BioNest Analytics

Conservation of Shoalhaven's Endangered Shorebird Species

By BioNest Analytics:

Dhuvaraha Baheerathan (z5361999)

Joceline Rijanto (z5394645)

Rohini Manohar (z5367092)

Sean Yeoh (z5358754)

Zhouxing Yang (z5314108)

BioNest Analytics

Our mission is to solve complex biodiversity problems by applying advanced statistical techniques to guarantee the conservation of endangered species.

Table of Contents

Executive Summary 3
Background 4
Scope 7
Plan 9
Findings 12
Discussion 29
Peer Review 31
Appendices 32
References 44

Executive Summary

The Shoalhaven region, located on the South Coast of NSW is a rich haven for biodiversity, home to a diverse range of species, many of which are threatened and endangered. However, the region is becoming more vulnerable to the detrimental impacts of climate change, which endangers the unique species and biodiversity of the area. The purpose of this study is to provide recommendations for a government revegetation project that is to take place in NSW's Shoalhaven region to ensure the project supports the local ecology and future climate conditions, maximising long-term biodiversity.

To maximise long-term biodiversity, our team at BioNest Analytics aimed to prevent the extinction of existing threatened and endangered species and chose to prioritize securing the future of five flagship shorebird species: Little Tern, Eastern Curlew, Bar-Tailed Godwit, Great Knot, and Curlew Sandpiper.

To focus the revegetation efforts on a smaller subsection of Shoalhaven, our team mapped the occurrences of the flagship species across Shoalhaven and found that they predominantly resided in Lake Wollumboola. Upon refining the location, our team forecasted the climate in the region utilising recent climate data and external projections. Following this, our team identified potential flora for revegetation and pinpointed their exact geolocations to build a species distribution model. The species distribution model enabled our team to analyze the suitability of potential flora species under climate conditions and select five candidates to implement in the revegetation scheme.

The findings from our analysis suggest that the government revegetation project should select the following species: Banksia Robur, Boronia Deanei, Avicennia Marina, Carpobrotus Edulis, and Carpobrotus Glaucescens. In the discussion of our analysis, we addressed potential limitations with the data we relied upon, and the method our team employed to provide recommendations for the revegetation project.

Hence, we advise that our recommendation be incorporated with caution and advocate for future analysts to implement improvements to the data collection process to produce more nuanced and accurate analysis that could provide invaluable insights for the revegetation program.

Background

Topic Introduction

Biodiversity refers to the variety and variability of plant, animal and microorganism species within a habitat and is crucial for supporting healthy ecosystems (NSW Development of Planning and Environment 2023). The Shoalhaven region, the revegetation scheme's designated location this project aims to implement, is home to a variety of habitats and ecological communities, 16 of which are listed as Endangered Ecological Communities under the NSW BC Act (Shoalhaven City Council n.d.). This biodiversity project is focused on the wetlands in Shoalhaven, as Australian wetlands play a pivotal role in maintaining global biodiversity as they are important nesting and feeding habitats for migratory shorebirds, amphibians, reptiles, and plant species (Department of Climate Change, Energy, the Environment and Water 2021). The greatest long-term threat to biodiversity is predicted to be climate change, with wetlands being amongst the most vulnerable ecosystems impacted (NSW Development of Planning and Environment 2023). Although wetlands can protect communities from floods and relieve droughts, degraded wetlands are more susceptible to climate change and may further contribute to the detrimental nature of climate change as they are large captors and storers of carbon (Department of Climate Change, Energy, the Environment and Water 2019).

Due to the critical role that wetlands have in supporting biodiversity, particularly shorebird species and their sensitivity to climate change, our project is focused on securing the future of five flagship endangered shorebird species listed under the NSW Threatened Species Conservation act and residing in the Shoalhaven area, which are the Little Tern, Eastern Curlew, Bar-Tailed Godwit, Great Knot, and Curlew Sandpiper (Lake Wollumboola Protection Association Inc n.d.). Upon analysing the spread of occurrences of the flagship species on the Atlas of Living Australia (ALA) map, our team have decided to take a case study of the Jervis Bay region within Shoalhaven, in particular, Lake Wollumboola located as part of the Jervis Bay National Park, situated on the South Coast of New South Wales, between the Shoalhaven-Crookhaven River estuary and Jervis Bay area (BirdLife Australia 2018).

Jervis Bay and Lake Wollumboola are listed as Key Biodiversity Areas (KBA) under the International Union for the Conservation of Nature's (IUCN) Global Standard for Identification (BirdLife Australia 2018). The KBA program is a global initiative established by a partnership of leading conservation NGOs that support the identification and conservation of critical sites for species and ecological communities (Key Biodiversity Areas n.d.). Our particular interest in Lake Wollumboola is due to its significance as a shorebird habitat, as supported by the observation of over 100 bird species, with at least 16 of those species listed under the NSW Threatened Species Conservation Act, including the five flagship species identified (Lake Wollumboola Protection Association Inc n.d.). The global significance of Lake Wollumboola is particularly recognised as a breeding habitat for the Little Tern, a key species in our case study, as part of the East Asian-Australasian Flyway Partnership, which covers the species' protection under migratory bird agreements with China, Japan and South Korea (Shoalhaven City Council 2023).

Prior Studies

The conservation of Lake Wollumboola has predominantly been upheld by the Shoalhaven City Council and local conservation groups, such as Birdlife Shoalhaven and the Lake Wollumboola Protection Association Inc. The previous work conducted in the area includes campaigning against development proposals that risk the conservation and protection of the environment. One of the more recent campaigns includes providing recommendations to the draft Illawarra-Shoalhaven Regional Plan 2042, primarily focusing on the proposed conservation efforts outlined in the plan (Lake Wollumboola Protection Association Inc 2020). There has been prior work done regarding the conservation of shorebird species in Shoalhaven as part of a NSW Government flagship program, Saving Our Species (SoS) Migratory Shorebird Project (Department of Planning and Environment 2023). The project focused on securing the future of six threatened species, which include the same species that our team selected for this case study, solidifying the decision to focus on the five selected flagship species. The project involved implementing beach warden patrols, sculptures, and murals at the Shoalhaven estuary to foster community engagement and education (Department of Planning and Environment 2023). However, there does not appear to be any recent projects in Lake Wollumboola that have implemented a revegetation scheme to ensure the future of threatened and endangered shorebird species.

Project Objectives

The primary objective of our project is to recommend 5 native flora species that can be implemented in the government revegetation scheme to prevent the extinction of the flagship species we've identified and other endangered shorebird species. To ensure the suitability of the recommended flora species and long-term success of the revegetation scheme, our biodiversity project consists of four main objectives:

1. **Spatial Distribution of the Flagship Shorebird Species**
   The spatial distribution of the following species, *Little Tern, Eastern Curlew, Bar-Tailed Godwit, Great Knot, and Curlew Sandpiper* will be mapped using ALA and the data collected will be used to select a region to conduct a case study.

2. **Future Climate Conditions**
   The future climate conditions for the 2023-2039 period in the Shoalhaven region will be produced using 2001-2023 climate data in conjunction with external climate change projection maps.

3. **Geolocation of Potential Vegetation**
   Potential flora species for revegetation will be collected spatially on ALA through an observation of the floral species in wetlands. Then, their geolocation i.e. latitude and longitude will be collected for use for Objective 4.

4. **Selection of Flora Species**
   The data collected in Objective 3 will be used to observe their current climate conditions, and tested against the future climate conditions collected from Objective 2 to see their viability in the future climate conditions. Then, the top 5 best candidates will be selected for the revegetation scheme.

Scope

Deliverables

BioNest Analytics aims to deliver a 30-page report on the revegetation scheme in Shoalhaven, detailing the methodology of the four main project objectives and our proposed recommendations. Our findings within the report will be presented to Xylos Systems representatives and the wider DATA3001 cohort. We anticipate that the report and presentation will be completed within three months of commencement of the project.

Information Provided & Data Used

The following data and model used from the information provided and external sources include:

- Spread of occurrences of 5 flagship endangered shorebird species within Shoalhaven from the period of 2018-2022 sourced from ALA
- An interactive climate change projections map sourced from AdaptNSW
- Spread of occurrences of 5 native flora species within Australia from the period of 2018-2022 sourced from ALA
- Climate conditions based on latitude and longitude sourced from WorldClim
- Temperature and rainfall of Jervis Bay from the period of 2001-2023 from the Bureau of Meteorology

Assumptions

For the validity of our report, the following are assumptions that we've considered for our project include:

1. **No introduction of foreign invasive species into the newly revegetated areas.**

   The introduction of an invasive species, i.e., a species whose abundance is greater than an accepted distribution within an area undermines the local ecosystem, threatening biodiversity and reducing total native species (Department of Climate Change, Energy, the Environment and Water 2022). Our team assumes that no foreign invasive species are introduced to the newly revegetated areas, due to the impractical nature of determining potential invasive species, given they may be brought in from anywhere. In addition, the scale of damage caused to the native fauna and flora by foreign species is highly unpredictable.

2. **No human intrusions or disturbances in the newly revegetated area, particularly construction.**

   To preserve the newly revegetated area, human intrusions including development projects, water management and agricultural practices are excluded as part of our assumptions, due to the detrimental effects they would have on the area. Our team has highlighted the prohibition of construction, as the construction industry is the greatest contributor to Shoalhaven's economic output, accounting for 19.48% of total output

(Shoalhaven City Council n.d.). Since the newly revegetated area has been specifically designated for the conservation of the flagship species we've identified, our team assumes that the economic interests do not take precedence over the conservation of these species and biodiversity in the area. Hence, our team assumes that no human intrusions or disturbances will occur in the newly revegetated area.

3. **The revegetated flora species mature without any obstructions.**

   Our team assumes that the revegetated flora species will mature to adult age without any obstructions to suit the environment for the flagship species identified and other endangered shorebird species.

Criteria Used to Measure Success

For Objective 1 and Objective 3, they were used as supporting data exploration for the other objectives, hence a measurement of success is impossible to determine due to the fact that there are no error measurements to be evaluated. Hence, the criteria of success for these 2 objectives will heavily rely on the validity, reliability and accuracy of the second-hand data collected from ALA.

The criteria we've used to measure the success of our second object is to compare the predicted values we've produced against an external projection model. We built a time series forecasting model on average rainfall and temperature based on data collected from 2001 - 2023 and will compare it against an existing climate prediction model adopted by the NARCliM initiative. The primary measure of success for our forecasting endeavor is the identification of a model that minimizes all three key metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

The criteria we've used to measure the success of our fourth objective is the number of spatial points where a species has at least a 50% chance of occurring. This has been based on visual analysis of the predicted species presence heatmaps for each species obtained from the Generalised Linear Model (GLM). Species which have met this criterion have then been evaluated based on the number of spatial points that lie above the GLM threshold. The GLM threshold has been computed by R, and species that best meet this criterion have again been selected by visual analysis of predicted presence/absence maps for each species.

Plan

Approach to Problem

The major stages of this project include:

Stage 1: Literature Research and Data Collection and Preprocessing

By conducting this stage, BioNest Analytics has garnered a clear understanding of the relevant, available data which determined the direction of this project.

This stage involved conducting research on endangered species in the Shoalhaven region to determine five flagship endangered shorebird species, *Little Tern, Eastern Curlew, Bar-Tailed Godwit, Great Knot, and Curlew Sandpiper* to focus on improving the biodiversity of Shoalhaven by preventing their extinction. Following this, the objective, *1. Spatial Distribution of the Flagship Shorebird Species* was produced by performing preliminary exploratory data analysis on the data collected from ALA pertaining to the flagship shorebirds species. This enabled the team to focus the revegetation efforts on the Lake Wollumboola and Jervis Bay region by understanding the spread of the flagship species and their ideal climate conditions.

