Marine Ecological Modelling
Global Climate Change

Principles of Ecological Niche Modelling

Jorge Assis, PhD // jmassis@ualg.pt // biodiversitydatascience.com
2023, Centre of Marine Sciences, University of Algarve

Ecological Niche Modelling**

Process of using **computer algorithms to estimate and predict** the **relationship between the distribution of biodiversity and the environmental conditions**.

Provides insights about **species environmental tolerances and habitat preferences**, and allows **making spatial predictions** of geographical distributions.

** also known as environmental niche modelling, species distribution modelling, habitat distribution modelling, …

Main approaches in ENM

Mechanistic modelling

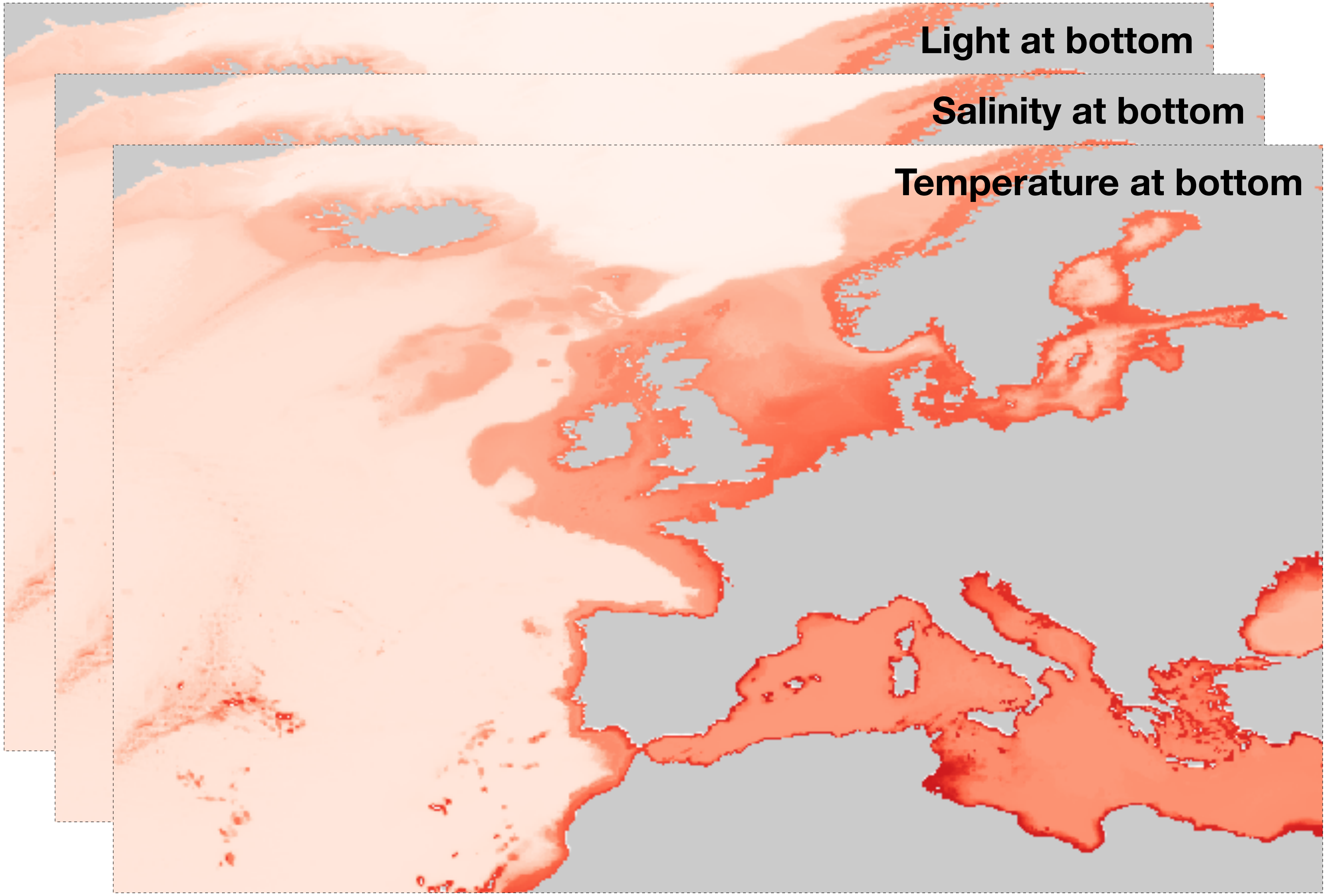
Uses **information about the physiological response of species to environmental conditions.**

(e.g., needs data on the effect of temperature on species survival; not always available).

Correlative modelling

Based on the statistical **correlation between presence records and the environment**, under the assumption that the **distribution of a species is an indicator of its environmental requirements.**

(i.e., niche theory; the fact that a species occurs is a particular place is linked to its tolerance to the conditions of such place).



