From presence-only to abundance species distribution models using transfer learning

B. Bourel, A. Joly, M. Servajean, S. Bettinger, J. A. Sanabria-Fernández, D. Mouillot

National Institute for Research in Digital Science and Technology

marine biodiversity exploitation & conservation

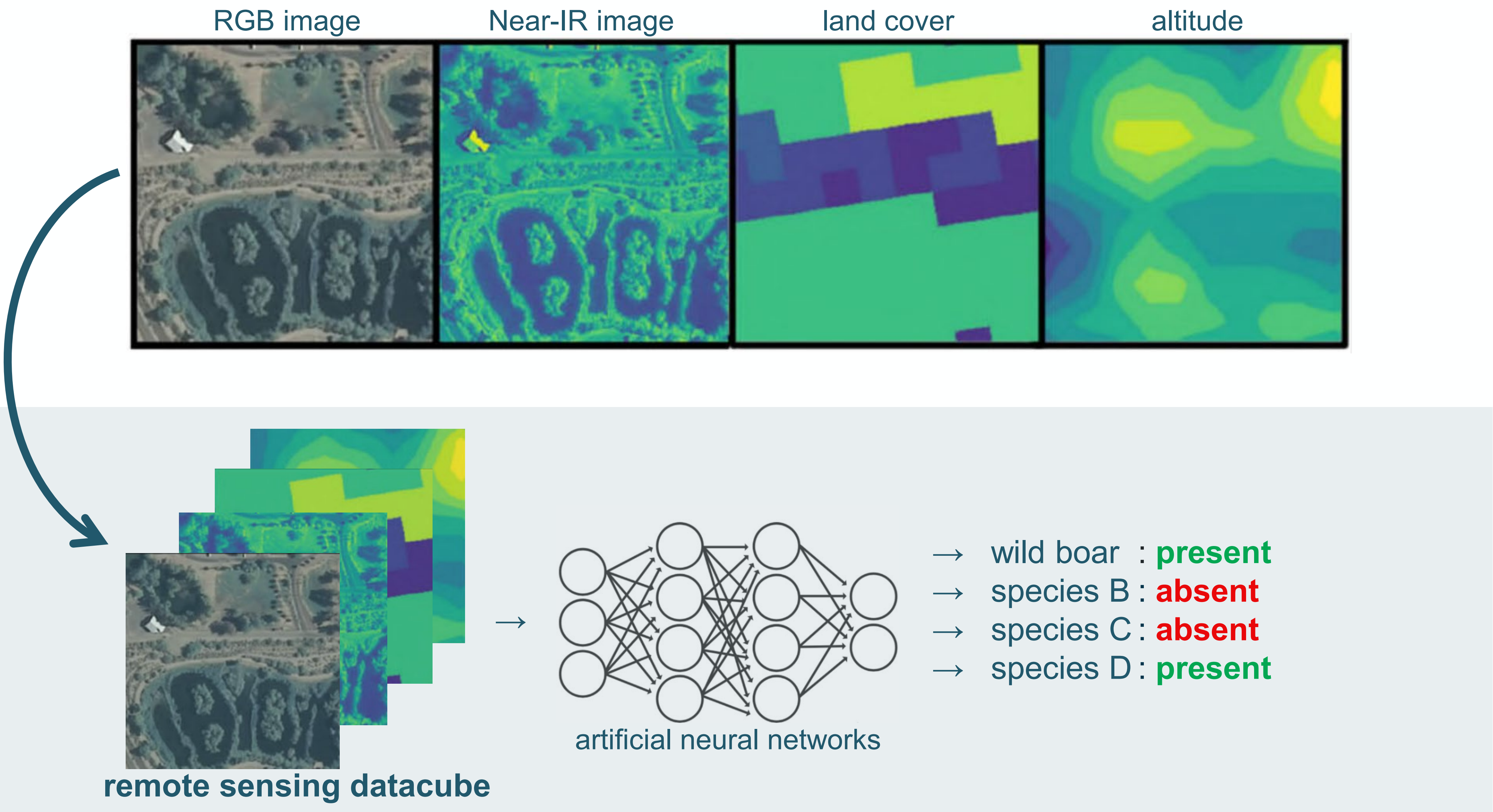
Laboratory of Computer Science, Robotics and Microelectronics of Montpellier

Le calcul intensif au service de la connaissance

agence nationale de la recherche

AGENCE FRANÇAISE DE DÉVELOPPEMENT

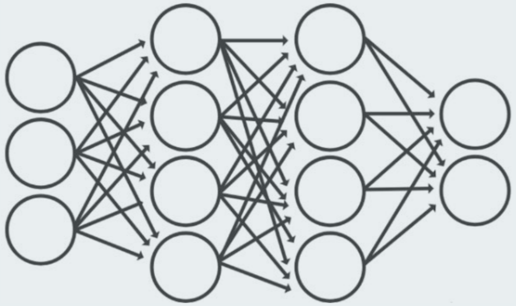
Why focus on species distribution models using artificial neural networks?



Current limitations: abundance data sets too small for neural networks

Large
presence-only
dataset

→



artificial neural networks

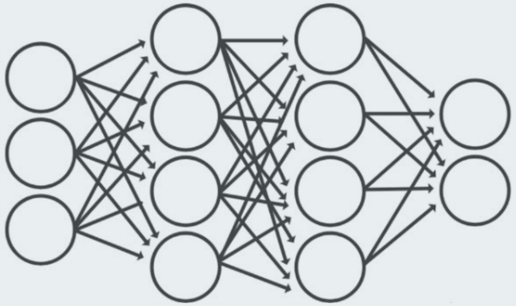
→

Good performance
for occurence
predictions

Better than
classic models

Small
abundance
dataset

→



artificial neural networks

→

Bad performance
for
occurence
predictions

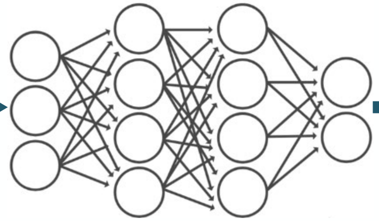
Not as good as
classic models

Overcoming limitations with transfer learning technology

Large abundance dataset

	Occurrence
site A	specie 2
site B	specie 3
site C	specie 2
site D	specie 3
site E	specie 2
site F	specie 4
site G	specie 1
site H	specie 1
[...]	[...]

