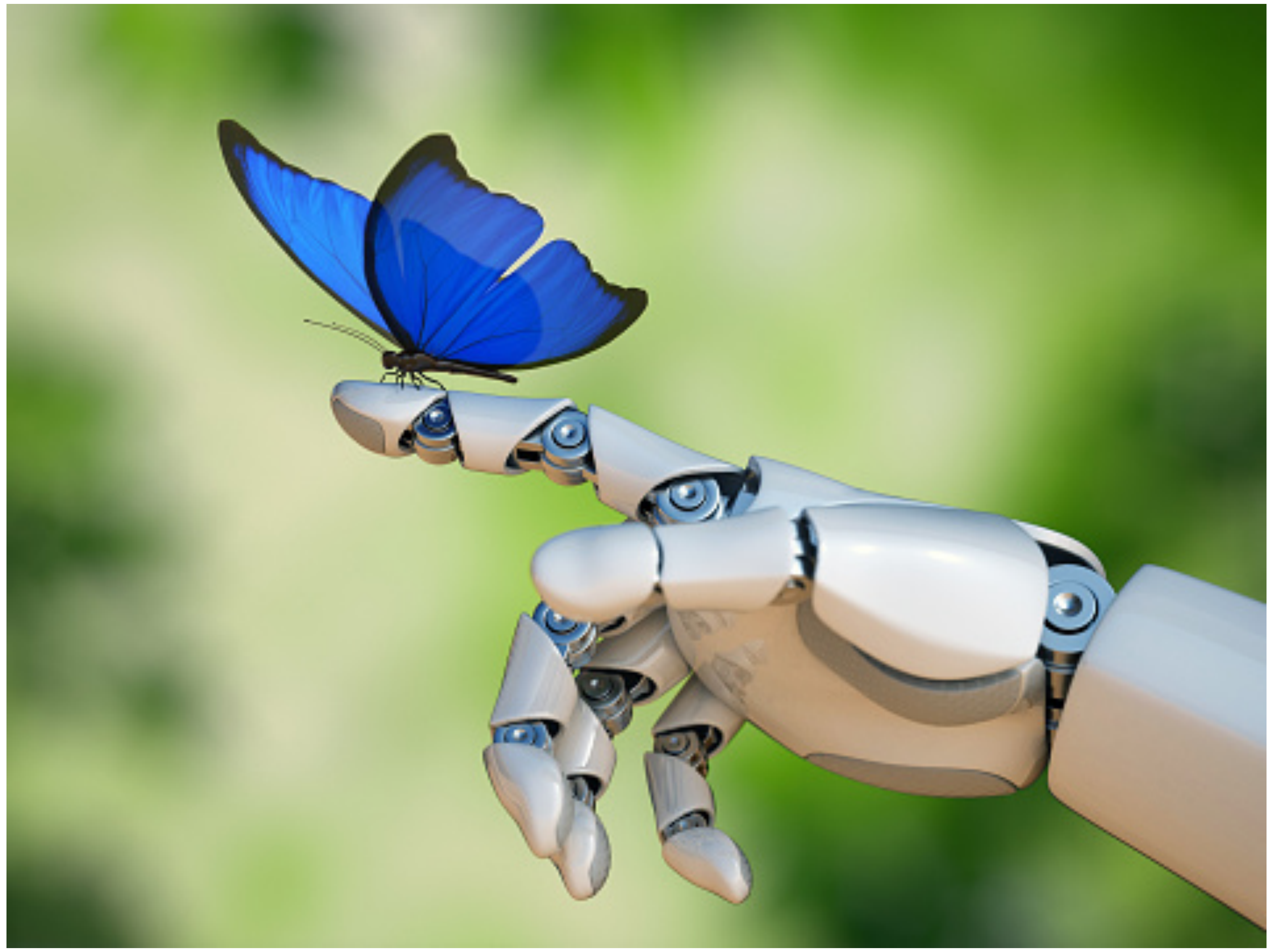
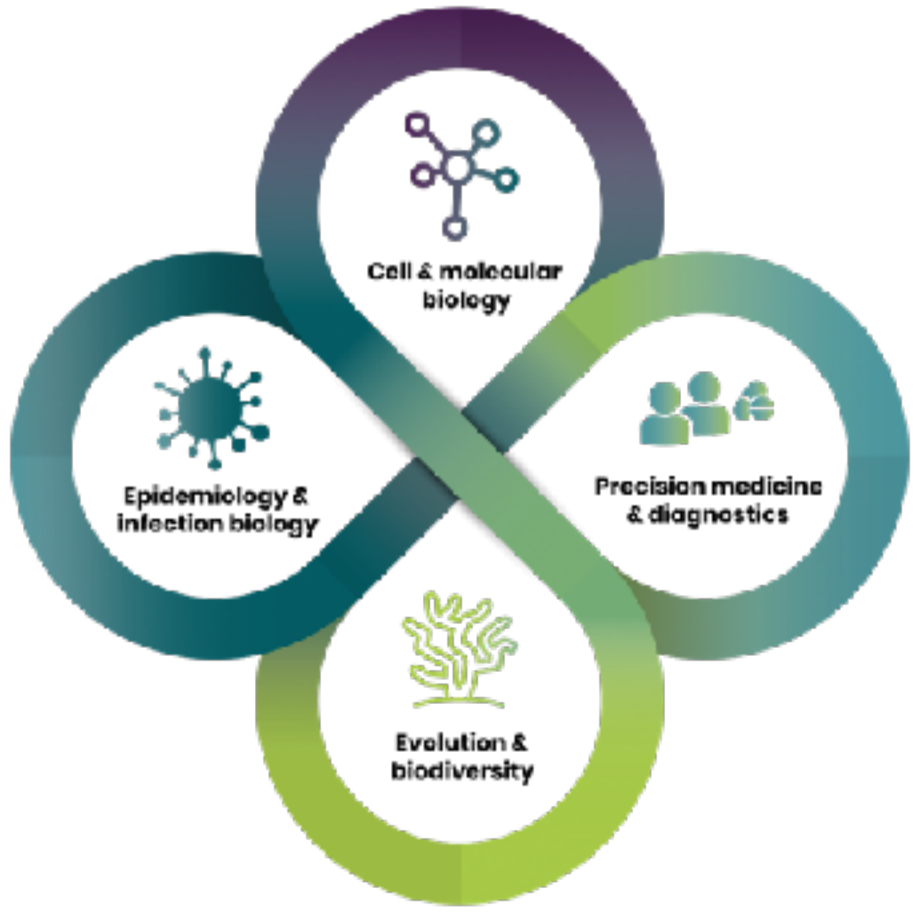
Neural Networks for biodiversity research: challenges and opportunities

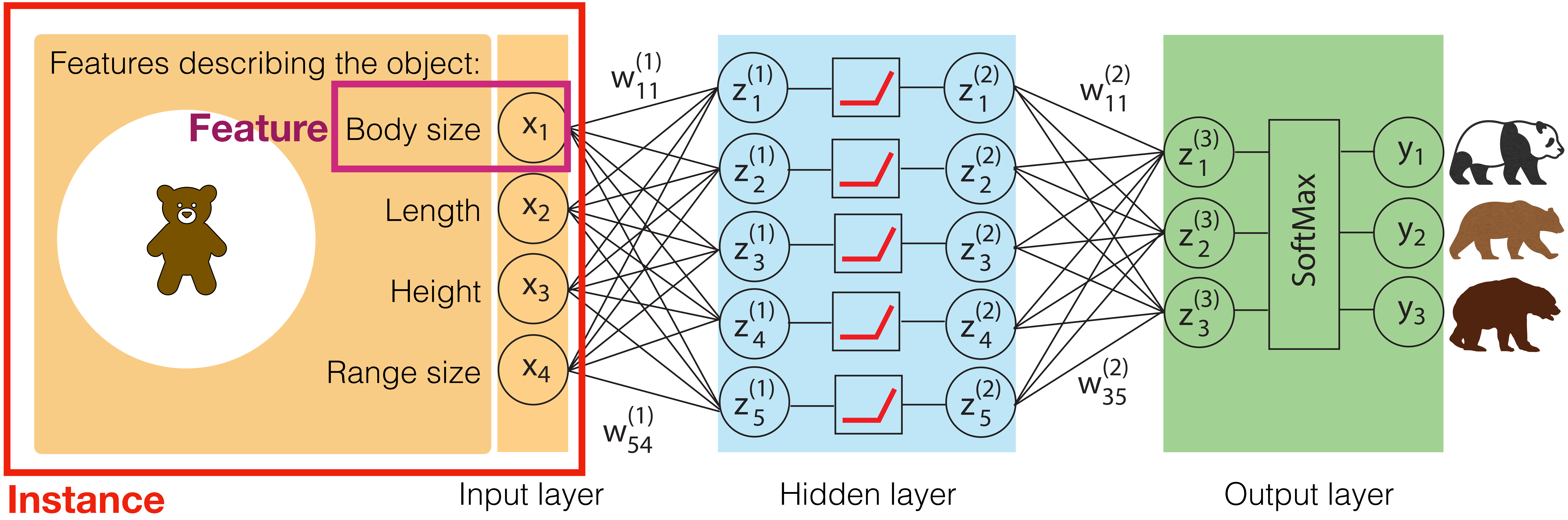


UPPSALA UNIVERSITET

SciLifeLab



Neural Network overview



Numerical representation of your data in multidimensional space: using weights and activation functions to turn input into output

Contraction of the hidden layer into a vector (tensor) of the size of the desired output (here 3)

Slide created by Daniele Silvestro, 2021

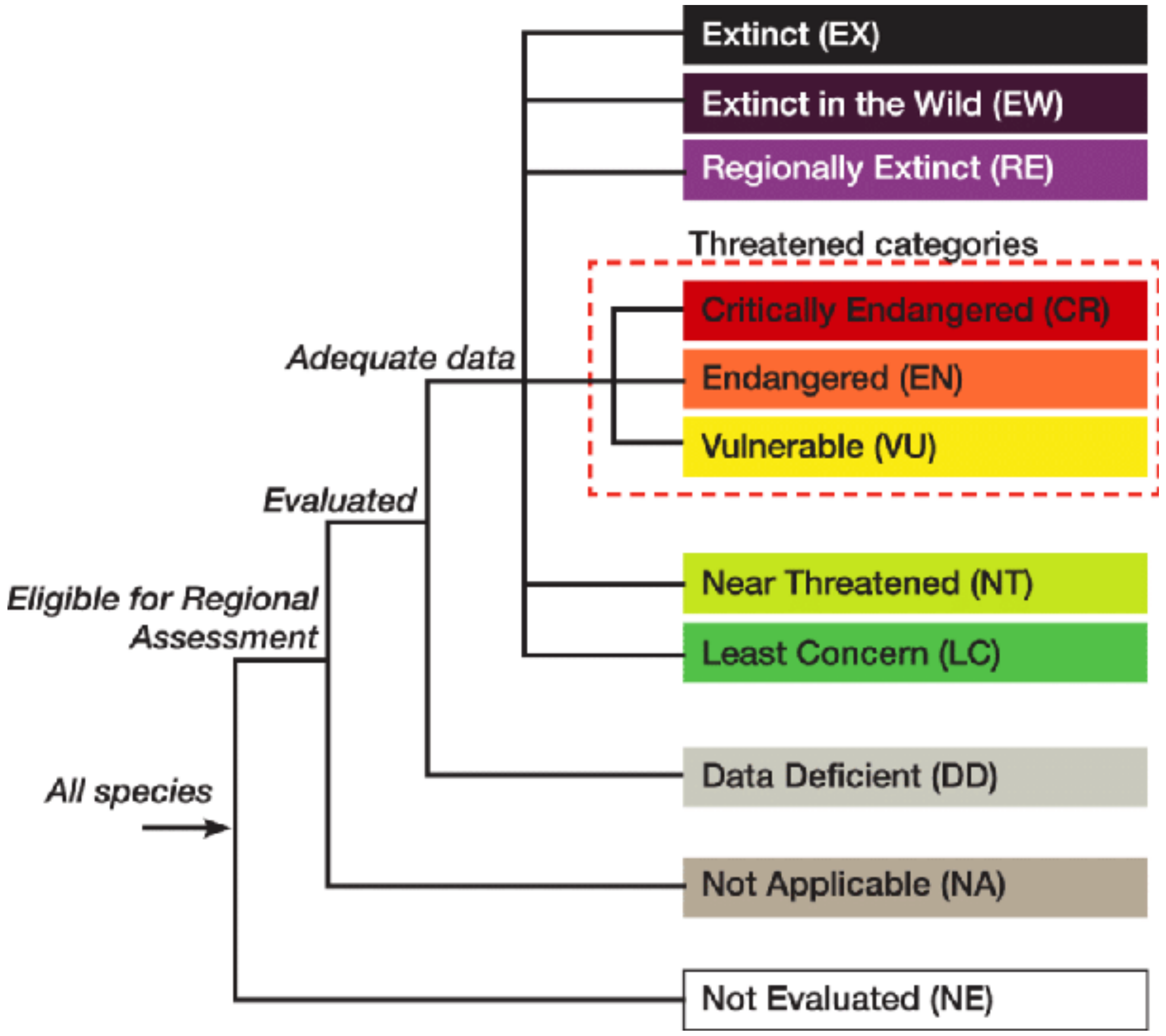
Strengths and weaknesses of neural network models

- versatil/flexible
- allow combination of different data types (e.g. images and numerical data)
- model is being learned, does not require defining a specific model (e.g. linear regression) a priori
- can find signal in complex data
- explores intercorrelations and dependencies between different input features

- "black box" model
- not useful if interested in parameter estimates (e.g., correlation coefficient)
- requires substantial training data, not useful for small datasets (>100 instances)
- over-parameterized, simpler models are often better