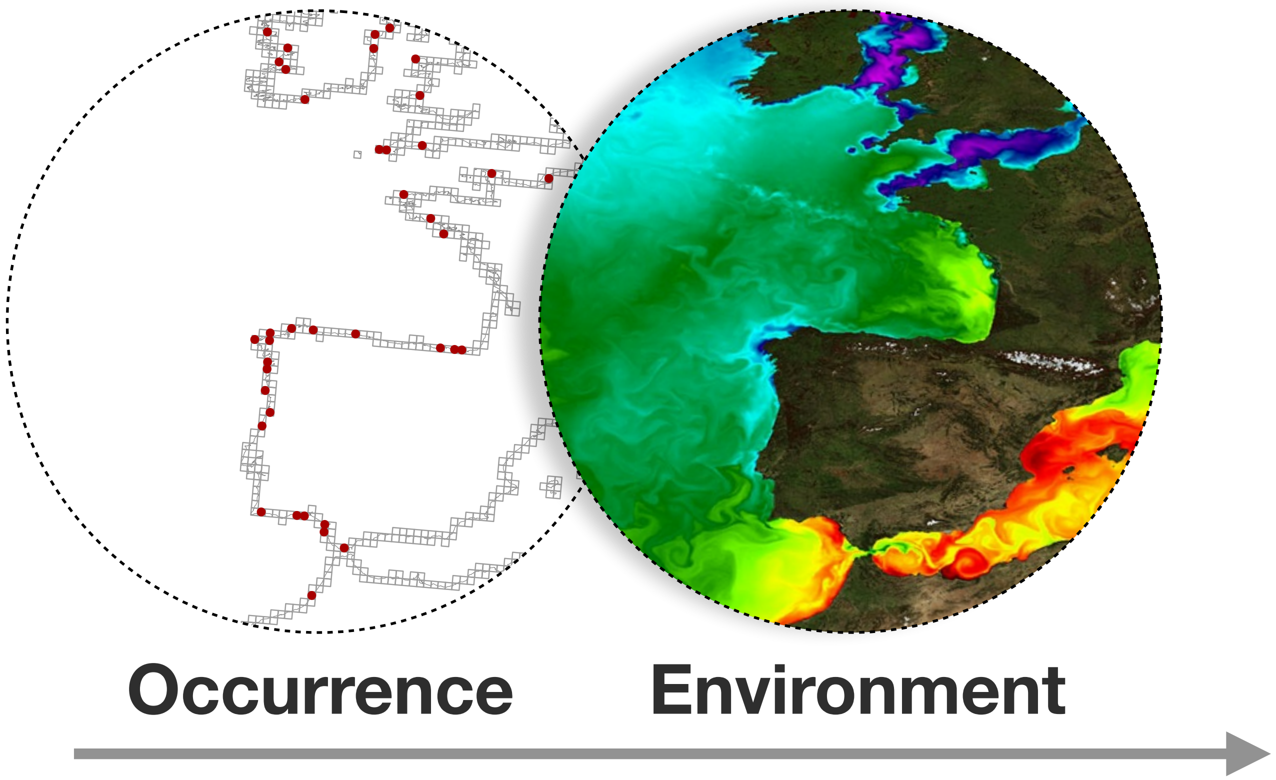
Mechanistic distribution models

Built by **reclassifying environmental gradients with tolerance limits inferred from physiological experiments.**

Mechanistic distribution models

A straightforward approach to predict the distribution of species.

Presence = [**Light** > 5 $E.m^2.year^{-1}$] ∩ [5ºC <= **Temperature** <= 20.5ºC]

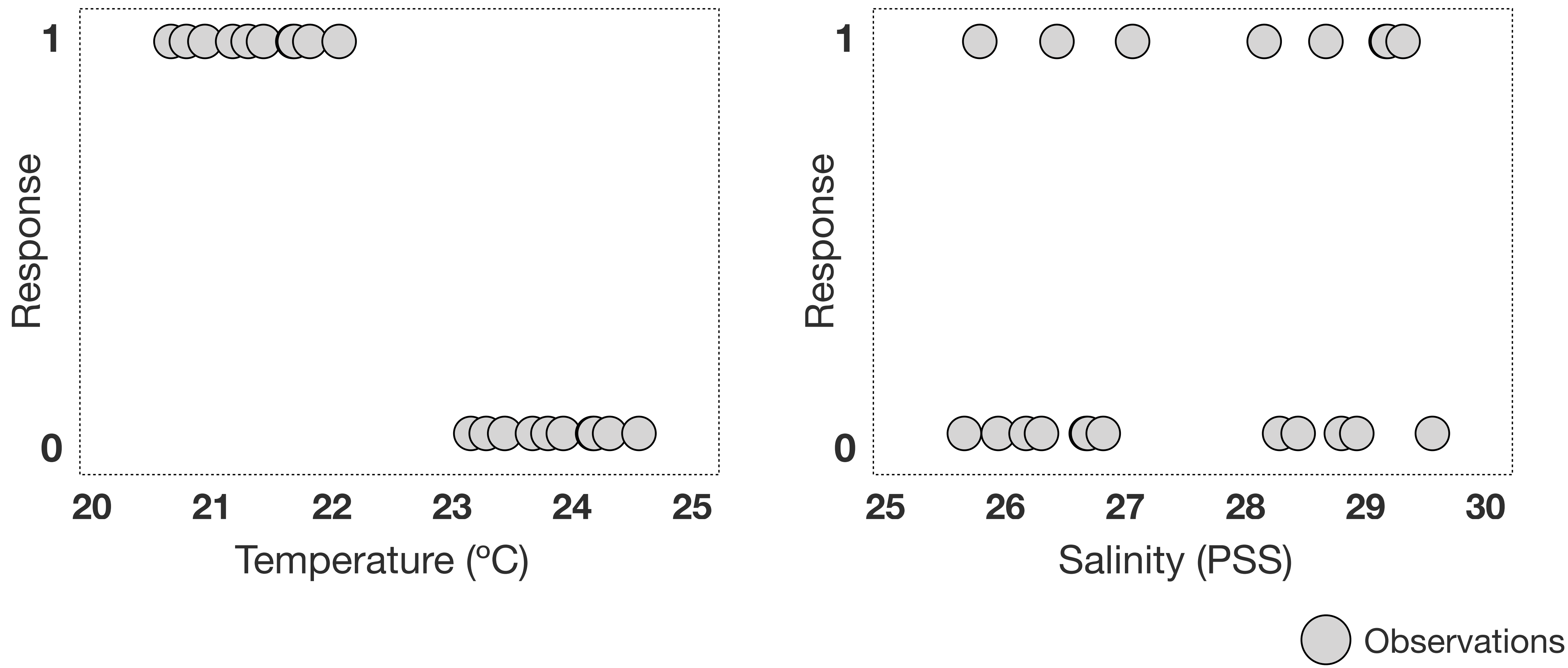


Resp	TempMax	Nitrate	Salinity
1	21	3	27
1	22	2	28
1	21	3	30
1	20	3	26
1	21	2	26
1	22	2	26
0	23	1	27
0	24	0	30
0	23	0	28
0	25	1	27
0	23	0	26
0	23	0	26