Training the model
to predict occurrences



artificial neural networks

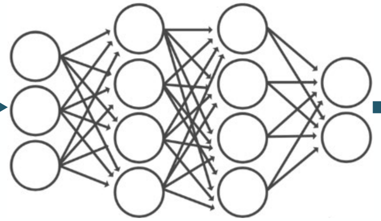
specie 1 → 0
specie 2 → 1
specie 3 → 0
.
.
specie x → 1

Overcoming limitations with transfer learning technology

Large abundance dataset

	Occurrence
site A	specie 2
site B	specie 3
site C	specie 2
site D	specie 3
site E	specie 2
site F	specie 4
site G	specie 1
site H	specie 1
[...]	[...]

Training the model
to predict occurrences

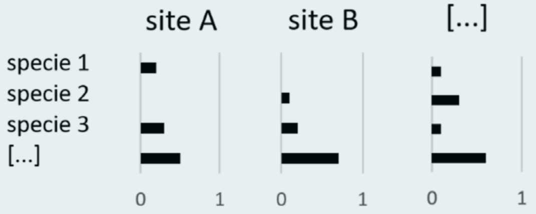


artificial neural networks

specie 1 → 0
specie 2 → 1
specie 3 → 0
.
.
specie x → 1

Transfer learning of
learned features on
presence-only dataset

Small abundance dataset



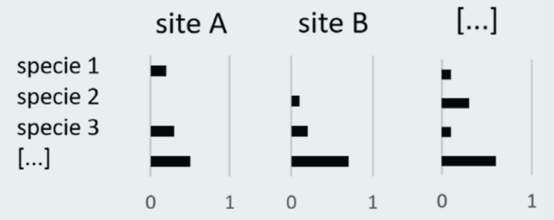
artificial neural networks

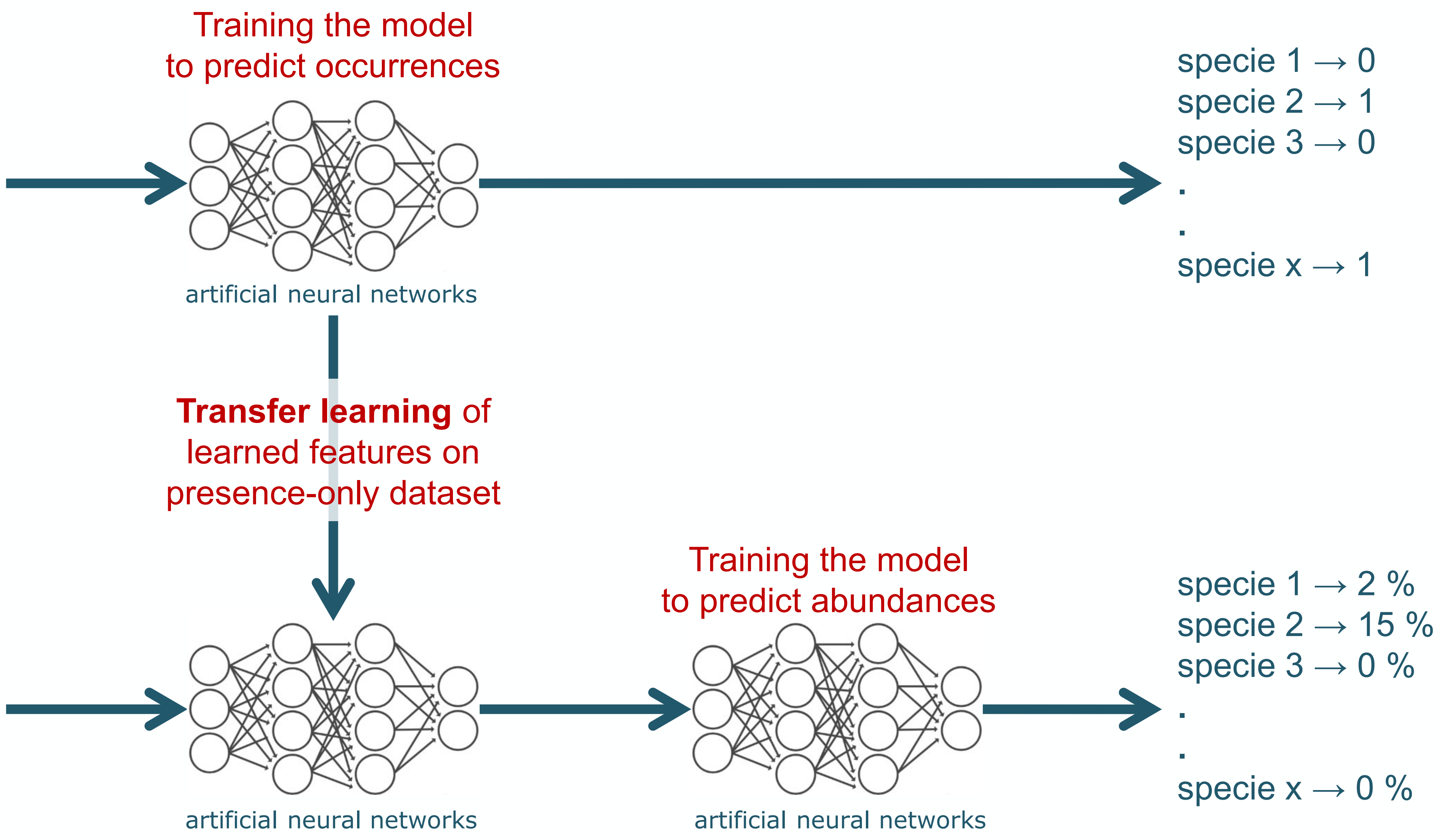
Overcoming limitations with transfer learning technology

