https://github.com/patch105/ICCB_koala_sdm

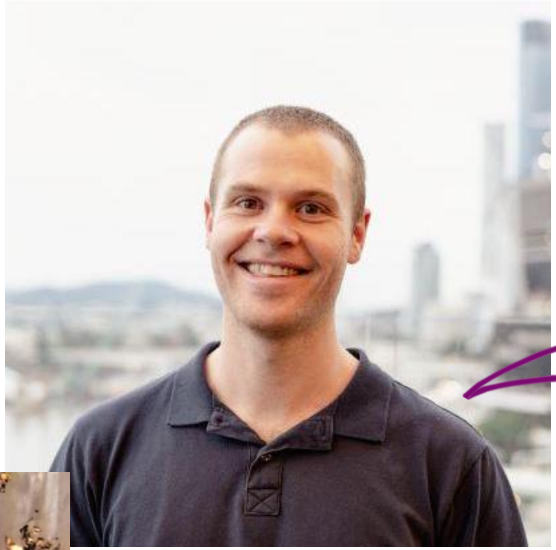
Modelling the SEQ koala distribution under current and future climate scenarios

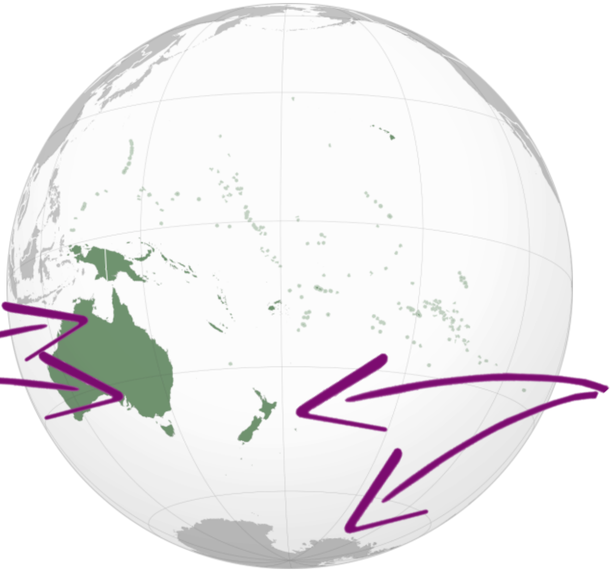
Charlotte Patterson, QUT
Scott Forrest, QUT, CSIRO

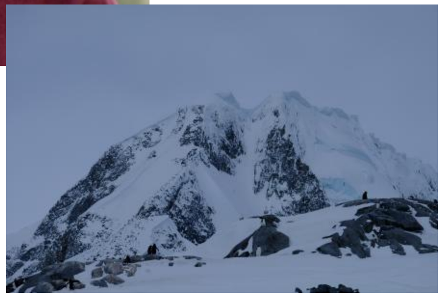
Acknowledgement of Traditional Owners

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

Who are we?







Who are you?

Open-source geospatial tools for conservation under climate change - A Koala case study

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

How will this workshop run?

- Workshop outline
  - Intro to SDMs
  - How to access the data/environmental layers
  - Climate projections
  - (Some) Koala ecology
  - Coded example (interspersed throughout)
    - Downloading data (koala records/environmental layers)
    - Environmental variable selection
    - Background sampling
    - Modelling
    - Prediction
    - Validation

What are we hoping you'll get out of it?

- Overview of SDMs
  - When they might be useful
- Some steps to prepare data and environmental layers
- Brief overview of methods, model fitting, prediction and validation
- Insight into some key considerations when developing models based on our experience and best-practice
- Resources to continue the journey

🤔 Flag spots where careful thought is needed for your SDM application

💡 Highlight key messages

Take home message

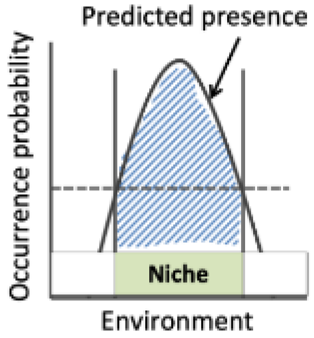
Modelling is often more art than science. There are many ways to model species, none of them the 'right way'. Some can, however, be more appropriate for a study question or species.

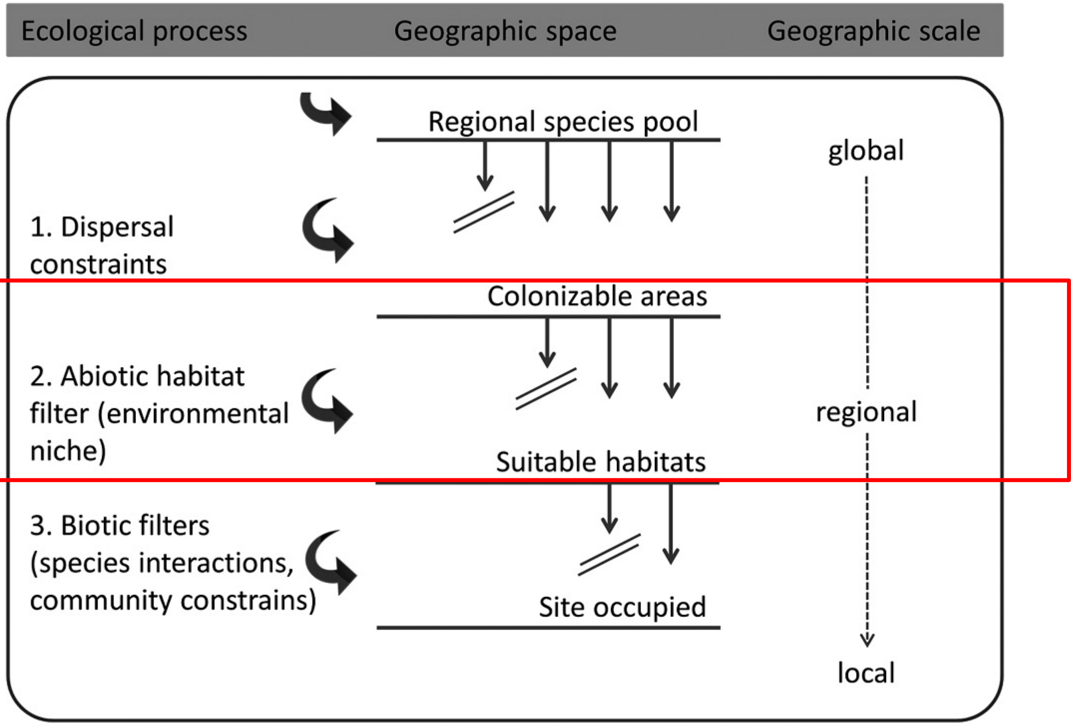
We make many decisions along the way and need to be transparent and clear about these decisions.

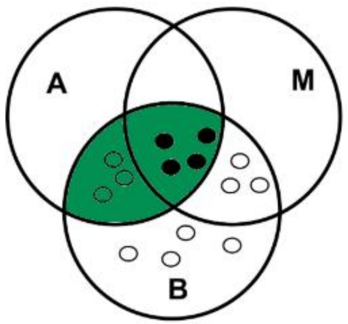
We also need to spend time understanding and communicating the effect of our decisions.

