University of Southern Queensland
Faculty of Health, Engineering and Sciences

Species Distribution Modelling of the Glossy Black Cockatoo in Queensland’s Condamine Region

A dissertation submitted by
Michael Walter John Evans

In fulfilment of the requirements of
ENG4111 and 4112 Research Project
Towards the degree of
Bachelor of Spatial Science (Honours) (Geographic Information Systems)
Submitted October, 2016
Abstract

This project undertakes species distribution modelling of the Glossy Black Cockatoo (*Calyptorhynchus lathami*). Species distribution modelling is a GIS application that has been used in a number of different studies. In this instance it will be used to predict and map the habitat suitability of areas across the Condamine River catchment in Queensland for the subject species and thus provide information valuable for developing conservation strategies.

The maximum entropy (Maxent) modelling software program was utilised in species distribution modelling. The data for the project has been sourced from a variety of spatial data custodians and then processed through the ArcGIS software to achieve the required data format for analysis. Two main types of datasets were required: the samples (sightings) data of the species, and the environmental variables that provide information to derive the prediction. These variables include; land use, DEM, slope, aspect, regional ecosystems, roads and drainage data.

The modelling has produced a satisfactory and valuable set of results. The main output is a species distribution map, in which every area is assigned specific habitat suitability values for the cockatoo in the area. Another important result is the contribution made by each variable to the final model. In this project land use (46.2%) and elevation (34.9%) were the most important variables in the model, while aspect (1.3%) was the most inconsequential.

Conducting this research has opened up avenues for further work such as expanding the scope to alternate species or different areas. Doing so would further assist in preventing the demise of an endangered species. Doing further research into the specific characteristics of the identified areas of high suitability would also be enlightening.

The project has been successful in indentifying suitable habitat for the Glossy Black Cockatoo. The information obtained from this study could be useful in future conservation efforts for this species.
University of Southern Queensland
Faculty of Health, Engineering and Sciences
ENG4111/ENG4112 Research Project

Limitations of Use

The Council of the University of Southern Queensland, its Faculty of Health, Engineering & Sciences, and the staff of the University of Southern Queensland, do not accept any responsibility for the truth, accuracy or completeness of material contained within or associated with this dissertation.

Persons using all or any part of this material do so at their own risk, and not at the risk of the Council of the University of Southern Queensland, its Faculty of Health, Engineering & Sciences or the staff of the University of Southern Queensland.

This dissertation reports an educational exercise and has no purpose or validity beyond this exercise. The sole purpose of the course pair entitled “Research Project” is to contribute to the overall education within the student’s chosen degree program. This document, the associated hardware, software, drawings, and other material set out in the associated appendices should not be used for any other purpose: if they are so used, it is entirely at the risk of the user.
University of Southern Queensland
Faculty of Health, Engineering and Sciences
ENG4111/ENG4112 Research Project

Certification of Dissertation

I certify that the ideas, designs and experimental work, results, analyses and conclusions set out in this dissertation are entirely my own effort, except where otherwise indicated and acknowledged.

I further certify that the work is original and has not been previously submitted for assessment in any other course or institution, except where specifically stated.

Michael Evans
Student Number: 0061046466
Acknowledgments

I would like to thank Professor Armando Apan for all his valuable assistance and guidance throughout the undertaking of this project and also to my family for their support and encouragement throughout this project and all my years at university.
## Contents

Abstract .............................................................................................................................................. i
Acknowledgments ............................................................................................................................... iv
List of Figures ........................................................................................................................................ vii
List of Tables ........................................................................................................................................ viii

Chapter 1 - Introduction ....................................................................................................................... 1

1.1 Introduction ................................................................................................................................... 1
1.2 Statement of the Problem ............................................................................................................... 2
1.3 Significance of the Study ............................................................................................................... 2
1.4 Aims and Objectives ..................................................................................................................... 3
1.5 Scope and Limitation of the Study ................................................................................................. 4
1.6 The Organisation of the Dissertation ............................................................................................ 5

Chapter 2 - Literature Review .............................................................................................................. 6

2.1 Introduction ................................................................................................................................... 6
2.2 Species Distribution Modelling .................................................................................................... 6

2.2.1 Overview .................................................................................................................................. 6
2.2.2 Development and Processes ..................................................................................................... 7
2.2.3 Types and Caveats .................................................................................................................... 7

2.3 Glossy Black Cockatoo .................................................................................................................. 8

2.3.1 Overview .................................................................................................................................. 8
2.3.2 Biology, Ecology, Identification and Distribution ....................................................................... 8
2.3.3 Prior Studies .............................................................................................................................. 10

2.4 MaxEnt ......................................................................................................................................... 11

2.4.1 Overview .................................................................................................................................. 11
2.4.2 Advantages and Disadvantages ............................................................................................... 12
2.4.3 Case Studies of MaxEnt Applications ....................................................................................... 12

2.5 Summary ....................................................................................................................................... 17

Chapter 3 - Research Methods ............................................................................................................. 19

3.1 Introduction ................................................................................................................................... 19
3.2 The Study Area .............................................................................................................................. 20
3.3 Data Acquisition and Pre-processing ............................................................................................ 23
3.4 Data Analysis ................................................................................................................................. 28

3.4.1 Overview .................................................................................................................................. 28
3.4.2 Preparation for Analysis ........................................................................................................... 28
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.3 Environmental Variables</td>
<td>29</td>
</tr>
<tr>
<td>3.4.4 MaxEnt Specifics and Settings</td>
<td>30</td>
</tr>
<tr>
<td>3.4.5 Output</td>
<td>32</td>
</tr>
<tr>
<td>3.4.6 Further Analysis</td>
<td>32</td>
</tr>
<tr>
<td>3.6 Summary</td>
<td>32</td>
</tr>
<tr>
<td>Chapter 4 - Results</td>
<td>33</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>33</td>
</tr>
<tr>
<td>4.2 Results of Subject</td>
<td>33</td>
</tr>
<tr>
<td>4.2.1 Pictures of the Model</td>
<td>33</td>
</tr>
<tr>
<td>4.2.2 Response Curves</td>
<td>34</td>
</tr>
<tr>
<td>4.2.3 Analysis of Variable Contributions</td>
<td>36</td>
</tr>
<tr>
<td>4.3 Results of Model</td>
<td>38</td>
</tr>
<tr>
<td>4.3.1 Analysis of Omission/ Commission</td>
<td>38</td>
</tr>
<tr>
<td>4.4 Conclusion</td>
<td>39</td>
</tr>
<tr>
<td>Chapter 5 - Discussion</td>
<td>40</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>40</td>
</tr>
<tr>
<td>5.2 Accuracy of Model</td>
<td>40</td>
</tr>
<tr>
<td>5.3 Distribution Map</td>
<td>40</td>
</tr>
<tr>
<td>5.4 Variable Contributions</td>
<td>42</td>
</tr>
<tr>
<td>5.5 General Observations</td>
<td>45</td>
</tr>
<tr>
<td>5.6 Sensitivity Analysis</td>
<td>46</td>
</tr>
<tr>
<td>5.6.1 Overview</td>
<td>46</td>
</tr>
<tr>
<td>5.6.2 Process</td>
<td>46</td>
</tr>
<tr>
<td>5.6.3 Results</td>
<td>46</td>
</tr>
<tr>
<td>5.6.4 Limited Variables Model</td>
<td>48</td>
</tr>
<tr>
<td>5.7 Summary</td>
<td>49</td>
</tr>
<tr>
<td>Chapter 6 - Conclusions and Recommendations</td>
<td>50</td>
</tr>
<tr>
<td>6.1 Conclusions</td>
<td>50</td>
</tr>
<tr>
<td>6.2 Recommendations</td>
<td>51</td>
</tr>
<tr>
<td>References</td>
<td>53</td>
</tr>
<tr>
<td>Appendix A</td>
<td>56</td>
</tr>
<tr>
<td>Appendix B</td>
<td>57</td>
</tr>
<tr>
<td>Appendix C</td>
<td>65</td>
</tr>
<tr>
<td>Appendix D</td>
<td>73</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1-1 The Glossy Black Cockatoo (Glossy Black Conservancy 2016) ...................... 1
Figure 1-2 Glossy Black Cockatoo (Glossy Black Conservancy 2016) ......................... 5
Figure 2-1 Diagram illustrating the basic process of species distribution modelling (University of Lausanne 2016) ................................................................. 7
Figure 2-2 Image showing the records for the Glossy Black Cockatoo (Atlas of Living Australia 2016) ................................................................................................. 9
Figure 2-3 Casuarina trees, the Glossy Black Cockatoo's preferred habitat (Glossy Black Conservancy 2016) .............................................................. 9
Figure 2-4 Casuarina Seed Cones, the birds preferred food source (Glossy Black Conservancy 2016) .............................................................. 10
Figure 2-5 Glossy Black Cockatoos, notice the yellow on the bird on the right (Glossy Black Conservancy 2016) .............................................................. 10
Figure 2-6 The strength of association (as test area under curve) between each species' presence and individual environmental variables at different spatial scales (Bellamy, Scott & Altringham 2013) ................................................................. 14
Figure 2-7 Representative MaxEnt response curves showing the probability of a species' presence at a location for a range of parameters (Bellamy, Scott & Altringham 2013) ................................................................. 14
Figure 2-8 Habitat suitability maps made using each species' pruned set of variables (Bellamy, Scott & Altringham 2013) ................................................................. 15
Figure 2-9 MaxEnt results of Model A (Aguilar, Farnworth & Winder 2015) .................. 16
Figure 2-10 MaxEnt results of Model B (Aguilar, Farnworth & Winder 2015) ............. 17
Figure 3-1 Identification of the study area ................................................................. 20
Figure 3-2 View in Main Range National Park ......................................................... 21
Figure 3-3 Forest area in Main Range National Park ................................................. 21
Figure 3-4 Casuarina trees near Toowoomba ............................................................. 22
Figure 3-5 View at the Bunya Mountains National Park .......................................... 22
Figure 3-6 Flowchart illustrating the steps involved with data processing ............... 24
Figure 3-7 Flowchart showing the processing steps in ArcGIS .................................. 24
Figure 3-8 Datasets 1 .................................................................................................. 26
Figure 3-9 Datasets 2 .................................................................................................. 27
Figure 3-10 Main MaxEnt settings ............................................................................ 31
Figure 3-11 Additional MaxEnt settings .................................................................... 31
List of Figures

