Spatiotemporal database and floristic vegetation mapping to support honeybee product development

Project Report: Ryan Kerr

August 2021

Supervisor: A/Prof Bryan Boruff

CRCHBP
FOR HONEY BEE PRODUCTS

Introduction

Honey bees (*Apis mellifera*) have a significant role in the commercial production of honey and pollination services in Australia (Benecke, 2007). As a source of crop pollination, many farmers believe that the feral bees are sufficient for these agricultural systems (Gordon & Davis, 2003), however, several studies have shown that at an industry level, honey bees have a much higher economic footprint in pollination (Klein et al., 2007). Similarly, honey produced from honey bees has also grown into a sizeable industry with beekeepers undergoing several trainings and enrolling with multiple organisations to find suitable apiary sites. Locating appropriate sites for this large-scale production in Australia is particularly challenging due to the diverse landscape (Crane, 1999). For this reason beekeepers have been known to migrate in an attempt to move to apiary sites in both natural environments and agricultural crop regions (Arundel et al., 2016).

Honey bee populations have been declining around the world due to a number of reasons including climate change (Lebuhn et al., 2013) and additionally, flowering events for native flora in Australia are also extremely variable (Arundel et al., 2016). This has led to a large number of resources being allocated in preserving honey bee populations and identifying suitable vegetation within accessible areas for migratory beekeepers. A significant amount of research has been conducted in different parts of Australia by compiling databases for flowering flora (Somerville, 2020) as well as by using advanced scientific and machine learning methods to detect and predict phenology patterns (Dixon et al., 2021). However, there have been limited attempts in collating multiple resources for flowering flora around Australia to geospatially identify regions suitable for apiary sites at different times of the year with relation to their varying levels of honey and pollen.

Australia has developed a key tool for planning and conservation through the introduction of Interim Biogeographic Regionalisation for Australia (IBRA) regions almost three decades ago (DAWE, 2021) which are based on a number of unique characteristics of the region. Further, a more localised subdivision of these regions are the IBRA subregions, each of which consists of distinct geomorphology. Thus, by identifying the floristic vegetation useful to honey bees within these 419 IBRA subregions, a comprehensive database and visual tool can be utilised by beekeepers to focus their apiary movements. Alternatively, many sites are also influenced by horticultural and agricultural crops thus, requiring a concerted effort in evaluating dominant regions in this regard. As honey bees have been studied to have a significant effect on crop pollination (Delaplane et al., 2000), it is advantageous to recognise suitable locations for not only honey production but also for this extremely high valued insect pollination.

Therefore, this report has been prepared in accordance with the Cooperative Research Centre for Honey Bee Products and B-Qual Australia with an aim of providing a spatiotemporal database for improving management of honey bee products throughout Australia. This required a focus on first identifying the relevant herbarium species in relation to *Apis mellifora*, with the subsequent development of an extensive database of their presence locations, occurrences, flowering times as well as honey and pollen production capacities. Furthermore, a methodology for crop species has also been identified for creating this comprehensive database based on Statistical Area Level 2 (SA2) boundaries from the Australian Bureau of Statistics and the annual crop productions within these areas.

Scope and Challenges

Native flora and crop species that are relevant to honey and pollen production have already been identified for Western Australia (WA) using multiple historical research papers. Additionally, the phenological periods and habitat types have also been recorded for each individual species within WA. The scope of this project was limited to the variety of flora associated with the eastern states of Australia, however, using these herbarium collections, the IBRA subregions across Australia more suited to every species were mapped and included in the database. Moreover, the project is also limited to IBRA subregions in all states and territories except for Northern Territory based on the requirement of B-qual. Although a total of 586 species which were identified, they quite possibly do not cover the entire spectrum of different flora found in the region but do include about 139 Eucalypt species and subspecies which rely heavily on bees for pollination (Moncur et al., 1995).

