Modelling the SEQ koala distribution under current and future climate scenarios

Charlotte Patterson, QUT
Scott Forrest, QUT, CSIRO

Acknowledgement of Traditional Owners

- QUT – Turrbal and Yugara
- Lands on which data were collected

Who are we?

Not koala experts !

How will this workshop run?

- Workshop outline:
  - Intro to SDMs
  - (Some) Koala ecology
  - Coded example of SDM workflow

    https:// geospatial-community.github.io/ICCB_geospatial_tools_conservation

- Look out for:

  🤔 Spots where careful thought is needed for your SDM application

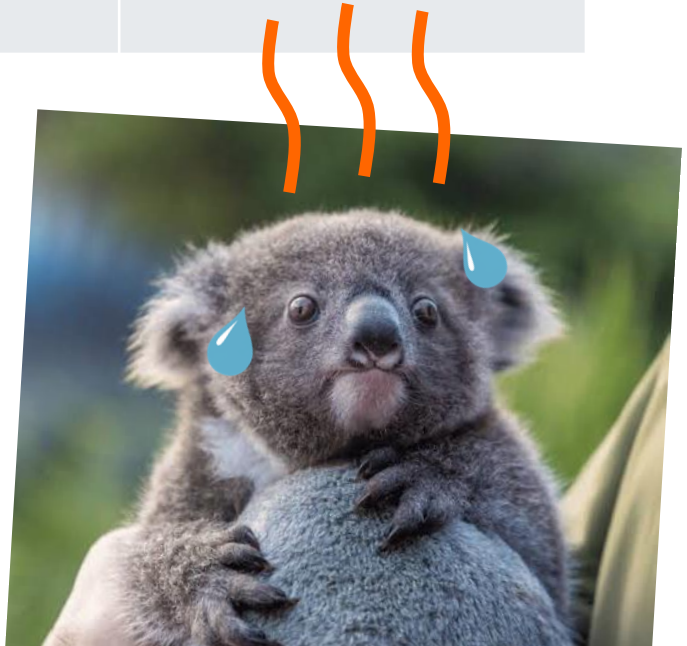
  💡 Key messages

  📖 Foundational resources

Where have we been and where are we headed?

Session 1	Session 2	Session 3	Session 4	Session 5
Intro to geospatial data and tools	Downscaled climate projections	Koala SDMs	Spatial conservation planning	Making maps with QGIS
Jason Flower, Mitch Rudge, Catherine Kim, EcoCommons team	Ralph Trancoso, Sarah Chapman, Rohan Eccles	Charlotte Patterson, Scott Forrest	Brooke Williams, Caitie Kuemple	Emma Hain, Nyall Dawson, Jason Flower

Where are we going to place protected areas to secure future populations of koalas under climate change scenarios?



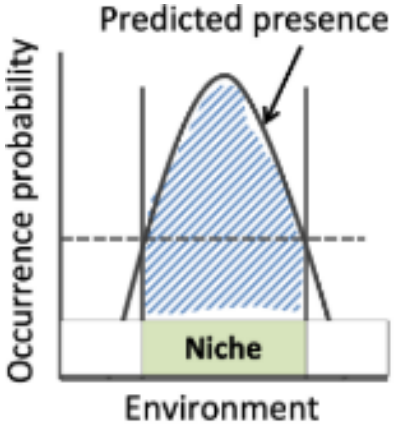
Take home message

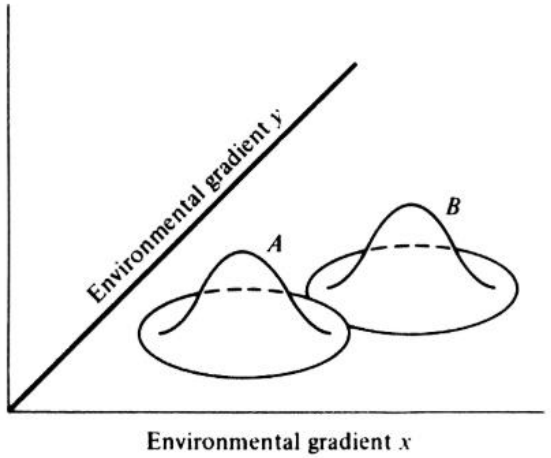
Modelling is often more art than science. There are many ways to model species, none of them the ‘right way’. Some can, however, be more appropriate for a study question or species.

We make many decisions along the way and need to be transparent and clear about these decisions.

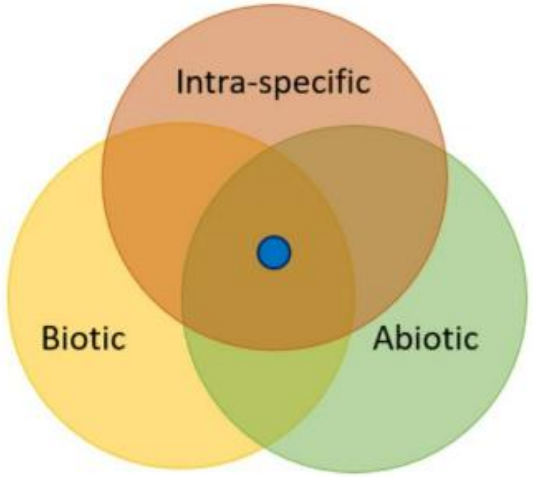
The SDM question: Where are species and why are they distributed as they are?