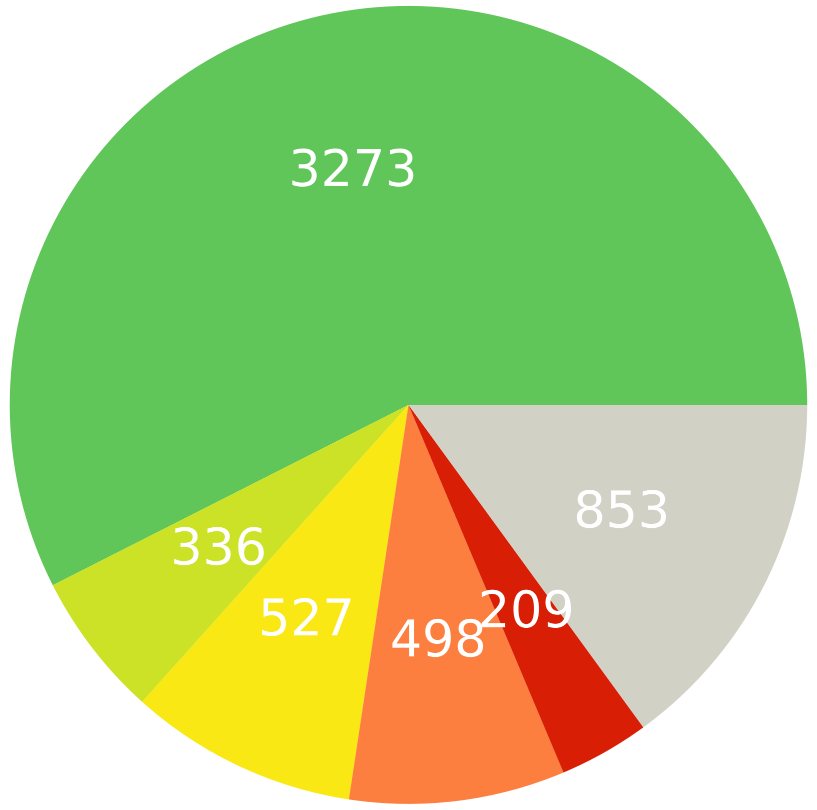
Modeling extinction risks

IUCN Redlist

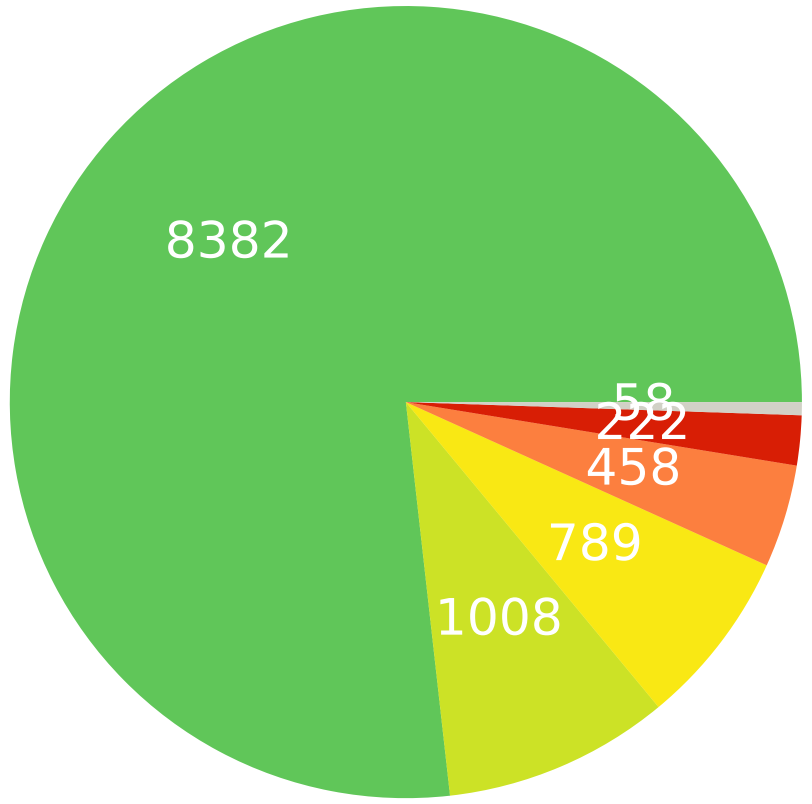


www.iucnredlist.org

Mammalia

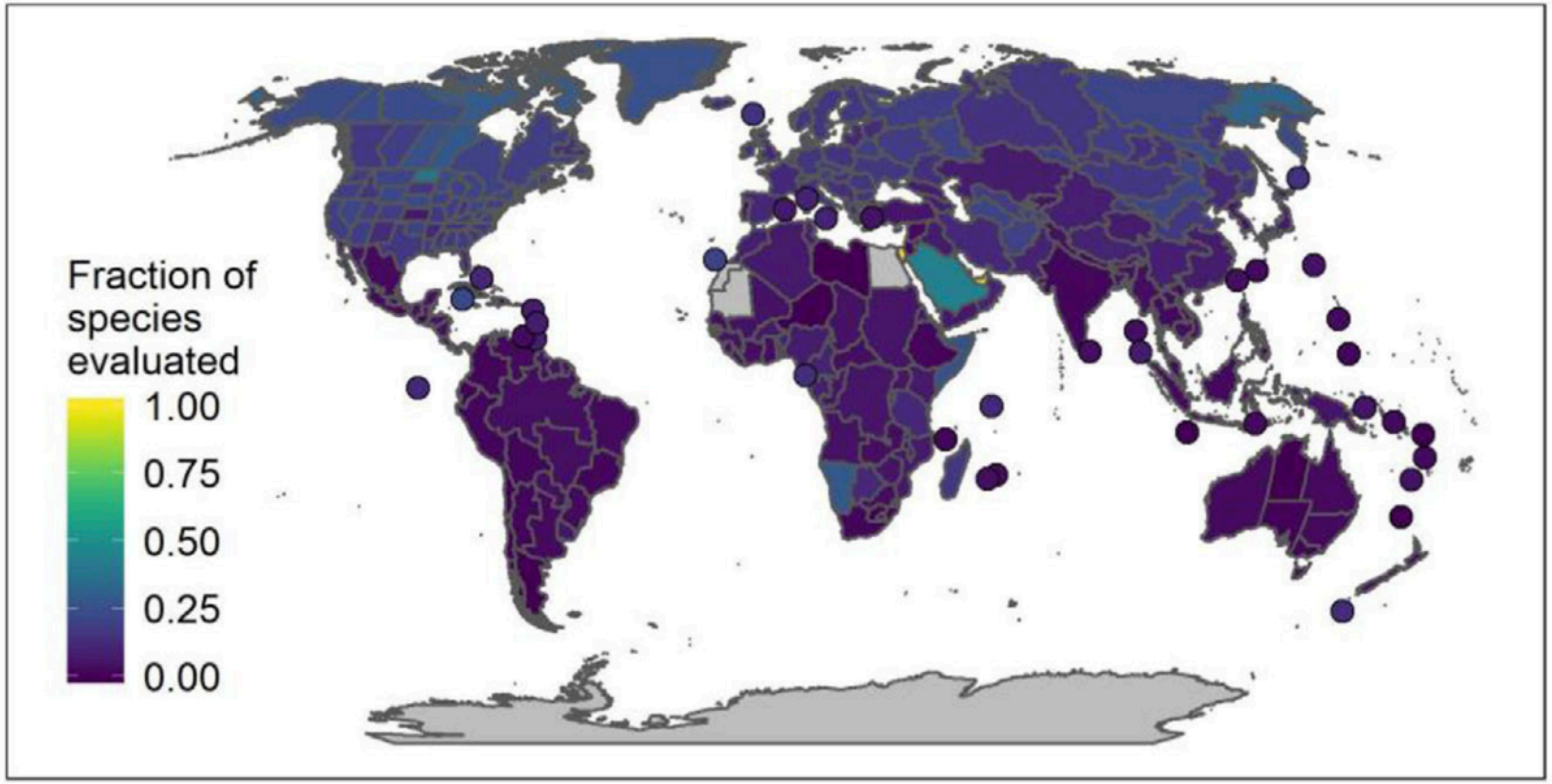


Aves



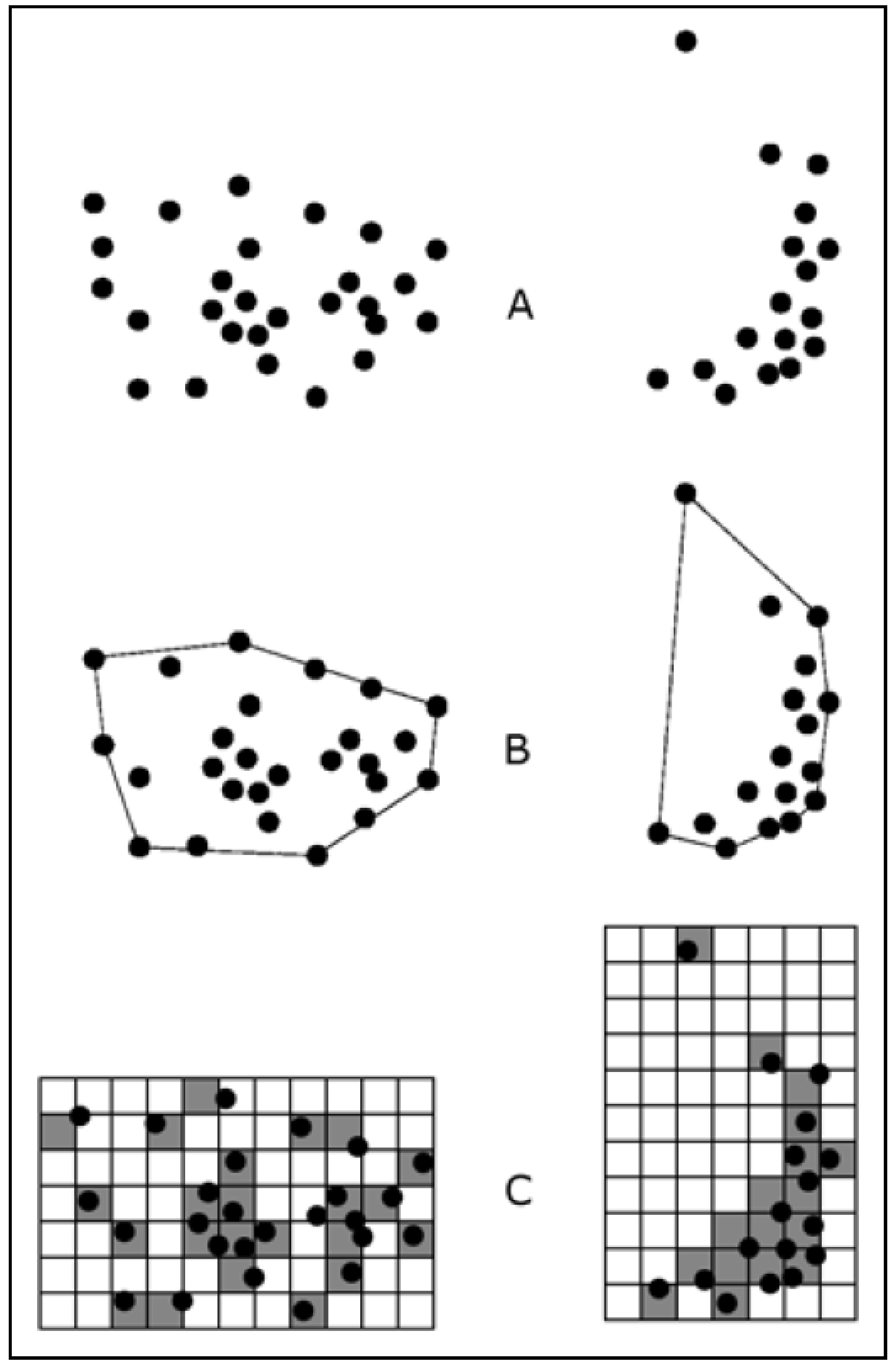
Much is still unknown

Fraction of orchid species assessed by IUCN



Zizka et al., 2020, Conservation Biology

Utility of distribution data

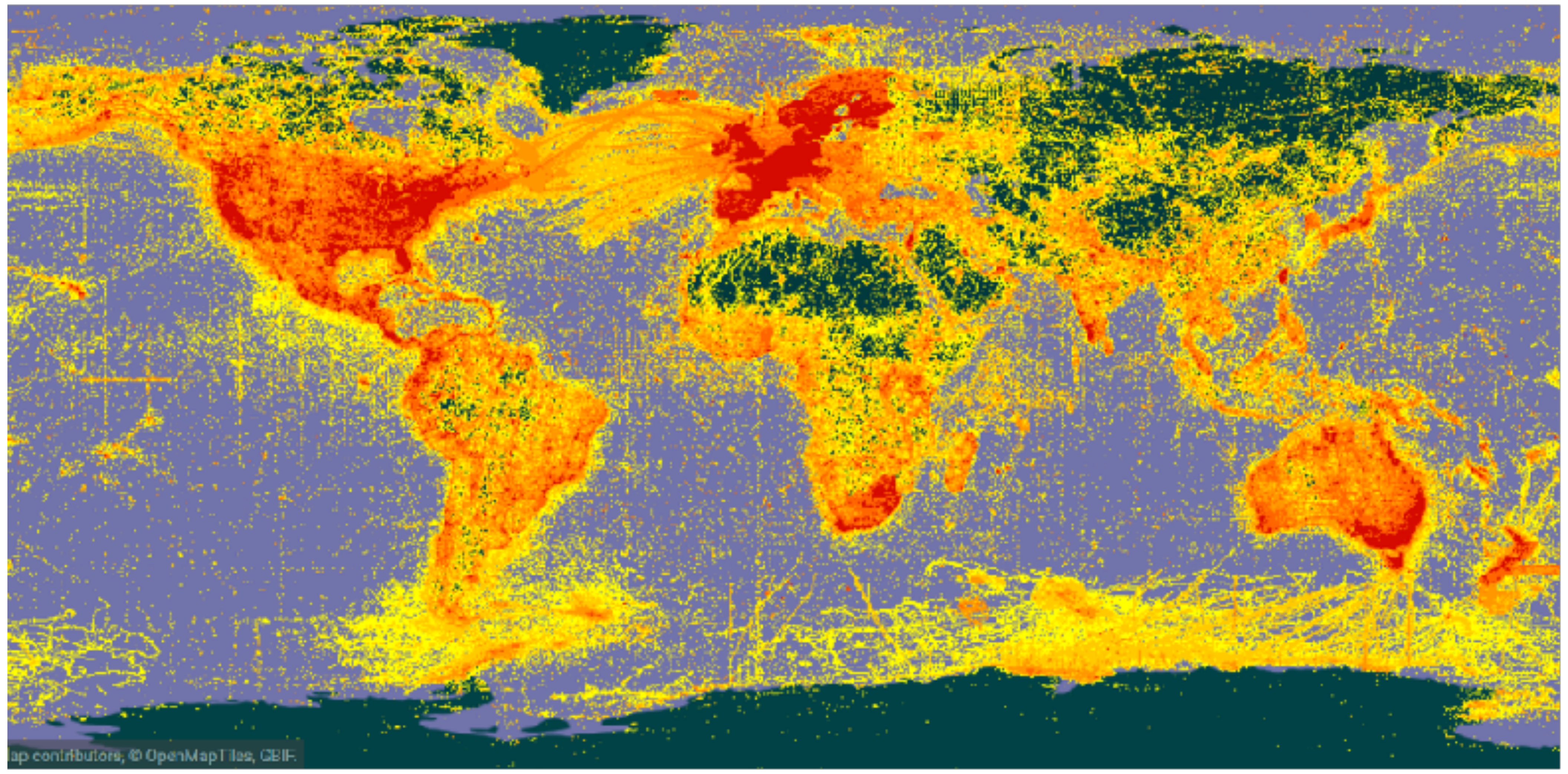


Occurrence records

Extent of occurrence

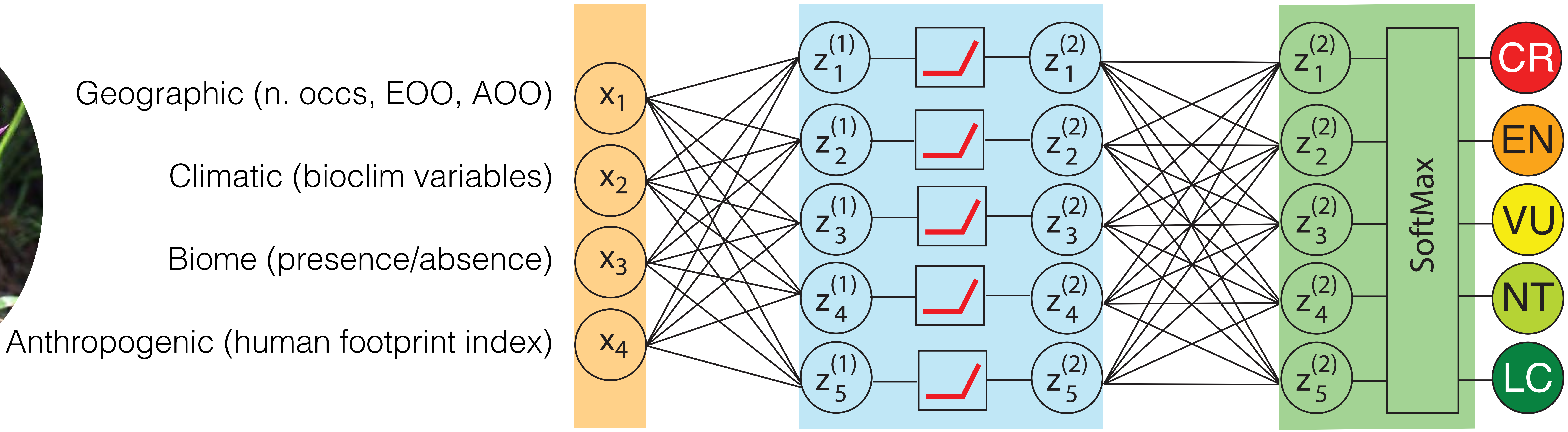
Area of occupancy

IUCN Red List categories and criteria
www.iucnredlist.org



GBIF SBDI

Threat status classification



Model implemented in :

Diversity and Distributions A Journal of Conservation Biogeography

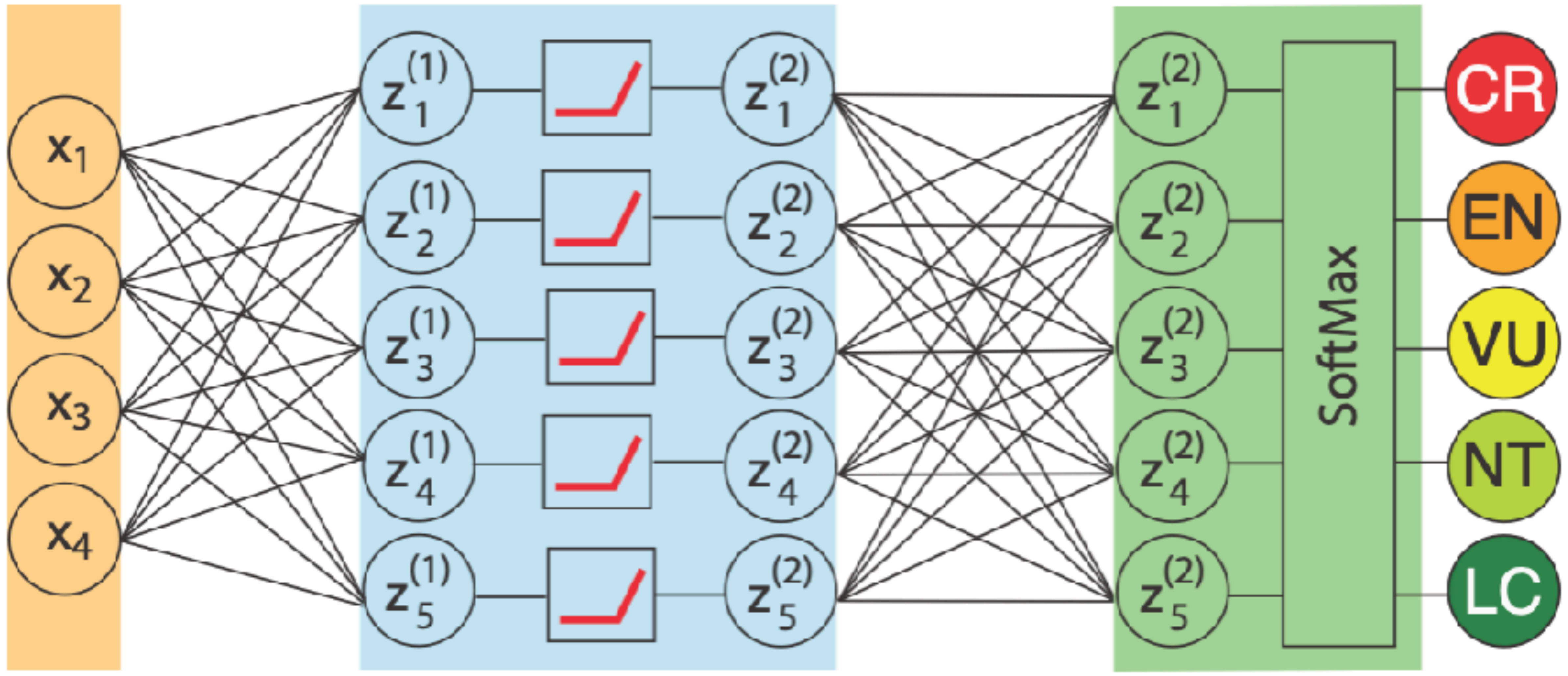
METHOD | Open Access

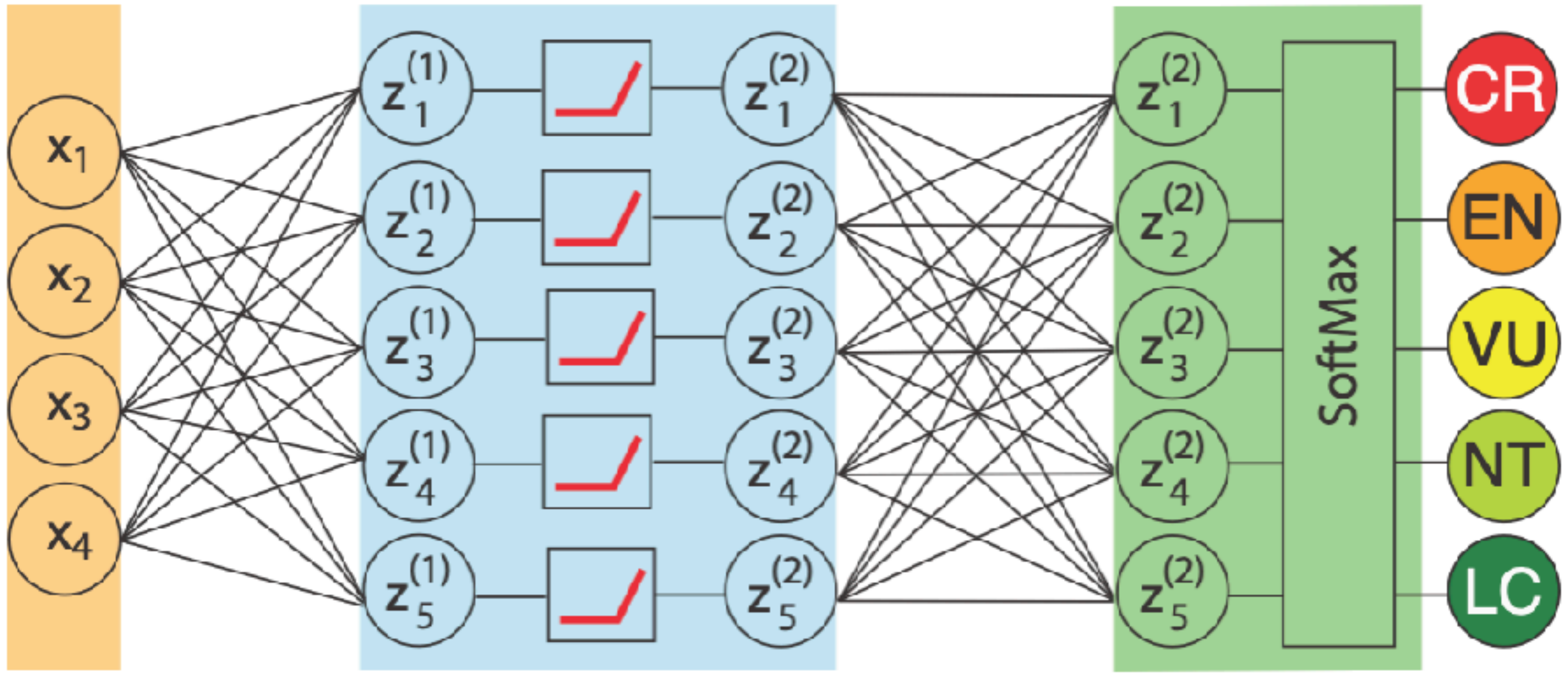
IUCNN – Deep learning approaches to approximate species' extinction risk

Alexander Zizka, Tobias Andermann, Daniele Silvestro

First published: 16 December 2021 | https://doi.org/10.1111/ddi.13450 | Citations: 1

```
# Example for classification model

architecture = []

# Input layer
architecture.append(tf.keras.layers.Flatten(input_shape=[4]))
# 1st hidden layer
architecture.append(tf.keras.layers.Dense(5, activation='relu',use_bias=False))
# Output layer
architecture.append(tf.keras.layers.Dense(5, activation='softmax',use_bias=False))

