

Interpretable ML for biodiversity

An introduction using species distribution models

Timothée Poisot

Université de Montréal

September 28, 2024



MAIN GOALS

1. How do we produce a model?
2. How do we convey that it works?
3. How do we talk about how it makes predictions?
4. How do we use it to guide actions?



THE STEPS

1. Get data about species occurrences
2. Build a classifier and make it as good as we can
3. Measure its performance
4. Explain some predictions
5. Generate counterfactual explanations
6. Briefly discuss ensemble models



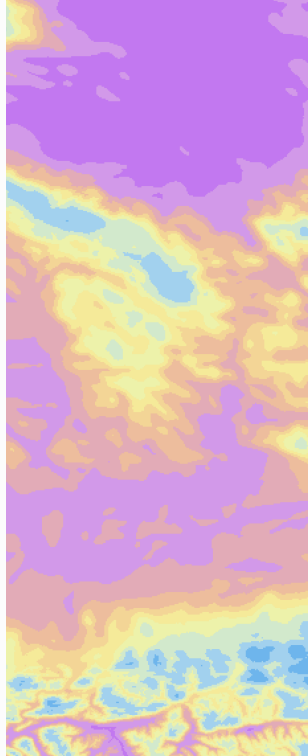
BUT WHY...

... think of SDM as a ML problem? Because they are (and would be better if we accepted this)

... the focus on explainability? We cannot ask people to *trust* - we must *convince*

§ 1

Introduction





GETTING DATA



GETTING A POLYGON



CHELSA2 DATA



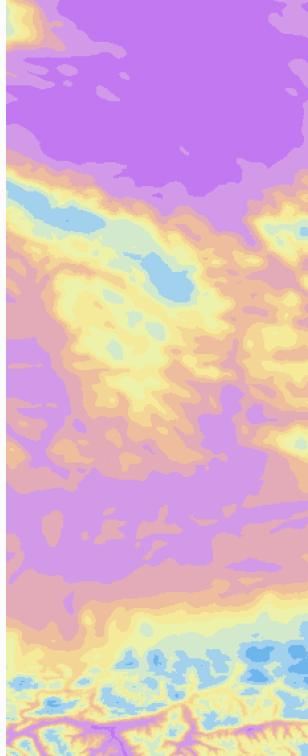
TRIMMING POLYGON



DOWNLOAD DATA FROM GBIF

§ 2

Validation





PSEUDO-ABSENCES

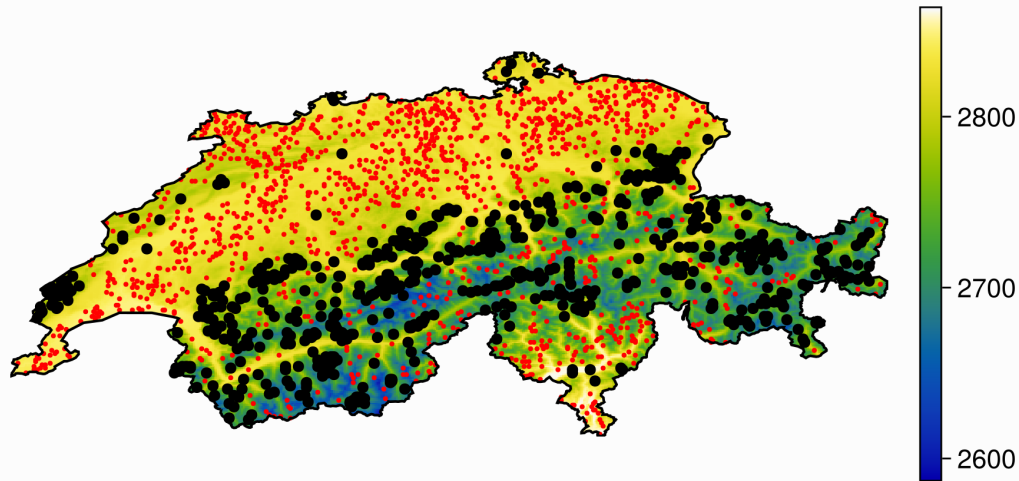
SDM Layer with 45336 Bool cells

Proj string: +proj=longlat +datum=WGS84 +no_defs

Grid size: (239, 543)