The biggest challenge was centred around obtaining information for the crop production for which the herbarium collection with spatial information is not available. In most cases, production is carried out intensively in either controlled locations or green houses, therefore, this project is limited to only orchards and large-scale plantations in determining locations as intensive farming is extremely variable. For other native flora, presence locations were retrieved from on an open-source platform, Atlas of Living Australia, which includes data with location uncertainties, environmental outliers etc. Hence, developing a framework to allow maximum number of records without compromising on precision and correctness was of utmost importance. Additionally, honey and pollen information from multiple sources varied significantly due to insufficient samples being tested for some resulting in lower accuracy. Thus, quantifying these attributes for individual species based on past literature has been limited.

Methods

The overall methodology used for this project can be encompassed in three broad procedures. The first includes identifying relevant literature sources and compiling the phenology periods as well as honey, pollen and occurrences of the melliferous flora species in the eastern states of Australia. The second step involves collecting records of presence locations of these species which are freely available on Atlas of Living Australia, an open-source, collaborative database online. However, this is only applicable for other flora including native plants, herbs, trees and shrubs rather than crop species for which an alternate approach is required. Using data from the Australia Bureau of Statistics, the SA2 administrative boundaries were identified which had a production value for individual crops. Finally, for both, the other flora and crop species, the associated IBRA subregions were noted and added to the database.

In the first case, three sources of literature were identified along with one website, listing over 550 records of flowering plants suited to honey bees. A database was created in MS Excel registering the scientific names along with their family, genus and common names. Additionally, the APNIIDs were retrieved to identify other native flora in Australia. Duplicates were added to separate spreadsheets to avoid repetition but have a record in case of any discrepancies. Flowering periods throughout the year were noted along with their associated colours for identification. The monthly intensity of flowering; low, medium and high, was also used as a measure for classification. For each species regardless of whether crops or other flora, the occurrences or habitat, honey and pollen production were also included in the database.

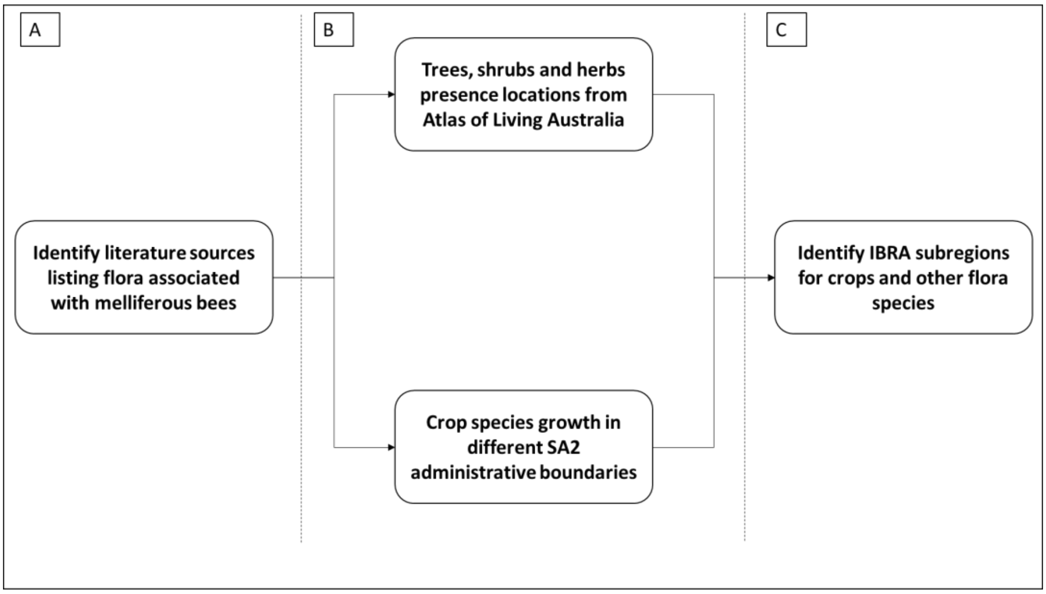


Figure 1 – Overall methodology to develop the spatiotemporal database divided into three broad procedures; (A) Identifying literature sources and extracting relevant information, (B) Identifying SA2 boundaries for crop species and obtaining presence locations for all other flora and (C) overlaying this information onto IBRA subregions for each species.