Large abundance dataset

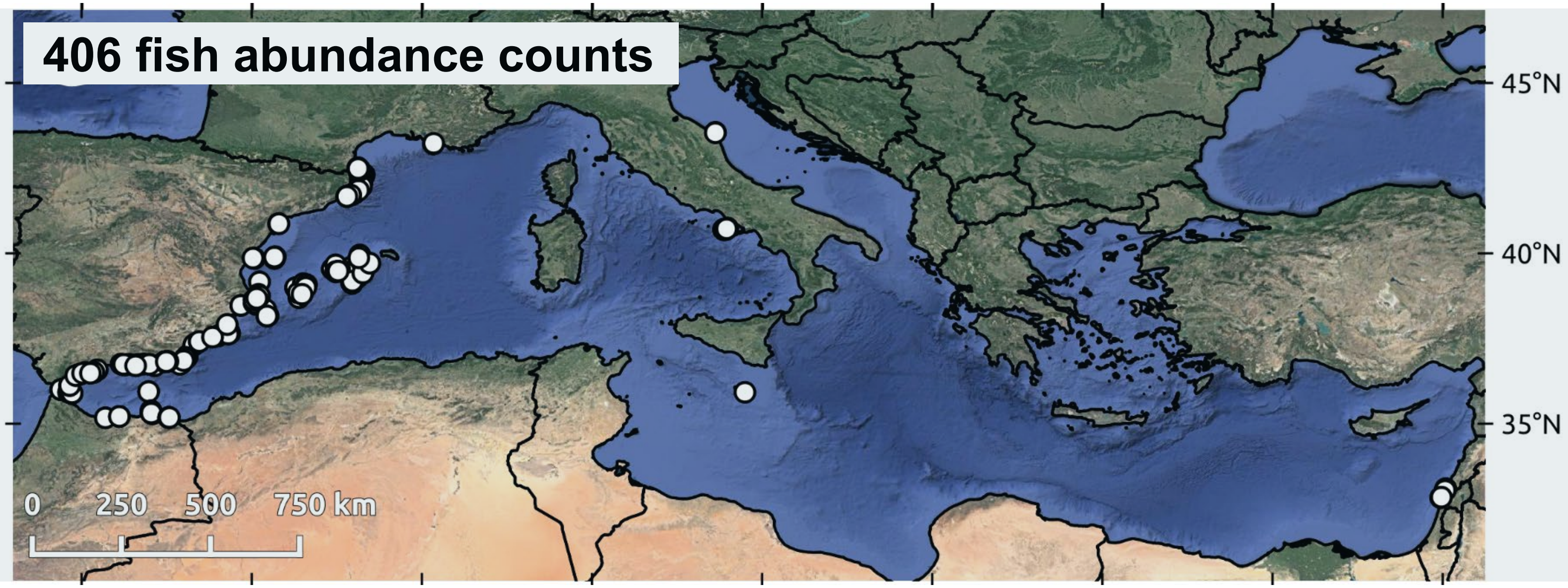
	Occurrence
site A	specie 2
site B	specie 3
site C	specie 2
site D	specie 3
site E	specie 2
site F	specie 4
site G	specie 1
site H	specie 1
[...]	[...]

