that the mating range sampled by test queens is significantly greater than previously assumed.

5.2.2 Introduction

In this study we mated a total of 29 queens at three locations in the state of Victoria, Australia (Table 14). The locations encompassed three of the major ecosystems in Victoria. (1) The sub-alpine wet sclerophyll forest between 600 and 1300 m above sea level. This forest type is dominated by mountain ash (E. regnans), and is irregularly exploited by beekeepers for species such as messmate (E. obliqua). The region provides important habitat for over 40 species of native arboreal vertebrates (Lumsden, 1991), that may compete with honey bees for nest sites (Lindenmayer et al., 2009). (2) The central box-ironbark forest. This region is between 150-600 m above sea level, and is characterized by a sparse canopy of box, ironbark and smooth-barked eucalypts. The ecosystem is an important honey producing region from species such as yellow box (E. melliodora), grey box (E. microcarpa) and red ironbark (E. tricarpa). (3) The mallee. This low-rainfall region is dominated by short (5-8 m) multi-stemmed eucalypts. It is a major honey producing region from species such as dumosa mallee (E. dumosa), yellow mallee (E. costata) and white mallee (E. gracilis). Along water courses beekeepers exploit species such as black box (E. largiflorens) and river gum (E. camaldulensis).
At each location we established two mating apiaries, one in a disturbed habitat and one in as undisturbed habitat as possible. To infer the range of possible feral colony densities from the results of the genetic analysis, we use the sampling distributions developed by Arundel et al. (2012). Our goal was to determine if feral honey bee colonies are present at similar densities throughout Victoria, and to examine the effect of land use on feral nest density.
Table 14: Locations of mating sites, listed from north to south

<table>
<thead>
<tr>
<th>Site Name</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Land use</th>
<th>Disturbed/ undisturbed</th>
<th>Nearest mating site</th>
<th>Distance between paired sites</th>
<th>Ecosystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wyperfeld</td>
<td>35°34'43.94&quot;S</td>
<td>142° 1'4.93&quot;E</td>
<td>National park</td>
<td>Undisturbed</td>
<td>Yaapeet</td>
<td>15 km</td>
<td>Mallee</td>
</tr>
<tr>
<td>Yaapeet</td>
<td>35°42'23.97&quot;S</td>
<td>142° 3'58.57&quot;E</td>
<td>Cereal and canola farming</td>
<td>Disturbed</td>
<td>Wyperfeld</td>
<td>15 km</td>
<td></td>
</tr>
<tr>
<td>Dookie</td>
<td>36°23'41.12&quot;S</td>
<td>145°43'56.17&quot;E</td>
<td>Mixed farming/grazing</td>
<td>Disturbed</td>
<td>Puckapunyal</td>
<td>85 km</td>
<td></td>
</tr>
<tr>
<td>Dookie East</td>
<td>36°53'16.29&quot;S</td>
<td>144°59'56.16&quot;E</td>
<td>Military base</td>
<td>Undisturbed</td>
<td>Puckapunyal North</td>
<td>12 km</td>
<td>Box-ironbark</td>
</tr>
<tr>
<td>Puckapunyal North</td>
<td>36°56'45.79&quot;S</td>
<td>144°53'33.01&quot;E</td>
<td>Military base</td>
<td>Undisturbed</td>
<td>Puckapunyal South</td>
<td>2 km</td>
<td></td>
</tr>
<tr>
<td>Puckapunyal South</td>
<td>36°57'29.24&quot;S</td>
<td>144°53'8.29&quot;E</td>
<td>Military base</td>
<td>Undisturbed</td>
<td>Puckapunyal North</td>
<td>2 km</td>
<td></td>
</tr>
<tr>
<td>Eildon</td>
<td>37°18'32.45&quot;S</td>
<td>145°55'25.83&quot;E</td>
<td>Forestry</td>
<td>Undisturbed</td>
<td>Marysville</td>
<td>30 km</td>
<td>Wet sclerophyll alpine forest</td>
</tr>
<tr>
<td>Marysville</td>
<td>37°32'30.68&quot;S</td>
<td>145°44'45.61&quot;E</td>
<td>Forestry</td>
<td>Disturbed (burnt)</td>
<td>Eildon</td>
<td>30 km</td>
<td></td>
</tr>
</tbody>
</table>
5.2.3 Materials and Methods

5.2.3.1 Establishment of mating nuclei

Mating nuclei (n = 38) were established in early October 2009 in Castlemaine, Victoria. To create a nucleus, three frames containing brood and one frame containing honey were removed from a strong colony. All bees were removed from the frames, and the frames returned to the colony in a box placed above a queen excluder on top of the source colony. Workers moved up onto the brood frames through the queen excluder, but the larger queen and drones could not.

After 24 hours the four frames, now densely covered exclusively with workers, were transferred to a standard nucleus hive. During this transfer, any drone pupae were destroyed, and the nucleus colony was furnished with a queen cell that was about to emerge.

The nuclei were then transported to the seven sites (Table 14) where they were left for 21-30 days. During this time the queens (hereafter ‘test queens’) emerged, mated, laid eggs and produced a first batch of brood. Some test queens failed to emerge from their cells – in most of these cases the nucleus colonies successfully reared a queen from their own female brood. We treated the offspring of these self-raised queens in the same way as the artificially-reared test queens. Worker brood were successfully reared from 29 of the test queens. Brood (n = ~200 pupae per queen) were harvested and stored in 100% ethanol.
5.2.3.2 Location of the sites

Sites were paired to reflect disturbed and undisturbed habitat within the same landscapes (Table 14). In the mallee region, we placed colonies within Wyperfeld National Park, more than 4 km from the park boundary. Commercial beekeeping has been excluded from the park for over 40 years, yet it has very high (up to 160 colonies/km²) densities of feral bee colonies in the red gum - black box forests that line Outlet Creek (Oldroyd et al., 1997). The site chosen for the mating, at Black Flat, was repeatedly surveyed for the presence of feral colonies for the period 1993-1995 by Oldroyd et al. (Oldroyd et al., 1994, 1995, 1997). A matched, disturbed site was selected 5 km outside the park towards the township of Yaapeet. Here the predominant mallee vegetation has been cleared to make way for wheat, barley, oats and canola cropping. Remnant mallee vegetation remains along fence lines and road verges.

Within the box-ironbark region of central Victoria we established two sites (Puckapunyal North and South) 2 km apart within the 400 km² Puckapunyal army training base, and a third site, Puckapunyal East, 12 km north east of Puckapunyal North and South. Because of military training exercises on the base, there has been no logging or beekeeping activity there since the 1970s. We established adjacent sites at Puckapunyal so that we could replicate our sampling (Table 14) and determine if queens mating 2 km apart would mate with drones from the same colonies. We established a matching disturbed site within the box-ironbark region at Dookie Agricultural College. At this site only remnant
tree cover remains. Principal agricultural activities include cattle grazing and mixed cropping.

The remaining two sites were established in wet sclerophyll alpine forest in south east Victoria. One site near the township of Marysville had been heavily burned during the widespread bushfires of January 2009, and we classified this site as disturbed. At the Marysville site, all trees and ground flora had been severely burnt or destroyed by the fire, severely limiting the forage available for bees. The other site, near Eildon, was largely unaffected by the 2009 bushfires. At this site, approximately 20% of the vegetation within 5 km of the site was burnt in the 2009 bushfires, and the remainder had not been burnt in the last 10 years. The extent of the 2009 bushfires can be viewed online at http://nremap-sc.nre.vic.gov.au/MapShare.v2/imf.jsp?site=forestexplorer (Department of Environment and Primary Industries, 2013b).

5.2.3.3 DNA extraction and genotyping

DNA was extracted from 48 progeny from each queen. A single hind leg from each pupa was used to extract DNA using a high-salt precipitation method (Aljanabi and Martinez, 1997). Each pupa was then genotyped at two sets of linked microsatellites (Shaibi et al., 2008): UN351, HB-SEX-01, HB-SEX-02 and HB-SEX-03 (hereafter ‘LG-3’), and HB-THE-01, HB-THE-02, HB-THE-03 and HB-THE-04 (hereafter ‘LG-13’). PCRs were performed using the conditions given in Shaibi et al. (2008). PCR products were fluorescently labelled and run on an Applied Biosystems 3130xl Genetic Analyzer. Genotyping was performed using Genemapper software (Applied Biosystems). Measures of allelic richness and expected
heterozygosity were determined for each site using GENEPOP (Raymond and Rousset, 1995) based on the inferred paternal allele of each genotyped worker.

5.2.3.4 Analysis

A test queen carries two haplotypes at both LG-3 and at LG-13 (assuming that she is not homozygous at all loci in either linkage group), and each of her worker offspring inherits one of the queen’s two haplotypes for each linkage group. Identifying the two haplotypes that are represented in all of the worker progeny of a test queen allows the identification of the test queen’s haplotypes at LG-3 and LG-13. The paternal haplotype for each worker, derived from the feral drone with which the test queen mated, is then determined by subtraction of the maternal haplotype (Oldroyd et al., 1996, Oldroyd and Wongsiri, 2006).

A queen heterozygous at LG-3 and LG-13 produces haploid drones with one of four genotypes (i.e., the four possible combinations of her two pairs of haplotypes). Thus the number of wild colonies present at a location can be estimated by dividing the number of unique paternal genotypes observed in the worker progeny at a site by four. However it can happen that the number of haplotypes observed in one linkage group significantly exceeds the number observed for the other linkage group. When this occurs it is appropriate to take the linkage group with the maximum number of drone haplotypes and divide this number by two as an improved estimate of the number of colonies present.
Having determined a range for the number of feral colonies represented in the broods of test queens we used the agent based modelling approach of Arundel et al. (2012) to estimate the number of colonies per square kilometre, based on likely assumptions about the flight range of queens (8 km) and drones (7 km) and mean mating frequency (n = 14). This method produces a lower estimate of feral colony numbers than does the traditional method where the area sampled by the test queens is assumed to be 4.5 km² (Jaffé et al., 2010). To allow comparison of the two kinds of estimates and with previous work, we have estimated feral colony density using both the agent based modelling approach and by simply dividing the number of unique feral colonies represented in broods by 4.5.

5.2.3.5 Non detection error

The genetic technique used here makes the simplifying assumption that all males present in a test queen’s colony are sampled without error (Arundel et al., 2012). This assumption is close to reality, provided that the number of worker bees genotyped from each test queen progeny is large. The probability of not sampling a patriline (i.e., at least one worker daughter of each male in a mating) is given by \((1 - p)^n\) where \(p\) is the proportion of the progeny sired by a male, and \(n\) is the sample size (Foster et al., 1999). Thus if a male is the father of 5% of progeny, the probability of not sampling his progeny in a sample of 48 is 0.0085. Therefore non-sampling of a patriline in a particular test queen brood is a negligible source of error in this study.

Non-detection error also arises from the possibility that two unrelated males share the same haplotypes by chance rather than because they are
brothers (Foster et al., 1999). This probability is the frequency of each haplotype in the population. Here we have used two groups of three linked microsatellite that are highly polyallelic. If the loci used for genotyping were unlinked, the probability that two random males in a population would share the same haplotype by chance would be \( \prod q_i \), where \( q_i \) is the frequency of the allele shared by the two random males at the \( i \)th locus. Because the loci used here are linked, we cannot legitimately use \( \prod q_i \), to estimate this probability, but must use the frequency of haplotypes in the sampled population (Jaffé et al., 2010).

The frequency of each multi-haplotype genotype in the population is \( (q_{Thel})(q_{Sex}) \), the probability that two random males will share the same multi-haplotype genotype. We have calculated this probability for every multi-haplotype male genotype observed in our data, and compute the average non-detection error per fathering male as the average of these probabilities per site (Jaffé et al., 2010).

5.2.4 Results

Measures of alleleic diversity and heterozygosity are given in Table 15. The minimum number of feral colonies represented in the progeny of test queens is given in Table 16. Treating Puckapunyal North and South as the same site, the mean minimum number of feral colonies represented in the progeny of test queens per site (i.e., the number of unique feral queens that contributed drones to the matings) was 18.0 ± 2.5 s.e. We performed two-way ANOVAs of the random effects of land type (disturbed vs undisturbed) and ecosystem (mallee, box-iron bark, alpine) on the minimum number of feral colonies represented in the brood.
Levene’s test of equality of variances between treatment groups showed that the data satisfied this assumption of ANOVA ($P > 0.05$). There was no significant interaction between land type and ecosystem, so we report an ANOVA of main effects only. Land type had no significant effect on the number of colonies detected ($F_{1,3} = 1.2, P = 0.36$). Disturbed sites had a mean of 17.7 feral colonies represented in broods, whereas undisturbed sites had 19.5 feral colonies. Ecosystem had a significant effect ($F_{2,3} = 17.0, P = 0.02$), with the mallee region ($\bar{x} = 27.0$) having a higher density of feral nests than the other two ($\bar{x} = 14.15$).
Table 15: Measures of genetic variability in honey bee populations in Victoria, Australia. Where the worker carried the same genotype as the queen both the paternal and maternal alleles were entered into the data set.

<table>
<thead>
<tr>
<th>Population</th>
<th>HBSex1</th>
<th>HBSex3</th>
<th>UN351</th>
<th>HBTh2</th>
<th>HBTh3</th>
<th>HBTh4</th>
<th>Ar</th>
<th>HE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wyperfeld</td>
<td>11</td>
<td>3</td>
<td>16</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>7.5</td>
<td>0.73</td>
</tr>
<tr>
<td>Yaapeet</td>
<td>7</td>
<td>3</td>
<td>18</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>7.2</td>
<td>0.72</td>
</tr>
<tr>
<td>Dookie</td>
<td>6</td>
<td>3</td>
<td>15</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>6.5</td>
<td>0.71</td>
</tr>
<tr>
<td>Puckapunyal</td>
<td>14</td>
<td>5</td>
<td>21</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>9.5</td>
<td>0.74</td>
</tr>
<tr>
<td>Eildon</td>
<td>5</td>
<td>3</td>
<td>16</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>6.6</td>
<td>0.72</td>
</tr>
<tr>
<td>Marysville</td>
<td>6</td>
<td>3</td>
<td>11</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>5.8</td>
<td>0.72</td>
</tr>
</tbody>
</table>

4 number of microsatellite alleles  
5 average allelic richness  
6 Average $H_e$ expected heterozygosity over all loci.
Table 16: Estimates of the number of feral honey bee colonies at 8 sites in Victoria, Australia. The number of unique drone haplotypes identified at each mating site is given for two linkage groups.

<table>
<thead>
<tr>
<th>Site</th>
<th>Number of broods examined</th>
<th>Average mating number (± s.e.)</th>
<th>NDE(^7)</th>
<th>Unique haplotypes identified</th>
<th>Minimum number of feral colonies</th>
<th>Mid-range estimate of feral colonies</th>
<th>Maximum number of feral colonies</th>
<th>Approximate feral colony density/km(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wyperfeld</td>
<td>5</td>
<td>16.8 ± 0.6</td>
<td>4.8 × 10(^4)</td>
<td>73 23 58 29 51 73</td>
<td>0.2 0.9</td>
<td>&gt; 10</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>Yaapeet</td>
<td>5</td>
<td>17.2 ± 0.9</td>
<td>5.2 × 10(^4)</td>
<td>68 28 50 25 46.5 68</td>
<td>0.15 0.7</td>
<td>&gt; 10</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>Dookie</td>
<td>4</td>
<td>13.5 ± 1.8</td>
<td>1.8 × 10(^4)</td>
<td>51 28 27 14 32.5 51</td>
<td>0.1 0.4</td>
<td>4</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>Puckapunyal East</td>
<td>4</td>
<td>14.5 ± 0.5</td>
<td>1.4 × 10(^4)</td>
<td>58 22 35 18 38 58</td>
<td>0.15 0.7</td>
<td>&gt; 10</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>Puckapunyal North</td>
<td>2</td>
<td>14.5 ± 8.5</td>
<td>5.6 × 10(^4)</td>
<td>26 12 21 11 18.5 26</td>
<td>- -</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Puckapunyal South</td>
<td>1</td>
<td>8 ± 0</td>
<td>7.2 × 10(^4)</td>
<td>8 4 6 3 5.5 8</td>
<td>- -</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Puckapunyal North + South</td>
<td>3</td>
<td>12.3 ± 5.4</td>
<td>6.3 × 10(^4)</td>
<td>34 16 24 12 23 34</td>
<td>0.1 0.3</td>
<td>1.5</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>Eildon</td>
<td>4</td>
<td>15.3 ± 1.3</td>
<td>1.7 × 10(^4)</td>
<td>59 22 19 15 37 59</td>
<td>0.1 0.6</td>
<td>&gt; 10</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>Marysville</td>
<td>4</td>
<td>13.8 ± 1.3</td>
<td>2.5 × 10(^4)</td>
<td>52 20 25 13 32.5 52</td>
<td>&lt; 0.1 0.4</td>
<td>5</td>
<td>2.9</td>
<td></td>
</tr>
</tbody>
</table>

\(^7\) Nondetection errors, calculated as in Jaffé et al. (2010)
\(^8\) Based on the agent model of (Arundel et al., 2012) and the number of test broods sampled, this is the mean density of feral colonies that is compatible with the number of feral colonies observed in broods (n = 100 simulations). See .
\(^9\) Based on the assumption that queens sample the feral colony population in a 4.5 km\(^2\) area centred on the test apiary (Jaffé et al., 2010). This estimate is the minimum number of feral colonies represented in test queen broods divided by 4.5.
Remarkably, the number of feral colonies represented in the progeny of queens that mated at Eildon (unburnt) and Marysville (burnt) was numerically similar (Table 16). There was no significant difference ($P > 0.05$) in the number of feral colonies detected between the two mountain ash sites combined ($\bar{x} = 14.0$) and the four box-ironbark sites combined ($\bar{x} = 14.3$).

The Wyperfeld and Yaapeet sites were separated by a distance of 15 km. Between the ten test queens from the two sites (five at each site), the brood shared eight drone genotypes – that is brood from four of the Wyperfeld queens shared both LG-3 and LG-13 paternal haplotypes with brood from one of three Yaapeet queens. These data suggest that drones from a single colony mated with queens from colonies that were 15 km apart.

Puckapunyal North and South were 2 km apart. Broods shared one drone genotype between the two sites. In addition, one LG-3 haplotype and two LG-13 haplotypes were common to brood from both sites. For this site pair, the results are consistent with drones from six feral colonies being represented in broods at both Puckapunyal North and South.

The sum of the minimum number of feral colonies identified in broods from Puckapunyal North and South (3 and 11, respectively – Table 16) is not identical to the minimum number of feral colonies calculated when the genotypes from both sites are combined (12, Table 16). This is because the patrilines from Puckapunyal North and South that share identical drone haplotypes are considered to have originated from a
single queen when the sites are combined, but from separate queens when the sites are considered separately. If Puckapunyal North and South are considered a single mating site, the number of feral colonies inferred across all sites ranged from 11 to 29 (Table 16).

Based on the agent based modelling approach of Arundel et al. (2012) the likely minimum density of feral colonies is about 0.1 colonies/km² at the box ironbark and mountain ash sites, and about 0.2 colonies/km² at the mallee sites (Table 16). Based on the ‘area sampled’ approach of Jaffé et al. (2010) the estimates of feral colony density range from 2.7 colonies/km² at the combined Puckapunyal North and South sites, to 6.4 colonies/km² at Wyperfeld (Table 16).

5.2.5 Discussion

The number of unique drone haplotypes represented in test queen progenies shows that feral honey bee colonies are present throughout Victoria, in both disturbed sites and in areas where the original tree cover is still present. We found no evidence for relative differences in feral colony densities between disturbed and undisturbed sites. Even at Marysville, where the vegetation had been recently destroyed by fire, feral colonies are present in densities similar to the unburned mountain ash site at Eildon and to the box-ironbark sites at Puckapunyal.

Our agent-based modelling (Arundel et al., 2012) shows that there is a log-linear relationship between feral colony densities and the count of unique colonies with which a queen mates. Consequently, at lower densities (≤ 1 colonies/km²), three or more test queens will detect order
of magnitude differences in the underlying feral colony density. Figure 12 shows the relationship between feral colony density and the number of colonies represented in test queen broods assuming (a) 3, (b) 4 and (c) 5 test queens. The shaded grey area represents the region within 2 standard deviations of the mean (thus accounting for 95% of observations) for each of the densities for which simulations were performed. This region shows the range of actual densities associated with a predicted maximum likelihood density. The count of unique feral colonies with which these queens mated (Y axis) can be used to estimate the number of feral colonies per km² (X axis). The number of feral colonies detected in our experiment is in all cases consistent with these lower feral colony densities. Based on Figure 12, the observed number of feral colonies represented in broods at Wyperfeld (29; Table 16) is compatible with a minimum feral colony density of 0.2 colonies/km². A mid-range estimate of feral colony density of 0.9 colonies/km² was obtained by using Figure 12 to infer the density from a count (of 51) halfway between the minimum (29) and maximum (73). Importantly, Arundel et al. (2012) showed that the spatial distribution of feral colonies and drone congregation areas has no effect on the number of feral colonies represented in test queen broods. Rather, the only important factor affecting the number of feral colonies represented in test queen broods is the variable of interest: the density of feral colonies. This finding is unsurprising when we consider that the mating system of the honey bee evolved to maximise intra-colonial genetic diversity (Crozier
Figure 12: The relationship between feral colony density and the number of colonies represented in test queen broods assuming (a) 3, (b) 4 and (c) 5 test queens based on agent-based modelling (Arundel et al., 2012).
Given these caveats, we note that our agent-based estimates (0.1 – 0.5 colonies/km²) are an order of magnitude lower than those reported on the basis of field surveys in Botswana (4.2, McNally and Schneider, 1996), New York (4.2, Morse et al., 1990), Mexico (6, Ratnieks et al., 1991), Panama (4-7, Boreham and Roubik, 1987) and Texas (12.5, Baum et al., 2005). More importantly, they are 2-3 orders of magnitude lower than field counts of feral colonies conducted from 1993 to 1995 in Wyperfeld National Park (Oldroyd et al., 1997). At the specific location of Black Flat where the queens for this study were mated, Oldroyd et al. (1997) found densities of 120 colonies/km² in 1993, 100 in 1994, and 20 in 1995. Note, however, that these estimates based on physical searches refer to the narrow band of riparian woodland that surrounds Black Flat. At broader scales, the density of feral colonies is much lower, and probably more similar to that estimated here.

Using the simple device of dividing the minimum number of feral colonies represented in broods by 4.5 (Jaffé et al., 2010), the estimate of the density of feral colonies is considerably higher than the estimate obtained from Figure 12, and closer to estimates obtained internationally by the same method (Jaffé et al., 2010). However, our experiment has shown that dividing the count of feral colonies present in test queen brood by a sample area of either 4.5 km² (Jaffé et al., 2010) or 10.0 km² (Moritz et al., 2007) significantly overestimates the density by underestimating the true extent of the area being sampled.
Our results from the Wyperfeld and Yaapeet sites showed a significant overlap in the feral colonies detected by test sites separated at a distance of 15 km. This mating distance is consistent with findings of Jensen et al. (2005), and implies a minimum sampled area with a radius of half this distance (7.5 km); that is, an area of 177 km². The riparian area surveyed by Oldroyd et al. (1997) is 7.7 km from Yaapeet at its closest point. This same point is 8.3 km from the Wyperfeld site. This suggests that colonies in this area were sampled by the test queens from both the Wyperfeld and Yaapeet sites. It further implies that four of the five test queens from Wyperfeld took one or more mating flights in a southerly direction, and that at least three of the four test queens from Yaapeet took one or mating flights in a north-westerly direction. This suggestion could potentially be confirmed by direct sampling of colonies in the area between the two sites.

