University of Zurich[UZH]

Department of Mathematical Modeling and Machine Learning (DM^{3}L)

EcoVision

ML4PS
Biodiversity: Species Distribution Modeling

Johannes Dollinger

Nov 2025

15 LIFE ON LAND

THE GLOBAL GOALS

How climate cha... worser...
...twaves, dro...
f...

29 AUG 2023 |

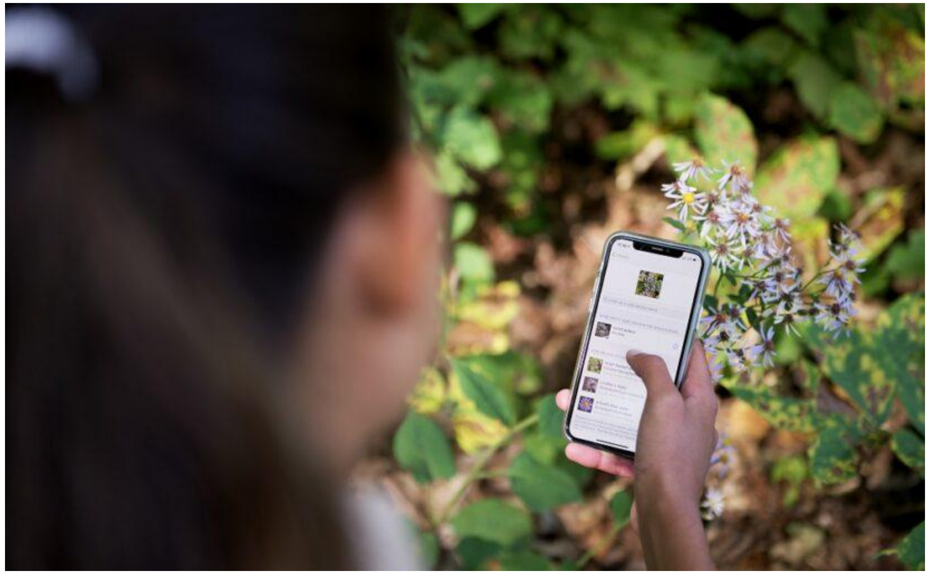
Kunming-Montreal Global

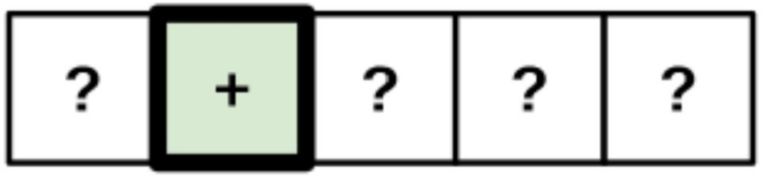
Accelerating extinction rate triggers domino effect of biodiversity loss