Table 1 – Data sources for each procedure in development of the spatiotemporal database of melliferous flora.

Procedure	Data Type	Name	Author	URL
A	Book	Honey and Pollen Flora of South Eastern Australia	Doug Somerville (2020)	
	Literature	Long-term flowering patterns of melliferous Eucalyptus Myrtaceae species	Birtchnell and Gibson (2006)	
	Literature	Flowering phenology of myrtaceous trees and their relation to climatic, environmental and disturbance variables in northern New South Wales	Law et. al (2000)	
	Website/Blog	Guide for selecting the best Australian suited flowers and plants to benefit the bees, pollinators and your garden.	Flowers of Melbourne	https://blog.flowersacrossmelbourne.com.au/flowers-for-bees/
B	Web database	Species Location Records	Atlas of Living Australia	https://www.ala.org.au/
	Web database	Agricultural Commodities 2015-16 for SA2 boundaries	Australian Bureau of Statistics	https://www.abs.gov.au/statistics/industry/agriculture/agricultural-commodities-australia/latest-release
C	Web database	Australia's bioregions and subregions	Department of Agriculture, Water and Environment, Australian Government	https://www.environment.gov.au/land/nrs/science/ibra

For the second step of the methodology, other flora species occurrence data was obtained from Atlas of Living Australia. Where possible only the species data was used unless the database contained a record about the subspecies. Each record was filtered for Australian boundaries along with the inclusion of location uncertainties, environmental outliers and spatial quality issues. These datasets were downloaded as comma-separated values (.csv) files containing latitudes and longitudes for every record. Alternatively, 'Agricultural Commodities' data from the 2015-16 financial year was downloaded from the Australian Bureau of Statistics (ABS). This included data for many different statistical boundaries including SA2, SA4 and States of which the former was chosen as the smallest scale to overlay with the 368 IBRA subregions excluding NT. For each crop species the associated commodity code for number of trees of bearing age where possible or the total plantation area in hectares for crop production were identified and if the condition of it being 'greater than 0' was satisfied, the related SA2 boundaries were selected. As the ABS predicts this information, the relative standard errors were also part of the expression, as those 'greater than 50%' were too unreliable for general use. To assist with this analysis, an ArcGIS Pro model was developed to automate the selection and shapefile conversion. This also included the third step for the crop species where the IBRA subregions were extracted by overlaying onto the selected SA2 boundaries using the 'Select by Location' tool. Similarly, for other herbarium collections, the IBRA subregions corresponding to the presence locations retrieved were obtained and saved as shapefiles. This was done using the 'Spatial Analyst' extension in ArcGIS Pro along with the 'os', 'sys', and 'arcpy' libraries in python notebook. The location based '.csv' files were converted to a point feature class and then IBRA subregions were extracted as shapefiles using the 'Select by Location' tool in python code.

