Species Distribution Modelling: Honours Survival Guide

Clarke van Steenderen

vsteenderen@gmail.com / clarke.vansteenderen@ru.ac.za

Centre for Biological Control, Department of Zoology and Entomology, Rhodes University

February 10, 2025

---

1 Introduction

Species Distribution Models (SDMs) are statistical tools designed to predict the geographic areas where species might occur. These predictions can be extrapolated over time (e.g., using future climate models that simulate various climate change scenarios) and across different environments and spatial scales (e.g. targeted areas within a country, whole provinces/districts or countries, continents, or the whole globe). Many SDM studies focus on invasive species to identify new areas that might be climatically suitable for their establishment and spread. Similarly, studies on threatened species aim to predict distribution changes under future climate conditions, helping prioritise areas for conservation efforts. In addition, SDMs are increasingly applied in biological control programs to predict the likelihood of suitable climatic conditions for biological control agents under consideration for release.

The applications for SDMs are broad, and once you have the skills required to build and test your own models, you will be able to create predictive maps for any taxon with available GPS records. While the process may seem complex at first—especially aspects like model parameterisation and R syntax —don't be discouraged. Keep reading, experimenting with your code, engaging with classmates, and reflecting on new concepts. With practice, you'll be able to successfully add SDMs to your ecological research toolkit.

2 Module outline

This Honours module will comprise a mix of discussion sessions, tutorials, brief lectures, and a final assignment. We will have one class discussion per week for the first three weeks of the module, themed on one or more prescribed readings (Section 5). Be sure to engage with each reading, and make notes. Tutorial sessions will cover the basics of SDM analysis pipelines in R. The scripts for each session are available on GitHub ().

3 What SDMs are not

Many people erroneously treat climatic suitability scores as direct probabilities of establishment. However, SDMs predict the relative climatic suitability of a particular site, reflecting the likelihood that the model will classify a species as present (scored as 1) or absent (scored as 0), based on environmental conditions (and/or other variables of interest). It is important to note that actual establishment depends on factors beyond climate alone, including biotic interactions, dispersal, habitat availability, and adaptive ability.

4 What SDMs are, and how they work

"*If you can't explain it simply, you don't understand it well enough*" - Einstein ☺

Suppose we have six confirmed GPS records for species A, and six sites where it does not occur (absence points), shown as green circles and red crosses in Figure 1. We aim to use average temperature and rainfall as variables (also termed 'covariates') to characterise this species' distribution. By extracting these data at each presence and absence point, we can train a predictive model to distinguish between presence (scored as 1) and absence (scored as 0). If the model performs well, we can project it onto new sites, providing probability values between 0 and 1 to indicate climatic suitability.

The new area in Figure 1 represents a climatic suitability map, where areas are coloured by probability, from 0 (least suitable; blue) to 1 (most suitable; red). The map suggests that if species A were to arrive in the new area, it would prefer the eastern and western regions, with the central areas being less suitable.

At this point, you may have several questions about the model-building process and the reliability of these predictions. These could include:

1. How should we generate pseudo-absence points if no true absences are confirmed? Which approach to generating pseudo-absences is most reliable? What if a species does indeed occur at a generated pseudo-absence point, but just hasn't been recorded there yet?

2. How do we verify that all presence points are real?

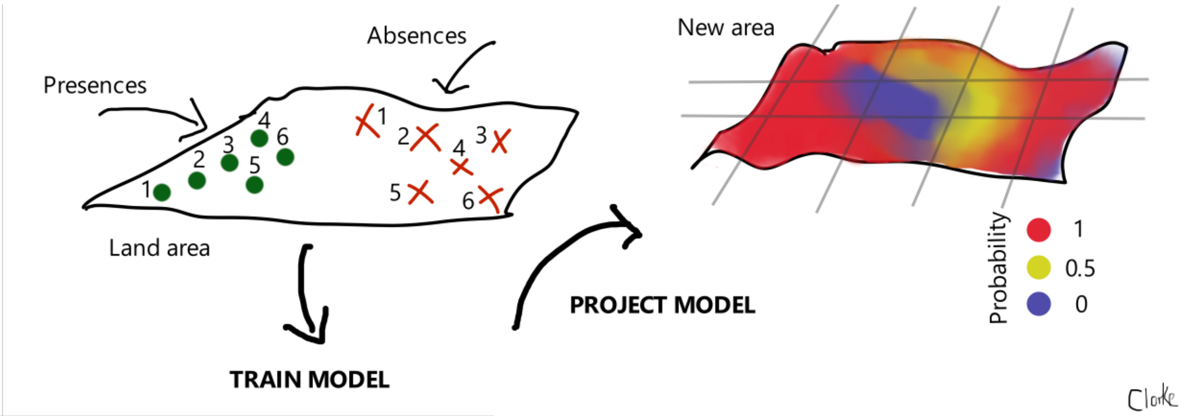
3. How do we determine which variables to include in the SDM? How might predictions change if only temperature, or temperature + rainfall + disturbance, or some other combination are used?

