Exploring the potential of **neural networks** for **Species Distribution Modeling**

Robin Zbinden* robin.zbinden@epfl.ch
Nina van Tiel* nina.vantiel@epfl.ch
Benjamin Kellenberger (Yale)
Lloyd Hughes (EPFL)
Devis Tuia (EPFL)

* Equal contribution

eceo EPFL
Yale University

Introduction to Species Distribution Models (SDMs)

- Relate species occurrence data with environmental variables
- Used to understand and predict the geographic distribution of a species
- Used to support decision-making for conservation and restoration

Species occurrence data

Presence-absence data: systematic field surveys

Presence-only data: incidental observations, combined with generated negative samples

Environmental variables

Climate, soil and topography

Vegetation indices, land use, land cover

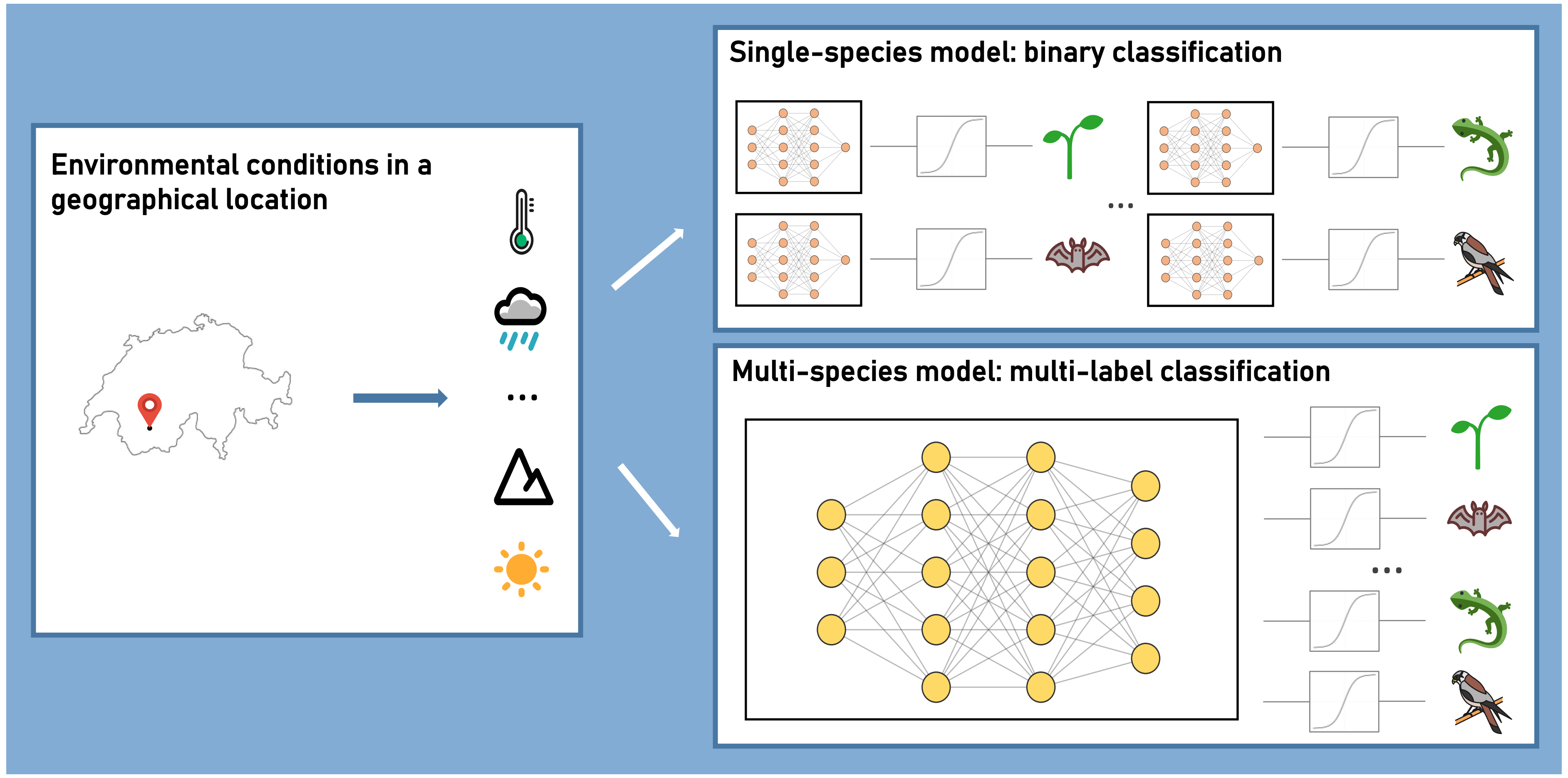
Remote sensing imagery

Models

Traditional statistical methods

ML methods

Neural networks, deep learning



Our work

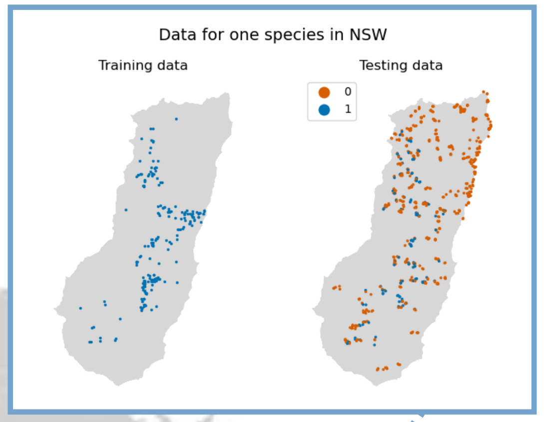
How do neural networks perform compared to well-established methods for SDMs?