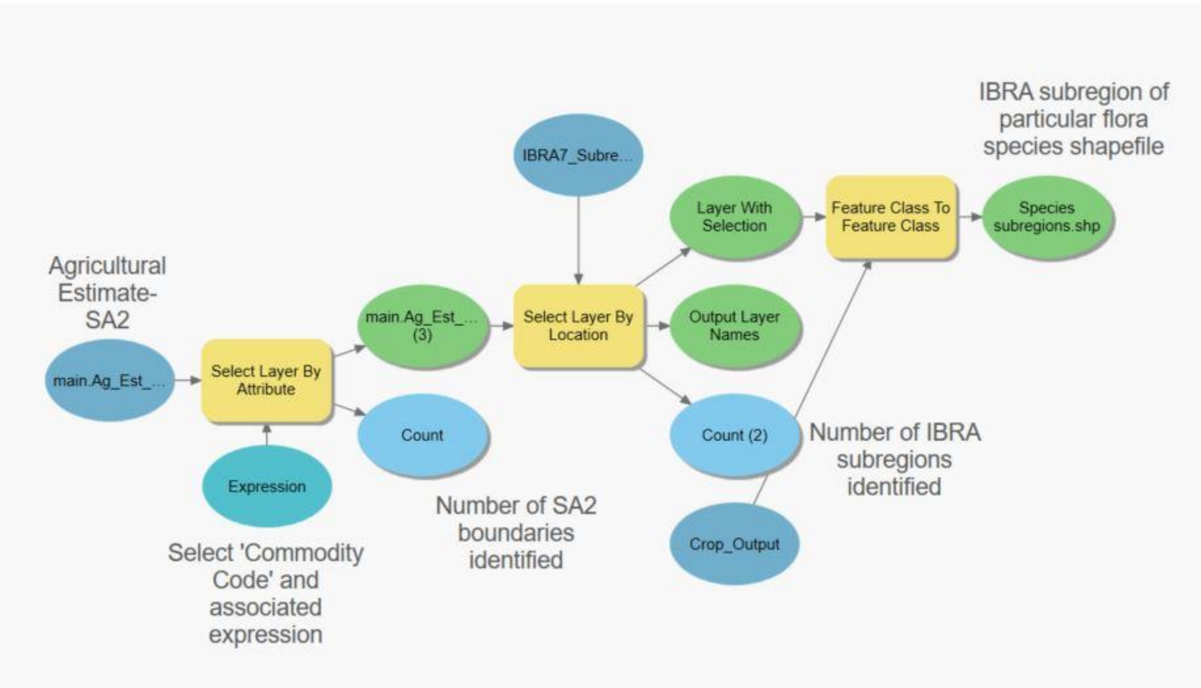


Figure 2 – Model developed in ArcGIS Model Builder for selecting and extracting SA2 boundaries based on crop species estimates obtained from ABS.

```
In [ ]:
#import libraries
import os, sys, arcpy, math
from arcpy import env
from arcpy.sa import *
# Check out the ArcGIS Spatial Analyst extension license
arcpy.CheckOutExtension("Spatial")
```

```
In [ ]:
#import variables
sceneFolder = r"F:\Semester4\SciencePracticum\DATA\NewALA"
outFolder = r"F:\Semester4\SciencePracticum\DATA\NewOutput"
outFolder2 = r"F:\Semester4\SciencePracticum\DATA\Points.gdb"
AOI = r"F:\Semester4\SciencePracticum\DATA\IBRA_Subregions\IBRA7_Subregions_Selected.shp"
```

```
In [ ]:
#check scene folder
print (sceneFolder)
print (outFolder)
print (outFolder2)
print (AOI)
```

```
In [ ]:
for file in os.listdir(sceneFolder): #accessing all filed in the scene folder using os library
    print (file)
    folderName = file
    print (folderName)
    path = sceneFolder + "\\" + folderName
    print (path)
    in_species1 = path + "\\" + folderName + ".csv"
    out_species1 = outFolder2 + "\\" + folderName + "_points"
    arcpy.management.XYTableToPoint(in_species1, out_species1,
                                    "decimalLongitude", "decimalLatitude")
    print ("created points")
    AOIselected = arcpy.SelectLayerByLocation_management(AOI, 'INTERSECT', out_species1)
    print ("selected locations")
    out_subregions1 = folderName + "_subregions"
    arcpy.FeatureClassToFeatureClass_conversion(AOIselected, outFolder, out_subregions1)
    print ("saved subregion shp" + out_subregions1)

print ("Finished processing all species")
```

Figure 3 – Python code from ArcGIS Pro notebook used to select and extract IBRA subregions for all other flora based on their presence locations obtained from ALA.

