# Use Case 1 - Ecological Niche Modeling

Please note that this use case is fictitious and built for instructional purposes using data downloaded from [www.gbif.org](http://www.gbif.org). Any reference to countries and structures/organisations, real or otherwise, within those countries are used merely to facilitate the use of the data and do not reflect the reality within those countries.

## Description of the use case

Gymnosarda unicolor, the Dogtooth tuna, is a medium-sized fish species in the family Scombridae, which includes mackerels, tunas, and other popular commercial fish species. They are found predominantly in reef environments throughout the tropical Indo-Pacific region, with smaller fish being more commonly found near shallow reef areas and larger ones haunting deep reef drop off areas, seamounts and steep underwater walls. Rainbow tuna prey primarily on smaller schooling fishes and squids. Usually they occur in small schools to a depth of 10–300 metres, but tagging studies in the Indian ocean have shown that individuals are able to disperse across distances of 4,000 km. Gymnosarda unicolor is harvested extensively across its range, both supplying local markets and as part of international trade. In the last 30 years, there has been a 200% increase in catch size for this species due to the increased mechanisation of the fishing fleet and increased demand from more affluent urban communities on the islands that can afford the fish. Further, in the last 5 years US retailer M&Vs has incorporated Gymnosarda unicolor within its supply chain in response to a new trend in the USA for the fish in “unicorn” sushi. With increasing targeted fishing of this species to supply domestic and international markets, there is concern over the continued persistence of the species and there is no information on its conservation status. Though this species is well-known by commercial fishermen, there has been no systematic study of the distribution of the species at global scale. A distribution map for the species would provide a starting point from which fishery management plans could be developed to ensure the sustainability of stocks.

## Data description

Box .- GBIF.org (22 November 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.5fhynv -

*Figure STYLEREF 1 \s 0. SEQ Figure \\* ARABIC \s 1 1 Distribution of available GBIF-mediated georeferenced Gymnosarda unicolor occurrence data.*

## You will need to download the current dataset for Gymnosarda unicolor from [www.gbif.org](http://www.gbif.org/).

## 

### Exercise 1 - Data Processing

You should download datasets from [www.gbif.org](http://www.gbif.org) for *Gymnosarda unicolor .* Not all of the data within the download is fit-for-purpose, and in this exercise you should process your data download to only include those data points that you will be using to build your ecological niche model. You should use whichever processing tool you are most comfortable e.g. Excel, Open Refine, R etc. We have provided you with a step-by-step guide in Excel and an R markdown document if you would like to take a more programmatic approach to data processing.

|  |
| --- |
| Q1. What are the taxonkeys for each of the species and what is the taxonomic status of each species? |
|  |
| Q2. In what format did you download the data and what are the unique properties of this type of download? |
|  |
| Q3. What are the DOIs of your downloads? |
|  |
| Q4. Which of the following key data quality processing steps did you use for cleaning both datasets? For each requirement, what is your justification? |
| Default geospatial issues  Absence records  Fossils and living specimens  Establishment Means  Old records  Uncertain location  Bad default values for coordinate uncertainty  Points along the Equator or prime meridian  Country centroids  Duplicate removal  Outliers  Metagenomics  Outside Native ranges  Gridded datasets  Automated identification |
| Q5. What additional data processing steps might you want to validate taxonomic identification of species? |
|  |

## Exercise 2 - Starting Wallace and Loading Occurrences

Let’s start by launching Wallace and giving you a quick overview of the steps you will go through to generate a niche model. You have been provided with a step-by-step guide to this exercise.

## 

|  |
| --- |
| Q1. What is the overall goal of your ecological niche model? That is, what question or hypothesis are you exploring? |
|  |
| Q2. What type of niche are we modeling? What limits our ability to model the kind of niche we hope to model? |
|  |

## 

## Exercise 3 - Determining a training region

|  |
| --- |
| Q1. On the map below, draw what you think would be a good training region for the model. |
|  |
| Q2. Why did you choose this area? Is it based on where the species is known to occur, as well as areas it could access? Are there features of the species’ natural history that help inform your choice? |
|  |

## Exercise 4 - Partitioning Occurrence Data

In this exercise, you will partition your occurrence data in Wallace. You have been provided with a step-by-step guide to this exercise.

# **Exercise 5 – Model Evaluation**

|  |
| --- |
| Q1. Record the AICc score for each model. |
| * 1. LQHP\_1:   2. LQHP\_2: |
| Q2. Which model performed better according to AICc |
|  |
| Q3. Fill in the following table with the model evaluation statistics for your model |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Model | OR10\_bin.1 | OR10\_bin.2 | OR10\_bin.3 | OR10\_bin.4 | ORmin\_bin.1 | ORmin\_bin.2 | ORmin\_bin.3 | ORmin\_bin.4 | | LQHP\_1 |  |  |  |  |  |  |  |  | | LQHP\_2 |  |  |  |  |  |  |  |  | |
| Q4. Based on the overall omission rate for all the bins, which model performed better? Does this match the conclusion reached using AICc? |
|  |
| Q5. Based on AICc and omission rate, which model do you think will be the best to continue working with? |
|  |

# 

## Exercise 6 - Visualizing, Thresholding, and Projecting Niche Models

In this exercise, you will process results from Exercise 6 to produce maps of *Gymnosarda unicolor* suitable habitat, as well as several plots to diagnose model performance. You will also project your Rainbow tuna model into different bioclimatic conditions. You have been provided with a step-by-step guide to this exercise.

|  |
| --- |
| Q1. What similarities do you see across the four visualizations? What are the major differences? |
|  |
| Q2. Look closely at your projected model. Based on what you know about our study species, *Gymnosarda unicolor*, do your model results make sense? Are there any areas of predicted absence or presence that are questionable? What areas? Why do you question the model prediction in these areas? |
|  |