Ecological Niche Models; Habitat Suitability Models; Climate Envelope Models





http://www.zo.utexas.edu/courses/bio373/chapters/Chapter13/Chapter13.html



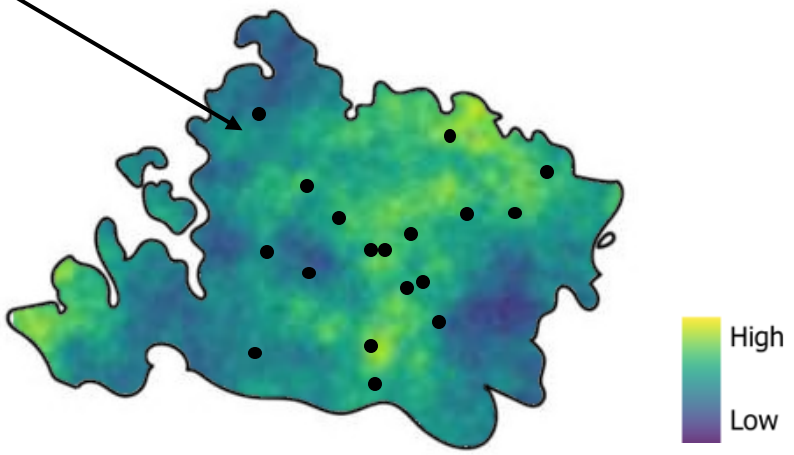
Takola & Schielzeth (2022)

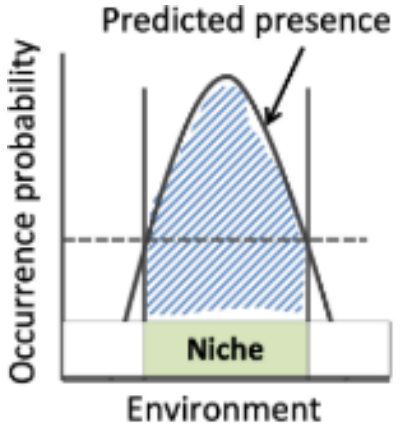
Guisan et al. (2017). ***Habitat suitability and distribution models: with applications in R***. Cambridge University Press.

The SDM question: Where are species and why are they distributed as they are?

Ecological Niche Models; Habitat Suitability Models; Climate Envelope Models

Environmental variables
-Soil moisture
-Temperature
etc.

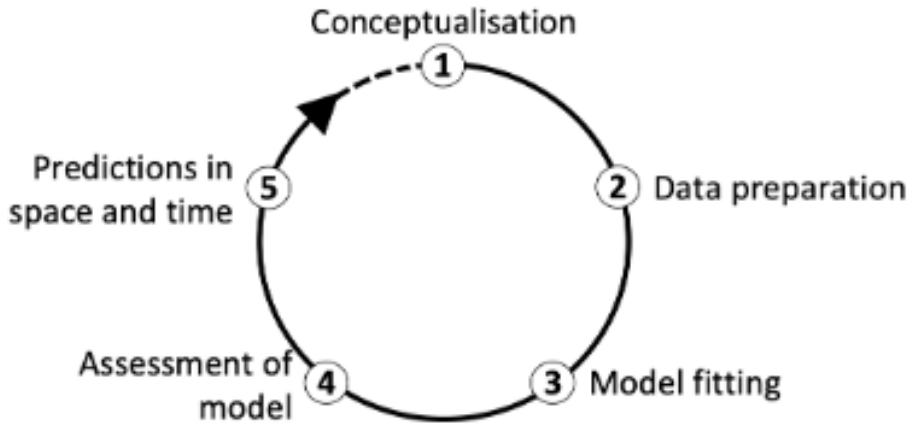




Interpolated probability of spp. presence

Guisan et al. (2017). ***Habitat suitability and distribution models: with applications in R***. Cambridge University Press.

An SDM workflow



Zurell et al. (2020). **A standard protocol for reporting species distribution models.** *Ecography*.

🤔 Conceptualisation: Some questions to ask yourself

1. What is the aim of my model?

- Explanation
- Mapping
- Transfer (spatial and/or temporal)

Araújo et al. (2019). **Standards for distribution models in biodiversity assessments**. *Science advances*.