4. How do we assess the impact of adding more training points? What if there are very few presence points available?

5. How accurate are predictions based on broad variables like temperature and rainfall, which do not capture microhabitats?

6. Should we be considering the thermal physiology of the target species in the SDMs? How might the species' ability to adapt to new environments confound predictions?

7. How reliable are future climate models, considering the vast variety of potential scenarios?



	PRESENCES		ABSENCES	
Site	Rain	Temperature	Rain	Temperature
1	2.2	22	10.2	5
2	3.5	25	11.3	10
3	6	23	15.3	12
4	4.5	22	9.6	8
5	4.2	21	13.2	9
6	6.1	20	14.3	11

1 **0**

Figure 1: A simplified diagram explaining the essence of species distribution modelling.

5 Readings

Week 1: the basics

1. Not as good as they seem: the importance of concepts in species distribution modelling (Jiménez-Valverde et al., 2008)

2. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time (Elith and Leathwick, 2009)

3. Species Distribution Modeling (Jennifer Miller, 2010)

4. New trends in species distribution modelling (Zimmermann et al., 2010)

Week 2: invasion context

1. Oh the places they'll go: improving species distribution modelling for invasive forest pests in an uncertain world (Srivastava et al. 2020)

2. Species distribution models have limited spatial transferability for invasive species (Liu et al, 2020)

3. The dos and don'ts for predicting invasion dynamics with species distribution models (Hui, 2023)

Week 3: global climate change and conservation

1. Improving species distribution models for climate change studies: variable selection and scale (Austin and van Niel, 2010)

2. Integrating species distribution modelling into decision-making to inform conservation actions (Villero et al, 2017)

6 Tutorials

Tutorial sessions will cover the following:

1. Setting up the R environment for SDM and GIS-related tasks

2. Sourcing and filtering GBIF data, and creating a database of GPS records for a target species

3. Selecting appropriate variables to include in the modelling step - running collinearity tests

4. Generating random, but climatically-meaningful, background GPS points

5. Extracting the variable data for each presence and background GPS point

6. Tuning the SDM model (i.e. finding the best settings that optimise performance)

7. Creating the best-suited SDM model

8. Projecting the SDM model onto a desired area, and plotting the suitability maps

9. Exporting data, and interpreting output

10. A bried introduction to GitHub for code sharing and storage

Assignment

Using the skills learnt in this module, complete these tasks in R:

1. Use the global distribution of the Argentine ant (*Linepithema humile* Mayr; Hynemoptera: Formicidae) to train an SDM using MaxEnt. Use variables from the WorldClim database, and also include global elevation as a predictor.

2. Project your MaxEnt model onto **South Africa** and its neighbouring countries (Namibia, Botswana, Zimbabwe, and Mozambique), and produce a climate suitability map. Overlay the current distribution records (GPS points) for the ant onto this map.

3. Produce a succinct summary report in which you present your methods and findings.

4. Discuss how you think the Argentine ant's distribution might change in the future, and highlight some potential suitable corridors that you think might enable its dispersal.

Bonus marks for an extra 20%: Repeat the analysis, but this time train the SDM on only the ant's **native South American range** (as opposed to its full global range). Present a new suitability map for South Africa and its neighbours, and discuss how it differs from the global range-trained model. Add a brief paragraph to your report in which you relate your modelling approaches to the fundamental versus realised niche of a species.