Data for modelling

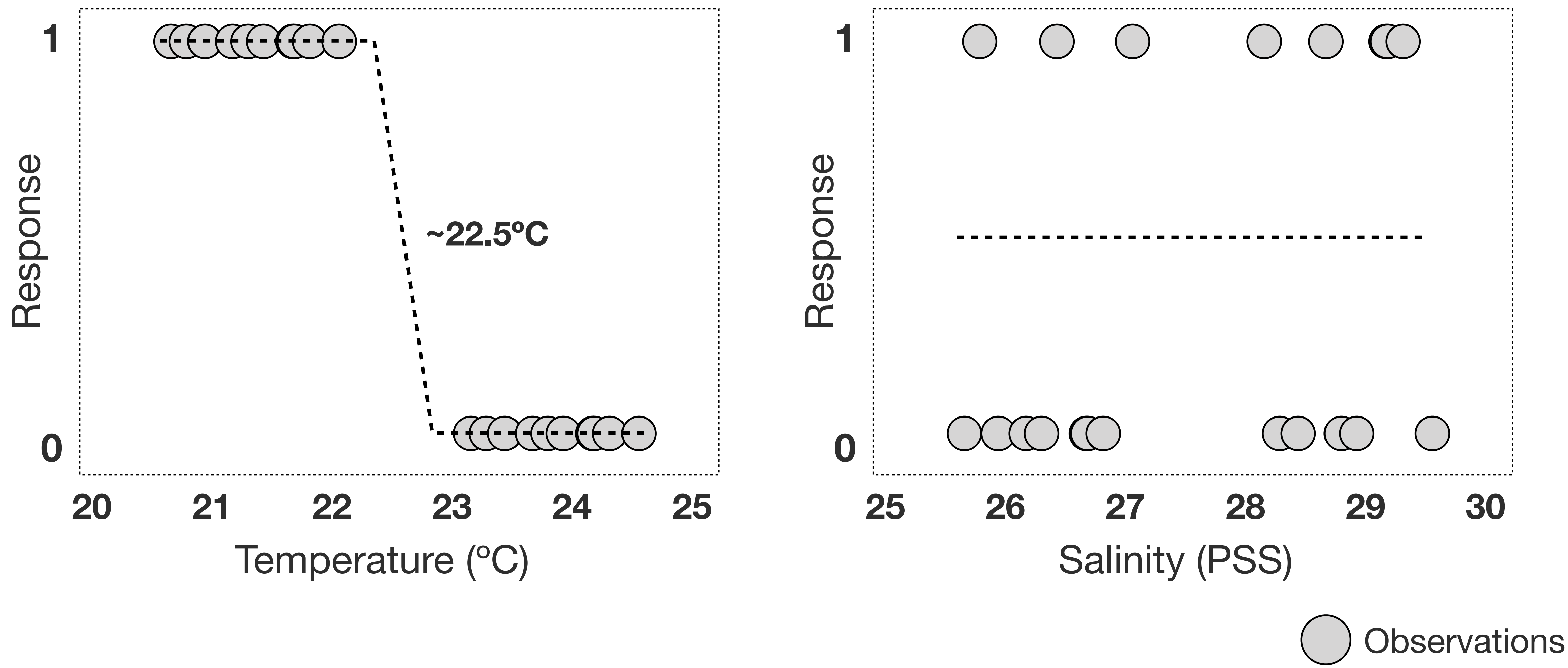
Correlative distribution models

Describe the **statistical relationship between distribution records and environmental conditions at those sites.** The **models should be evaluated for "ecological realism" - consistency with ecological knowledge of limiting factors** and species response curves.



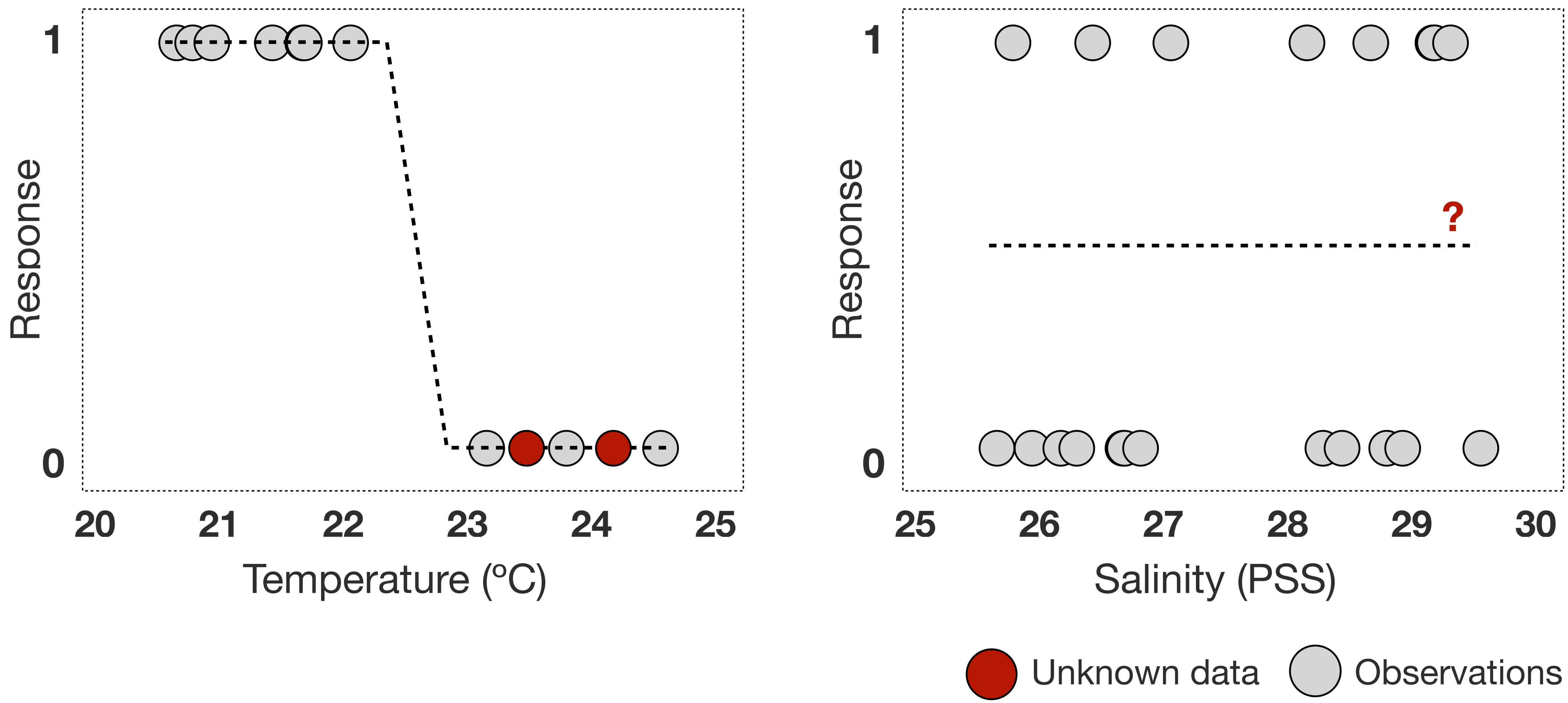
Correlative distribution models

Describe the **statistical relationship between distribution records and environmental conditions at those sites.** The **models should be evaluated for "ecological realism" - consistency with ecological knowledge of limiting factors** and species response curves.