Subsequently, appropriate climate data of the Jervis Bay region was collated and underwent a cleaning process to be used for further data analysis. In addition, this stage involved research on feasible methodologies for further analysis of the data collected and similar prior work done.

Stage 2: Data Analysis and Modelling

This stage primarily involved conducting exploratory data analysis and building models from the data collected from the prior stage, detailed under section *2. Future Climate Conditions*.

BioNest Analytics developed ARIMA models on the climate data of Jervis Bay to study the different climate variables as well as the trends. These models were then used in conjunction with the interactive climate projection map from AdaptNSW to project the future climate variables for the period 2023-2039 in the Illawarra region. Using future climate predictions allows us to better select vegetation to repopulate and determine their climate resistance for such future conditions.

Stage 3: Floral Geolocation

This section involved producing a spatial distribution of potential vegetation, which collects the occurrence of potential vegetation that is suitable for Lake Wollumboola, selected from an area of a wetland. The latitude and longitude data will then be collected and parsed into Stage 4: Floral Species Selection for the benchmark testing on future conditions.

Stage 4: Floral Species Selection

This stage involved producing a species distribution model to determine the resilience of flora species in future climate conditions. The model incorporated the species occurrence within Shoalhaven data from the prior stage and bioclimatic data, including current and forecasted rainfall and temperature averages from the geodata package. The methodology used to produce the species distribution model is detailed under section *4. Flora Species Selection*. The species distribution model enabled the team to select the five most suitable flora species based on the number of geographical locations where the plants are predicted to occur in the future.

Deviation from the Initial Plan

MLR Limitations

The main difference between the current plan and original plan is the model chosen to predict the future climate conditions of the area. The initial plan incorporated a Multiple Linear Regression model (MLR) that would predict the climate conditions for the 2023-2028 period in Jervis Bay. The team faced difficulties in implementing an MLR model to predict the future climate condition, calling for a new methodology instead. The team recognised that the forecasting period 2023-2028 would not show any significant change in the climate conditions, which led to extending the forecasting period to 2023-2039. In a parallel manner, climate projections require extensive meteorological, geographical, and astronomical data over a span of decades. Owing to the colossal amount of data involved and the complex interrelationships between climate variables, most formal climate projections use supercomputers for analysis & computation.

Hence, our team chose to predict the future climate conditions of the region for the period 2023-2039, by building ARIMA models and consulting an external climate projection map.

Inclusion of Objective 2,3 and 4

From the project proposal, our team found that a singular model was insufficient to accurately predict and provide an analysis on what flora to best use for the revegetation scheme, as outlined in the MLR limitations section. Hence, a revision of the methodology of our teams' approach was required for this project.

Our team had finalised on four objectives that enabled us to better answer the question, with each objective providing a different scope on the outcome of our results. They will be used in conjunction with each other, hence providing a backbone for our studies.

Issues Encountered

The following are issues that the BioNest Analytics team faced during our project:

1) Incorporation of Multiple Linear Regression Model

   In the Stage 2 period, BioNest Analytics encountered the problem of predicting the future climate conditions of Jervis Bay by the MLR method. After consideration, our team adapted our approach to predicting future climate conditions by incorporating ARIMA models in conjunction with an external climate projection model.

2) Correlation between the Climate Variables

   In the data analysis period in Stage 2, BioNest Analytics found that there was little correlation between the tidal variables and temperature. Since the external projection model models future temperature conditions, the little correlation makes predicting future tidal conditions difficult. Through external climate models and a deeper exploration of the flora species, we have chosen our flora species based on our projections of future temperature and rainfall with an ARIMA model.

3) Devising a Data-Driven Approach for Native Flora Species Selection

   BioNest Analytics encountered the difficulty of devising a data driven approach to determine appropriate native flora species for our revegetation scheme. To overcome this, a species distribution model based on a GLM was developed to predict suitable regions and species presence in future climate conditions. Current and forecasted climate data was sourced from WorldClim and suitable species were selected from visual analysis of plots based on the computed GLM threshold.

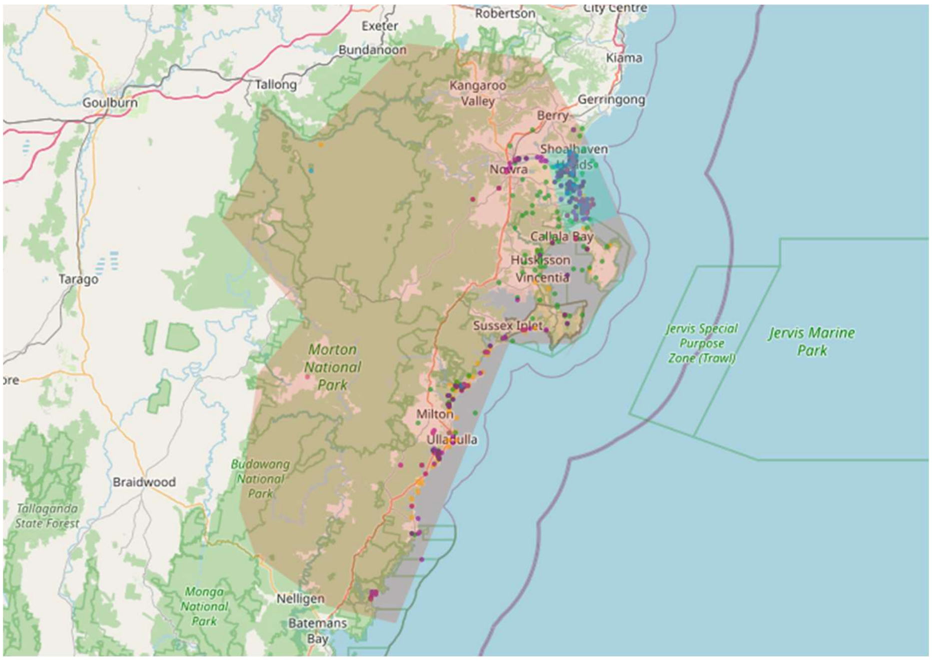
4) Incorporation of Forecasted Data from Objective 2 to Objective 4

   BioNest Analytics encountered difficulties in embedding the forecasted climate data found in Objective 2 into the species distribution model in Objective 4. To overcome this, forecasted climate data was obtained from WorldClim and used in the GLM. We decided to use the range of forecasted climate conditions from Objective 2 as well as analysis of species maps in Objective 4 to select native flora species most suitable to survive in the region.

Findings

1. Spatial Distibution of Flagship Shorebird Species

By analysing the spatial distribution of the flagship endangered shorebird species, the revegetation efforts can be focused on a sub-region of Shoalhaven. This was achieved through mapping the 5 flagship species through the spatial portal in ALA. Figure 1 displays the spatial distribution of the species within Shoalhaven, the orange shaded region. By observation, the species appear to centralise around Lake Wollumboola, the aqua shaded region in Figure 1.



Color	Species name/Area name
Aqua shaded region	Lake Wollumboola
Orange shaded region	Shoalhaven
Purple dot	Bar-Tailed Godwit
Yellow dot	Little Tern
Green dot	Eastern Curlew
Blue dot	Great Knot
Pink dot	Curlew Sandpiper

Figure 1: Spatial Arrangement of Five Flagship Endangered Shorebird Species in Shoalhaven

	No. of Occurences in			% Spread	
Species Name	Lake Wollumboola	Shoalhaven Area	Australia	% Spread in Shoalhaven compared to Australia	% Spread in Lake Wollumboola compared to Shoalhaven
Great Knot	381	384	41174	0.009326274	0.9921875
Curlew Sandpiper	63	664	85232	0.007790501	0.962349398
Bar-tailed Godwit	3249	3575	89985	0.036106018	0.908811189
Far Eastern Curlew	2257	2519	78629	0.032036526	0.895990472
Little Tern	2346	2604	45712	0.056965348	0.900921659

Figure 2: No. Of Occurrences in Selected Regions

To cement the decision of choosing Lake Wollumboola as a case study, we pulled the occurrences of these flagship species and compared the two areas. We observed that 85-90% of the population of the selected endangered shorebird species that are in Shoalhaven reside in Lake Wollumboola, as displayed in Figure 2. The full spatial distribution in ALA is available in the references section.

Hence, the decision to choose Lake Wollumboola was done based on the data analysis of the spatial distribution of flagship.

2.1 Future Climate Conditions: ARIMA Model

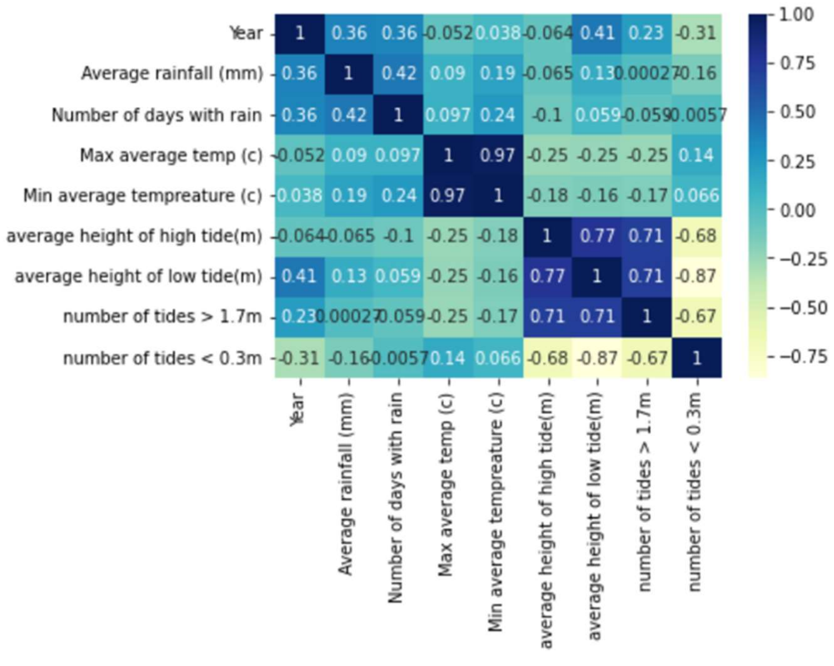


Figure 3: Correlation Matrix of Climate Variables

As shown in the above figure, there is little correlation between temperature, rainfall and tides. Therefore, we will use ARIMA model instead to predict future climate data. This choice is informed by its ability to capture intricate time-dependent patterns in the data. Leveraging two decades of historical climate data (2001-2023) from the Bureau of Meteorology, our ARIMA models, optimized with Root Mean Square Error (RMSE) metrics, provide monthly maximum value forecasts for each year.

ARIMA Model

ARIMA, which stands for Autoregressive Integrated Moving Average, is a popular and widely used statistical method for analyzing and forecasting time series data. It is a powerful tool for understanding, modeling, and predicting various types of time-dependent data. ARIMA models can capture a wide range of temporal structures such as trends, seasonality, and irregular fluctuations in the data. ARIMA models are effective for making short to medium-term forecasts based on historical time series data.