Small abundance dataset



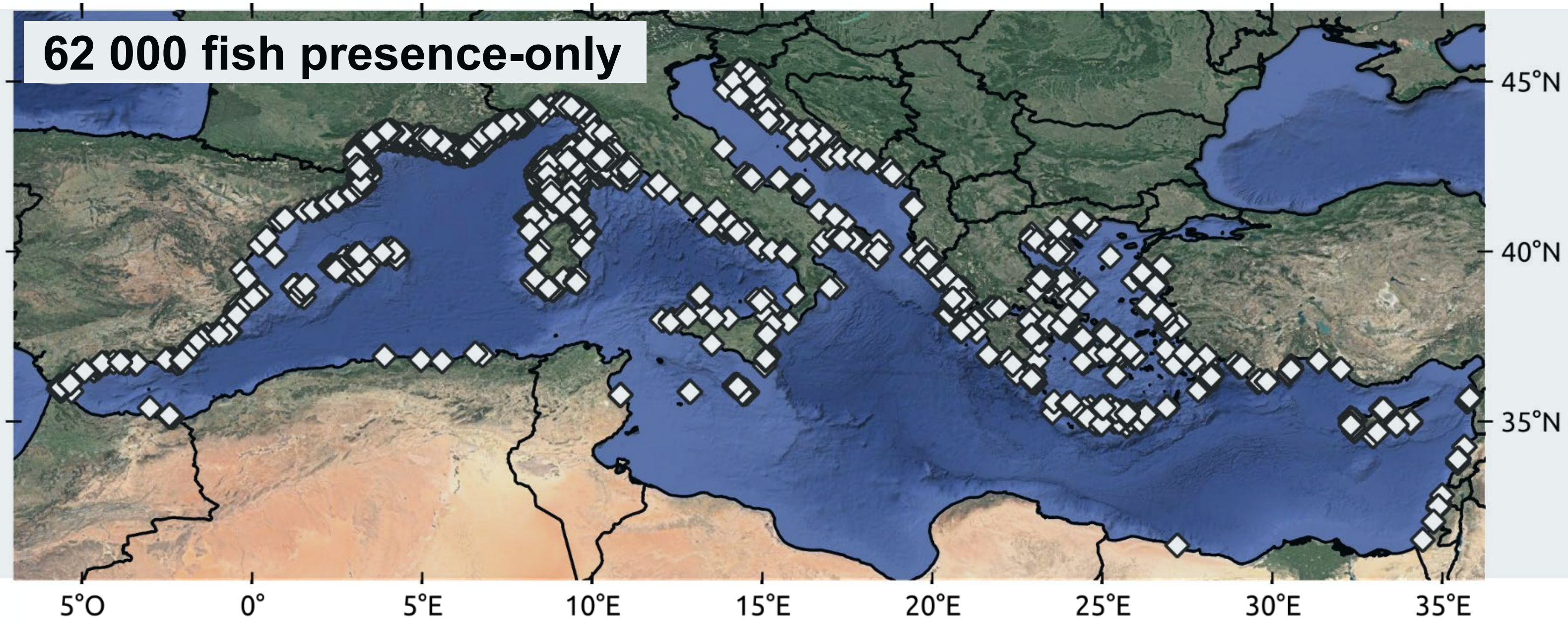


Fish occurrence and fish abundance datasets



Reef Life Survey

Abundance of 47 fish species for each count

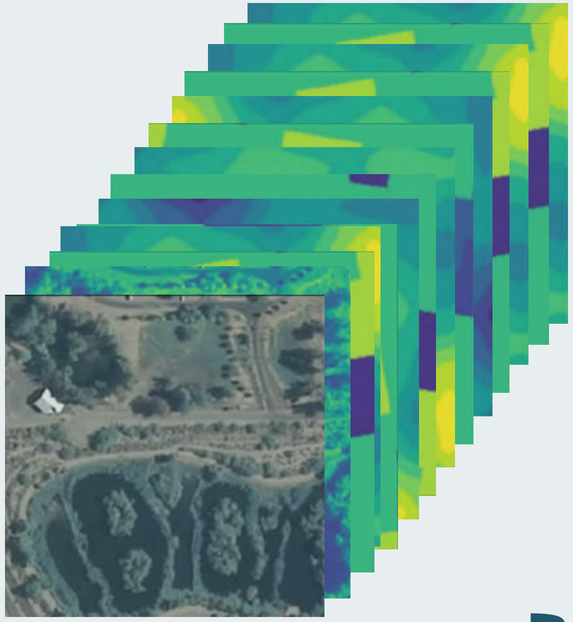


GBIF
Global Biodiversity Information Facility

Occurrences of 181 fish species in total

Environmental dataset and benchmark

Species distribution models using **artificial neural networks**



15 rasters :
- 14 environmental covariates
- one satellite image

30 bands

Species distribution models using **Random Forest** (for benchmark)

Value of the central pixel of the 30 bands

Standard deviation of the 30 bands

Evaluation Metric

$$D2log = 1 - \frac{\sum_{i=1}^{n} \left| log(y_i + 1) - log(\hat{y}_i + 1) \right|}{\sum_{i=1}^{n} \left| log(y_i + 1) - log(\tilde{y} + 1) \right|}$$

where n = number of abundance of species, y_i = true abundance of species *i of* a given species in a given site, $\hat{y}_i$ = predicted abundance of species *i* of a given species in a given site, $\tilde{y}$ = median of true abundance of species

D2log = 0 → the model explains nothing

D2log = 1 → the model predicts the data perfectly

Model performances on fish abundance prediction

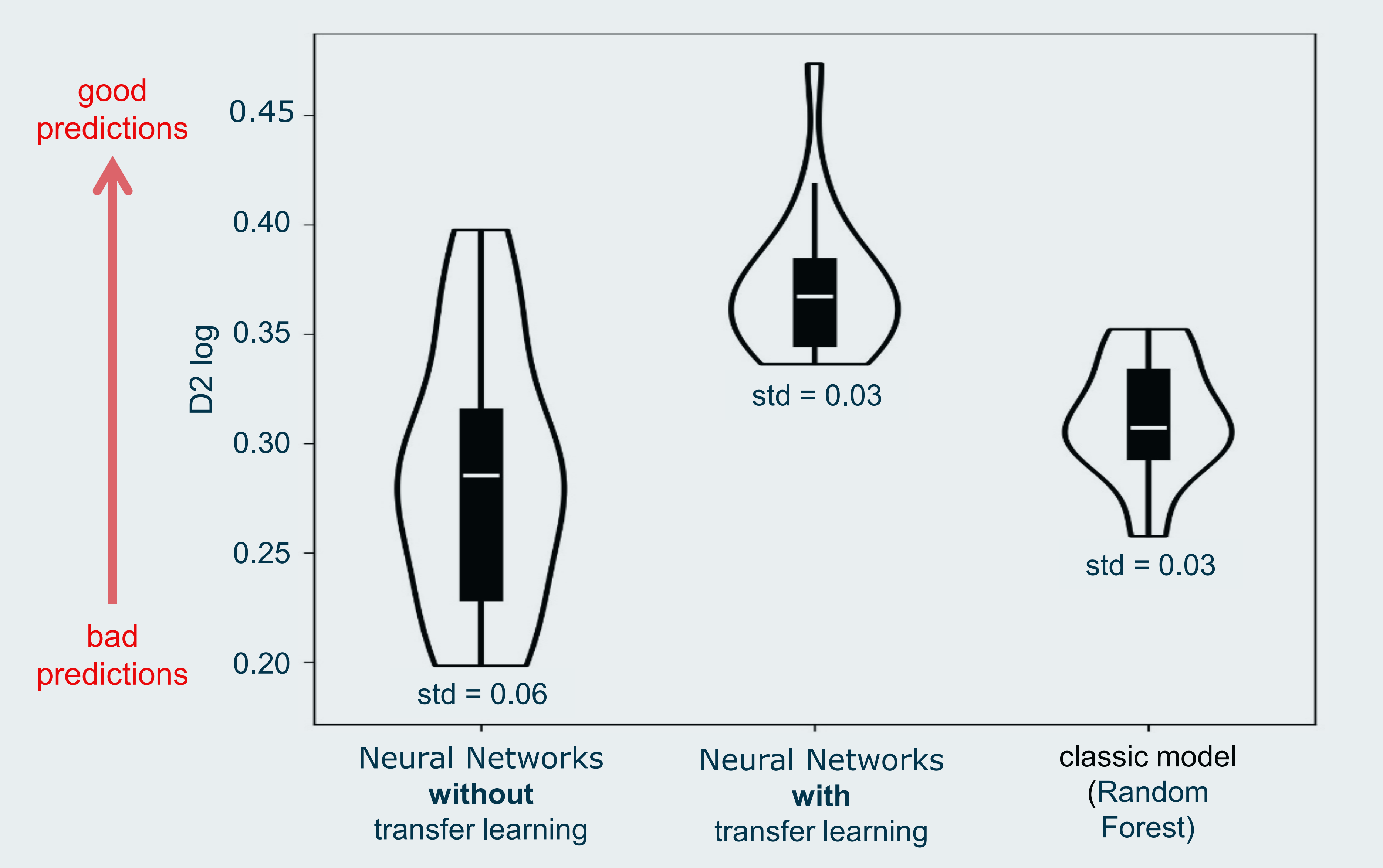


Figure : Violin plots showing the model performances on the fish abundance test set over 20 folds for D2 log. Std = Standard deviation.