# Compile the model
model = tf.keras.Sequential(architecture)
model.compile(loss='categorical_crossentropy',
                  optimizer='adam', metrics=['accuracy'])

# Get overview of model architecture
model.summary()
```

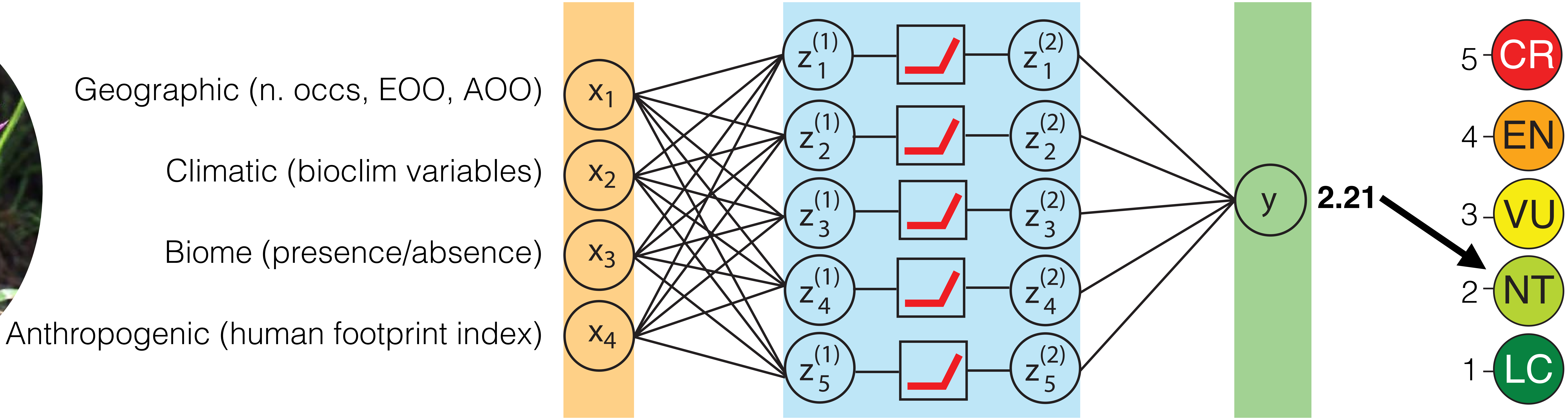





```python
# Example for classification model

architecture = []

# Input layer
architecture.append(tf.keras.layers.Flatten(input_shape=[4]))
# 1st hidden layer
architecture.append(tf.keras.layers.Dense(5, activation='relu',use_bias=False))
# Output layer
architecture.append(tf.keras.layers.Dense(5, activation='softmax',use_bias=False))

# Compile the model
model = tf.keras.Sequential(architecture)
model.compile(loss='categorical_crossentropy',
              optimizer='adam', metrics=['accuracy'])

# Get overview of model architecture
model.summary()
```

```
>>> model.summary()
Model: "sequential_9"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
flatten_9 (Flatten)          (None, 4)                 0
_________________________________________________________________
dense_21 (Dense)             (None, 5)                 20
_________________________________________________________________
dense_22 (Dense)             (None, 5)                 25
=================================================================
Total params: 45
Trainable params: 45
Non-trainable params: 0
_________________________________________________________________
```