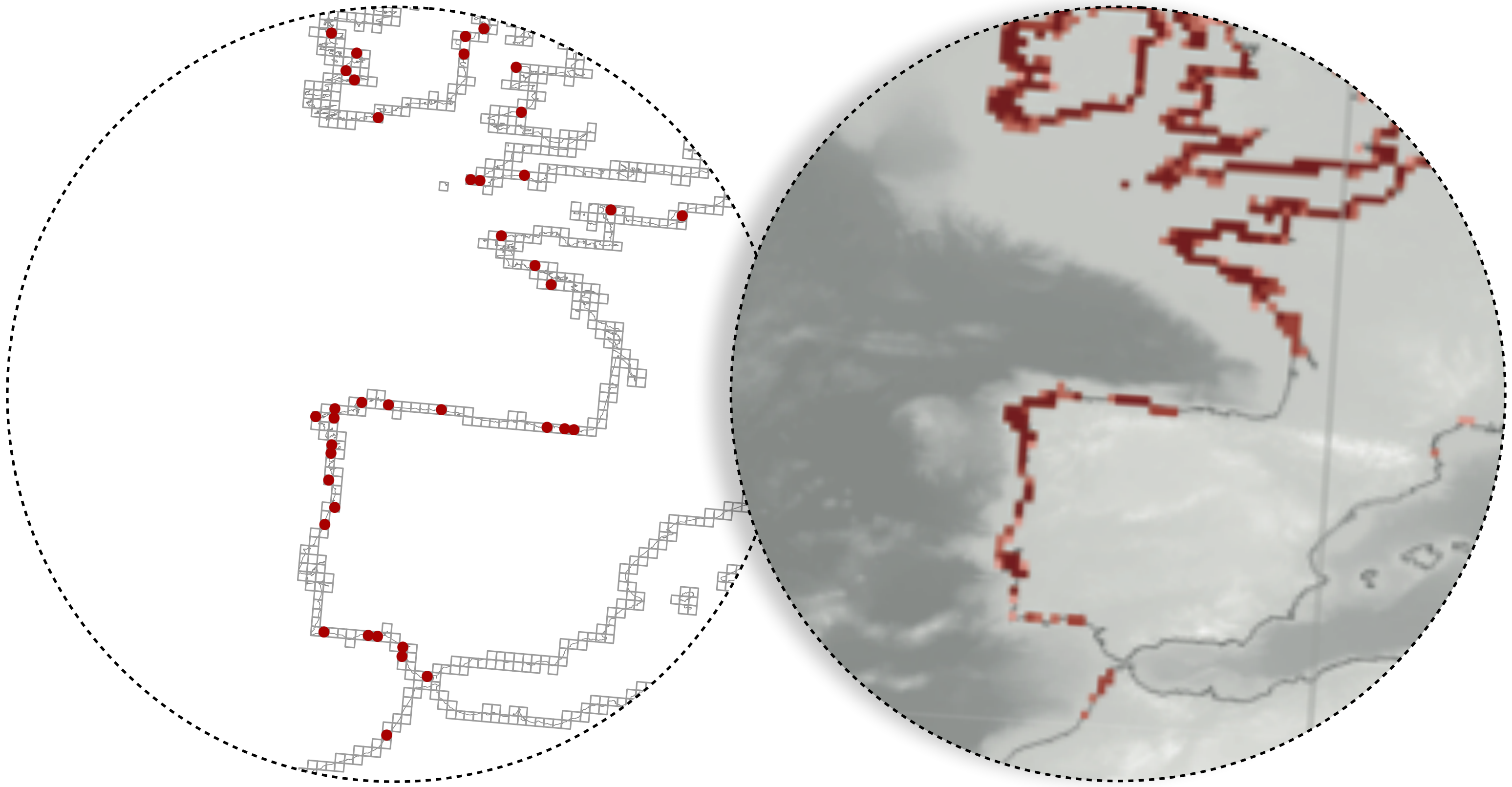
Correlative distribution models

Describe the **statistical relationship between distribution records and environmental conditions at those sites.** The **models should be evaluated for “ecological realism” - consistency with ecological knowledge of limiting factors** and species response curves.



If a model can explain the relationship between distribution records and environmental variables, we can **make predictions to unknown samples (unsurveilled regions)**.

e.g., Temp. = 23.5°C or 24.5°C, response is 0 (i.e., absence).

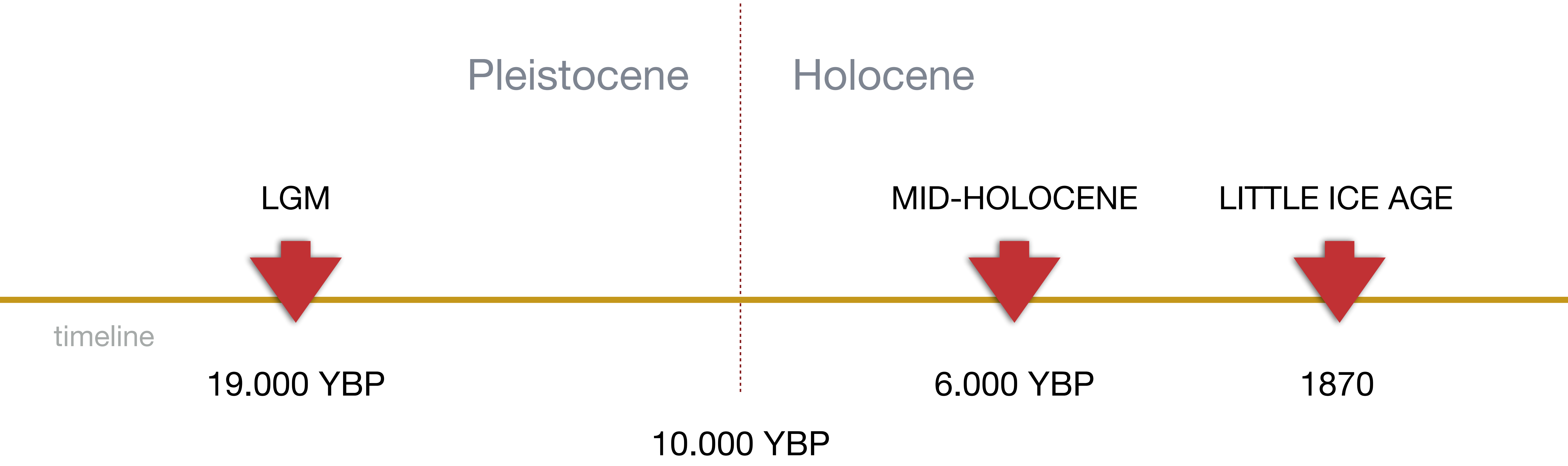


Predictive model-based interpolation

Made to **new sites within the range of values of environmental conditions sampled in the training data** and within the **same time window** of sampling.

From **scattered records in space to continuous distribution surfaces**.

Typical **applications** include **mapping species' present distributions, important tools for ecology and conservation planning**.