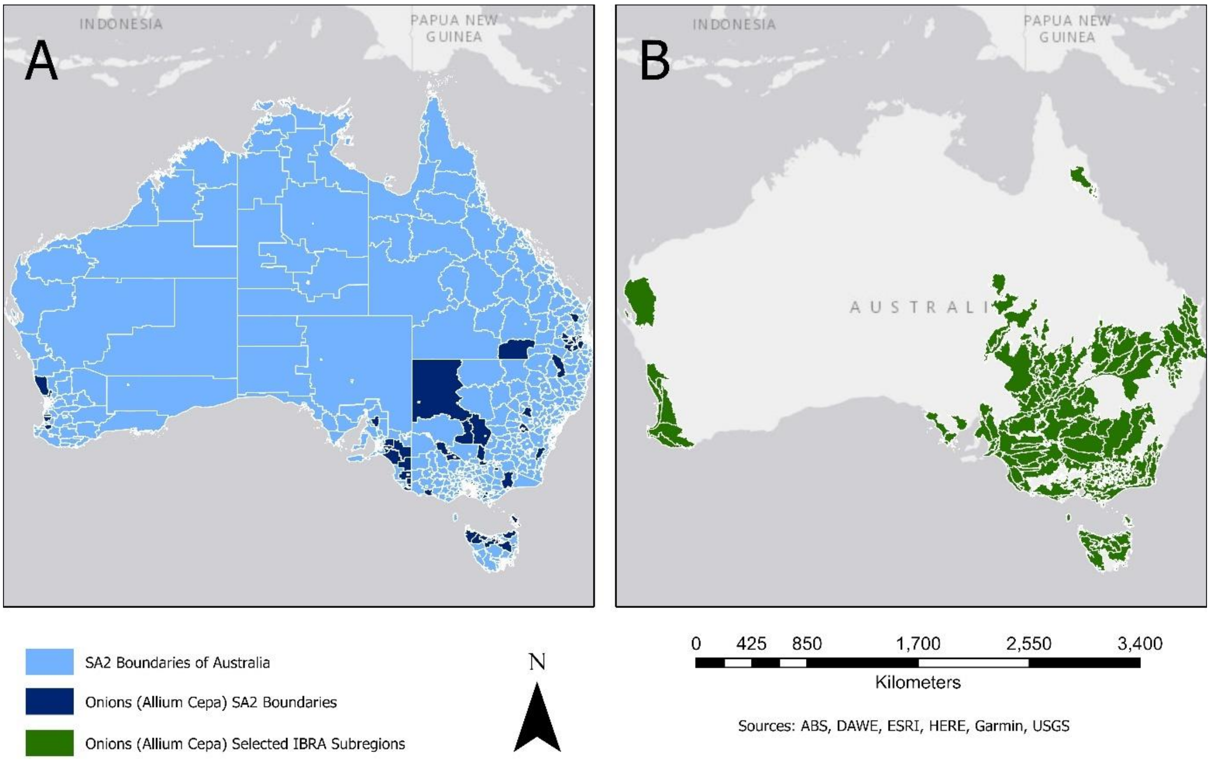
Finally for Procedure 'C', all the species' associated IBRA subregions were extracted from their respective shapefiles. This was automated separately for both crop species and other flora. The 'os', 'pandas' and 'geopandas' libraries were used within 'Spyder' IDE using python version 3.8 to read in the shapefiles and extract the IBRA subregions before writing them as a list along with the species name in the separate worksheet. The IBRA subregions codes that were obtained were separated by commas to be readable for future analysis and visual representation. These rows were then copied into the original database in two separate columns: one for crop species and other flora depending on the source of information.

```python
import os
import geopandas as gpd
import pandas as pd


Folder = r"F:\Semester4\SciencePracticum\DATA\NewCrops"

Files = [i for i in os.listdir(Folder) if i.endswith('.shp')]
print(Files)

#newdata = gpd.GeoDataFrame()
#newdata['Species'] = None
#newdata['Subregions'] = None


#x=0
myList = []

for f in Files:
    #print(f)
    full_name = os.path.join(Folder, f)
    print(full_name)

    g = gpd.read_file(full_name)
    sub_code_7 = list(g['SUB_CODE_7'])
    subregion = ','.join([str(item) for item in sub_code_7 ])

    species = f.strip('.shp')
    #species = f.split('_')[0] + ' ' + f.split('_')[1]

    myList.append((species, subregion))


res1 = pd.DataFrame(myList, columns = ['species','subregions'])
res1.to_csv(r"F:\Semester4\SciencePracticum\DATA\subregions_newcrops.csv", index=False)
```