Model performances on fish abundance prediction

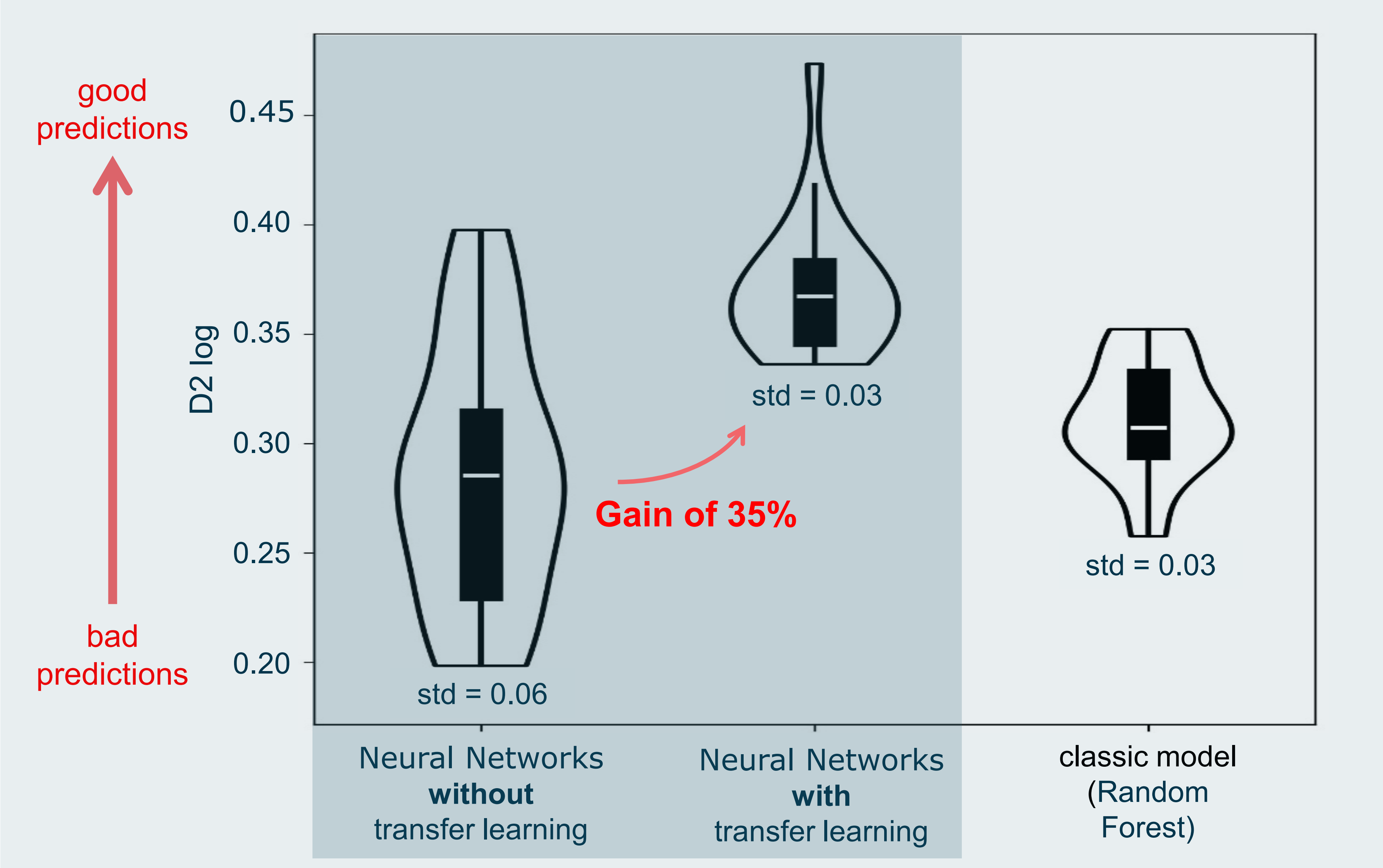


Figure : Violin plots showing the model performances on the fish abundance test set over 20 folds for D2 log. Std = Standard deviation.