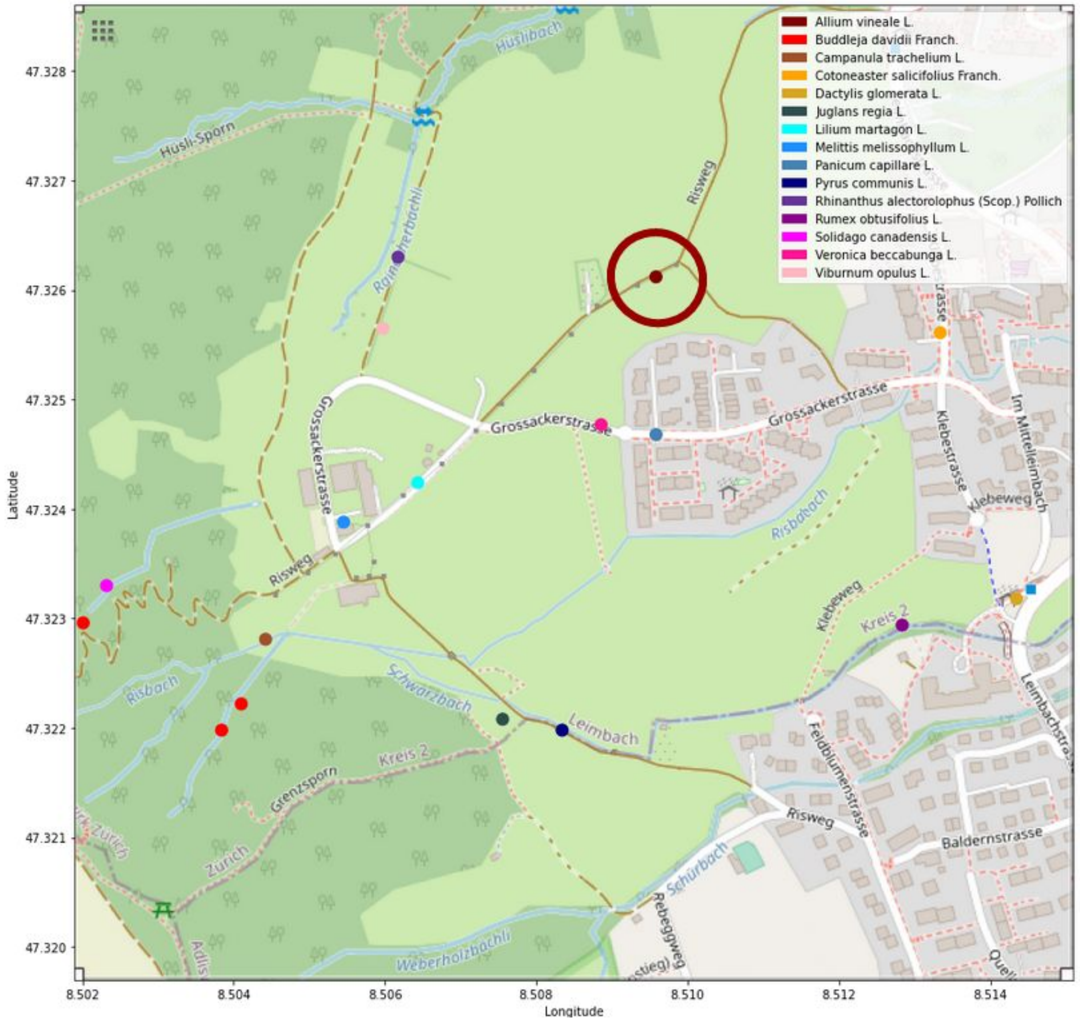
21 May 2024

Sources:
https://www.bbc.com/news/science-environment-58073295
https://news.un.org/en/story/2024/05/1150056
https://www.unep.org/news-and-stories/speech/kunming-montreal-global-biodiversity-framework-how-halt-and-reverse

Noisy, biased, crowd-sourced presence-only data (PO)