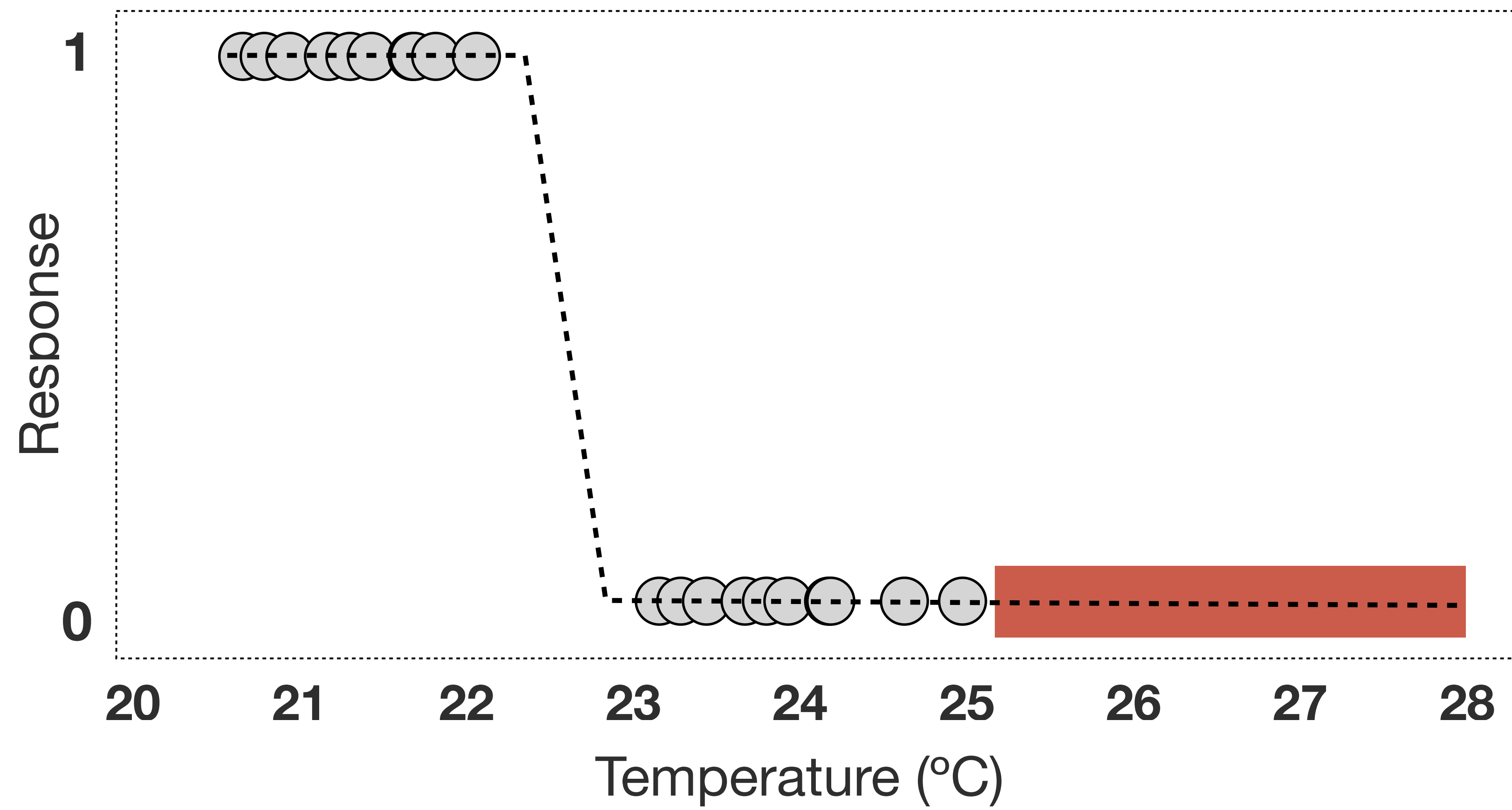


Predictive model-based transferability

Made to **unsurveilled geographic or temporal domains.**

No information on the similarity between the environment of training data and the predictions (present-day vs future conditions).