Threat status regression model



Model implemented in :


Diversity and Distributions
A Journal of Conservation Biogeography
METHOD | Open Access

IUCNN – Deep learning approaches to approximate species' extinction risk

Alexander Zizka, Tobias Andermann, Daniele Silvestro

First published: 16 December 2021 | https://doi.org/10.1111/ddi.13450 | Citations: 1

Species occurrence data
from public databases

Longitude	Latitude
136.5329	-28.4025
150.8236	-24.8218
136.2989	-25.2995
133.7812	-25.9481
141.1054	-17.6605
141.7411	-31.0796
134.3736	-14.1506
133.6164	-22.2898
133.8822	-14.6768
141.817	-20.0843
135.409	-15.5381
140.8475	-17.4635
137.0116	-17.8959
143.7243	-20.1088
133.2699	-22.2836
136.5648	-25.7836
135.4576	-26.4805
133.6333	-23.8357
135.83	-16.8239
150.0949	-23.6582
141.1789	-17.295
139.53	-34.32
141.0037	-23.6464
135.8615	-28.0871
135.5817	-26.3667
134.3555	-14.177
133.4385	-25.7375
134.4207	-14.0479
142.4779	-20.0137
136.8653	-18.7861
135.3074	-26.5741

- Total occurrences
- Longitudinal range
- Mean Longitude
- Latitudinal range
- Mean Latitude ...

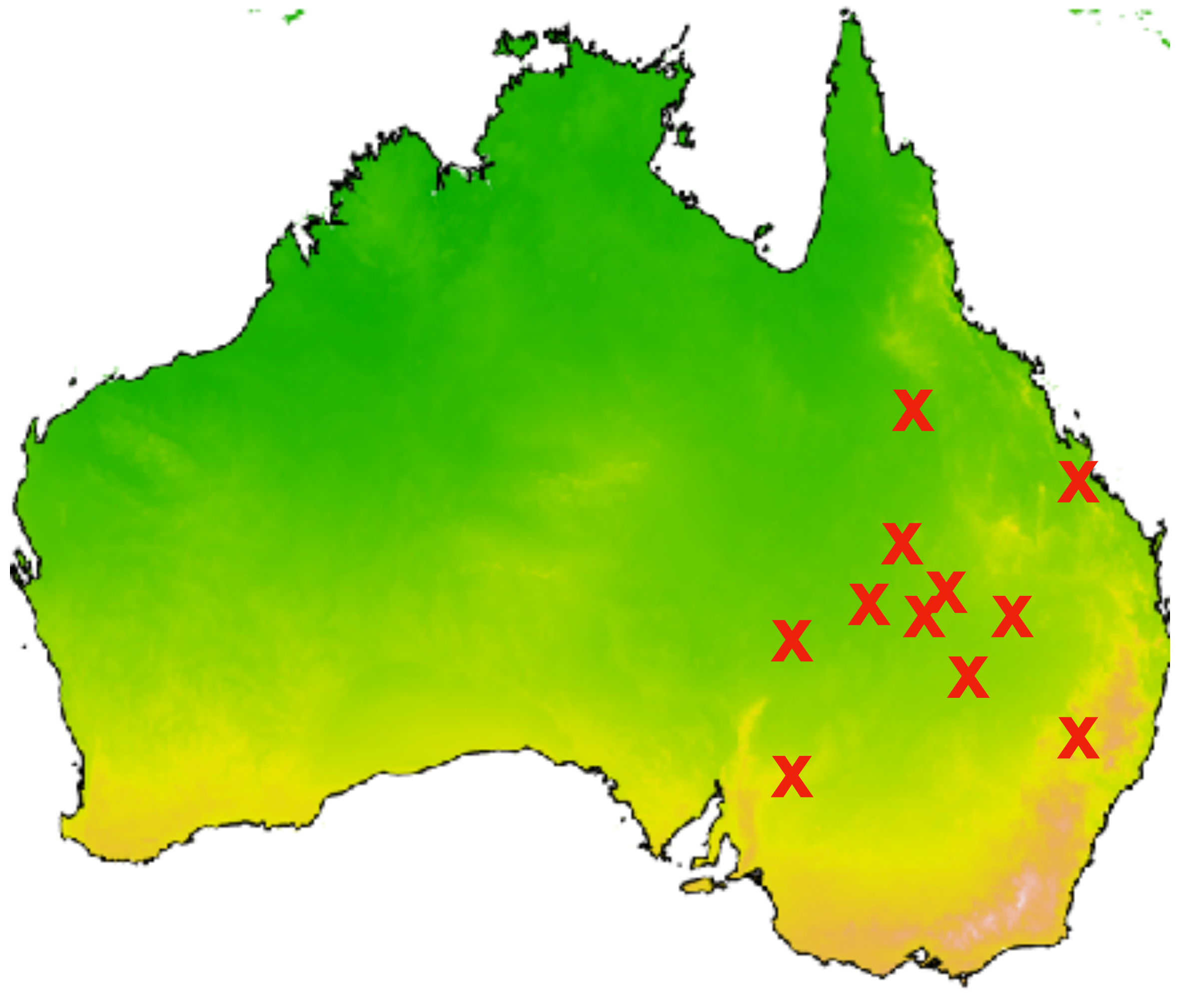
Geographic features

Species occurrence data
from public databases

Longitude	Latitude
136.5329	-28.4025
150.8236	-24.8218
136.2989	-25.2995
133.7812	-25.9481
141.1054	-17.6605
141.7411	-31.0796
134.3736	-14.1506
133.6164	-22.2898
133.8822	-14.6768
141.817	-20.0843
135.409	-15.5381
140.8475	-17.4635
137.0116	-17.8959
143.7243	-20.1088
133.2699	-22.2836
136.5648	-25.7836
135.4576	-26.4805
133.6333	-23.8357
135.83	-16.8239
150.0949	-23.6582
141.1789	-17.295
139.53	-34.32
141.0037	-23.6464
135.8615	-28.0871
135.5817	-26.3667
134.3555	-14.177
133.4385	-25.7375
134.4207	-14.0479
142.4779	-20.0137
136.8653	-18.7861
135.3074	-26.5741

- Total occurrences
- Longitudinal range
- Mean Longitude
- Latitudinal range
- Mean Latitude ...
- Median temperature
- Range of temperature
- Median precipitation ...

Geographic features



Climatic features

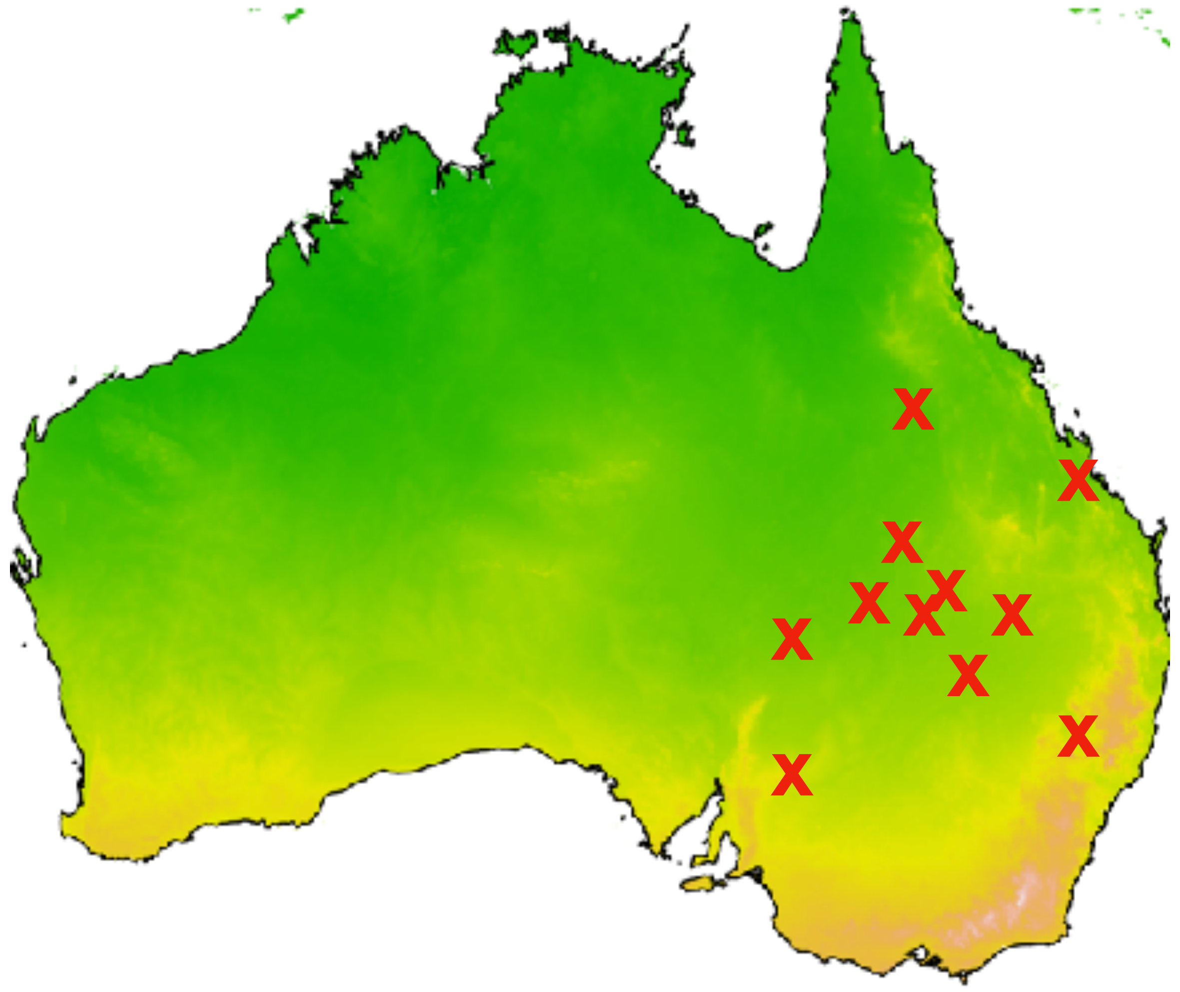
GBIF

Species occurrence data from public databases

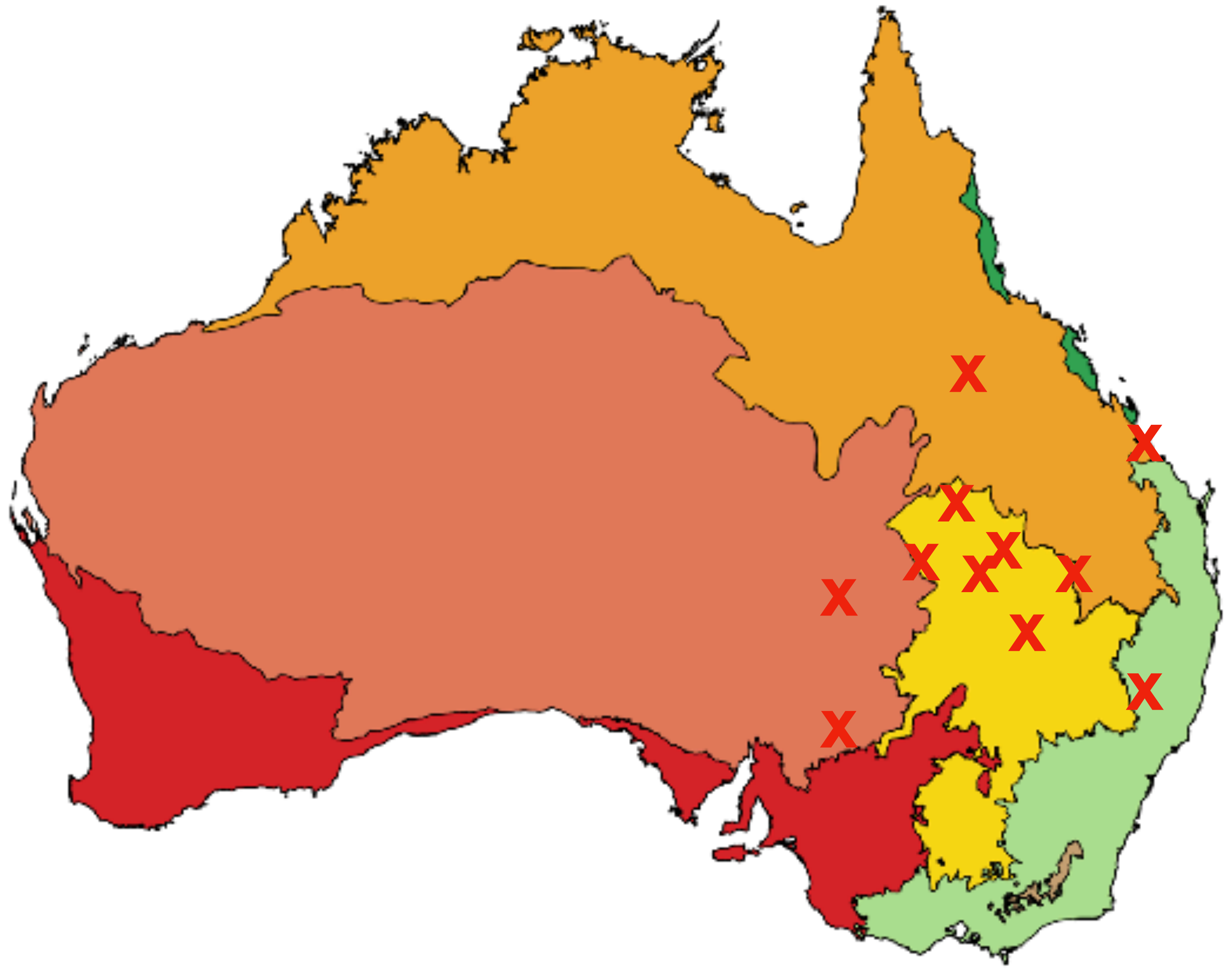
Longitude	Latitude
136.5329	-28.4025
150.8236	-24.8218
136.2989	-25.2995
133.7812	-25.9481
141.1054	-17.6605
141.7411	-31.0796
134.3736	-14.1506
133.6164	-22.2898
133.8822	-14.6768
141.817	-20.0843
135.409	-15.5381
140.8475	-17.4635
137.0116	-17.8959
143.7243	-20.1088
133.2699	-22.2836
136.5648	-25.7836
135.4576	-26.4805
133.6333	-23.8357
135.83	-16.8239
150.0949	-23.6582
141.1789	-17.295
139.53	-34.32
141.0037	-23.6464
135.8615	-28.0871
135.5817	-26.3667
134.3555	-14.177
133.4385	-25.7375
134.4207	-14.0479
142.4779	-20.0137
136.8653	-18.7861
135.3074	-26.5741

- Total occurrences
- Longitudinal range
- Mean Longitude
- Latitudinal range
- Mean Latitude ...
- Median temperature
- Range of temperature
- Median precipitation ...
- Presence in biome 1
- Presence in biome 2 ...