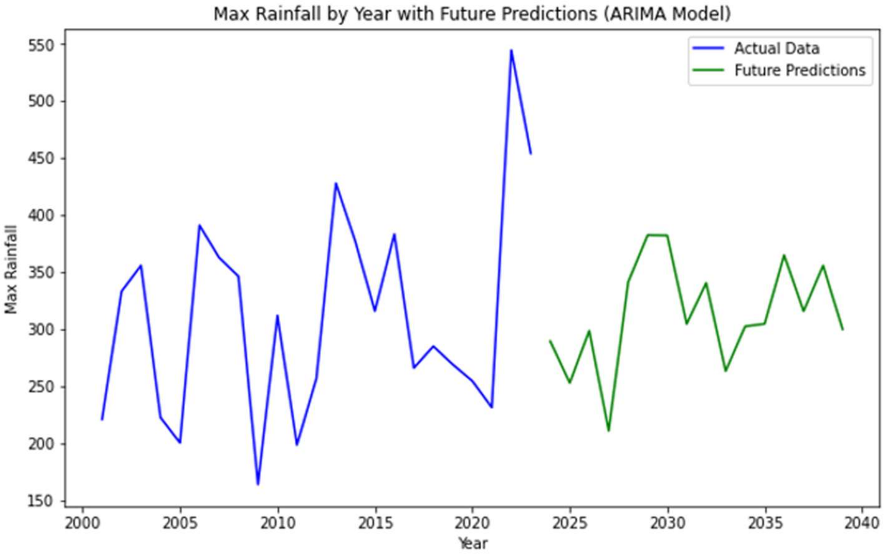
Methodology

Data Cleaning: The climate data must be cleaned as the first stage. To handle null values, the group sample mean for the corresponding month is substituted. By doing this, the temporal structure of the data is preserved, and missing values are represented more accurately.
Data Selection: Only the maximum values for each month are taken into consideration for analysis in order to handle seasonality and concentrate on relevant climate factors. This reduction captures the most important features of monthly climate patterns while streamlining the modelling procedure.
Data Splitting: A 4:1 ratio is maintained when the dataset is divided into training and test sets. This helps to validate the model's prediction capabilities by ensuring a thorough assessment of the model's performance on unobserved data.
ARIMA Model Orders Determination: The orders of the ARIMA model play a critical role in its forecasting accuracy. A two-fold approach is taken to determine these orders:

- Manual Testing: In the initial stages of research, several orders are carefully tested in accordance with subject expertise and data-driven insights. This stage establishes a baseline understanding and functions as a qualitative assessment.
- Auto Order Selection: After that, an automatic order selection procedure is used to maximize the model. This stage refines the manual research with data-driven methods by employing algorithms to test and find the best orders in a methodical manner.

Final Model Selection: The final step in the process is to combine the knowledge gained from both auto order selection and manual testing. The ARIMA model selected for future climate data prediction is the one that performs best when qualitative and quantitative evaluations are taken into account.

Time Series Forecasting

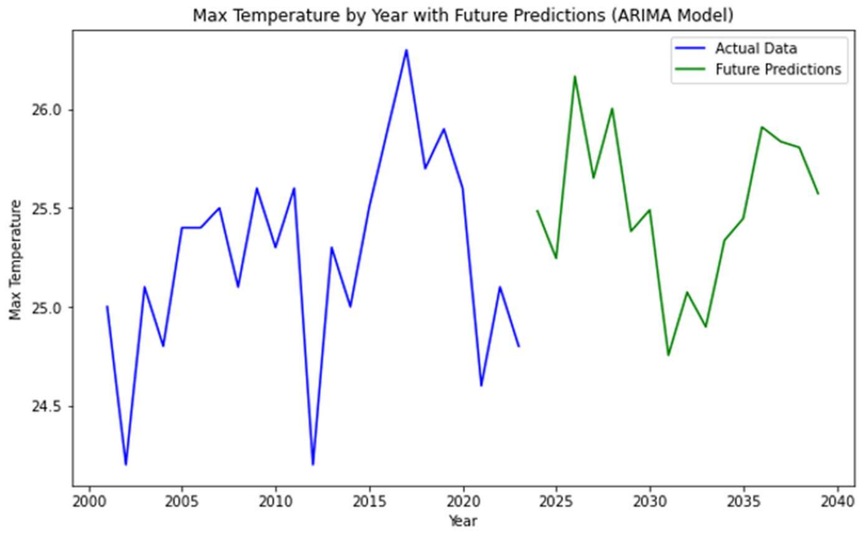



```
mae - manual: 0.24946389056903034
mape - manual: 0.041899519193116644
rmse - manual: 0.3009647742317148

mae - auto: 0.30213174296659184
mape - auto: 0.050397604516272766
rmse - auto: 0.3660018211449392
```

Figure 4: Max Rainfall ARIMA Manual Model

Examining our ARIMA model in the figure above, we project anticipated monthly rainfall (2024-2040) to be between 200-400 mm. Figure 4 indicates a stable pattern at around 300 mm, guiding the rainfall conditions for optimal plant choices. After log transformation, the RMSE (0.3) of the manual model surpasses the auto order, enhancing model performance.




```
mae - manual: 0.0031287733106051263
mape - manual: 0.0009744396328719526
rmse - manual: 0.0031287733106051263
```

```
mae - auto: 0.015130947030062902
mape - auto: 0.004712452135475365
rmse - auto: 0.015130947030062902
```

Figure 5: Max Temperature ARIMA Manual Model

As exhibited in Figure 5, temperature forecasts reveal an upward trend (2024-2040), ranging from 24.7-26.3°C. As indicated in the above figure, optimal temperature conditions around 25.5°C facilitate the plant selection process. After log transformation, the RMSE (0.003) of the manual model outperforms the auto order, highlighting enhanced predictive accuracy.

2.2 Future Climate Conditions: External Climate Change Projections Model

Climate projections require extensive meteorological, geographical, and astronomical data over a span of decades. Owing to the colossal amount data involved and the complex interrelationships between climate variables, most formal climate projections use supercomputers for analysis & computation. Hence, for our project we've consulted an external projection model to anticipate future climate conditions.

In 2015, the NSW Office of Environment and Heritage launched the NSW and Australian Regional Climate Modelling (NARCliM) initiative with the assistance of pertinent professional scholars & researchers from the Australian National University (ANU). The initiative's main objective is to generate Australia-wide surfaces & grids for various climate variables, changes in temperature & rainfall included.

Methodology

Below is a stepwise summarized description of the methodology adopted by the NARCliM initiative.

1. Standardization of monthly mean climate data lays the foundation for the subsequent processes.
2. Fitting elevation-dependent climate surfaces to the standardized monthly mean data. This is done using the ANUSPLIN (Hutchinson & Xu 2013) thin plate smoothing spline package.
3. Calculation of fine-scale gridded monthly mean climatologies for not only the baseline period 1990–2009 but also future periods, specifically 2020–2039, the near future, and 2060–2079, the far future. This is achieved by applying the ANUCLIM (Hutchinson & Xu 2011) package to elevation-dependent climate surfaces and the monthly mean data at 9-second and 0.01-degree spatial resolution. The researchers chose to use regular grid digital elevation models at 9-second and 0.01-degree spatial resolution. To convert the 1990–2009 climatologies to the projected fine-scale future climatologies, "delta" change factor downscaling method was used, as described by CSIRO and Bureau of Meteorology (2015).
4. The summary results of the regression process for daily minimum and maximum temperature, pan evaporation, and rainfall are highlighted. The regression process increased the number of stations with standard period mean estimates, improving the accuracy of the interpolated surfaces.
5. A different process was used for solar radiation due to the smaller solar radiation network. The method incorporates both astronomical controls on solar radiation and its dependence on rainfall occurrence at the monthly time step.

Projection Outcomes

Pictures of the NARCliM initiative's main accomplishment, which is the interactive climate change projections map, are presented below with interpretations.

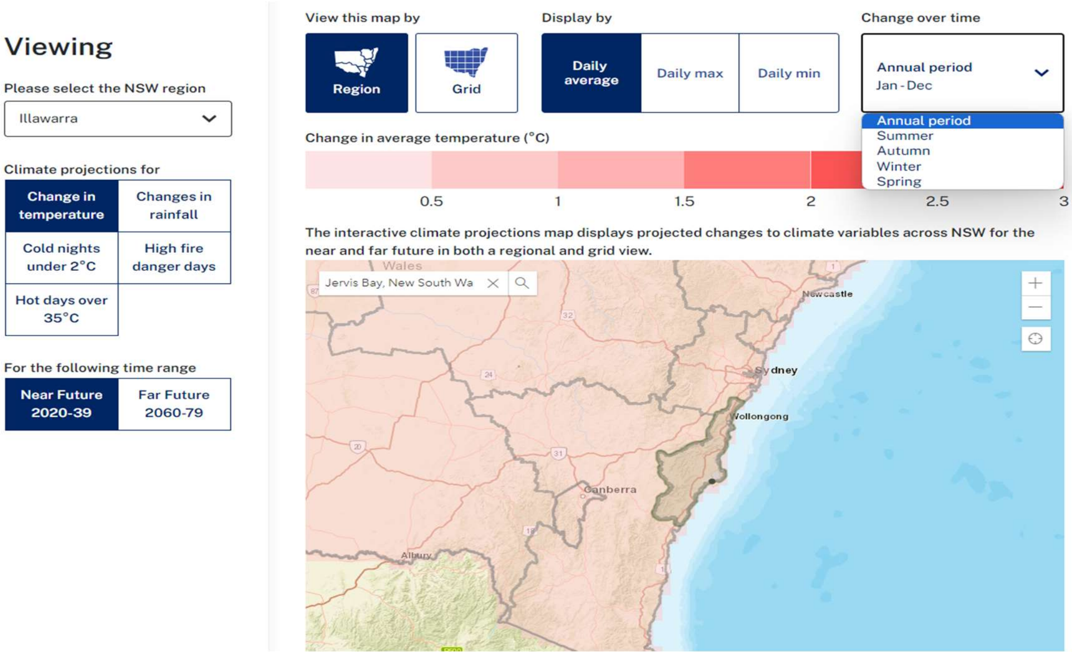


Figure 6: Interactions on the Regional Map

Users of the interactive climate change projections map may view by region or grid. To obtain an overview of the Shoalhaven, view this map by "region" and select "Illawarra" as the NSW region. Looking at the region colored in a darker shade on the map, users can confirm that Shoalhaven is included in Illawarra geographically. As exhibited in Figure 6, the dot pinpoints Jervis Bay, our group's specific area of interest, but it doesn't do much other than indicating Jervis Bay's location because the interactive map narrows down to climate projections of an NSW region at best, meaning that smaller territories like Jervis Bay don't have their specific climate projections.

To the left of the map, users may also choose from 5 different climate indicators and 2 separate time ranges. Our group is primarily interested in short to medium term changes in temperature & rainfall since they exert substantial external influences on the growth & well-being of local flora and fauna so here, we select "change in temperature" and "Near Future 2020-39" as an example.

Above the bar indicating daily change in temperature, which darkens the shade of red as temperature increases from 0 to 3°C (degrees Celsius), users may choose to have the daily average, maximum, or minimum on display. On a side note, most regions in Figure 6 are shaded in light red, indicating a relatively small change in daily average temperature across the year, below 1°C. This shade may grow darker or lighter depending on which season users choose to be on display.

In Figure 6, based on the options selected, users are looking at the change in daily average temperature across the year in the Illawarra region in 2020-2039.