Model **transferability may lead to extrapolation.**

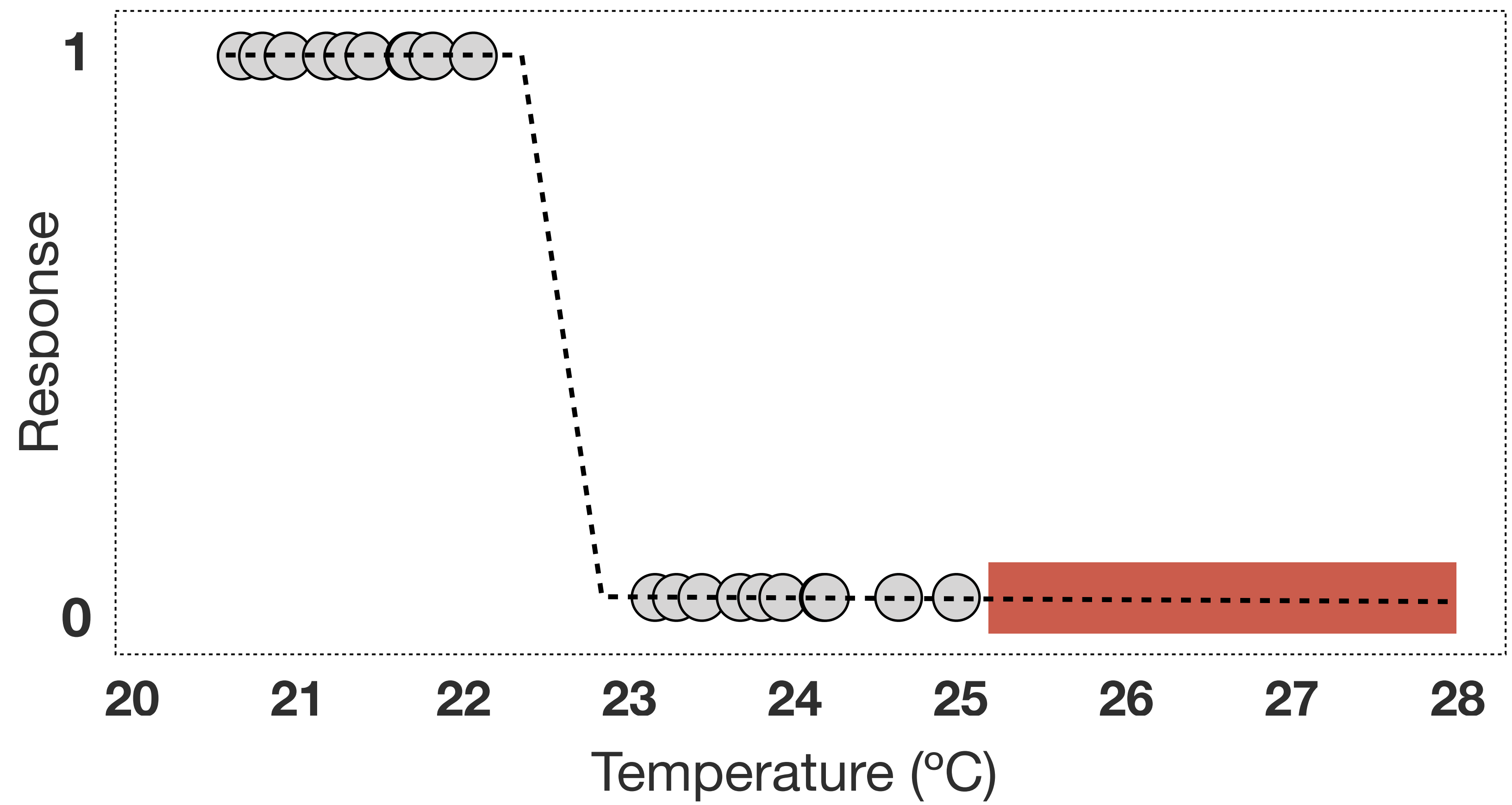


Extrapolation refers to **predictions for environmental values** that are **outside the range of environmental values used to fit the model**.

e.g.,

A **model used records with temperatures of 20-25ºC**.

If **predictions** are **made for temperatures > 25ºC**, then the model will extrapolate. No information exists on the probability of occurrence at > 25ºC, so the **predictions will be uncertain**.



Avoid extrapolation in favour of interpolation

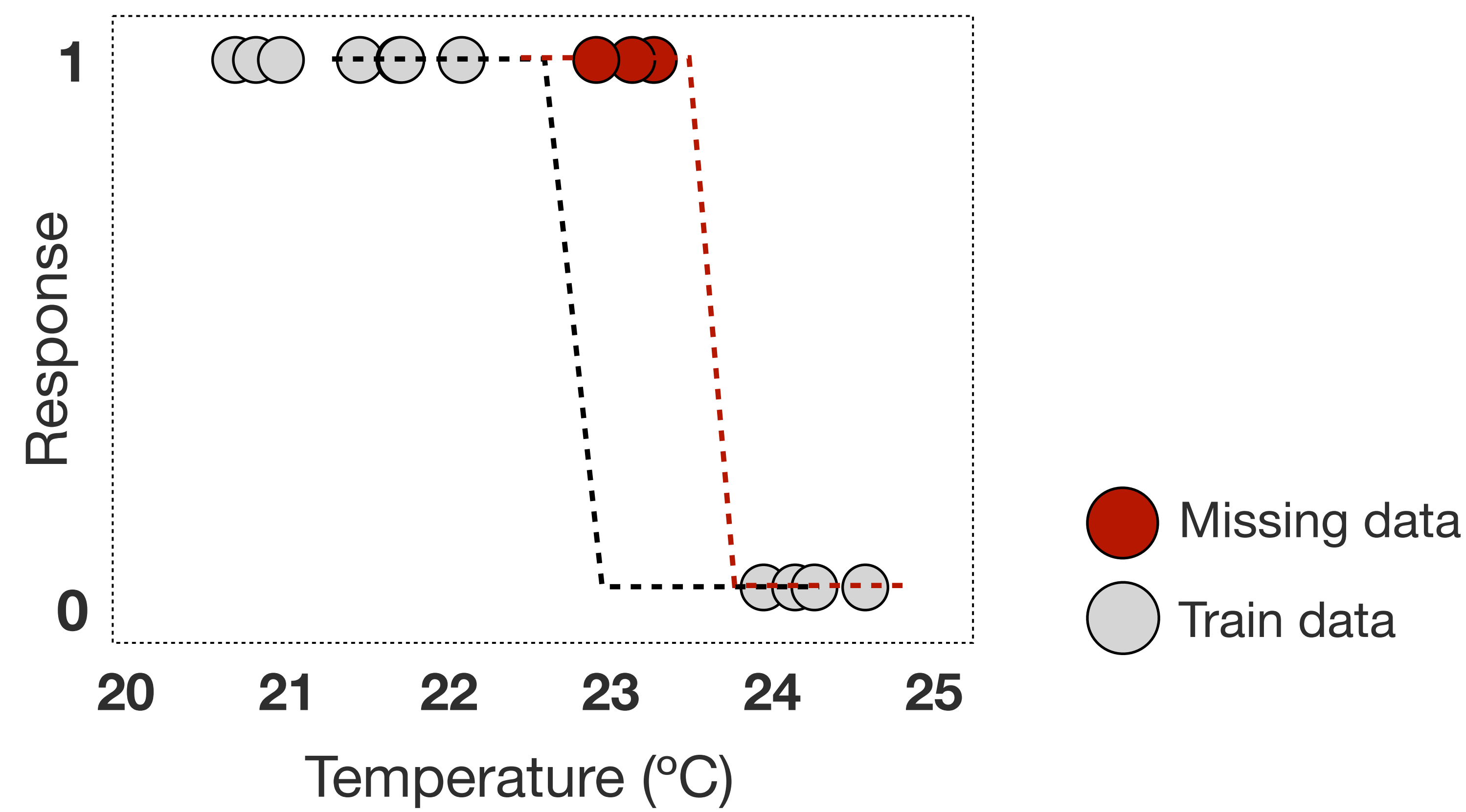
But when extrapolation exists (e.g., future climate changes), model should be treated with a great deal of caution.

1. Avoid predicting with complex functions;
2. Use a parsimonious models (reduced number of predictors).
3. Interpret models with sound ecological knowledge;

Corrective models can identify the **niche of a species only if the records used to fit the models cover the distribution of the species**.

When mapped, it represents the **potential distribution or the habitat suitability**.