Geographic features



Climatic features



Biome features

GBIF

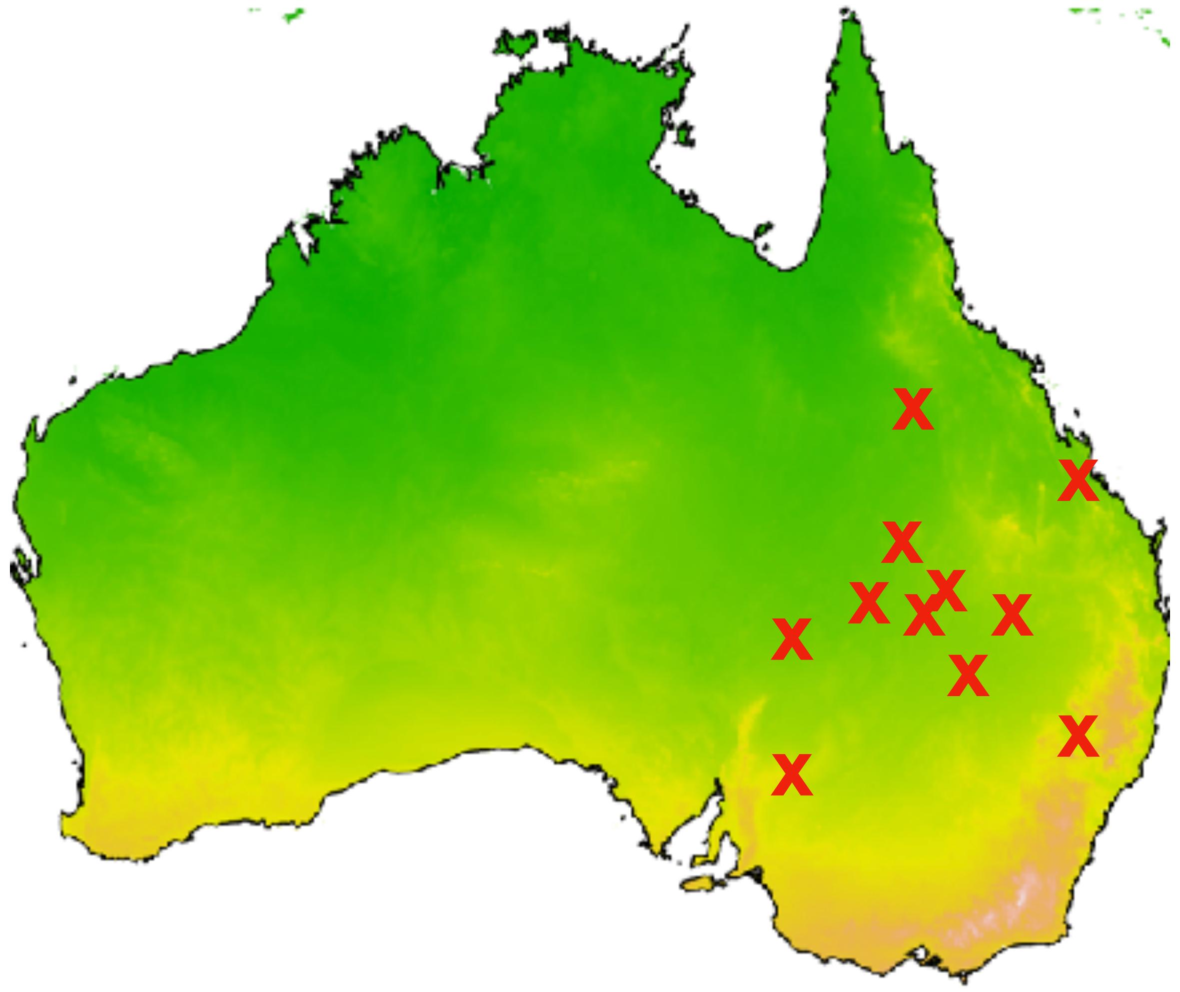
Species occurrence data
from public databases

Longitude	Latitude
136.5329	-28.4025
150.8236	-24.8218
136.2989	-25.2995
133.7812	-25.9481
141.1054	-17.6605
141.7411	-31.0796
134.3736	-14.1506
133.6164	-22.2898
133.8822	-14.6768
141.817	-20.0843
135.409	-15.5381
140.8475	-17.4635
137.0116	-17.8959
143.7243	-20.1088
133.2699	-22.2836
136.5648	-25.7836
135.4576	-26.4805
133.6333	-23.8357
135.83	-16.8239
150.0949	-23.6582
141.1789	-17.295
139.53	-34.32
141.0037	-23.6464
135.8615	-28.0871
135.5817	-26.3667
134.3555	-14.177
133.4385	-25.7375
134.4207	-14.0479
142.4779	-20.0137
136.8653	-18.7861
135.3074	-26.5741

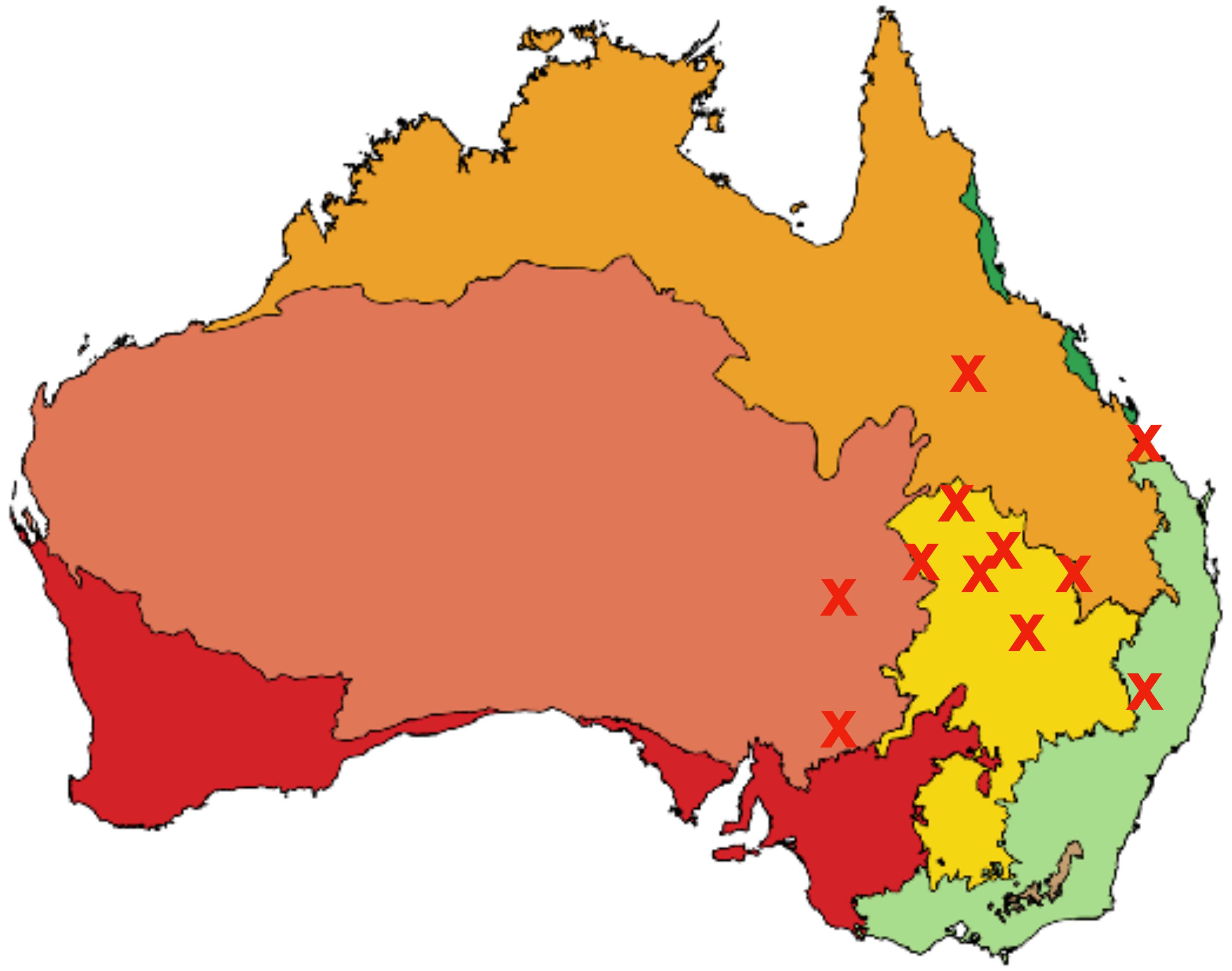
- Total occurrences
- Longitudinal range
- Mean Longitude
- Latitudinal range
- Mean Latitude ...
- Median temperature
- Range of temperature
- Median precipitation ...
- Presence in biome 1
- Presence in biome 2 ...
- Fraction of records with high human footprint
- Fraction of records with medium human footprint

...