Figure 4-1 Species Distribution Map ................................................................. 34
Figure 4-2 Response curve for Aspect variable ................................................. 35
Figure 4-3 Response curve for Land Use variable .............................................. 35
Figure 4-4 Variable contribution table .............................................................. 36
Figure 4-5 Jackknife Test 1 ............................................................................. 36
Figure 4-6 Jackknife test 2 .............................................................................. 37
Figure 4-7 Jackknife test 3 .............................................................................. 37
Figure 4-8 Omission rate graph ....................................................................... 38
Figure 4-9 AUC graph ..................................................................................... 39
Figure 5-1 Protected Areas in Condamine Region ........................................... 41
Figure 5-2 DEM with correlated area highlighted ........................................... 42
Figure 5-3 Drainage map showing area of interest .......................................... 43
Figure 5-4 Land Use map highlighting key areas ............................................. 44
Figure 5-5 Slope data with notable areas marked .......................................... 45
Figure 5-6 Species Distribution Map created during sensitivity analysis ...... 47
Figure 5-7 Sensitivity analysis variable contribution table ............................ 48
Figure 5-8 Comparison of SDMs created using all seven variables (Left) and just two variables (Right). ................................................................. 48

List of Tables

Table 2-1 Advantages and disadvantages of using Maxent ................................ 12
Table 3-1 Dataset Properties ......................................................................... 24
Table 3-2 List of primary and derived datasets used in this study ................. 25
Table 5-1 Comparison of AUC values .............................................................. 47
Chapter 1 - Introduction

1.1 Introduction

In a country, such as Australia, with its vast array of wildlife it is important that species are preserved and their well being and habitats looked out for. Conservation efforts are an important activity in today’s society and there are many strategies in place to help protect endangered species (Lindenmayer & Hobbs 2004). The Glossy Black Cockatoo (*Calyptorhynchus lathami*) is just one species present in the Condamine region in Queensland’s south-east. This species is identified as being vulnerable by Queensland’s Department of Environment and Heritage Protection (2015). This species has been chosen as the subject of this research project where its ideal habitat characteristics will be determined through the use of GIS related technologies.

![The Glossy Black Cockatoo](image)

*Figure 1-1 The Glossy Black Cockatoo (Glossy Black Conservancy 2016)*

The research and analysis will take the form of Species Distribution Modelling. Species Distribution Modelling (SDM) is a common form of GIS analysis. This modelling involves the collection and systematic analysis of location data for a certain species be it flora or fauna. This location data is analysed based on environmental variables. A model is created which draws parallels between the observed locations of the species and the makeup of their environment. Then it makes predictions as to the viability of areas as habitat for the species. For this study the modelling software MaxEnt will be used to make this prediction.
Species Distribution Models are developed in order to reveal important statistics and information on the study object allowing for a greater understanding of its behaviour and role in the ecosystem. These models provide information integral to conservation efforts of endangered species (Liu et al. 2013).

Because Species Distribution Modelling is a common and already established GIS application, using it as the basis of a research topic allows the research to have a manageable goal but still provide new and important results that could help save a threatened species. Elaborating on this point, keeping the project within the scope of a major area of GIS will avoid the research from becoming too complicated and unachievable while still providing satisfying and unique results to add to the knowledge bank of professionals and interested parties.

Species Distribution Modelling will be used to analyse the Glossy Black Cockatoo in the Condamine region. The result of which will produce a species distribution map which will identify the preferred habit locations for the species based upon the environmental characteristic of the area. The benefits of this study will be felt by those involved in conservation efforts and also professionals interested in the application of Species Distribution Modelling and MaxEnt.

1.2 Statement of the Problem
Species Distribution Modelling has been carried out by numerous parties studying a vast range of species in many portions of the globe (Booth 2014). This research aims to add to the wealth of information about certain species by providing important information about one more species in a specific area of Australia, thus filling a research gap.

The proposed study fills a niche in the scientific community by adding information on another species to the existing knowledge bank. The scope of this project has been set to ensure its feasibility and successful completion. By reducing this scope it is aimed to achieve a higher quality of information for what is studied. The range of the project is thus so it may be accomplished using readily accessible resources and data.

1.3 Significance of the Study
The studies significance will stem from its determination of the ideal habitat conditions for the Glossy Black Cockatoo. The study will be beneficial due to its compilation of information showing the behaviour and habit preferences of the subject species. This knowledge will be valuable in ensuring the species continued survival.
Chapter 1

Introduction

The research will contribute to the knowledge bank of GIS professionals. It will do this by adding an additional study on the application of Species Distribution Modelling and the software program MaxEnt for professional study. The more studies that exist which use these processes will aid in the future development of their technologies. This will allow consolidation of their effectiveness and help ensure the best results are being achieved in any application.

The study will be useful for those dealing with wildlife in the Condamine region and those concerned with fauna conservation. The well being of the environment and it its fauna are a matter of public social consciousness, therefore contributing information that will help preserve the species will be of interest those people. This study will help minimise the harm affecting the subject species as the information gained will be able to be used in conservation efforts.

The study will determine the species realised ecological niche in the subject area (Phillips, Anderson & Schapire 2006). This refers to the environmental characteristics most suitable to ensure the species continued survival. Species distribution modelling studies have been carried out on many species, some examples include those studies by Bellamy, Scott and Altringham (2013), Diao and Wang (2014) and Evans et al. (2011). The specifics and key concepts involved in these types of studies will be explained in the following chapter.

1.4 Aims and Objectives

The objectives of this research are to undertake species distribution modelling of the Glossy Black Cockatoo, in the Condamine Region of Queensland, to map its habitat suitability which will help in conservation efforts for the species. The full programme can be found in the project specification submitted as Appendix A.

The aim of the study’s output is to help assist with conservation practices pertaining to the Glossy Black Cockatoo in the Condamine region of Queensland. The benefits of this study may be useful for a broader range of species and areas than just the specific circumstances defined in this project. This is because some information obtained may be applicable in other circumstances. To keep the project within the bounds of what can be realistically accomplished with the time and resources available the project has a narrow, focussed scope to ensure the highest quality of what is accomplished.
Chapter 1

Introduction

It is aimed to derive information related to the habitat and behaviour of the subject species so the information may assist with the continued projection of the species and inform those with an interest in local wildlife or species distribution modelling techniques. The Condamine region is the study area for this task. It was chosen as it is near to where the study is being undertaken. Previous studies have been consulted to inform these aims and objectives and so keep the project in line with current practice in the industry (Merow, Smith & Silander 2013).

More specific objectives of the research are provided below:

- To review the literature related the Glossy Black Cockatoo, Species Distribution Modelling and other related concepts.
- To develop a habitat suitability model for the Glossy Black Cockatoo using appropriate environmental variables.
- To assess the importance of the environmental variables in the prediction of habitat suitability for the Glossy Black Cockatoo making evaluations of the achieved results.

1.5 Scope and Limitation of the Study

The project topic had to be decided on with respect to the time constraints applied by the course. The project needed to have a manageable and achievable goal that could be completed in the allotted timeframe. This impacted on the scope of the study, particularly in terms of the study area chosen and the subject species. Therefore, the study was limited to the Condamine River catchment and solely the Glossy Black Cockatoo within it.

The project was also restricted to the data that was available. The project was carried out using pre-existing and freely available datasets. No primary data was specifically obtained for this project. This impacted the number and specificity of environmental variables which could be used. However, due to the wealth of data available on the various spatial data infrastructures this was not a great impediment to the completion of the research.

Despite the many modelling tools available, only MaxEnt was employed for this project. MaxEnt is a tool that has been specifically designed for Species Distribution Modelling application and is one of the most preferred and trusted. MaxEnt was decided to be the ideal tool to use for this study.
1.6 The Organisation of the Dissertation

The following dissertation is presented according to the standard conventions of this format of writing. Firstly a review of the literature relevant to this project will be presented and the key concepts will be explained. The third chapter will begin to detail the specifics of this project, such as the formation of the topic. The study area and the data preparation involved will also be explained along with the analysis methods used. Results will be presented in the chapter after that. This will mainly include the output from the software that performed the analysis. These results will be interpreted and discussed in the next chapter with the key findings being outlined. The dissertation will close with the conclusions and recommendations derived from conducting this research. In order for the reader to obtain a clearer picture as to the nature of this project, descriptions of its major aspects are presented in the succeeding chapter.

![Glossy Black Cockatoo](Glossy Black Conservancy 2016)

*Figure 1-2 Glossy Black Cockatoo (Glossy Black Conservancy 2016)*
Chapter 2 - Literature Review

2.1 Introduction
As per the project objectives, research on the key areas of the topic was undertaken. The focus of the literature review was on similar studies that may have been carried out on different subjects in the past. This was done to provide an understanding of how others have carried out Species Distribution Modelling in the past. The review also gave information on technologies and procedures that will be used during the research, so the research may be carried out as efficiently as possible and to the highest achievable standard. The review of existing information highlighted three key concepts or areas related to the project:

- Species Distribution Modelling,
- Maxent, and
- The Glossy Black Cockatoo.

Important information found on these topics are summarised below.

2.2 Species Distribution Modelling

2.2.1 Overview
The utilisation of species distribution models is becoming greater in scientific literature and public outreach products. Their capacity to display locations, abundance and suitable environmental conditions result in their immense value in management and conservation, research and ecological forecasting of harmful invasive species, threatened and endangered species, or species of special concern (Jarnevich et al. 2015). The utilisation of species distribution modelling has increased dramatically in recent years (Booth 2014). GIS and species distribution mapping techniques can be used to: evaluate and assess current distributions; identify suitable habitat locations; and use models to project future scenarios (Aguilar, Farnworth & Winder 2015). Species distribution modelling is a widely used approach that is used to generate maps indicating suitable areas for a particular species. The software tools and algorithms involved only continue to improve as the form of study continues to be undertaken (Aguilar, Farnworth & Winder 2015). The modelling can use many different variables
using different geographic and temporal scales and environmental, socio-economic and non-biological layers (Booth 2014).

2.2.2 Development and Processes
A study by Booth (2014) delved into the beginnings of species distribution modelling and how it evolved to become a key method of analysis. The process that would eventually be termed species distribution modelling has developed since the 1980’s where it was aimed to find a reliable way to estimate mean climatic conditions, in conjunction with this, methods of analysing flora and fauna developed also (Booth 2014). The general procedure of SDM unfolds as follows: geocoded information is used by software packages to relate the data to the real world and map likely species distributions, this process may be done under current or future conditions (Booth 2014). The following figure outlines the basic process of species distribution modelling.