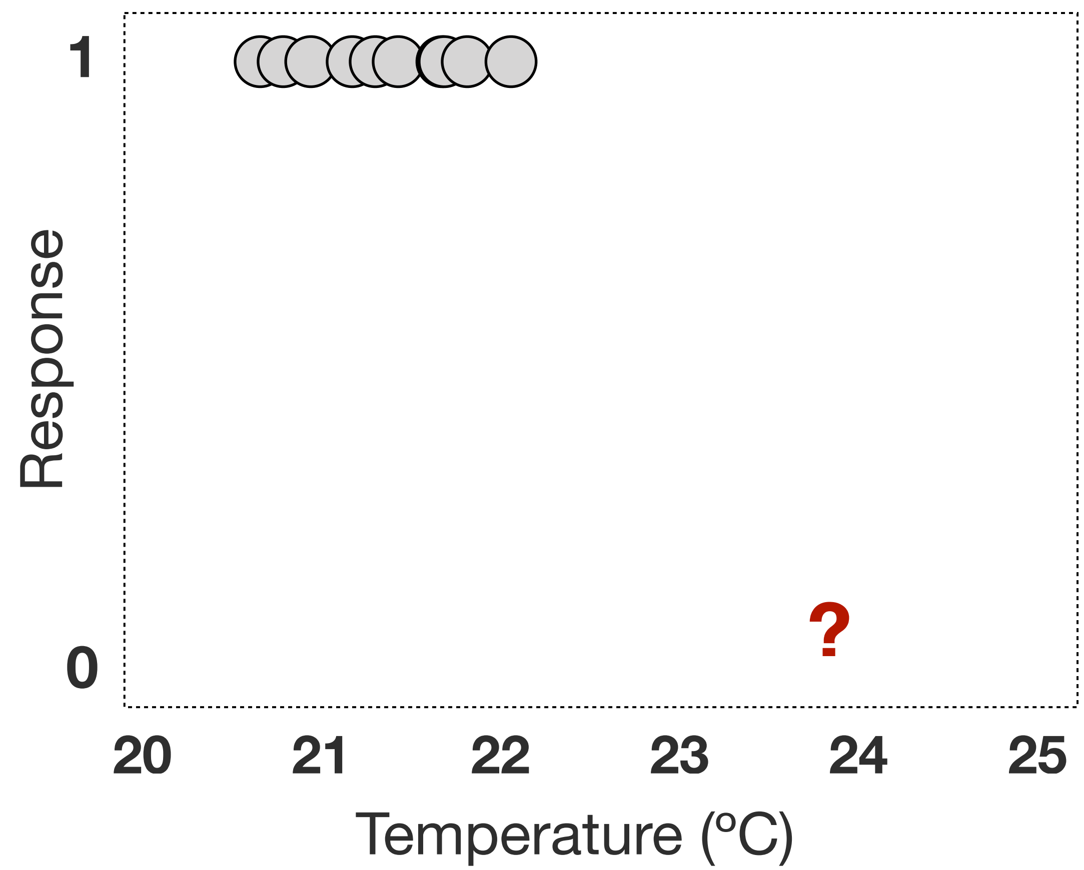
If records are insufficient, the models do not identify the fundamental niche; **the model fits only to the portion of the niche that is represented by the observed records** (truncated niche).



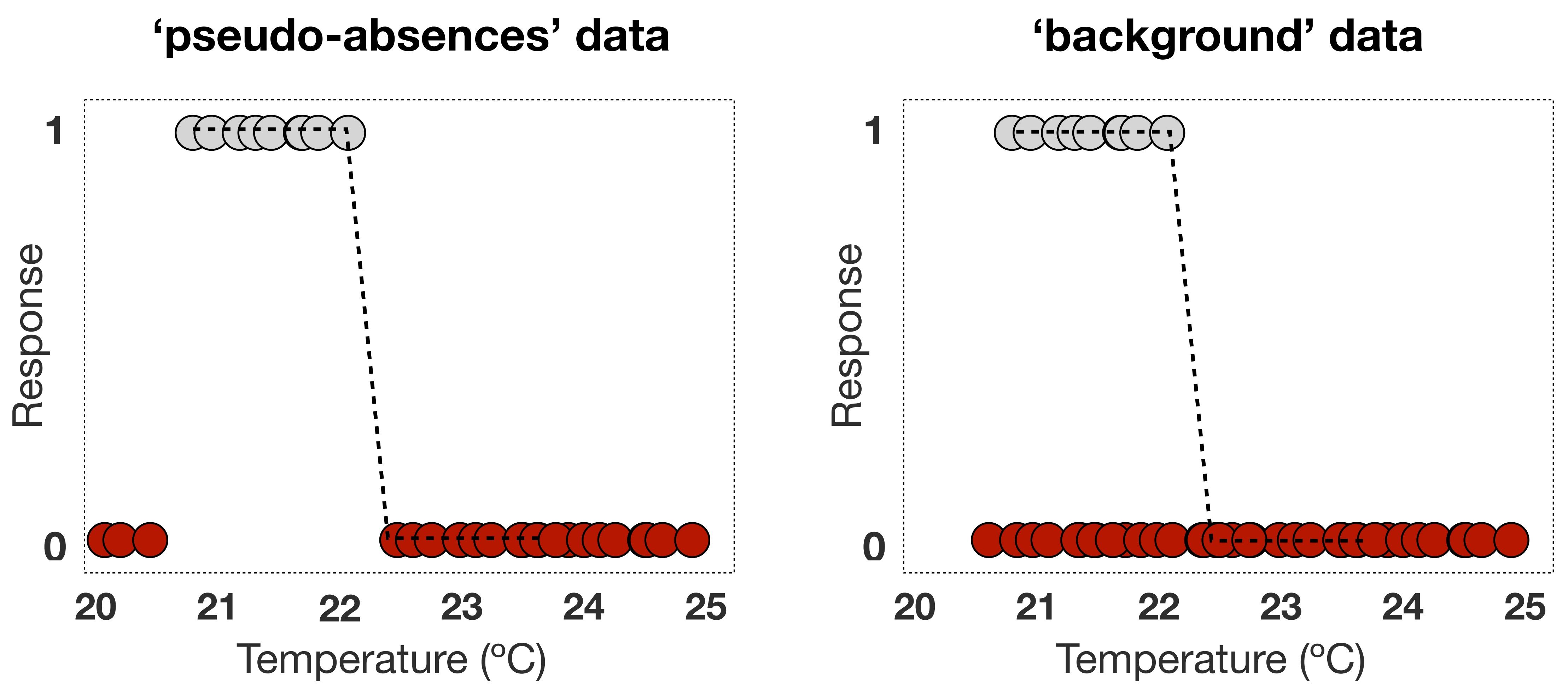
Corrective models can identify the **niche of a species only if the records used to fit the models cover the distribution of the species**.

When mapped, it represents the **potential distribution or the habitat suitability**.

If records are insufficient, the models do not identify the fundamental niche; **the model fits only to the portion of the niche that is represented by the observed records** (truncated niche).



Correlative models infer the **relationship between the occurrence of biodiversity and the environment**, but **absence records are often unavailable** or unknown, leading to **presence-only datasets** and to the need of developing **presence-only models**.



Models based on 'pseudo-absences', generated from the **study area where occurrences do not exist.** Any regression or machine learning algorithm can be implemented (e.g., GLM and BRT).

Models based on 'background', generated from the entire study area. Focus on **how the environment where the species occurs relates to the environment across** the rest of **the study area** (e.g., MaxEnt).