Our study provides the first relative measures of feral bee densities across Australian ecosystems. We find that land use has no significant effect on the density of bees, but that there are significantly more feral colonies in the mallee than elsewhere. Our estimates are based on assumptions about mating flight distances supported by our experimental findings, and we thus believe our model provides a realistic range for the density of colonies at a site. If we make the assumption that our density estimates are accurate within an order of magnitude, what can we say about the environmental effects of feral colonies on Australian agricultural and natural ecosystems? First, our data suggest that the density of feral bees is probably insufficient to provide adequate
pollination in a horticultural setting. Typical recommended stocking
rates are 100-200 colonies/km² (e.g., Free, 1970), whereas our estimates
are < 10 colonies/km². We suggest that without supplementation with
domestic colonies, it is unlikely that any crops requiring insect pollination
are adequately pollinated. Second, concerns about the impacts of feral
honey bee colonies on natural ecosystems (Goulson, 2003) are likely to
be unfounded in most areas, because the density of feral colonies is quite
low. While there may be some competition for forage, competition for
nest sites with native fauna seems unlikely except in specific sites such as
riparian woodland where the density of feral colonies can be very high.

5.2.6 Acknowledgements

We acknowledge the financial support of the Sustainable Futures
initiative of the University of Sydney, and the Rural Industries Research
and Development Corporation. Bob and Peter McDonald, commercial
beekeepers from Castlemaine, Victoria provided nucleus hives and
assistance for the field experiment, and their support is greatly
appreciated. Michael Duncan assisted with the field work and queen
rearing. We are also grateful to Sean-Paul Smith from Puckapunyal
Military Area for his help.

5.3 Estimating feral colony densities in South East Australia
from captured drones

Prior to the surveys conducted by Hinson et al. (2015), surveys of feral
colony densities using genetic analysis of captured drones had been
performed in Europe (Jaffé et al., 2010), Africa (Moritz et al., 2007, Moritz
et al., 2008, Jaffé et al., 2010) and Central America (Moritz et al., 2013) but never in Australia.

5.3.1 Site selection and land use

To assess the effect of land use on feral colony densities, Hinson et al. (2015) selected survey sites in three regions categorised as either undisturbed (native vegetation) or disturbed (land cleared for grazing or agricultural purposes), similar to the approach used by Arundel et al. (2014). Figure 13, Figure 14, and Figure 15 show land use in the vicinity of sampling sites within the Barrington Tops, Weddin Shire and Wimmera regions respectively. Each map is constructed from a 0.01 degree raster of land use provided by the Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) (2006). The sites are described in detail in (Hinson et al., 2015); the value of the maps is to illustrate the homogeneity (or otherwise) of the land use within the flight range of each survey site. For the Barrington Tops sites (Figure 13) land use is largely homogeneous within 5 km of each site. Allyn River is located within a forestry site but surrounding areas are largely undisturbed. For the Weddin Shire sites (Figure 14), the two sites in undisturbed environments (Ben Hall’s Cave Camp and Holy Camp) are both close to the edge of the National Park and thus any effect of land use (undisturbed vs. disturbed) is likely to be lessened for these survey sites. Land use around the disturbed sites (Grenfell and Tyagong) is largely homogeneous. Within the Wimmera region (Figure 15), the sites at Black Flat and Lake Brambruk within Wyperfeld National Park have previously been surveyed for feral colonies in 1993, 1994 and 1995 (Oldroyd et al.,
1994, Oldroyd *et al.*, 1995, Oldroyd *et al.*, 1997) using a direct search for nests. The Black Flat and Yaapeet sites were also surveyed in spring 2009 using the queen mating method; see Section 5.2 and Arundel *et al.* (2014). It is thus appealing to be able to also survey the same sites using the remaining main method for feral colony surveys; drone capture. The landscape in the vicinity of each of the survey sites in the Wimmera region is largely homogeneous; while Figure 15 shows the land adjacent to the Lake Albacutya site as being water, Hinson *et al.* (2015) note that at the time of the surveys in early 2011 that the lake had not seen flows since 1980 and is thus dry.
Figure 13: Land use in the vicinity of Barrington Tops sampling sites
Figure 14: Land use in the vicinity of Weddin Shire sampling sites
Figure 15: Land use in the vicinity of Wimmera region sampling sites
5.3.2 Generating sampling distributions

My main contribution to the experiment was to generate the synthetic sampling distributions used to infer densities from the results of the genetic analysis of the drones captured at each site. An immediate challenge was that the number of drones sampled at each site varied from 14 to 278; only two of the twelve sites shared the same number of drone samples (70 drones). This necessitated re-running the simulations outlined in Chapter 4 and (Arundel et al., 2013) for each unique sample size. Sampling distributions were initially generated using feral colony densities of 0.1, 0.5, 1, 5 and 10 colonies/km$^2$ and DCA densities of 0.1, 0.5, 1, 5 and 10 DCAs/km$^2$. Each individual simulation run used a unique aggregated spatial distribution of feral colonies, and a unique random spatial distribution of DCAs. For each combination of feral colony density and DCA density, 100 simulations were performed; in total this amounted to 30,000 executed simulations. Initial interpolation of results showed that for many sites densities were at the lower end of the modelled range (~ 0.5 colonies/km$^2$). For this reason, a second set of simulations was performed to generate additional data points for interpolation. This necessitated generation of additional aggregated spatial distributions with feral colony densities of 0.2, 0.3, 0.4, 0.7, 1.5, 2, 3, and 4 colonies/km$^2$. As DCA density was shown by (Arundel et al., 2013) to have little effect on the sampling distributions it was held constant at 1 DCA/km$^2$ for the second set of simulations. In total, a further 7200 simulations were performed. The full set of log-linear plots used for interpolation of the density from the estimated number of unique...
colonies at each site are reproduced in Figure 55 - Figure 75. For large sample sizes the linearity of the log-linear plot starts to show signs of breaking down (Figure 70, Figure 72, Figure 74).

5.3.3 Results

The results of the interpolation applied to the estimated number of colonies at each survey site are summarised in Table 17. The density range is derived from the intercepts of the horizontal line plotted for the estimated number of unique colonies with the grey region. An artefact of the method used to generate this region of uncertainty (discussed in Arundel et al. (2013)) is that it occasionally exceeds the magnitude of the sample size (Figure 55, Figure 56, Figure 57); a practical impossibility that can be safely ignored in the interpretation of data.
### Table 17: Estimates of the density of feral honey bee colonies at 12 sites in south-eastern Australia (Hinson et al., 2015)

<table>
<thead>
<tr>
<th>Site name</th>
<th>Region</th>
<th>Disturbed /undisturbed</th>
<th>Number of drones sampled</th>
<th>Number of unique haplotypes</th>
<th>Estimated number of unique colonies</th>
<th>Estimated density (colonies / km²)</th>
<th>Density range (colonies / km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloucester Tops</td>
<td>BT</td>
<td>Undisturbed</td>
<td>79</td>
<td>36</td>
<td>18</td>
<td>0.4</td>
<td>0.5 to 1.5</td>
</tr>
<tr>
<td>Stratford Park</td>
<td>BT</td>
<td>Disturbed</td>
<td>123</td>
<td>28</td>
<td>14</td>
<td>0.25</td>
<td>0.1 to 1.5</td>
</tr>
<tr>
<td>Allyn River</td>
<td>BT</td>
<td>Undisturbed</td>
<td>72</td>
<td>47</td>
<td>23.5</td>
<td>0.65</td>
<td>0.3 to 2.5</td>
</tr>
<tr>
<td>Monkerai</td>
<td>BT</td>
<td>Disturbed</td>
<td>70</td>
<td>43</td>
<td>21.5</td>
<td>0.6</td>
<td>0.3 to 2.5</td>
</tr>
<tr>
<td>Ben Hall's Cave Camp</td>
<td>WS</td>
<td>Undisturbed</td>
<td>62</td>
<td>38</td>
<td>19</td>
<td>0.55</td>
<td>0.2 to 2</td>
</tr>
<tr>
<td>Grenfell</td>
<td>WS</td>
<td>Disturbed</td>
<td>70</td>
<td>41</td>
<td>20.5</td>
<td>0.55</td>
<td>0.25 to 2</td>
</tr>
<tr>
<td>Holy Camp</td>
<td>WS</td>
<td>Undisturbed</td>
<td>278</td>
<td>75</td>
<td>37.5</td>
<td>0.8</td>
<td>0.25 to 2</td>
</tr>
<tr>
<td>Tyagong</td>
<td>WS</td>
<td>Disturbed</td>
<td>74</td>
<td>36</td>
<td>18</td>
<td>0.45</td>
<td>0.1 to 1.5</td>
</tr>
<tr>
<td>Black Flat</td>
<td>WR</td>
<td>Undisturbed</td>
<td>222</td>
<td>98</td>
<td>49</td>
<td>1.5</td>
<td>0.4 to 3</td>
</tr>
<tr>
<td>Lake Albacutya</td>
<td>WR</td>
<td>Disturbed</td>
<td>29</td>
<td>19</td>
<td>9.5</td>
<td>0.25</td>
<td>0.1 to 1.5</td>
</tr>
<tr>
<td>Lake Brambruk</td>
<td>WR</td>
<td>Undisturbed</td>
<td>241</td>
<td>84</td>
<td>42</td>
<td>1</td>
<td>0.25 to 2.5</td>
</tr>
<tr>
<td>Yaapeet</td>
<td>WR</td>
<td>Disturbed</td>
<td>14</td>
<td>12</td>
<td>6</td>
<td>0.15</td>
<td>&lt; 0.1 to 0.6</td>
</tr>
</tbody>
</table>
5.3.4 Discussion

Taking into account the re-estimated densities for other surveys using the drone capture method (discussed in Chapter 4 and Arundel et al. (2013)), the inferred densities are largely consistent with other surveys conducted globally. There was a statistically significant difference between the inferred densities for disturbed and undisturbed environments for the Wimmera region sites, and for the pooled (all sites combined) estimates only. The size of any effect of environment at the Weddin Shire sites was likely reduced by the choice of undisturbed sites; approximately one third of the within flight range area constituted disturbed habitat. As noted by Hinson et al. (2015), none of the surveyed sites would provide an adequate number of colonies to perform pollination of agricultural or horticultural crops.
Given that the Wimmera region sites had been surveyed previously, an obvious question to ask is whether any conclusions can be drawn about the long term temporal variation in colony densities. Unfortunately the answer is no, the key problem with interpretation being the vastly different areas surveyed by each of the respective techniques. Both the
Black Flat and Lake Brambruk DCAs from which drones were captured could have comprised colonies located at most of the survey sites from the surveys conducted by Oldroyd et al. (1994, 1995, 1997); in particular, sites 3, 4, 5, 6, and 7 (Figure 17). The extended range of the queen mating method used by Arundel et al. (2014) would also potentially pick up colonies located at sites 1 and 2 (Figure 17). The results of all previous surveys conducted at Wyperfeld are summarised in Table 18. Using an identical survey method in three consecutive years, Oldroyd et al. (1997) found a marked decline in colony numbers and attributed this to low rainfall and high temperatures.

Table 18: Comparison of feral colony survey results within Wyperfeld National Park

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>North of Lake Werrebean</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>North of The Kidneys</td>
<td>2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>South of Black Flat Lake</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>29 - 73</td>
<td>49</td>
</tr>
<tr>
<td>Lake Jerriwirrup</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>West of Lake Brambruk</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Wonga Lake</td>
<td>6</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>South of Lake Brambruk</td>
<td>7</td>
<td>15</td>
<td>18</td>
<td>4</td>
<td>N/A</td>
<td>42</td>
</tr>
<tr>
<td>Total colonies</td>
<td>34</td>
<td>34</td>
<td>10</td>
<td>29 - 73</td>
<td>42-49</td>
<td></td>
</tr>
<tr>
<td>Area surveyed (km²)</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>&gt; 177</td>
<td>&gt; 50</td>
<td></td>
</tr>
<tr>
<td>Density (colonies/km²)</td>
<td>136</td>
<td>136</td>
<td>40</td>
<td>0.2 to &gt; 10</td>
<td>0.25 to 3</td>
<td></td>
</tr>
</tbody>
</table>

Figure 18 shows the rolling average of daily rainfall over the preceding year, and monthly maximum temperatures over the last 25 years. While the summer of 1995 is notable for its low rainfall, the summer of 1993 actually provided unusually wet and (relatively) cool conditions that would likely have led to increase in colony numbers that year. The year
leading up to the time of the survey of Hinson et al. (2015) was also unusually wet, and flooding was experienced across the Wimmera region in response to heavy rainfall received in January (Figure 18).

Unfortunately though no conclusions can be drawn though on whether colony densities were either higher or lower than in previous surveys. Even putting aside the near certainty of significant year-to-year temporal variations in colony density the differences in the measured densities across the all three surveys could have arisen entirely as an artefact of the differences in the areas from which the samples were drawn.
Figure 18: Rainfall and maximum temperatures over 25 years at Wyperfeld National Park
One point of comparison between the different survey methods worth exploring is the relative ease with which the Yaapeet site was surveyed by Arundel et al. (2014), versus the difficulty experienced by Hinson et al. (2015) in even capturing a sufficient sample size of drones. As noted by Hinson et al. (2015) there is likely to be some positive relationship between the rate of capture at a DCA and the colony density. In areas of low colony density, the most viable survey technique to use is thus queen mating, as queens are more clearly more adept at locating DCAs and "capturing" drones than humans.
6 THE SPATIO-TEMPORAL DISTRIBUTION OF MANAGED COLONIES

6.1 Introduction

The spatial distribution of honey bee colonies managed by beekeepers changes over time, as beekeepers move hives in response to changes in resource availability. This chapter presents a model of how beekeepers move their hives in response to changing resource availability to evaluate the importance of their ability to predict future flowering events accurately.

The basis of this chapter is a paper presented at the GeoComputation conference in 2009, namely:

6.2 Motivation

There are several reasons why knowledge of the spatio-temporal distribution of managed colonies is important. Firstly, as outlined in Chapter 5, there are insufficient densities of feral colonies in most environments to provide adequate pollination of agricultural and horticultural crops. For pollination of these crops to occur managed hives must be used, and in many cases these must be brought onsite by beekeepers. Secondly, understanding the movement of beekeepers helps facilitate the development of tools to assist the beekeeping industry generally, such as the prototype system discussed in Chapter 7. Thirdly, and as outlined in Section 2.7, it is critical to understand the movement of beekeepers to respond to biosecurity incidents. The imposition of quarantine zones and contact tracing between infested premises both require detailed knowledge of hive movements.

There are two fundamental methods by which we can seek to understand the movement of hives by beekeepers; measurement or prediction. As discussed in Section 2.7, it is mandatory for commercial beekeepers to record hive movements but there is no centralised collection or analysis of this data currently. For biosecurity purposes, measured data on the actual movement of hives is superior to predictions of movement. However, until this actual measured data becomes available, predictions of hive movements based on resource availability still provide the ability to gain insights into the dynamics of beekeeper hive movements.

Predicting the movement of hives by beekeepers requires understanding how they perform their seasonal planning in response to their predictions.
and observations of flowering, and how they update these plans when new information becomes available. The dimensions of this optimisation problem are summarised in Table 5 and the problem itself is introduced in Section 3.5 as having the characteristics of a One-commodity Pickup and Delivery Travelling Salesman Problem (1-PDTSP) (Hernández-Pérez and Salazar-González, 2004).

The managed hive movement optimisation problem can be viewed as representative of a broader class of what can be termed harvesting problems. In harvesting problems, there may be multiple sites at which harvesting can be performed, and multiple existing deployments of harvesting equipment. The operator must efficiently redeploy the harvesting equipment from their current sites to the new harvesting locations. The concept of harvesting is a general one; the object performing the harvesting may be an animal (honey bee, cow, sheep etc.) or a machine (combine harvester, logging equipment, monitoring equipment etc.). The common factor is that homogeneous objects are redeployed into an environment in which the harvested resource is renewed over time.

The fundamental differentiator between the 1-PDTSP and the harvesting problem solved by beekeepers is that every chosen delivery location will become a pickup location at some future date, and vice versa. As human beings, the choices we have available to us in the future often depend on the decisions we made in the past. This phenomenon can also apply to route choices, yet most research focuses only on the case of how good routes are chosen in a single time period. It is important to understand
the qualitative and quantitative differences between routes chosen for a single time period and routes chosen over an extended period when developing realistic models of human behaviour. Traditional pickup and delivery problems assume no dependence between the solutions for any two problems. However, in harvesting problems the delivery locations for any given solution assume the role of pickup locations when the problem is next solved. When there are more options for locating an object than there are objects to occupy the locations, this introduces the possibility that adopting the optimal solution in the current iteration may effectively penalise the next iteration and thus lead to a suboptimal solution overall.

A second consideration is the case where the delivery locations visited are not considered to be equally rewarding. Thus the attractiveness of a delivery site depends on both the travel costs to reach it and the profit gained from it (Feillet et al., 2005). In this sense, routes are planned in such a way to maximise economic utility.

The hypothesis of this experiment anticipates the interaction of these two effects, and is thus:

- Solution performance for the harvesting variant of the TSP improves with increasing foresight.
- In the case where TSP locations have unequal attractiveness, the significance of foresight on performance is reduced.

The logic behind the first of these points is intuitive; improved capabilities to predict the future and react accordingly are always
associated with improved outcomes in many fields of human endeavour. However, the interpretation of the second point is less obvious; it raises the possibility that focussing on the more profitable (and ignoring the less profitable) locations in the current iteration can yield good performance and thus reducing the need to plan ahead.

6.3 Model development

6.3.1 Simulation design

In order to examine the effects of foresight and heterogeneity on the harvesting variant of the TSP, a simulated Euclidean space was created and populated by randomly generated locations. The model was developed in Java, using the JUNG Framework (JUNG Framework Development Team, 2009). Here, locations represent the sites at which harvesting is performed, and are occupied by notional objects representing the harvesting equipment to be relocated at each iteration. The locations are grouped into iterations as shown in Figure 19, with the numbers within each circle indicating the iteration. A central depot, indicated by the circle labelled “0” is defined. Routes are defined to start at the depot then pickup objects from the delivery locations chosen in the last iteration, deliver these to locations chosen in the current iteration and then return to the depot.
The choice of the number of objects (designated \( n \)), the number of iterations \( (i) \) and the number of possible delivery locations in each iteration \( (m) \) is arbitrary; it is assumed that any properties of good solutions to the optimisation problem are relatively insensitive to either choice. In the simulation shown in Figure 19 there are four objects \( (n=4) \) to be relocated over three iterations \( (i=3) \) and their starting (pickup) locations in the first iteration are given by the points labelled “1”. In each iteration there are seven possible delivery locations \( (m=7) \) for the four objects picked up in that iteration. These delivery locations are labelled
for first, second and third iterations as "2", "3", and "4" respectively. As the vehicle is assumed to have unit capacity, each route must interleave pickup and delivery locations. That is, in every solution a visit to a chosen delivery location will always be performed directly after a visit to one of the mandatory pickup locations. The temporal linking between iterations occurs because the chosen delivery locations in any particular iteration become the pickup locations in the next iteration as that is now where the harvesting equipment has been placed. Thus the sequence of the node labels traversed by the agent/harvester for the values chosen is as follows:

- Iteration 1: 0, 1, 2, 1, 2, 1, 2, 1, 2, 0
- Iteration 2: 0, 2, 3, 2, 3, 2, 3, 2, 3, 0
- Iteration 3: 0, 3, 4, 3, 4, 3, 4, 3, 4, 0

There are two categories of optimal routes; iteration optimal solutions and overall optimal solutions. In the case of iteration optimal solutions, the algorithm seeks the optimal solution for the current iteration only; this then determines the starting locations for the next iteration where the process is repeated. The overall optimal solution algorithm finds the single best solution across all iterations. Thus while in any particular iteration the overall optimal solution may have a cost greater than the iteration optimal solution, it is guaranteed to have a cost less than or equal to the sum of iteration optimal solution across all iterations.

Foresight is introduced into the simulation by varying the number of iterations over which the algorithm has visibility of possible delivery
locations. Thus is in the situation illustrated in Figure 19, with maximum foresight the algorithm may have access to all the site locations for all future iterations (labels 0, 1, 2, 3, 4). If the algorithm can only "look ahead" one iteration, then in the first iteration the algorithm has knowledge of locations with labels 0, 1, 2, and 3, and in the second iteration then gains knowledge of the locations labelled with 4. Where the algorithm has no foresight and thus only possesses information about the sites in the current iteration, the route chosen will by definition be the iteration optimal route.

6.3.2 Formal specification

The formal specification of the 1-PDTSP is outlined in many papers (Hernández-Pérez, 2004, Hernández-Pérez and Salazar-González, 2004, Berbeglia et al., 2007, Parragh et al., 2008, Battarra et al., 2014). The harvesting variant of the TSP can additionally be defined as follows.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_0$</td>
<td>the set of all possible locations as vertices, $v$, across all iterations</td>
</tr>
<tr>
<td>$V$</td>
<td>the set of all possible locations as vertices, $v$, across all iterations</td>
</tr>
<tr>
<td>$V_r$</td>
<td>the set of all locations in iteration $r$</td>
</tr>
<tr>
<td>$V_{r}^{P}$</td>
<td>the set of all pickup vertices used in iteration $r$</td>
</tr>
<tr>
<td>$V_{r}^{D}$</td>
<td>the set of all delivery vertices used in iteration $r$</td>
</tr>
<tr>
<td>$i$</td>
<td>the number of iterations</td>
</tr>
<tr>
<td>$r$</td>
<td>the iteration number; ($r = 1$ to $i$)</td>
</tr>
<tr>
<td>$n$</td>
<td>the number of objects to be relocated/number of pickup locations in each iteration; $</td>
</tr>
<tr>
<td>$m$</td>
<td>the number of delivery locations in each iteration; $</td>
</tr>
<tr>
<td>$c_{ij}$</td>
<td>the cost to traverse arc or edge $(v_i, v_j)$</td>
</tr>
<tr>
<td>$c_{IO}$</td>
<td>the total cost of the iteration optimal solution</td>
</tr>
<tr>
<td>$c_{OO}$</td>
<td>the total cost of the overall optimal solution</td>
</tr>
<tr>
<td>$\delta$</td>
<td>proportional difference of iteration optimal solution cost to overall optimal solution cost</td>
</tr>
<tr>
<td>$x_{ij}$</td>
<td>${1$, if arc $(v_i, v_j)$is traversed $}$ ${0$, otherwise $}$</td>
</tr>
</tbody>
</table>
In general, the pickup locations in any iteration are the delivery points in the previous iteration:

\[ V_r^P = V_{r-1}^D \quad (r > 1) \]

The vertices present in each iteration include the depot and are:

\[ V_r = \{v_0\} \cup V_r^P \cup V_r^D \]

The aim of the iteration optimal solution is to minimise the route cost of within each iteration, summed across all iterations:

\[ c^{IO} = \sum_{r=1}^{i} \left( \min \sum_{(v_i, v_j) \in V_r} c_{ij} x_{ij} \right) \]

The aim of the overall optimal solution is to minimise the overall cost of routes:

\[ c^{OO} = \min \sum_{r=1}^{i} \sum_{(v_i, v_j) \in V_r} c_{ij} x_{ij} \]

Thus:

\[ c^{IO} \geq c^{OO} \]

and this can be expressed as a proportional difference:

\[ \delta = \frac{c^{IO} - c^{OO}}{c^{OO}} \]
6.3.3 Computational complexity

The TSP has been shown to be NP-hard, and this is assumed to hold for most variants of the TSP. In this experiment, it is assumed that the harvesting variant of the TSP is also NP-hard.