Conceptualisation: Some questions to ask yourself

2. What are my model outputs going to be used for?

- Testing hypotheses about a species' ecology
- Spatial prioritisation for protection
- Choosing where to survey

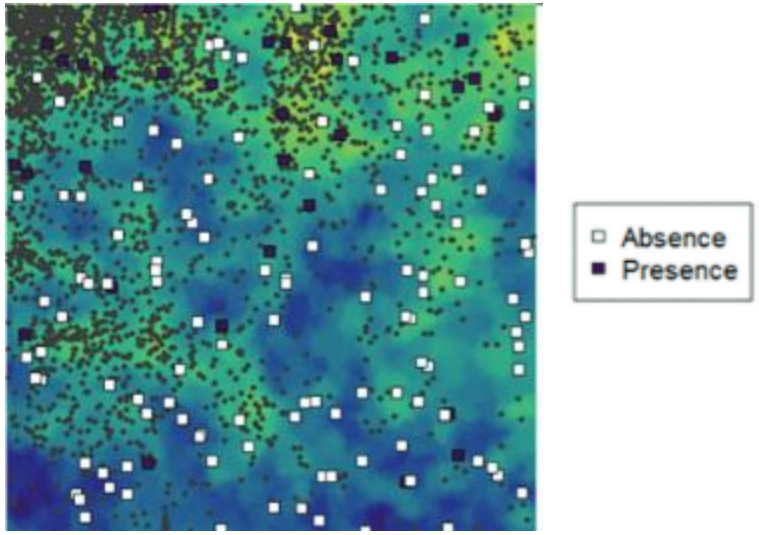
Clearly defining your model's purpose guides your workflow

Guillera-Arroita et al. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global ecology and biogeography*.

🤔 Data for modelling a species distribution

- Presence-only
  - Occurrence/incidental records
  - Citizen science databases

- Presence-absence
  - Systematic surveys
  - Atlases



💡 *Your data limit what is possible with SDMs. Prioritise quality over quantity.*