![Diagram illustrating the basic process of species distribution modelling](University of Lausanne 2016)

Species and environmental data are used to create distribution prediction based on statistical models. SDM is important for biogeography and phylogeography applications (Iturbide et al. 2015). Species distribution models can be defined as the methods used to relate known locations of biological entities with the makeup of their environment to then predict and map potential distributions of the entity in the geographic space (Guisan & Zimmermann 2000).

2.2.3 Types and Caveats
Iturbide et al. (2015) identifies two types of SDM building techniques; profile and group discrimination. Profile techniques use only known presence data, while group
discrimination utilises information on where the species does not occur, this is called absence data (Iturbide et al. 2015). Group discrimination techniques are said to provide better results (Iturbide et al. 2015). There are a number of different methods of constructing SDM’s, including maximum entropy, generalised linear models and multivariate adaptive regression splines (Iturbide et al. 2015).

Jarnevich et al. (2015) puts forth four important caveats for species distribution modelling which will be important to consider when conducting research in this area. The caveats are:

- All sampling data are incomplete and potentially biased,
- Predictor variables must capture distribution constraints,
- No single model works best for all species in all areas at all spatial scales and over time, and
- The results of species distribution models should be treated like a hypothesis to be tested and validated with additional sampling and modelling in an iterative process.

Species distribution modelling is an efficient tool for analysing species data.

### 2.3 Glossy Black Cockatoo

#### 2.3.1 Overview

The subject of this research is the Glossy Black Cockatoo. While the exact nature of this bird is not relevant to the technical logistics of the project it seems necessary to give a brief description of the animal, its characteristics and what studies have been done on it previously. The focus of the research is to show the importance of habitat mapping of the species for conservation planning. MaxEnt will be the tool used to conduct this mapping.

#### 2.3.2 Biology, Ecology, Identification and Distribution

The Glossy Black Cockatoo is found solely in the east of Australia mainly in southeast Queensland down through northern New South Wales and also on Kangaroo Island in South Australia. The South Australian mainland population has become extinct due to land clearing for urban developments (Glossy Black Conservancy 2010). The bird lives in pairs or family groups, together raising the single egg produced in each breeding season once a year (Dubbo Field Naturalist & Conservation Society 2010). The accompanying
figure is taken from the Atlas of Living Australia. It shows the locations of all the Glossy Black Cockatoo sightings in Australia. This is presented just to show the extent of its habitat.

![Occurrence records map](image)

**Figure 2-2 Image showing the records for the Glossy Black Cockatoo (Atlas of Living Australia 2016)**

Casuarina trees are the bird’s dominant food and nesting source. They often nest in these hollow dead trees. Hence why habitat destruction is of critical importance to the species because such trees are often cleared as their significance is not realised (Dubbo Field Naturalist & Conservation Society 2010). The bird feeds only on the seed cones of the Casuarina trees hence the ease in which the species could become seriously endangered (Glossy Black Conservancy 2010). The following figures show some Casuarina trees and their seed cones.

![Casuarina trees](image)

**Figure 2-3 Casuarina trees, the Glossy Black Cockatoo's preferred habitat (Glossy Black Conservancy 2016)**
Glossy Blacks are smaller than other Black Cockatoo species. They are a brownish black with a small crest. Males have red coloured tail feathers, females black. The females head and neck also feature yellow spots (Dubbo Field Naturalist & Conservation Society 2010).

2.3.3 Prior Studies
Cameron (2006) conducted a study of the nesting ecology of a population of Glossy Black Cockatoo’s in central New South Wales. This study highlighted a number of points that, while they are made about the situation in New South Wales, could still impact this study. The Glossy Black Cockatoo resides in woodlands and open forests of eastern
Australia and is known by the scientific name *Calyptorhynchus lathami* (Higgins 1999). As the bird nests in the hollows of eucalypt trees it has seen much of its habitat destroyed though clearing efforts since the advent of European settlers (Benson & Redpath 1997). This destruction of habitat has caused populations to decline over the years (Newton 1994). The study puts forth the conclusion that suitable nesting sites are key to the species continual survival (Cameron 2006). While the focus of this study was quite different to the one at hand, there is still insight to be sought from its findings.

### 2.4 MaxEnt

#### 2.4.1 Overview

The Maxent software has been used to model and predict the distribution of a wide range of species and organisms. These include: invasive species, endangered and threatened flora and fauna, organisms of economic significance and ancient species (Aguilar, Farnworth & Winder 2015). Phillips, Anderson and Schapire (2006) describe Maxent as a general-purpose method with a simple and precise mathematical formulation that allows predictions and inferences to be made from incomplete information. The method comes with a number of features that make it suitable for species distribution modelling. Maxent is able to estimate target probability distributions by utilising maximum entropy. This procedure is carried out dependant on certain constraints imposed by the data available (Phillips, Anderson & Schapire 2006). MaxEnt calculates the probability that the species is present based upon the environment (Phillips, Anderson & Schapire 2006).

Three key components of species distribution modelling are an ecological model, a data model and a statistical model. Maxent is an example of the latter (Phillips, Anderson & Schapire 2006). Maxent uses the maximum entropy principle to approximate unknown probability distributions (Phillips, Anderson & Schapire 2006). Maxent is a widely used tool to aid in species distribution modelling.
2.4.2 Advantages and Disadvantages

Phillips, Anderson and Schapire (2006) identifies some advantages and drawbacks of Maxent, they are outlined in the following table.

Table 2-1 Advantages and disadvantages of using Maxent

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requires only presence data, with environmental information</td>
<td>Not a mature statistical method, so there are fewer guidelines</td>
</tr>
<tr>
<td>Can utilise continuous and categorical data, and incorporate interactions between variables</td>
<td>Further study is required with regards to regularisation and over-fitting</td>
</tr>
<tr>
<td>Algorithms have been developed to guarantee optimal probability distribution</td>
<td>Maxent using an exponential model for probabilities</td>
</tr>
<tr>
<td>Maxent probability distribution is well defined which aids in analysis</td>
<td>Special-purpose software is required</td>
</tr>
<tr>
<td>Over-fitting can be avoided by using regularisation techniques</td>
<td>Reliant upon quality of sample data</td>
</tr>
<tr>
<td>Due to the explicit nature of the Maxent probability distribution there is the capacity for sampling bias to be addressed in the future</td>
<td></td>
</tr>
<tr>
<td>The output is continuous which allows for more detailed analysis</td>
<td></td>
</tr>
<tr>
<td>Maxent could be applied to species presence/absence data using a conditional model</td>
<td></td>
</tr>
<tr>
<td>Maxent is a generative approach not discriminative</td>
<td></td>
</tr>
<tr>
<td>Maximum entropy modelling is an ongoing research area</td>
<td></td>
</tr>
<tr>
<td>Maxent’s flexibility allows it to be used for many applications</td>
<td></td>
</tr>
</tbody>
</table>

2.4.3 Case Studies of MaxEnt Applications

The following section presents two case studies of species distribution modelling using MaxEnt. These are provided to give the reader an idea of the main processes involved in this research and also to highlight some of the many different forms this type of research can take. While not identical to the current study, there are many similarities between the methodologies and aims of these studies that it is still a valuable exercise to become familiar with what has already been accomplished.
2.4.3.1 Case Study 1 - Multiscale, presence-only habitat suitability models: fine-resolution maps for eight bat species

One example of using MaxEnt for creating habitat suitability models is in the paper by Bellamy, Scott and Altringham (2013). In this instance different species of bat in the UK were the subject for which the models were created. It was aimed to determine whether the models were capable of contributing to habitat management decision-making. Eight different species were included in this study and multiple maps were produced at different scales in contrast to the present study which only uses one species.

In this study MaxEnt was used to determine the significance of each variable. It was found that the best models had fewer variables which made them of more use to conservation planning. The authors also made the conclusion that the approach was relevant for any species provided the appropriate presence data was available. The present study will aim to confirm this. The maps produced were able to identify areas of conservation concern allowing assessment of environmental change and providing conservation managers with valuable information for habitat creation and improvement plans. Many of the processes used and results achieved in this study are similar to what is to be achieved in this present study.

As the study used MaxEnt similar techniques were used to prepare the data in the same vein as the present study. This means that the environmental data was collected and processed using ArcGIS and put into a raster format because this is what is required by MaxEnt. These variables were then run through the software along with the species records to create the models. The default settings for MaxEnt were used.

The results achieved included AUC graphs which show the predictive power of the environmental variables. Graphs were also created showing the probability of species presence based upon the environmental variables. Habitat suitability maps were also created for all species. Excerpts from the results of this paper are presented below.
Figure 2-6 The strength of association (as test area under curve) between each species' presence and individual environmental variables at different spatial scales (Bellamy, Scott & Altringham 2013)

Figure 2-7 Representative MaxEnt response curves showing the probability of a species' presence at a location for a range of parameters (Bellamy, Scott & Altringham 2013)
2.4.3.2 Case Study 2 - Mapping the stray domestic cat (Felis catus) population in New Zealand: Species distribution modelling with a climate change scenario and implications for protected areas

A second case study is that of a New Zealand study on stray cat populations in the country by Aguilar, Farnworth and Winder (2015). This study used two different types of occurrence data along with the environmental variables, which included climate elements, to produce the suitability maps. Unlike the present study this study looked at a species that is considered a nuisance with the aim being to try and manage the un-owned cat populations to help prevent the spread of diseases, to prevent harm to native fauna and to address the welfare concerns of some persons. The study used GIS and species distribution modelling to:

- Evaluate and assess current un-owned cat distributions,
Identify natural areas that are proximal to urban areas to evaluate colonisation risk, and

Build future scenarios using climate models.

In addition to analysing the current situation this study also investigated possible future distributions by way of factoring in climate change.

ArcGIS and MaxEnt were used to process the data just as it is in the present study. Of the two models produced the first used location data solely collected from Auckland that was then projected to the entire country. The second model used human population data to act as an equivalent for the cat presence data along with the Auckland data. This was done as it was thought that stray populations are often centred round urban areas. This theory was based on prior research (Aguilar & Farnworth 2012). The models used Bioclim data, which is a dataset featuring numerous climatic variables.

The results of the study showed that the North Island had more favourable conditions for the cat populations with both islands showing increased suitability in the future scenario. The study also found which protected environmental areas would be most susceptible to the rise of stray cat populations. The following figures showcase some of the obtained results from this study.

Figure 2-9 MaxEnt results of Model A (Aguilar, Farnworth & Winder 2015)
This study used a number of different methods to the current project, namely the extrapolated location data and the climate information used for predicting future situations. Despite these contrasts the heart of the study is still similar and just showcases the many options available to researchers using MaxEnt and species distribution modelling techniques.