In a given iteration, there are many more unique solutions than there are unique finishing configurations for that iteration. In this experiment, where \( n = 4 \) and \( m = 7 \), there exists \( P(n, n) = n! = 4! = 24 \) possible order permutations for the pickup locations and \( P(m, n) = P(7, 4) = 840 \) possible order permutations for the delivery locations. These are interleaved, resulting in \( n! \times P(m, n) = 20160 \) unique routes in each iteration. The number of unique sets of routes increases exponentially with \( r \). Even in the case of \( r = 5 \) the number of unique sets of routes is huge: \((20160)^5 = 3.3 \times 10^{21}\). However, from the perspective of searching for an optimal solution across all iterations, there can only be a single optimal solution for each finishing configuration; this reduces the number of routes that need to be retained for each iteration to \( C(m, n) = C(7, 4) = 35 \). Each of these \( C(m, n) \) routes has \( n! \times P(m, n) \) routes in the next iteration, resulting in \( C(m, n) \times n! \times P(m, n) \) routes to be evaluated, but even this number still only generates a single optimal solution for each of the \( C(m, n) \) finishing configurations. Thus across \( r \) iterations, there are \( n! \times P(m, n) + C(m, n) \times n! \times P(m, n)(r - 1) \) routes to be evaluated in an exhaustive search of the solution space. Importantly for the overall optimal solution, this search space grows linearly rather than exponentially with the number of iterations. Thus the number of routes that needs to be evaluated for \( r = 5 \) is only \( 2.8 \times 10^6 \).
6.3.4 Heterogeneous attractiveness

Thus far in the introduction of the harvesting problem there has been an implicit assumption that all locations are equally profitable or attractive, and optimisation has been performed only with the goal of minimising the distance travelled. There are numerous ways in which the profitability of attractiveness of individual locations can be adjusted with the consequence of influencing route choice. The most common method is to assign a profit to each location and subtract from this the cost of travel to it from a given origin. The resulting quantity, commonly known as economic utility, is one possible measure of the location’s attractiveness. To determine the optimum route, the route selection algorithm must seek to maximise the (positive) utility.

This experiment seeks to transform a problem defined by Euclidean distances and heterogeneous attractiveness into an equivalent problem defined by homogeneous attractiveness and modified $c_{ij}$ such that solving the transformed problem is equivalent to maximising the economic utility in the original problem.

![Diagram](image)

**Figure 20: Transformation of homogeneously attractive locations into heterogeneously attractive locations**
In Figure 2(a), all sites have an equal attractiveness (of value 1) as indicated by their label. Therefore the choice of next destination from the origin depends only on the cost associated with travel to each location. On this basis, the preferred destination site is the closest site with a cost of 90. In Figure 2(b), the same travel costs apply but the three points have each been given a different profit or attractiveness; 1, 2, or 3. In Figure 2(c), the bi-criteria problem is transformed into a single criteria problem via a subtractive transformation using a weighting coefficient, $\beta$, for each attractiveness of $\beta = 30$. The new costs are given by:

$$\text{cost}_{\text{subtracting transformation}} = \text{cost}_{\text{original}} - (\beta \times \text{profit})$$

In the transformed situation of Figure 2(c) the closest site becomes the one with new cost 80. An alternative to the subtractive transformation, and the approach used in this experiment, is a dividing transformation. The transformed costs in this case are given by division, as the name suggests:

$$\text{cost}_{\text{divisor transformation}} = \frac{\text{cost}_{\text{original}}}{\text{profit}}$$

In the transformed situation of Figure 2(d) the closest site becomes the one with new cost 65.

6.4 Experiments

To investigate the performance of the harvesting variant of the TSP, 1000 random simulations were performed for each set of parameters. The values of $n = 4$ and $m = 7$ were used in each experiment. The number of
locations generated was determined by the number of iterations, according to the formula:

\[ |\{v_0\}| + |V_r^P| + i \times |V_r^D| = 1 + n + i(m) = 1 + 4 + i(7) = 5 + 7i \]

With the exception of the depot, \(v_0\), which was assigned a fixed position in the centre of the square simulation space, all other points were given locations where both their \(x\) and \(y\) coordinates were chosen from a uniform random distribution. Random seeds from 1 to 1000 were used to generate location distributions for the 1000 simulation runs. In this manner, all experiments were conducted against the same set of randomly generated location distributions.

Rather than fixing the number of iterations and varying the degree of foresight, it is functionally equivalent and more computationally efficient to give the route finding algorithm visibility across all locations in all iterations, and simply vary the number of iterations in the simulation.

6.4.1 Experiment 1: Effect of foresight on performance

The aim of the first experiment was to quantify the effect of foresight on performance, and quantify this effect for different degrees of foresight. The degrees of foresight used to compute the overall optimal solution ranged from a minimum of two iterations \((i = 2)\) to a maximum of six iterations \((i = 6)\). The case of \(i = 1\) corresponds to visibility of the current iteration only, which equates to the iteration optimal solution. For each value of \(i\), 1000 simulations were run and the iteration optimal and overall optimal solutions across all iterations were computed and compared. The degree of foresight for the optimal overall solution in each
case matched precisely the number of iterations in the simulation. The results are shown under the heading of “Homogeneous attractiveness” in Table 20 and Figure 22.

Iteration optimal and overall optimal solutions to the problem outlined in Figure 19 are illustrated below in Figure 21. In the first iteration, the iteration optimal solution yields a path with length 2504 units. One of the delivery locations chosen is a point in the lower left corner of the simulation space. The iteration optimal solution algorithm, lacking any foresight, ignores the fact that this location is a comparatively large distance away from any possible delivery points in the next iteration. However, the overall optimal solution algorithm has complete visibility of all future iterations and the ability to determine the optimal solution across the entire duration of the simulation. It chooses a path with length 2537 in the first iteration. By sacrificing performance in the first iteration it is able to outperform the iteration optimal solution by the end of the next iteration (4158 vs. 4871 units).
Figure 21: Iteration optimal and overall optimal solutions for a particular random configuration across three iterations.
6.4.2 Experiment 2: Effect of heterogeneous attractiveness on performance

In the second experiment, the effect of heterogeneous attractiveness was quantified by repeating the comparison of iteration optimal and overall optimal solutions after modifying the profitability of locations. Locations were randomly assigned profits of either 1, 2 or 3 with equal weighting for each profit. To determine the resulting single criteria costs for each location, the dividing transformation was then used. The proportional differences between iteration optimal and overall optimal solution were then determined for the differing degrees of foresight, as in Experiment 1. The results are shown under the heading of “Heterogeneous attractiveness” in Table 20 and Figure 22.

![Boxplots for 1000 random distributions and differing numbers of iterations](image)

*Figure 22: Boxplots for 1000 random distributions and differing numbers of iterations*
Table 20: Results for 1000 randomly generated location distributions

<table>
<thead>
<tr>
<th>i</th>
<th>$\bar{\delta}$, homogeneous attractiveness</th>
<th>$\bar{\delta}$, heterogeneous attractiveness</th>
<th>Two-sample $t$-test</th>
<th>Two-sample Wilcoxon test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$p$</td>
<td>95% CI</td>
</tr>
<tr>
<td>2</td>
<td>2.67%</td>
<td>2.22%</td>
<td>$&lt; 0.05$</td>
<td>0.12% 0.80%</td>
</tr>
<tr>
<td>3</td>
<td>4.09%</td>
<td>3.61%</td>
<td>$&lt; 0.05$</td>
<td>0.07% 0.88%</td>
</tr>
<tr>
<td>4</td>
<td>4.90%</td>
<td>4.47%</td>
<td>$&lt; 0.05$</td>
<td>0.02% 0.83%</td>
</tr>
<tr>
<td>5</td>
<td>5.37%</td>
<td>4.74%</td>
<td>$&lt; 0.05$</td>
<td>0.26% 1.00%</td>
</tr>
<tr>
<td>6</td>
<td>5.60%</td>
<td>5.17%</td>
<td>$&lt; 0.05$</td>
<td>0.09% 0.79%</td>
</tr>
</tbody>
</table>

6.5 Discussion

As shown in Table 20 and Figure 22, a solution optimised across multiple iterations significantly improves upon that optimised for a particular iteration alone. This result applies both for the cases of homogeneous and heterogeneous location attractiveness. While the magnitude of the performance improvement grows with increasing foresight ability, the rate of increase of improvement slows.

The results also support the second hypothesis that when locations are given heterogeneous attractiveness that the performance difference between overall optimal and iteration optimal solutions would reduce. A decrease of the mean performance difference was observed for all numbers of iterations, and this decrease was significant at the 5% level using both the two-sample t-test and two-sample Wilcoxon rank sum test. However, as the range of the mean difference between the two samples was less than 1%, the impact of heterogeneous attractiveness was small in absolute terms.

Human performance finding tours in the traditional TSP tends to be in the range of no more than 10% worse than the optimal tour. Graham et al. (2000) found that for problems of up to 30 cities, performance was
typically within 3% of optimal. MacGregor et al. (1999) found that in 48-
city problems performance was within 10% of optimal. Even for 120-city
problems, performance is typically within 11% of optimal (Dry et al.,
2006). Thus while heterogeneous attractiveness only reduces the need
for foresight by <1%, this can still be viewed as being significant in
simplifying the problem to be solved by beekeepers and improving their
chances of finding optimal (or near optimal) solutions.

Many of the characteristics of heuristic solutions for optimal harvesting
routes will necessarily be different to those used for the TSP. There are
extraneous points (non-chosen delivery locations) within each iteration,
and across multiple iterations. This precludes the finding of a Gestalt
figure or aesthetically pleasing solution (MacGregor et al., 2000,
MacGregor et al., 2004). As Figure 21 shows, both the iteration optimal
and overall optimal solution violate the principle of crossing avoidance
(MacGregor et al., 2004) used in human solutions to the TSP.

Furthermore, when the attractiveness of the locations is heterogeneous,
the distances between locations are no longer the sole consideration in
the formation of good routes. Human performance on the harvesting
variant of the TSP may therefore be comparatively worse than on the
classic TSP, but this has not yet been measured.

The model as implemented has some limitations. It represents a very
rigid view of reality, which the same number of sites available in each
iteration and fixed time intervals with events precisely synchronised
against these intervals. Due to computational limitations, the values of \( n, m \) and \( r \) investigated are also relatively small. In the model, foresight is
always accurate and events do not start earlier or later than planned. Any or all of these constraints could be relaxed in further research. Only a single model for heterogeneous attractiveness has been investigated in this experiment; there is a possibility that other models or different distributions of attractiveness may yield a different result.

It should also be highlighted that the model of attractiveness itself is simplistic. Locations may be attractive for a variety of reasons; in the case of a beekeeper the profit associated with a location could be due to nectar, to pollen, to paid pollination services or to some combination of these. The attractiveness of locations may therefore not be independent of one another; if a beekeeper chooses a location because it provides strong nectar (but little pollen), then in the next iteration locations that provide pollen will become more attractive. The simple model used in this experiment lacks the ability to model different dimensions of attractiveness, but this could again be a worthwhile area of future investigation.

6.6 Conclusion

In the harvesting variant of the TSP, performance increases with increasing foresight, and provides benefits over repeatedly choosing an optimal route for the current iteration in the order of 2-6%. It is important for beekeepers to be able to predict a couple of iterations into the future and plan their movements on this basis. However, beyond the first couple of iterations the incremental benefits of additional foresight diminish. Where locations are not equally attractive the significance of foresight is likely to be reduced, but only by a small amount. In practical
terms, a good beekeeper may be able to make some medium term predictions of flowering and plan his or her operations on this basis. An expert beekeeper, possessing more foresight, will be able to predict further into the future, but the savings in travel costs will be relatively small.

Where differences arise between the performance of the iteration optimal and overall optimal solutions, this is typically because the overall optimal solution chooses a set of delivery locations that are not optimal for the current iteration, but are optimal over the longer term. For any set of chosen delivery locations, $V^D_r$, there are many possible routes which can be chosen. The variation in the cost of the routes far exceeds the variation between iteration and overall optimal solution, irrespective of whether homogeneous or heterogeneous attractiveness of locations is used. To summarise, it is far more important to choose a good tour for an arbitrary set of delivery locations in the current iteration than it is to choose the optimal set of delivery locations to be used as pickup locations in the next iteration. In constructing a simplified spatio-temporal model of beekeeper hive movements, it would therefore be valid to ignore the effects of predicted flowering events on the movement of hives by beekeepers and use only actual flowering events.
7 THE SPATIO-TEMPORAL DISTRIBUTION OF EUCALYPT FLOWERING

7.1 Eucalypts and remote sensing

7.1.1 Introduction

The dominant sources of food for honey bees in Australia are eucalypt species. The remote sensing of patterns in eucalyptus growth holds great promise for understanding the spatio-temporal distribution of floral resources for honey bees in Australia, which in turn will lead to a greater understanding of beekeeper seasonal movements and feral colony densities.

7.1.2 The distribution of eucalypts

A key element is deriving a spatial-temporal distribution of floral resources in Australia is accurate information on the distribution of eucalypt species.
The species within foraging range of a particular apiary site will determine the times of year a beekeeper is likely to visit the site. Precise knowledge of actual flowering of these species is needed before hives will be moved to the site. While an individual beekeeper can make observations at their site of prevailing species, to model beekeeper movements requires a similarly detailed knowledge of species but at a landscape scale.

Feral colony densities are highly likely to be influenced by the level of available melliferous resources. To fulfil the promise that remote sensing offers in modelling the spatio-temporal distribution of such resources, accurate knowledge of species distributions is required at the same level of spatial resolution as the remote sensing data; 500 m × 500 m in the case of the MODIS satellite data discussed in Section 7.1.6.

To illustrate the challenges in reliably establishing the distribution of a particular eucalypt species, Figure 23 shows published distributions for one of Victoria's most important melliferous eucalypts; *Eucalyptus melliodora* (yellow box).

Figure 23(a) shows the modelled extent of vegetation highly likely to be native woody vegetation as of 2005 (Department of Environment and Primary Industries, 2015b) shaded in green. The dominant tree in these native landscapes is the eucalypt. As the extent of native vegetation has itself been established using time series LANDSAT satellite imagery and ground-truthed site data (Department of Sustainability and Environment, 2007b), it can be regarded to be accurate at remote sensing resolutions.
Figure 23(b) and Figure 23(c) show the Victorian Department of Environment and Primary Industries' biodiversity atlas *E. melliodora* records for sites with Figure 23(b) showing records with moderate to low (±500 m to ±10 km) spatial accuracy (Department of Environment and Primary Industries, 2014d) and Figure 23(c) showing high (<±500 m) spatial accuracy (Department of Environment and Primary Industries, 2014c) records. Each occurrence record is marked with a green circle.

Figure 23(d) shows consolidated herbaria and other occurrence records for *E. melliodora* extracted from the Atlas of Living Australia (2015). There is significant correspondence between this dataset and the combined datasets of Figure 23(b) and Figure 23(c), with the added advantage that species' distribution data extends across state boundaries. Models predicting beekeeper movements using species distribution data require data for multiple states as beekeepers will move their hives across state boundaries in pursuit of flowering events.

Figure 23(e) shows digitisation of the occurrence records for *E. melliodora* from Brooker (2002). The relative paucity of occurrence records is most likely because occurrence records from only a single government herbarium has been used. Analysis of Atlas of Living Australia (2015) data suggests that the source of the records in this case is the Australian National Herbarium.

Figure 23(f) reflects a digitised georeferenced view of the distribution of *E. melliodora* published by Brooker and Kleinig (2006). Comparing this with Figure 23(e) it appears that Brooker and Kleinig (2006) derived this
distribution by taking the Australian National Herbarium occurrence records and enclosing them in a minimum boundary geometry. As a consequence, this map overestimates the distribution of *E. melliodora* in the Australian Alps, South Eastern Highlands and Southern Volcanic Plains bioregions, and underestimates the distribution in the South East Corner, South East Coastal Plain, Victorian Midlands and Riverina bioregions (Department of the Environment, 2015).

The distribution of *E. melliodora* in Figure 23(g) has been digitised and georeferenced from Nicolle (2006). This map encloses the majority of the consolidated herbaria records from Figure 23(d) but overestimates the distribution of *E. melliodora* in the Australian Alps, South Eastern Highlands and Southern Volcanic Plains bioregions (Department of the Environment, 2015).

Figure 23(h) illustrates the digitised and georeferenced view of the distribution of *E. melliodora* from Costermans (2006). This map underestimates the distribution in the South East Corner, South East Coastal Plain and Riverina bioregions (Department of the Environment, 2015).

Finally, Figure 23(i) shows the digitised, georeferenced distribution of *E. melliodora* in Goodman (1973). On balance, and when compared with herbaria records, this is the best representative view of the distribution of *E. melliodora* from all the available field guides. However, the map still underestimates the distribution in the South East Corner and South East Coastal Plain bioregions (Department of the Environment, 2015).
While this comparison focuses on only a single species; *E. melliodora*, the disagreement between sources is readily apparent in the analysis of distributions of other important commercial beekeeping melliferous species. The methodology used to develop the maps in many texts lacks rigour and precision and the resulting distributions are of limited value in the development of spatially-explicit models.

Rather than adopting a bottom-up method of building distributions from occurrence records, an alternative top-down approach is to divide the landscape into bioregions, sub-regions and eventually into characteristic ecological and floristic communities. Nationally, the Australian government has adopted a scheme called Interim Biogeographic Regionalisation for Australia (IBRA) which classifies Australia’s landscapes into 89 large geographically distinct bioregions based on common climate, geology, landform, native vegetation and species information (Department of the Environment, 2015). In Victoria, bioregions have been defined in accordance with the IBRA-defined sub-regions (Department of Environment and Primary Industries, 2015c); see also Figure 1. Within each bioregion, Ecological Vegetation Classes (EVC) are used as the standard unit for classifying vegetation types in Victoria (Department of Environment and Primary Industries, 2014a). There are currently 601 EVCs defined in Victoria (Department of Environment and Primary Industries, 2014b). At the lowest level of the hierarchy, floristic communities are defined (Department of Environment and Primary Industries, 2014a). Across a study of 14 EVCs in the box-ironbark forests of Victoria, (Mac Nally *et al.*, 2002) found that the use of EVCs as
biodiversity management units may account reasonably well for trees, birds and mammals but not reptiles or invertebrates. For the purposes of modelling eucalypt distributions this is encouraging, but the lack of interoperable classification schemes across state boundaries is a limiting factor.

Another top-down approach to modelling vegetation distributions is embodied in the National Vegetation Information System (NVIS), which is a collaborative initiative between the Australian and state and territory governments to manage national vegetation data to help improve vegetation planning and management within Australia (Department of the Environment, 2012). In New South Wales, a top-down classification of vegetation based on Keith (2004) has been adopted in the Vegetation Information System (VIS) (Office of Environment & Heritage, 2015a). While the 16 top-level flora categories from VIS do not directly align with the 32 top-level flora categories in NVIS, mapping of attributes facilitates interoperability (Office of Environment & Heritage, 2015b). In Victoria, each bioregion-EVC combination equates to a unique NVIS description, again providing a degree of interoperability (Department of Environment and Primary Industries, 2014b).
Figure 23: Comparison of published distributions of *Eucalyptus melliodora* in Victoria, Australia
7.1.3 The phenology of eucalypts

Phenology involves the recording of recurring natural events (Hudson et al., 2010a), and for eucalypts this includes patterns of seasonal growth and flowering. Flowering patterns of less than 30 of the 800 eucalypt species have been studied (Birtchnell and Gibson, 2006). The flowering periods of species (timing and duration of flowering) is listed in field guides (Brooker, 2002, Holliday and Watton, 2002, Boland et al., 2006, Brooker and Kleinig, 2006, Costermans, 2006, Nicolle, 2006) and texts on honey flora targeted at apiarists (Goodman, 1973, Clemson, 1985). Timing and durations for important melliferous eucalypt and other species are reproduced in Table 23. To the extent that there is agreement on timing and duration between the various different texts, the determination of whether a particular species in a particular location will flower in a given year at a particular time depends on a multitude of factors.

Following the "spring flush" of new growth (Jones et al., 2011), adult eucalypts develop flower buds in inflorescences borne in leaf axils (Potts and Gore, 1995). The buds develop enclosed in a bract, which is shed as the buds swell. The bud consists of a hollow receptacle covered by a cap-like opercula, which when shed at anthesis results in what is referred to as the flower. In common with other members of the family Myrtaceae the eucalypt flower features a large number of stamens. The filaments of these stamens are generally white in eucalypt species from eastern Australia, meaning that the flower itself appears white or yellow.
The time between floral initiation and anthesis varies amongst eucalypt species. Few scientific studies into budding to anthesis durations have been undertaken, but those that have found durations ranging from 9.5 months for *E. melliodora* (Moncur and Boland, 1989) to 29 months in the case of *E. regnans* (Ashton, 1975). Texts used by beekeepers highlight budding to anthesis durations for a much larger number of eucalypt species, and these are said to range between 6 to 24 months in Victoria (Goodman, 1973) and a few weeks to 2 years in New South Wales (Clemson, 1985). Across a selection of key melliferous species, there is reasonable agreement between scientific studies and beekeeping texts (Table 21). In a study of 51 species of eucalypts, Keatley and Hudson (1998) found that the development time of buds was longer in inflorescences which had a larger total volume, and that this volume is positively correlated with the timing of bud appearance. A consequence of the lengthy period between bud development and flowering is that climate influences on budding are therefore different to flowering (Hudson and Keatley, 2013).
Table 21: Budding to anthesis duration of a selection of eucalyps

<table>
<thead>
<tr>
<th>Eucalypt species</th>
<th>Budding to anthesis duration according to various sources (months)</th>
<th>Goodman (1973)</th>
<th>Clemson (1985)</th>
<th>Scientific studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. camaldulensis</td>
<td></td>
<td>9-12</td>
<td>11-12</td>
<td>N/A</td>
</tr>
<tr>
<td>E. cladocalyx</td>
<td></td>
<td>12-15</td>
<td>12</td>
<td>14(^{10})</td>
</tr>
<tr>
<td>E. gracilis</td>
<td></td>
<td>3-6</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>E. leptophylla</td>
<td>(^{11})</td>
<td>8-12</td>
<td>N/A</td>
<td>14(^{12})</td>
</tr>
<tr>
<td>E. leucoxylon</td>
<td></td>
<td>4-12</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>E. macroชนyla</td>
<td></td>
<td>15-18</td>
<td>27</td>
<td>N/A</td>
</tr>
<tr>
<td>E. melliodora</td>
<td></td>
<td>10-12</td>
<td>8-13</td>
<td>9.5(^{13})</td>
</tr>
<tr>
<td>E. microcarpa</td>
<td></td>
<td>3-6</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td>E. regnans</td>
<td></td>
<td>27</td>
<td>N/A</td>
<td>27-29(^{14})</td>
</tr>
<tr>
<td>E. spathulata</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
<td>26(^{15})</td>
</tr>
<tr>
<td>E. tricarpa</td>
<td>(^{16})</td>
<td>5-6</td>
<td>5-9</td>
<td>4-10(^{17})</td>
</tr>
</tbody>
</table>

The most intensively analysed long-term records of eucalypt flowering are those from the mid-20\(^{th}\) century recorded by the Forest Commission of Victoria and forest overseer Bill Sheen (Keatley et al., 2002). From the 1920s to until the early 1980s, State Forest Commissions in New South Wales, Queensland, Victoria and Western Australia kept records of eucalypt flowering but few of these records survived (Keatley et al., 2002). A list of studies based on the surviving Victorian data is provided in Table 22. In more recent years, there have only been two other studies conducted over multiple years and at a similar spatial scale. In New South Wales, Law et al. (2000) recorded monthly flowering data from 1982-

\(^{10}\) (Ellis and Sedgley, 1992)
\(^{11}\) Listed as E. foecunda
\(^{12}\) (Ellis and Sedgley, 1992)
\(^{13}\) (Moncur and Boland, 1989)
\(^{14}\) (Ashton, 1975)
\(^{15}\) (Ellis and Sedgley, 1992)
\(^{16}\) Listed as E. sideroxylon
\(^{17}\) (Porter, 1978), classified as E. sideroxylon
1992 for 20 species across 23 sites. In Victoria, flowering of 3 species was recorded at more than 60 sites during the period of 1995-2008 in six major survey programs (Mac Nally and Horrocks, 2000, Mac Nally and Horrocks, 2002, Radford et al., 2005, Radford and Bennett, 2007, Thomson et al., 2007, Mac Nally et al., 2009)
### Table 22: Eucalypt phenology studies based on Forest Commission of Victoria data

<table>
<thead>
<tr>
<th>Study</th>
<th>Location and duration of data used</th>
<th>Species studied</th>
<th>Factors identified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Havelock</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rushworth</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maryborough</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bendigo</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1950-1976</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Porter (1978)</td>
<td>1940-1970</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Keatley et al. (1999)</td>
<td>1940-1971</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Keatley et al. (2002)</td>
<td>1940-1962</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Keatley and Hudson (2007)</td>
<td>1940-1971</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>1945-1970</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Kim et al. (2009)</td>
<td>1940-1971</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hudson et al. (2010a)</td>
<td>1938-1972</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hudson et al. (2010b)</td>
<td>1940-1971</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hudson and Keatley (2013)</td>
<td>1940-1962</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Most eucalypt species do not flower every year. Ashton (1975) found that for *E. regnans* flowering followed a 2-year cycle with a tendency for heavier flowering every fourth year. Flowering intensity in heavier years was 25-65 times the magnitude of flowering intensity in poor years. Keatley and Hudson (2007) studied flowering data from the Forest Commission of Victoria for two sites in central Victoria; Rushworth and Havelock Nature Conservation Reserves. They found that *E. leucoxylon* flowered every year at both locations, *E. tricarpa* flowered in 9 out of 10 years at Rushworth and 8 out of 10 years at Havelock, and *E. microcarpa* flowered roughly 8 out of 10 years at both sites. *E. polyanthemos* was the least reliable flowering 7 out of 10 years at Rushworth, and 9 out of 10 years at Havelock. *E. melliodora* was very reliable at Havelock (only failing to flower once in 32 years) but less so at Rushworth (flowering 9 out of 10 years). These statistics do not account for variations in flowering intensity from year to year; peak flowering intensities were spread over the entire year between species and sites, with peak flowering for each species at Havelock tending to follow peak flowering at Rushworth by 1-2 months. In a survey of commercial beekeepers, Birtchnell and Gibson (2006) recorded that flowering frequency ranged from 1-7 years, although most species flowered every 2-4 years. The mode and range of flowering of important melliferous eucalypts and other species are shown in Table 23. Even in an off-cycle year where less than 70% of individuals within a species produce buds that persist to anthesis, it is likely that there will be some individual trees within each forest block that will flower.


