Species Distribution Modelling Overview

11 October 2022

What is Species Distribution Modelling?

- SDM extrapolates species distribution data (e.g. occurrence) in space and time, usually based on a statistical model
- Using species distribution data and environmental variables hypothesized to influence habitat suitability for said species, a species distribution can be approximated

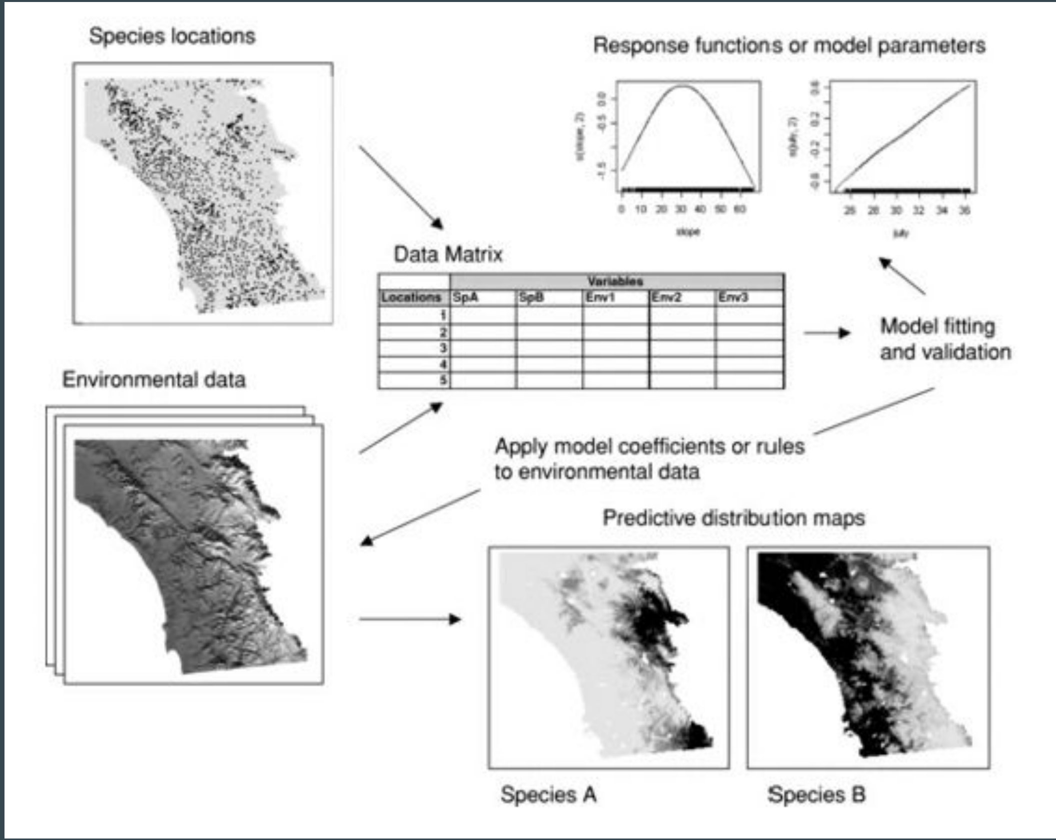


Figure credit: Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.

Goal

What is the aim of species distribution modeling (SDM)?

- Estimate the similarity of the conditions at any site to the conditions at the locations of known occurrence (and maybe known absence) of a phenomenon
- This method is commonly used to predict species ranges with climate and other environmental data as predictors
- The model utilized, assuming good fit, can provide insight into habitat and environmental preferences and enables potential spatial prediction of species distribution