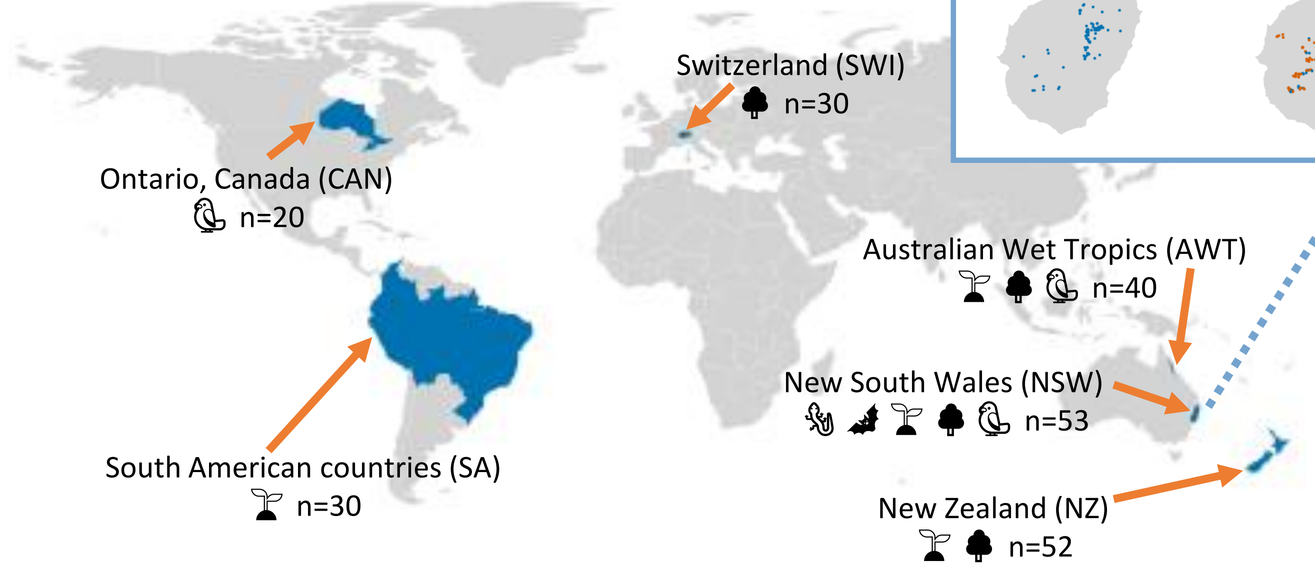
Dataset: benchmark dataset [1]: tabular data for 225 species from 6 regions. Train on presence-only, test on presence-absence.

Model: multi-layer perceptrons (MLPs) with:

- binary classification for single-species models
- multi-label classification for multi-species models

Loss function: weighted binary cross-entropy

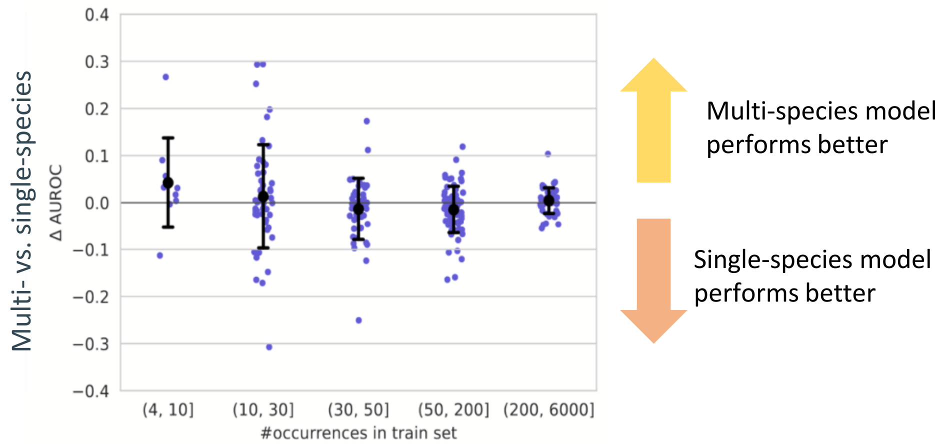




Results

Mean **AUROC** across species for each region

		AWT	CAN	NSW	NZ	SA	SWI
[2]	MaxEnt	**0.686**	0.584	0.713	0.738	0.804	0.809
	XGBoost	0.653	0.568	0.706	0.720	0.788	0.815
	Random Forest	0.675	0.572	0.718	0.746	**0.813**	**0.818**
	Ensemble	0.683	0.580	**0.723**	**0.749**	0.806	0.812
Ours	Single-species MLP	0.666	0.589	0.688	0.715	0.799	0.808
	Multi-species MLP	0.617	**0.605**	0.708	0.714	0.803	0.815

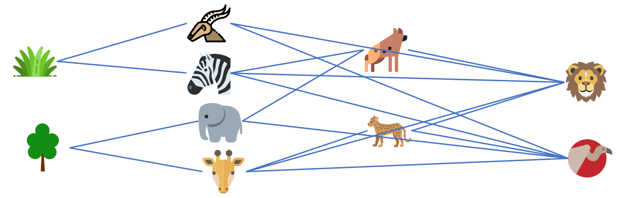


Future directions

More complex deep learning based SDMs involving and combining:

- **Multi-modal data** providing, for example, geospatial and temporal context

- Biological and ecological information through **knowledge-guided** machine learning



- Transfer learning, model **pre-training**

References

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

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