https://galah.ala.org.au/R/articles/quick_start_guide.html

Koala data from the ALA

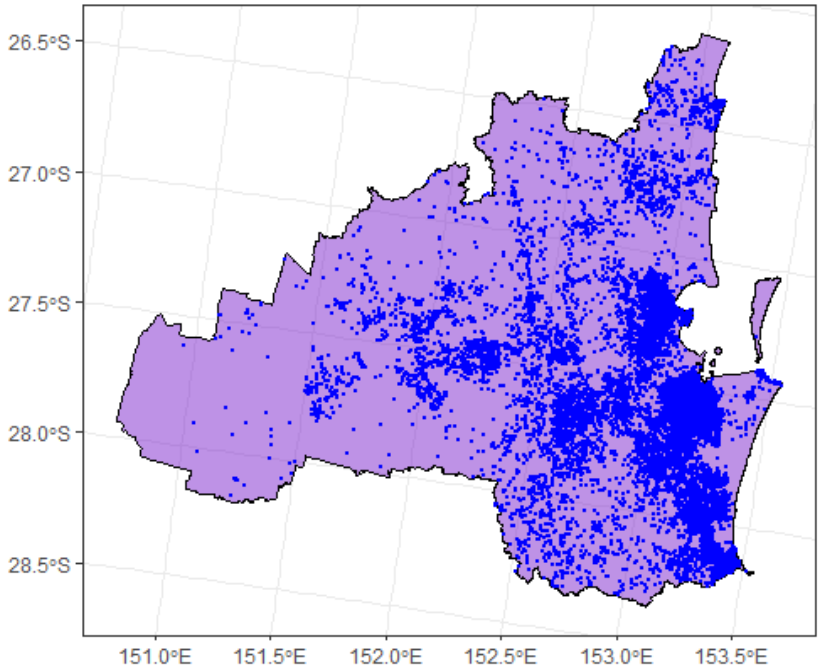


Photo by DAVID ILIFF.
License: https://creativecommons.org/licenses/by-sa/3.0/

GBIF
Global Biodiversity Information Facility

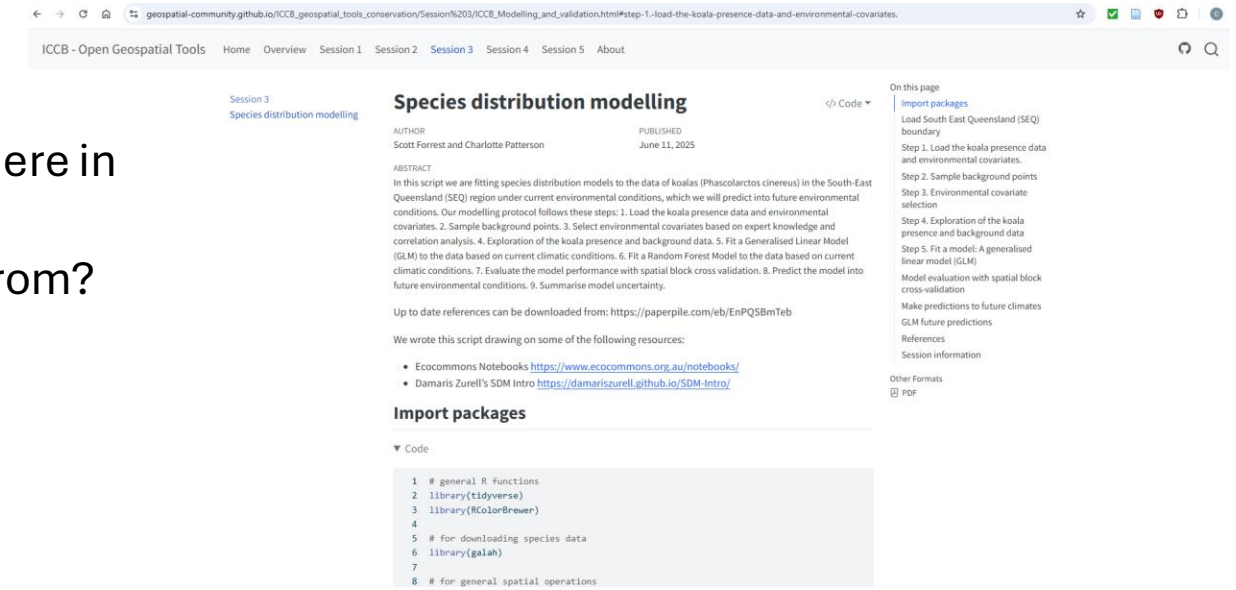
EcoCommons

Atlas of Living Australia
ala.org.au
sharing biodiversity knowledge

Have a go: Data preparation

https://geospatial-community.github.io/ICCB_geospatial_tools_conservation

- Questions:
  - How many koala presences are there in QLD? How about SEQ?
  - What sources do the data come from?
  - What years are they from?



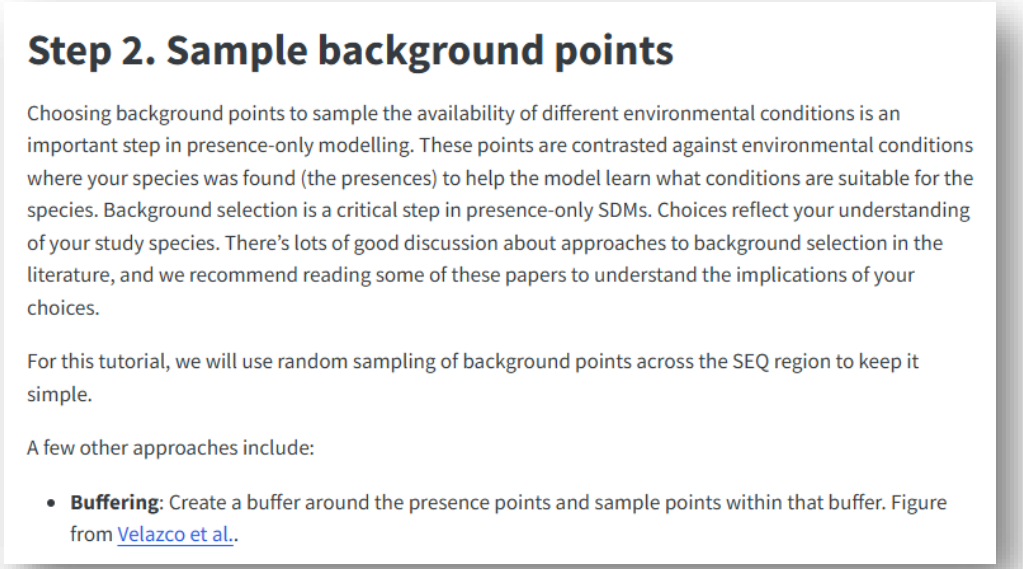
🤔 Background selection

- Many ‘presence-only’ approaches rely on the selection of background or ‘pseudo-absence’ points
- These points are contrasted against environmental conditions where your species was found

💡 *Background selection is a critical step in presence-only SDMs. Choices reflect your understanding of your study species.*

Have a go: Background selection

https://geospatial-community.github.io/ICCB_geospatial_tools_conservation



Step 2. Sample background points

Choosing background points to sample the availability of different environmental conditions is an important step in presence-only modelling. These points are contrasted against environmental conditions where your species was found (the presences) to help the model learn what conditions are suitable for the species. Background selection is a critical step in presence-only SDMs. Choices reflect your understanding of your study species. There's lots of good discussion about approaches to background selection in the literature, and we recommend reading some of these papers to understand the implications of your choices.

For this tutorial, we will use random sampling of background points across the SEQ region to keep it simple.