Model performances on fish abundance prediction

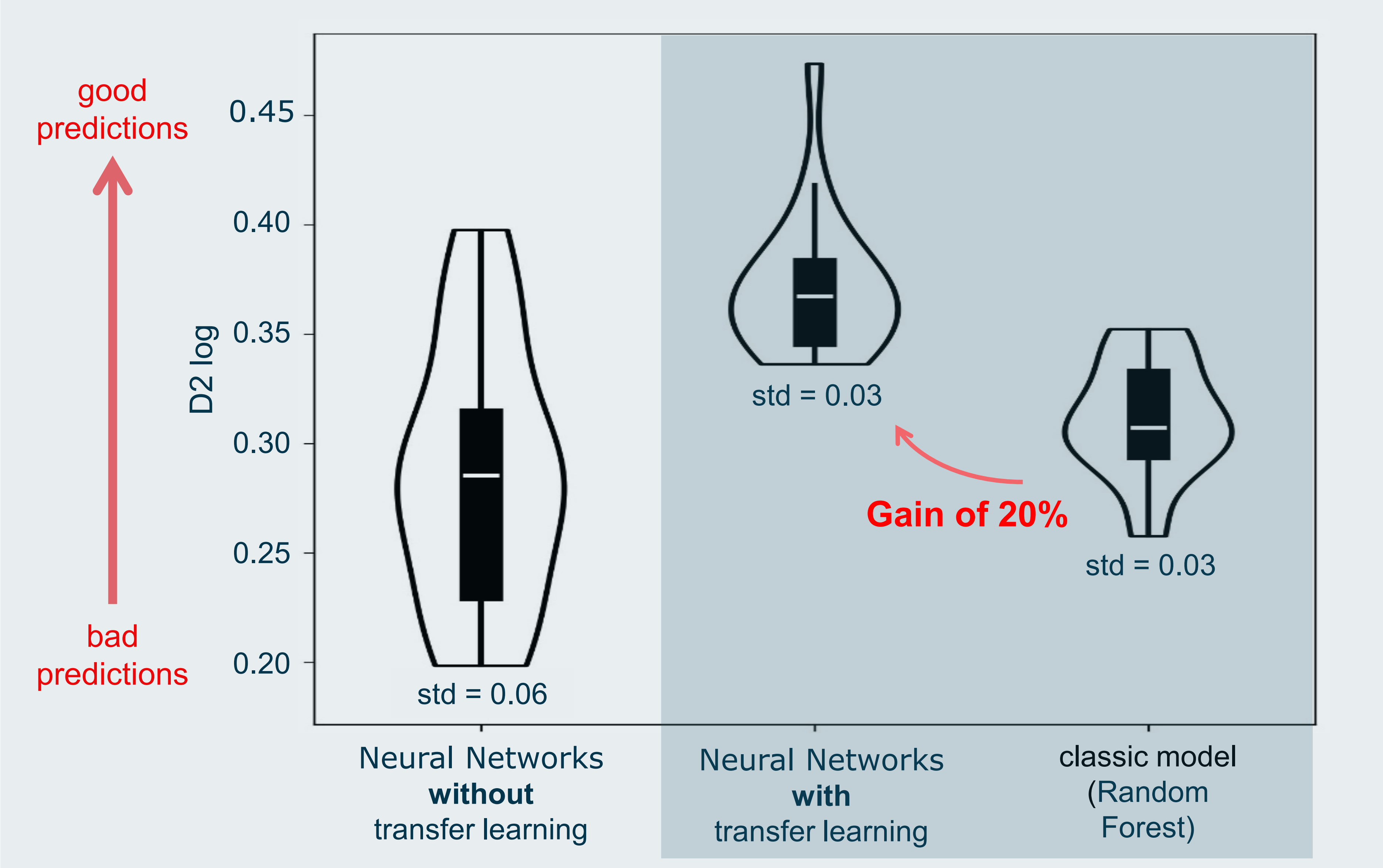


Figure : Violin plots showing the model performances on the fish abundance test set over 20 folds for D2 log. Std = Standard deviation.

Model performances on fish abundance prediction

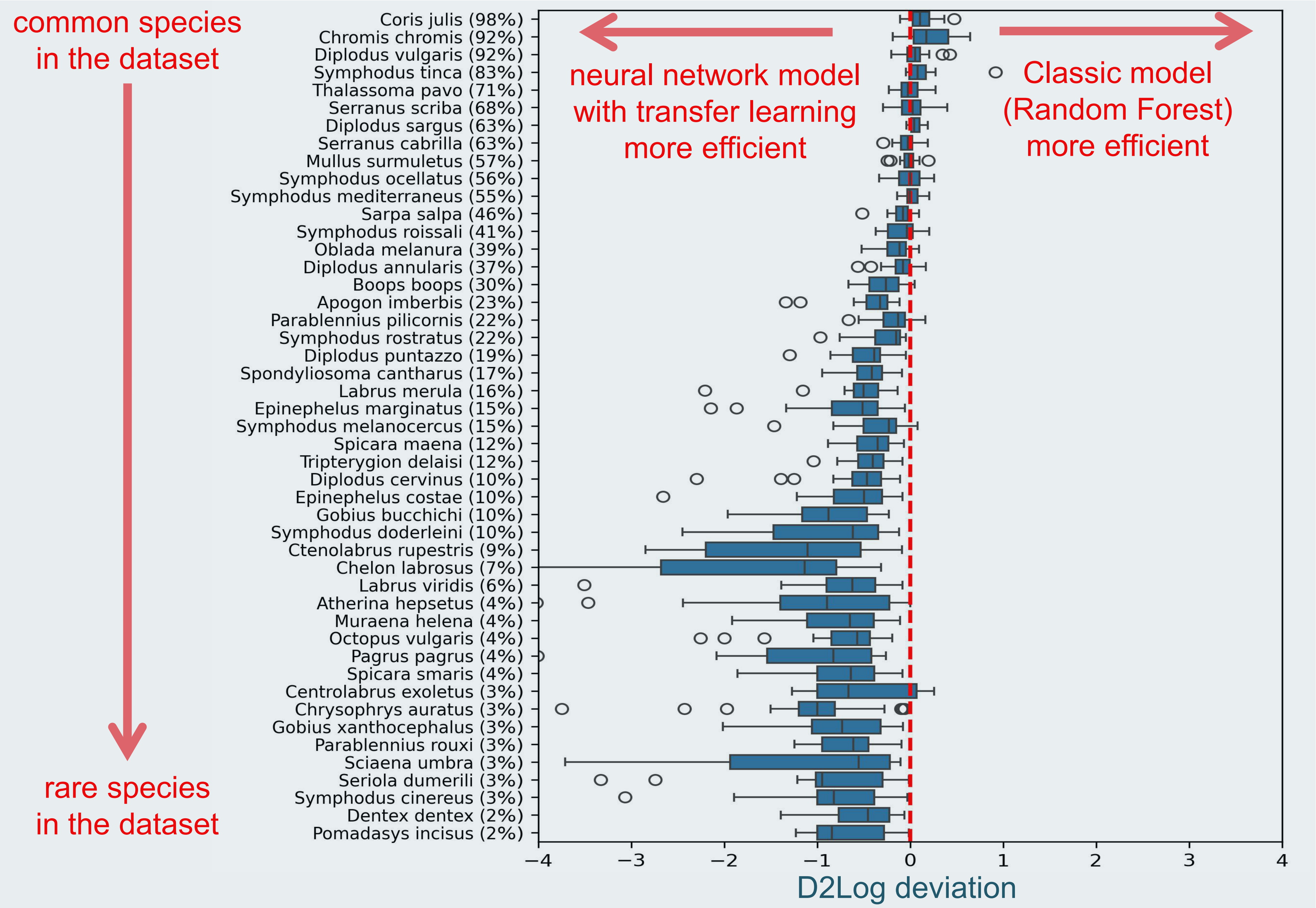


Figure: D2Log deviation by species between the neural network model with transfer learning and classic model calculated for each of the 20 folds.