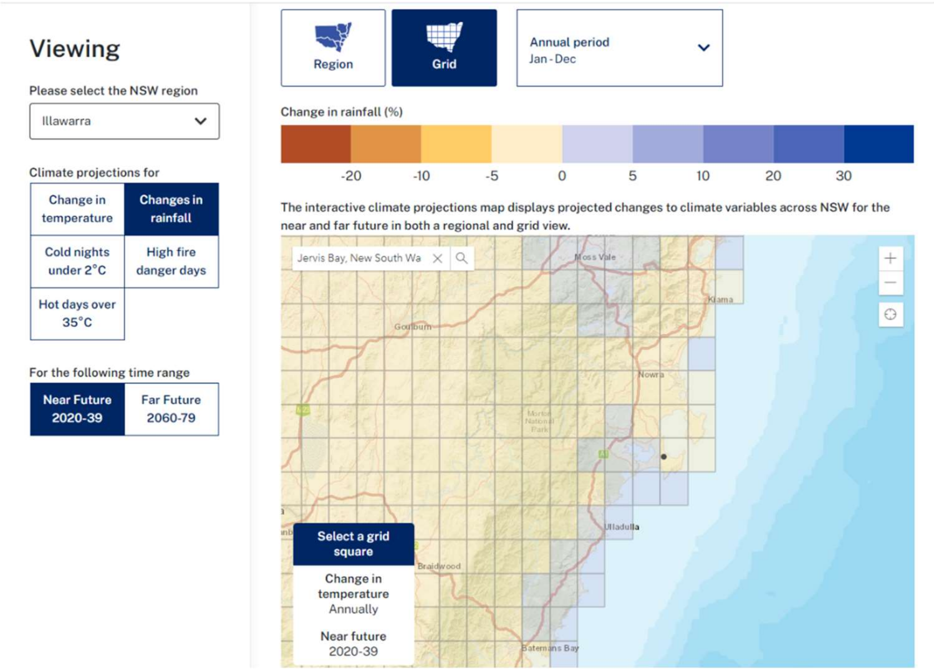


Figure 7: Interactions on the Gridded Map

The interactive map by grid view provides a more detailed look of climate conditions in and around the Jervis Bay territory. Here, "change in rainfall" is selected as the climate variable. According to the bar indicating average change in rainfall, the darker the shade of blue, the greater the increase in rainfall; the darker the shade of red, the greater the decrease in rainfall.

As displayed in Figure 7, grids colored in blue are mostly distributed along the coast, with few situated inland, which is consistent with the common belief that coastal territories, Jervis Bay included, receive more rainfalls because of their proximity to oceans. In addition, the Great Dividing Range runs the entire length of Australia's eastern coastline through NSW. The presence of this mountain range close to the coast can cause orographic rainfall, a phenomenon where moist air from the ocean is forced to rise by the mountains. As the air rises, it cools and condenses, leading to precipitation.

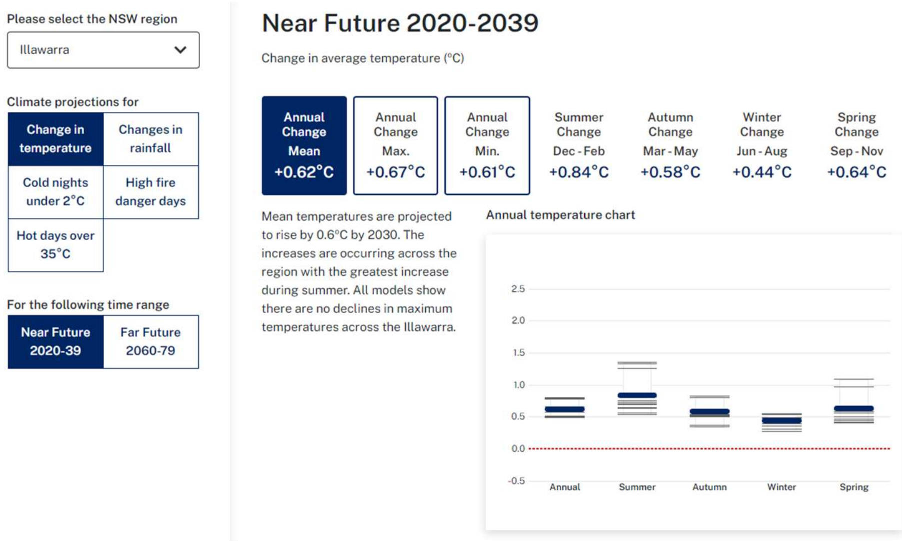


Figure 8: Quantitative Temperature Predictions under the Interactive Map

The annual temperature chart on the bottom right of Figure 8 presents the average results from 12 different models for the Illawarra region. The white bar's length represents the range of values obtained from the 12 models for Illawarra. Each thin black horizontal line depicts the average value from an individual model. It's important to note that there are 12 such lines for every bar, and they don't correspond to any specific location within Illawarra. The thin blue line indicates the combined average of all 12 models for Illawarra. The dotted red horizontal line, corresponding to 0.0 on the vertical axis, signifies no variation in the data.

From Figure 8, the NARCliM initiative predicts a 0.62°C increase in mean annual temperature in the Illawarra region by 2039, with 0.67°C being the maximum and 0.61°C being the minimum. The projected average seasonal temperature increases in summer, autumn, winter, and spring are 0.84°C, 0.58°C, 0.44°C, and 0.64°C respectively.

Among the listed values, 0.84°C increase in summer is the most alarming one because average temperature rising by almost 1 degree Celsius can have pronounced influences on the local environment and the revegetation scheme. According to projections of the NARCliM, the number of extreme hot days over 35°C is expected to increase by 1.9 per year, 1.7 of them occurring in summer.

Deane's Boronia (Boronia Deanei), which is on the list of our group's recommendations, grows in shaded, moist environments (rainforests, gullies, etc.) and prefers cooler weather in addition to more consistent microclimates. Drastic increases in temperature and extreme hot days could be detrimental to Deane's Boronias' growth & survival.

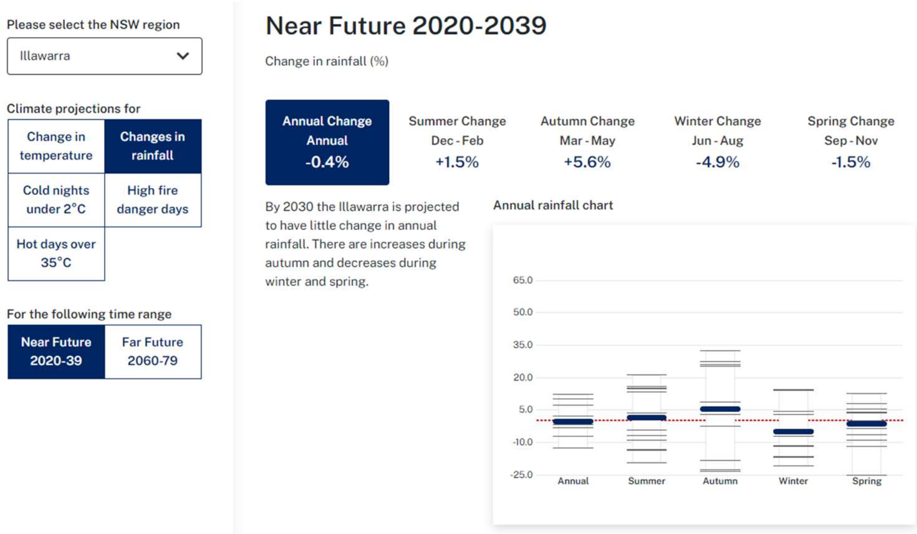


Figure 9: Quantitative Rainfall Predictions under the Interactive Map

From Figure 9, the NARCliM initiative predicts an overall 0.4% decrease in mean annual rainfall in the Illawarra region by 2039. The average seasonal rainfall is anticipated to increase by 1.5% & 5.6% in summer & autumn respectively but decrease by 4.9% & 1.5% in winter and spring respectively.

Combining these results with our group's list of species recommendations, half of the plants on the list are not likely to be remarkably impacted by the projected changes in rainfall. Hottentot Fig, one of our species recommendations, are succulents adapted to store water, leading to their higher resilience to changing rainfall patterns. Grey Mangroves (Avicennia Marina), one of our selected species' primary water sources is tidal influx instead of rainfall.

As for the other half, Deans Boronia (Boronia Deanei) prefers cooler, shaded, and consistent environments. Reduced rainfall in winter & spring may stress the plant. Moreover, Boronia Deanei is inherently sensitive to changing patterns in weather. Broad-Leaved Banksias (Banksia Robur) are drought-tolerant to an extent, but depending on the timing, their unique pollination cycles might be undermined by the changing rainfall patterns.

However, the predicted alterations in rainfall in the Illawarra region are insignificant in general. They are not expected to have pronounced effects on local flora & fauna.

Summary of External Climate Predictions

In closing, based on the projected changes in temperature and rainfall, the Illawarra region is experiencing an incremental warming trend and an alternating trend in precipitation. Each of the increases in temperature can have specific and cumulative impacts on the local environment and the species within, whereas the changes in rainfall may not have significant impacts. It's essential to consider these changes in the context of broader global warming trends and their potential long-term consequences.

3. Flora Geolocation

Background

To select optimal plants that will help improve the conditions in wetlands as well as climate resistance, we utilized ALA to collect our data. Firstly, we analyzed a wetland area within Shoalhaven and collected the occurrence of each different floral species. From this, we obtained a total of 151 plant species.