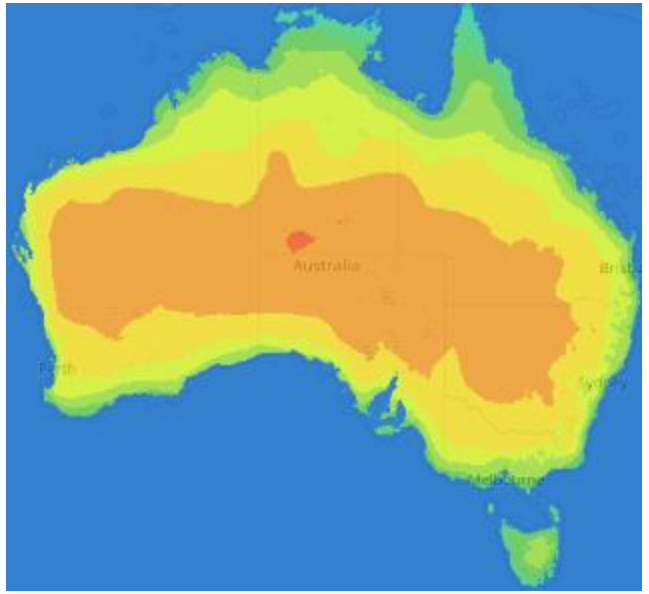
A few other approaches include:

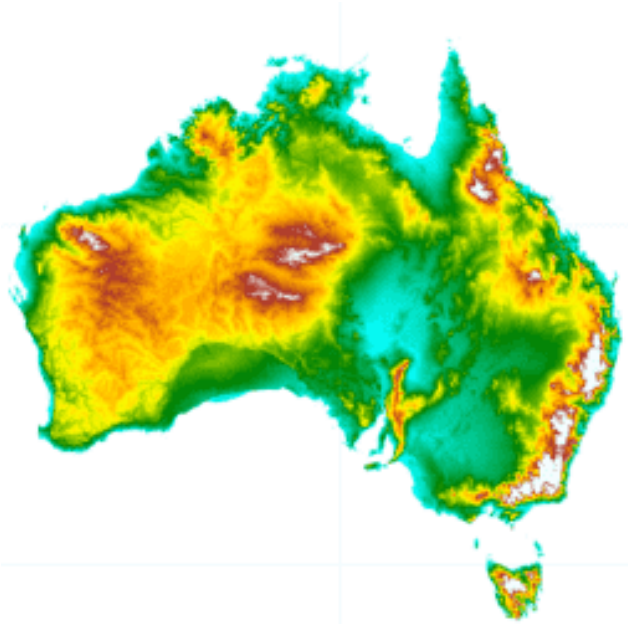
- **Buffering**: Create a buffer around the presence points and sample points within that buffer. Figure from Velazco et al..

🤔 Environmental layers

- Raster data representing covariates that can be used to describe a species' niche.

- Common examples are:
  - 'Bioclim' variables related to temperature and precipitation
  - Topographic variables like elevation
  - Satellite-derived measures of vegetation





Koala (*Phascolarctos cinereus*)

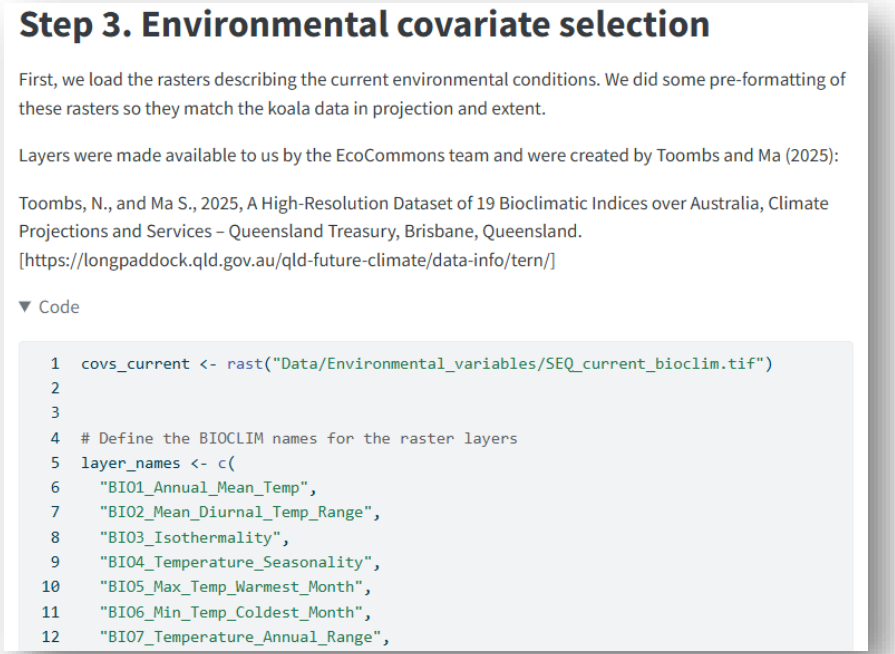
- Found in open forest and woodland
- Dependent on specific feeding trees*
  - ~ 30 species of *Eucalyptus*
- Sensitive to land-use change

Photo by DAVID ILIFF.
License: https://creativecommons.org/licenses/by-sa/3.0/

*a spatial layer of these trees can be used as a **mask**

Let’s look at our covariates & explore our data

https://geospatial-community.github.io/ICCB_geospatial_tools_conservation



Step 3. Environmental covariate selection

First, we load the rasters describing the current environmental conditions. We did some pre-formatting of these rasters so they match the koala data in projection and extent.