2.5 Summary

This information has highlighted the fact that studies of this nature are continuing to be carried and that such things are encouraged. The objectives of species distribution modelling keep in line with the intended objectives of the project itself. The concepts and tools involved in species distribution modelling are quite vast, so the appropriate methods that are most relevant to this study will have to be selected.

The principles of Maxent have been outlined to give a baseline understanding of the concepts involved and how that software is ideal for undertaking this form of modelling. While other software and methods are available and do exist, Maxent is the ideal choice for this study. It was also the software recommended by the supervisor. There is a growing interest in the modelling of species, and using Maxent allows for consistency across various studies.
A brief look at information on the species in question shows that the results of this study would be beneficial as the bird has not been fully studied before. This study would be one of the first of its type to focus on the Condamine region.

The major purpose of this presentation of information was to give some background on the major concepts and ideas present in the research in order for one to familiarise themselves with these issues to make understanding the research more straightforward.
Chapter 3 - Research Methods

3.1 Introduction
Species distribution modelling has been undertaken using data collected from a variety of sources. These sources include; QGIS, Atlas of Living Australia and other spatial data sites. This raw data was processed using the GIS software ArcGIS. The processed data was then analysed using the MaxEnt software and the findings determined and reported on. A brief overview of the steps involved in carrying out the research is as follows.

1. Conducting a literature review. The aim of this was to assess the current research carried out in the relevant area. Conducting a literature review allowed one to become further acquainted with the research area and showed how it could be built upon. The review also provided vital guidance in certain concepts and procedures that may be relevant to the task.

2. Collection of raw data. Data was collected from a number of sources to be used for a number of applications. Occurrence data in point form was the input of the species. Other variables and data layers were obtained which were used to inform the study. This data had already been collected and simply needed to be sourced from freely accessible spatial data infrastructures.

3. Processing of Data. The obtained data was processed through methods of converting formats, deriving additional data layers and all other steps required to prepare it for analysis. It was at this stage that the variables, against which the species data was to be compared, were determined. In this study the important variables included; road networks, drainage patterns, land use types and various derivatives of a digital elevation model.

4. Analysis of data. The processed data was then analysed and the findings recorded. This was the stage where the main objectives of the study were met. The analysis was undertaken using Maxent which is a standard program in the industry for undertaking such tasks. The capabilities of Maxent were used to make justifiable conclusions based upon the input data which had been prepared in previous phases of the research project.
3.2 The Study Area

The study area for the project is the Condamine River Catchment in South East Queensland. The region was named for the river which runs through the area which has predominately been used for agricultural pursuits due to the high quality land available. The extent of the region is defined by the catchment area of the Condamine River, a significant waterway which eventually connects with the Murray-Darling River System in the South of Australia. The principal urban centres encompassed by the region are Toowoomba, Warwick, Dalby and Chinchilla (Condamine Alliance 2016). This area was chosen because of its proximity to where the research is being conducted. As such this study area was selected before the subject species. The following figure identifies the study area.

![Figure 3-1 Identification of the study area](image)

The following figures are photos taken at various places around the study area which are presented to give one an idea of the landscape of the region. These are personal photographs of the researcher.
Figure 3-2 View in Main Range National Park

Figure 3-3 Forest area in Main Range National Park
Chapter 3  
Research Methods

Figure 3-4 Casuarina trees near Toowoomba

Figure 3-5 View at the Bunya Mountains National Park
3.3 Data Acquisition and Pre-processing

During research the task of data acquisition and pre-processing were collectively referred to as the data phase. The major tasks of the Data phase were:

- Identification of datasets required,
- Acquisition of said datasets, and
- Pre-processing of data.

The datasets required for the project were determined by considering the nature of the project, its aims and objectives and what was to be achieved by it. The data was selected in order to provide the greatest array of variables which would achieve the best results. The data was sourced from two main locations:

- The Queensland Spatial Catalogue – A public access Spatial Data Infrastructure provided by the Queensland Government and managed by the Department of Natural Resources and Mines.
- The Atlas of Living Australia – An Australian Government initiative providing biodiversity data on all Australian species.

The principal datasets obtained are as follows:

- Presence data of bird
- Condamine boundary
- Queensland Land Use Map
- Queensland Digital Elevation Model (DEM)
- Drainage
- Regional Ecosystems
- Roads

The data phase was primarily concerned with taking the data from its raw form to one that is suitable for analysis. The major analysis involved MaxEnt and so steps needed to be taken to ensure that the data would work properly with the software. This involved making all the datasets properties uniform. The utilised values are shown in the following table.
Chapter 3

Research Methods

Table 3-1 Dataset Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projected Coordinate System</td>
<td>Albers Equal Area Qld</td>
</tr>
<tr>
<td>Geographic Coordinate System</td>
<td>GCS GDA 1994</td>
</tr>
<tr>
<td>Spatial Reference</td>
<td>GDA 1994 MGA Zone 56</td>
</tr>
<tr>
<td>Data Format</td>
<td>Raster</td>
</tr>
<tr>
<td>Cell Size</td>
<td>30, 30m</td>
</tr>
<tr>
<td>Extent</td>
<td>Condamine Boundary</td>
</tr>
</tbody>
</table>

These specific properties were chosen for the following reasons. The Condamine Boundary was used as a template and so all layers were clipped to its extent and put into its projection. The raster format is required by MaxEnt and all layers needed to be in the same data format. In selecting a cell size a number of factors needed to be considered. As the area is quite large a very high resolution is impractical, but a decent level of detail is still required. The cell sizes of the raw data also impact this as resolution cannot be increased. The data was prepared in ArcMap using the software’s many spatial tools. The properties were selected that were most suitable to the task at hand.

The following flowchart outlines the process which the data takes. From downloading it from the appropriate spatial data infrastructure (SDI), processing it in ArcGIS, inputting it for analysis in MaxEnt and then obtaining the output results.

![Flowchart illustrating the steps involved with data processing](image)

*Figure 3-6 Flowchart illustrating the steps involved with data processing*

This second chart further details the ArcGIS processing and the steps and tools required to prepare the data for use with MaxEnt.

![Flowchart showing the processing steps in ArcGIS](image)

*Figure 3-7 Flowchart showing the processing steps in ArcGIS*
Once the environmental variable datasets were processed they were converted to ASCII format, which is the format required by Maxent. This task was simply done using the appropriate tool in ArcGIS.

The samples data containing the point location information on the bird had to be processed somewhat separately from the rest of the datasets due to its different structure. In order to make the samples data compatible with the other layers, the raw, downloaded csv file was displayed in ArcGIS and reprojected to the appropriate coordinate system. The format of the points was also changed from latitude and longitude to an x and y point.

Below a table describing the datasets that have been used in this project is presented. Then the following pages present maps of these processed datasets. As these layers have been designed to run with the computer software some of them are not intuitively interpretable simply by visual means. Note that town names have been added to the samples map.

Table 3-2 List of primary and derived datasets used in this study.

<table>
<thead>
<tr>
<th>Spatial Dataset</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species Samples Data</td>
<td>The data, in point form, identifying the individual sightings for the bird in the region. There were 162 records.</td>
<td>Atlas of Living Australia</td>
</tr>
<tr>
<td>Digital Elevation Model (DEM)</td>
<td>Elevation data for region.</td>
<td>Queensland Government</td>
</tr>
<tr>
<td>Drainage</td>
<td>Water course locations, expressed as distance from.</td>
<td>Queensland Government</td>
</tr>
<tr>
<td>Land Use</td>
<td>Land Use types in region, from Queensland Land Use Map (QLUMP).</td>
<td>Queensland Government</td>
</tr>
<tr>
<td>Regional Ecosystems</td>
<td>Data on geology, landform and soil.</td>
<td>Queensland Government</td>
</tr>
<tr>
<td>Roads</td>
<td>Road locations, expressed as distance from.</td>
<td>Queensland Government</td>
</tr>
<tr>
<td>Aspect</td>
<td>Orientation of slopes.</td>
<td>Derived from DEM</td>
</tr>
<tr>
<td>Slope</td>
<td>Magnitude of slopes.</td>
<td>Derived from DEM</td>
</tr>
</tbody>
</table>
Figure 3-8 Datasets 1
Figure 3-9 Datasets 2
3.4 Data Analysis

3.4.1 Overview
The second and most critical phase which formed the crux of the research involved the actual analysis of the data that was previously obtained and processed. This component of the research was termed the analysis phase. Tasks of the Analysis Phase included carrying out the analysis using Maxent and recording the results. These results were then used to inform whether additional analysis was required to achieve an appropriate result. This would be done by means of collecting additional data or carrying out certain procedures once more perhaps in the form of sensitivity analysis. As the main part of the research, aspects of this phase could have been repeated until the optimal results were produced. Before the analysis was run, further research into the capabilities and settings of MaxEnt was conducted to ensure a firm grasp was held on the critical concepts.

3.4.2 Preparation for Analysis
Tutorials on the MaxEnt software were undertaken and relevant papers that use the software for a similar purpose were read in order to gain an understanding of how to use the software in preparation for the analysis portion of the project. This was important due to the myriad ways of using MaxEnt and to ensure accurate results were obtained and able to be synthesised appropriately.

The specifics of this phase can be found in the relevant literature which has been investigated in the literature review. Many articles have been published concerning species distribution modelling and MaxEnt and so the methodology of this project has been derived from these previous studies. MaxEnt is used to model species distributions using presence only data (Phillips, Anderson & Schapire 2006). This means that the software calculates the probability of occurrence in areas where there is no presence data. For this project the presence data of the Glossy Black Cockatoo has been sourced from the Atlas of Living Australia. In addition to this occurrence data environmental variables need to be carefully selected as they provide the other major part of the analysis.
3.4.3 Environmental Variables

Ideal environmental variables are critical to species distribution modelling because they have a great affect on the final result. These variables need to be chosen with care and logic to ensure that the predictions gained are based around factors relevant to the condition of the species. Some variables may be too specific or not relevant and using such variables would produce misleading results.

Essentially the analysis involves the software predicting the species distribution and environmental requirements based on factors included in the variables that influence this environmental suitability (Phillips, Anderson & Schapire 2006). The variables, or features, need to be selected discerningly as they are of critical importance to the model. The features chosen for this project relate mainly to the physical characteristics of the subject area. Road and drainage data has been included to see if proximity to these features has an influence on the species preferred habitat. The regional ecosystems, DEM and land use information will inform what type of environment is best suited to the subject. These variables provide a decent overview of the main aspects of the study environment and were sufficient for this study.