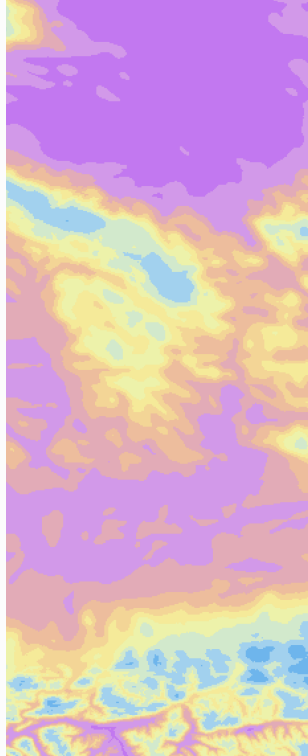


VISU



§ 3

Model setup





SETUP

```
SDeMo.MultivariateTransform{MultivariateStats.PCA} → SDeMo.NaiveBayes → P(x  
) ≥ 0.5
```



CROSS-VALIDATION

0.5620979518435002



RE-TRAINING

```
SDeMo.MultivariateTransform{MultivariateStats.PCA} → SDeMo.NaiveBayes → P(x  
) ≥ 0.281
```



INITIAL PRED

SDM Layer with 69967 Float64 cells

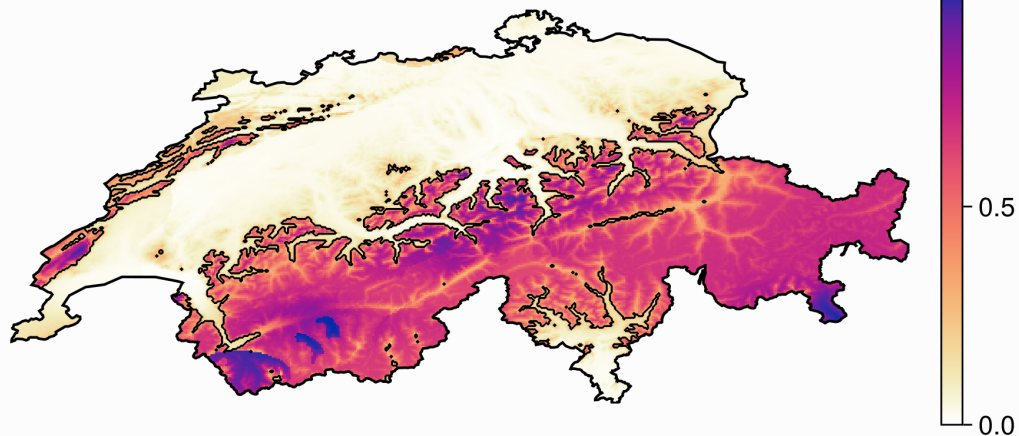
Proj string: +proj=longlat +datum=WGS84 +no_defs

Grid size: (239, 543)



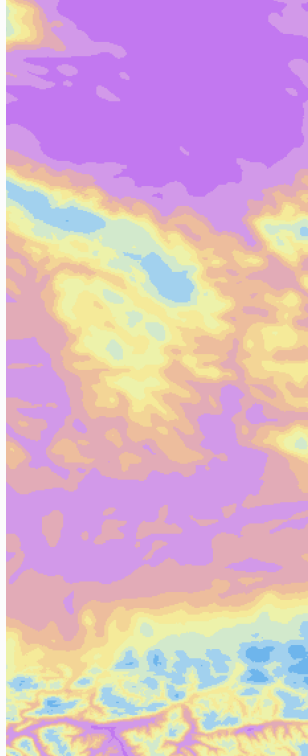
VISU

Prediction



§ 4

Why?