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

Ecological Niche Models; Habitat Suitability Models; Climate Envelope Models

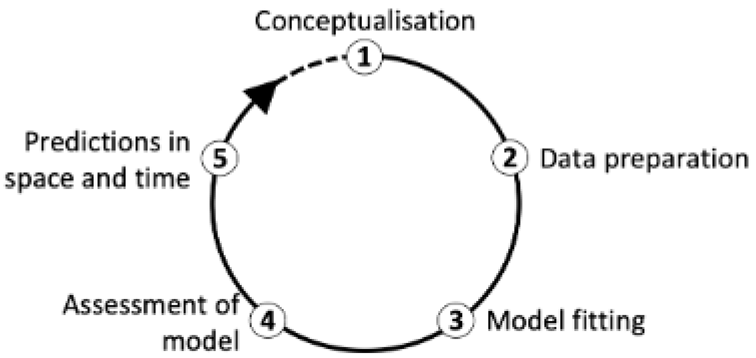




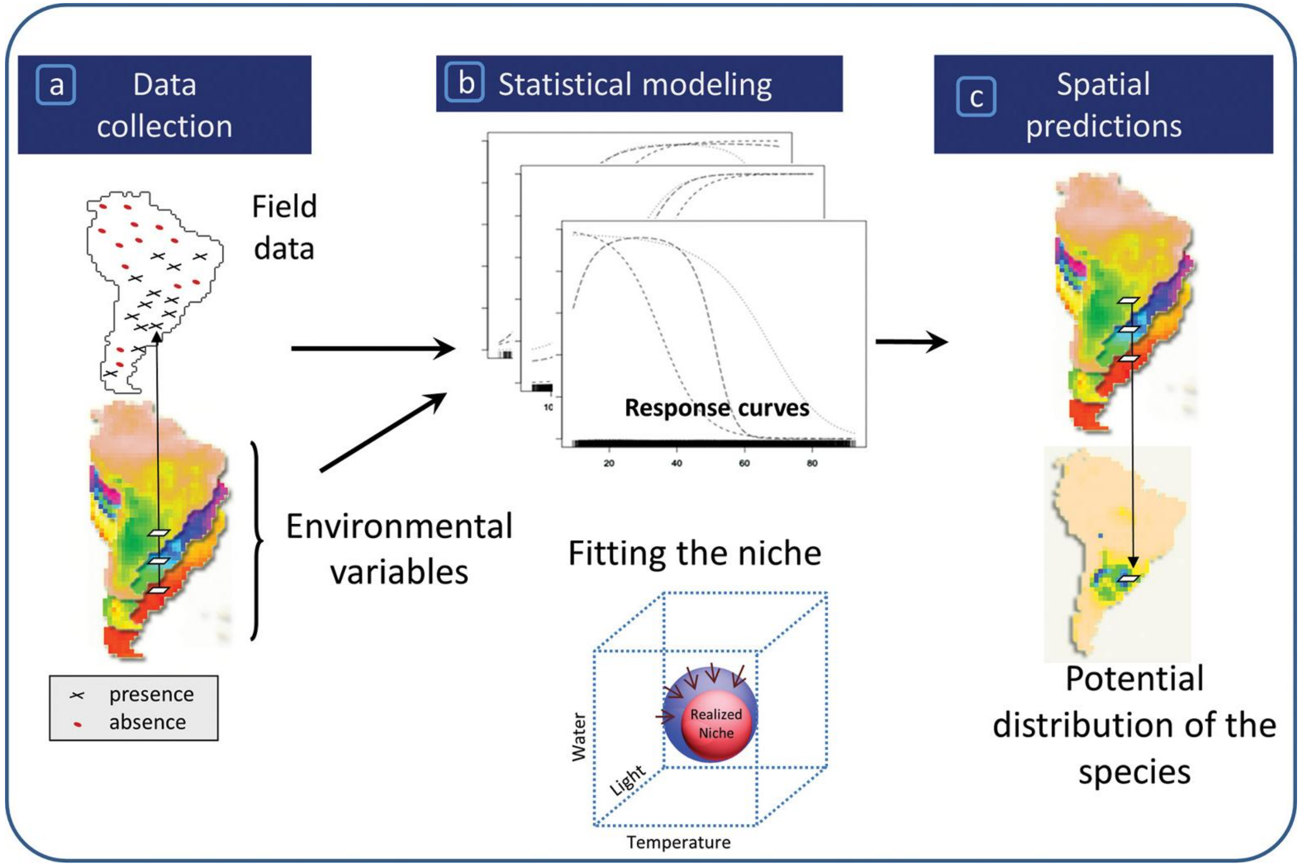


Guisan et al. (2017)

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



Zurrell et al. (2020)



Guisan et al. (2017)

So, you think you want to make an SDM…?

Perhaps you want to:

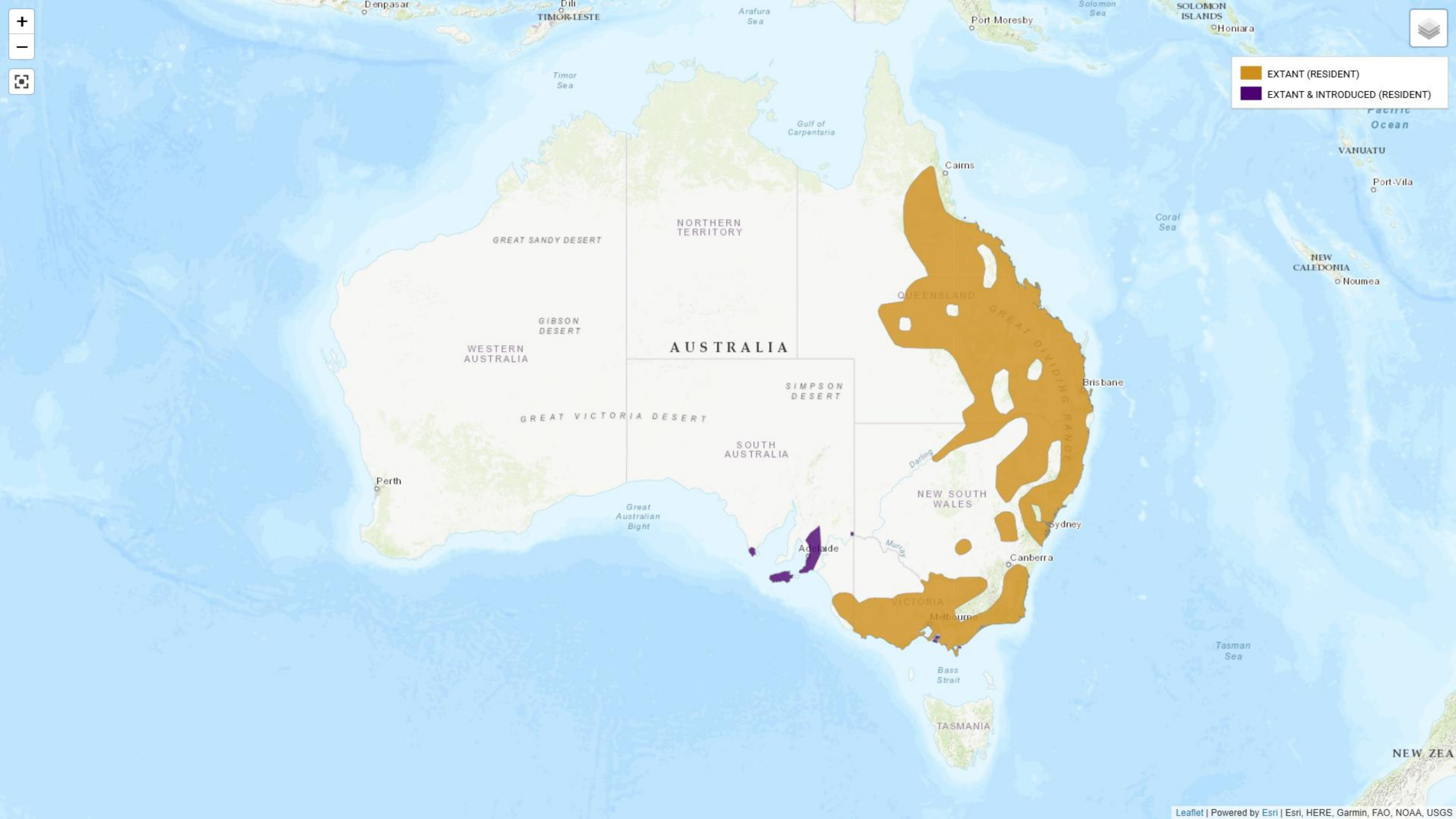
- Predict the distribution of a threatened species (e.g., for surveys)
- Plan where to put a protected area
- Assess the risk of a new invasive species
- Understand better what drives the dynamics of a study species

Koala (*Phascolarctos cinereus*)

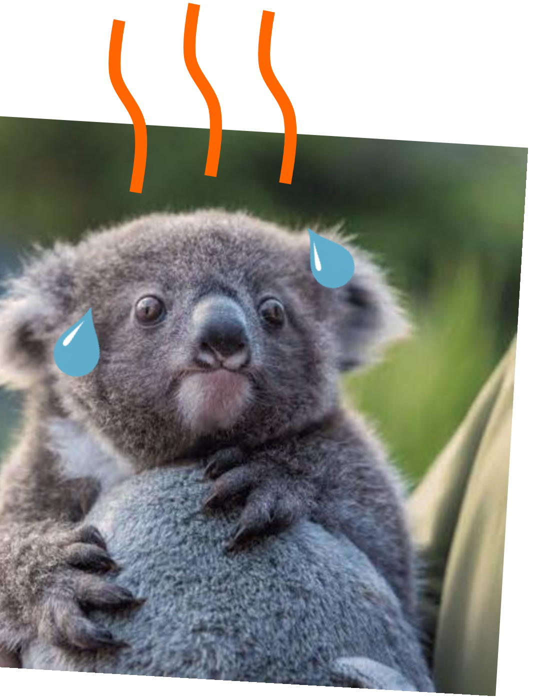
- The word "koala" comes from the Dharug *gula*, meaning 'no water’