Components of SDM

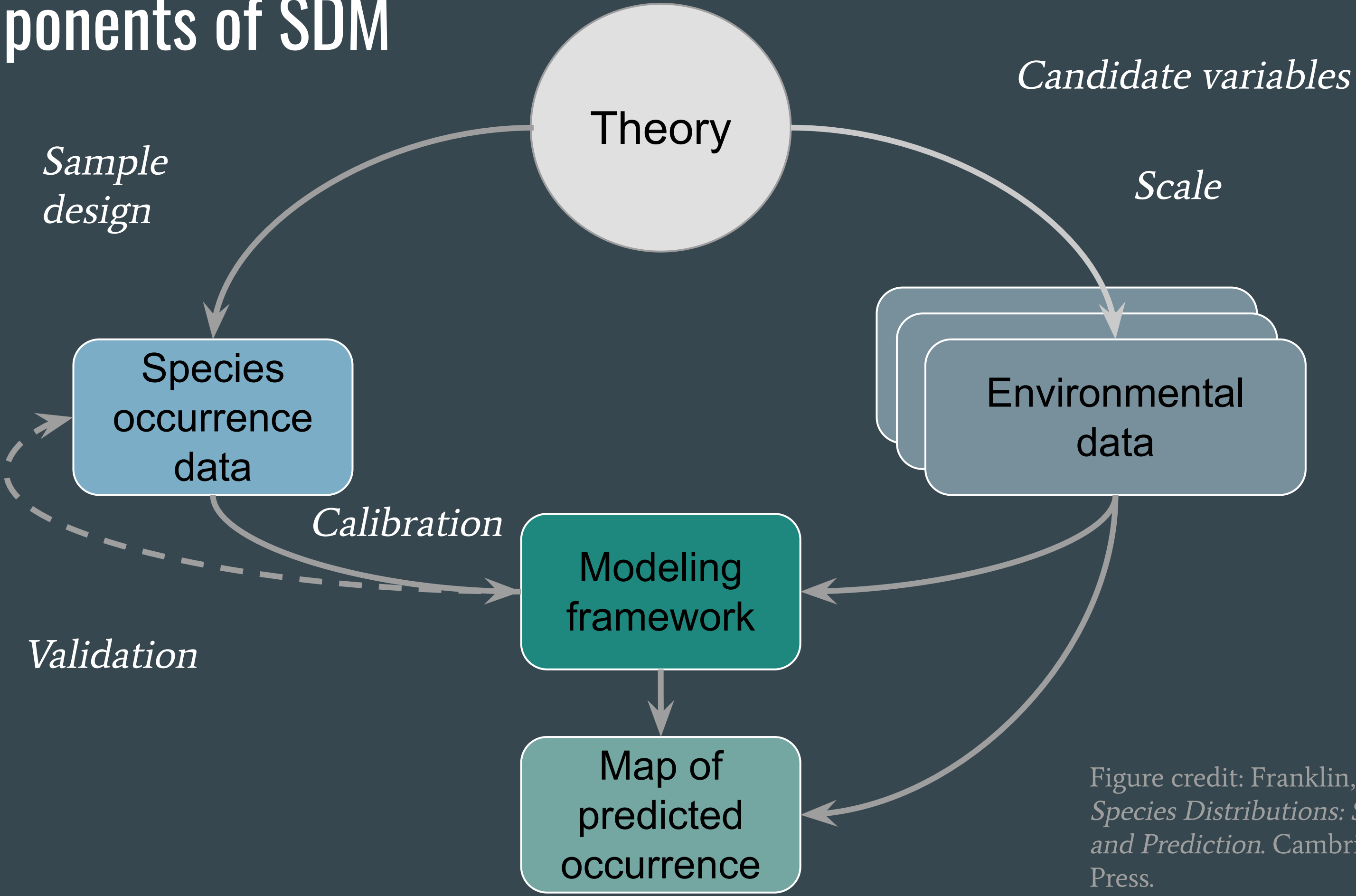


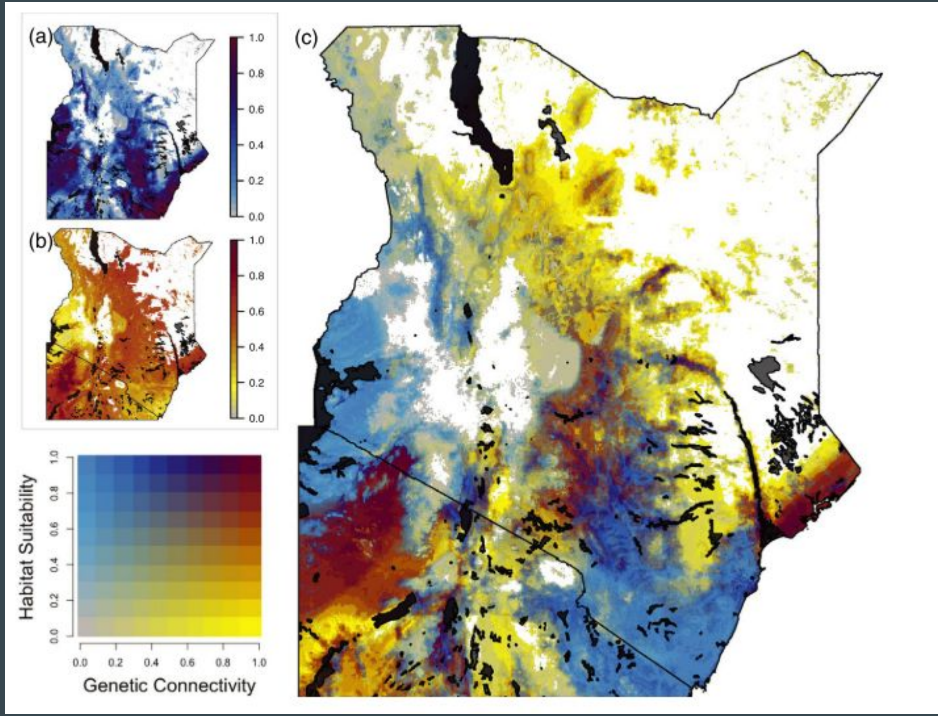
Figure credit: Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.

Traditional SDM implementation

1. Prepare occurrence data
2. Prepare predictors
3. Fit model
4. Predict

1. Prepare occurrence data	2. Prepare predictors	3. Fit model	4. Predict
Locations of species occurrence compiled	Curate suite of environmental predictor variables and extract values at species occurrence sites	Use environmental predictor values to fit a model estimating similarity to occurrence sites, or an alternative measure such as species abundance	Using the model created in previous step, predict species occurrence across geography of interest

Why Species Distribution Modelling?



- Exciting space that has grown rapidly in recent decades
- Largely due to :
  - Growing need for information on geographic distribution of biodiversity
  - Improved techniques
    - Remote sensing
    - Global positioning systems
    - Statistical learning methods
  - New data availability
    - Imagery
    - Entomological surveillance
    - Genetic lineage data
- Particularly relevant for vector-borne disease research and intervention
  - If we can better understand where the non-human vector exists, we are better positioned to intervene productively

Figure credit: Bishop, A. P., Amatulli, G., Hyseni, C., Pless, E., Bateta, R., Okeyo, W. A., Mireji, P. O., Okoth, S., Malele, I., Murilla, G., Aksoy, S., Caccone, A., & Saarman, N. P. (2021). A machine learning approach to integrating genetic and ecological data in tsetse flies (Glossina pallidipes) for spatially explicit vector control planning. *Evolutionary Applications*, *14*(7), 1762. https://doi.org/10.1111/EVA.13237

What you will learn in this course

- Understand SDM processes & concepts
- Understand how to choose the proper data and model for a given scenario
- SDM implementation and algorithms
  - Classical regression models
    - Logistic regression
    - Generalized Linear Models (GLM)
    - Generalized Additive Models (GAM)
  - Machine learning methods
    - MaxEnt
    - Random forest
- Accuracy assessment
- How to connect SDM outputs with overall disease epidemiology to improve disease control