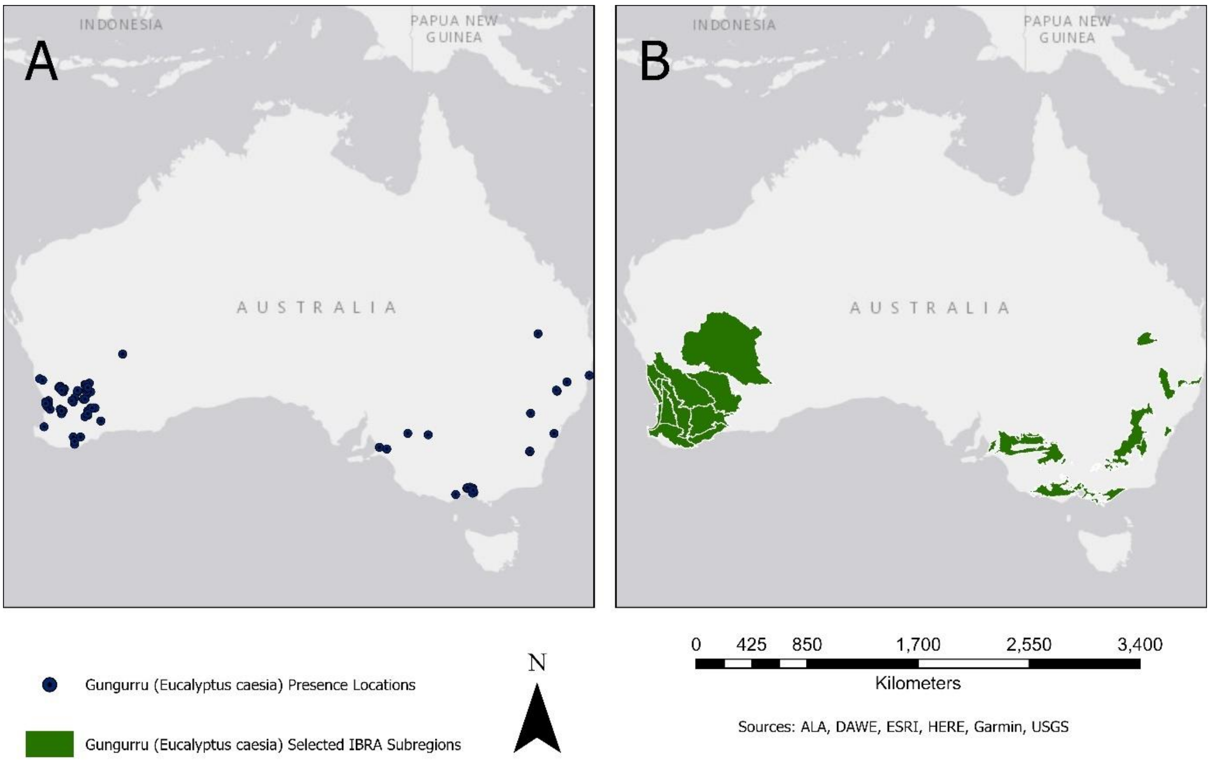
Figure 4 – Python code from Spyder IDE using 'pandas' and 'geopandas' to bring lists of IBRA subregion into the spatiotemporal database.