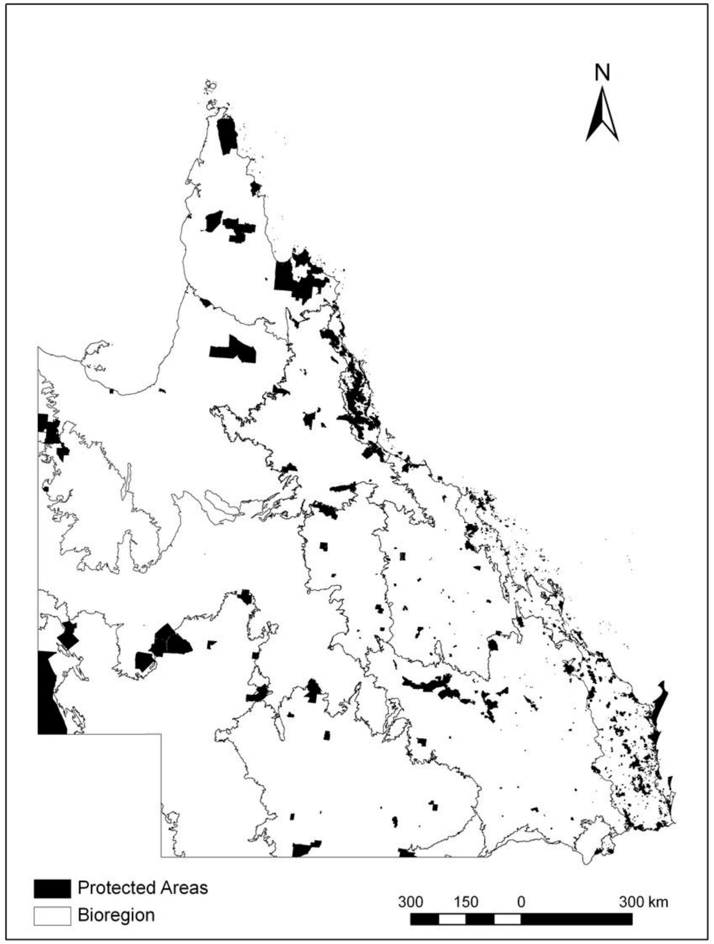
- *Phaskolos* = pouch, *arktos* = bear (Greek)
- *Cinereus* = ash coloured (Latin)

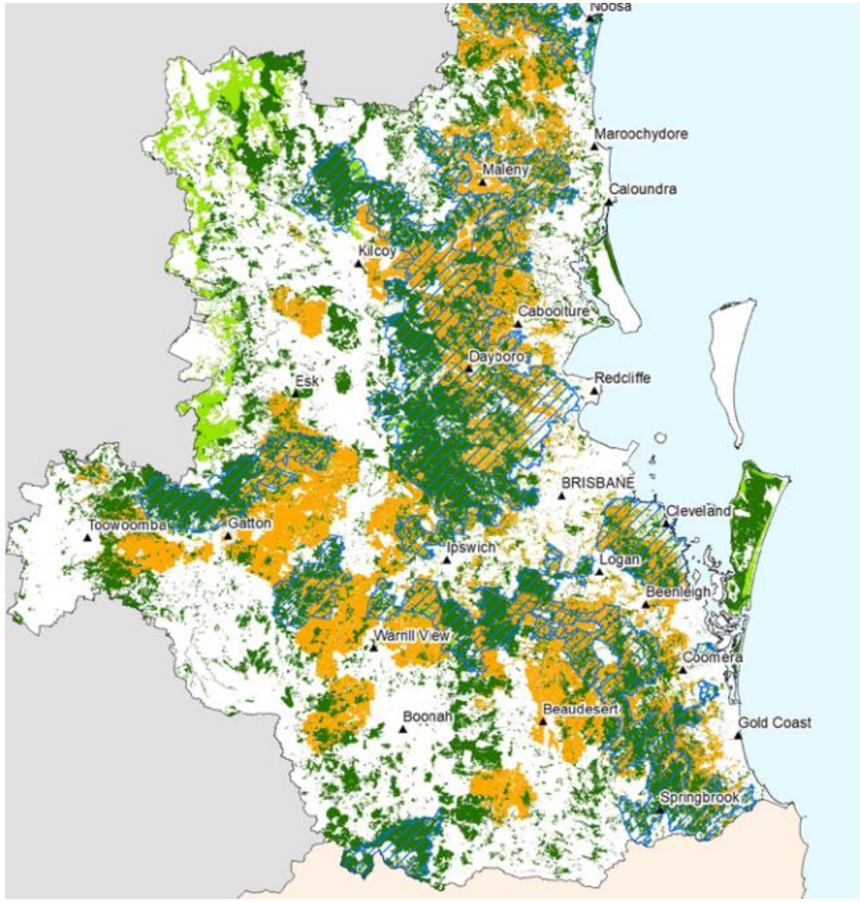
Photo by DAVID ILIFF.


EXTANT (RESIDENT)
EXTANT & INTRODUCED (RESIDENT)
AUSTRALIA
WESTERN AUSTRALIA
NORTHERN TERRITORY
QUEENSLAND
SOUTH AUSTRALIA
NEW SOUTH WALES
VICTORIA
TASMANIA
GREAT SANDY DESERT
GIBSON DESERT
GREAT VICTORIA DESERT
SIMPSON DESERT
GREAT DIVIDING RANGE
Perth
Adelaide
Melbourne
Canberra
Sydney
Brisbane
Cairns
Darling
Murray
Great Australian Bight
Bass Strait
Tasman Sea
Coral Sea
Gulf of Carpentaria
Timor Sea
Arafura Sea
Solomon Sea
Pacific Ocean
Denpasar
Dili
TIMOR-LESTE
Port Moresby
SOLOMON ISLANDS
Honiara
VANUATU
Port-Vila
NEW CALEDONIA
Noumea
NEW ZEA
Leaflet | Powered by Esri | Esri, HERE, Garmin, FAO, NOAA, USGS

Case Study: Protecting future koala habitat in South-East Queensland







South-East Queensland Koala Conservation Strategy, Qld Govt.

🤔 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

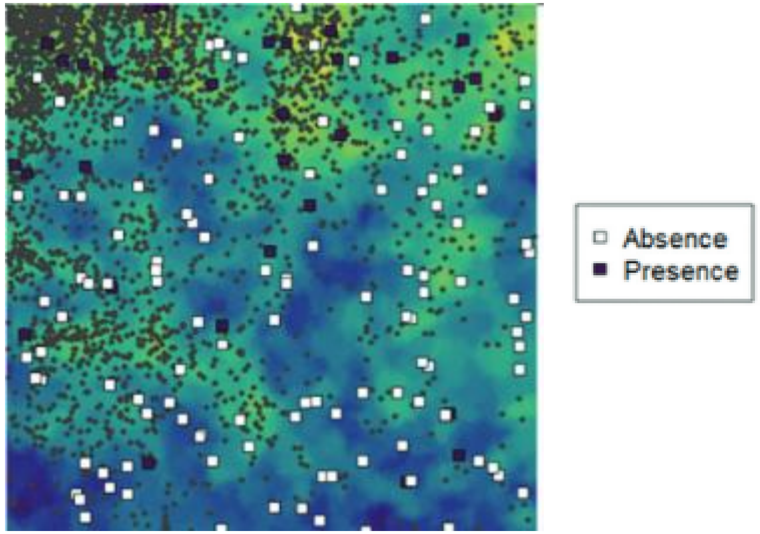
💡 *Clearly defining your model's purpose guides your workflow*

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

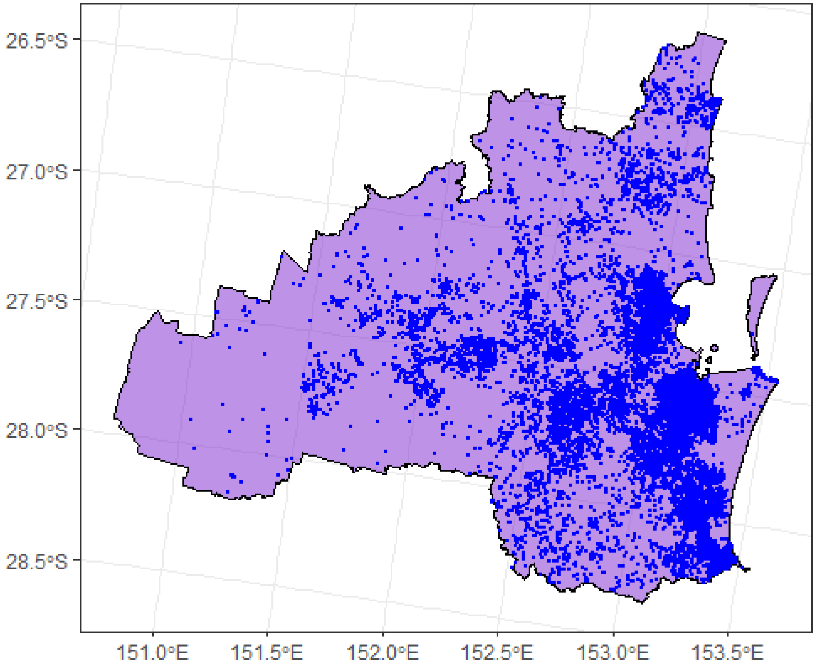


💡 *Your data limit what is possible with SDMs. Prioritise quality over quantity.*