How many absence records?

Linear models, additive models and maximum entropy models

A large number of pseudo-absences / background (e.g., 10,000);

Boosted Regression Trees models

The same number of pseudo-absences as presences records (but never less than 1,000).

Barbet-Massin et al., 2012

Which environmental predictors for modelling?

Use **large datasets**, an approach with stronger criticism.

Use **pre-selected datasets**, linked to known physiological rules**.

** the choice of predictors should be **guided by the objectives and the hypotheses raised** about the species-environment relationship.

Bouchet et al., 2019

Modelling with abiotic conditions

Ecophysiological knowledge should guide the modeller in process of choosing which factors may determine the distribution of species.

My main questions

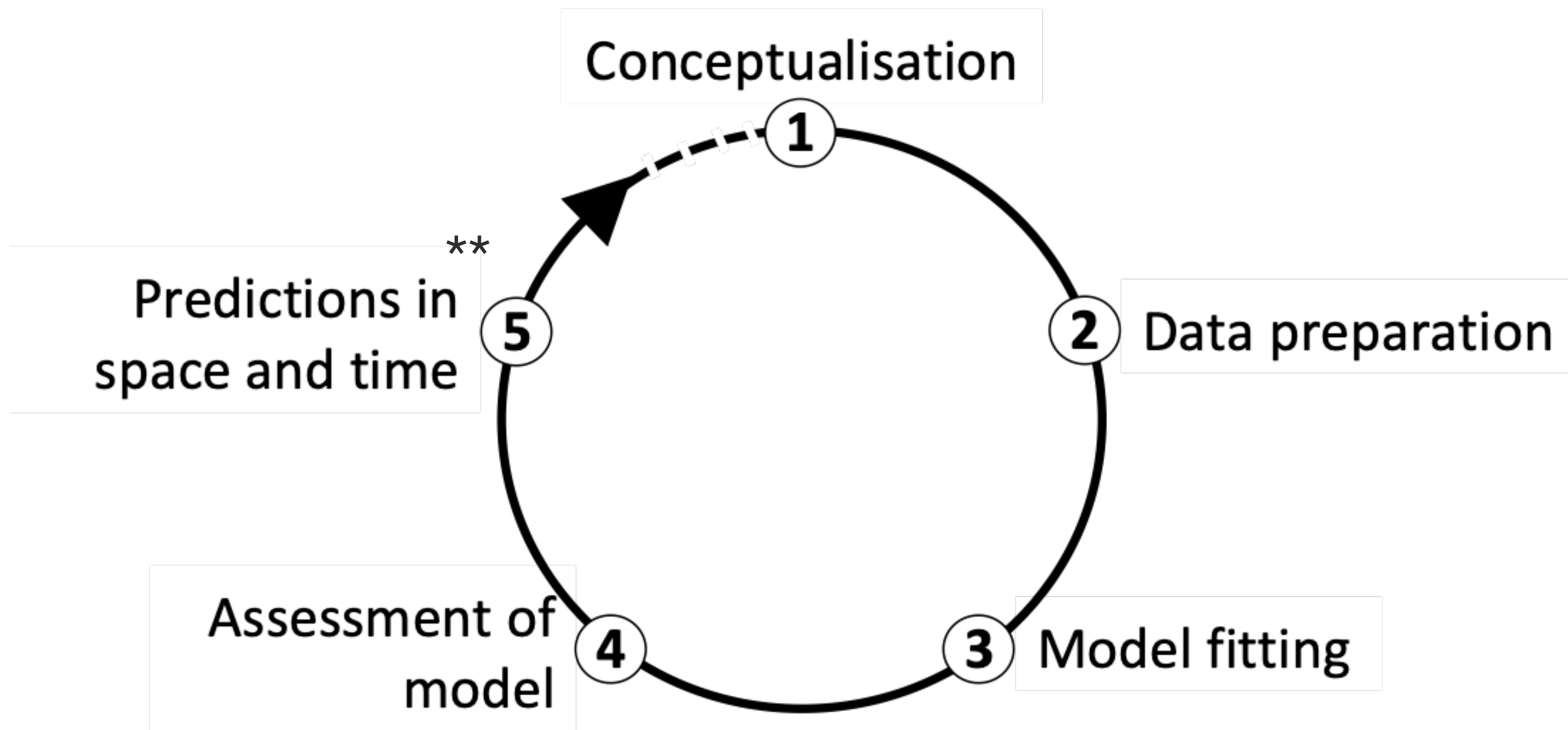
Which environmental variables drive the distribution of my model species?

Which response I expect for each environmental variable?

My main hypotheses must be based on ecophysiological knowledge

(…)

Steps for model building



Model building is an iterative process and there is much to learn on the way (a loop rather than a straightforward approach).

You may want to revisit and improve certain steps (e.g., improve biodiversity data collection or remove surplus environmental layers).

** not always part of ENM studies; depends on the model objective.

Model fitting

Key steps in good modelling practice include the following:

1. Gathering relevant biodiversity and environmental data, and assessing its adequacy (relevance and completeness, avoiding truncated niches);

2. Selecting an appropriate modelling algorithm;

3. Fitting the model to the training data and evaluating performance, including the realism of fitted response functions, the model's fit to data, and predictive performance on test data;

4. Mapping predictions to geographic space;

5. Apply thresholds if continuous predictions need reduction to a binary map;

6. Iterating the process to improve the model in light of knowledge gained throughout the process.

(Elith & Leathwick 2009)

Assumptions of ENM associated with data

Observed distributions are indicative of environmental tolerances and resource requirements since species are in equilibrium with the environmental conditions - species occur in all suitable areas and are absent from all unsuitable areas (niche theory).

Niche space assumption : The study contains the full range of conditions that the species can inhabit (for the examined abiotic variables).

Dispersal / demographic assumption : Factors related to dispersal, establishment, and persistence do not drive the species not to occupy an environmentally suitable area.

Biotic assumption : Biotic interactions do not drive the species not to occupy an environmentally suitable area.

Desired properties of ENM fitting

Deductive: develop hypotheses about the causes of the pattern explained and predicted by the models.

Distinct between patterns and process: distinguish between the patterns observed and the mechanisms involved.

Simplicity: highlight a few mechanisms without becoming entangled in complex interactions and correlations between many variables.

Parsimonious: preferring the simpler of two equally adequate models; favor simple explanations over complex.

Generality: aimed to achieve general broad ecological conclusions.

Main assumption of ENM

The observed distributions are an indication of the environmental tolerances and resource requirements, since species are in equilibrium with the environmental conditions.

i.e.,

Species occur in all suitable areas and are absent from all unsuitable areas (niche theory).