Model performances on fish abundance prediction

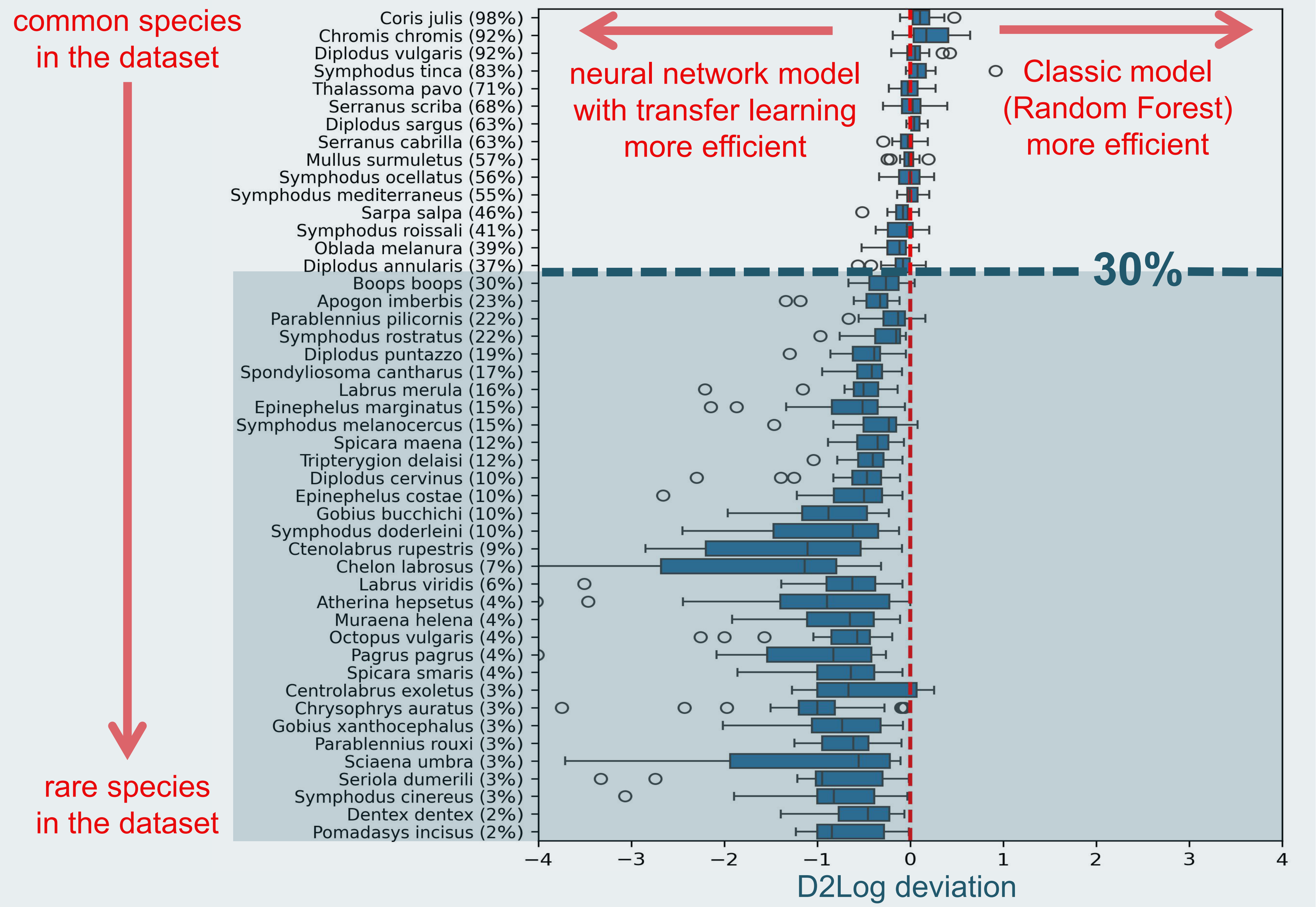


Figure: D2Log deviation by species between the neural network model with transfer learning and classic model calculated for each of the 20 folds.

Conclusion and future work

Species distribution models using neural networks are able to:

- extract relevant information for predicting species abundance from large presence-only
- re-use this information to outperform classic models for predicting species abundance
- better abundance prediction of species poorly represented in datasets than classic models

Conclusion and future work

Species distribution models using neural networks are able to:

- extract relevant information for predicting species abundance from large presence-only
- re-use this information to outperform classic models for predicting species abundance
- better abundance prediction of species poorly represented in datasets than classic models

Question for future work: what is the nature of the information extracted from the species data alone by the neural networks used to optimise predictions of abundance?

Important patterns ?

Interspecific relationships ?

If species A is present,
then species B is absent,
and species C must be present.

Other ?

?

Why focus on species distribution models using artificial neural networks?

Original use

wild boar

Derived use for species prediction models

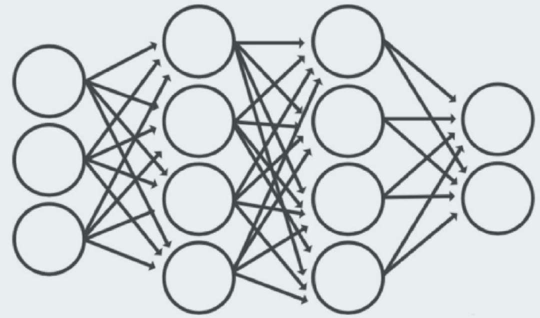
wild boar is present

2 wild boar per km^2

Why focus on species distribution models using artificial neural networks?

Species distribution models using artificial neural networks for predict present

→

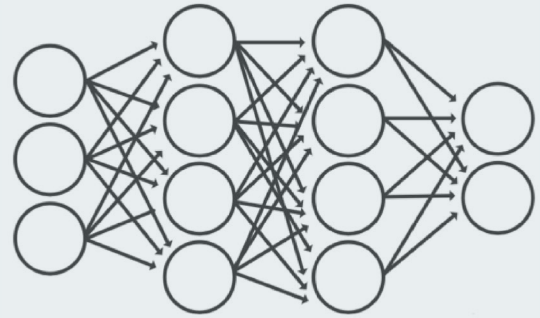


artificial neural networks

→ wild boar : **present**
→ species B : **absent**
→ species C : **absent**
→ species D : **present**

Species distribution models using artificial neural networks for predict abundance