	A	B	C	D	E	F	G	H	I	J	K
1	names_ar	Species N	Scientific	Taxon Rar	Kingdom	Phylum	Class	Order	Family	Genus	Vernacular Name
2	Avicennia	Avicennia	(Forssk.) \	species	Plantae	Charophy	Equisetop	Lamiales	Acanthace	Avicennia	Grey Mangrove
3	Casuarina	Casuarina	Sieber ex	species	Plantae	Charophy	Equisetop	Fagales	Casuarina	Casuarina	Swamp Oak
4	Lantana ca	Lantana ca	L.	species	Plantae	Charophy	Equisetop	Lamiales	Verbenac	Lantana	Common Lantana
5	Guioa sen	Guioa sen	(F.Muell.)	species	Plantae	Charophy	Equisetop	Sapindale	Sapindace	Guioa	Wild Quince
6	Pittospor	Pittospor	Vent.	species	Plantae	Charophy	Equisetop	Apiales	Pittospora	Pittospor	Snowdrop Tree (Ic
7	Podocarp	Podocarp	R.Br. ex Er	species	Plantae	Charophy	Equisetop	Pinales	Podocarpa	Podocarp	Brown Pine
8	Syzygium	Syzygium	(Poir.) Nie	species	Plantae	Charophy	Equisetop	Myrtales	Myrtaceae	Syzygium	
9	Ripogonu	Ripogonu	R.Br.	species	Plantae	Charophy	Equisetop	Liliales	Ripogona	Ripogonu	White Supplejack
10	Endiandra	Endiandra	Nees	species	Plantae	Charophy	Equisetop	Laurales	Lauraceae	Endiandra	Corkwood
11	Cryptocar	Cryptocar	R.Br.	species	Plantae	Charophy	Equisetop	Laurales	Lauraceae	Cryptocar	Native Laurel
12	Juncus kra	Juncus kra	(Buchena	subspecie	Plantae	Charophy	Equisetop	Poales	Juncaceae	Juncus	Sea Rush
13	Sarcocorn	Sarcocornia quinque		subspecie	Plantae	Charophy	Equisetop	Caryophyl	Chenopoc	Sarcocorn	Beaded Glasswort
14	Claoxylon	Claoxylon	Baill.	species	Plantae	Charophy	Equisetop	Malpighia	Euphorbia	Claoxylon	Brittlewood
15	Livistona a	Livistona a	(R.Br.) Ma	species	Plantae	Charophy	Equisetop	Arecales	Arecaceae	Livistona	Cabbage Tree Paln
16	Oplismen	Oplismen	(L.) P.Beat	species	Plantae	Charophy	Equisetop	Poales	Poaceae	Oplismen	Australian Basket-
17	Suaeda au	Suaeda au	(R.Br.) Mo	species	Plantae	Charophy	Equisetop	Caryophyl	Chenopoc	Suaeda	Austral Seablite
18	Zostera|h	Zostera	L.	genus	Plantae	Charophy	Equisetop	Alismatale	Zosterace	Zostera	KarepÅ
19	Eupomatia	Eupomatia	R.Br.	species	Plantae	Charophy	Equisetop	Magnolial	Eupomatia	Eupomatia	Copper Laurel
20	Glochidio	Glochidion ferdinan		variety	Plantae	Charophy	Equisetop	Malpighia	Phyllantha	Glochidion	
21	Hydrocoty	Hydrocoty	Lam.	species	Plantae	Charophy	Equisetop	Apiales	Araliaceae	Hydrocotyle	
22	Marsdenia	Marsdenia	R.Br.	species	Plantae	Charophy	Equisetop	Gentianal	Apocynac	Marsdenia	Common Milk Vin
23	Pittospor	Pittospor	W.T.Aiton	species	Plantae	Charophy	Equisetop	Apiales	Pittospora	Pittospor	Rough-fruit Pittos
24	Banksia in	Banksia in	L.f.	species	Plantae	Charophy	Equisetop	Proteales	Proteacea	Banksia	Honeysuckle
25	Cenchrus	Cenchrus	(Hochst. e	species	Plantae	Charophy	Equisetop	Poales	Poaceae	Cenchrus	
26	Clerodenc	Clerodenc	(Vent.) R.I	species	Plantae	Charophy	Equisetop	Lamiales	Lamiaceae	Clerodenc	Lolly Bush
27	Eucalyptu	Eucalyptu	Sm.	species	Plantae	Charophy	Equisetop	Myrtales	Myrtaceae	Eucalyptu	Southern Mahogar

Figure 10: Potential Flora Candidates for the Revegetation Scheme

Then, we mapped these species individually through the spatial portal of ALA. Since Lake Wollumboola is a coastal area, such candidate species are required to be resistant to high salinity levels, as well as irregular rainfall patterns, observed in the second objective. Doing this will allow us to observe the spread of each species and determine its climate resilience.

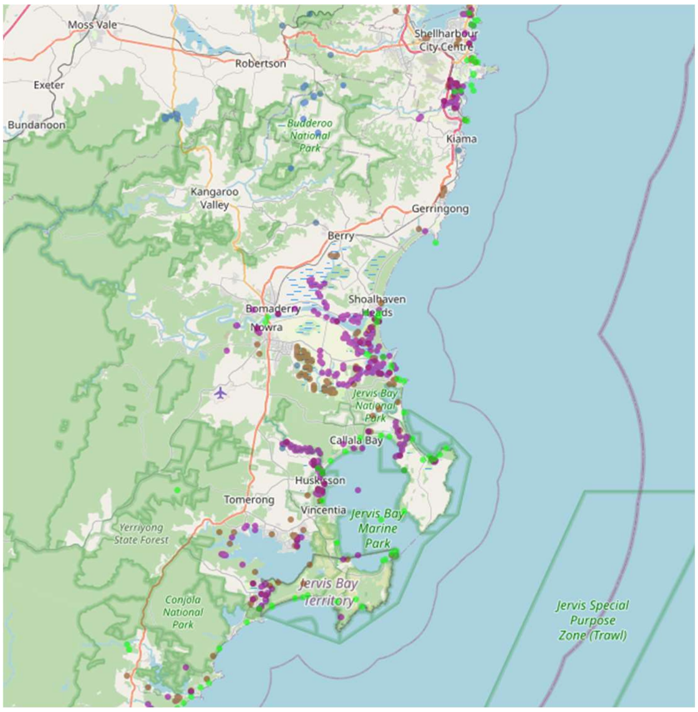


Figure 11: Spread of 5/151 Candidate Floral Species to be Revegetated within Shoalhaven

From these points, we exported the latitude and longitude into a spreadsheet and this data will be passed into Objective 4 to be used to select our 5 floral species to revegetate.

4: Selection of Flora Species

To determine the resilience of the native flora species explored in future climate conditions, a species distribution model was created. This model was created using species occurrence data sourced from ALA and spatial data extracted from the spatial mapping of native flora species conducted in Objective 3. The model has been constructed using a GLM, which is a combination of conventional linear regression models for a continuous response variable. This model was chosen to prevent overfitting of the model, ensuring equal and optimal performance on both training and testing data.

Methodology

The species distribution model was constructed as follows:

1. *Data Cleaning* – Before commencing the analysis, all duplicate and NULL data values were removed. Species occurrences at spatial points outside of the Shoalhaven region were also removed, separating the remaining data for each species into individual data files. Only species with at least 5 or more occurrences were chosen, reducing the number

of species to be analysed from 20 to 11. Data on all 19 bioclimatic variables available on WorldClim was downloaded using the geodata package and stored as a data frame.

2. *Initial Species Occurrence Map* – A geographical map of the Shoalhaven region was sourced using geodata package, serving as a canvas for visualising plant distribution. The species occurrence data for one species was plotted on the map to ensure the data has been imported correctly as seen in Figure 12.

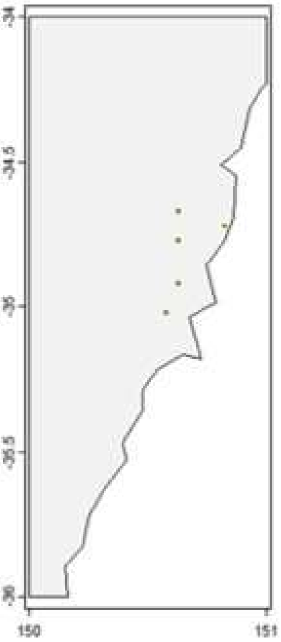


Figure 12: Map of Banksia Robur Occurrences

3. *Inclusion of Pseudo-Absence Points* – Also known as background points, 1000 pseudo-points were randomly sampled and added to the data. Pseudo-absence points have been incorporated to account for all available environments in the Shoalhaven region and to compare against observed presence environment when building the model. Bioclimatic data was used to determine the spatial resolution of the pseudo-absence points and added to the data. A binary presence/absence column was added to the data, 1 indicating presence and 0 indicating absence. The data with pseudoabsence points was plotted as seen in Figure 13.

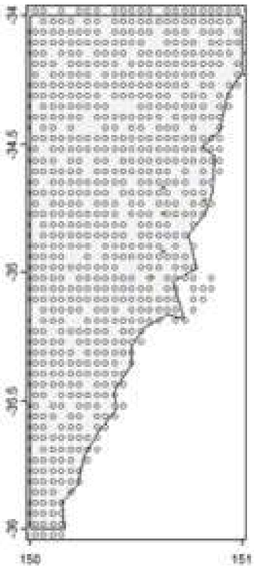


Figure 13: Map of Pseudo-absence Points

4. *Model Building* – To build the model, 20% of the data was randomly allocated to a training dataset and the remaining 80% to a testing dataset. The model was then built using the GLM with the training dataset, using the binary presence/absence values as the continuous variable. Predicted values were obtained from this dataset and plotted as a heatmap as seen in Figure 14. A higher value indicates greater probability the species is present in that location.

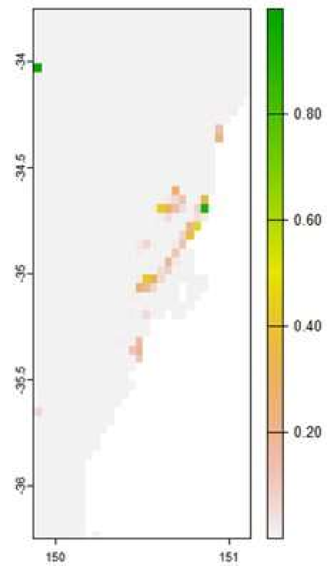


Figure 14: Heatmap of Predicted Values

5. *Incorporation of GLM Threshold* – The model was evaluated using the training data and predicted values were obtained. Using these predicted values and the "thresholds" element of "glm_eval" command, a threshold cutoff was determined to find whether a particular location was suitable or not for the selected species. Only spatial points where occurrence probability is higher than the threshold were plotted, as seen in Figure 15. The green areas indicate regions that are suitable for the selected species in future climate conditions.

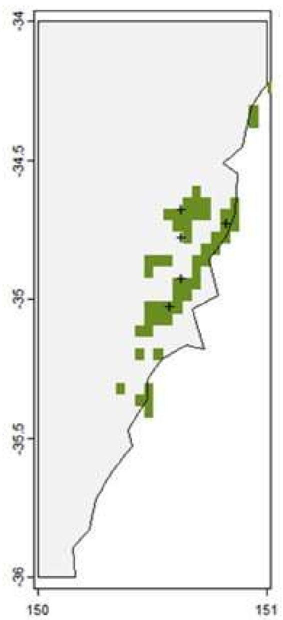


Figure 15: Map of Spatial Points Greater than Threshold

6. *Forecasting Species Occurrences* – Forecasted bioclimatic data for the 2024–2039 period in Shoalhaven was extracted from World Climate Data using the geodata package. Steps 4 and 5 were repeated using the forecasted climate data, as seen in Figure 16.

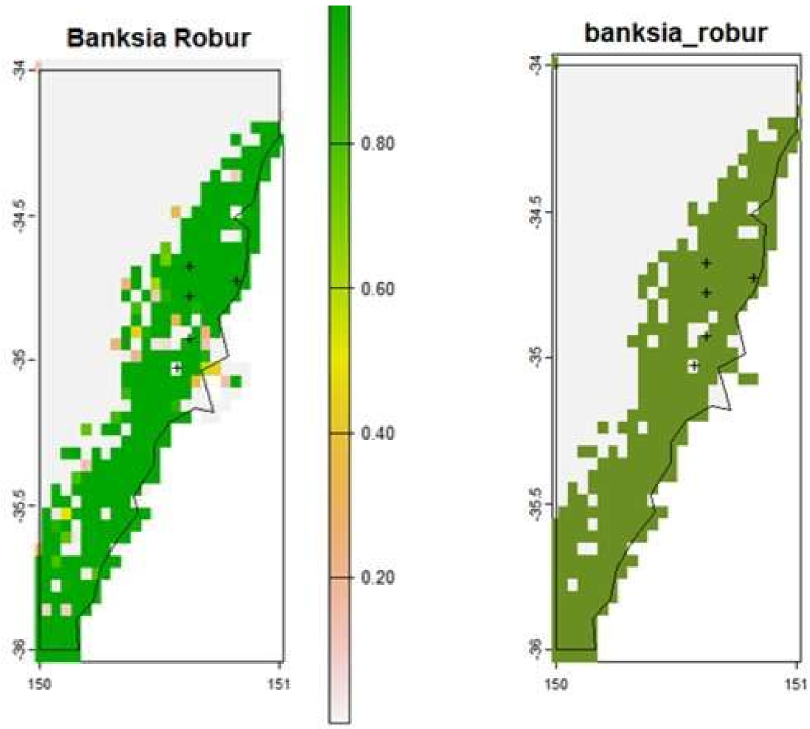


Figure 16: Maps of Banksia Robur

Steps 2-6 were then repeated for all 11 species and plots created for each species.