Results

The resulting database prepared from the above-mentioned methodology is an extensive list of flora species and there associated flowering times, occurrences, honey and pollen production as well as IBRA subregions. The data accompanying this database includes vector data of presence locations of over 500 species along with shapefiles of the subregions identified for all 586 species. Thus, the results of only one crop species and one eucalypt species are visually represented to ascertain the outputs of the project. A total of 77 SA2 boundaries that were predicted to have a plantation area 'greater than 0' hectares for Onions (*Allium Cepa*) were extracted from the SA2 layer from ABS. Furthermore, these boundaries were overlayed onto the IBRA subregions, resulting in 121 associated subregions majority of which were in south-eastern Australia with only a small percentage in Queensland and Western Australia. Similarly, the IBRA subregions for Gungurra (*Eucalyptus caesia*) were also identified by selecting the locations based on the point vector dataset of presence locations from ALA. 234 herbarium collections for Gungurra were found to occur in only 24 IBRA subregions with the species predominantly found in south-west Western Australia regions.



Map 1 – SA2 boundaries for Onions (*Allium Cepa*) identified in (A) before subsequently using an overlay to obtain the relevant IBRA subregions across Australia (B).



Map 2 – Point location dataset for Gungurru (*Eucalyptus caesia*) obtain from ALA (A) before subsequently using an overlay to obtain the relevant IBRA subregions across Australia (B).

Limitations

A number of limitations need to be addressed with the completion of this project. Firstly, the exclusion of Northern Territory IBRA subregions from this analysis is a disadvantage for beekeepers in that area. There are 51 subregions in Northern Territory and thus, an extensive number of possible locations have not been considered. Secondly, the ALA presence locations contain location uncertainties, environmental outliers and spatial quality issues which in some cases, account for a sizeable proportion of the datasets. This could even result in errors and inaccuracy when considering endangered flora or those species which are low in numbers. Apart from this, there is also the need to mention that the information retrieved from the multiple literature sources is framed differently in each case, and thus compiling them together into a systematic and structured database required certain assumptions and approximations. For example, only phenology periods from 'Flowers of Melbourne' webpage/blog were retrieved without the extent of the flowering intensity. Similarly, from Birtchnell and Gibson's 2006 paper, flowering time for each month was divided into three sections and intensity was rated on five different levels based on budding percentage. Therefore, the monthly flowering intensities were averaged across the three periods to find the optimal approximation.

Finally, a total of 56 crop species were identified according to the initial resources, however, six species' were not listed within the ABS predictions and a further 16 were categorised based on the vegetable and fruit groups shown in Table 2. Furthermore, the SA2 boundaries were limited to the accuracy of ABS predictions as relative standard errors (RSE) above 50% were not considered suitable for general use. Another limitation with regards to crops is the prediction used from 2015-16 financial year which is the earliest available dataset for the SA2 boundaries.

Table 2 – Limitations and assumptions based on ABS Agricultural Commodities data for 22 crop species.

Common Name	**Scientific Name**	**General Category Assumption**
Parsnip	*Trachymene species*	All other vegetables for human consumption
Mediterranean Turnip	*Brassica tournefortii*	
Raddish	*Raphanus sativus*	
Cucumber	*Cucumis sativus*	
Squash	*Cucurbita pepo*	
Soyabean	*Glycine max*	All other oil seeds (broadacre – non-cereal crops)
Sunflower	*Helianthus annuus*	
Common Vetch	*Vicia sativa*	Other pulses
Hairy Vetch	*Vicia villosa*	
Hungarian Vetch	*Vicia pannonica*	
Lychee	*Litchi chinensis*	Plantation and other fruits
Logans	*Dimocarpus longan*	
Cumquats	*Citrus fortunella*	Other citrus fruits
Tangelo	*Citrus tangelo*	
Currants	*Ribes species*	Other berries
Carambola	*Averrhoa carambola*	Other orchard fruit
Passionfruit	*Passiflora edulis*	Unlisted
Brazilian Pepper Tree	*Schinus terebinthifolius*	
Indian Hemp	*Cannabis sativa*	
Horehound	*Marrubium vulgare*	
Coffee	*Coffea arabica*	
Grammas	*Cucurbita moschata*	