Study region – South East Queensland

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

Koala data from the ALA



GBIF Global Biodiversity Information Facility

EcoCommons

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

galah ala.org.au

🤔 Background selection

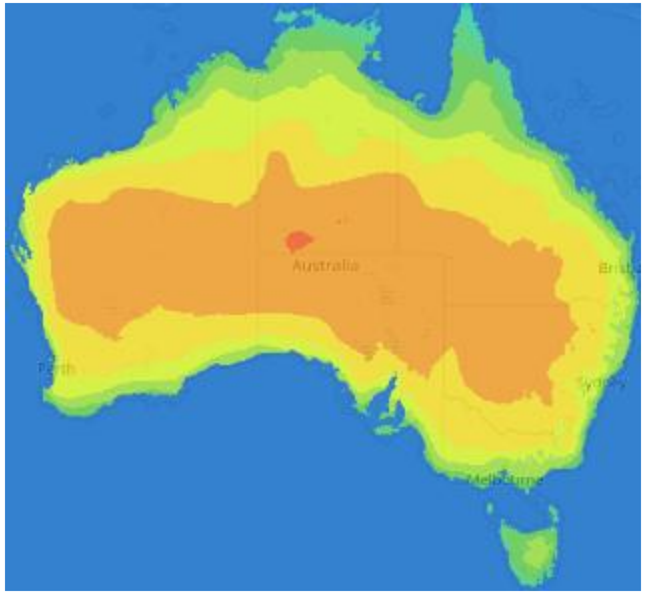
- Many ‘presence-only’ approaches rely on the selection of background or ‘pseudo-absence’ points
- These points are contrasted against environmental conditions where your species was found

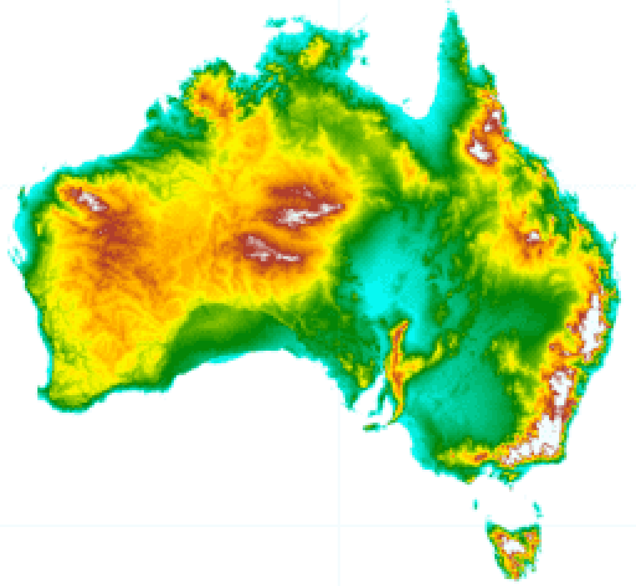
💡 *Background selection is a critical step in presence-only SDMs. Choices reflect your understanding of your study species.*

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

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

Photo by DAVID ILIFF.

Environmental covariate selection

Option 1. Expert consultation or selection based on species knowledge

Option 2. Correlation & multicollinearity checks

Option 3. Model selection

BIOCLIM Layers

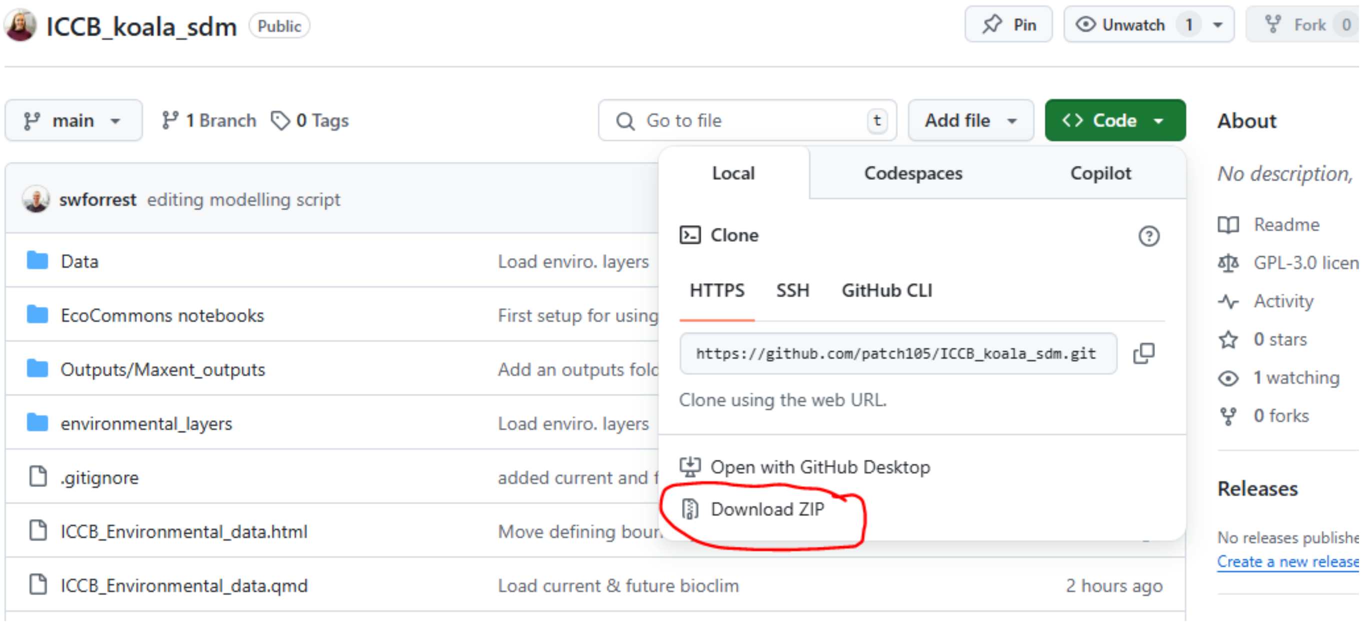
- **BIO1** = Annual Mean Temperature
- **BIO2** = Mean Diurnal Range (Mean of monthly (max temp - min temp))
- **BIO3** = Isothermality (BIO2/BIO7) (×100)
- **BIO4** = Temperature Seasonality (standard deviation ×100)
- **BIO5** = Max Temperature of Warmest Month
- **BIO6** = Min Temperature of Coldest Month
- **BIO7** = Temperature Annual Range (BIO5-BIO6)
- **BIO8** = Mean Temperature of Wettest Quarter
- **BIO9** = Mean Temperature of Driest Quarter
- **BIO10** = Mean Temperature of Warmest Quarter
- **BIO11** = Mean Temperature of Coldest Quarter
- **BIO12** = Annual Precipitation
- **BIO13** = Precipitation of Wettest Month
- **BIO14** = Precipitation of Driest Month
- **BIO15** = Precipitation Seasonality (Coefficient of Variation)
- **BIO16** = Precipitation of Wettest Quarter
- **BIO17** = Precipitation of Driest Quarter
- **BIO18** = Precipitation of Warmest Quarter
- **BIO19** = Precipitation of Coldest Quarter

Conceptualisation & Data Preparation Resources

- Guillera-Arroita et al. (2015). **Is my species distribution model fit for purpose? Matching data and models to applications.** *Global ecology and biogeography*.
- Guisan et al. (2017). ***Habitat suitability and distribution models: with applications in R*.** Cambridge University Press.
- Araújo et al. (2019). **Standards for distribution models in biodiversity assessments.** *Science advances*.
- Zurell et al. (2020). **A standard protocol for reporting species distribution models.** *Ecography*.