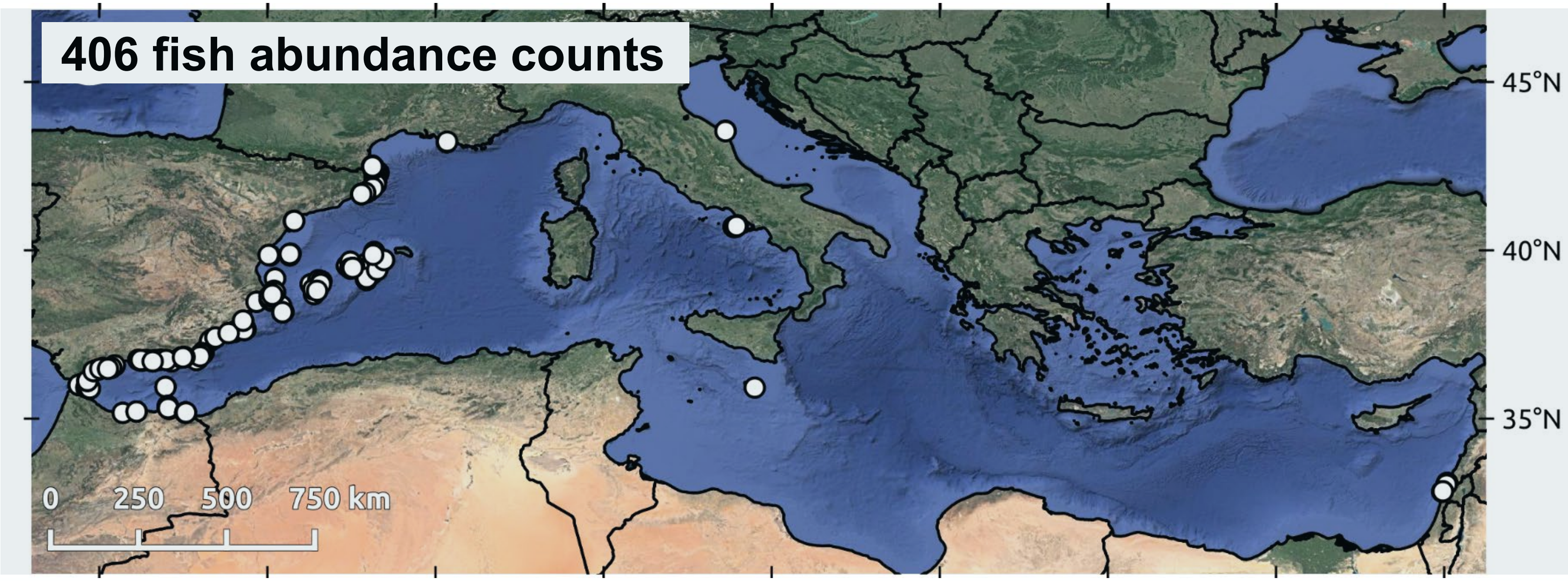
→



artificial neural networks

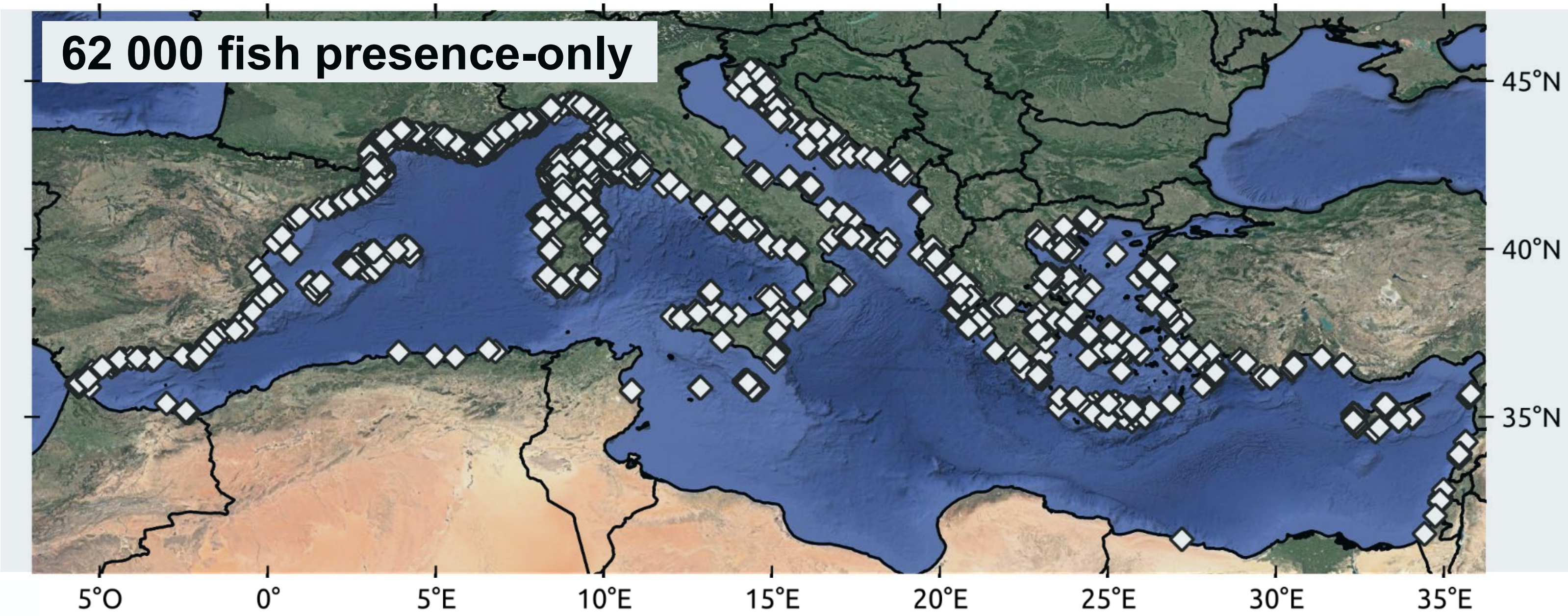
→ wild boar : **2.1 per km²**
→ species B : **0.0 per km²**
→ species C : **0.1 per km²**
→ species D : **5.4 per km²**

Fish occurrence and fish abundance datasets



Reef Life Survey

Abundance of 47 fish species for each count



GBIF
Global Biodiversity Information Facility

Occurrences of 181 fish species in total