Table 23: Mode and timing of flowering of Victorian melliferous species

<table>
<thead>
<tr>
<th>Scientific name</th>
<th>Eucalyptus camaldulensis</th>
<th>Eucalyptus microcarpa</th>
<th>Eucalyptus tenuifolia</th>
<th>Eucalyptus melliodora</th>
<th>Eucalyptus gracillima</th>
<th>Eucalyptus macrocarpa</th>
<th>Eucalyptus obliqua</th>
<th>Eucalyptus tricarpa</th>
<th>Brassica napus</th>
<th>Arctotheca calendula</th>
<th>Echium plantagineum</th>
<th>Prunus amygdalus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common name</td>
<td>River red gum</td>
<td>Grey box</td>
<td>Yellow gum</td>
<td>Yellow box</td>
<td>White mallee</td>
<td>Red stringybark</td>
<td>Messmate</td>
<td>Ironbark</td>
<td>Canola</td>
<td>Capeweed</td>
<td>Paterson's curse</td>
<td>Almond</td>
</tr>
<tr>
<td>Mode</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Range</td>
<td>2-5</td>
<td>2-3</td>
<td>1-3</td>
<td>2-3</td>
<td>1-5</td>
<td>3-7</td>
<td>2-4</td>
<td>2-5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>January</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>February</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>March</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>April</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>May</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>June</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>July</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>August</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>September</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>October</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>November</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>December</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
</tr>
</tbody>
</table>

Potter et al. (1999)  
Goodman (1973)  
Parsons and Cuthbertson (2001)  
Gordon et al. (2014)
Environmental factors, especially temperature and rainfall, are important predictors of eucalypt flowering. Rainfall was found to be a significant factor in five out of eight Victorian regions studied by Porter (1975). By manipulating environmental temperatures using greenhouses, Moncur (1992) found that *E. lansdowneana* commenced floral development when released from a period of cold stress. Similarly, Law et al. (2000) established that cool temperatures prior to floral budding were a strong predictor of flowering for nine eucalypt species. Across all species, the period of greatest flowering occurred 9 months after the highest monthly rainfall recorded during the 10 year survey period. Using Forest Commission of Victoria flowering data for four species (*E. leucoxylon*, *E. microcarpa*, *E. polyanthemos* and *E. tricarpa*) across 23 years (from 1940-1962), Keatley et al. (2002) found that the combination of temperature and rainfall (with temperature exerting the greatest influence) was significantly correlated with flowering. Using data from 1940-1971, Hudson et al. (2010a) showed that flowering was significantly, and non-linearly, influenced by temperature but that rainfall was not a significant factor. This contrasts with the earlier study of Keatley et al. (2002) where rainfall was found to be significant for flowering commencement of both *E. tricarpa* and *E. leucoxylon*, and Kim et al. (2009), where rainfall had a direct positive impact on *E. tricarpa*. Hudson et al. (2010a) determined that the main temperature driver of *E. leucoxylon* flowering is minimum temperature, maximum temperature for *E. polyanthemos*, both minimum and maximum temperature for *E. tricarpa*, and mean temperature for *E. microcarpa*. In a study of six eucalypt species, Rawal
et al. (2014) determined that temperature, soil moisture availability, air humidity and photoperiod length were all important variables controlling the growth phenology, although the effects of each variable varied between species suggesting different ecological niches and adaptability.

Few studies have addressed the effects of fire and logging on the flowering of eucalypts. Law et al. (2000) did not find any evidence of the effect of logging on flowering at two sites where logging occurred, and found the effect of fire on flowering to vary between species. Logging was reported to increase flowering of *E. microcarpa* by Birtchnell and Gibson (2006).

In most states of Australia, for a commercial beekeeping operation to remain viable a beekeeper must become an expert in predicting eucalypt flowering at their apiary sites. Birtchnell and Gibson (2008) surveyed the top tools used by apiarists to predict the likelihood of flowering event occurring in melliferous flora. In order of importance, these were:

1. Rainfall patterns/timing of seasonal rainfall (n=32),
2. History of site (n=25),
3. Rainfall events (n=25),
4. Length of growth spike (n=21),
5. General budding (n=20),
6. Tree health, specifically the colour and look (n=16),
7. Buds (n=15), and
8. Budding intensity (n=10),
where \( n \) = number of beekeepers using the tool out of 44 survey respondents.

For a beekeeper to use these each of these tools requires access to data or direct observation. Rainfall data (tools 1 and 3) are readily available from the Bureau of Meteorology (Bureau of Meteorology, 2015a). The history of each site in terms of flowering and nectar production (tool 2) is recorded by individual beekeepers for their own sites, and for commercial reasons is rarely shared. There is no regulatory requirement for beekeepers to report individual site history, even for those sites on public land. At an aggregate level honey packers record production data for their suppliers, but with the exception of statistics collected in Victoria from 1957-1978 (Commonwealth Bureau of Census and Statistics Victorian Office, 1957, Australian Bureau of Statistics Victorian Office, 1970) there is no detailed information on honey production available in the public domain. The remaining tools (Tools 4-8) are all phenomena that are correlated with data available via remote sensing.

7.1.4 Remote sensing of vegetation

Remote sensing studies the radiation emanating from the earth's surface, and in the case of the remote sensing of vegetation much of this radiation is reflected radiation (Jones and Vaughan, 2010). In a forest, solar radiation incident on the canopy encounters leaves, stems, soil and water each of which has different spectral properties in terms of reflectance, absorption and transmission (Jones and Vaughan, 2010).
Leaf structures vary between species, ranging from flat broad leaves to needle shaped leaves in conifer species. To prevent water loss, most leaves have an epidermis covered in a thick layer or cuticle which is often waxy (Jones and Vaughan, 2010). Along with a waxy coating, a blueish grey-green colour is characteristic of the leaves of a number of *Eucalyptus* species and of the juvenile leaves of many others, and this adaptation benefits the plant by increasing light reflectance, decreasing transpiration rates, promoting water shedding, inhibiting fungal attacks and increasing frost resistance (Cameron, 1970). The leaves of many plants are covered in hairs on one or both surfaces (Jones and Vaughan, 2010). While seedling and juvenile eucalypts sometimes are covered with hairs, the leaves of adult plants are generally hairless (Brooker and Kleinig, 2006). The epidermis also features pores (stomata) through which gas exchange occurs, and these stomata are flanked by guard cells which can close the stomata thus preventing water loss (Mauseth, 2014). While for most species stomata occur on the lower epidermis, eucalypts hold their leaves vertically and thus stomata occur on both sides of the leaf (Mauseth, 2014). Characteristic of most eucalypt leaves in southern Australia are oil glands, and four characteristic patterns of gland reticulation exist (Brooker and Kleinig, 2006). Chlorophyll in the leaves is concentrated in chloroplasts (Mauseth, 2014), and while most higher plant leaves are flat to optimally spread out this photosynthetic tissue some species such as conifers have needle or cylindrical shaped leaves (Jones and Vaughan, 2010).
The spectral properties of healthy leaves are dominated in visible wavelengths by chlorophyll, which strongly absorbs incident radiation in the visible wavelengths (400 nm-700 nm) with a dip in absorption in the green (500 nm) resulting in the human-perceived green colour of vegetation (Jones and Vaughan, 2010). Vegetation absorbs relatively little radiation in the infrared, with the exception of water absorption bands at 1450 nm, 1950 nm and 2500 nm (Jones and Vaughan, 2010).

The presence or absence, arrangement, density and orientation of structures within leaves of individual species collectively determine their specific radiative properties (Jones and Vaughan, 2010). Reflection and absorption occur not just at the leaf surface but also within the leaf, and volume scattering occurs from the internal cell surfaces (Jones and Vaughan, 2010). To abstract the complexity of modelling leaf radiative properties, parameterised models such as PROSPECT (for broadleaves) and LIBERTY (for conifers) have been developed (Jones and Vaughan, 2010). PROSPECT has been used to model the radiative properties of *E. globulus* leaves (Barry *et al.*, 2009b).

Brooker and Kleinig (2006) observe that the leaves of eucalypt seedlings often are a different shape, size and also colour to those of mature plants. England and Attiwill (2006) found that there are significant changes in *E. regnans* leaf form with increasing tree age and height. As trees grew and aged, leaf size decreased, leaf thickness increased, cuticle thickness and waxiness increased, and leaves became narrower relative to their length. In contrast, Greaves and Spencer (1993) states that the leaves of many juvenile eucalypts are not just broader but also thicker than leaves from
mature trees with a waxy coating. All of these changes in leaf form with age are likely to impact the radiative properties of leaves and hence of the canopy.

Patterns of seasonal growth and senescence can be observed in leaves by monitoring variations in the concentration of different growth pigments. The dominant pigment influencing changes in the spectral properties of eucalyptus leaves across the growing season is chlorophyll (Datt, 1998, Datt, 1999b, Datt, 1999a, Coops et al., 2003, Barry et al., 2009b). New foliage growth; the "spring flush", is often tinged red reflecting concentrations of the pigment anthocyanin (Barry et al., 2008, Barry et al., 2009a). Finally, stresses on the plant are often reflected in a change of pigment concentrations and spectral properties with examples including drought, pests and diseases (Stone et al., 2000, Stone et al., 2001, Coops et al., 2004).

Just as the structure of leaves determines their radiative properties, the structure of canopies determines their radiative properties at remote sensing resolutions. Each element in the observed ecosystem such as leaves, stems, branches, and soil contributes to the observed reflected radiation. The spatial orientations of not just the elements, but also the incident radiation and sensor are important (Jones and Vaughan, 2010). Radiation is scattered by leaves in different orientation and secondary and even tertiary reflectances can occur before radiation is reflected back to the sky (Jones and Vaughan, 2010). As eucalyptus leaves generally hang near vertically, their canopies are semi-transparent and therefore reflectance measurements from above are influenced by reflectance from
the understory (Greaves and Spencer, 1993). Open and overlapping canopies also create challenges in remote sensing of eucalypt vegetation (Youngentob et al., 2011).

Radiation passing through the canopy and hitting the ground will be either absorbed or reflected. Soils typically exhibit a smooth increase in reflectance from the visible to the near infrared, with lighter sandy soils reflecting more radiation than clays, which in turn reflect more radiation than loams (Jones and Vaughan, 2010). Water reflects very little radiation, typically less than 3% across visible, near and mid-infrared regions (Jones and Vaughan, 2010). For this reason, wet soils reflect less radiation than dry soils.

Atmospheric effects must also be accounted for in the remote sensing of vegetation from satellite platforms. Atmospheric effects can be categorised into two classes; absorption and scattering (Gobron et al., 1999). By carefully choosing the sensor spectral bands, absorption of radiation by specific gases can generally be avoided (Gobron et al., 1999). Rayleigh scattering occurs when atmospheric particles have small diameters relative to the wavelength of the radiation, and can occur with larger molecules of atmospheric gases such as oxygen and nitrogen even in the absence of atmospheric impurities (Campbell and Wynne, 2011). Both Rayleigh and aerosol scattering are wavelength dependent, with ultraviolet light scattered about 16 times as much as red light, and blue light scattered about 4 times as much as red light (Campbell and Wynne, 2011). Eucalypt forests are abundant sources of gaseous molecules such as terpenes (Ristovski et al., 2010) and condensable vapours (Suni et al.,
Chapter 7: The Spatio-Temporal Distribution of Eucalypt Flowering

2008) but no studies have been done on the specific effect of these on remote sensing observations.


While direct sensing of flowering of some melliferous species is possible, no study has yet proven the feasibility of directly observing eucalypt flowering using remote sensing data. Remote sensing data has been used to observe Paterson's Curse (Echium plantagineum) (Ullah, 1989, Bulman, 2004), canola (Brassica napus) (Sulik and Long, 2015) and lucerne (Medicago sativa) (Wardlow et al., 2007, Wardlow and Egbert, 2008, Wardlow and Egbert, 2010), all of which are important nectar and pollen plants for managed honey bees. The challenge of directly observing flowering in a eucalypt forest is considerably greater due to
heterogeneity of eucalypt species, heterogeneity of flowering intensity within a eucalypt species, and sparseness of canopy cover in many forests all contributing to a lower density of flowering per pixel. A further challenge is that most of the flowers of important honey-producing eucalypts such as *E. tricarpa* and *E. leucoxylon* hang downwards (Wilson, 2003) and therefore cannot be observed from above.

While direct observations of flowering may be useful for deriving historical predictions of beekeepers movements and for determining resource availability for feral colonies, it has limited use in predicting the future movements of beekeepers. Beekeepers need to know ahead of time when flowering events are going to occur to plan their migratory movements; for this reason it becomes necessary to predict rather than simply observe flowering events. The prediction of flowering events using remote sensing data requires proxies to be developed for the main tools (Birtchnell and Gibson, 2008) used by beekeepers to predict flowering. The best candidate for predicting flowering is observed vegetation growth tracked using vegetation indices.

### 7.1.5 Vegetation indices

Bannari *et al.* (1995, p.96) define an index as a "number quantifying the intensity of a phenomenon which is too complex to be decomposed into known parameters". Vegetation indices (VI) attempt to measure biomass or vegetative vigour and are "formed from combinations of several spectral values that are added, divided, or multiplied in a manner designed to yield a single value that indicates the amount or vigour of
Vegetation indices are designed such that they are more clearly related to the biophysical parameters of interest such as vegetation cover, leaf pigments, leaf proteins or leaf water than any of the original bands (Jones and Vaughan, 2010). While vegetation indices are most directly related to fractional vegetation cover or the fraction of absorbed photosynthetically active radiation (fAPAR), they are also commonly used as estimators of leaf-area index (LAI) (Jones and Vaughan, 2010).

The reflectance of vegetation changes sharply moving from the red into the near-infrared at around 700 nm, and the point of maximum slope, the red edge position (REP) is strongly correlated with foliar chlorophyll content (Dawson and Curran, 1998). Unfortunately there is no generally accepted technique for estimating REP and so each technique produces a different value for REP from the same set of data (Dash et al., 2009). Furthermore, there is an asymptotic relationship between REP and chlorophyll content and therefore estimates of foliar chlorophyll content at high concentrations of chlorophyll concentration are inaccurate (Dash et al., 2009).

Due to the complexity in interpreting the spectral response of vegetated areas, more than 50 vegetation indices have been developed in the last 40 years with tabulations of indices available in many papers and texts (Bannari et al., 1995, le Maire et al., 2004, Jones and Vaughan, 2010, Croft et al., 2014).
Optimised vegetation indices such as those calculated for the MODIS product sets (outlined in 7.1.6.2) are typically calculated by a processing flow as shown in Figure 24.

![Figure 24: Vegetation index processing flow (Huete et al., 1999)](image)

In the following sections we focus on the vegetation indices of most interest to the study of eucalypt phenology on the basis of pre-processed data for more than a decade being available directly from NASA and ESA.

7.1.5.1 The Normalised Difference Vegetation Index (NDVI)

The normalised difference vegetation index (NDVI) transforms the NIR to red reflectance ratio, $\rho_{\text{NIR}}/\rho_{\text{red}}$ to a normalised form where VI values range from -1 to +1, and is commonly expressed as:

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}}}$$

where $\rho_x$ are the full or partially atmospheric-corrected surface reflectances (Solano et al., 2010).

The NDVI is straightforward to compute for remote sensing data from a large number of different sensors and one of the key advantages of the NDVI is hence the length of the data record available. Brown et al. (2006) conducted a study of NDVI records derived from Advanced Very High Resolution Radiometer (AVHRR), SPOT-Vegetation, SeaWiFS, Moderate
Resolution Imaging Spectroradiometer (MODIS), and Landsat ETM+ sensors. They found that continuity of the NDVI data-record is achievable given the similarity of the datasets. The Land Long Term Data Record (LTDR) realises this possibility and provides a 34-year continuous NDVI dataset spanning the AVHRR and MODIS sensors (Masuoka, 2015).

As a ratio index, the NDVI also provides the advantage of reducing certain types of positively-correlated band-correlated noise due to clouds, sun and view angles, topography, and calibration errors (Solano et al., 2010). However, as a ratio index one of the most significant disadvantages of the NDVI is its inherent non-linear relationship with biophysical characteristics such as vegetation fraction (fV), leaf area index (LAI), and aboveground biomass (Jiang and Huete, 2010) and asymptotic behaviours (Jiang and Huete, 2010, Solano et al., 2010). For highly vegetated areas, the NDVI can saturate and thus become insensitive to changes in chlorophyll content (Jiang and Huete, 2010). Ratio indices also fail to account for canopy bidirectional reflectance anisotropies, especially in the case of canopy shadowing (Solano et al., 2010).

A further disadvantage of the NDVI is that is strongly influenced by soil background brightness and atmospheric contaminations (Jiang et al., 2008, Jiang and Huete, 2010). The soil-adjusted vegetation index (SAVI) and atmospherically resistant vegetation index (ARVI) have both been proposed to mitigate these limitations of the NDVI (Jiang et al., 2008).
7.1.5.2 The Enhanced Vegetation Index (EVI)

The enhanced vegetation index (EVI) was developed to compensate the major shortcomings of the NDVI and build on the improvements offered by both SAVI and ARVI. The EVI thus aims to provide improved sensitivity in high biomass areas, de-coupling of canopy background signals and resistance to atmospheric influences (Huete et al., 2002). The EVI includes adjustment factors to compensate for nonlinear, differential red and NIR transfer through a canopy (Huete et al., 2002) similar to SAVI (Jiang et al., 2008). As noted in Section 7.1.4, atmospheric scattering for blue wavelengths is larger than for red wavelengths, and hence when aerosol concentrations are higher the difference between the two bands increases (Solano et al., 2010). This information can then be used in vegetation indices to stabilise the index against variations in aerosol concentration levels (Solano et al., 2010). ARVI and EVI both incorporate this concept (Jiang et al., 2008, Solano et al., 2010).

EVI is determined based on the following formula:

$$
EVI = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + C_1 \cdot \rho_{\text{red}} - C_2 \cdot \rho_{\text{blue}} + L} \cdot (1 + L)
$$

where $\rho_x$ are the full or partially atmospheric-corrected (for Rayleigh scattering and ozone absorption) surface reflectances, $L$ is the canopy background adjustment for correcting nonlinear, differential NIR and red radiant transfer through a canopy; $C_1$ and $C_2$ are the coefficients of the aerosol resistance term (which uses the blue band to correct for aerosol influences in the red band) (Solano et al., 2010). The coefficients adopted in the EVI have been empirically determined as $L = 1$, $C_1 = 6$ and $C_2 = 7.5$. 
(Huete et al., 1999). Over high reflectance surfaces such as clouds, snow and ice, the EVI is replaced by a modified 2-band EVI which does not make use of the blue band (Jiang et al., 2008, Solano et al., 2010):

$$2\text{-band EVI} = 2.5 \cdot \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}} + 1}$$

The high reflectance of the blue band over such surfaces would otherwise lead to an atmospheric over-correction condition.

7.1.5.3 The MERIS Global Vegetation Index (MGVI)

The MERIS Global Vegetation Index (MGVI) has been designed to have a positive linear relationship with the fraction of absorbed photosynthetically active radiation (fAPAR) (Zurita-Milla et al., 2009). fAPAR is defined as the fraction of photosynthetically active radiation (PAR) absorbed by green elements of healthy vegetation (Pickett-Heaps et al., 2014). fAPAR is a variable directly related to canopy functioning, especially productivity (Jones and Vaughan, 2010). While most scaled vegetation indices are related to fAPAR, the relationships are not usually linear (Jones and Vaughan, 2010). MGVI is hence a useful but little-used source of data for phenological studies (Boyd et al., 2011).

The MGVI is computed using top of atmosphere reflectances in the blue, red and NIR MERIS bands (Table 24), where similar to EVI the blue band is used to correct the red and NIR reflectances for atmospheric scattering (Gobron, 2011).

fAPAR can also be calculated for other platforms and sensors and NASA provide fAPAR data as a standard MODIS product. In an evaluation of six
fAPAR products across the Australian continent using multi-year records, Pickett-Heaps et al. (2014) found significant differences between products, especially in the case of forest-classified biomes. A similar study in Europe led D’Odorico et al. (2014) to conclude that fAPAR products cannot yet be reliably fed into existing biogeochemical process models.

7.1.5.4 The MERIS Terrestrial Chlorophyll Index (MTCI)

The MTCI is derived as the ratio of the differences between three bands in the red/NIR (Dash and Curran, 2007):

\[
\text{MTCI} = \frac{\rho_{753.75} - \rho_{708.75}}{\rho_{708.75} - \rho_{681.25}}
\]

For the MERIS sensor, the reflectances \(\rho_{681.25}, \rho_{708.75}\) and \(\rho_{753.75}\) correspond to the reflectances of the centre wavelengths of sensor bands 8, 9 and 10 respectively (Table 24).

The MTCI has shown to have a close correspondence with chlorophyll content, notably at high values, (Dash and Curran, 2006, Dash and Curran, 2007, Dash et al., 2009).

7.1.6 Satellite platforms and sensors

There are many different satellite platforms launched since the 1970s which provide data relevant to the remote sensing of vegetation. Trade-offs are inherent between the spatial, temporal and often spectral resolutions of the sensors on the different platforms.
Compared to land-based phenomenon, atmospheric and oceanographic phenomenon exist over larger spatial scales and vary more rapidly (Jones and Vaughan, 2010). For this reason, many satellite systems designed for meteorological and oceanographic remote sensing feature both high temporal resolution and low spatial resolution (typically 1 km – 5 km).

One of the earliest constellations of weather satellites also used for vegetation studies is the Polar-orbiting Operational Environmental Satellite (POES) operated by the National Oceanic and Atmospheric Administration (NOAA) (Jones and Vaughan, 2010). The 5-channel Advanced Very High Resolution Radiometer (AVHRR/2) was carried on NOAA-7 which launched in 1981, and subsequently on NOAA-9, NOAA-11, NOAA-12, NOAA-13 and NOAA-14. NOAA-15, featuring the 6-channel AVHRR/3, launched in 1998 (Pichel et al., 2001). The AVHRR was well suited to production of NDVI data, with a resolution at nadir of 1.1 km. Individual satellites have a revisit time (at the equator) of 12 hours, although with two satellites in orbit at any one time this effectively provides a revisit time of 6 hours (Jones and Vaughan, 2010).