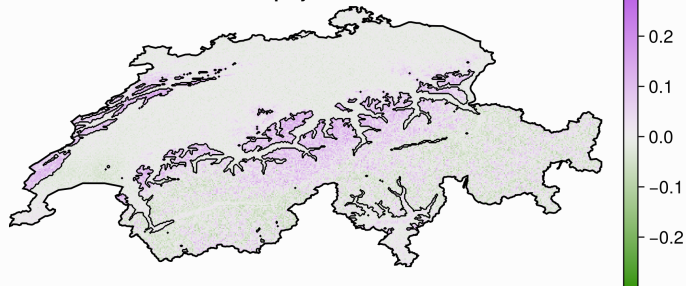


CODE

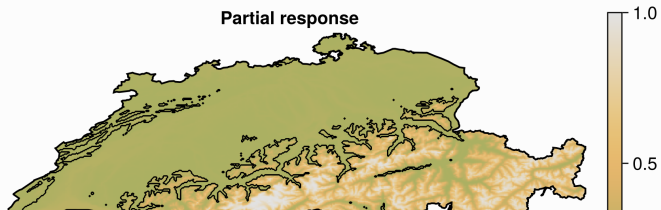


VISU

Shapley values



Partial response





MOSAIC

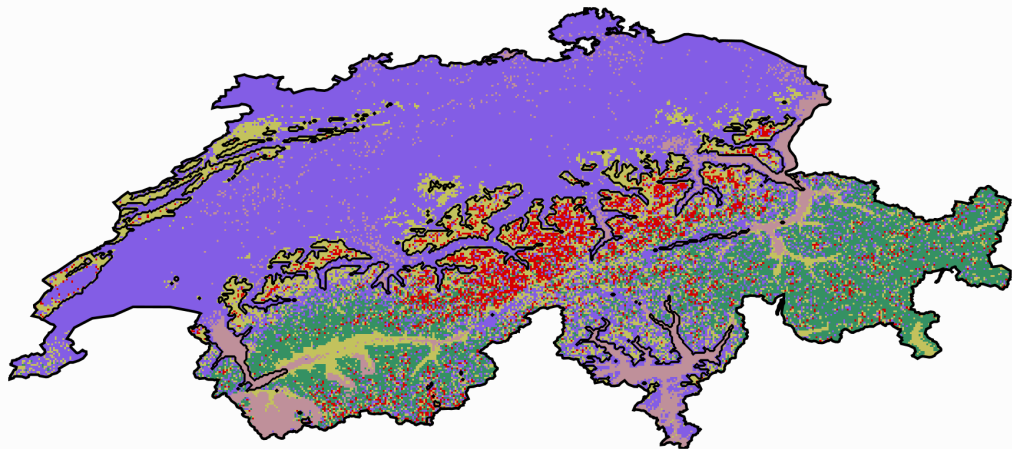
SDM Layer with 69967 Int64 cells

Proj string: +proj=longlat +datum=WGS84 +no_defs

Grid size: (239, 543)

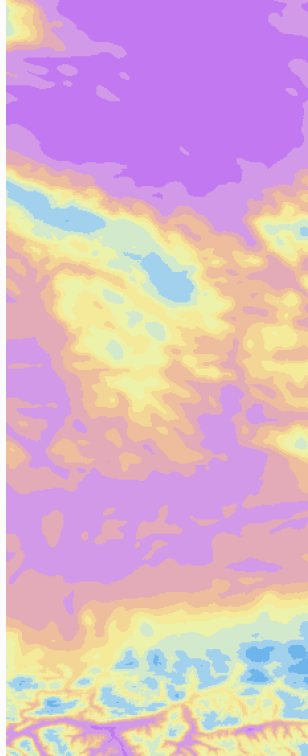


VISU



§ 5

About ensemble models





UNCERTAINTY

0.6537641876644507



ADD PRED

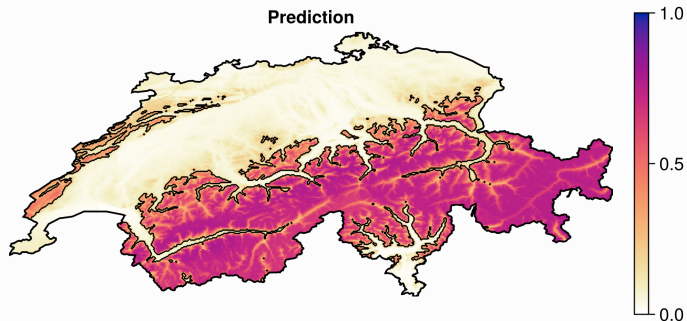
SDM Layer with 69967 Float64 cells

Proj string: +proj=longlat +datum=WGS84 +no_defs

Grid size: (239, 543)



Prediction



Uncertainty