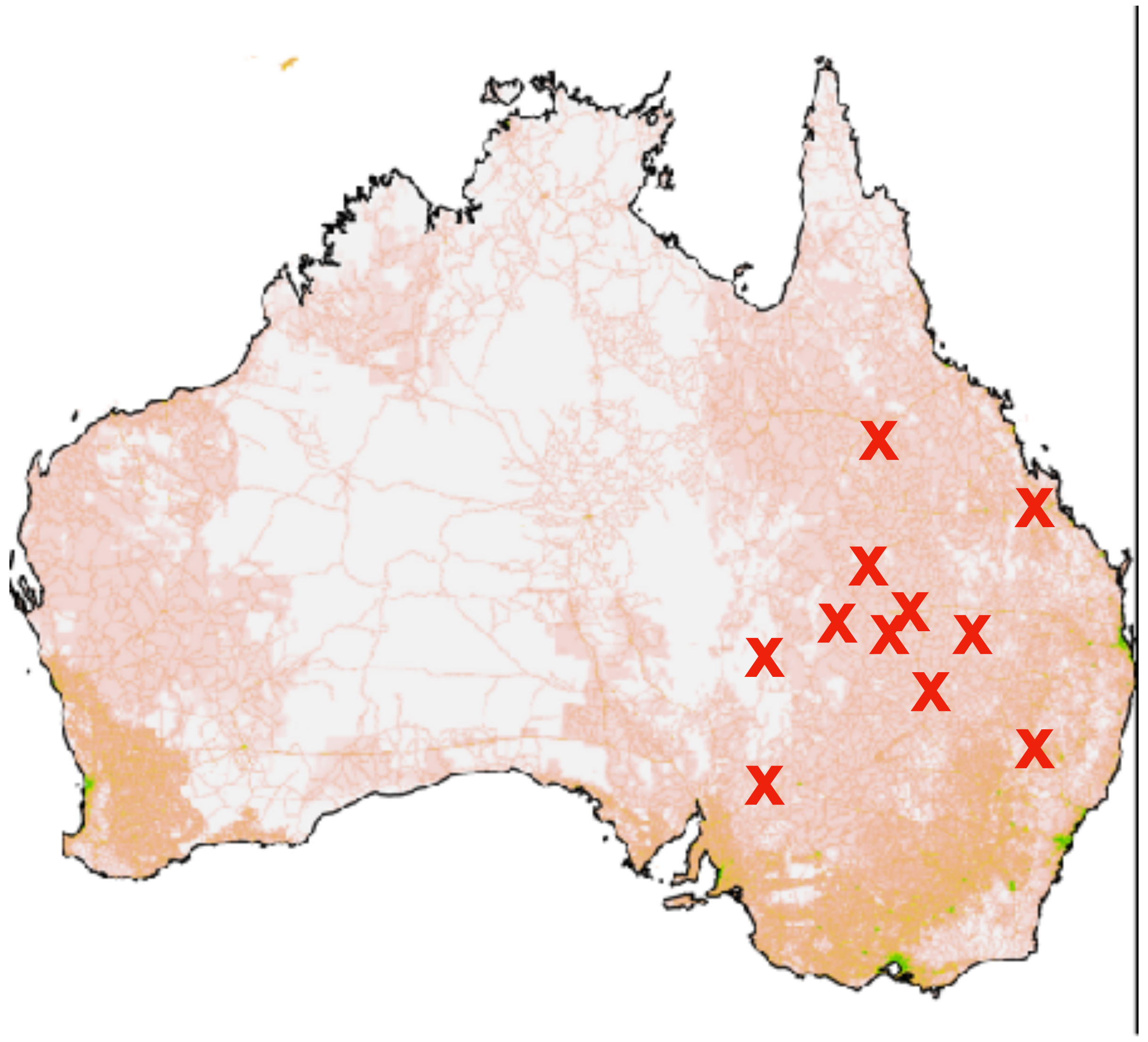
Geographic features



Climatic features



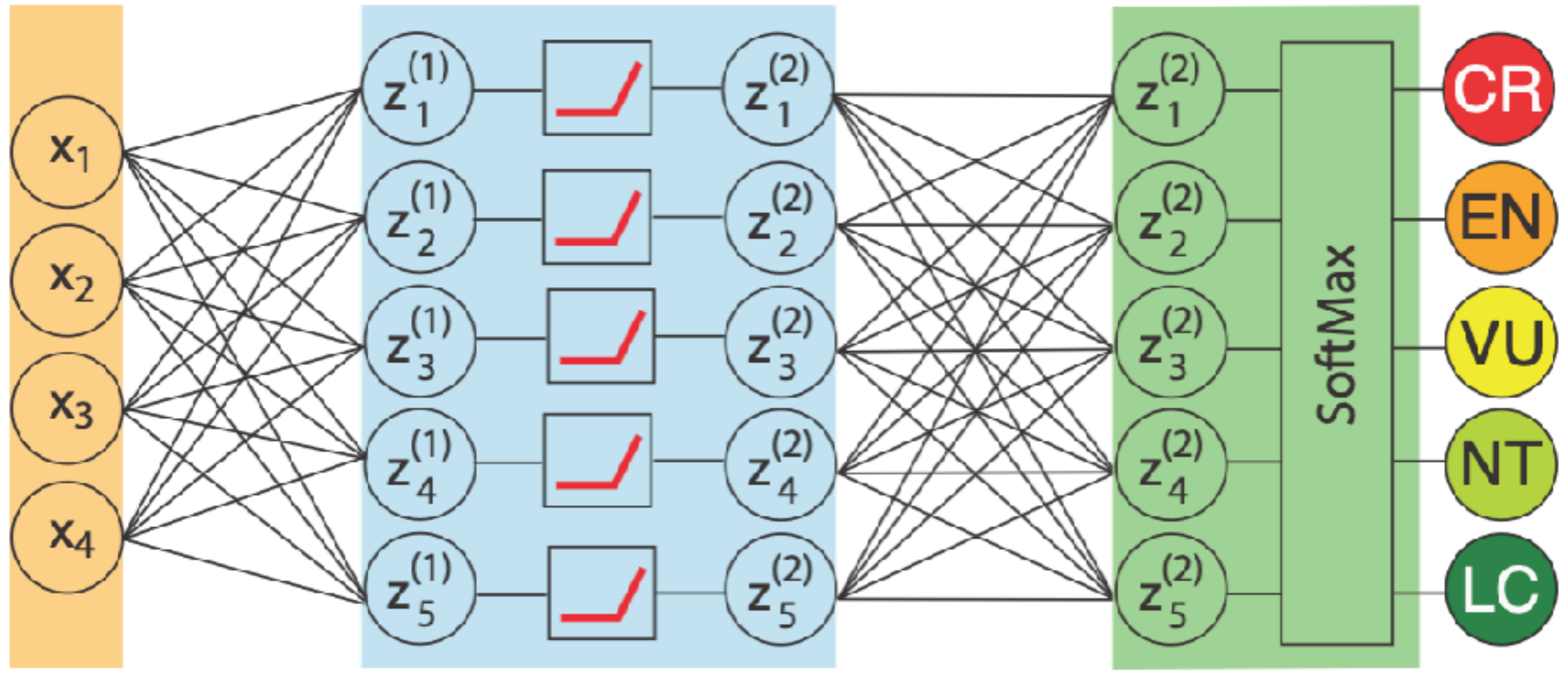
Biome features



Human footprint features

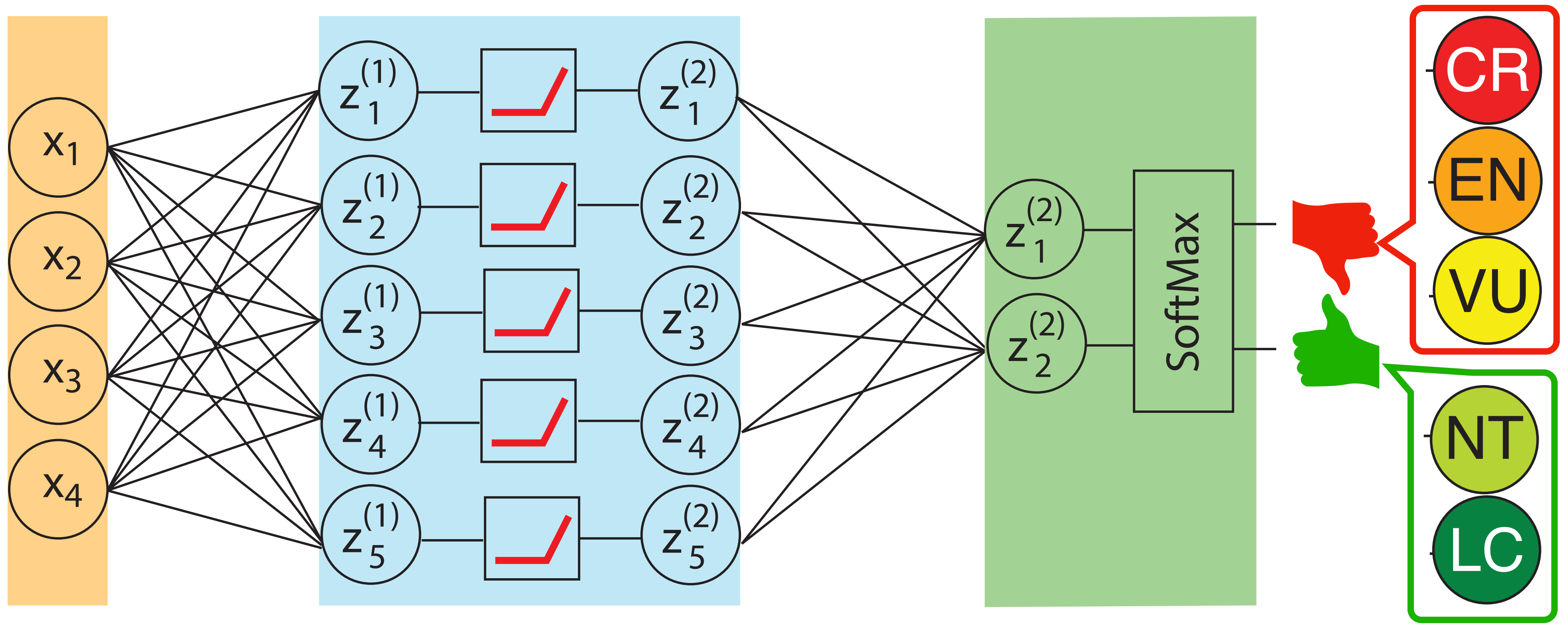
Model performance

5-class model



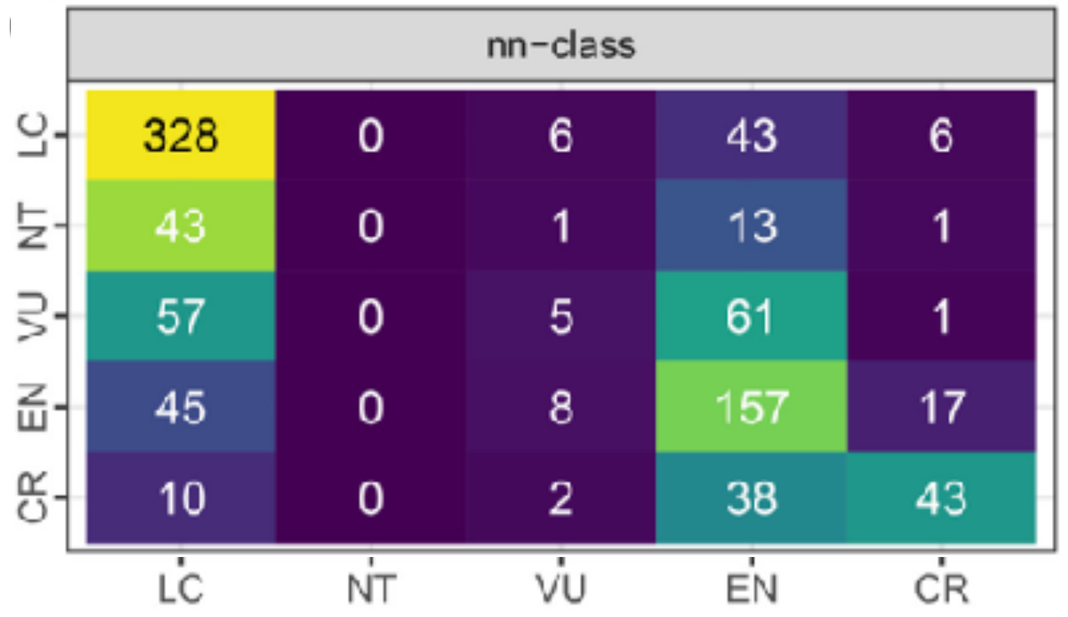
64% test-set accuracy

2-class model: Threatened or not threatened

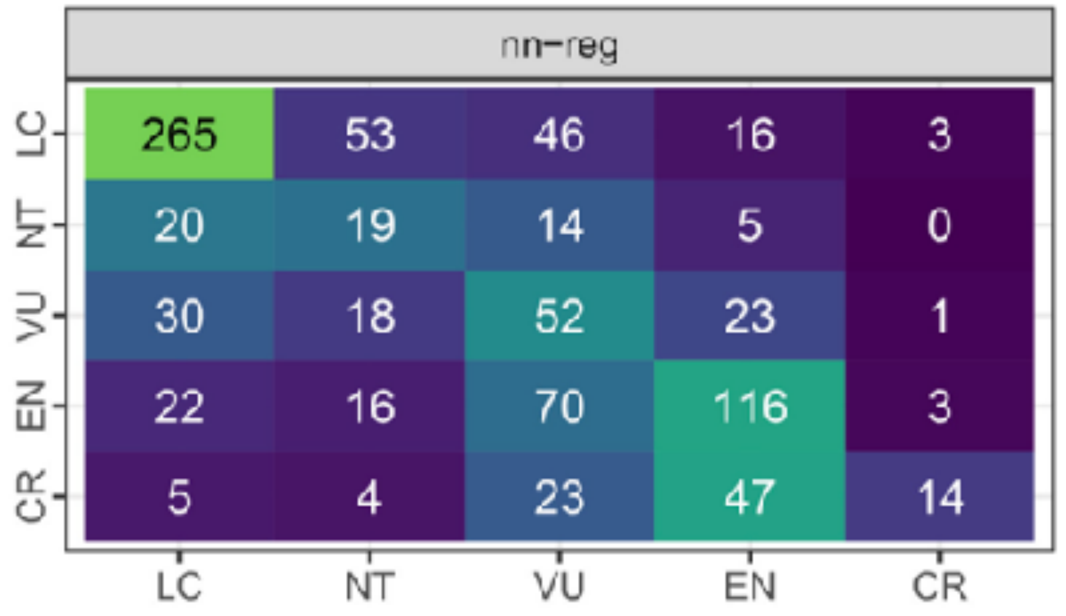


84% test-set accuracy

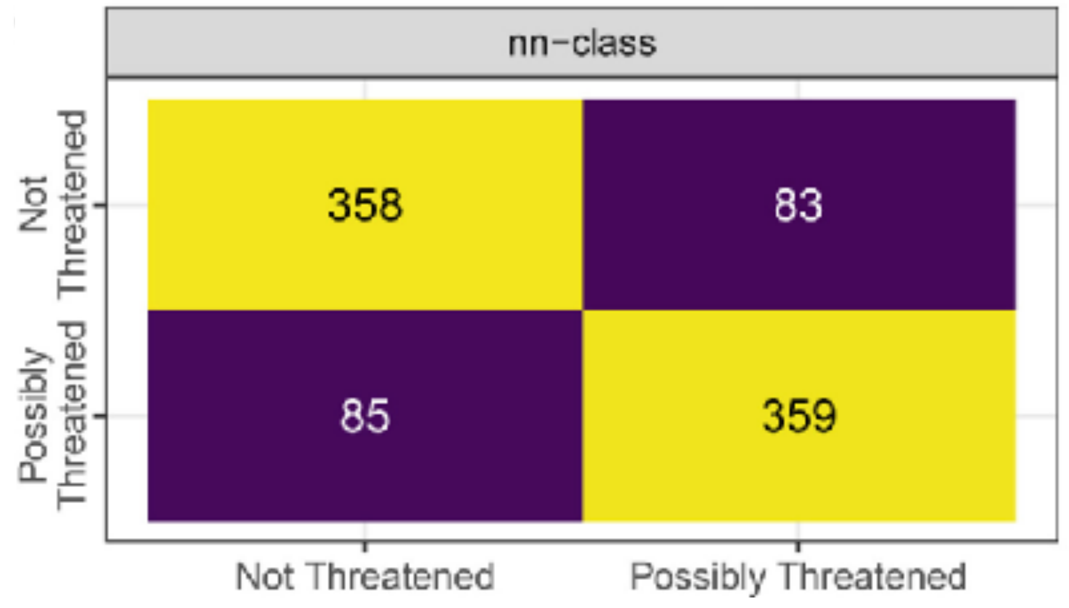
Classification



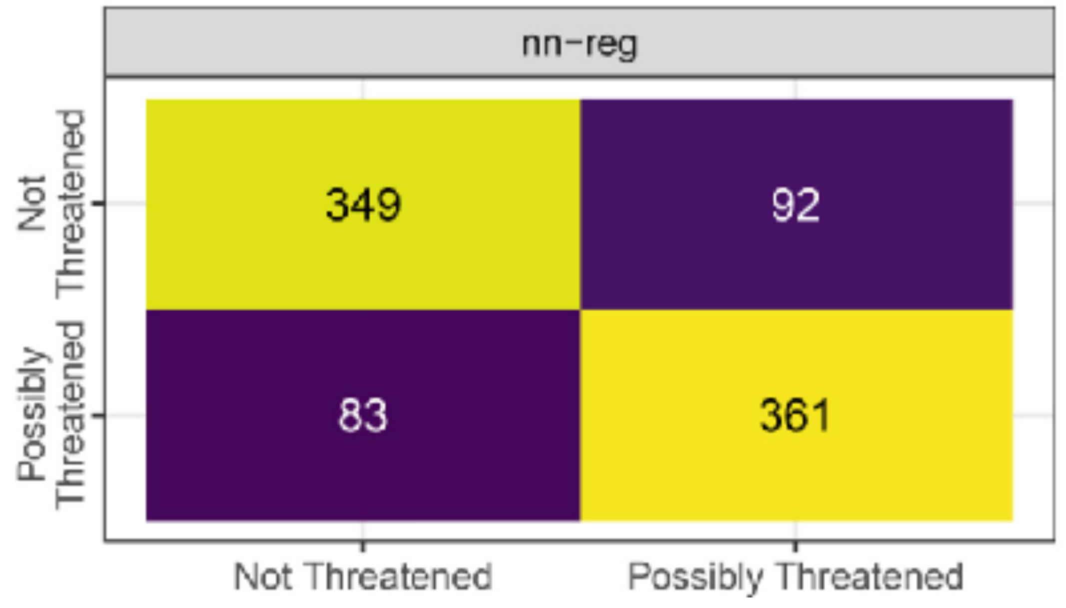
Regression



Classification



Regression



IUCNN R-package

repo status Active | DOI 10.5281/zenodo.5510580

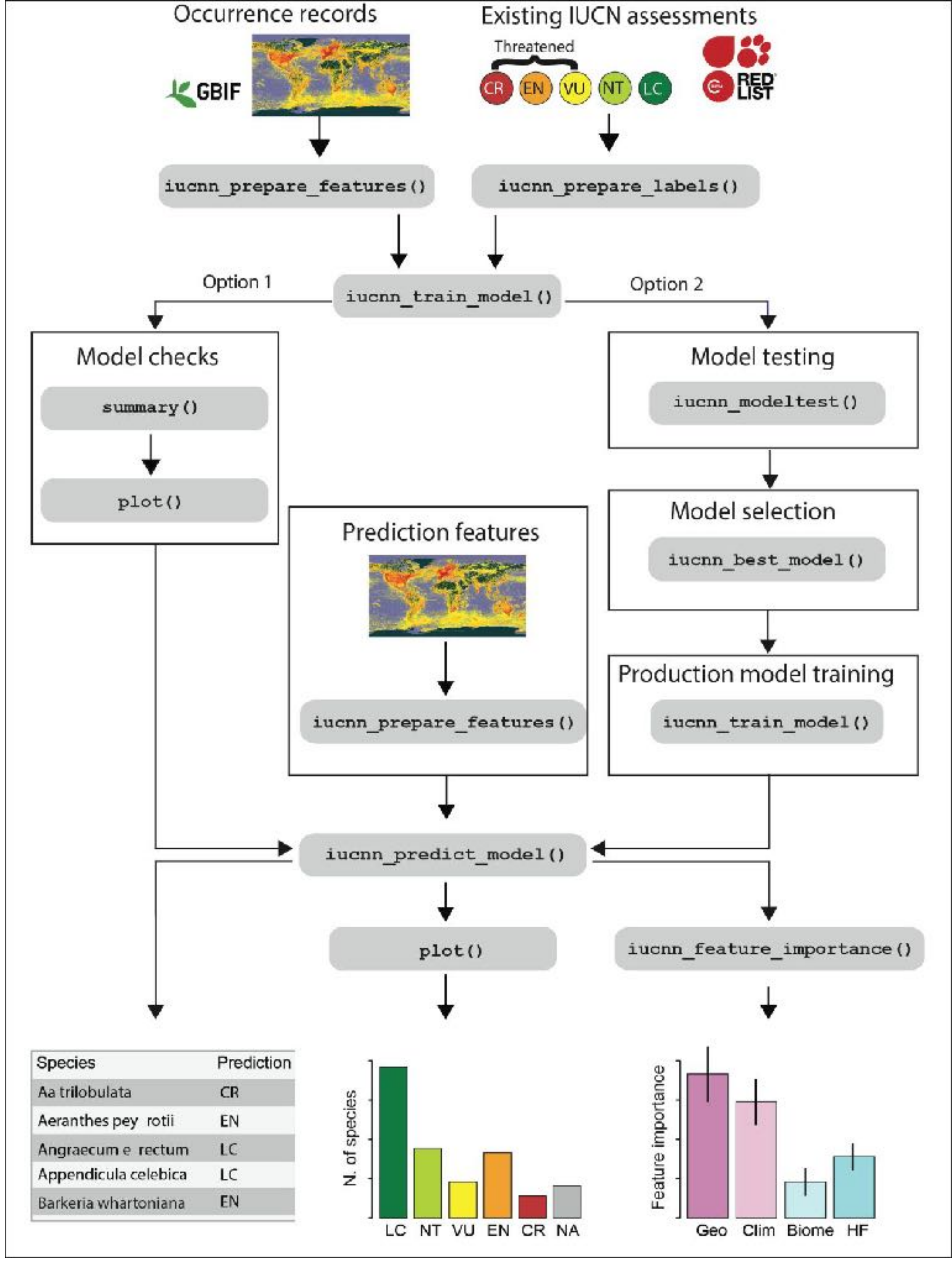
IUCNN

Batch estimation of species' IUCN Red List threat status using neural networks.