Layers were made available to us by the EcoCommons team and were created by Toombs and Ma (2025):

Toombs, N., and Ma S., 2025, A High-Resolution Dataset of 19 Bioclimatic Indices over Australia, Climate Projections and Services – Queensland Treasury, Brisbane, Queensland. [https://longpaddock.qld.gov.au/qld-future-climate/data-info/tern/]

▼ Code

```
covs_current <- rast("Data/Environmental_variables/SEQ_current_bioclim.tif")


# Define the BIOCLIM names for the raster layers
layer_names <- c(
  "BIO1_Annual_Mean_Temp",
  "BIO2_Mean_Diurnal_Temp_Range",
  "BIO3_Isothermality",
  "BIO4_Temperature_Seasonality",
  "BIO5_Max_Temp_Warmest_Month",
  "BIO6_Min_Temp_Coldest_Month",
  "BIO7_Temperature_Annual_Range",
```

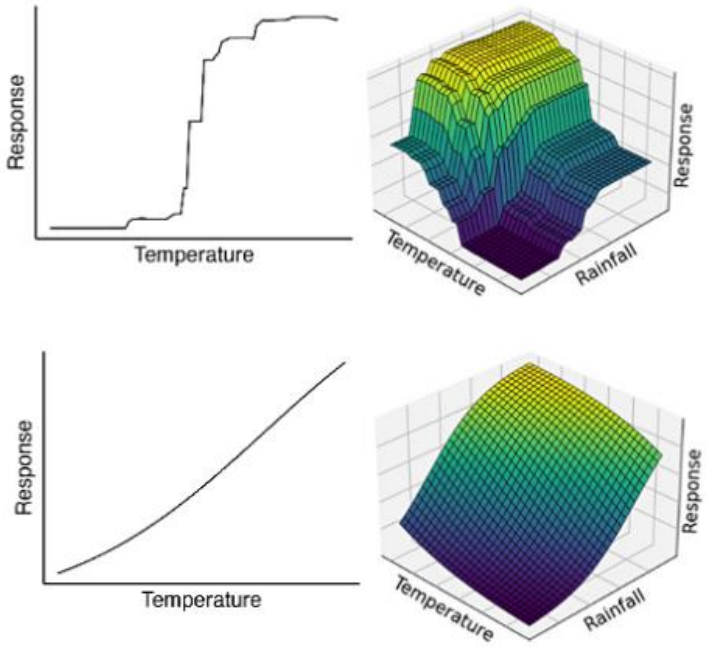
🤔 Models and algorithms

- Models differ in their flexibility and interpretability

- Spectrum from linear to highly non-linear
- Different strengths depending on model aim

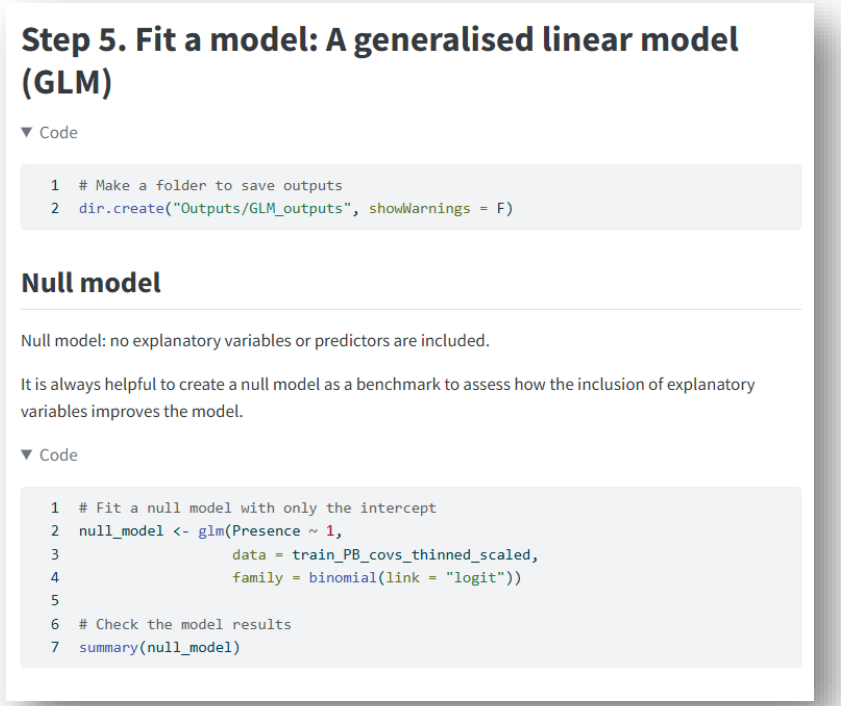
- Generalised Linear Model (GLM)
- Generalised Additive Model (GAM)
- Random Forest (RF)
- Maximum entropy modelling (Maxent)
- Deep learning approaches
  - Convolutional neural networks

💡 *There's no one 'right' model to use – different models are best for different contexts. Sometimes multiple models must to be tried to find the best approach.*



Time to fit some models!

https://geospatial-community.github.io/ICCB_geospatial_tools_conservation



Step 5. Fit a model: A generalised linear model (GLM)

▼ Code

```
# Make a folder to save outputs
dir.create("Outputs/GLM_outputs", showWarnings = F)
```

Null model

Null model: no explanatory variables or predictors are included.