The main strength of the AVHRR data sets lie in the length of their temporal record (Boyd et al., 2011). To provide continuity for AVHRR data beyond the mission life of current NOAA platforms, EUMETSAT's series of three Metop satellites (Metop-A, Metop-B and Metop-C) carry the AVHRR/3 sensor. Metop-A was launched in 2006 and Metop-B launched in 2012; Metop-C is yet to launch (Pinzon and Tucker, 2014).
Continuity for NOAA’s POES system is also provided by Suomi-NPP, launched in 2011. Suomi-NPP is equipped with a new sensor the Visible/Infrared Imaging Radiometer Suite (VIIRS), which significantly outperforms AVHRR in all aspects of spatial, spectral, and radiometric resolution and accuracy (Changyong et al., 2014). The VIIRS Environmental Data Record (EDR) provides both NDVI and EVI data globally, with NDVI at 375 m resolution and EVI at 750 m resolution (Miura, 2015).

Medium resolution systems have typically been developed to provide more specialised support of land remote sensing applications. Because they provide a reasonable trade-off in terms of spatial, temporal, and spectral resolution, medium resolution satellite platforms and their sensors are the most valuable for the study of eucalypt phenology.

Onboard the Terra (EOS) satellite launched by NASA in 1999 is the Moderate-resolution Imaging Spectroradiometer (MODIS) (Maccherone, 2015). MODIS features 36 bands with resolutions ranging from 250 m to 1 km (Maccherone, 2015). The Terra satellite has a revisit time of 1-2 days (Jones and Vaughan, 2010), and because MODIS images the entire earth’s surface the combination of high temporal with medium spectral and spatial resolutions make it a strong choice for phenology studies.

The European Space Agency (ESA) Envisat satellite was launched in 2002 and in 2012 contact with the satellite was lost and the end of mission declared (ESA, 2012a). Amongst other sensors, Envisat supported the Medium Resolution Imaging Spectrometer (MERIS) which had 15 bands
Chapter 7: The Spatio-Temporal Distribution of Eucalypt Flowering

(ESA, 2012b) at a resolution at nadir of 260 m × 300 m (Rast et al., 1999) with a revisit time of 3 days (Dash and Curran, 2007). The Sentinel-3 system will provide data continuity for the Envisat system and MERIS instrument with the first launch scheduled for late 2015 (ESA, 2015).

The SPOT-4 and SPOT-5 satellites, launched in 1998 and 2002 respectively, both carried the VEGETATION (VGT) sensor with three bands in the visible and NIR and one band in the shortwave infrared with a 1 km resolution (Kramer, 2015). As both satellites have now been retired, continuity is being provided by the PROBA-V satellite with three bands in the visible and NIR at 300 m resolution and one band in the shortwave infrared with 600 m resolution (Kramer, 2015). SPOT VGT and PROBA-V both provide near daily revisit frequencies (Saint, 1994, Kramer, 2015).

With over 40 years of accumulated data, the Landsat satellites (Landsat 1, 2, 3, 4, 5, 7 and 8) provide the longest temporal record of space-based land surface observations (Roy et al., 2014). The Landsat 7 Enhanced Thematic Mapper + features three bands in the visible, one in the NIR, two in the SWIR and two in the thermal IR all at 30 m resolution, and a panchromatic band at 15 m resolution (Roy et al., 2014). The Operational Land Imager (OLI) and Thermal Infrared Sensor (TIR) on Landsat 8 together offer broadly comparable bands to the ETM+ and add a shorter wavelength blue band and new SWIR band (Roy et al., 2014). The revisit time of the Landsat satellites is 16 days, and this combined with cloud cover can limit the reliable documentation of phenology (Roy et al., 2014).
Higher spectral resolution is achieved by the Hyperion sensor, on board the EO-1 spacecraft launched in 2000 (Thenkabail et al., 2012). Hyperion features 196 bands in the visible, NIR and SWIR with 30 m resolution and a 16-day revisit time (Thenkabail et al., 2012). Unlike other sensors such as MODIS, Hyperion does not acquire data over the entire land surface. Since launch, only 35 swathes (each 7.5 km × 100 km) have been captured of Victoria, Australia (USGS, 2015); each swath equates to only 0.3% of the state’s area. This lack of historical data limits the usefulness of Hyperion data for phenology studies in Australia.

DigitalGlobe™ is a commercial operator of several satellite platforms including IKONOS, Quickbird, GeoEye-1, Worldview-1, Worldview-2, and Worldview-3 (DigitalGlobe, 2015a). DigitalGlobe™ provide multispectral data suitable for remote sensing of vegetation at high spatial resolution with short revisit times; up to 1.24 m resolution with a revisit time of < 1 day in the case of Worldview-3 (DigitalGlobe, 2015b). The cost of acquiring the large datasets required for phenology analysis and the lack of historical data mean that commercial data is seldom used in non-commercial research studies.

7.1.6.1 MODIS, MERIS and SPOT VGT/PROBA-V compared

For study of eucalypt phenology, medium resolution spectrometers such as MODIS and MERIS and SPOT VGT/PROBA-V currently offer the most useful data sets.

From a temporal resolution perspective, predicting flowering events in real-time is likely to require sensor data from a satellite with multiple
revisits per week; a precondition satisfied by all three sensors. As there is no history of use of vegetation indices within the beekeeping industry there is currently no baseline for how frequently beekeepers require or desire growth data for sites of interest, but we can reasonably assume the frequency needs to be more than monthly but less than daily. MODIS data are available in a pre-composited format based on a 16-day composite, (NASA Land Processes Distributed Active Archive Center (LP DAAC), 2015), MERIS is available in a weekly composite (UK Multi-Mission Product Archive Facility et al., 2013) and SPOT VGT/PROBA-V is available in a 10-day composite (VITO, 2010). The goal of the compositing methodology is to extract a single value per pixel from all the retained filtered data, which is representative of each pixel over the compositing period (Solano et al., 2010). In the filtering process only nadir view pixels without atmospheric or cloud cover contamination are retained. In the case of MODIS data and with global mean cloud cover in the range of 50-60% the number of retained pixels from a 16-day compositing period is typically less than 10 and often varies between 1 and 5 (Solano et al., 2010). Cloud cover in Australia typically ranges from 3 to 3.25 oktas (Jovanovic et al., 2011), where each okta represents 1/8th (12.5%) of the sky covered by cloud. This suggests that cloud cover in remote sensing imagery over Australia will average around 40% and that the pre-processed composite imagery will be of greater value in phenology studies than individual images. Composite imagery also significantly reduces the storage and processing requirements on the computing systems used for their analysis.
With spatial resolutions of 500 m, 260 m × 300 m, 1 km and 300 m respectively, pixels from MODIS, MERIS, SPOT VGT and PROBA-V represent areas ranging from 8 ha to 100 ha in size. The average areas of public forests in Victoria in which beekeeping is permitted range from 70 ha for NCR to 8063 ha for State Forests (Environment Conservation Council, 2001). The 4 km foraging range of honey bee workers (Seeley, 1995) translates to a foraging area of approximately 5000 ha. On the basis of typical forest size and honey bee foraging areas, MODIS, MERIS and PROBA-V spatial resolutions appear adequate for analysing honey bee floral resource phenology.

All three sensors have spectral bands suitable for observing vegetation properties (Table 24), with MERIS's higher spectral resolution in the red/NIR region allowing production of the MTCI (Dash and Curran, 2007). While MERIS bands are also suitable for production of NDVI and EVI, these are not provided as standard products and must therefore be processed from other data sets (Fensholt et al., 2006, Croft et al., 2014). The bandwidth of the SPOT-VGT/PROBA-V spectral bands is considerably broader than either MODIS or MERIS, with bands in the red and NIR overlapping. Just as MODIS supports the construction of less VIs than MERIS, so too SPOT-VGT/PROBA-V supports less VIs than MODIS.
Table 24: Comparison of MODIS and MERIS spectral bands

<table>
<thead>
<tr>
<th>Band</th>
<th>Band no.</th>
<th>MODIS Bandwidth (nm)</th>
<th>Spatial resolution</th>
<th>NDVI</th>
<th>EVI</th>
<th>MERIS Bandwidth (nm)</th>
<th>Spatial resolution</th>
<th>MGVI</th>
<th>MTCI</th>
<th>SPOT-VGT/PROBA-V Bandwidth (nm)</th>
<th>Spatial resolution</th>
<th>NDVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>3</td>
<td>459 - 479</td>
<td>500 m</td>
<td>not used</td>
<td>used</td>
<td>2</td>
<td>437.5-447.5</td>
<td>260 m x 300 m</td>
<td>not used</td>
<td>not used</td>
<td>0</td>
<td>415-500</td>
</tr>
<tr>
<td>Red</td>
<td>1</td>
<td>620 - 670</td>
<td>250 m</td>
<td>used</td>
<td>used</td>
<td>8</td>
<td>677.5-685</td>
<td>260 m x 300 m</td>
<td>used</td>
<td>used</td>
<td>1</td>
<td>580-770</td>
</tr>
<tr>
<td>NIR</td>
<td>2</td>
<td>841 - 876</td>
<td>250 m</td>
<td>used</td>
<td>used</td>
<td>13</td>
<td>855-875</td>
<td>260 m x 300 m</td>
<td>used</td>
<td>not used</td>
<td>2</td>
<td>730-960</td>
</tr>
</tbody>
</table>

18 Bandwidth is programmable between 2.5 and 30 nm (Boyd et al., 2011).
19 While several sources (Zurita-Milla et al., 2009, ESA, 2012b) label the band centred on 865 nm as band number 13, Gobron (2011) contradicts this, labelling this band as band number 14.
All three sensors can provide at least a decade’s worth of data. SPOT-VGT, and now PROBA-V data is available continuously since 1998. MODIS data is available continuously since 2000. MERIS data is available for the period from 2002 until 2012, and unfortunately there will be a gap in the data history until Sentinel-3 commences collection in late 2015.

In summary, given the lack of continuous data history for MERIS and the courser spectral and spatial resolution of SPOT-VGT data, the best possible choice of medium resolution sensor for the study of honey bee floral resource phenology is MODIS.

7.1.6.2 MODIS products and tools

Satellite data products are generally categorised by levels ranging from 0 to 4, based on the amount of processing undertaken to produce the resulting data set. Data from raw satellite feeds (Level 0) is radiometrically calibrated (Level 1) then atmospherically corrected to yield a surface reflectance product (Level 2) (Vermote et al., 2011). MODIS vegetation indices products are built from daily Level 2 surface reflectance products (MOD09) (Solano et al., 2010). The MODIS VIs (MOD13) are Level 3 products as they are gridded onto a map projection (sinusoidal in this case) and composited over either 16-day or monthly periods (Solano et al., 2010).

Each MODIS VI tile is approximately 1200 km × 1200 km in dimension and uniquely identified by a row and column number. The coverage of
MODIS VI tiles across Australia is displayed in Figure 25; a total of 16 tiles are required to provide full national coverage.

Figure 25: MODIS VI tile coverage for Australia (Google Earth, 2013)

The MODIS 16-day composite VI at 250 m resolution is designated MOD13Q1. The data set is currently in its fifth edition; Collection 5 with improvements over previous collections discussed in Solano et al. (2010).

A further consideration with any data set is the latency between acquisition of the data by the sensor and availability of data on internet-accessible servers. For Layer 2 collections such as MOD09 the latency is generally less than 2 hours (Masuoka, 2010). However, for MOD13Q1 the latency for the 16-day composite can be up to 20 days (Gusso et al., 2014). Thus remote sensing observations from, for instance, early
January may not be available in composite form until early February. Conversely, predictions of flowering events within the current compositing period must be made without data from the previous compositing period.

7.1.7 Remote sensing for phenology

For over 30 years, satellite based remote sensing has been used in the study of seasonal vegetation dynamics (Boyd et al., 2011). The seasonal pattern of variation in vegetated land surfaces observed from remote sensing has become generally known as land surface phenology (Friedl et al., 2006). Friedl et al. (2006) are careful to distinguish land surface phenology from traditional vegetation phenology which refers to specific life cycle events such as budbreak, flowering and leaf senescence. Land surface phenology is related to landscape-level phenological event changes (Boyd et al., 2011), and climate processes operating at seasonal and inter-annual time scales are identifiable in the phenology of vegetation over large areas of the earth (Friedl et al., 2006).

Land surface phenology analysis is frequently performed with the aid of time series constructed from vegetation indices, which provide a powerful tool for following phenological development of crops and natural ecosystems (Jones and Vaughan, 2010). A frequent aim of such analyses is to extract key phenological metrics of interest, defined by Zhang et al. (2006) as:

1. **greenup**, the date of onset of greenness increase,
Chapter 7: The Spatio-Temporal Distribution of Eucalypt Flowering

2. **maturity**, the date at which canopy greenness approaches its seasonal maximum

3. **senescence**, the date at which canopy greenness begins to decrease, and

4. **dormancy**, the date at which canopy greenness reaches a minimum.

There are many different platforms and for each of these platforms one or more vegetation indices that can be used for land surface phenology. The choice of platform will govern the duration and resolution of data available for analysis. The choice of vegetation index will impact the sensitivity to changes in the phenological events of interest and the availability or otherwise of pre-processed VI data. Perhaps unsurprisingly given the long data record and pre-processed data sets available, phenology studies using NDVI data from the AVHRR dominate the literature (Table 25). In a comparison of phenological profiles and key phenological event dates derived from MERIS MTCI, MERIS MGVI, MODIS NDVI and MODIS EVI, Boyd et al. (2011) found that the uncertainties arising from the differences were as large as those arising from climatic perturbations. (Jeganathan et al., 2010) found spatial similarity between EVI and MTCI estimates of phenological variables, and that on average MTCI had an earlier detection of the onset of greenness (OG). Among the 15 biome classes analysed, the greatest differences between EVI and MTCI estimates were for evergreen needle leaf, mixed forest and cropland. Unfortunately, no comparative studies of land
surface phenology using different vegetation indices have been conducted for Australian ecosystems.

As discussed in Section 7.1.6.2, cloud cover and other data gaps in daily images favour the use of composite products in time series analysis. Time series constructed from composite products do still contain noise due to clouds and poor atmospheric conditions, and this type of noise typically depresses VI values (Chen et al., 2004). Additional white noise characterised by a zero mean is also introduced by BRDF and soil background effects (Atzberger and Eilers, 2011b).

A number of strategies have been developed to filter the noise from time series of vegetation indices (Table 25). The effects of noise reduction techniques vary for different land cover types and study areas, and for this reason (Michishita et al., 2014) recommend that an evaluation of techniques should be conducted for each study to determine the most appropriate choice. Only a single comparative study of filtering techniques has been conducted using Australian data (Lovell and Graetz, 2001).
Table 25: Noise filtering techniques used in land surface phenology studies

<table>
<thead>
<tr>
<th>Type of filter</th>
<th>Region</th>
<th>Satellite/VI</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harmonic series and Fourier transform</td>
<td>Global</td>
<td>AVHRR NDVI</td>
<td>Sellers et al. (1994)</td>
</tr>
<tr>
<td></td>
<td>Europe</td>
<td>AVHRR NDVI</td>
<td>Roerink et al. (2000)</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>AVHRR fAPAR</td>
<td>McCloy and Lucht (2004)</td>
</tr>
<tr>
<td></td>
<td>Southeast Asia</td>
<td>MVC SPOT VGT-S</td>
<td>Chen et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>North Scandanavia</td>
<td>MODIS NDVI</td>
<td>Beck et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>Northwest China</td>
<td>AVHRR NDVI</td>
<td>Ma and Veroustraete (2006)</td>
</tr>
<tr>
<td></td>
<td>West US</td>
<td>AVHRR NDVI</td>
<td>Bradley et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>Swiss Alps</td>
<td>AVHRR, SPOT &amp; MODIS NDVI</td>
<td>Fontana et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>Sichuan province, China</td>
<td>MODIS NDVI</td>
<td>Bian et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>MERIS MTCI</td>
<td>Atkinson et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Southeast Asia</td>
<td>MVC SPOT VGT-S</td>
<td>Chen et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>Africa</td>
<td>AVHRR NDVI</td>
<td>Jönsson and Eklundh (2004)</td>
</tr>
<tr>
<td></td>
<td>Swiss Alps</td>
<td>AVHRR, SPOT &amp; MODIS NDVI</td>
<td>Fontana et al. (2008)</td>
</tr>
<tr>
<td>Savitzky-Golay</td>
<td>West Central Alberta, Canada</td>
<td>MODIS NDVI</td>
<td>Hird and McDermid (2009)</td>
</tr>
<tr>
<td></td>
<td>Sichuan province, China</td>
<td>MODIS NDVI</td>
<td>Bian et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>North Queensland, Australia</td>
<td>Landsat, MODIS NDVI</td>
<td>Schmidt et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Murray-Darling basin, Australia</td>
<td>MODIS EVI</td>
<td>Broich et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>Eastern China</td>
<td>MODIS NDVI</td>
<td>Michishita et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>Australia</td>
<td>MODIS EVI</td>
<td>Broich et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Mongolia</td>
<td>MODIS NDVI</td>
<td>Eckert et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>Jiangxi Province, China</td>
<td>MODIS EVI</td>
<td>Li et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>MODIS EVI</td>
<td>Zhang et al. (2014)</td>
</tr>
<tr>
<td>Double logistic</td>
<td>North Scandanavia</td>
<td>MODIS NDVI</td>
<td>Beck et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>West Central Alberta, Canada</td>
<td>MODIS NDVI</td>
<td>Hird and McDermid (2009)</td>
</tr>
<tr>
<td></td>
<td>India</td>
<td>MERIS MTCI</td>
<td>Atkinson et al. (2012)</td>
</tr>
<tr>
<td></td>
<td>Eastern China</td>
<td>MODIS NDVI</td>
<td>Michishita et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>China</td>
<td>AVHRR NDVI, MERIS MTCI</td>
<td>He et al. (2015)</td>
</tr>
<tr>
<td>Asymmetric Gaussian</td>
<td>Africa</td>
<td>AVHRR NDVI</td>
<td>Jönsson and Eklundh (2002)</td>
</tr>
<tr>
<td>Wavelets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Best Index Slope Extraction (BISE)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonlinear running median</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean value iteration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean compositing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whittaker smoother</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanning smoothing (RMMEEH)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4253HT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMD3-ARMA5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Sensor/Instrument/Algorithm</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>AVHRR NDVI</td>
<td>Jönsson and Eklundh (2004)</td>
</tr>
<tr>
<td>North Scandanavia</td>
<td>MODIS NDVI</td>
<td>Beck et al. (2006)</td>
</tr>
<tr>
<td>West Central Alberta, Canada</td>
<td>MODIS NDVI</td>
<td>Hird and McDermid (2009)</td>
</tr>
<tr>
<td>South England</td>
<td>MERIS MTCI, MODIS NDVI, MODIS EVI</td>
<td>Boyd et al. (2011)</td>
</tr>
<tr>
<td>India</td>
<td>MERIS MTCI</td>
<td>Atkinson et al. (2012)</td>
</tr>
<tr>
<td>Eastern China</td>
<td>MODIS NDVI</td>
<td>Michishita et al. (2014)</td>
</tr>
<tr>
<td>US</td>
<td>AVHRR NDVI</td>
<td>Li and Kafatos (2000)</td>
</tr>
<tr>
<td>Spain</td>
<td>MODIS EVI</td>
<td>Sakamoto et al. (2005)</td>
</tr>
<tr>
<td>Jiangxi Province, China</td>
<td>MODIS EVI</td>
<td>Martínez and Gilabert (2009)</td>
</tr>
<tr>
<td>West Africa</td>
<td>AVHRR NDVI</td>
<td>Viovy et al. (1992)</td>
</tr>
<tr>
<td>Australia</td>
<td>AVHRR NDVI</td>
<td>Lovell and Graetz (2001)</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>MVC SPOT VGT-S</td>
<td>Chen et al. (2004)</td>
</tr>
<tr>
<td>Northwest China</td>
<td>AVHRR NDVI</td>
<td>Ma and Verostraete (2006)</td>
</tr>
<tr>
<td>Central West Africa, Southwest Europe</td>
<td>MERIS NDVI</td>
<td>Vancutsem et al. (2007)</td>
</tr>
<tr>
<td>United States</td>
<td>AVHRR NDVI</td>
<td>Reed et al. (1994)</td>
</tr>
<tr>
<td>Northwest China</td>
<td>AVHRR NDVI</td>
<td>Ma and Verostraete (2006)</td>
</tr>
<tr>
<td>West Central Alberta, Canada</td>
<td>MODIS NDVI</td>
<td>Hird and McDermid (2009)</td>
</tr>
<tr>
<td>Sichuan province, China</td>
<td>MODIS NDVI</td>
<td>Bian et al. (2010)</td>
</tr>
<tr>
<td>Eastern China</td>
<td>MODIS NDVI</td>
<td>Michishita et al. (2014)</td>
</tr>
<tr>
<td>Central West Africa, Southwest Europe</td>
<td>MERIS NDVI</td>
<td>Vancutsem et al. (2007)</td>
</tr>
<tr>
<td>South America</td>
<td>SPOT VGT NDVI</td>
<td>Atzberger and Eilers (2011a)</td>
</tr>
<tr>
<td>South America</td>
<td>SPOT VGT NDVI</td>
<td>Atzberger and Eilers (2011b)</td>
</tr>
<tr>
<td>India</td>
<td>MERIS MTCI</td>
<td>Atkinson et al. (2012)</td>
</tr>
<tr>
<td>Eastern China</td>
<td>MODIS NDVI</td>
<td>Michishita et al. (2014)</td>
</tr>
<tr>
<td>West Central Alberta, Canada</td>
<td>MODIS NDVI</td>
<td>Hird and McDermid (2009)</td>
</tr>
<tr>
<td>Eastern China</td>
<td>MODIS NDVI</td>
<td>Michishita et al. (2014)</td>
</tr>
<tr>
<td>West Central Alberta, Canada</td>
<td>MODIS NDVI</td>
<td>Hird and McDermid (2009)</td>
</tr>
</tbody>
</table>
## Chapter 7: The Spatio-Temporal Distribution of Eucalypt Flowering

<table>
<thead>
<tr>
<th>Method</th>
<th>Region</th>
<th>Data Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cubic spline</td>
<td>Eastern China</td>
<td>MODIS NDVI</td>
<td>Michishita et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>MODIS APAR</td>
<td>Marsden et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Brazil</td>
<td>MODIS NDVI</td>
<td>le Maire et al. (2011a)</td>
</tr>
<tr>
<td>Singular spectrum analysis</td>
<td>Northern Australia</td>
<td>MODIS EVI</td>
<td>Ma et al. (2013)</td>
</tr>
<tr>
<td>Not specified</td>
<td>Amazon rainforest, South America</td>
<td>MODIS EVI</td>
<td>Huete et al. (2006b)</td>
</tr>
<tr>
<td></td>
<td>Sahel and Soudan, Africa</td>
<td>AVHRR NDVI</td>
<td>Heumann et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>Sydney basin, Australia</td>
<td>MODIS NDVI, EVI, VARI and SR</td>
<td>Caccamo et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>Pakistan</td>
<td>MODIS NDVI</td>
<td>Ahmad (2013)</td>
</tr>
</tbody>
</table>
For the purposes of this thesis, and in the absence of any region specific prior studies, the Savitzky-Golay filter has been chosen to filter vegetation indices time series. Chen et al. (2004) used a Savitzky-Golay filter (Savitzky and Golay, 1964) to smooth a time series constructed from a 10-day NDVI composite and found that the fitted result eliminated displacements and spurious oscillations associated with Fourier-based fitting, and was considerably easier and quicker to apply than the BISE algorithm (Viovy et al., 1992). In contrast, Hird and McDermid (2009) found that the Savitzky-Golay filter was outperformed by the asymmetric Gaussian (Jönsson and Eklundh, 2002) and double-logistic (Beck et al., 2006) function-fitting techniques in a study of six different filtering methods. In a study conducted tracking grassland growth in the Swiss Alps, Fontana et al. (2008) established that the Fourier adjustment created better NDVI time series than the Savitzky-Golay filter. However, in a comparison of seven different noise reduction techniques in Eastern China, Michishita et al. (2014) found the Savitzky-Golay filter and Hanning smoothing (RMMEH) filter (Jin and Xu, 2013) to be the most appropriate techniques for their study area.