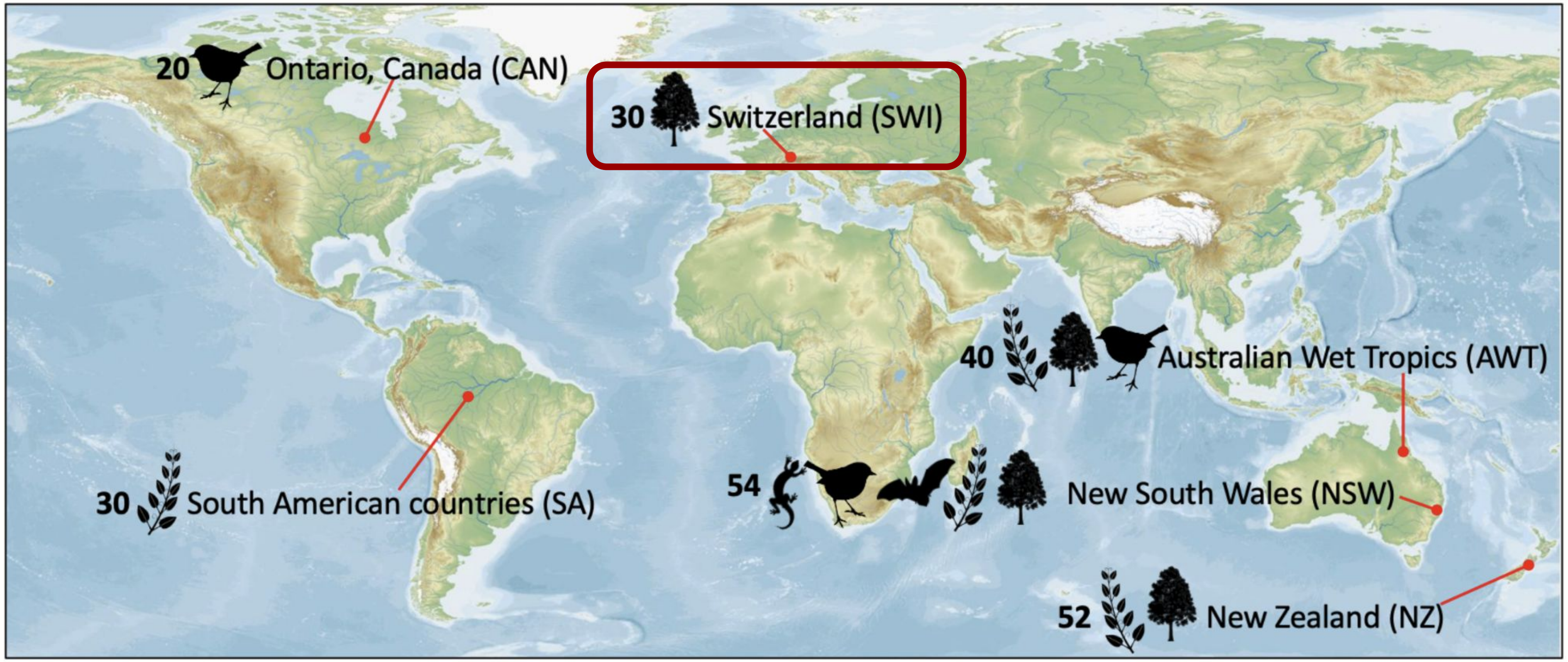


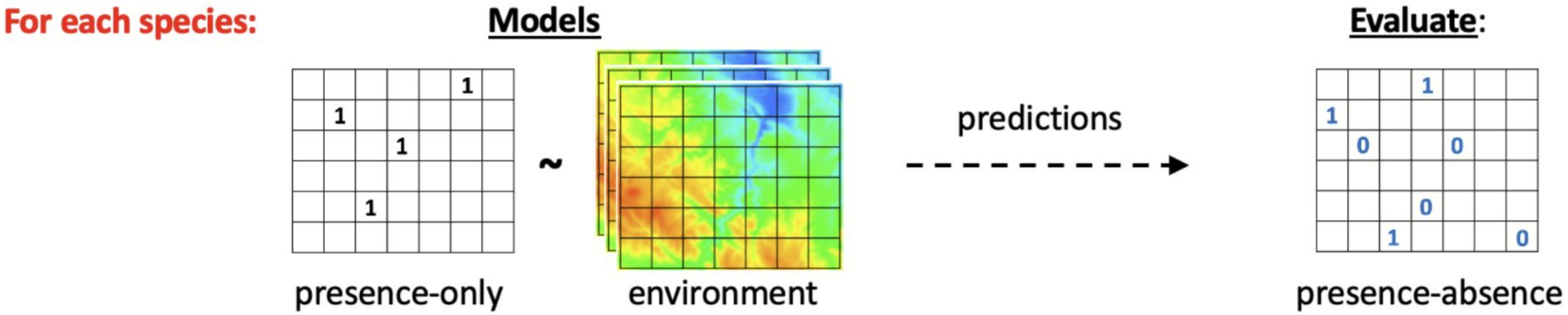

Predicted Presences

Testing on presence-absence surveys (PA)

-	+	-	-	+

Source:
Example Foto - My sister and her group


20 Ontario, Canada (CAN)
30 Switzerland (SWI)
40 Australian Wet Tropics (AWT)
54 New South Wales (NSW)
30 South American countries (SA)
52 New Zealand (NZ)


For each species:
Models
Evaluate:
presence-only
~
environment
predictions
presence-absence

```python
from tester import Tester
import numpy as np
sf = Tester("./data/test_pa/SWItest_pa.csv")
sf.test(np.ones((10013, 30)))
print("-----")
sf.test(np.zeros((10013, 30)))
```

Last executed at 2025-05-14 12:06:50 in 6.53s

```
Accuracy: 0.0809
F1-score: 0.1497
Mean predicted per location: 30.0
-----
Accuracy: 0.9191
F1-score: 0.0000
Mean predicted per location: 0.0
```

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2\frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

Data

Load the data (*train_po/SWItrain_po.csv*, *test_env/SWItest_env.csv* and *test_pa/SWItest_pa.csv*) and familiarize yourself with them, including analyzing their value & class distribution. See *train_po/01_metadata_SWItrain_po.csv* for an explanation of the predictors. You will need to pre-process the data for your modeling, but do not need to apply filtering.

Classic Modeling

Train two classic (non-neural) models on the PO-data. You may take inspiration from previous labs or the list of methods used in the original publication[1]. Produce predictions for the PA-data and receive the accuracy and F1-score from the Tester-class provided in tester.py. Make sure to try out different hyperparameters in both this and the next section to improve your models performance.

Deep Learning Model

Now design and train a small neural network on the PO-data, scoring it again on the PA-data.

```
Accuracy: 0.9306
F1-score: 0.3925
Mean predicted per location: 1.0
```

Spatial Implicit Neural Representations

Test out training a neural network using the SINR AN-full loss[2]. You may use the *train_bg* folders files as the samples at random locations.

Discussion

Compare the results from the different models and propose improvements.

Bonus: Results Table

Run the experiments for all six regions in the provided data and put the results into a proper (Latex) table as commonly seen in publications.

The data used in this lab

[1] Elith*, Jane, et al. "Novel methods improve prediction of species' distributions from occurrence data." *Ecography* 29.2 (2006): 129-151.

Elith, Jane, et al. "Presence-only and presence-absence data for comparing species distribution modeling methods." *Biodiversity informatics* 15.2 (2020): 69-80.

Valavi, Roozbeh, et al. "Predictive performance of presence-only species distribution models: a benchmark study with reproducible code." *Ecological monographs* 92.1 (2022): e01486.

Deep SDMs

[2] Cole, Elijah, et al. "Spatial implicit neural representations for global-scale species mapping." *International conference on machine learning*. PMLR, 2023.

Botella, Christophe, et al. "The GeoLifeCLEF 2023 dataset to evaluate plant species distribution models at high spatial resolution across Europe." *arXiv preprint arXiv:2308.05121* (2023).

Dollinger, Johannes, et al. "Sat-sinr: High-resolution species distribution models through satellite imagery." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 10 (2024): 41-48.

Zbinden, Robin, et al. "On the selection and effectiveness of pseudo-absences for species distribution modeling with deep learning." *Ecological Informatics* 81 (2024): 102623.

Dollinger, Johannes, et al. "Climplicit: Climatic Implicit Embeddings for Global Ecological Tasks." *arXiv preprint arXiv:2504.05089* (2025).