One major type of dataset present in similar studies but absent from this one is climate related data. This data has been left out as though a relatively large area there would be little difference in climate variables across specifically the Condamine region. As the data would likely be uniform across the area it is not valuable for this project. However, if in the future similar studies of the species were carried out in other areas climate data could be used as a means of comparison, helping to identify the preferred habitat at a larger scale.
3.4.4 MaxEnt Specifics and Settings
MaxEnt uses mathematical algorithms to determine a probability distribution for the area. Many of the very technical aspects of this are laid out in literature such as that by Phillips, Anderson and Schapire (2006). Some important concepts are as follows. MaxEnt divides the subject area into a number of pixels, hence the need for raster data. Each pixel is assigned a probability whose value is very small as all the values must add to one. Each occurrence point is considered a sample point and its environmental features are used to inform the prediction of the other cells (Elith et al. 2011).

There are a number of parameters to be set in the MaxEnt software before running it. These parameters are shown below along with their default value.

- Convergence threshold: $10^{-5}$
- Maximum iterations: 1000
- Regularisation value $\beta$: $10^{-4}$
- Use of Linear, Quadratic, Product and Binary features

These default values were to be used initially, but could have been changed after inaugural analysis if deemed appropriate to do so. Deciding upon these parameters is an ongoing subject of research (Dudík, Phillips & Schapire 2004; Merow, Smith & Silander 2013). For the environmental layers Maxent differentiates between continuous and categorical data. For this project the Land Use data was categorical and the remainder were continuous. In order to allow the software to run some statistical analysis a portion of the samples can be put aside as a random test percentage. In this instance 25% was the value set aside. The following two figures are screenshots of the settings used in running the analysis.
Chapter 3  
Research Methods

Figure 3-10 Main MaxEnt settings

Figure 3-11 Additional MaxEnt settings
3.4.5 Output
MaxEnt provides a wealth of output options including the distribution map and many graphs assessing variable contribution and the accuracy of the prediction. With regard to producing the species distribution map. Because the software gives a value to each pixel in the area, once the analysis is completed it was easy to display the data as a map classified by appropriate breaks in the probability values. These graphical depictions of the results provided the main visual result of the project from which the conclusions were drawn. These maps were constructed with the appropriate environmental layers so the connection between the environmental and the species distribution could be easily appreciated. The other various graphs produced were also valuable in making conclusions about the success of the model.

3.4.6 Further Analysis
Having obtained the main results it was decided to conduct further analysis in a smaller portion of the study area. This is called sensitivity analysis. This analysis involved selecting a portion of the study area which had been identified as being highly suitable and producing further results to see if more detailed results could be obtained.

3.6 Summary
A clear and well defined methodology was developed to ensure a clear focus was maintained at all times during research. Elements of the methods taken from previous studies were studied thoroughly to ensure the best possible procedures and results. Having absorbed the background information on how the program worked and how to run the analysis, thereby gaining an understanding of the mechanics of Maxent, the analysis was simply a matter of running the program to acquire the results. These results will be showcased in the next chapter.
Chapter 4 - Results

4.1 Introduction

Following the completion of the data pre-processing, the datasets were ready to be run through the Maxent software. Maxent provides the user with a wealth of results on all aspects of the analysis. This allows one to examine the effectiveness of the model itself along with obtaining the results. This capacity of the software allows the validity of the results to be assessed immediately. The first couple of attempts at running the model failed due to an error with the data. Once this issue was rectified the program ran successfully. Key excerpts from the analysis are presented below. Due to the wealth of documentation provided by the software the remainder will be in Appendix B. All these results and graphs have been generated by the Maxent software.

4.2 Results of Subject

4.2.1 Pictures of the Model

The first item presented is the picture of the Maxent model displaying the suitability of the area for the species. This is the species distribution map. Warmer colours (red, orange, yellow) indicate more suitable conditions for the bird and green colours being typical areas where the bird is found (Phillips 2005). This image is in the logistic format which is the easiest to interpret from a visual standpoint. In the logistic format each cell is assigned a probability, the total of these probability values adds to one. The white dots indicate the actual presence locations of the species as per the input samples data.
4.2.2 Response Curves

Response curves are derived from the model; these response curves indicate which variables are best at predicting the distribution at different values. Presented here are curves for the most important variable, Land Use, and the least important, Aspect. The y-axis represents the predicted probability of suitable conditions with respect to changes in the variable. Note that because the land use data is categorical not continuous its graph is in a different format.
Chapter 4

Results

Figure 4-2 Response curve for Aspect variable

Figure 4-3 Response curve for Land Use variable
4.2.3 Analysis of Variable Contributions

The following presents the variable contribution information in table form. Variable contribution is important as it allows an understanding of the key features that impact the ideal distribution.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>landuse</td>
<td>46.2</td>
<td>15.5</td>
</tr>
<tr>
<td>dem</td>
<td>34.9</td>
<td>45</td>
</tr>
<tr>
<td>slope</td>
<td>6.5</td>
<td>4.7</td>
</tr>
<tr>
<td>roads</td>
<td>5.2</td>
<td>5.1</td>
</tr>
<tr>
<td>re</td>
<td>3.4</td>
<td>6.5</td>
</tr>
<tr>
<td>drainage</td>
<td>2.4</td>
<td>21</td>
</tr>
<tr>
<td>aspect</td>
<td>1.3</td>
<td>2.2</td>
</tr>
</tbody>
</table>

This graph shows the results of the jackknife test for variable importance. This test works by running the process multiple times and creating a model omitting one variable on each instance. A second model is then created for each variable solely on its own. These models are then compared with the original one to assess the importance of each variable (Phillips 2005).
This second jackknife plot uses test gain rather than training gain. In this graph it shows DEM as the most important variable. Also note the changes to regional ecosystems and aspect. In this plot aspect appears with a negative gain. This means that the variable does not contribute to the model and in fact harms it. It could also mean that on its own the variable produces a model not better than random. With this in mind it may be beneficial to remove this variable from the model entirely.

The final jackknife plot uses AUC on test data. Here it shows that the model would produce a similar result if only the DEM was used. Also it is interesting to see the regional ecosystems and slope data take prominence over the land use. This plot also suggests that the gain would be increased if land use was omitted.
4.3 Results of Model

The next series of results pertain to the quality of the model derived and how accurately it predicts the sample. These results are more useful in determining whether the results achieved are accurate or not, allowing for comparisons between other MaxEnt models to be made. Because of this, these results are more theoretical and lack the more practical end of the research. The software is capable of running some statistical analysis, these are the results presented.

4.3.1 Analysis of Omission/Commission

First is the omission rate which refers to the fraction of test samples that are located in pixels that have not been identified by the model as being suitable habitat for the species (Phillips, Anderson & Schapire 2006). A low omission rate is necessary as it adds to the validity of the prediction. In this instance, the omission on training samples stays relatively in line with the predicted omission which is ideal.

![Figure 4-8 Omission rate graph](image)
Next is the receiver operating curve. In this graph the black line indicates what the model would look like if it performed no better than random. The closer the red line is to the top left the better the predictions are (Phillips 2005). Here the greater the area under the curve is the more successful the model. This information is useful for comparing different models created with MaxEnt.

![Sensitivity vs. 1 - Specificity for Calyptorhynchus lathamii](image)

**Figure 4-9 AUC graph**

### 4.4 Conclusion

The results of the software have successfully produced a species distribution map indicating habitat areas of high and low suitability. It has carried this out with a satisfactory degree of accuracy, so the information presented is of a sufficient quality to pass judgements about. These results have produced a number of interesting findings which will be discussed and interpreted in greater detail in the following chapter.
Chapter 5 - Discussion

5.1 Introduction
Having presented and examined the crux of the results in the preceding chapter the results will now be synthesised and their meaning explained and examined. This chapter will also include some general discussion on the results and identify further avenues for future research which could be pursued.

5.2 Accuracy of Model
The software allows the quality of the prediction to be determined. It is shown in the analysis of omission/commission that the model did produce a distribution that is better than random. This indicates success in the basics of the analysis and that the results can be considered representative of the conditions of the study area.

In the analysis of omission/commission it can be seen that the omission on training samples and test samples follow the predicted omission for the most part. This indicates that the prediction does comply with the samples data and so the model is validated and the results can be considered reliable. Ultimately this part performs to an acceptable standard.

The receiver operating curve for this model is quite satisfactory. The maximum value for the AUC is 1, for the training data in this instance it is 0.834. This graph requires intuitive interpretation as the optimal value of the AUC cannot be known (Phillips, Anderson & Schapire 2006). Based upon this value and the look of the graph it is evident that the model has produced a prediction better than random.

5.3 Distribution Map
Discussed now is the species distribution map. This is the main output for the project and based upon the previously learnt knowledge of the model’s success the findings of this map can be considered valid and accurate. The distribution map that has been created shows a limited amount of suitable areas, because Maxent identifies highly suitable areas. The Bird may still live in the ‘cooler’ areas but habitat/conservation efforts should be concerned with maximising the identified highly suitable areas.

The areas identified as containing highly suitable conditions are located near Bell, around the Bunya Mountains National Park, around Toowoomba and around Main
Range National Park. The most significant of these is Main Range National Park which is north-west of Killarney. As it is the most prominent area identified, the highly suitable area identified around Killarney will be the focus of further sensitivity analysis. This analysis is discussed in section 5.6. It should be noted that two of the areas identified are national parks which are known for being a refuge for endangered species. The following figure identifies the protected areas in the region including the aforementioned national parks.

![Protected Areas in the Condamine Region](image)

*Figure 5-1 Protected Areas in Condamine Region*
The highly suitable areas are on the outer ends of the boundary extent. This indicates that perhaps there are additional suitable habitat areas for the species outside of this projects subject area. Some possible future work may be to extend the analysis into these areas to determine if this is the case.

5.4 Variable Contributions

The response curves highlight the most important variables for making the predictions. They show that little impact is made by variables such as roads, aspect and slope, while the land use data has greatest influence with the DEM being of the second most importance. These response curves dictate sharp spikes in the Regional Ecosystems and slope curves indicating that for those certain group of values an accurate prediction can be made using those variables. The Jackknife graphs again show the layer aspect to be virtually negligible, with land use of primary consideration.

Though its influence is not as significant as land use, the prominence of the DEM could suggest that the species prefer areas of higher terrain, such as those in southern portion of the study area. This theory is supported when comparing the distribution map to the DEM input layer. It is most distinct in the southern portion where there is a direct correlation between the high elevation of the DEM and midrange to high suitable conditions of the distribution map as show in the figure.