Installation

1. Install IUCNN directly from Github using devtools.

```
install.packages("devtools")
library(devtools)

install_github("IUCNN/IUCNN")
```



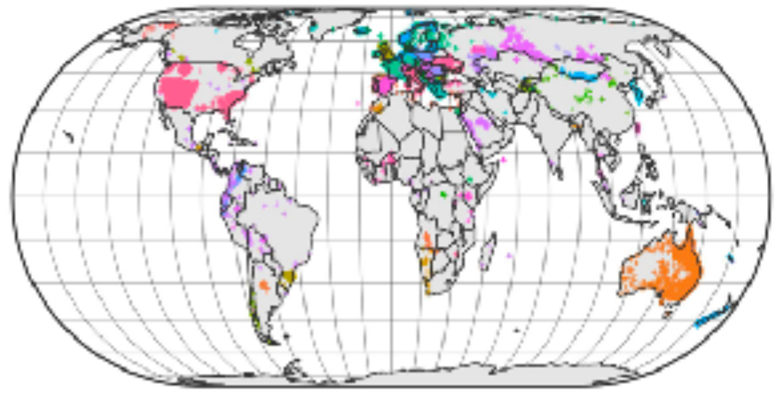
Zizka, A., Andermann, T., and Silvestro, D. IUCNN – Deep learning approaches to approximate species' extinction risk. *Diversity and Distributions* 28, 227–241. doi:10.1111/ddi.13450.

Modeling species richness

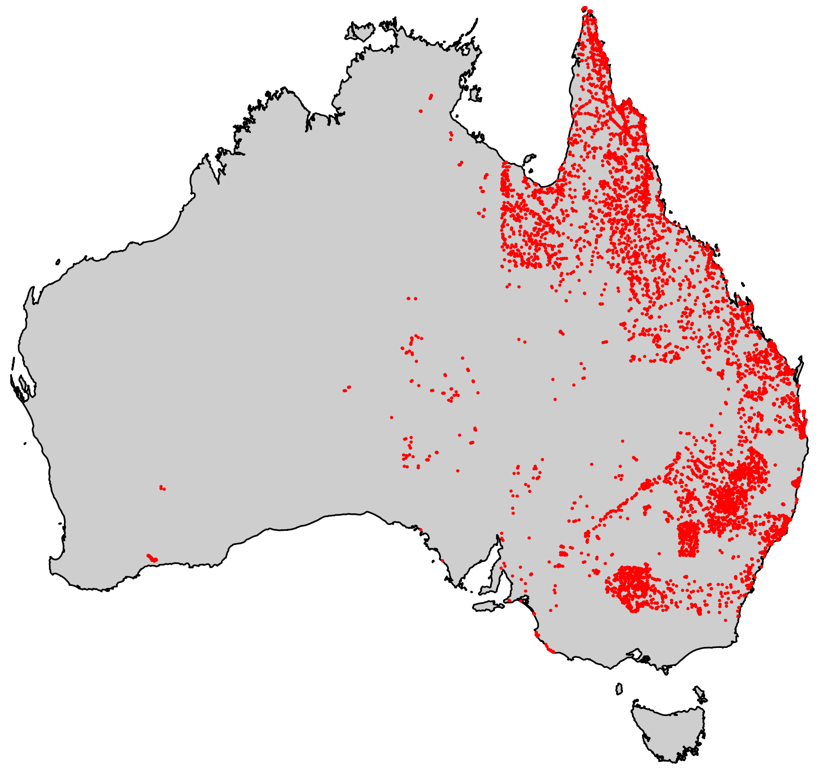
Modeling species richness

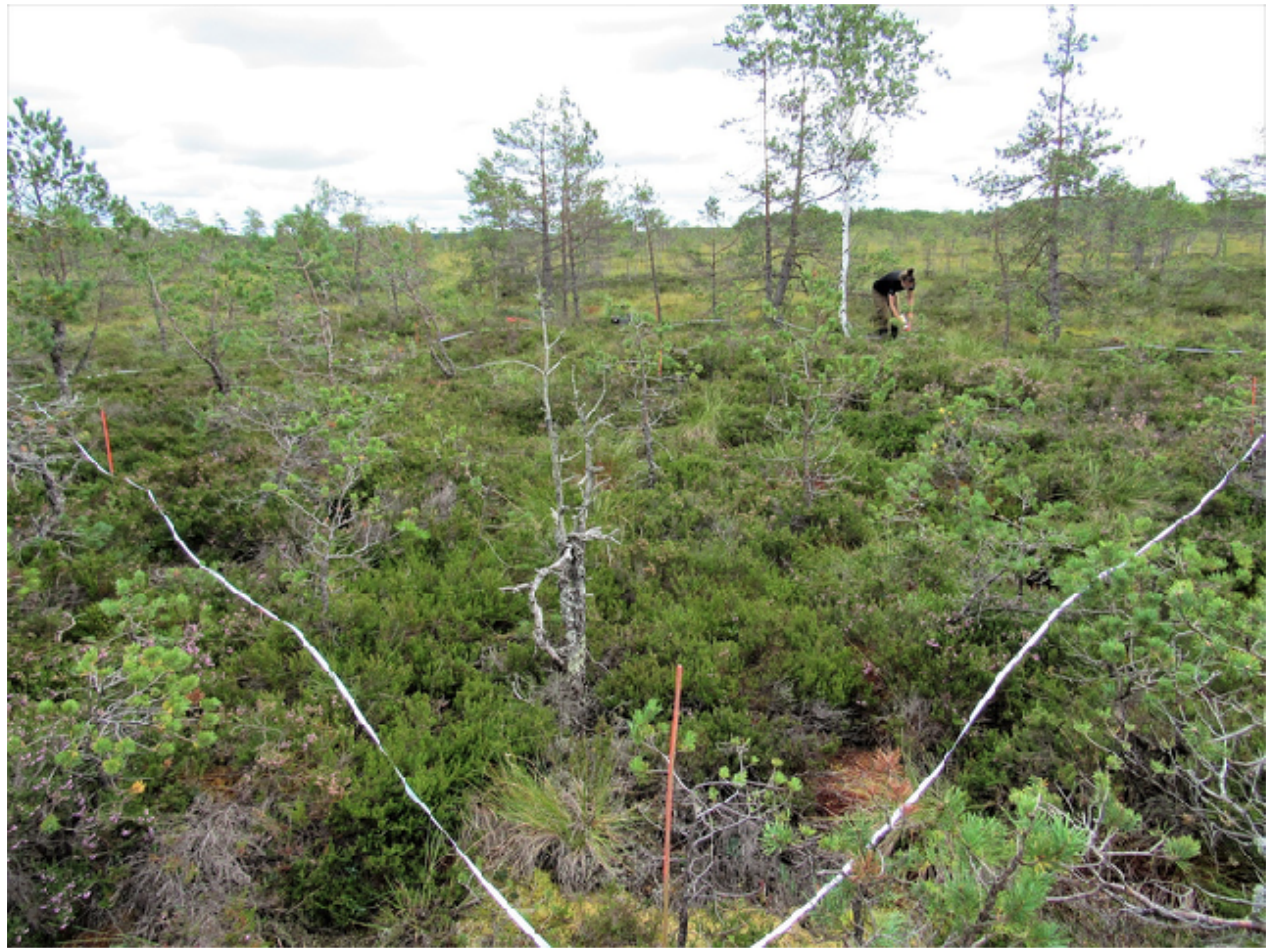
Extrapolating from point estimates

sPlot



- Train neural network (**machine learning**) with vegetation plot data (species inventory)
- Point measurements of plant species richness