The Savitzky-Golay filter for a VI time series is defined as:

$$Y_i^* = \frac{\sum_{i=-m}^{i=m} C_i Y_{j+i}}{N}$$

where $Y$ is the original VI value, $Y^*$ the resultant VI value, $C_i$ is the coefficient for the $i$th VI value of the filter, $j$ is the running index of the original ordinate data table and $N$ is the number of convoluting integers and is equal to the smoothing window $(2m+1)$ (Chen et al., 2004). The
half-width of the smoothing window \((m)\) and the degree of the smoothing polynomial \((d)\) are chosen to provide an optimum result; larger values of \(m\) provide a smoother series but flatten sharp peaks and smaller values of \(d\) produce a smoother series but may introduce bias.

The \(m\) and \(d\) values used are typically determined iteratively and empirically from study to study. Not all studies using the Savitzky-Golay filter report the values of \(m\) and \(d\) used making replication of their results difficult. Values reported from previous studies include \((m, \ d) = (4, 6)\) (Chen et al., 2004), \((m, \ d) = (3, 2)\) (Jönsson and Eklundh, 2004), \(m=5,6,7\) (Hird and McDermid, 2009), \(m=2\) (Schmidt et al., 2012), \((m, \ d) = (3,3)\) (Michishita et al., 2014).

Only a single study into the phenology of nectar production has previously been conducted. Nightingale et al. (2008) attempted to directly establish whether satellites can track the honey bee nectar flow in the US by correlating vegetation phenology trends derived from time series of AVHRR NDVI data with 15-years of hive weight measurements and found a strong correspondence.

7.1.8 Remote sensing of eucalypts and eucalypt phenology

As outlined in Section 7.1.4, eucalypt leaves and canopies have properties not shared by many other broadleaf species and this raises the question of whether vegetation indices developed for other species are applicable for eucalypts. In a study across twenty \textit{Eucalyptus} species and one closely related \textit{Angophora} species, Datt (1998) found poor correlation between the NDVI and chlorophyll pigment content when using a red
bandwidth of 680 nm and a NIR bandwidth of 800 nm. In contrast, Cunningham et al. (2007) found that NDVI was the variable that contributed most to a model of plant area index for river red gum (*E. camaldulensis*) stand health. Using Landsat 7 data, Cunningham et al. (2007) determined that the model of vigour was contributed to most by reflectance in the near infrared (760-900 nm) and the far infrared (2080-2350 nm), and re-affirmed these findings in a follow-up study (Cunningham et al., 2010). Datt (1999a, 1999b) used the visible/NIR reflectance properties leaves from twenty eucalypt species to develop new spectral indices for the remote sensing of chlorophyll content in eucalypts. The two best performing indices were:

$$\text{Datt99} = \frac{\rho_{850} - \rho_{710}}{\rho_{850} - \rho_{680}}$$  \text{and}

$$\text{Datt derivative (DD)} = \frac{D(\rho_{754})}{D(\rho_{704})},$$  

where $D$ is the first derivative of the smoothed reflectance spectrum at the specified bandwidths.

The indices proposed by Datt (1999a, 1999b) and developed using eucalypts have also been shown to perform well for other broad leaf and needle leaf species. Le Maire et al. (2004) found a modified version of Datt99, replacing $\rho_{850}$ with $\rho_{780}$ after Maccioni et al. (2001) the best performing leaf index in a sample of 53 northern hemisphere broad leaf species. In a study conducted in Canadian forests, Croft et al. (2014) found the DD index to be the best performing index in combined broadleaf and needle-leaf forests using data acquired from the MERIS...
sensor, with MTCI second best for combined forests and for broadleaf forests. Datt99 performed well at leaf level, but less so at canopy level. Neither EVI nor NDVI at the bandwidths used by MODIS were tested as part of the comparison. The findings of Croft et al. (2014) are echoed by Coops et al. (2003) who found that in a study of a forest featuring dominated by *E. pilularis*, *E. saligna* and *E. paniculata* in southern NSW there was poor correlation at the canopy scale between Datt99 and leaf relative chlorophyll content. The DD index was not assessed as part of this study.

Much of the prior work done on remote sensing of eucalypt phenology has been focussed on eucalypt plantations in Brazil (Marsden et al., 2010, le Maire et al., 2011a, le Maire et al., 2011b, le Maire et al., 2012, le Maire et al., 2014). This is presumably due to the prevalence of eucalypt plantations in Brazil; compared to Australia's 1 million hectares of hardwood eucalypt plantation (Montreal Process Implementation Group for Australia and National Forest Inventory Steering Committee, 2013), Brazil has five times the planted area; five million hectares (le Maire et al., 2014).

Studies of eucalypt phenology in natural forest environments in Australia are few in number. Hill et al. (2006) showed that MODIS LAI majorly overestimated field measurements of LAI in some eastern Australian open forests and woodlands. Similarly, Leuning et al. (2005) found that the MODIS LAI algorithm overestimated LAI by a factor of two in a southeast Australian eucalypt forest. le Maire et al. (2012) attempted to overcome the LAI estimation limitation by developing a species-specific
eucalypt LAI index called EucVI which they tested successfully in eucalypt plantations in Brazil using MODIS time-series data. Caccamo et al. (2011) investigated the sensitivity of MODIS data to monitor drought in the Sydney basin bioregion by comparing time series constructed from vegetation greenness indices (NDVI, EVI, VARI and SR) with vegetation water indices. They found that all indices showed significant correlation with the standardised precipitation index (SPI) (McKee et al., 1993) but that the vegetation water indices using bands in the SWIR performed better in high biomass ecosystems. Schmidt et al. (2012) fused high spatial resolution Landsat TM/ETM+ data with high temporal resolution MODIS NDVI data in a study of the phenology of a Queensland savannah populated with eucalypts.

Relatively few high spectral resolution remote sensing studies of eucalypt forests and plantations have been undertaken. Exceptions include studies using HYPERION data (Coops et al., 2001), the airborne CASI-2 spectrograph (Coops et al., 2004, Goodwin et al., 2005), and the airborne HyMap sensor (Youngentob et al., 2011).

Studies of other species have confirmed that flowering can impact the accuracy of biomass estimates from NDVI and EVI (Miaogen et al., 2010), and it is possible that the intensity of eucalypt flowering would also impact these estimates. There have not been any studies focussed on remote sensing (either from satellite, aircraft or ground based) of eucalypt budding or flowering directly.
A final characteristic of eucalypts suitable for study via land surface phenology is the changing concentration of anthocyanin with time. The presence of foliar anthocyanin has been found to impact the performance of vegetation indices used to estimate chlorophyll content (Viña and Gitelson, 2011). Gitelson et al. (2001) has proposed a two-band index; the Anthocyanin Reflectance Index (ARI) to detect anthocyanin concentrations. Anthocyanin content in eucalypts can vary due to stresses on the plant (Close, 2001, Stone et al., 2001, Close and Beadle, 2003, Barry et al., 2008, Barry et al., 2009a). Barry et al. (2008) observed correlation between defoliation in *E. globulus* and the ARI, and between low nutrients and low temperatures in *E. pilularis* and the ARI. Barry et al. (2009a) similarly found a strong correlation between potassium deficiency in the leaves of *E. globulus* and the ARI. No analysis has yet been performed for correlation between eucalypt stressors and the three-band anthocyanin index developed by Gitelson et al. (2006). Anthocyanin content also increases during the "spring flush" (Barry et al., 2008, Barry et al., 2009a). This is an adaptation to suppress photosensitivity in the delicate new foliage (Close and Beadle, 2003). Barry et al. (2009a) notes the challenges in distinguishing between increased concentrations due to stress compared with spring growth, and recommends the use of a time series of a range of vegetation indices to separate stress and phenology in time and space.

### 7.1.9 Summary

While the eucalypt is the dominant tree species in most natural environments in Australia, relatively little quantitative data exists on
long-term patterns of eucalypt flowering. There are multiple platforms and vegetation indices, and lengthy data records available that could be used to analyse eucalypt phenology. Specialised vegetation indices for eucalypts have also been developed but not yet applied in phenology studies. There is significant potential to provide remotely-sensed growth information to beekeepers as an additional tool for seasonal planning.

7.2 Visualising patterns of eucalypt growth

7.2.1 Introduction

This chapter highlights the relationship between remotely-sensed observations of eucalypt growth, flowering and honey production. A prototype web-based application named BeeBox has been developed for beekeepers to visualise these same patterns of growth, and is discussed in a journal paper currently under peer review by an international journal:


A further demonstration of the use of BeeBox to examine patterns of flowering across similarly forested sites is also provided.

7.2.2 A web-based application for beekeepers to visualise patterns of growth in floral resources using MODIS data

7.2.2.1 Abstract

The honey bee industry is of immense importance to global agriculture. In many countries beekeepers are migratory and move their hives between flowering events. Predicting such flowering events is
particularly difficult in Australia due to the irregular flowering of eucalypts. We have developed a web-based application for Victorian beekeepers to visualise patterns of growth in floral resources using MODIS and other data, and thus make remote predictions about whether flowering will occur at their apiary sites. We demonstrate the use of this application through comparing ironbark (*Eucalyptus tricarpa*) growth patterns with flowering and honey production records. While the scientific community as a whole has embraced the use of satellite imagery as a tool for phenological studies, our prototype represents the first attempt to make this same information available to a more general audience.

### 7.2.2.2 Software availability

<table>
<thead>
<tr>
<th>Software name:</th>
<th>BeeBox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developers:</td>
<td>Guan Gui, Jonathan Arundel, Xi Liang</td>
</tr>
<tr>
<td>Year first official release:</td>
<td>2012</td>
</tr>
<tr>
<td>Hardware required:</td>
<td>PC, tablet, mobile</td>
</tr>
<tr>
<td>Software required:</td>
<td>Web browser. Chrome, Firefox, Safari and Internet Explorer v11 supported.</td>
</tr>
<tr>
<td>Program language:</td>
<td>JavaScript, HTML5, Java, Perl, IDL, R</td>
</tr>
<tr>
<td>License:</td>
<td>Free for non-commercial use</td>
</tr>
</tbody>
</table>

### 7.2.2.3 Introduction

Pollinator decline poses a significant risk to agriculture and thus requires a strong focus on improving management of the industry, and further exploiting the value of geospatial information (Rogers and Staub, 2013). Any tool which therefore empowers beekeepers to make improved
remote predictions about flowering will significantly reduce cost and risk to beekeeping operations.

To visualise patterns of growth in floral resources and hence allow beekeepers to make better decisions in planning their seasonal operations we have developed a prototype application called "BeeBox". BeeBox is a web-based application, providing an interactive and user-friendly client interface using modern JavaScript libraries (Walker and Chapra, 2014, Vitolo et al., 2015) and free and open source software components (Swain et al., 2015). We have used freely available geo-referenced data, in particular MODIS remote sensing imagery (NASA Land Processes Distributed Active Archive Center (LP DAAC), 2015), to create a prototype for beekeepers – a general audience not trained in the use of geographic information systems (GIS) – in the state of Victoria, Australia.

The central concept behind this tool is that beekeepers use the application to remotely view patterns of growth in areas where they hold apiary sites, and hence make predictions of whether flowering is likely to occur at each site.

The seasonal pattern of variation in vegetated land surfaces observed from remote sensing is generally known as land surface phenology (Friedl et al., 2006), and in its 30-year history (Boyd et al., 2011) few attempts have been made to communicate findings to a non-scientific audience. A comprehensive review of current software available for land surface phenology is provided by Eerens et al. (2014); in general such applications require users to possess specialised knowledge. In Australia, initiatives such as the Atlas of Living Australia (ALA) (Atlas of Living
Chapter 7: The Spatio-Temporal Distribution of Eucalypt Flowering

Australia, 2015) and Terrestrial Ecosystem Research Network (TERN) (Terrestrial Ecosystem Research Network, 2015) aggregate environmental data from multiple sources and present this online to a broader audience. Although TERN has recently added a phenology product based on MODIS data for the entire Australian continent for the period from 2000-2012 (Broich et al., 2015), accessing and using this data still requires specialist software and knowledge. In contrast, BeeBox allows beekeepers to produce phenological time series interactively for any region of interest using only a web-browser, and on any platform (PC, tablet, mobile).

In this chapter we describe the architecture and data sets used by BeeBox, and provide a demonstration of functionality by comparing BeeBox-generated vegetation index time series with historical flowering and honey production records.

7.2.2.4 Data and functionality

Beekeepers make predictions of flowering for a eucalypt species at a particular apiary site based on:

(1) knowledge of eucalypt species within foraging range of the site
(2) observations of growth within individual eucalypt species
(3) flowering history for individual species at the site, and
(4) environmental conditions (e.g., rainfall, temperature)

In attempting to systemise this knowledge for each site possessed by beekeepers, we must first start with the sites themselves.
7.2.2.4.1 Apiary site locations
Australian beekeepers will typically hold a mixture of sites on both public and private land of which only a relatively small proportion of sites will be in use at any one time. Data on public land apiary sites is available, but to avoid competition from other beekeepers for the same floral resources individual beekeepers tend not to divulge the use of these sites or the location of their sites on private land. Beekeepers make use of GPS for recording the location of sites and online navigation tools such as Google Maps for planning their driving routes to these sites. For this reason we have developed the application to allow beekeepers to view Google Maps data as a background layer and upload their site locations or areas of interests using a KML file format via the region analysis wizard.

7.2.2.4.2 Eucalypt species
Most of the native forests used by beekeepers in Australia are comprised of multiple eucalypt species as well as understory plants and grasses. Honey bee colonies routinely forage several kilometres from their hives, and thus a single apiary site will harvest from an area in the order of 100 km² (Visscher and Seeley, 1982). Across this foraging range there are many different floristic communities (Environment Conservation Council, 2001). The distribution maps for eucalypt species vary widely between sources (Brooker and Kleinig, 2006, Costermans, 2006, Nicolle, 2006) so for the purposes of our prototype we have instead opted for herbaria records from the Atlas of Living Australia (ALA) (2015). We separately

---

display the modelled extent of native vegetation (Department of Environment and Primary Industries, 2015a).

7.2.2.4.3 Observations of growth
While beekeepers directly observe eucalypt growth, budding and health as predictors of flowering (Birtchnell and Gibson, 2006), for the BeeBox application we make use of MODIS vegetation indices (Solano et al., 2010) as a proxy for these observations at landscape scale. MODIS sensor data provides good temporal resolution for land surface phenology with a revisit time 1-2 days (Jones and Vaughan, 2010). MODIS provides better spectral and spatial resolution than SPOT-VGT (Kramer, 2015), and has the advantage over MERIS (Rast et al., 1999) of a continuous data record since 2000 (ESA, 2015). Land surface phenology studies frequently utilise MODIS vegetation index data; either MODIS NDVI (Beck et al., 2006, Fontana et al., 2008, Hird and McDermid, 2009, Bian et al., 2010, Marsden et al., 2010, Boyd et al., 2011, Caccamo et al., 2011, le Maire et al., 2011a, Schmidt et al., 2012, Ahmad, 2013, Jin and Xu, 2013, Michishita et al., 2014, Eckert et al., 2015) and/or MODIS EVI (Zhang et al., 2003, Sakamoto et al., 2005, Huete et al., 2006a, Huete et al., 2006b, Zhang et al., 2006, Boyd et al., 2011, Caccamo et al., 2011, Ma et al., 2013, Broich et al., 2014, Li et al., 2014, Zhang et al., 2014, Broich et al., 2015). MODIS data has also been used in the study of natural environments and eucalypt forests in Australia (Leuning et al., 2005, Hill et al., 2006, Cleugh et al., 2007, Caccamo et al., 2011, Caccamo, 2012, Schmidt et al., 2012, Ma et al., 2013, Pickett-Heaps et al., 2014), and in eucalypt plantations globally.
The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia (Lopes et al., 2009, Marsden et al., 2010, le Maire et al., 2011a, le Maire et al., 2012, le Maire et al., 2014).

To serve the needs of migratory beekeepers based in Victoria, the coverage of BeeBox extends into New South Wales, Queensland and South Australia. We are using five MODIS MOD13Q1 tiles (Huete et al., 1999) stitched together to provide this coverage; full coverage of the Australian subcontinent would require a further eleven tiles. This dataset provides a 16-day composite of two vegetation indices with a pixel resolution of 250 m; the Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), both of which are viewable using BeeBox. The advantages of using a Level 3 composited data set rather than a Level 2 daily data set such as MOD09 (Vermote et al., 2011) are twofold; composited images are less effected by cloud cover or atmospheric contamination and the reduction in storage and processing power needed is significant. At the time of writing, raw and processed data for the five tiles used by BeeBox consumes ~1TB of storage.

The impact of choosing to use composited images is most noticeable in the on-demand generation of EVI time series across all images collected since February 2000 within the BeeBox region analysis wizard. This wizard enables beekeepers to define a region of interest and produce a time series of averaged EVI values across the region. The region of interest can be either directly drawn on the map, or a pre-existing KML
file produced using software such as Google Earth\textsuperscript{21} can be uploaded. While it is straightforward to provide time series for both indices, our prototype deliberately features only a single index to simplify the user experience. We chose EVI in preference to NDVI based on its improved sensitivity in high biomass areas, decoupling of canopy background signals and resistance to atmospheric influences (Huete \textit{et al.}, 2002). The time series is displayed after excluding low quality pixels from time series, and using a third-order Savitzky-Golay filter (Savitzky and Golay, 1964) to reduce noise (Chen \textit{et al.}, 2004).

\subsection*{7.2.2.4.4 Flowering and honey records}
Within the last decade, there are no data collected and therefore able to be displayed about the flowering of eucalypts at a landscape scale. The large scale and long term surveys of flowering in the box-ironbark forests conducted by the former Forest Commission of Victoria from 1930 until 1981 (Keatley, 1999) predate the time period covered by AVHRR and MODIS data. The best available data are the surveys conducted by MacNally \textit{et al.} (2009). These are analysed against BeeBox data in this paper, but not made generally available through the BeeBox application. At an aggregate level, honey production yields from individual species are recorded by the honey packers who buy, pack and distribute the honey from individual commercial beekeepers. While honey packers do not commonly share this data for reasons of commercial confidentiality, the

largest honey packer in Australia has shared production data for analysis in this paper.

7.2.2.4.5 Environmental conditions
Beekeepers in Australia commonly make use of climate and forecast data provided by the Australian Bureau of Meteorology through their website (Bureau of Meteorology, 2015a). For the purposes of our prototype, we have not sought to make this same information available through BeeBox.

7.2.2.5 Architecture and implementation
The architecture for the BeeBox application was developed with the following objectives:

(1) Provide a rich, interactive user experience across multiple device types – to aid adoption we wanted to provide a rich user experience across the maximum number of device types. By using modern JavaScript frameworks we ensured maximum exposure for the application while providing a consistent user experience across platforms. BeeBox works with all modern browsers on desktop, laptops, smartphones and tablets. An added benefit of using JavaScript frameworks is the reduced development time and costs.

(2) Minimise costs through adoption of open source software – with a limited budget for development, we sought to minimise costs by making heavy use of open source software and frameworks. With the exception of pre-existing IDL code from the lead author re-used in the development of the region analysis wizard, no commercial software has been used in the
construction of the application. It would be possible to use open source alternatives to replicate the functionality of the IDL code if desired.

The overall architecture for BeeBox is shown in Figure 26.
The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia

Figure 26: BeeBox application architecture

---

1 Atlas of Living Australia
2 Victorian Department of Sustainability and Environment
3 United States Geological Survey

---

262 Jonathan Arundel - November 2015
External data is made available either directly in the presentation layer (Google Maps), via WMS (herbaria records), or pre-processed (native vegetation and MODIS data). Every 16 days the latest MODIS data is automatically downloaded and pre-processed. The BeeBox controller (written in Java) serves as the bridge between data and presentation layers. The controller helps load map data into Geoserver\(^{22}\), maintains active connections to client browsers, and synchronises and forwards requests and responses between clients and server. Both the BeeBox controller and Geoserver run as servlets in Tomcat\(^{23}\). The presentation layer leverages OpenLayers\(^{24}\), jQuery\(^{25}\), ExtJS\(^{26}\), GeoExt\(^{27}\) and Raphael\(^{28}\) to provide a dynamic interactive user interface which works across browsers on different client devices. jQuery (in conjunction with the BeeBox controller) is also used to maintain real-time server communication such that new data can be actively pushed clients without a need for a client browser refresh. On-demand processing is required for the region analysis wizard to generate and display the filtered EVI time series for a specified region, and for the compare dates module to perform raster calculations for pairs of images.

\(^{22}\) http://geoserver.org/, accessed 5\(^{th}\) June 2015
\(^{23}\) http://tomcat.apache.org/, accessed 5\(^{th}\) June 2015
\(^{24}\) http://openlayers.org/, accessed 5\(^{th}\) June 2015
\(^{25}\) https://jquery.com/, accessed 5\(^{th}\) June 2015
\(^{27}\) http://geoext.org/, accessed 5\(^{th}\) June 2015
\(^{28}\) http://raphaeljs.com/, accessed 5\(^{th}\) June 2015
The source code for BeeBox is available on request, and from late-2015 the application will be hosted by TERN, the Terrestrial Ecosystem Research Network (Terrestrial Ecosystem Research Network, 2015).

7.2.2.6 Case study and discussion

At its most basic level, BeeBox provides the functionality for beekeepers to interact with the various map layers and navigate backwards and forwards in time through consecutive 16-day composite colourised vegetation index images. As an aid for visualising the differences between any two images (e.g., consecutive images or year-to-year comparisons) BeeBox provides the functionality for beekeepers to generate an on-demand raster subtraction where positive values (growth) are colourised in red and negative values (generally senescence, but may also include harvesting/logging or fire) are colourised in blue. A sample comparison for the 16-day period spanning the Victorian Black Saturday bushfires on February 7, 2009 is shown in Figure 27. The burn scar can be seen in blue, with a small area of faint growth east of the burnt area appearing pink.
Figure 27: BeeBox visualisation comparing MODIS 16-day images before and after the Victorian Black Saturday bushfires
As a demonstration of how BeeBox can be used to predict honey production, we have focussed on production of ironbark \((Eucalyptus tricarpa, \text{and to a lesser extent}, Eucalyptus sideroxylon)\) honey in Victoria. \(Eucalyptus tricarpa\) is distributed through the box-ironbark woodlands of Central Victoria, and across East Gippsland through into southern coastal New South Wales (Brooker and Kleinig, 2006, Nicolle, 2006). \(Eucalyptus sideroxylon\) occurs on the western slopes and plains of New South Wales into south-eastern Queensland, with a small extension into Victoria near the New South Wales border (Brooker and Kleinig, 2006, Nicolle, 2006).