Analysis of Model Output

The predicted species presence heatmaps and predicted presence/absence map based on GLM threshold output from Step 6 were used to determine native flora species suitable for future climate conditions. To do so, these plots were visually analysed and qualitatively evaluated against the following criteria.

Number of spatial points where a species has at least a 50% likelihood to occur – through visual analysis, species with a high number of points where a species has a 50% or higher chance of occurring were selected, evidenced by the yellow and green points on the heatmap. From this analysis, Carpobrotus glaucesens, Banksia Robur, Boronia Deanei subspecies acutifolia, Boronia deanei, Avicennia marina, Avicennia marina subspecies australa, Carpobrotus edulis and Carpobrotus were deemed to have a suitable number of spatial points meeting this criterion. However, Carpobrotus aequilaterus, Carpobrotus rossii and Carpobrotus chilensis did not meet this criterion and were subsequently discarded as potential flora species for revegetation.

Species with high number of spatial points above the GLM threshold – the plots of forecasted species presence and absences based on the GLM threshold for the eight species chosen above were visually analysed. Green regions indicate areas lying above the GLM threshold that represent suitable future climate conditions for a particular species. Plots with a greater number of green regions indicate a species is more likely to be successful across the Shoalhaven region. Based on this analysis, Banksia Robur, Boronia Deanei, Avicennia Marina, Carpobrotus Edulis, and Carpobrotus Glaucescens were seen to have the highest number of green regions, indicating these species will be most suitable in future climate conditions in the Shoalhaven based off predictions from this model.

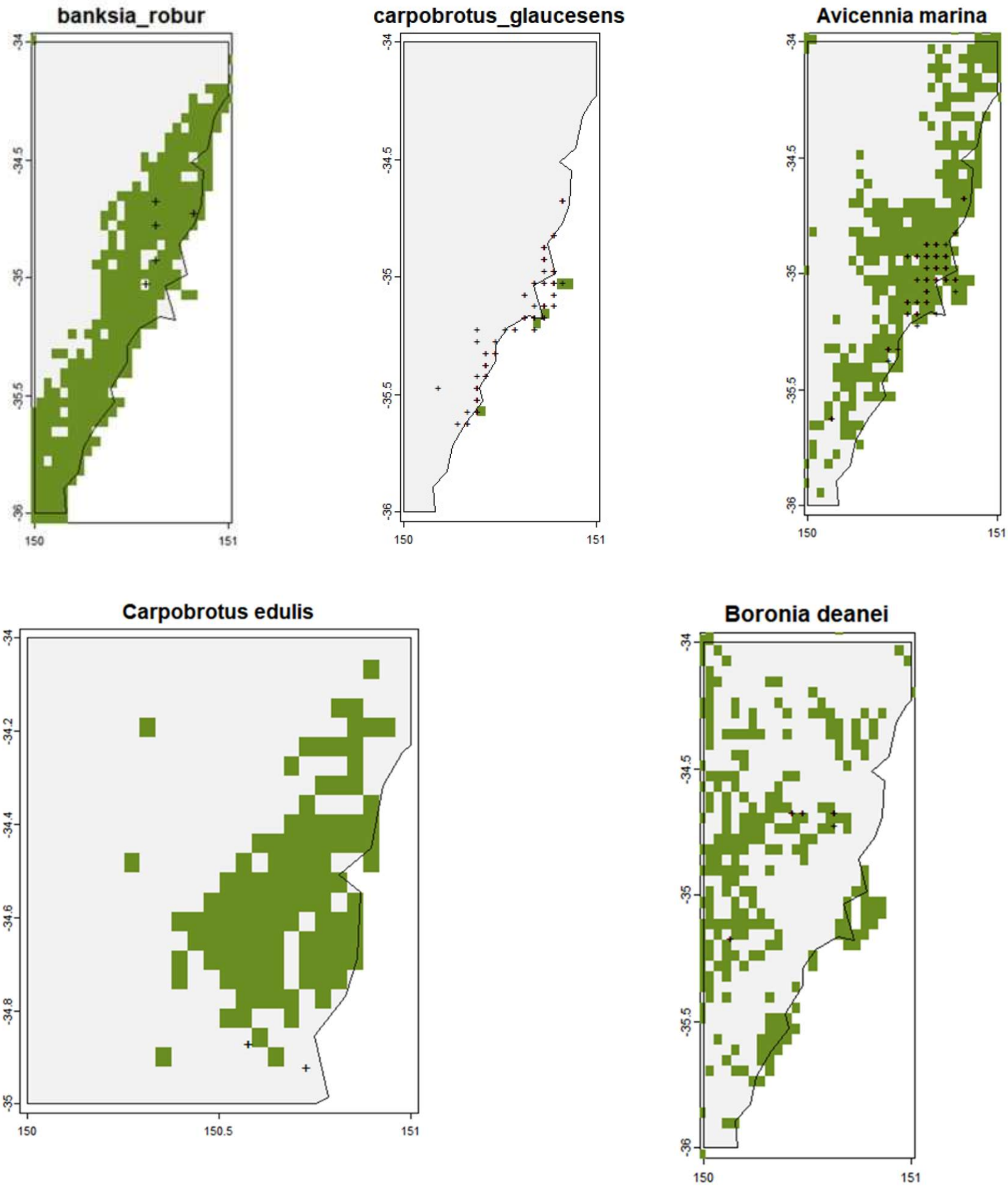


Figure 17: Maps of Selected Flora Species

Discussion

Summary of Results

Based on the ARIMA model, by 2040, the temperature is anticipated to increase, with predictions showing a range from 24.7°C to 26.3°C, with optimal conditions around 25.5°C. Rainfall predictions expect a monthly range between 200-400 mm, with a stable pattern around 300 mm.

According to the NARCliM collaboration's projections, the average yearly temperature will increase by 0.62°C and the average yearly rainfall will decrease by 0.4% by 2039.

Based on the GLM, the Banksia Robur, Boronia Deanei, Avicennia Marina, Carpobrotus Edulis, and Carpobrotus Glaucescens were found to be the most suitable species for future climate conditions. In addition to the GLM, the suitability of these species is supported by the predictions from the ARIMA Model, NARCLiM collaboration's projections and qualitative background research on the native flora species.

Limitations and Lessons Learnt for Future Analyses

Many researchers don't grant visitors immediate access to their intellectual properties, which are climate models in this case, limiting our choices since it is imprudent to use the results generated by a model without knowing what exactly is going on inside. It normally takes researchers weeks to respond to requests for access to their models, meaning that analysts should express their interests further in advance given the timeframe.

Despite confidentiality of the model, what sets the NARCliM project apart is the provision of a detailed description of methodology employed and sponsorship from ANU and Australian governmental entities, enhancing its credibility, which motivated us to eventually choose to cite the NARCliM initiative over other options.

However, the outcomes of the chosen NARCliM project don't offer temporal precision. Limited by the inherent nature of climate projections, which span from decades to centuries, predicting the overall seasonal change in temperature/rainfall of a region in the next 20 years (2020-2039) is as specific as the NARCliM project could get, meaning that obtaining the exact weather forecast of a specific month or season in the future is impossible.

In addition, another limitation our team faced was during the data processing stage. Our team integrated various types of statistics, including geographic information, species distribution data, and climate data, which was a challenging task. To harmonize numerous data sources with varying formats, scales, and accuracy levels, minor amendments must be made, which may have impacted the dependability of our work.

To address this issue, in future endeavours, analysts could establish or adopt standardized data formats and protocols when processing data gathered from different sources. If amendments are inevitable, implement rigorous quality control measures at each stage of data collection and recording to ensure maximum consistency & integrity in the data.

Lastly, another limitation our team faced was a lack of species occurrences data, which could undermine the accuracy of predictions, potentially impacting the effectiveness of species recommendations and the subsequent conservation strategies designed for the endangered shorebird species and their habitat.

Conclusion

Hence, our team at BioNest Analytics advises that the government department undertaking the revegetation project incorporate the following five floral species: Banksia Robur, Boronia Deanei, Avicennia Marina, Carpobrotus Edulis, and Carpobrotus Glaucescens, with caution. Improvements to our approach, such as incorporating more data from licensed sources and more rigorous quality control measures, would enable a more nuanced and accurate assessment of potential species to be considered for revegetation. With these improvements, future analysts could provide invaluable insights for the government revegetation program and similar future programs.

Peer Review

Review 1

We liked how the report included a vast range of references and how detailed the recommendations were for the project – recommend 5 native flora species that can be implemented in a vegetation scheme to prevent the extinction of the flagship species and other endangered shorebird species. We are interested in seeing the complete results for the recommended flora species in the final report.

The research used to determine the flora species recommended holds great promise, but we did notice the description of visuals were too in-depth and focused on just describing the graph, rather than establishing connections to how this can impact the selection of flora species. Further, we noticed it was hard to find sections of the report and we advise a table of contents to make it easier to find contents.

Suggestion

The description of visuals was too in-depth and focused on just describing the graph

Response

BioNest Analytics excluded from the main body of the report unnecessary visualisations produced during exploratory data analysis that did not add value to the report.

Suggestion

It was hard to find sections of the report.

Response

BioNest Analytics included a table of contents at the beginning of the report as a response. We also sectioned out our objectives into clearer outcomes and purposes.

Review 2

The draft of your report looks promising, especially in its use of external data sources. It's great that you already have such solid findings! However, more work needs to be done in the methodology section. Your draft focuses more on the outcome of the model instead of the scientific process behind the model. Focusing on this will enhance the clarity of how conclusions were reached and why the results are plausible. It would be beneficial to provide more information about the original dataset and the external ones, including exploratory data analysis, feature engineering, feature selection and model selection to better understand the foundation of the analysis. Additionally, the report should introduce a measure of success for evaluating the model's performance. This metric will help assess the model's effectiveness, irrespective of the final outcomes.

Suggestion

More work needs to be done in the methodology section. Your draft focuses more on the outcome of the model instead of the scientific process behind the model.

Response

BioNest Analytics expanded on these sections to explicitly detail the methods conducted to produce the findings. We ensured to keep logs of the methodologies used in the appendix and references.

Suggestion

Introduce measure of success for evaluating the model's performance.

Response

BioNest Analytics explicitly stated the measure of success for each model. For example, our team expanded on the criteria of success for objective 2.