Have a go: Data preparation

https://github.com/patch105/ICCB_koala_sdm



Have a go: Data preparation

https://github.com/patch105/ICCB_koala_sdm

ICCB_koala_sdm.Rproj

ICCB_Environmental_data.qmd

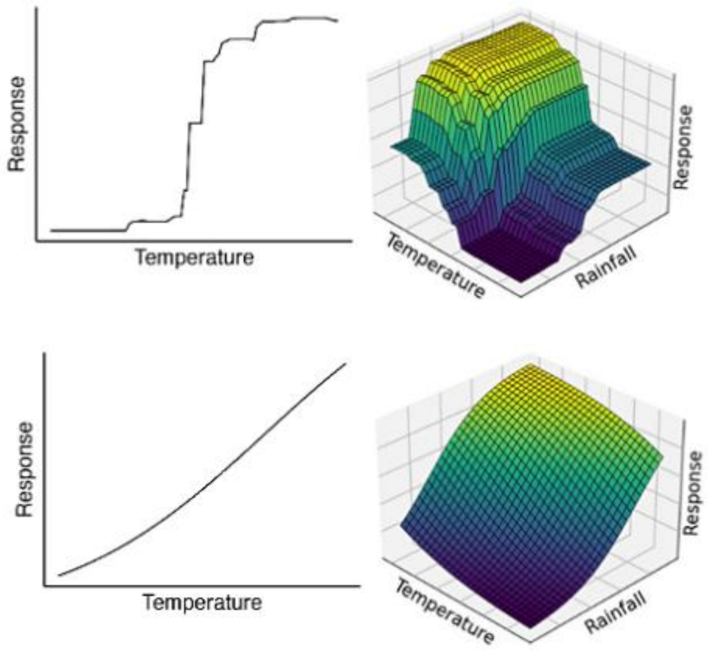
ICCB_Species_data.qmd

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



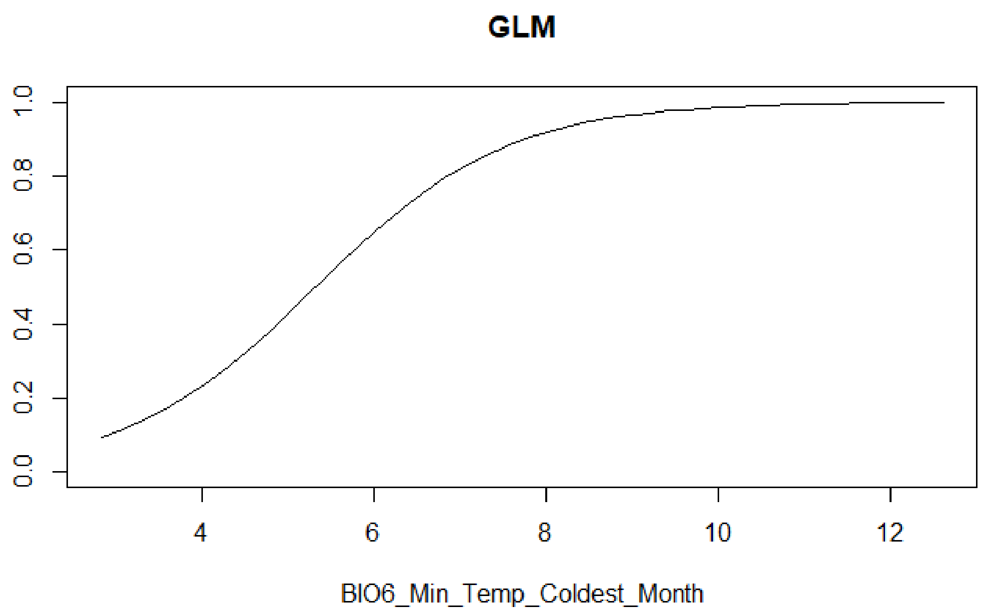
Generalised Linear Model (GLM)

- Regression
- Can fit
  - Linear: $x_1 + x_2 + \ldots + x_n$
  - Quadratic: $\mathbf{x_1 + x_1^2} + x_2 + \ldots + x_n$
  - Higher order polynomial terms: $\mathbf{x_1 + x_1^2 + x_1^3} + x_2 + \ldots + x_n$

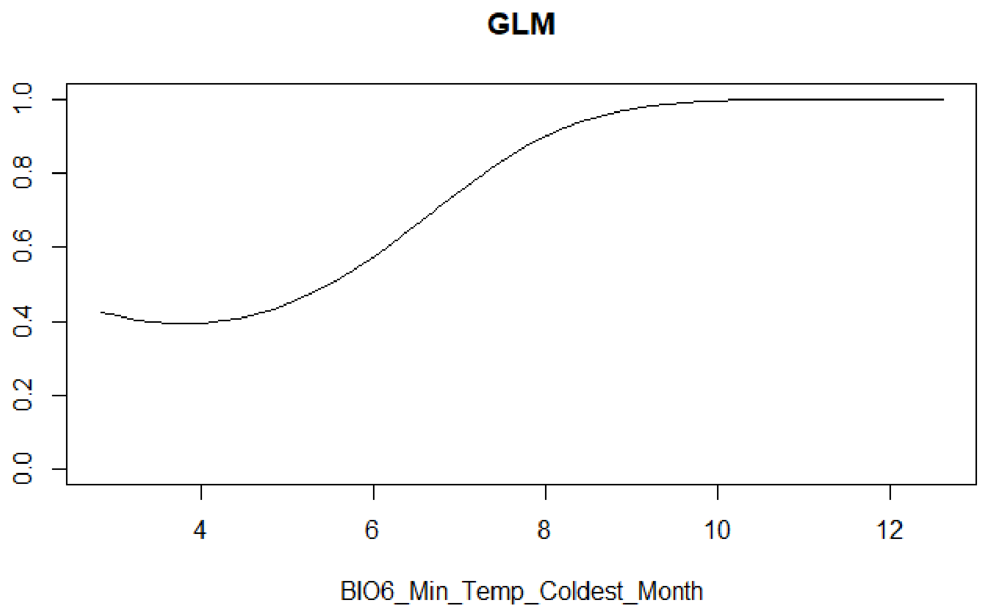
  - Interactions between covariates: $x_1 + x_2 + x_1 * x_2 + \ldots + x_n$

Partial response plots (response curves)