![Figure 5-2 DEM with correlated area highlighted](image-url)
The drainage data also indicates a logical conclusion that suitable areas are not located in central area where there is greater distance from drainage lines. Obviously the species requires regular access to water sources. This is also strongly evident in the far northeast portion of the maps which shows that at the greatest distance from water the software predicts the lowest suitability, as identified by the figure.

![Drainage map showing area of interest](image)

**Figure 5-3 Drainage map showing area of interest**

The model shows that Land Use was the most critical variable in the prediction. It can be seen that there is correlation between the Land use and the prediction; suitable areas for the bird are not found in the central area with the ‘high’ value. The high values refer to land designated for cropping and other agricultural pursuits. It stands to reason that this land that has been interfered with and probably contains developed areas, such as those used for farming, which would not be suitable habitat. As land use has been found to be the most influential variable it would be of most interest to extra sensitivity analysis. The land use data also shows that the areas marked by the lowest value are in those areas of greatest suitability. These low types are areas marked as nature conservation and managed resource protection. These comments can be confirmed by the following figure.
The urban centre of Toowoomba is highly evident on the land use map and is in stark contrast to the makeup of the other areas of high suitability. It would be worth investigating why exactly this area has been identified as featuring highly suitable conditions as it does not necessarily seem to be an ideal spot. The most likely reasons could be because of Toowoomba’s location on the edge of the Great Dividing Range, due to the bird’s preference for sloped terrain and also the presence of Casuarina trees on private property in the region, perhaps due to residents encouraging native flora.

The variable importance can be slightly misleading at times when working with highly correlated variables (Phillips 2005). This could explain why aspect and to an extent slope were variables which provided a far smaller contribution to the overall model. Both these variables are derivatives of the DEM which is regarded as the second most important variable. It seems that perhaps the elevation model on its own provides enough information to form a reasonable model without the derivative variables. The process could be run again excluding the two derivative variables to assess if it makes much impact. Alternatively the reverse could be done in which case it could be expected that in the absence of the DEM the slope and aspect features would take on greater importance.
The slope data more or less just identifies where the two major national parks are rather than identifying anything specific about the characteristics of the bird, as can be seen in the next figure. The national parks are located at higher altitudes and therefore have greater slope hence the correlation to the DEM and the model.

![Slope data with notable areas marked](image)

The two variables with the greatest contribution to the model were land use and the DEM. Correlated variables could account for this as well. Because, just as aspect and slope are related to the DEM, components of the roads and drainage data would also be reflected in the land use variable.

### 5.5 General Observations

It would seem that the bulk of the Condamine region is not suitable habitat for the Glossy Black Cockatoo. However the small suitable areas that have been identified may be the most expansive suitable areas for the species. Without expanding the analysis further into the areas it cannot be stated with absolute certainty whether adjacent areas provide more highly suitable areas or just the same.

Elevation and land use has an impact on climate and vegetation, hence their importance to the bird’s habitat. Ideal habitat would also require sufficient vegetation which is reflected in the land use variable. Arid areas or open cropping fields would not be
satisfactory habitat. The comment on elevation could prove pertinent if the study was expanded in the future to other areas. Doing this would allow the examination of whether the bird prefers higher elevation areas or the climate and vegetation characteristics that come with such areas. Research could be done on the exact characteristics that make up the highly suitably areas.

5.6 Sensitivity Analysis

5.6.1 Overview
The most prominent highly suitable habitat region in the study area was in the southeast of the study area around Main Range National Park. For this reason the area was chosen as the subject of further sensitivity analysis.

5.6.2 Process
To undertake this additional analysis a new, smaller study area was identified based upon the findings in the main model. The area was chosen because it encompassed the most extensive highly suitable area. With this new boundary selected, the environmental and location datasets were clipped so their extent matched the new boundary. The analysis was then run once more through MaxEnt using the same settings.

5.6.3 Results
This analysis provided some interesting results; the majority of which can be found in Appendix C. Firstly the new distribution map is presented in the figure which also shows the smaller boundary extent.
The omission rate on samples, particularly test samples, was a little more wayward than the main model. This is perhaps due to the smaller scale on which the analysis was now being run. Previously suitable areas may no longer be identified as such and it would be more likely for the bird to move around this smaller area. In the AUC test the sensitivity analysis produced a better result, keeping in mind that a value of 1 is the optimal result. This is shown in the following table.

\begin{table}[h]
\centering
\begin{tabular}{ |c|c|c| }
\hline
 & \textbf{Sensitivity Analysis} & \textbf{Main Model} \\
\hline
\textbf{Training Data} & AUC = 0.859 & AUC = 0.834 \\
\textbf{Test Data} & AUC = 0.801 & AUC = 0.777 \\
\hline
\end{tabular}
\caption{Comparison of AUC values}
\end{table}

Looking at the distribution map it shows that the focus of the prediction has been narrowed and that there are fewer highly suitable areas, as was expected. The variable contribution shows land use to still be the dominant one. The DEM was still the second most critical variable but its influence was lessened while the Regional Ecosystems saw increased prominence. In this model roads became the least important variable; this is probably because the study area is now focussed on a region primarily featuring nature conservation areas. There would be fewer roads to influence the model in national parks.
compared to metropolitan areas. The following figure shows the variable contributions of this new model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>landuse</td>
<td>51.6</td>
<td>26.6</td>
</tr>
<tr>
<td>dem</td>
<td>17.7</td>
<td>30.5</td>
</tr>
<tr>
<td>re</td>
<td>11.2</td>
<td>3.4</td>
</tr>
<tr>
<td>slope</td>
<td>8.6</td>
<td>25.1</td>
</tr>
<tr>
<td>drainage</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>aspect</td>
<td>3.8</td>
<td>9</td>
</tr>
<tr>
<td>roads</td>
<td>2.1</td>
<td>5.4</td>
</tr>
</tbody>
</table>

Figure 5-7 Sensitivity analysis variable contribution table

5.6.4 Limited Variables Model

Out of interest to see what affect it would have on the model another run of the analysis was performed using just the DEM and land use variables. The result produced a much less stable prediction both in terms of the accuracy of the model and the level of detail achieved by the prediction. Evidently while they may be the most critical variables they can only succeed in bringing the prediction to a certain degree of accuracy from which the other variables assist in refining the distribution and strengthening the model. Presented below is a comparison of the two species distribution maps created during the sensitivity analysis, the remainder of the results from the second run are in Appendix D.

Figure 5-8 Comparison of SDMs created using all seven variables (Left) and just two variables (Right).
5.7 Summary

This chapter has presented a discussion of the results achieved. Drawing logical conclusions based upon what the analysis has shown. The analysis has been successful in predicting the habitat location of the Glossy Black Cockatoo but there are still things that could be done to enhance this research. Some of these factors have already been touch upon, but more will be discussed in the next chapter.
Chapter 6 - Conclusions and Recommendations

6.1 Conclusions

Finally some further thoughts are presented to wrap up the key points of this research. In particular, possibilities for future study on this topic are presented along with some final recommendations and conclusions.

The software has little room for customisation and this study is just an example of applying standard industries practices to an application. There may be room for more detailed specific study using specifically tailored methods. One is reliant upon Maxent’s statistical models when using the software. This could be an issue as the software’s methodology may not be the most suitable for the task at hand. However, having said this, the results achieved are satisfactory.

The research has produced a species distribution map for the Glossy Black Cockatoo in Queensland’s Condamine region. The model that has been produced relied on some variables far more than others. This may indicate that some variables were simply not as relevant as others and perhaps there are alternate ones that could be used in a future study. The model obtained reveals that the land use and elevation of an area are of primary consideration when dealing with habitat preferences for the Glossy Black Cockatoo.

The project has succeeded in producing a species distribution map for the Glossy Black Cockatoo in Queensland’s Condamine region. The results achieved were of a high standard and worthy of interpretation. Given the time and scope restriction on the study some additional research pursuits were unable to be carried out but have been suggested as a possibility for future research. In addition to the construction of the species distribution map the other specific objectives of the research have been achieved. The importance of the environmental variables used has been extensively discussed. The conducted literature review provided insight into processes being applied in this research along with information on the species itself. This information influenced the research as well assisting in identifying paths for future research.

The project has been successful in filling a research gap. It has done so by providing an additional study featuring an application of species distribution modelling and MaxEnt as well as providing valuable information on the Glossy Black Cockatoo that may be used
to influence future conservation efforts for the species. This project has used the GIS technologies of species distribution modelling and MaxEnt to a practical end, modelling the preferred habit of the Glossy Black Cockatoo, one of many endangered species in Australia’s vast ecological landscape.

6.2 Recommendations
While some comments have already been made with regards to what future work could be carried out on this topic, some additional thoughts are presented here. Carrying out further study in different areas would allow for greater understanding of the best conditions for the species, doing this would allow for comparisons to be made between the different areas.

Another interesting future avenue to pursue may be to run the same model using different species data to see if there is correlation between habitat preferences of these species, particularly other birds. More runs could also be done to further validate the results. Doing this would help to further assess the success of the MaxEnt model and ensure that it has predicted the ‘correct’ areas. Such runs could be done with similar data. This study only used the MaxEnt software for analysis. As there are a number of different programs available it may be interesting to test the data using alternate software as well.

Tying back into a primary aim of the research it is yet to be seen what specific impact this information may bring to conservation efforts. It will need to be determined how practical it would be to focus conservation efforts for the species to this area. Considering that areas around Killarney are already classified as national park, it seems that the species interests are being catered to whether directly or indirectly. It was found during the literature review that the Glossy Black Cockatoo’s primary food source and nesting site is provided by Casuarina trees. Therefore this would be an interesting variable with which to run the model, if data on the location of these trees could be obtained.

In order to assist with conservation efforts for the species, attention should be focused initially on not degrading the identified areas in any way. Further to that movements could be made to enhance the characteristics of these areas that make them suitable in order to promote them as habitat for the species. The Glossy Black Cockatoo is heavily reliant upon Casuarina trees; hence it would be beneficial to encourage the growth of that tree, particularly in the highly suitable areas. The species distribution map that has
been produced will be useful for those involved in protecting the species and also gives an indication as to where the bird is most likely to be sighted for more recreational purposes, such as for bird watching groups.
References


Cameron, M 2006, 'Nesting habitat of the glossy black-cockatoo in central New South Wales', *Biological Conservation*, vol. 127, no. 4, pp. 402-10.