It is always helpful to create a null model as a benchmark to assess how the inclusion of explanatory variables improves the model.

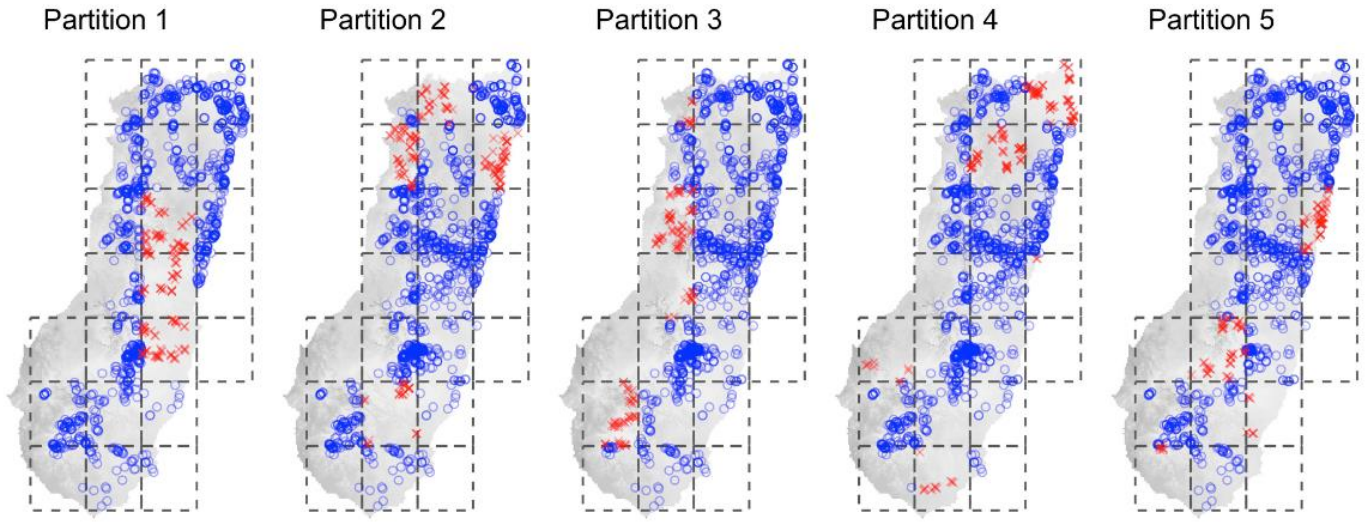
▼ Code

```
# Fit a null model with only the intercept
null_model <- glm(Presence ~ 1,
                  data = train_PB_covs_thinned_scaled,
                  family = binomial(link = "logit"))

# Check the model results
summary(null_model)
```

🤔 Validating SDMs

- Cross-validation (CV)
  - Hold out a subset of the data to test the model's predictions against
  - Spatial block CV



💡 *Truly independent presence-absence data is the gold standard*

Valavi et al. (2023)

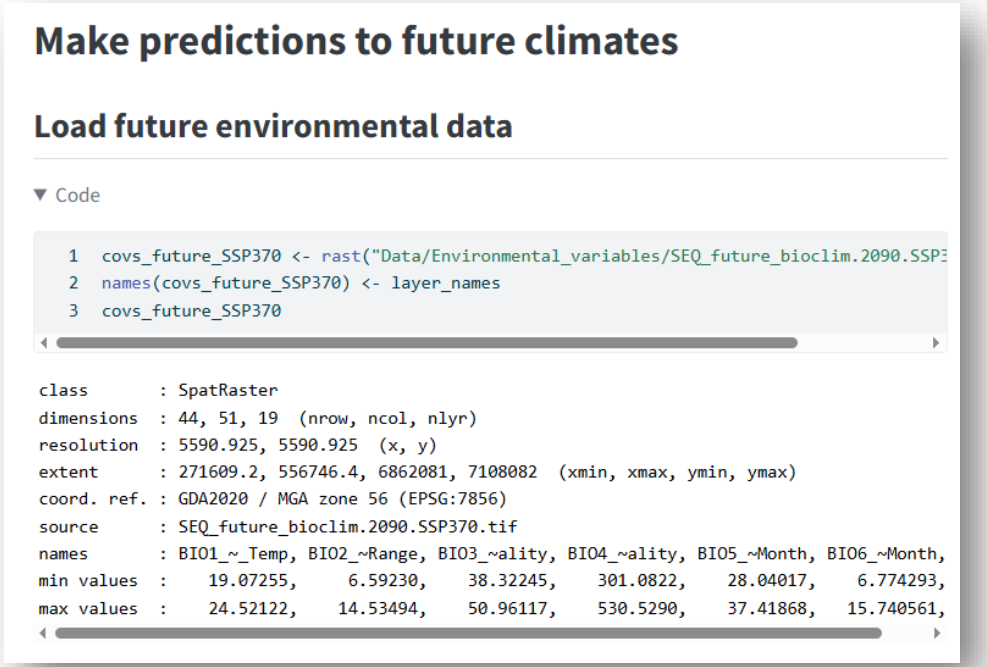
🤔 Validating SDMs – Evaluation metrics

- ‘Threshold-dependent’ or ‘Threshold-independent’
- Calibration (e.g., Boyce Index)
- Discrimination (e.g., AUC ROC)