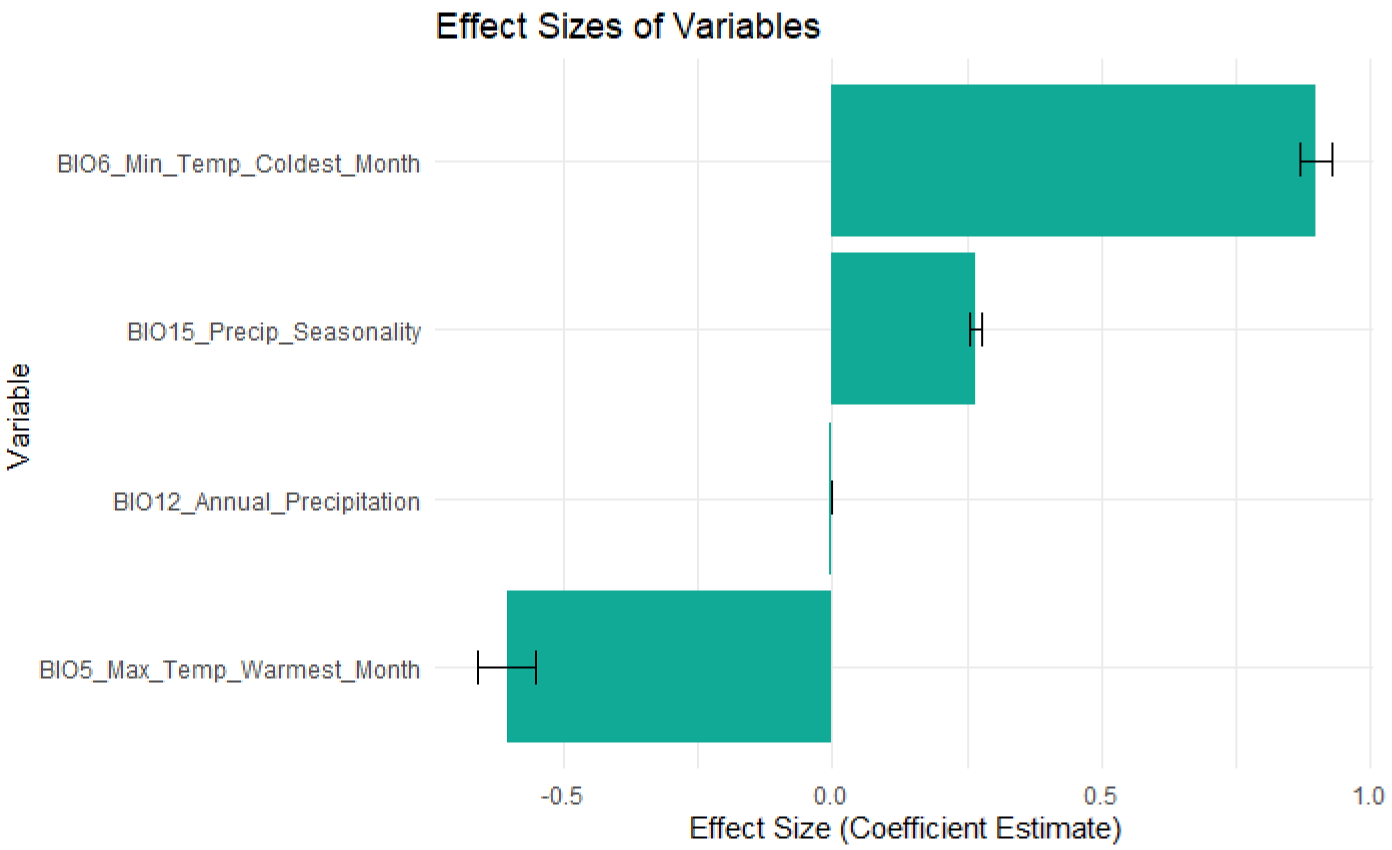
Linear term



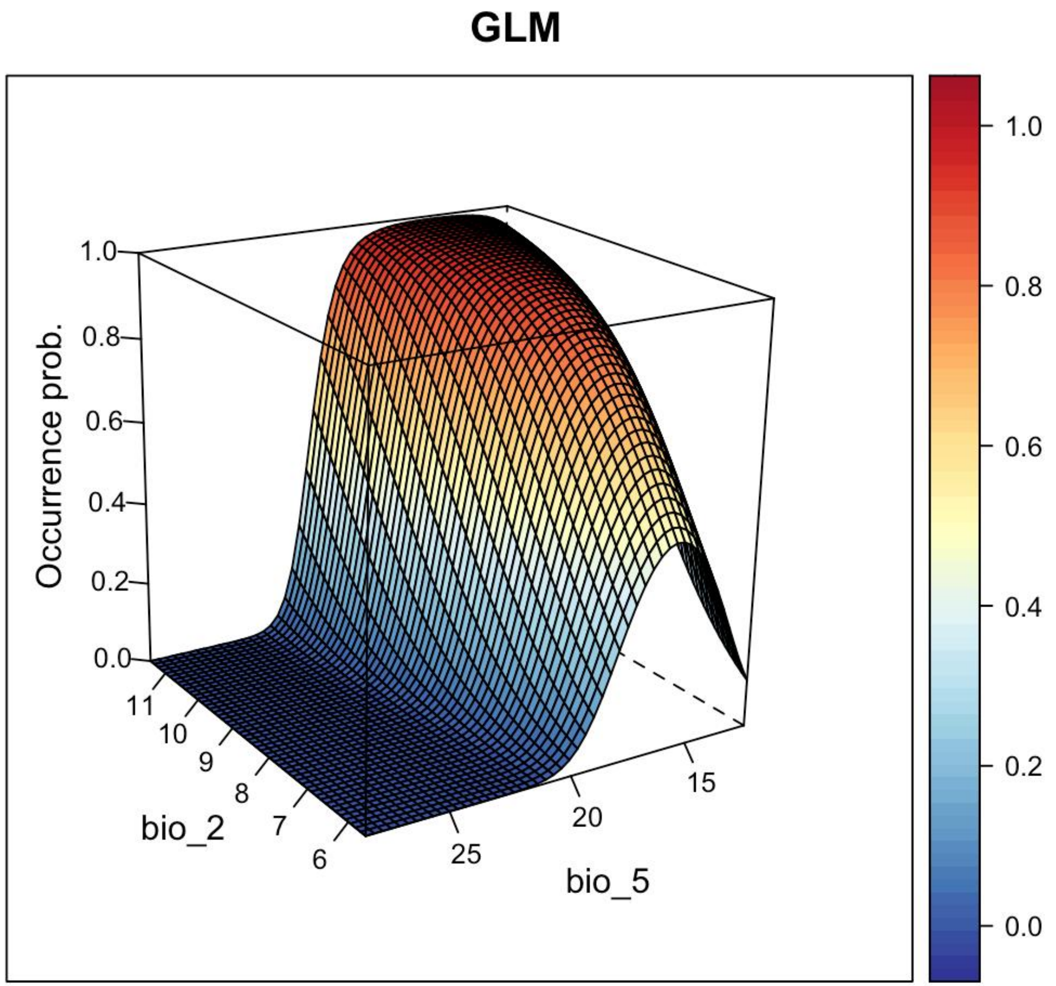
Linear + quadratic term



Effect sizes



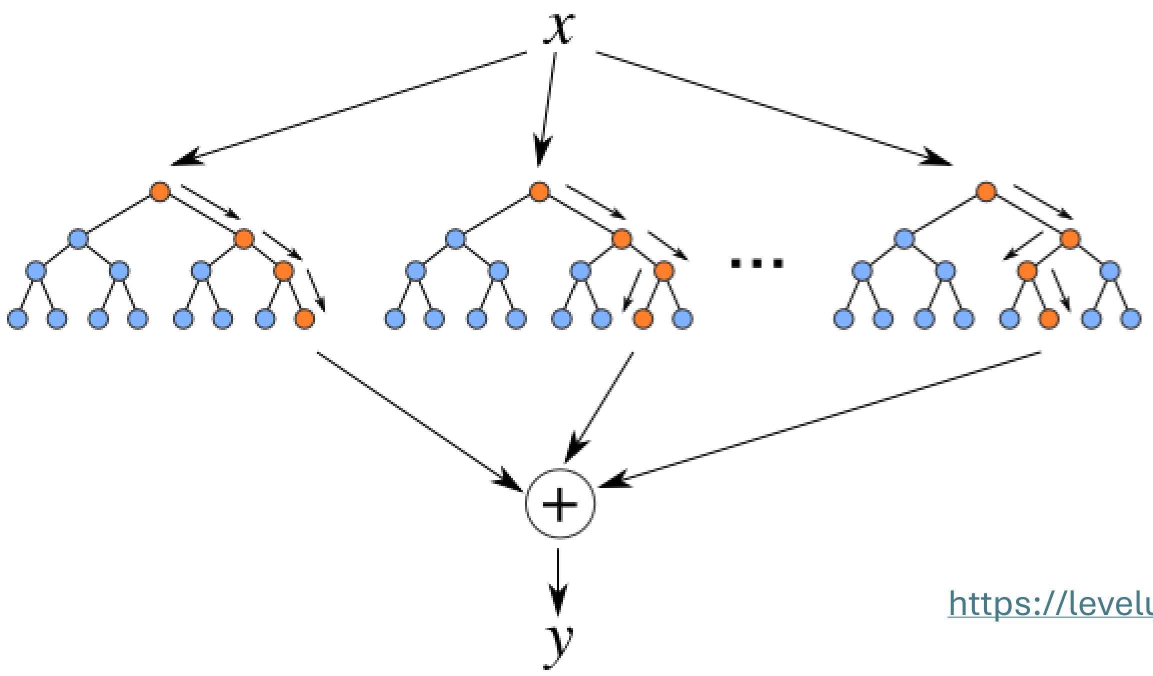
Interactions between covariates



from: https://damariszurell.github.io/SDM-Intro/

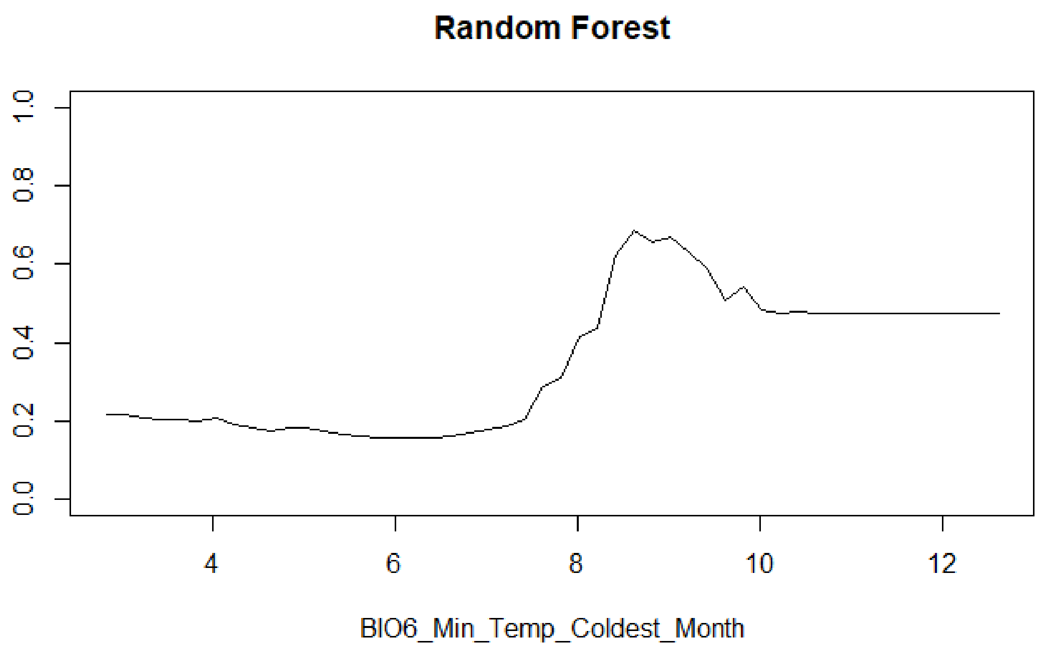
Random Forest (RF)