Appendices

Appendix A

```
# Rainfall Testing

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_squared_error

from statsmodels.tsa.seasonal import seasonal_decompose

from statsmodels.graphics.tsaplots import plot_acf, seasonal_plot

from statsmodels.tsa.arima.model import ARIMA


df_rainfall = pd.read_csv('Rainfall.csv', parse_dates=True)

df_rainfall_altered = df_rainfall.drop('Annual', axis=1)

df_rainfall_altered = df_rainfall_altered.drop('Product code', axis=1)
```

```
df_rainfall_altered = df_rainfall_altered.drop('Station Number', axis=1)
df_rainfall_altered.head()

df_rainfall_altered['Jan']=df_rainfall_altered['Jan'].fillna(df_rainfall_altered['Jan'].mean())
df_rainfall_altered['Feb']=df_rainfall_altered['Feb'].fillna(df_rainfall_altered['Feb'].mean())
df_rainfall_altered['Mar']=df_rainfall_altered['Mar'].fillna(df_rainfall_altered['Mar'].mean())
df_rainfall_altered['Apr']=df_rainfall_altered['Apr'].fillna(df_rainfall_altered['Apr'].mean())
df_rainfall_altered['May']=df_rainfall_altered['May'].fillna(df_rainfall_altered['May'].mean())
df_rainfall_altered['Aug']=df_rainfall_altered['Aug'].fillna(df_rainfall_altered['Aug'].mean())
df_rainfall_altered['Oct']=df_rainfall_altered['Oct'].fillna(df_rainfall_altered['Oct'].mean())
df_rainfall_altered['Nov']=df_rainfall_altered['Nov'].fillna(df_rainfall_altered['Nov'].mean())
df_rainfall_altered['Dec']=df_rainfall_altered['Dec'].fillna(df_rainfall_altered['Dec'].mean())
df_rainfall_altered['Max_Rainfall'] = df_rainfall_altered.drop('Year', axis=1).max(axis=1)

df_rainfall_altered['Max_Rainfall'] = df_rainfall_altered.drop('Year', axis=1).max(axis=1)

selected_columns = ['Year', 'Max_Rainfall']
df = df_rainfall_altered[selected_columns]
df = df.set_index('Year')
print(df)

df = np.log(df)
df.plot()

msk = (df.index < df.index[-5])
df_train = df[msk].copy()
df_test = df[~msk].copy()
```

```
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
acf_original = plot_acf(df_train)
pacf_original = plot_pacf(df_train, lags=8)

from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(df_train, order=(6,1,2))
model_fit = model.fit()
print(model_fit.summary())

import matplotlib.pyplot as plt
residuals = model_fit.resid[1:]
fig, ax = plt.subplots(1,2)
residuals.plot(title='Residuals', ax=ax[0])
residuals.plot(title='Density', kind='kde', ax=ax[1])
plt.show()

import pmdarima as pm
auto_arima = pm.auto_arima(df_train, stepwise=False, seasonal=False)
auto_arima

forecast_test = model_fit.forecast(len(df_test))
df['forecast_manual'] = [None]*len(df_train) + list(forecast_test)
df.plot()

forecast_test_auto = auto_arima.predict(n_periods=len(df_test))
df['forecast_auto'] = [None]*len(df_train) + list(forecast_test_auto)
df.plot()
```

```
from sklearn.metrics import mean_absolute_error, mean_absolute_percentage_error,
mean_squared_error

mae = mean_absolute_error(df_test, forecast_test)

mape = mean_absolute_percentage_error(df_test, forecast_test)

rmse = np.sqrt(mean_squared_error(df_test, forecast_test))

print(f'mae - manual: {mae}')

print(f'mape - manual: {mape}')

print(f'rmse - manual: {rmse}')


print(f'mae - auto: {mae}')

print(f'mape - auto: {mape}')

print(f'rmse - auto: {rmse}')


# Temperature Testing

df_temperature = pd.read_csv('Temperature.csv')


df_temperature_altered = df_temperature.drop('Annual', axis=1)

df_temperature_altered = df_temperature_altered.drop('Product code', axis=1)

df_temperature_altered = df_temperature_altered.drop('Station Number', axis=1)


#Replace null values based on the group/sample mean

df_temperature_altered['Jan']=df_temperature_altered['Jan'].fillna(df_temperature_altered['Jan'].
mean())

df_temperature_altered['Feb']=df_temperature_altered['Feb'].fillna(df_temperature_altered['Feb']
.mean())

df_temperature_altered['Mar']=df_temperature_altered['Mar'].fillna(df_temperature_altered['Mar'
].mean())

df_temperature_altered['Apr']=df_temperature_altered['Apr'].fillna(df_temperature_altered['Apr']
.mean())
```

```
df_temperature_altered['Jun']=df_temperature_altered['Jun'].fillna(df_temperature_altered['Jun'].
mean())

df_temperature_altered['Oct']=df_temperature_altered['Oct'].fillna(df_temperature_altered['Oct'].
mean())

df_temperature_altered['Nov']=df_temperature_altered['Nov'].fillna(df_temperature_altered['Nov
'].mean())

df_temperature_altered['Dec']=df_temperature_altered['Dec'].fillna(df_temperature_altered['Dec'
].mean())


df_temperature_altered['Max_Temperature'] = df_temperature_altered.drop('Year',
axis=1).max(axis=1)

# Keep only the desired columns

selected_columns = ['Year', 'Max_Temperature']

df = df_temperature_altered[selected_columns]

df = df.set_index('Year')

# Display the updated DataFrame

print(df)


df = np.log(df)

df.plot()


msk = (df.index < df.index[-5])

df_train = df[msk].copy()

df_test = df[~msk].copy()


from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

acf_original = plot_acf(df_train)


pacf_original = plot_pacf(df_train, lags=10)
```

```
from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(df_train, order=(8,1,5))
model_fit = model.fit()
print(model_fit.summary())

import matplotlib.pyplot as plt
residuals = model_fit.resid[1:]
fig, ax = plt.subplots(1,2)
residuals.plot(title='Residuals', ax=ax[0])
residuals.plot(title='Density', kind='kde', ax=ax[1])
plt.show()

import pmdarima as pm
auto_arima = pm.auto_arima(df_train, stepwise=False, seasonal=False)
auto_arima

forecast_test = model_fit.forecast(len(df_test))
df['forecast_manual'] = [None]*len(df_train) + list(forecast_test)
df.plot()

forecast_test_auto = auto_arima.predict(n_periods=len(df_test))
df['forecast_auto'] = [None]*len(df_train) + list(forecast_test_auto)
df.plot()

from sklearn.metrics import mean_absolute_error, mean_absolute_percentage_error,
mean_squared_error
mae = mean_absolute_error(df_test, forecast_test)
mape = mean_absolute_percentage_error(df_test, forecast_test)
rmse = np.sqrt(mean_squared_error(df_test, forecast_test))
```

```
print(f'mae - manual: {mae}')
print(f'mape - manual: {mape}')
print(f'rmse - manual: {rmse}')

mae = mean_absolute_error(df_test, forecast_test_auto)
mape = mean_absolute_percentage_error(df_test, forecast_test_auto)
rmse = np.sqrt(mean_squared_error(df_test, forecast_test_auto))
print(f'mae - auto: {mae}')
print(f'mape - auto: {mape}')
print(f'rmse - auto: {rmse}')

#Rainfall and Temperature Prediction
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA

# 1. Data Preprocessing
# Convert 'Year' to a datetime format if needed
df_temperature_altered['Year'] = pd.to_datetime(df_temperature_altered['Year'], format='%Y')

# 2. Grouping Data
# Extract the maximum values for each year
max_values =
df_temperature_altered.groupby(df_temperature_altered['Year'].dt.year)['Max_Temperature'].ma
x().reset_index()

# 3. Fitting an ARIMA Model
model = ARIMA(max_values['Max_Temperature'], order=(8, 1, 5))
model_fit = model.fit()
```

```
# 4. Forecasting Future Years
future_years = list(range(2024, 2040))
future_predictions = model_fit.forecast(steps=len(future_years))

# 5. Visualizing the Results
plt.figure(figsize=(10, 6))
plt.plot(max_values['Year'], max_values['Max_Temperature'], color='blue', label='Actual Data')
plt.plot(future_years, future_predictions, color='green', label='Future Predictions')
plt.title('Max Temperature by Year with Future Predictions (ARIMA Model)')
plt.xlabel('Year')
plt.ylabel('Max Temperature')
plt.legend()
plt.show()

# 1. Data Preprocessing
# Convert 'Year' to a datetime format if needed
df_rainfall_altered['Year'] = pd.to_datetime(df_rainfall_altered['Year'], format='%Y')

# 2. Grouping Data
# Extract the maximum values for each year
max_values =
df_rainfall_altered.groupby(df_rainfall_altered['Year'].dt.year)['Max_Rainfall'].max().reset_inde
x()

# 3. Fitting an ARIMA Model
model = ARIMA(max_values['Max_Rainfall'], order=(6, 1, 2))
model_fit = model.fit()
```

```
# 4. Forecasting Future Years
future_years = list(range(2024, 2040))
future_predictions = model_fit.forecast(steps=len(future_years))

# 5. Visualizing the Results
plt.figure(figsize=(10, 6))
plt.plot(max_values['Year'], max_values['Max_Rainfall'], color='blue', label='Actual Data')
plt.plot(future_years, future_predictions, color='green', label='Future Predictions')
plt.title('Max Rainfall by Year with Future Predictions (ARIMA Model)')
plt.xlabel('Year')
plt.ylabel('Max Rainfall')
plt.legend()
plt.show()

# Merge datasets on the 'Year' column
df_merged = pd.merge(max_temperature_values, max_rainfall_values, on='Year')
# Calculate correlation
correlation = df_merged['Max_Temperature'].corr(df_merged['Max_Rainfall'])
print(f'Correlation between Temperature and Rainfall: {correlation}')
```