53


Glossy Black Conservancy 2010, 'Glossy Black-Cockatoo Conservation Guidelines for South-Eastern Queensland and Far North-Eastern New South Wales', Glossy Black Conservancy,


Newton, I 1994, 'The role of nest sites in limiting the numbers of hole nesting birds: a review', *Biological Conservation* 70, pp. 265–76.


Appendices

Appendix A

Project Specification

For: Michael Evans

Title: Species Distribution Modelling of the Glossy Black Cockatoo in Queensland’s Condamine Region

Major: Geographic Information Systems

Supervisor: Armando Apan

Enrolment: ENG4111 – ONC S1, 2016
ENG4112 – ONC S2, 2016

Project Aim: To undertake species distribution modelling of the Glossy Black Cockatoo (Calyptorhynchus lathami), in the Condamine Region of Queensland, to map its habitat suitability that will help in conservation efforts for the species.

Programme: Issue B

1. Research existing information on Species Distribution Modelling and other related concepts in order to provide a knowledge foundation on which to build.

2. Review the literature related to the biology, particularly on habitat preferences, of the Glossy Black Cockatoo.

3. Collect the appropriate and relevant spatial datasets to be used for the project.

4. Process the data in GIS, putting them in the format appropriate to undertake species distribution modelling.

5. Analyse the data using Maxent software and to produce species distribution (habitat suitability) map for the species.

6. Evaluate the results and make conclusions and recommendations based upon them.

7. Write and submit the dissertation.
Maxent model for Calyptorhynchus_lathami

This page contains some analysis of the Maxent model for Calyptorhynchus_lathami, created Thu Jul 28 15:28:14 AEST 2016 using Maxent version 3.3.3k. If you would like to do further analyses, the raw data used here is linked to at the end of this page.

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.

![Omission and Predicted Area for Calyptorhynchus_lathami](image)

The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.803 rather than 1; in practice the test AUC may exceed this bound.
Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes $6 \times$ training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

<table>
<thead>
<tr>
<th>Cumulative threshold</th>
<th>Logistic threshold</th>
<th>Description</th>
<th>Fractional predicted area</th>
<th>Training omission rate</th>
<th>Test omission rate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.060</td>
<td>Fixed cumulative value 1</td>
<td>0.864</td>
<td>0.016</td>
<td>0.000</td>
<td>5.331E-2</td>
</tr>
<tr>
<td>5.000</td>
<td>0.122</td>
<td>Fixed cumulative value 5</td>
<td>0.697</td>
<td>0.048</td>
<td>0.050</td>
<td>7.17E-3</td>
</tr>
<tr>
<td>10.000</td>
<td>0.161</td>
<td>Fixed cumulative value 10</td>
<td>0.569</td>
<td>0.095</td>
<td>0.200</td>
<td>2.788E-2</td>
</tr>
<tr>
<td>0.812</td>
<td>0.054</td>
<td>Minimum training presence</td>
<td>0.877</td>
<td>0.000</td>
<td>0.000</td>
<td>7.207E-2</td>
</tr>
<tr>
<td>10.250</td>
<td>0.163</td>
<td>10 percentile training presence</td>
<td>0.563</td>
<td>0.095</td>
<td>0.200</td>
<td>2.488E-2</td>
</tr>
<tr>
<td>31.550</td>
<td>0.281</td>
<td>Equal training sensitivity and specificity</td>
<td>0.233</td>
<td>0.238</td>
<td>0.350</td>
<td>8.279E-5</td>
</tr>
<tr>
<td>42.779</td>
<td>0.381</td>
<td>Maximum training sensitivity plus specificity</td>
<td>0.134</td>
<td>0.317</td>
<td>0.450</td>
<td>1.294E-5</td>
</tr>
<tr>
<td>25.899</td>
<td>0.245</td>
<td>Equal test sensitivity and specificity</td>
<td>0.300</td>
<td>0.206</td>
<td>0.300</td>
<td>2.611E-4</td>
</tr>
<tr>
<td>51.829</td>
<td>0.481</td>
<td>Maximum test sensitivity plus specificity</td>
<td>0.083</td>
<td>0.429</td>
<td>0.450</td>
<td>1.06E-7</td>
</tr>
</tbody>
</table>
## Specificity

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Value</th>
<th>Standard Error</th>
<th>Omission</th>
<th>Area</th>
<th>Threshold Value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance training omission, predicted area and threshold value</td>
<td>0.812</td>
<td>0.054</td>
<td>0.877</td>
<td>0.000</td>
<td>0.000</td>
<td>7.207E-2</td>
</tr>
<tr>
<td>Equate entropy of thresholded and original distributions</td>
<td>17.540</td>
<td>0.201</td>
<td>0.424</td>
<td>0.159</td>
<td>0.250</td>
<td>3.237E-3</td>
</tr>
</tbody>
</table>

---

### Pictures of the model

This is a representation of the Maxent model for Calyptorhynchus_lathami. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

Click [here](#) to interactively explore this prediction using the Explain tool. If clicking from your browser does not succeed in starting the tool, try running the script in C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Run 5\Output\Calyptorhynchus_lathami_explain.bat directly. This tool requires the environmental grids to be small enough that they all fit in memory.
Response curves

These curves show how each environmental variable affects the Maxent prediction. The curves show how the logistic prediction changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version. Note that the curves can be hard to interpret if you have strongly correlated variables, as the model may depend on the correlations in ways that are not evident in the curves. In other words, the curves show the marginal effect of changing exactly one variable, whereas the model may take advantage of sets of variables changing together.

In contrast to the above marginal response curves, each of the following curves represents a different model, namely, a Maxent model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. They may be easier to interpret if there are strong correlations between variables.
Analysis of variable contributions

The following table gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>landuse</td>
<td>46.2</td>
<td>15.5</td>
</tr>
<tr>
<td>dem</td>
<td>34.9</td>
<td>45</td>
</tr>
<tr>
<td>slope</td>
<td>6.5</td>
<td>4.7</td>
</tr>
<tr>
<td>roads</td>
<td>5.2</td>
<td>5.1</td>
</tr>
<tr>
<td>re</td>
<td>3.4</td>
<td>6.5</td>
</tr>
<tr>
<td>drainage</td>
<td>2.4</td>
<td>21</td>
</tr>
<tr>
<td>aspect</td>
<td>1.3</td>
<td>2.2</td>
</tr>
</tbody>
</table>

The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is landuse, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is landuse, which therefore appears to have the most information that isn’t present in the other variables.
The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.

Lastly, we have the same jackknife test, using AUC on test data.

Raw data outputs and control parameters
The data used in the above analysis is contained in the next links. Please see the Help button for more information on these.

**The model applied to the training environmental layers**

**The coefficients of the model**

**The omission and predicted area for varying cumulative and raw thresholds**

**The prediction strength at the training and (optionally) test presence sites**

**Results for all species modeled in the same Maxent run, with summary statistics and (optionally) jackknife results**

Regularized training gain is 0.859, training AUC is 0.834, unregularized training gain is 1.126. Unregularized test gain is 0.874. Test AUC is 0.777, standard deviation is 0.058 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm terminated after 500 iterations (16 seconds).

The follow settings were used during the run:

- 63 presence records used for training, 20 for testing.
- 10063 points used to determine the Maxent distribution (background points and presence points).
- Environmental layers used: aspect dem drainage landuse(categorical) re roads slope
- Regularization values: linear/quadratic/product: 0.156, categorical: 0.250, threshold: 1.370, hinge: 0.500
- Feature types used: hinge linear quadratic
- respnecurves: true
- jackknife: true
- outputdirectory: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Run 5\Output
- samplesfile: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Run 5\bird_gda.csv
- environmentallayers: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Run 5\ASCII
- randomtestpoints: 25
- Command line used:

```
Command line to repeat this species model: java density.MaxEnt nowarnings noprefixes -E "" -E Calyptorhynchus_lathami respnecurves jackknife 
"outputdirectory=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Run 5\Output"
"samplesfile=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Run 5\bird_gda.csv"
"environmentallayers=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Run 5\ASCII"
randomtestpoints=25 -N roads2 -t landuse
```

file://C:/Users/Michael/Documents/Uni/2016/ERP/MaxEnt/ForAnalysis/Run%205/Output/Calyptorhynchus_lathami.html
Appendices

Appendix C
Maxent model for Calyptorhynchus_lathami

This page contains some analysis of the Maxent model for Calyptorhynchus_lathami, created Sun Sep 04 12:50:47 AEST 2016 using Maxent version 3.3.3k. If you would like to do further analyses, the raw data used here is linked to at the end of this page.

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.

The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.784 rather than 1; in practice the test AUC may exceed this bound.
Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes $6 \times \text{training omission rate} + 0.04 \times \text{cumulative threshold} + 1.6 \times \text{fractional predicted area}.$

<table>
<thead>
<tr>
<th>Cumulative threshold</th>
<th>Logistic threshold</th>
<th>Description</th>
<th>Fractional predicted area</th>
<th>Training omission rate</th>
<th>Test omission rate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.066</td>
<td>Fixed cumulative value 1</td>
<td>0.874</td>
<td>0.000</td>
<td>0.000</td>
<td>2.981E-1</td>
</tr>
<tr>
<td>5.000</td>
<td>0.149</td>
<td>Fixed cumulative value 5</td>
<td>0.701</td>
<td>0.000</td>
<td>0.000</td>
<td>4.081E-2</td>
</tr>
<tr>
<td>10.000</td>
<td>0.200</td>
<td>Fixed cumulative value 10</td>
<td>0.578</td>
<td>0.000</td>
<td>0.111</td>
<td>5.453E-2</td>
</tr>
<tr>
<td>15.358</td>
<td>0.239</td>
<td>Minimum training presence</td>
<td>0.479</td>
<td>0.000</td>
<td>0.111</td>
<td>1.439E-2</td>
</tr>
<tr>
<td>20.800</td>
<td>0.275</td>
<td>10 percentile training presence</td>
<td>0.398</td>
<td>0.069</td>
<td>0.222</td>
<td>2.428E-2</td>
</tr>
<tr>
<td>31.057</td>
<td>0.335</td>
<td>Equal training sensitivity and specificity</td>
<td>0.276</td>
<td>0.276</td>
<td>0.222</td>
<td>2.525E-3</td>
</tr>
<tr>
<td>20.012</td>
<td>0.270</td>
<td>Maximum training sensitivity plus specificity</td>
<td>0.409</td>
<td>0.034</td>
<td>0.222</td>
<td>2.844E-2</td>
</tr>
<tr>
<td>33.054</td>
<td>0.346</td>
<td>Equal test sensitivity and specificity</td>
<td>0.256</td>
<td>0.276</td>
<td>0.222</td>
<td>1.559E-3</td>
</tr>
</tbody>
</table>

Training data (AUC = 0.859) | Test data (AUC = 0.801) | Random Prediction (AUC = 0.5)
Pictures of the model

This is a representation of the Maxent model for Calyptorhynchus_lathami. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

Click here to interactively explore this prediction using the Explain tool. If clicking from your browser does not succeed in starting the tool, try running the script in C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Output\Calyptorhynchus_lathami_explain.bat directly. This tool requires the environmental grids to
be small enough that they all fit in memory.