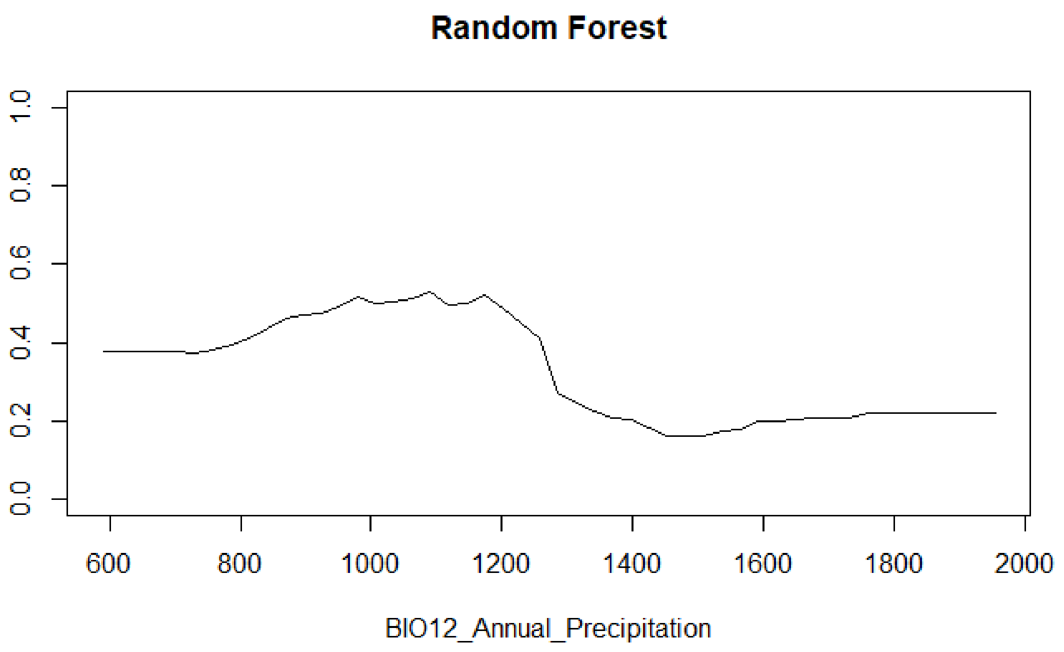
- Classification (binary) or regression
- Ensemble of many decision trees (hence ‘forest’)



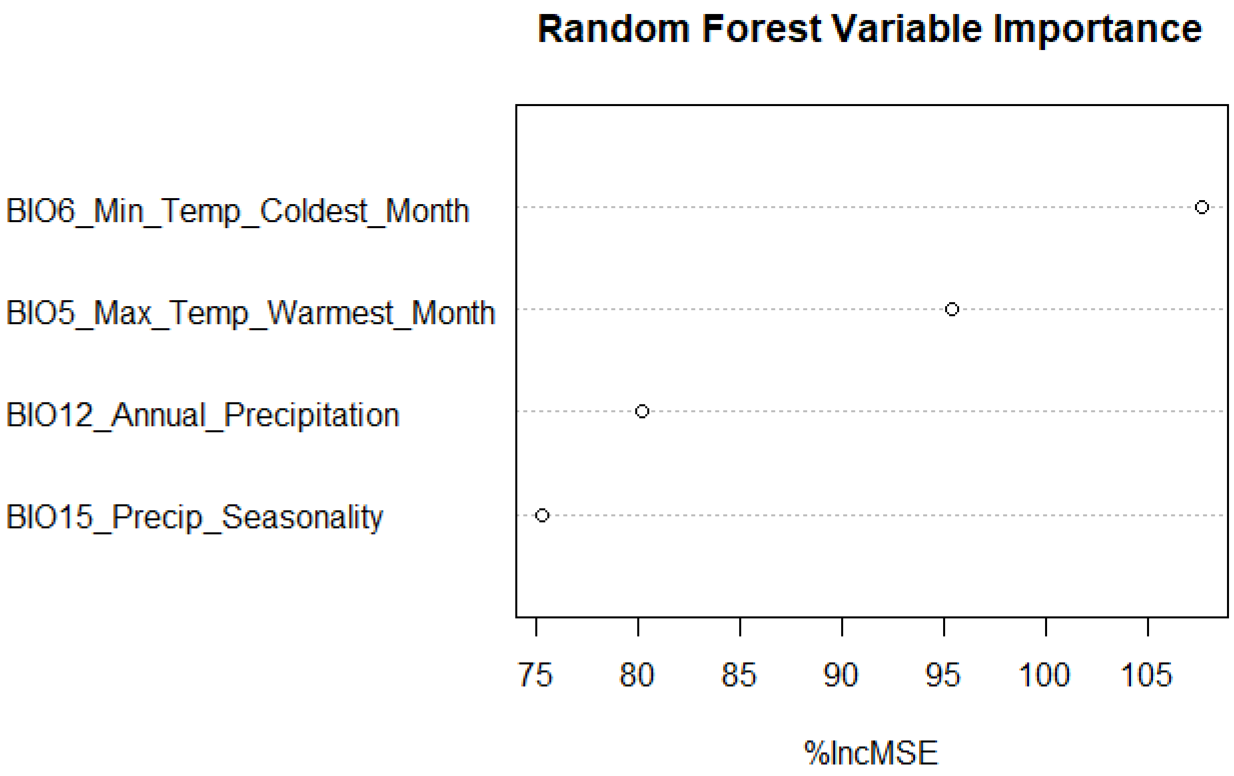
https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

Partial response plots





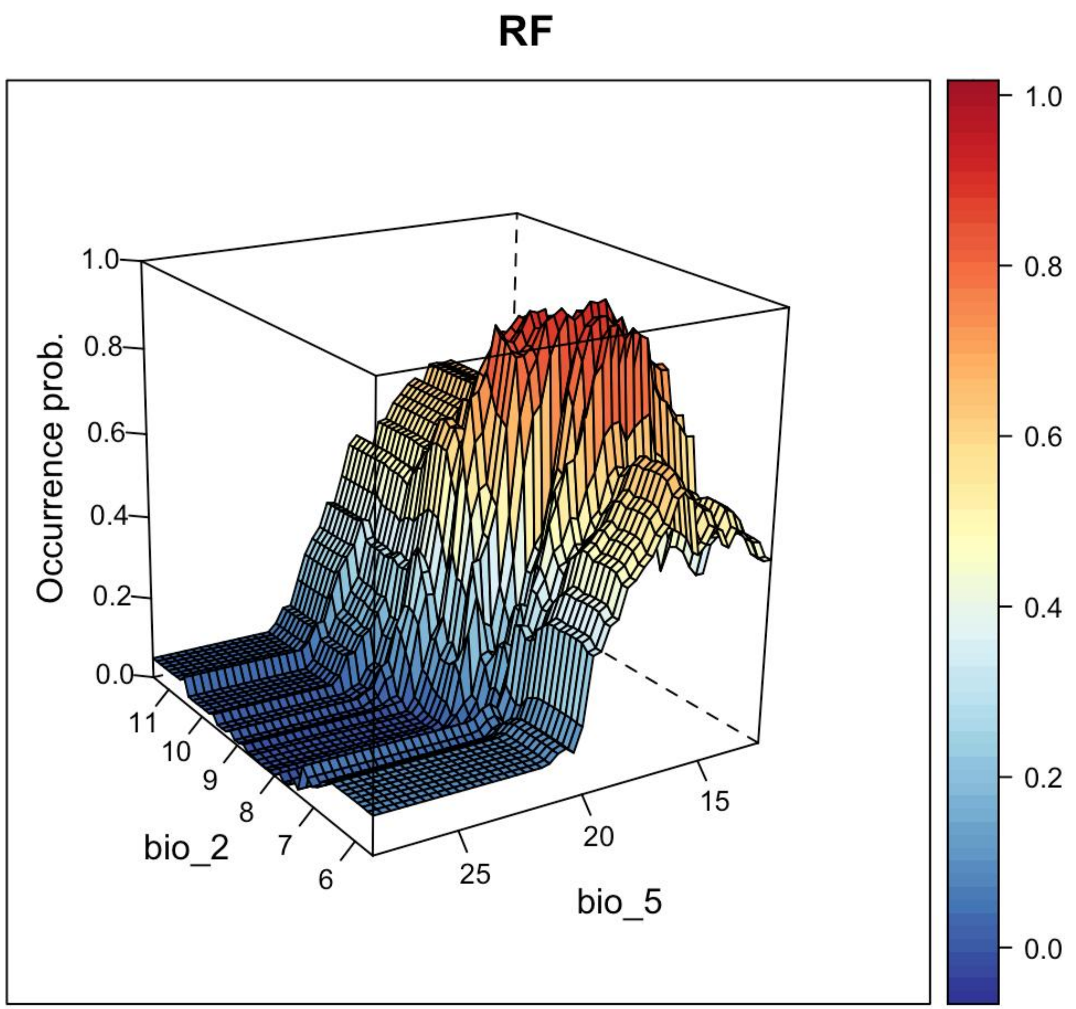
Variable Importance



%IncMSE = change in mean squared error when the variable is permuted (shuffled)

Essentially, how much does this variable affect the predictions?