Multiple metrics can describe different aspects of model performance – some more or less relevant to your study

Let’s predict the future for koalas

https://geospatial-community.github.io/ICCB_geospatial_tools_conservation



Make predictions to future climates

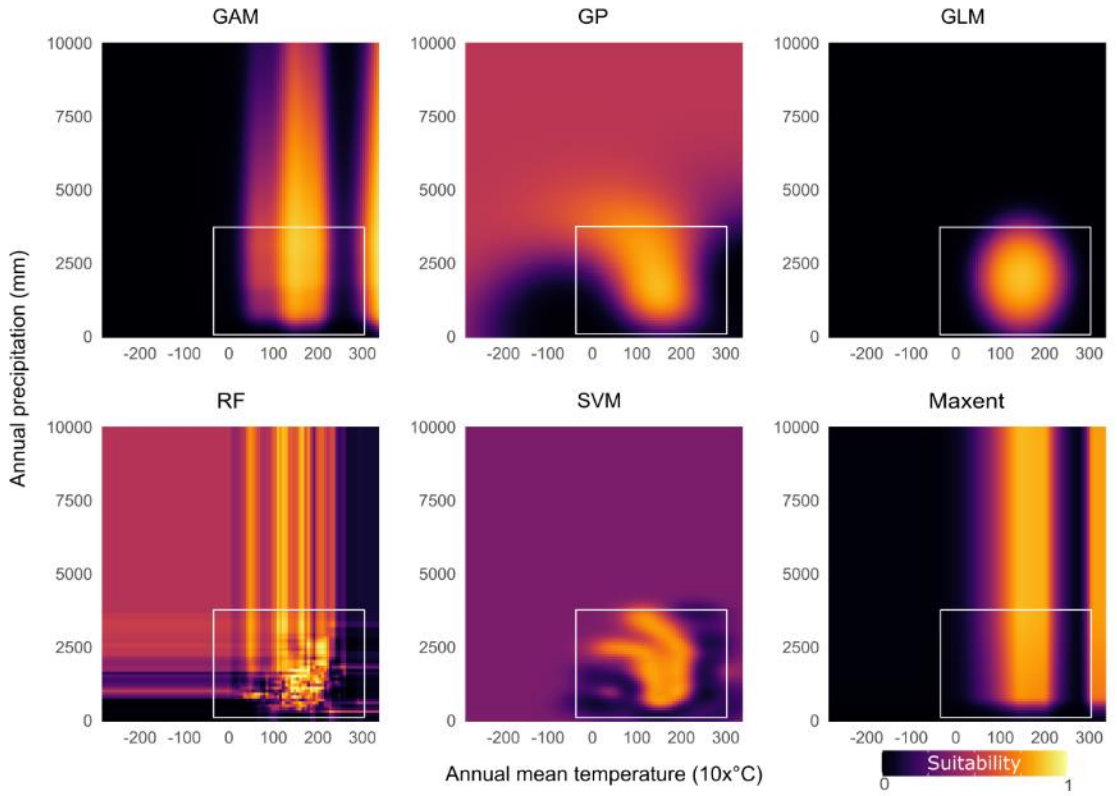
Load future environmental data

▼ Code

```
1  covs_future_SSP370 <- rast("Data/Environmental_variables/SEQ_future_bioclim.2090.SSP3
2  names(covs_future_SSP370) <- layer_names
3  covs_future_SSP370
```

```
class       : SpatRaster
dimensions  : 44, 51, 19  (nrow, ncol, nlyr)
resolution  : 5590.925, 5590.925  (x, y)
extent      : 271609.2, 556746.4, 6862081, 7108082  (xmin, xmax, ymin, ymax)
coord. ref. : GDA2020 / MGA zone 56 (EPSG:7856)
source      : SEQ_future_bioclim.2090.SSP370.tif
names       : BIO1_~_Temp, BIO2_~Range, BIO3_~ality, BIO4_~ality, BIO5_~Month, BIO6_~Month,
min values  :    19.07255,     6.59230,    38.32245,    301.0822,    28.04017,    6.774293,
max values  :    24.52122,    14.53494,    50.96117,    530.5290,    37.41868,   15.740561,
```

Predicting into the unknown & uncertainty



Velazco et al. (2023) **How far can I extrapolate my species distribution model? Exploring shape, a novel method.** Ecography.

Take home message

Modelling is often more art than science. There are many ways to model species, none of them the ‘right way’. Some can, however, be more appropriate for a study question or species.

We make many decisions along the way and need to be transparent and clear about these decisions.

Thanks everyone!