The BeeBox representation of \(E.\ tricarpa\) and \(E.\ sideroxylon\) herbaria records is shown in Figure 28. Victorian honey producers make use of \(E.\ tricarpa\) for honey production in preference to \(E.\ sideroxylon\) (Birtchnell and Gibson, 2006), and the flowering phenology of \(E.\ tricarpa\) has been extensively studied (Porter, 1978, Keatley and Hudson, 2007, Mac Nally \textit{et al.}, 2009, Hudson \textit{et al.}, 2010b, Hudson and Keatley, 2013, Rawal \textit{et al.}, 2014). The relatively limited and well-defined spatial distribution of \(E.\ tricarpa\), coupled with an available flowering record which overlaps with the period of MODIS data availability (Mac Nally \textit{et al.}, 2009) make it the best choice of species for demonstrating the use of BeeBox.
Figure 28: *Eucalyptus tricarpa* (green dots) and *Eucalyptus sideroxylon* (grey dots) herbaria records
The pattern of flowering for *E. tricarpa* across 12 remnant forest sites in central and northern Victoria was analysed by Mac Nally *et al.* (2009) for the period from 1997-2008. In the period overlapping with MODIS data availability, they observed strong (>60% of sample points) flowering in mid-2001, mid-2004 and mid-2006 with some (10-30% of sample point) flowering in mid-2000 and mid-2003 (Figure 29b). During this same period, ironbark honey production was strongest in the years of 2000, 2004 and 2006 with some production in 2001 (Figure 29c).
Figure 29: Comparison of EVI growth time series (a) with flowering data (b) and honey production (c) for *Eucalyptus tricarpa*.
To compare the flowering and production data with the land surface phenology for these regions, we make use of the BeeBox region analysis wizard. Taking the five remnant forest sites with the highest amount of tree cover (Radford et al., 2005) from the study of Mac Nally et al. (2009), we drew polygons for the forested areas in Google Earth which we then saved as a single KML file. We used the region analysis wizard to upload this KML file, shown in Figure 30, and produced a smoothed EVI time-series for this forested area, shown in Figure 29a. One of the key benefits in defining a polygonal region of interest is that it provides less-noisy time series than single-pixel analysis (Broich et al., 2015), without the limitations of rectangular-shaped areas (Bradley et al., 2010).
Figure 30: BeeBox region analysis wizard showing upload of Google Earth KML file against NDVI background
The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia

The EVI time-series shows strong growth coinciding with strong flowering and honey production in the years of 2000, 2004 and 2006. It also shows weak growth coinciding with weak flowering in the years of 2002 and 2003. There is strong growth and flowering in 2001, but only a relatively small amount of honey produced. In 2005 and 2007, there is again strong growth but no significant flowering or honey production recorded. A possible explanation for why the observed strong growth did not lead to flowering (and therefore honey production) in these years is that *E. tricarpa* is known to have a short interval between flowering of two years (Birtchnell and Gibson, 2006). In consecutive years of strong growth, flowering and honey production is only likely to occur every second year. In our data, the three year interval between flowering events in 2001 and 2004 happens to span the period of lowest growth, suggesting that unless growth exceeds a certain threshold, flowering will not occur. It is important to highlight that from 2001 until 2009 south-eastern Australia recorded the longest uninterrupted series of years with below median rainfall since at least 1900 in a period known as the "Millennium Drought" (van Dijk et al, 2013), and it is against this background that we need to interpret any trends in MODIS data.

Individual beekeepers can use their knowledge of site history, eucalyptus flowering intervals and BeeBox data in a similar manner to make predictions about flowering at each one of their individual sites. The linkage between growth and flowering is demonstrably clearer than the linkage between rainfall, temperature and flowering and thus tools like
BeeBox can provide a valuable additional resource to the migratory beekeeping industry.

7.2.2.7 Conclusions

While the scientific community as a whole has embraced the use of satellite imagery as a tool for phenological studies, BeeBox represents the first attempt to make this same information available to a more general audience; in this case beekeepers. We have removed the need for specialised knowledge, software and hardware to access and process the data by developing an intuitive web-based system for data visualisation.

We also see significant potential beyond the honey bee industry for the use of BeeBox in other primary industries and natural resource management. The only barrier to the expansion of geographic coverage is the availability of compute, storage and network resources and we believe that providing national or even global coverage would be possible without significant changes to the current code base or user interface.

Challenges remain in systemising the specialised knowledge accumulated by beekeepers over decades and generations. We have demonstrated the agreement of growth with honey production for *E. tricarpa*, and to do this we used flowering data and honey production data not generally available for all species. While individual beekeepers can synthesise information from BeeBox with their own local knowledge to predict flowering, we do not foresee any immediate progress being able to do this at landscape scale. Under current industry policies, there is no driver for site-specific production data collection and reporting, and this limits the scientific
community’s ability to build predictive models of flowering. Such a model would have immense value to understanding the eucalypt forest ecosystems in Australia more generally and the impacts of climate change. Nevertheless, we believe that BeeBox still represents a significant advancement of environmental software in the public domain.

7.2.2.8 Acknowledgements

This work was supported by the Rural Industries Research and Development Corporation (RIRDC) under Project Number 004872. Joseph Leach, Xi Liang and Edward Webber provided valuable contributions to the project. Finally, we acknowledge the cooperation of Capilano Honey who provided the production data used in the analysis.

7.3 Growth, temperature, rainfall and their relationship across sites

7.3.1 Introduction

The relationship of temperature and rainfall with eucalypt growth and flowering has been explored in many studies (see Section 7.1.8 for a summary), but few studies have made use of remote sensing data as the basis for this analysis. Section 7.2.2 provided an example demonstrating correspondence between eucalypt growth (as measured by remote sensing data), flowering and honey production. This section provides a further example of the use of BeeBox to analyse time series of vegetation indices data across several similarly forested sites in Victoria. The hypothesis is that for forest blocks characterised by the same mix of species, those closer together (and therefore experiencing more similar
rainfall and temperature patterns) will exhibit more closely matched patterns of growth.

The choice of forest type for this analysis has been determined by selecting an environment for which there is good baseline knowledge of growth and flowering patterns. Two of the most studied areas of eucalyptus growth and flowering in Victoria are the Rushworth State Forest and Havelock Nature Conservation Reserve (NCR); previous studies are summarised in Table 22. These three areas are located within the Goldfields bioregion (Department of Environment and Primary Industries, 2015c). The forest blocks primarily comprise box-ironbark forest. This type of forest is designated by the Ecological Vegetation Class (EVC) 61. Department of Sustainability and Environment (2004, p.21) defines box ironbark forest under EVC 61 as:

*Occurs in low rainfall areas on gently undulating rises, low hills and peneplains on infertile, often stony soils derived from a range of geologies. The open overstorey to 20 m tall consists of a variety of eucalypts, often including one of the Ironbark species. The mid storey often forms a dense to open small tree or shrub layer over an open ground layer ranging from a sparse to well-developed suite of herbs and grasses.*

Large trees occur with a density of approximately 15/ha, with tree canopy cover approximating 30%. The dominant species are all eucalypts; *E. microcarpa, E. tricarpa, E. polyanthemos,* and *E. leucoxylon* (Department of Sustainability and Environment, 2004).
Three additional forested areas were also chosen for analysis. Box ironbark forest in Puckapunyal was included, as this area had previously been surveyed for feral honey bee colonies; see Chapter 5 and Arundel et al. (2014) for details. Tunstalls NCR was also chosen as it was one of the sites surveyed by Radford et al. (2005). Both Tunstalls NCR and Puckapunyal are within the Goldfields bioregion. Finally, in the north-east of Victoria lies the Warby-Ovens National Park, within the Northern Inland Slopes bioregion. A large forested block of box-ironbark was chosen for analysis from within the National Park. EVC61 within the Northern Inland Slopes bioregion is defined similarly to within the Goldfields bioregion, but there are some differences in the dominant eucalypt species; *E. sideroxylon*, *E. macrorhyncha*, *E. polyanthemos*, and *E. microcarpa* (Department of Sustainability and Environment, 2007a). The five sites chosen in total are illustrated in Figure 37.

### 7.3.2 Method

The first step in the analysis was to select homogeneous areas of forest for analysis at each of the sites. For each of the selected sites, areas identified as EVC61 vegetation were extracted from the NV2005 EVC data set (Department of Environment and Primary Industries, 2014b). The resolution of MODIS EVI satellite data is 250 m. A negative 250 m buffer was therefore applied to each area to ensure that there would be no edge effects in the resulting time series. The buffered areas were intersected with the forest or park boundary polygon/s derived from Department of Environment (2015). For each site, the resulting polygon was then exported in a KML file format ready for analysis in BeeBox.
BeeBox was used to produce an EVI time series for each of the sites. The KML file for each site was uploaded within BeeBox’s region analysis wizard and a graphical plot of the EVI time series was produced. The raw (unfiltered) data used to produce the EVI time series graph was downloaded from BeeBox. Filtering of the time series was performed using a third-order Savitzky-Golay filter (Savitzky and Golay, 1964). Comparisons were made between sites using both the raw and filtered time series data.

Climate data for each site was also obtained from the Bureau of Meteorology (2015b). The location of the nearest Bureau of Meteorology weather station for rainfall and temperature records was manually determined for each of the five forested sites. The weather station sites are shown in Figure 37.

The zoo (Zeileis and Grothendieck, 2005), ggplot2 (Wickham, 2009), scales (Wickham, 2014), and hydroTSM (Zambrano-Bigiarini, 2014) packages within R (R Core Team, 2014) were used to generate plots comparing sites and variables.

7.3.3 Results

The EVI time series for each site is plotted against the corresponding seasonal rainfall (with spring rainfall highlighted in green) and monthly maximum daily average temperatures in Figure 31 - Figure 35.
Figure 31: Comparison of EVI time series data with seasonal rainfall and monthly maximum daily temperature averages for Tunstalls Nature Conservation Reserve
Figure 32: Comparison of EVI time series data with seasonal rainfall and monthly maximum daily temperature averages for Havelock Nature Conservation Reserve
Figure 33: Comparison of EVI time series data with seasonal rainfall and monthly maximum daily temperature averages for Puckapunyal Army Reserve
Figure 34: Comparison of EVI time series data with seasonal rainfall and monthly maximum daily temperature averages for Rushworth State Forest
The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia

Figure 35: Comparison of EVI time series data with seasonal rainfall and monthly maximum daily temperature averages for Warby Ovens National Park
A comparison of the raw and filtered EVI time series data for each of the five sites is shown in Figure 36. A visual analysis indicates close correspondence between the time series for the olive (Havelock NCR) and red (Tunstalls NCR) data, and an even stronger match between the pink (Warby-Ovens National Park) and blue (Rushworth State Forest) data. Figure 37 shows the sites exhibiting similar patterns of growth linked with green lines.
Figure 36: Raw and filtered EVI time series for box ironbark (EVC61) sites in Victoria
The distance between the Tunstalls NCR and Havelock NCR sites is 37 km.

The distance between the Rushworth State Forest and Warby-Ovens National Park sites is 105 km.
Figure 37: Selected box ironbark (EVC61) sites (yellow polygons/green labels), temperature data sites (red labels), and rainfall data sites (blue labels) across Victoria. Sites with matching EVI time series are shown linked with a green line.
As the closest match was observed between the Rushworth State Forest and Warby-Ovens National Park sites, a further set of graphs were generated to overlay EVI, temperature and rainfall data for the two sites on a single plot (Figure 38). As it is difficult to see differences in the timing of rainfall using only seasonal averages, a 90-day rolling daily average was used instead.
Figure 38: Comparison of EVI time series data, 90-day rolling daily average rainfall and monthly maximum daily temperature averages for the Rushworth State Forest and Warby-Ovens National Park sites.
7.3.4 Discussion

There is evidence against the hypothesis that closer sites of similar forest exhibit more closely matched patterns of growth. The Tunstalls NCR and Havelock NCR sites exhibit similar patterns of growth, and in this case are also close to one another (37 km). However, the patterns of growth are even more closely matched between the Rushworth State Forest and Warby-Ovens National Parks site, despite these sites being separated by a distance of over 100 km. The sites reside in different bioregions, and the dominant mix of species varies slightly between sites. In addition, the Rushworth State Forest site is logged (Environment Conservation Council, 2001) whereas the Warby-Ovens site, being National Park, is not. The closest site to Rushworth is Puckapunyal (35 km), yet its pattern of growth is markedly different to that of Rushworth’s.

In the case of the Rushworth and Warby-Ovens sites, the pattern of growth is similar despite differences in the temperature and rainfall received. As Warby-Ovens National Park is further north, it is generally hotter during summer; this can be seen in the average daily temperature maximums (Figure 38) where temperatures in February are typically 3-5% higher. While both sites over time tend to receive a similar rainfall, the 90-day rolling daily average shows significant variation in the timing of when rainfall is received. For example, in early 2005 the Warby-Ovens sees a spike in rainfall due to a significant downpour that was clearly not received at Rushworth. The reverse of this situation occurs late 2010. However, there is not any obvious visual link between variation in the timing and amount of rainfall and variation in the timing and magnitude
of vegetation growth. While the most anomalous period in terms of rainfall, 2010-2011, is followed by seasons of high growth, even higher growth occurs in 2004 without the rainfall event.

While the similarity of the pattern observed at the Rushworth and Warby-Ovens sites is remarkable, the dissimilarity between the Rushworth and Puckapunyal sites can potentially be explained by an artefact of the methodology. As EVCs represent a modelled construct, there may be variation in the classification and accuracy of vegetation types between two areas identified as the same EVC. As the canopy of EVC61 is relatively open (30% coverage), trends in growth could be heavily influenced by changes in the understorey rather as well as changes in the canopy. In the case of Puckapunyal, erosion problems prompted the deliberate introduction of an exotic grass, *Phalaris aquatica* which has become successfully naturalised. However, without ground truthing the species present at each of the selected sites, it is difficult to make any strong conclusions on to what extent differences might be explainable based on differences in vegetation.

7.3.5 Conclusions

Patterns of similar eucalypt forest show patterns of similar growth, and to some extent this phenomenon seems independent of distance. The relationship between patterns of eucalypt forest growth and the variables typically thought to influence growth; rainfall and temperature, is cryptic. The close correspondence of growth across sites more than 100 km apart indicates the operation of similar biological and ecological processes.
As this study has focussed on only one of the ~600 EVCs in Victoria, it would be valuable to extend the study to other species and across broader landscapes. There are challenges with constraining the analysis to modelled EVCs these may not identify similarly vegetated areas with the degree of accuracy needed for land phenology studies.

An alternative approach would be to use an analytical technique that reduces the dimensions of the data without assuming a priori knowledge of vegetation characteristics. Based on the size of the dataset (> 1TB) this would be computationally demanding and require the use of distributed computing and distributed algorithms. Techniques and frameworks have recently been developed to perform distributed principal component analysis (PCA) of zebrafish neuron activity time series (Freeman et al, 2014); similar techniques could be applied to identify other areas of vegetation separated by large distances exhibiting similar patterns of growth.
8 CONCLUSIONS

8.1 Summary
This thesis outlines the basis for applying spatial science to understanding the spatio-temporal distributions of honey bee populations (both feral and managed) and floral resources. This has entailed an examination of several spatial processes; point processes, honey bee mating, and directed movement. It has involved the development of spatially explicit agent-based models with visualisation and statistical analysis of their results. Finally, this thesis outlines a new application for visualising remote sensing which contributes to the vision of Digital Earth.

Chapter 4 describes a novel approach to using agent-based models to generate synthetic sampling distributions for genetics-based field surveys of honey bee colony densities. Colony densities vary with number of colonies observed in the sample in a log-linear relationship, rather than a linear relationship as previously thought. As a consequence, previous field surveys have overestimated densities by an order of magnitude. By using the synthetic sampling distributions, new interpretations of the results from previous field surveys have been provided. The results of simulations
conducted with different characteristic spatial distributions of colonies and DCAs (random, aggregated, and overdispersed) have shown that the count of unique colonies observed in a sample is relatively insensitive to the spatial distribution. This is suggestive of a highly evolved spatial process for mating that ensures genetic diversity in the offspring irrespective of landscape configuration. By using the results of the simulations, Chapter 4 also provides guidelines for the design of future genetics-based field surveys. While the distributions developed in Chapter 4 apply only to *A. mellifera*, the approach used to development them has applicability beyond just honey bees.

Chapter 5 describes the first field experiments using genetics-based survey methods of feral colonies in Australia. These surveys have shown that densities of feral honey bees are remarkably uniform across different types of environments, with some evidence that densities are marginally higher in undisturbed environments. Regardless of the type of environment, densities of feral honey bees are not high enough to provide pollination of most horticultural or agricultural crops.

Chapter 6 shows that in planning the movement of their hives from one set of flowering events to the next, beekeepers need to solve complex optimisation problems. Where a beekeeper is able to use foresight to predict the location of future flowering events, the routes they choose can be up to 2-6% shorter than those obtained from using current flowering information only. While this is not a large improvement in absolute terms, human performance on similar optimisation problems has been shown to be generally within 10% of the optimal solution so a 2-6% improvement is significant in this context.
Chapter 7 demonstrates the value of remote sensing for understanding changes in floral resources for the Australian honey bee industry. This chapter highlights for the first time agreement between patterns of eucalypt growth, flowering, and honey production at landscape scale. Chapter 7 also describes the first software application to make remotely-sensed vegetation indices data accessible to a general audience. This application, named BeeBox, has been developed using almost exclusively open source tools. While BeeBox has been developed to assist beekeepers plan their seasonal operations more effectively, it also provides an open platform for research beyond just the honey bee industry.

This final chapter brings together the three elements researched in this thesis, namely feral colonies, managed colonies, and eucalypt phenology to assess some of the risks to pollination security. There are still large gaps in our knowledge about the various elements affecting pollination security, and more work is needed to fully understand the relationships between climate, growth, eucalypt flowering, and the effects of flowering on feral colony densities and beekeeper movements.

8.2 Discussion and outlook

8.2.1 Feral colony densities

Genetic-based techniques are extremely valuable for surveying feral colony densities. Surveys can be conducted effectively in any type of terrain, overcoming a significant limitation of direct search methods. The area surveyed is large, and necessarily has biological meaning being equal to either the flight range of drones (in the case of the drone capture
method) or the combined flight range of queens and drones (in the case of the queen mating method). Both methods rely on using honey bees to find other honey bees, a task for which they are eminently suited (and we humans are not).

The results from simulations and from field surveys suggest that different genetic techniques may be appropriate in different circumstances. Extremely low densities are best surveyed by the queen mating method, on account of the larger area sampled and the ability of queens to locate DCAs with greater ease than humans. Experience from the field surveys conducted by Hinson et al. (2015) highlight how difficult it can be to locate DCAs in areas of low colony density. Higher relative densities are better suited to surveying via the drone capture method, as a lesser number of individuals need to be sequenced and there is less uncertainty in the parentage inferences as there is no need to infer the maternal haplotype. Section 4.2.6 (Arundel et al., 2012) recommends that 10 queens are needed to resolve order of magnitude differences in density, and Section 5.2.3 (Arundel et al., 2014) proposes sampling 48 brood per queen, thus requiring 480 individuals to be sequenced for the queen mating method. Section 4.3.6 (Arundel et al., 2013) suggests that a sample size of 192 drones is used in the drone mating method to resolve densities of up to 10 colonies/km². To survey the same location with the queen mating method therefore requires more than twice the number of samples to be collected, sequenced and analysed. If absolute certainty is needed that no managed colonies are captured in the sample, or if there is
a need to locate colonies precisely, then a direct search of the survey area for colonies is the only option.

Both the design of future surveys using genetic-based techniques and the interpretation of survey results have been improved through the use of the simulations discussed in Chapter 4. Recommendations for the number of queens to be used in the queen mating method and the drones to be analysed in the drone capture method are both based on simulation results across a large range of densities and characteristic spatial distributions. Unfortunately, neither of the surveys described in Chapter 5 implement these recommendations as both agent-based models described in Chapter 4 were developed after the surveys had already been conducted. Future surveys will however benefit from these survey design recommendations. The interpretation of results from genetic-based surveys is much improved by the use of synthetic sampling distributions. Previous interpretations of genetic-based survey results have relied on parameters derived from a single experiment. Synthetic sampling distributions have been derived using probabilistic models of queen and drone behaviour with results collected across tens of thousands of experiments. In reporting a range of possible densities for a survey conducted at a specific location, synthetic sampling distributions better capture the measurement uncertainty inherent in the sample size.

Simulations also suggest that the observed count of unique colonies in a sample, be it worker brood or drones, is relatively insensitive to the characteristic spatial distribution of feral colonies, and to the distribution
and density of DCAs. This suggests a highly evolved spatial process that ensures successful matings occur irrespective of landscape configuration.

One unresolvable challenge in the interpretation of surveys conducted by different researchers in different locations using different survey techniques is the meaning of the term "density". O'Sullivan and Unwin (2003) state that the problem with using density as a measure of a spatial point pattern intensity is the sensitivity to definition of the study area. It is perhaps unsurprising then that the highest reported feral colony densities come from surveys conducted using direct searches of relatively small areas (Oldroyd et al., 1997, Goodman and Hepworth, 2004). In another similarly sized area, some of the same authors (Goodman and Hepworth, 2004) failed to find any colonies, which resulted in the lowest reported colony density in the literature; 0 colonies/km². Landscapes are typically heterogeneous, and direct search has shown that feral colonies have a tendency to form aggregations. Under such circumstances, random transects of the size used by Oldroyd et al. (1997) (0.25 km²) or Goodman and Hepworth (2004) (0.06 – 0.27 km²) can yield significantly different results based on the choice of survey area. This is best illustrated visually. Figure 39 shows an aggregated spatial point pattern of colonies within a 50 km × 50 km area generated using the rMatClust package (Baddeley and Turner, 2005). The area surveyed by Oldroyd et al. (1997) corresponds to a kernel bandwidth of 0.28 km; the resulting local density estimates in Figure 39a range from 0 to 10 colonies/km². Using a kernel bandwidth of 4 km, corresponding to the flight range of foraging workers (Seeley, 1995), results in local density estimates in
Figure 39b range from 0 to 1.1 colonies/km². Finally, using a kernel bandwidth of 15 km, corresponding to the maximum mating distance of queens and drones (Jensen et al., 2005), results in local density estimates in Figure 39c ranging from 0.25 to 0.55 colonies/km². All reported densities are within the range reported in the literature (Table 4); the value reported from an actual survey conducted of an area with this configuration would depend on the survey area selected and the choice of survey technique. From the perspective of understanding colony densities at a landscape level, the larger range of the genetic-based survey techniques clearly provides a more meaningful and more robust measure of density than direct survey methods.
Figure 39: Aggregated spatial point pattern with kernel density bandwidths of (a) 0.28 km, (b) 4 km, and (c) 15 km
After using the synthetic sampling distributions to reinterpret the results of other genetic based studies, feral colony densities in South East Australia have been found to be similar to those in other parts of the world. Feral colony densities worldwide are typically within the order of magnitude ranging from 0.1 to 1 colonies/km². That the densities of feral colonies in Australia is similar to other regions is somewhat surprising given that Australia is one of the few regions worldwide free of varroa, and varroa is known to cause significant losses of feral populations. One possible explanation is the observations that over time feral populations do adapt to the presence of varroa (Seeley, 2007, Rosenkranz et al, 2010, Locke and Fries, 2011), and thus populations elsewhere in the world have recovered to levels similar to those observed in Australia.

There is a small, but statistically significant, difference in the feral colony density between disturbed and undisturbed habitats in Australia, with densities typically higher in the latter environments. This highlights the need to conserve habitat for feral honey bee populations, especially in areas of broad-scale agriculture or horticulture where crops are reliant of incidental population by feral honey bees.

Lebuhn et al (2013) recommend the establishment of pollinator monitoring programs, and Australia has a unique opportunity to commence such a program prior to the inevitable arrival of varroa. Monitoring changes in feral honey bee colony densities across space and across time would allow for an improved understanding of the relationship between abiotic and biotic factors and population distribution and abundance. Such a monitoring program would also
provide an opportunity to harmonise results obtained from different survey techniques; one or more of the survey techniques (queen mating, drone capture, direct search) could be applied simultaneously at a given location.

Finally, there is scope to improve the genetics-based survey techniques in future research based on advances in landscape genetics. At the moment, no specialised tools exist for the inference of sibship/parentage relationships using linked microsatellite markers. The development of specialised tools would allow better inferences to be made about the number of unique colonies present in a sample of either worker brood or drones. In addition, there is spatial information embedded in the frequency of observed haplotypes that could be used if techniques were developed to make this possible. For example, in the queen mating survey described in Section 5.2, the unique haplotypes observed in each sample are lumped together to determine a count of unique colonies. Given that each queen has chosen DCAs independently, observations about the frequency with which a particular haplotype occurs in the offspring of multiple queens may reveal information about the likely proximity of that colony. A haplotype appearing in the offspring of every queen is more likely to indicate a close colony than a haplotype appearing in the offspring of only one test queen. Finally, the paired site approach discussed in Chapter 5 shows that it is possible with multiple survey sites to make some inferences about the direction in which particular colonies are located. Section 5.2.5 describes how four of the Wyperfeld test queens had mated with drones from the same colony as one Yaapeet test
queens. Intuitively, it is likely that this colony is located somewhere between the Wyperfeld and Yaapeet sites, and based on the higher number of matings of Wyperfeld queens, located closer to Wyperfeld than to Yaapeet. Landscape genetic techniques could be used to thus generate probabilistic distributions of likely colony locations based on how frequently a haplotype has been observed at two (or even three) nearby sites.