Interactions between covariates



from: https://damariszurell.github.io/SDM-Intro/

Have a go: Data preparation

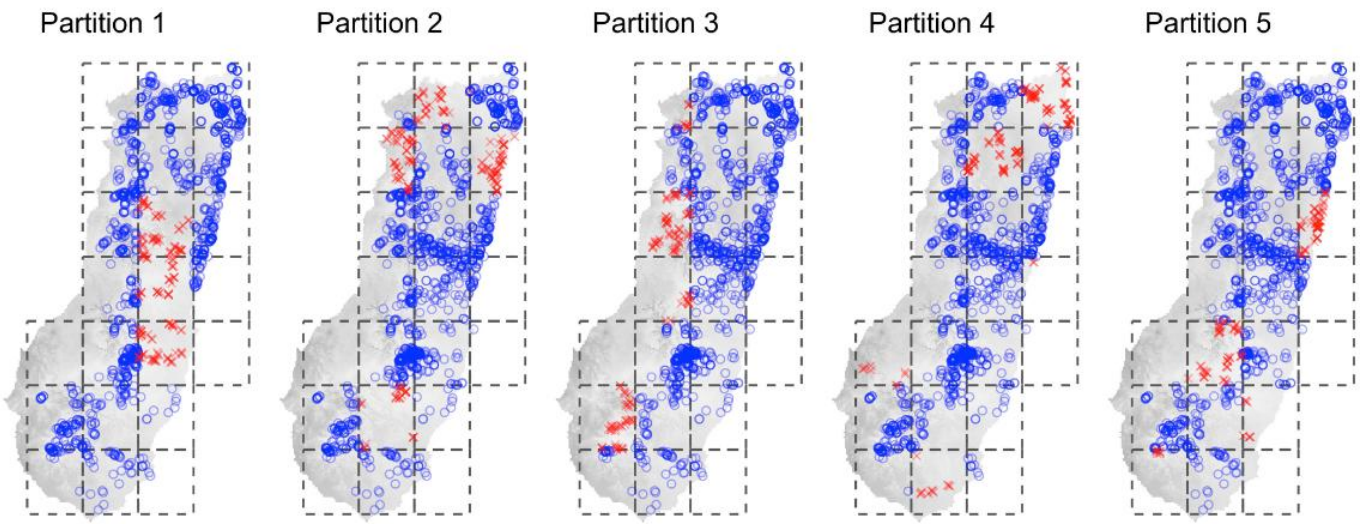
https://github.com/patch105/ICCB_koala_sdm

ICCB_Modelling_and_validation.qmd

Generating predictions

🤔 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)

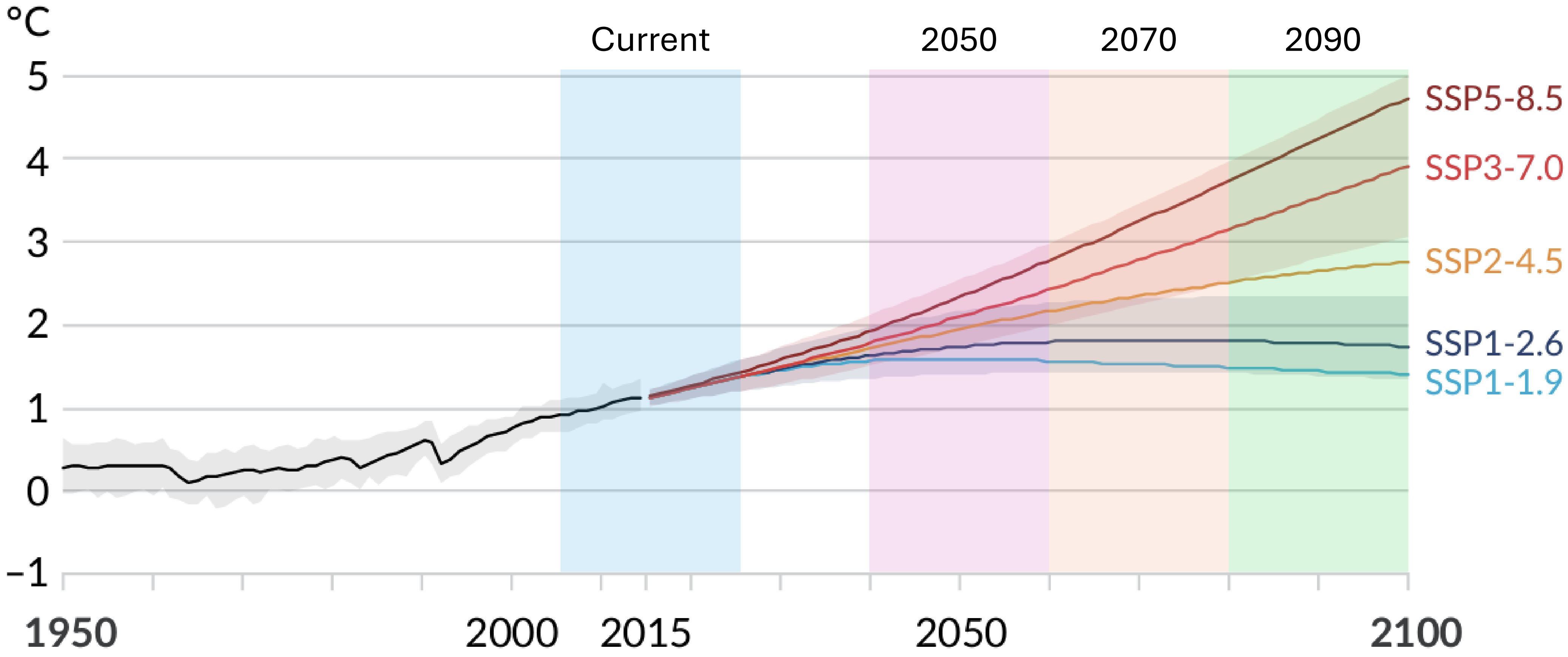
🤔 Validating SDMs – Evaluation metrics

- ‘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*

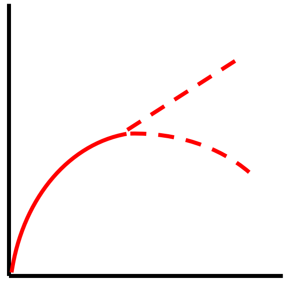
Valavi et al. (2023)

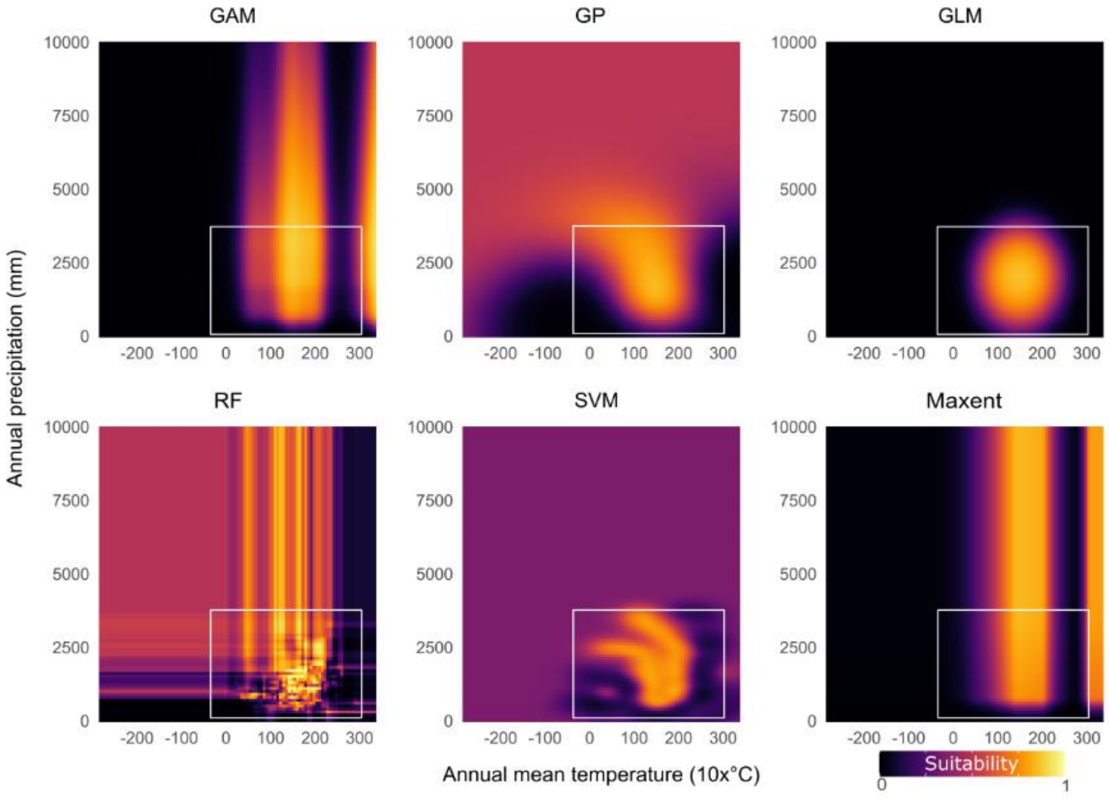
Climate projections



Global surface temperature change relative to 1850-1900. Projected global average surface temperature change in each of the five SSP scenarios. Source: [IPCC]

Predicting into the unknown





Velazco et al. (2023)

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.

We also need to spend time understanding and communicating the effect of our decisions.

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

Key resources

- Guillera-Arroita et al. (2015). **Is my species distribution model fit for purpose? Matching data and models to applications.** *Global ecology and biogeography*.
- Guisan et al. (2017). ***Habitat suitability and distribution models: with applications in R*.** Cambridge University Press.
- Araújo et al. (2019). **Standards for distribution models in biodiversity assessments.** *Science advances*.
- Zurell et al. (2020). **A standard protocol for reporting species distribution models.** *Ecography*.