---

**Response curves**

These curves show how each environmental variable affects the Maxent prediction. The curves show how the logistic prediction changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version. Note that the curves can be hard to interpret if you have strongly correlated variables, as the model may depend on the correlations in ways that are not evident in the curves. In other words, the curves show the marginal effect of changing exactly one variable, whereas the model may take advantage of sets of variables changing together.

![Response curves](image)

In contrast to the above marginal response curves, each of the following curves represents a different model, namely, a Maxent model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. They may be easier to interpret if there are strong correlations between variables.

![Response curves](image)
Analysis of variable contributions

The following table gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>landuse</td>
<td>51.6</td>
<td>26.6</td>
</tr>
<tr>
<td>dem</td>
<td>17.7</td>
<td>30.5</td>
</tr>
<tr>
<td>re</td>
<td>11.2</td>
<td>3.4</td>
</tr>
<tr>
<td>slope</td>
<td>8.6</td>
<td>25.1</td>
</tr>
<tr>
<td>drainage</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>aspect</td>
<td>3.8</td>
<td>9</td>
</tr>
<tr>
<td>roads</td>
<td>2.1</td>
<td>5.4</td>
</tr>
</tbody>
</table>

The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is landuse, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is landuse, which therefore appears to have the most information that isn't present in the other variables.
The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we’re looking at test data.

Lastly, we have the same jackknife test, using AUC on test data.

Raw data outputs and control parameters
The data used in the above analysis is contained in the next links. Please see the Help button for more information on these.

- The model applied to the training environmental layers
- The coefficients of the model
- The omission and predicted area for varying cumulative and raw thresholds
- The prediction strength at the training and (optionally) test presence sites
- Results for all species modeled in the same Maxent run, with summary statistics and (optionally) jackknife results

Regularized training gain is 0.658, training AUC is 0.859, unregularized training gain is 1.049. Unregularized test gain is 0.853. Test AUC is 0.801, standard deviation is 0.074 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2). Algorithm converged after 420 iterations (11 seconds).

The follow settings were used during the run:
- 29 presence records used for training, 9 for testing.
- 10029 points used to determine the Maxent distribution (background points and presence points).
- Environmental layers used: aspect dem drainage landuse(categorical) re roads slope
- Regularization values: linear/quadratic/product: 0.269, categorical: 0.250, threshold: 1.710, hinge: 0.500
- Feature types used: hinge linear quadratic
- responsecurves: true
- jackknife: true
- outputdirectory: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Output
- samplesfile: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Bird_sen.csv
- environmentallayers: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\ASCII
- randomtestpoints: 25

Command line used:
```
java density.MaxEnt nowarnings noprefixes -E "" -E Calyptorhynchus_lathami responsecurves jackknife
"outputdirectory=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Output"
"samplesfile=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Bird_sen.csv"
"environmentallayers=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\ASCII" randomtestpoints=25 -t landuse
```
Appendices

Appendix D
Maxent model for Calyptrorhynchus_lathami

This page contains some analysis of the Maxent model for Calyptrorhynchus_lathami, created Wed Sep 07 11:27:27 AEST 2016 using Maxent version 3.3.3k. If you would like to do further analyses, the raw data used here is linked to at the end of this page.

Analysis of omission/commission

The following picture shows the omission rate and predicted area as a function of the cumulative threshold. The omission rate is calculated both on the training presence records, and (if test data are used) on the test records. The omission rate should be close to the predicted omission, because of the definition of the cumulative threshold.

![Omission and Predicted Area for Calyptrorhynchus_lathami](image)

The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than 1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.727 rather than 1; in practice the test AUC may exceed this bound.
Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes $6 \times \text{training omission rate} + 0.04 \times \text{cumulative threshold} + 1.6 \times \text{fractional predicted area}.$

<table>
<thead>
<tr>
<th>Cumulative threshold</th>
<th>Logistic threshold</th>
<th>Description</th>
<th>Fractional predicted area</th>
<th>Training omission rate</th>
<th>Test omission rate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.091</td>
<td>Fixed cumulative value 1</td>
<td>0.902</td>
<td>0.000</td>
<td>0.000</td>
<td>3.964E-1</td>
</tr>
<tr>
<td>5.000</td>
<td>0.222</td>
<td>Fixed cumulative value 5</td>
<td>0.758</td>
<td>0.000</td>
<td>0.111</td>
<td>3.21E-1</td>
</tr>
<tr>
<td>10.000</td>
<td>0.271</td>
<td>Fixed cumulative value 10</td>
<td>0.669</td>
<td>0.034</td>
<td>0.111</td>
<td>1.462E-1</td>
</tr>
<tr>
<td>8.206</td>
<td>0.269</td>
<td>Minimum training presence</td>
<td>0.699</td>
<td>0.000</td>
<td>0.111</td>
<td>1.945E-1</td>
</tr>
<tr>
<td>13.681</td>
<td>0.278</td>
<td>10 percentile training presence</td>
<td>0.608</td>
<td>0.069</td>
<td>0.111</td>
<td>7.753E-2</td>
</tr>
<tr>
<td>38.094</td>
<td>0.378</td>
<td>Equal training sensitivity and specificity</td>
<td>0.324</td>
<td>0.310</td>
<td>0.333</td>
<td>3.677E-2</td>
</tr>
<tr>
<td>45.684</td>
<td>0.388</td>
<td>Maximum training sensitivity plus specificity</td>
<td>0.247</td>
<td>0.379</td>
<td>0.333</td>
<td>9.462E-3</td>
</tr>
<tr>
<td>37.152</td>
<td>0.377</td>
<td>Equal test sensitivity and specificity</td>
<td>0.333</td>
<td>0.310</td>
<td>0.333</td>
<td>4.242E-2</td>
</tr>
<tr>
<td>70.420</td>
<td>0.715</td>
<td>Maximum test sensitivity plus specificity</td>
<td>0.047</td>
<td>0.586</td>
<td>0.444</td>
<td>2.38E-5</td>
</tr>
</tbody>
</table>
### Specificity

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Omission Balance</th>
<th>Predicted Area</th>
<th>Threshold Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.066</td>
<td>0.199</td>
<td>0.780</td>
<td>0.000</td>
<td>0.111</td>
</tr>
<tr>
<td>12.816</td>
<td>0.276</td>
<td>0.622</td>
<td>0.069</td>
<td>0.111</td>
</tr>
</tbody>
</table>

### Pictures of the model

This is a representation of the Maxent model for Calyptorhynchus_lathami. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

Click [here](file:///C:/Users/Michael/Documents/Uni/2016/ERP/MaxEnt/ForAnalysis/Sensitivity%20Analysis/Run%202_7.9/Output/Calyptorhynchus_lathami.html) to interactively explore this prediction using the Explain tool. If clicking from your browser does not succeed in starting the tool, try running the script in `C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Run 2_7.9\Output\Calyptorhynchus_lathami_explain.bat` directly. This tool requires the environmental grids to be small enough that they all fit in memory.
**Response curves**

These curves show how each environmental variable affects the Maxent prediction. The curves show how the logistic prediction changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version. Note that the curves can be hard to interpret if you have strongly correlated variables, as the model may depend on the correlations in ways that are not evident in the curves. In other words, the curves show the marginal effect of changing exactly one variable, whereas the model may take advantage of sets of variables changing together.

In contrast to the above marginal response curves, each of the following curves represents a different model, namely, a Maxent model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. They may be easier to interpret if there are strong correlations between variables.

---

**Analysis of variable contributions**

The following table gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent contribution</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>landuse</td>
<td>76.3</td>
<td>76.2</td>
</tr>
<tr>
<td>dem</td>
<td>23.7</td>
<td>23.8</td>
</tr>
</tbody>
</table>
The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is landuse, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is landuse, which therefore appears to have the most information that isn't present in the other variables.

![Jackknife of regularized training gain for Calyptorhynchus_lathami](image)

The next picture shows the same jackknife test, using test gain instead of training gain. Note that conclusions about which variables are most important can change, now that we're looking at test data.

![Jackknife of test gain for Calyptorhynchus_lathami](image)

Lastly, we have the same jackknife test, using AUC on test data.

![Jackknife of AUC for Calyptorhynchus_lathami](image)

### Raw data outputs and control parameters

The data used in the above analysis is contained in the next links. Please see the Help button for more information on these.

- The model applied to the training environmental layers
- The coefficients of the model
- The omission and predicted area for varying cumulative and raw thresholds
- The prediction strength at the training and (optionally) test presence sites
- Results for all species modeled in the same Maxent run, with summary statistics and (optionally) jackknife results

Regularized training gain is 0.474, training AUC is 0.769, unregularized training gain is 0.658. Unregularized test gain is 0.874.
Test AUC is 0.782, standard deviation is 0.095 (calculated as in DeLong, DeLong & Clarke-Pearson 1988, equation 2).
Algorithm converged after 160 iterations (2 seconds).

The follow settings were used during the run:
29 presence records used for training, 9 for testing.
10029 points used to determine the Maxent distribution (background points and presence points).
Environmental layers used: dem landuse(categorical)
Regularization values: linear/quadratic/product: 0.269, categorical: 0.250, threshold: 1.710, hinge: 0.500
Feature types used: hinge linear quadratic
responsecurves: true
jackknife: true
outputdirectory: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Run 2_7.9\Output
samplesfile: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Run 2_7.9\Bird_sen.csv
environmentallayers: C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Run 2_7.9\ASCII
randomtestpoints: 25
Command line used:
Command line to repeat this species model: java density.MaxEnt nowarnings noprefixes -E "" -E Calyptorhynchus_lathami responsecurves jackknife
"outputdirectory=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Run 2_7.9\Output"
"samplesfile=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Run 2_7.9\Bird_sen.csv"
"environmentallayers=C:\Users\Michael\Documents\Uni\2016\ERP\MaxEnt\ForAnalysis\Sensitivity Analysis\Run 2_7.9\ASCII" randomtestpoints=25 -N aspect -N drainage -N re -N roads -N slope -t landuse