8.2.2 Managed colony movements

In theory, comprehensive data on the movement of managed colonies by beekeepers exists but this is held at the individual beekeeper level rather than in a form suitable for research. Beekeepers are required to keep records of hive movements for the last three years (Australian Honey Bee Industry Council, 2015). These records have never been collected or analysed at a large scale. The lack of any auditing of beekeeping records or penalty for non-compliance is likely to mean that many beekeepers are not keeping records beyond those that they use for their own purposes. Where beekeepers do keep records, it is highly likely that the format in which they keep them will vary from beekeeper to beekeeper. Published examples of seasonal movements, such as those described in Table 3 and Figure 2, are rare. While some state government departments do hold records of beekeeper registrations and associated leases on public land, these represent only a small proportion of the sites used. While beekeeper movements frequently occur across state boundaries, there is no aggregation of the information held by the state governments at the national level.
There are two possible avenues to collect data about beekeeper hive movements at a large scale. The first is to implement a system similar to the National Livestock Identification System (NLIS). The NLIS is currently used for recording the movement of cattle, sheep and goats (Meat & Livestock Australia, 2015). It comprises an animal identifier (e.g., an electronic tag), identification of the physical location of the animal (a property identification code), and a web accessible database to store and correlate location and movement data (Animal Health Australia, 2013). The NLIS is viewed as an important tool for biosecurity, food safety and market access (Animal Health Australia, 2013). There are no technical reasons why a similar system could not be implemented for hive tracking providing similar benefits to the honey bee and pollination industries. However, a number of barriers to adoption of such a system can be anticipated. Given that the greater economic benefits of honey bees is in the pollination services they provide rather than honey and beeswax products, there is likely to be considerable debate as to which parties should bear what proportion of the costs of introducing the system. The introduction would require policy and regulatory changes. Finally, implementing the system would require the support and cooperation of beekeepers. Beekeepers are not known as eager adopters of new technologies. A possible solution to gaining beekeeper support for any new technology is to demonstrate the economic incentives of adoption. This then leads to a second avenue for collecting data on hive movements; a production tool similar to BeeBox. If for an individual beekeeper a tool like BeeBox was able to make personalised predictions of flowering
events based on data entered by the beekeeper, it would likely gain considerable traction within the industry. At an aggregate level, the data collected by such a system, either actively (i.e., entered by beekeepers) or passively (e.g., geo-location of IP addresses used to access the system), could form the basis of a movement analytics platform.

Rather than rely on as yet unavailable data, Chapter 6 describes how some progress has been made towards predicting beekeeper movement in general terms. Beekeepers need to make predictions about flowering events then solve complex optimisation problems to efficiently plan their seasonal movements and maximise their profitability. However, actual beekeeper performance on solving the abstracted problem described in Chapter 6 is yet to be measured.

8.2.3 Linking eucalypt phenology and honey bee distributions

Underpinning both the temporal changes in feral colony densities and the seasonal movements of beekeepers is the flowering phenology of eucalypts. In Australia, Oldroyd et al. (1997) has hypothesised about the linkage between temperature and rainfall and colony densities. In practice, climatic variables such as temperature and rainfall regulate plant growth, which occurs as a precursor to flowering, which then results in nectar and pollen production. This simplified chain of causal relationship is illustrated in Figure 40. Linkages between some elements in the chain have been constructed through scientific studies, but we are still far from understanding the end-to-end relationships.
One of the challenges in linking the elements of the chain illustrated in Figure 40 is that the level of quantitative data available for each element varies considerably. Many decades of precise climatic data is available at a precise level over most of the Australian land mass. Since the discontinuation of forest commission record keeping in the early 1980s, little new data has been gathered on eucalypt growth and flowering. Only two new field studies have been conducted in this period; these are reviewed in Chapter 7. Given the lack of quantitative ground-truthed growth data, remote sensing observations can provide a useful proxy for vegetation growth as illustrated in Chapter 7. The best source of quantitative data on nectar production lies in the records kept by individual beekeepers, and perhaps more usefully, by honey packers. Unfortunately, no large scale quantitative data exists on pollen production. Finally, quantitative data on the spatio-temporal distribution of both managed and feral honey bee colonies is extensively reviewed and discussed in Chapters 2 and 4. Data exists, but only for relatively small areas over relatively short periods of times, and the precision of the estimates creates challenges in drawing conclusions about relationships of feral colony densities with other variables in the chain.
Following on from the suggestion in Section 8.2.1 to establish a pollinator monitoring program, and the recommendation in Section 8.2.2 to implement a system similar to the NLIS for the beekeeping industry, there is also considerable value in the re-establishment of monitoring programs for eucalyptus growth and flowering. Although we live in an era where there are increased concerns about the effects of climate change and threat of bushfires, there is no routine monitoring currently conducted of the growth and flowering of eucalypt forests in Australia. Previous generations saw the value in monitoring eucalypt growth and flowering, and the need to do so is more pressing now than then.

Chapter 7 begins an exploration of the linkages between climate, growth, flowering and honey production using MODIS EVI time series generated by the prototype BeeBox system. For *E. tricarpa* at least, there seems to be an observable relationship between growth, flowering and honey production. The relationship between climate variables and remotely sensed growth indices is cryptic, and it has not been possible in this thesis to do more than illustrate this fact. Further work is needed to generalise the findings of Chapter 7 across a larger range of species, and to perform statistical analyses of the relationships between causal variables.

It is possible that predictions of flowering from growth will only be possible if site histories are known. The best source of data on flowering at the landscape scale is beekeepers. If tools like BeeBox can be developed to the point where they can provide personalised site predictions based on information entered by beekeepers, then these same tools will necessarily also capture beekeeper knowledge of site history.
Beekeepers may also provide the means to effectively ground truth predictions of flowering at large scale.

For land surface phenology studies of eucalypt forests, it is highly likely that better alternatives to the EVI and NDVI exist. Most of the vegetation indices discussed in Chapter 7 have been developed using leaf reflectance measurements for northern hemisphere species (Datt, 1998). Based on the approach of Datt (1999a, 1999b), there may be merit in deriving a VI specifically for use in the study of eucalypt phenology. Neither the EVI or NDVI is particularly sensitive to the changes in anthocyanin concentration that signal new growth in many eucalypt forests. Many eucalypt forests also possess open canopies, and in time series analysis it would be highly desirable to discriminate between changes occurring at the canopy level and changes occurring in the understorey. Eucalypt-specific phenology indices would have value beyond the study of forests in Australia.

Eucalypts are an important timber product globally, and account for 10% of global forest plantation, second only in significance to the genus *Pinus* with 20% (Palmberg-Lerche *et al.*, 2001). In practice, it may be necessary to utilise multiple different indices tuned to different phenological events to best monitor eucalypt forests, and these might need to be tailored to the different eucalypt forest types. Moving away from standardised indices would also necessitate generation of the new indices from Level 2 data, and this would add cost and complexity to the analysis.

For land surface phenology studies, there is also a case to be made for utilising data from multiple sensors. Land surface phenology studies typically only use data from a single platform (Table 25), and for this
reason data is usually composited to remove the effects of cloud cover. The use of data from multiple sensors would increase the probability of obtaining a cloud-free pixel on any given day. In addition, the different spectral bands of the different sensors (Table 24) could be exploited to discriminate phenological events.

Advances in the design of phenological vegetation indices could also help resolve the uncertainty in eucalypt species’ distributions. Accurate species distributions are critical to the development of models predicting flowering from growth data, and models of flowering are critical to the prediction of feral colony density and beekeeper hive movements. Section 7.1.2 illustrates the extent to which different references disagreed on the distribution of eucalypt species. Section 7.3.4 shows that even within a single defined EVC, there is evidence of considerable variation in the EVC composition. Time series analysis of vegetation indices can highlight which pixels share similar signatures, and by correlating this with existing field survey data and other high spatial and spectral resolution data improved species distributions could be generated.

8.2.4 Consequences for pollination security

On the basis of the findings outlined in this thesis, what can we conclude about the current status of pollination security in Australia? Firstly, it is possible that there are more feral honey bee colonies in Australia today than managed colonies. Australia has around 1.5 million km² of forest (The World Bank, 2015), and around 570,000 managed honey bee colonies (Crooks, 2008). Densities in the ranges observed in Chapter 5
would mean that feral colonies exceed managed colonies in number. Secondly, only managed colonies exist at a density sufficient to provide pollination of agricultural and horticultural crops. Thirdly, the spread of varroa will in the short term at least significantly reduce feral colony numbers, but this is more likely to impact pollination in natural ecosystems than for agricultural or horticultural crops. Fourthly, feral eradication at a large scale in response to a biosecurity incident is unlikely to be feasible given the dispersed presence of colonies across many different kinds of landscapes. Finally, despite the investigations outlined in Chapters 6 and 7 we know remarkably little about the patterns of eucalypt flowering at landscape scale, and the movements of hives by beekeepers in response to these patterns. We cannot predict when devastating honey seasons such as those experienced in 2012-2013 (Victoria), 2013-2014 (Queensland) and 2014-2015 (Tasmania) are going to occur, and therefore we cannot predict when the pollination of agricultural and horticultural crops might suffer due to a lack of managed hives. This lack of knowledge about flowering and beekeeper movements is a threat to pollination security.

With more data on feral colony densities over time, flowering events and hive movements it would be possible to provide a better assessment of the risks to pollination, and even to mitigate some of these. The impacts of feral colonies on pollination and other pollinators in natural environments could be better understood, poor flowering seasons could be predicted and the risks of managed colony losses better managed, and
when biosecurity incidents occur responses involving both managed and feral colonies could be better planned and executed.
9 REFERENCES


pollination shortage but increasing pollinator dependency', Current Biology, 18(20), 1572-1575.

Aizen, M. A. and Harder, L. D. (2009) 'The global stock of domesticated honey bees is growing slower than agricultural demand for pollination', Current Biology, 19(11), 915-918.


Chapter 9: References


BeeAware (2015a) *Africanised honey bee* [online], available:  

[accessed 17/07/2015].

BeeAware (2015c) *Small hive beetle* [online], available:  

Beekman, M., Allsopp, M. H., Lim, J., Goudie, F. and Oldroyd, B. P. (2011)  
'Asexually produced Cape honeybee queens (*Apis mellifera capensis*) reproduce sexually',  
*Journal of Heredity*, 102(5), 562-566.

*Varroa destructor* in New Zealand bees: delimiting survey results and management options', in  


Bureau of Meteorology (2015b) *Climate data online* [online], available:  


Chapter 9: References


Capilano Honey Limited (2015) *Broker presentation*, February 2015, Brisbane:

Capilano Honey Limited, available:


The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia


eucalypt biochemistry: preliminary project design and specifications', in

IEEE International Geoscience and Remote Sensing Symposium

(IGARSS), Sydney, 9-13 July 2001, IEEE International Geoscience and

crown condition in eucalypt vegetation by remotely sensed optical indices',


Coops, N. C., Stone, C., Culvenor, D. S., Chisholm, L. A. and Merton, R. N.

(2003) 'Chlorophyll content in eucalypt vegetation at the leaf and canopy
scales as derived from high resolution spectral data', Tree Physiology,
23(1), 23-31.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B.,
Limburg, K., Naeem, S., O'Neil, R. V., Paruelo, J., Raskin, R. G., Sutton,
P. and van den Belt, M. (1998) 'The value of the world's ecosystem


Melbourne: Costermans Publishing.

Crane, E. (1999) The world history of beekeeping and honey hunting, New York:

Routledge.


The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia


Department of Environment and Primary Industries (2015b) *Native vegetation - modelled extent 2005* (NV2005_EXTENT) [database], ANZVI0803003404, Melbourne: Department of Environment and
The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia

Primary Industries, available:


Department of Environment and Primary Industries (2015c) Victorian bioregions
- mapped at 1:100,000 (version 3.0 - May 2004)

(VBIOREG100/VBIOREG) [database], ANZVI0803002636, Melbourne:
Department of Environment and Primary Industries, available:


management (PLM25) generalised [database], ANZVI0803003978, Melbourne: Department of Environment, L., Water & Planning, available:

[accessed 03/09/2015].


Department of Sustainability and Environment (2007b) 'Native vegetation information: overview of native vegetation spatial datasets', (Information sheet no. 1), Melbourne: Victorian Government Department of Sustainability and Environment.

Department of the Environment (2012) National vegetation information system (NVIS) [online], [accessed 11/05/2015].


DigitalGlobe (2015a) Satellite information [online], available:
https://www.digitalglobe.com/resources/satellite-information [accessed 29/05/2015].

DigitalGlobe (2015b) WorldView-3 [online], available:
https://www.digitalglobe.com/sites/default/files/DG_WorldView3_DS_for Web_0.pdf [accessed 29/05/2015].
Diss, K. (2011) *Doing it tough as honey production slumps* [online], available: 


Chapter 9: References

_Echium wildpretii_ (Boraginaceae) on Tenerife, Canary Islands', _Biological Conservation_, 118(3), 301-311.


The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia


ESA (2012a) *ESA declares end of mission for Envisat* [online], ESA website, available: http://www.esa.int/Our_Activities/Observing_the_Earth/Envisat/ESA declares_end_of_mission_for_Envisat [accessed 03/05/2015].

ESA (2012b) *MERIS frequently asked questions* [online], available: http://earth.esa.int/pub/ESA_DOC/ENVISAT/MERIS/VT-P017-DOC-005-E-01-01_meris.faq.1_1.pdf [accessed 03/05/2015].

ESA (2015) *Sentinel 3* [online], ESA Earth Online, available: https://earth.esa.int/web/guest/missions/esa-future-missions/sentinel-3 [accessed 02/05/2015].


[http://cce.nasa.gov/mtg2008_ab_presentations/Phenology_Friedl_whitelpdf](http://cce.nasa.gov/mtg2008_ab_presentations/Phenology_Friedl_whitepdf) [accessed 25/05/2015].


Gordon, R., Bresolin-Schott, N. and East, I. J. (2014) 'Nomadic beekeeper movements create the potential for widespread disease in the honeybee industry', *Australian Veterinary Journal*, 92(8), 283-90.


Hamblyn, J. (2008) *Varroa movement controls revoked* [online], available: 


The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia


Kim, S. W., Hudson, I. L. and Keatley, M. R. (2009) 'Modelling the flowering of four eucalypts species via MTDg with interactions', in Anderssen, R. S.,


Klein, A. M., Vaissiere, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C. and Tscharntke, T. (2007) 'Importance of pollinators in...


The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia


NDVI time-series data for the northwest of China', *Advances in Space

Ma, X., Huete, A. R., Yu, Q., Coupe, N. R., Davies, K., Broich, M., Ratana, P.,
'Spatial patterns and temporal dynamics in savanna vegetation phenology
across the North Australian Tropical Transect', *Remote Sensing of
Environment*, 139, 97-115.

Mac Nally, R., Bennett, A. F., Brown, G. W., Lumsden, L. F., Yen, A., Hinkley,
planning units represent different components of biodiversity?', *Ecological
Applications*, 12(3), 900-912.

Mac Nally, R., Bennett, A. F., Thomson, J. R., Radford, J. Q., Unmack, G.,
change appears to exacerbate habitat loss and degradation', *Diversity and

endangered migrant: the Swift Parrot (*Lathamus discolor*) in its winter
range', *Biological Conservation*, 92(3), 335-343.


The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia


Chapter 9: References


Miura, T. (2015) *VIIRS vegetation indices* [online], available: 
http://viirsland.gsfc.nasa.gov/Products/VI.html [accessed 02/05/2015].


Morse, R. A., Camazine, S., Ferracane, M., Minacci, P., Nowogrodzki, R.,
population density of feral colonies of honey bees (Hymenoptera: Apidae)
in a city in upstate New York', *Journal of Economic Entomology*, 83(1),
81-83.

Nabhan, G. P., Allen-Wardell, G., Bernhardt, P., Bitner, R., Burquez, A.,
Buchmann, S., Cane, J., Cox, P. A., Dalton, V., Feinsinger, P., Ingram, M.,
Inouye, D., Jones, C. E., Kennedy, K., Kevan, P., Koopowitz, H.,
Medellin, R., Medellin-Morales, S., Pavlik, B., Tepedino, V., Torchio, P.
on the conservation of biodiversity and stability of food crop yields',
*Conservation Biology*, 12(1), 8-17.

NASA Land Processes Distributed Active Archive Center (LP DAAC) (2015)
*MOD13Q1* [database], Sioux Falls, South Dakota: NASA Land Processes
Distributed Active Archive Center (LP DAAC).

National Bee Pest Surveillance Program (2014) *Operations manual* [online],


Office of Environment & Heritage (2015a) Vegetation information system: classification [online], available:
The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia

[accessed 11/05/2015].

Office of Environment & Heritage (2015b) Vegetation information system: maps
[online], available:


The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia


Plant Health Australia (2012b) Biosecurity online training [online], available: http://www.planthealthaustralia.com.au/resources/training/biosecurity-online-training/ [accessed 31/07/2015].


Chapter 9: References


Reynolds, A. M. (2009) 'Levy flight patterns are predicted to be an emergent property of a bumblebees' foraging strategy', Behavioral Ecology and Sociobiology, 64(1), 19-23.


Cascella, R., Kecskemeti, G., Jeannot, E., Cannataro, M., Ricci, L.,
Benkner, S., Petit, S., Scarano, V., Gracia, J., Hunold, S., Scott, S.,
Lankes, S., Lengauer, C., Carretero, J., Breitbart, J. and Alexander, M.,


Stevenson, M. A., Benard, H., Bolger, P. and Morris, R. S. (2005) 'Spatial epidemiology of the Asian honey bee mite (Varroa destructor) in the North Island of New Zealand', Preventive Veterinary Medicine, 71(3-4), 241-52.


Chapter 9: References

Australian eucalypt forest', *Atmospheric Chemistry and Physics*, 8(1), 129-139.


The University of Queensland (2011a) *Capeweed Arctotheca calendula* [online], Environmental Weeds of Australia for Biosecurity Queensland, available:
http://keyserver.lucidcentral.org/weeds/data/03030800-0b07-490a-8d04-0605030c0f01/media/Html/Arctotheca_calendula.htm [accessed 14/06/2015].

The University of Queensland (2011b) *Paterson's curse Echium plantagineum* [online], Environmental Weeds of Australia for Biosecurity Queensland, available: http://keyserver.lucidcentral.org/weeds/data/03030800-0b07-490a-8d04-0605030c0f01/media/Html/Echium_plantagineum.htm [accessed 14/06/2015].


The Spatio-temporal Distribution of Honey Bees and Floral Resources in Australia


van der Zee, R., Brodschneider, R., Brusbardis, V., Charriere, J.-D., Chlebo, R.,
Coffey, M. F., Dahle, B., Drazic, M. M., Kauko, L., Kretavicius, J.,
Kristiansen, P., Mutinelli, F., Otten, C., Peterson, M., Raudmets, A.,
Santrac, V., Seppala, A., Soroker, V., Topolska, G., Vejsnaes, F. and Gray,
A. (2014) 'Results of international standardised beekeeper surveys of
colony losses for winter 2012-2013: analysis of winter loss rates and
mixed effects modelling of risk factors for winter loss', Journal of
Apicultural Research, 53(1), 19-34.

van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y.,
and implications for water resources, ecosystems, economy, and society',
Water Resources Research, 49(2), 1040-1057.

'Minimum required number of specimen records to develop accurate
species distribution models', Ecography, 38, 1-11.

Vanbergen, A. J. and Insect Pollinators Initiative (2013) 'Threats to an ecosystem
service: pressures on pollinators', Frontiers in Ecology and the
Environment, 11(5), 251-259.


Chapter 9: References


10 APPENDICES

ADDITIONAL FIGURES ....................................................................................................................415
**ADDITIONAL FIGURES**

Figure 41: Simulation space with randomly distributed DCAs (green stars) and randomly distributed feral colonies (circles). Within range colonies sampled are blue, and not sampled are red.
Figure 42: Simulation space with randomly distributed DCAs (green stars) and aggregated feral colonies (circles). Within range colonies sampled are blue, and not sampled are red.
Figure 43: Simulation space with randomly distributed DCAs (green stars) and overdispersed feral colonies (circles). Within range colonies sampled are blue, and not sampled are red.
Figure 44: Results of 100 simulations for a sample size of 96 drones with aggregated feral colonies and randomly distributed DCAs
Figure 45: Results of 100 simulations for a sample size of 96 drones with hyper-aggregated feral colonies and randomly distributed DCAs
Figure 46: Results of 100 simulations for a sample size of 96 drones with overdispersed feral colonies and randomly distributed DCAs.
Figure 47: Results of 100 simulations for a sample size of 96 drones with randomly distributed feral colonies and aggregated DCAs
Figure 48: Results of 100 simulations for a sample size of 15 drones with aggregated feral colonies and randomly distributed DCAs
Figure 49: Results of 100 simulations for a sample size of 192 drones with aggregated feral colonies and randomly distributed DCAs
Figure 50: Distribution of mean flight distances of all drones present at the sampled DCA with randomly distributed feral colonies and randomly distributed DCAs
Figure 51: Distribution of flight distances for all drones present at the sampled DCA across all simulations for randomly generated feral colonies of density 0.1/km² and randomly distributed DCAs of density 0.1/km².
Figure 52: Distribution of the proportion of colonies detected in the sample versus within range of the DCA for a sample size of 15 drones with randomly distributed feral colonies and randomly distributed DCAs
Figure 53: Distribution of the proportion of colonies detected in the sample versus within range of the DCA for a sample size of 96 drones with randomly distributed feral colonies and randomly distributed DCAs.
Figure 54: Distribution of the proportion of colonies detected in the sample versus within range of the DCA for a sample size of 192 drones with randomly distributed feral colonies and randomly distributed DCAs
Figure 55: Results of 100 simulations at varying DCA densities for a sample size of 14 drones with aggregated feral colonies and randomly distributed DCAs
Figure 56: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 14 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 57: Results of 100 simulations at varying DCA densities for a sample size of 29 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 58: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 29 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 59: Results of 100 simulations at varying DCA densities for a sample size of 62 drones with aggregated feral colonies and randomly distributed DCAs
Figure 60: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 62 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 61: Results of 100 simulations at varying DCA densities for a sample size of 70 drones with aggregated feral colonies and randomly distributed DCAs
Figure 62: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 70 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 63: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 72 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 64: Results of 100 simulations at varying DCA densities for a sample size of 74 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 65: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 74 drones with aggregated feral colonies and randomly distributed DCAs
Figure 66: Results of 100 simulations at varying DCA densities for a sample size of 79 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 67: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 79 drones with aggregated feral colonies and randomly distributed DCAs
Figure 68: Results of 100 simulations at varying DCA densities for a sample size of 123 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 69: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 123 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 70: Results of 100 simulations at varying DCA densities for a sample size of 222 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 71: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 222 drones with aggregated feral colonies and randomly distributed DCAs
Figure 72: Results of 100 simulations at varying DCA densities for a sample size of 241 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 73: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 241 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 74: Results of 100 simulations at varying DCA densities for a sample size of 278 drones with aggregated feral colonies and randomly distributed DCAs.
Figure 75: Results of 100 simulations at a DCA density of 1 DCA/km² for a sample size of 278 drones with aggregated feral colonies and randomly distributed DCAs
Author/s:
Arundel, Jonathan Paul

Title:
The spatio-temporal distribution of honey bees and floral resources in Australia

Date:
2015

Persistent Link:
http://hdl.handle.net/11343/59612

File Description:
Main thesis