Future Research

Although, the initial aim of this project was linking herbarium sample collections of melliferous flora to IBRA subregions across Australia, additional analysis for the same is required. This includes researching supplementary resources to include other flora species at the first instance. Moreover, the inclusion of Northern Territory IBRA subregions in the analysis will give sufficient direction to beekeepers in that area as well. This will also broaden the scope of the study accounting for all the 419 IBRA subregions. Crop species identifications were limited to ABS Agricultural Commodities data from 2015-16. Although, using a more recent estimation of crop species estimation will increase the overall accuracy, it also may be worthwhile in identifying the IBRA subregions based on the land use data on the catchment scale from the Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES). The intensive land use can be used to determine agricultural production at a sharper scale which is produced through fine-scale satellite imagery. Furthermore, visualising this database is considered as the next step in this project which can be done based on a variety of diversity indices. A measure of richness, that counts the different number of species in each IBRA subregion is considered an important tool along with evenness, such as Shannon's Diversity Index or Simpson's Diversity Index, that considers the abundance of each species as well. These indices may provide significant measure of spatial variability of each species taking into account the contrasting underlying environmental variables.

References

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

Benecke, F. S. (2007). *Commercial beekeeping in Australia*. Rural Industries Research and Development Corporation.

Birtchnell, M. J., & Gibson, M. (2006). Long-term flowering patterns of melliferous Eucalyptus (Myrtaceae) species. *Australian Journal of Botany*, *54*(8), 745-754.

Crane, E. (1999). *The world history of beekeeping and honey hunting*. Routledge.

DAWE. (2021). *Australia's bioregions (IBRA)* Australian Government. Retrieved 18th May 2021 from https://www.environment.gov.au/land/nrs/science/ibra

Delaplane, K. S., Mayer, D. R., & Mayer, D. F. (2000). *Crop pollination by bees*. Cabi.

Dixon, D. J., Callow, J. N., Duncan, J. M., Setterfield, S. A., & Pauli, N. (2021). Satellite prediction of forest flowering phenology. *Remote Sensing of Environment*, *255*, 112197.

Gordon, J., & Davis, L. (2003). Valuing honeybee pollination. *Rural Industries Research and Development Corporation, Canberra*, *36*.

Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society. B, Biological sciences*, *274*(1608), 303-313. https://doi.org/10.1098/rspb.2006.3721

Law, B., Mackowski, C., Schoer, L., & Tweedie, T. (2000). Flowering phenology of myrtaceous trees and their relation to climatic, environmental and disturbance variables in northern New South Wales. *Austral Ecology*, *25*(2), 160-178.

Lebuhn, G., Droege, S., Connor, E. F., Gemmill-Herren, B., Potts, S. G., Minckley, R. L., Griswold, T., Jean, R., Kula, E., & Roubik, D. W. (2013). Detecting insect pollinator declines on regional and global scales. *Conservation biology*, *27*(1), 113-120.

Moncur, M., Mitchell, A., Fripp, Y., & Kleinschmidt, G. (1995). The role of honey bees (Apis mellifera) in eucalypt and acacia seed production areas. *The Commonwealth Forestry Review*, 350-354.

Somerville, D. (2020). *Honey and Pollen Flora of South Eastern Australia*. Tocal College, NSW DPI.