Appendix B

```
library(tidyverse)
library(sf)
library(mapview)
library(fpp3)
library(raster)
library(geodata)
library(predicts)
library(readxl)

bioclim_data <- worldclim_global(var = "bio",
                                 res = 2.5,
                                 path = "Model 3/")

carpobrotus_glaucescens <- read_excel("Latlongflora Full v2.xlsx", sheet = "Carpobrotus glaucescens")

mapview(carpobrotus_glaucescens, xcol = "Longitude", ycol = "Latitude", crs = 4269, grid = FALSE)


max_lat <- ceiling(max(carpobrotus_glaucescens$Latitude))
min_lat <- floor(min(carpobrotus_glaucescens$Latitude))
max_lon <- ceiling(max(carpobrotus_glaucescens$Longitude))
min_lon <- floor(min(carpobrotus_glaucescens$Longitude))

# Store boundaries in a single extent object
geographic_extent <- extent(x = c(min_lon, max_lon, min_lat, max_lat))


# Download data with geodata's world function to use for our base map
world_map <- world(resolution = 5,
                   path = "data/")

# Crop the map to our area of interest
my_map <- crop(x = world_map, y = geographic_extent)

# Plot the base map
plot(my_map,
     axes = TRUE,
     col = "grey95")

# Add the points for individual observations
points(x = carpobrotus_glaucescens$Longitude,
       y = carpobrotus_glaucescens$Latitude,
       col = "olivedrab",
       pch = 20,
       cex = 0.75)



# Make an extent that is 25% larger
sample_extent <- geographic_extent * 1.25

# Crop bioclim data to desired extent
bioclim_data <- crop(x = bioclim_data, y = sample_extent)

# Plot the first of the bioclim variables to check on cropping
plot(bioclim_data[[1]])
plot(bioclim_data[[12]])


# Set the seed for the random-number generator to ensure results are similar
set.seed(20210707)

# Randomly sample points (same number as our observed points)
background <- spatSample(x = bioclim_data,
                         size = 1000,    # generate 1,000 pseudo-absence points
                         values = FALSE, # don't need values
                         na.rm = TRUE,   # don't sample from ocean
                         xy = TRUE)      # just need coordinates

# Look at first few rows of background
head(background)



# Plot the base map
plot(my_map,
     axes = TRUE,
     col = "grey95")

# Add the background points
points(background,
       col = "grey30",
       pch = 1,
       cex = 0.75)

# Add the points for individual observations
points(x = carpobrotus_glaucescens$Longitude,
       y = carpobrotus_glaucescens$Latitude,
       col = "olivedrab",
       pch = 20,
       cex = 0.75)


# Pull out coordinate columns, x (longitude) first, then y (latitude) from
# carpobrotus rossii data
presence <- carpobrotus_glaucescens[, c("Longitude", "Latitude")]
# Add column indicating presence
presence$pa <- 1

# Convert background data to a data frame
absence <- as.data.frame(background)
# Update column names so they match presence points
colnames(absence) <- c("Longitude", "Latitude")
# Add column indicating absence
absence$pa <- 0
```

```
# Join data into single data frame
all_points <- rbind(presence, absence)

# Reality check on data
head(all_points)

# ADDING CLIMATE DATA

bioclim_extract <- extract(x = bioclim_data,
                           y = all_points[, c("Longitude", "Latitude")],
                           ID = FALSE) # No need for an ID column

# Add the point and climate datasets together
points_climate <- cbind(all_points, bioclim_extract)

# Identify columns that are latitude & longitude
drop_cols <- which(colnames(points_climate) %in% c("Longitude", "Latitude"))
drop_cols # print the values as a reality check

# Remove the geographic coordinates from the data frame
points_climate <- points_climate[, -drop_cols]

# TRAINING AND TESTING DATA

# Create vector indicating fold
fold <- folds(x = points_climate,
              k = 5,
              by = points_climate$pa)

table(fold)

testing <- points_climate[fold == 1, ]
training <- points_climate[fold != 1, ]

# MODEL BUILDING

# Build a model using training data
glm_model <- glm(pa ~ ., data = training, family = binomial())

# Get predicted values from the model
glm_predict <- predict(bioclim_data, glm_model, type = "response")

# Print predicted values
plot(glm_predict)

# Use testing data for model evaluation
glm_eval <- pa_evaluate(p = testing[testing$pa == 1, ],
                        a = testing[testing$pa == 0, ],
                        model = glm_model,
                        type = "response")

# Determine minimum threshold for "presence"
glm_threshold <- glm_eval@thresholds$max_spec_sens
```

```
# Plot base map
plot(my_map,
     axes = TRUE,
     col = "grey95")


# Only plot areas where probability of occurrence is greater than the threshold
plot(glm_predict > glm_threshold,
     add = TRUE,
     legend = FALSE,
     col = c(NA, "olivedrab")) # <-- Update the values HERE

# And add those observations
points(x = carpobrotus_glaucescens$Longitude,
       y = carpobrotus_glaucescens$Latitude,
       col = "black",
       pch = "+",
       cex = 1)

# Redraw those country borders
plot(my_map, add = TRUE, border = "grey5")




# Download predicted climate data
forecast_data <- cmip6_world(model = "MPI-ESM1-2-HR",
                             ssp = "245",
                             time = "2024-2039",
                             var = "bioc",
                             res = 2.5,
                             path = "data")

# Use names from bioclim_data
names(forecast_data) <- names(bioclim_data)

# Crop forecast data to desired extent
forecast_data <- crop(x = forecast_data, y = sample_extent)

# Predict presence from model with forecast data
forecast_presence <- predict(forecast_data, glm_model, type = "response")

# Plot base map
plot(my_map,
     axes = TRUE,
     col = "grey95",
     main = "Carpobrotus Glaucescens")

# Add model probabilities
plot(forecast_presence, add = TRUE)
```

```
# Redraw those country borders
plot(my_map, add = TRUE, border = "grey5")

# Add original observations
points(x = carpobrotus_glaucescens$Longitude,
       y = carpobrotus_glaucescens$Latitude,
       col = "black",
       pch = "+",
       cex = 1)

# Plot base map
plot(my_map,
     axes = TRUE,
     col = "grey95",
     main = "carpobrotus_glaucesens")

# Only plot areas where probability of occurrence is greater than the threshold
plot(forecast_presence,
     add = TRUE,
     legend = FALSE,
     col = c(NA, "olivedrab")
)

# And add those observations
points(x = carpobrotus_glaucescens$Longitude,
       y = carpobrotus_glaucescens$Latitude,
       col = "black",
       pch = "+",
       cex = 0.6)

# Redraw those country borders
plot(my_map, add = TRUE, border = "grey5")
```

References

AdaptNSW, n.d., *Interactive climate change projections map*, NSW Government, accessed 29 October 2023, <https://www.climatechange.environment.nsw.gov.au/projections-map>.

Atlas of Living Australia 2023, *ALA | Login*. ala.org.au, accessed 1 November 2023, <https://spatial.ala.org.au/?ss=1697731743607>.

Bray, F 2020, *Submission regarding the Draft Illawarra Shoalhaven Regional Plan 2041*, Lake Wollumboola Protection Association Inc, accessed 6 October 2023, <https://shared-drupal-s3fs.s3-ap-southeast-2.amazonaws.com/master-test/fapub_pdf/ISRP%20Submissions/00049%20Lake%20Wollumboola%20Protection%20Association%20Inc%20Part%201.pdf >.

CSIRO & Bureau of Meteorology 2015, Climate Change in Australia, Information for Australia's Natural Resource Management Regions: Technical Report, CSIRO and Bureau of Meteorology, Canberra, 216 pp.

Department of Agriculture and Fisheries 2018, *Grey Mangrove*, Queensland Government, accessed 1 November 2023, <https://www.daf.qld.gov.au/business-priorities/fisheries/habitats/marine-plants-including-mangroves/common-mangroves/grey-mangrove>.

Department of Climate Change, Energy, the Environment and Water 2021, *Wetlands and migratory shorebirds*, Australian Government, accessed 5 October 2023, <https://www.dcceew.gov.au/water/wetlands/about/migratory-shorebirds#:~:text=They%20provide%20food%20and%20habitat,role%20in%20maintaining%20global%20biodiversity>.

Department of Climate Change, Energy, the Environment and Water 2021, *Invasive species*, Australian Government, accessed 8 October 2023, <https://www.dcceew.gov.au/environment/invasive-species>.

Department of Climate Change, Energy, the Environment and Water n.d., *Lachnagrostis adamsonii — Adamson's Blown-grass, Adamson's Blowngrass*, Australian Government, accessed 1 November 2023, <http://www.environment.gov.au/cgi-bin/sprat/public/publicspecies.pl?taxon_id=19708>

Department of Environment and Energy 2019, *Wetlands and Climate Change*, Australian Government, accessed 5 October 2023, <https://www.dcceew.gov.au/sites/default/files/documents/wetlands-climate-change.pdf>.

Hutchinson, MF & Xu, T 2013, *ANUSPLIN Version 4.4 User Guide*, Fenner School of Environment and Society, Australian National University, Canberra, accessed 29 October 2023, <http://fennerschool.anu.edu.au/files/anusplin44.pdf>.

Key Biodiversity Areas n.d., *About KBAs*, accessed 6 October 2023, <https://www.keybiodiversityareas.org/about-kbas>.

Lake Wollumboola Protection Association Inc n.d., *Birdlife*, accessed 6 October 2023, <http://www.wollumboola.org.au/birdlife/> .

NSW Department of Planning and Environment 2018, *Plants in wetlands*, NSW Government, accessed 31 October 2023, <https://www.environment.nsw.gov.au/topics/water/wetlands/plants-and-animals-in-wetlands/plants#:~:text=Semi%2Dpermanent%20wetlands&text=They%20are%20dominated%20by%20sedges,as%20water%20primrose%20and%20nardoo.>.

NSW Department of Planning and Environment 2023, *About biodiversity*, NSW Government, accessed 5 October 2023, <https://www.environment.nsw.gov.au/topics/animals-and-plants/biodiversity/about-biodiversity>.

NSW Department of Planning and Environment 2023, *Impacts of climate change on biodiversity*, NSW Government, accessed 5 October 2023, <https://www.environment.nsw.gov.au/topics/animals-and-plants/biodiversity/about-biodiversity/climate-change-impacts>.

NSW Department of Planning and Environment 2023, *Migratory shorebirds*, NSW Government, accessed 6 October 2023, <https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Animals-and-plants/Threatened-species/sos-summaries/2016-21/saving-our-species-projects-2016-21-migratory-shorebirds-230068.pdf>.

Office of Environment and Heritage 2018, *Round-leafed Wilsonia – profile*, NSW Government, accessed 1 November 2023, <https://www.environment.nsw.gov.au/threatenedspeciesapp/profile.aspx?id=10839>.

Office of Environment and Heritage 2019, *Waterfall Greenhood – profile,* NSW Government, accessed 1 November 2023, <https://www.environment.nsw.gov.au/threatenedspeciesapp/profile.aspx?id=10704>.

Office of Environment and Heritage 2022, *Deane's Boronia – profile*, NSW Government, accessed 1 November 2023, <https://www.environment.nsw.gov.au/threatenedspeciesapp/profile.aspx?id=10097>.

Oliver J 2023, *A very brief introduction to species distributions models in R*, GitHub, accessed 20 November 2023, <https://jcoliver.github.io/learn-r/011-species-distribution-models.html>.

REMPLAN Economy, n.d., *Economy, Jobs and Business Insights*, accessed 8 October 2023, <https://app.remplan.com.au/shoalhaven/economy/summary?state=nQNXc8lBphy2b2JFq687BYU7IDIAVr>.

Roads and Maritime Services 2018, *NSW Tides 2018-2019*, NSW Government, accessed 2 October 2023, <https://www.nsw.gov.au/sites/default/files/2021-07/tide-tables-2018-2019.pdf>.

Shoalhaven City Council n.d., *Endangered Ecological Species (Communities)*, accessed 5 October 2023, <https://www.shoalhaven.nsw.gov.au/For-Residents/Our-Environment/Native-Animals-Plants/Endangered-Ecological-Communities#section-2>.

Shoalhaven City Council 2023, *Help protect our threatened shorebirds*, media release, accessed 6 October 2023, <https://www.shoalhaven.nsw.gov.au/Council/News/Help-protect-our-threatened-shorebirds>.

Sullivan, P 2018, *Objection to the Long Bow Point Golf course application (SSD 8406)*, BirdLife Australia, accessed 6 October 2023, <https://www.ipcn.nsw.gov.au/resources/pac/media/files/pac/projects/2018/07/long-bow-point-golf-course/comments-and-presentations-received-before-31st-july-2018/paul-sullivan.pdf>.

Transport for NSW 2019, NSW Tides 2019-2020, NSW Government, accessed 2 October 2023, <https://www.nsw.gov.au/sites/default/files/2021-07/tide-tables-2019-2020.pdf>.

Transport for NSW 2020, NSW Tides 2020-2021, NSW Government, accessed 2 October 2023, <https://www.nsw.gov.au/sites/default/files/2021-07/tide-tables-2020-2021.pdf>.

Transport for NSW 2022, NSW Tides 2022-2023, NSW Government, accessed 2 October 2023, <https://www.nsw.gov.au/sites/default/files/2022-06/NSW-tide-tables-2022-2023-Pdf.pdf>.

Victorian Resources Online Statewide n.d., Spike-Rush, Agriculture Victoria, accessed 1 November 2023, <https://vro.agriculture.vic.gov.au/dpi/vro/vrosite.nsf/pages/sip_spike_rushes>

WorldClim 2023, WorldClim climate data, accessed 30 October 2023, <https://rdrr.io/cran/geodata/man/worldclim.html>.

Xu, T & Hutchinson, MF 2011, ANUCLIM Version 6.1 User Guide, Fenner School of Environment and Society, Australian National University, Canberra, accessed 29 October 2023, <http://fennerschool.anu.edu.au/files/anuclim61.pdf>.