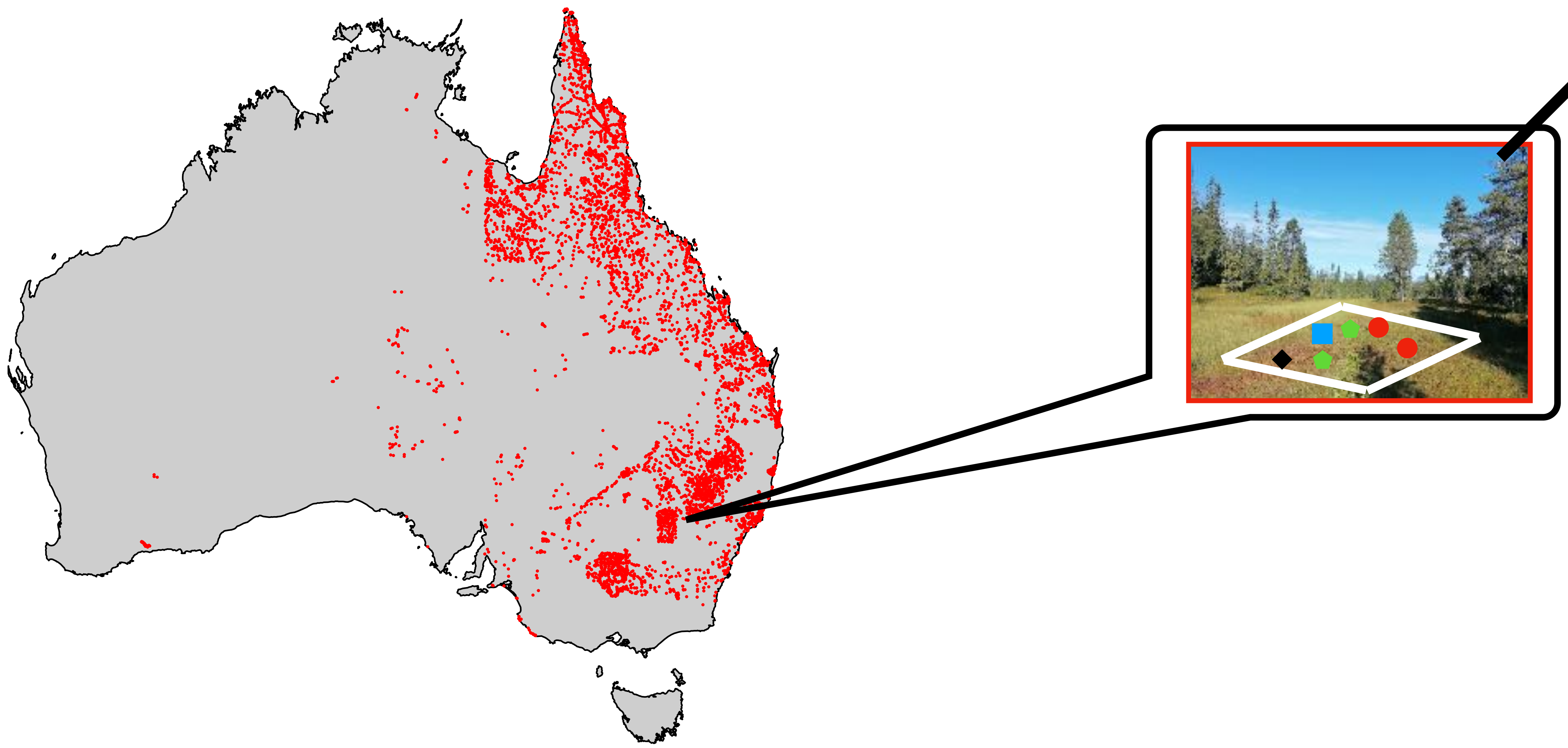


frontiers
in Plant Science

Estimating Alpha, Beta, and Gamma Diversity Through Deep Learning

Tobias Andermann[1,2,3,4*], Alexandre Antonelli[1,2,5,6], Russell L. Barrett[7,8] and Daniele Silvestro[1,2,3,4]

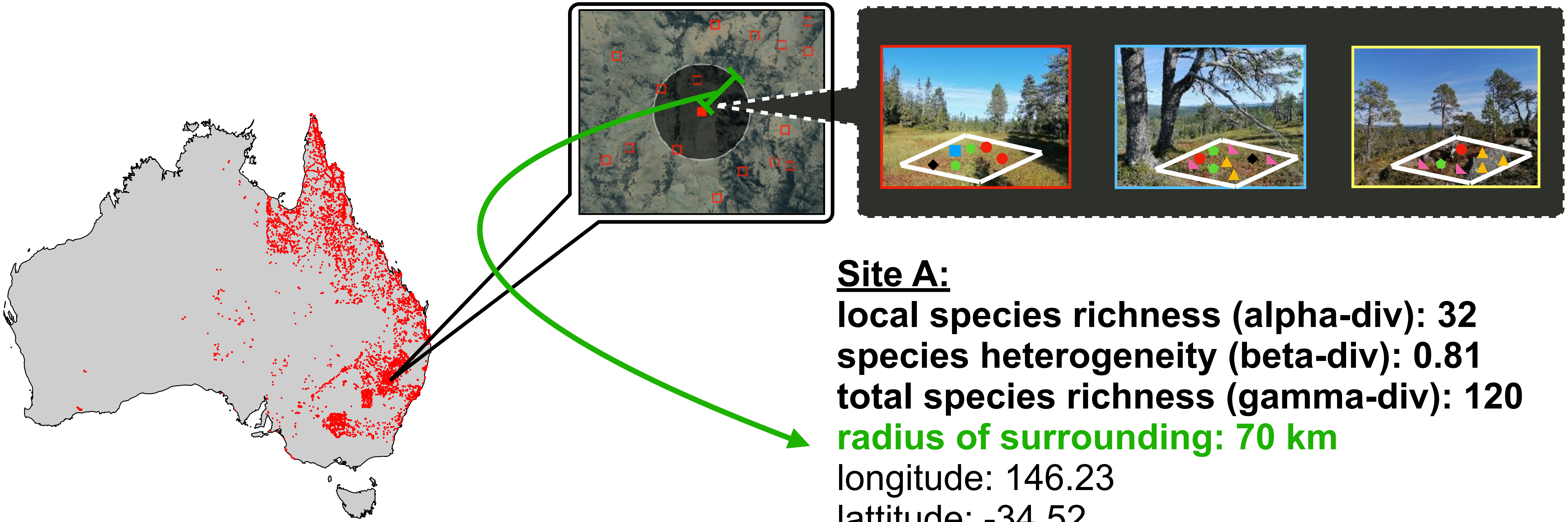
Preparing training data



<u>Site A:</u>
local species richness: 32
size of area: 20m²
longitude: 146.23
lattitude: -34.52
avg. temperature: 20
avg. rainfall: 500
human footprint: 0.94
soil category: 12
...

Andermann et al. (2022). ***Frontiers in Plant Science***

Preparing training data

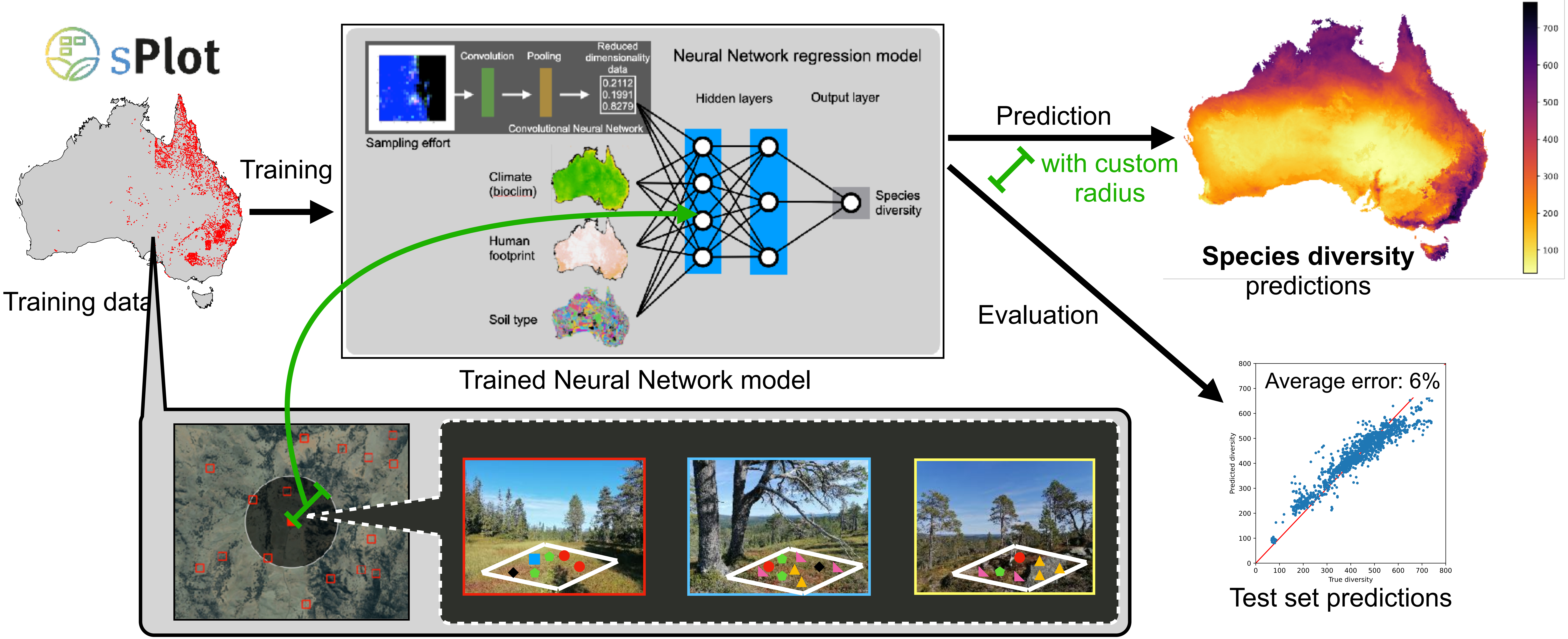


<u>Site A:</u>
local species richness (alpha-div): 32
species heterogeneity (beta-div): 0.81
total species richness (gamma-div): 120
radius of surrounding: 70 km
longitude: 146.23
lattitude: -34.52
avg. temperature: 20
avg. rainfall: 500
human footprint: 0.94
soil category: 12

Andermann et al. (2022). ***Frontiers in Plant Science***…

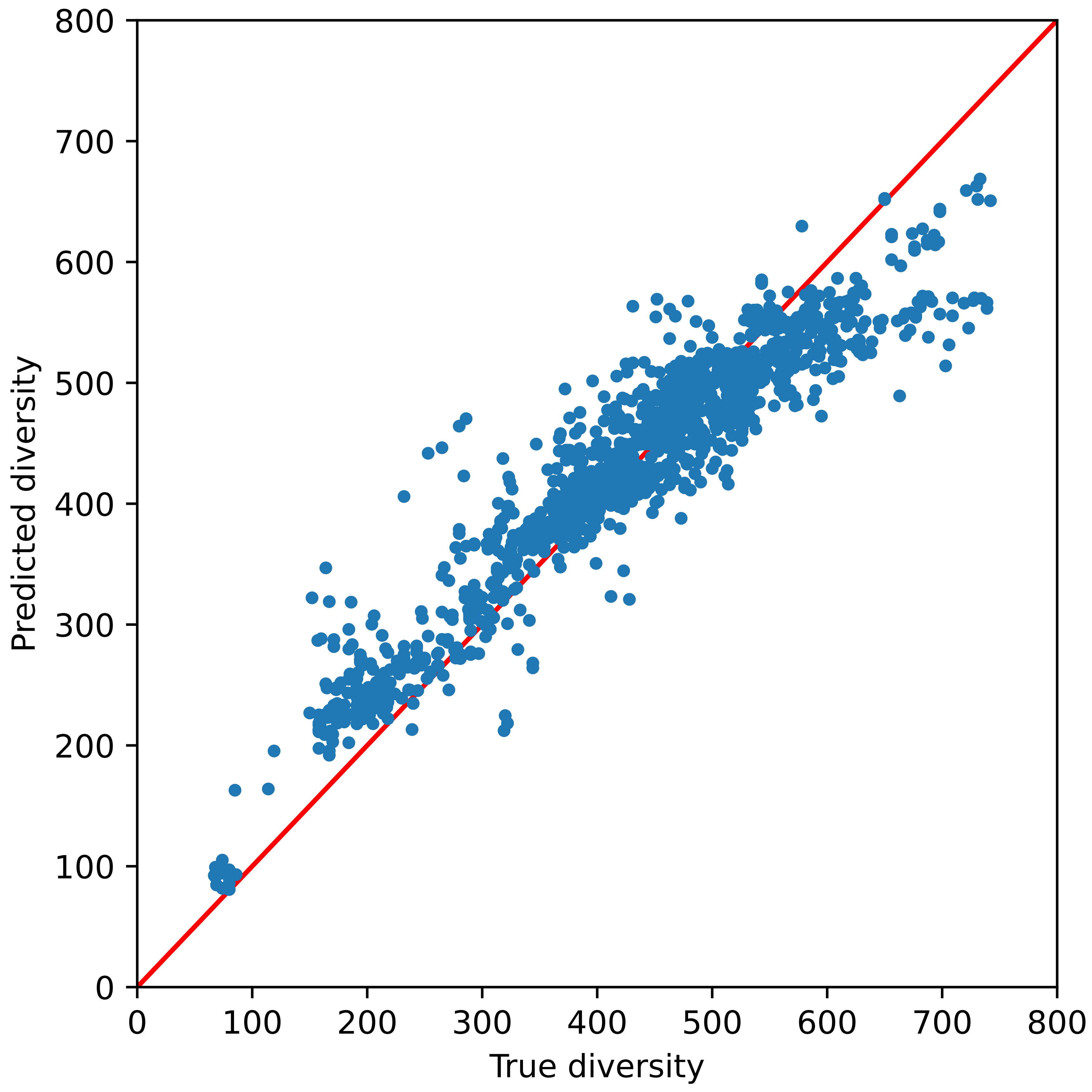
Modeling species richness

Extrapolating from point estimates

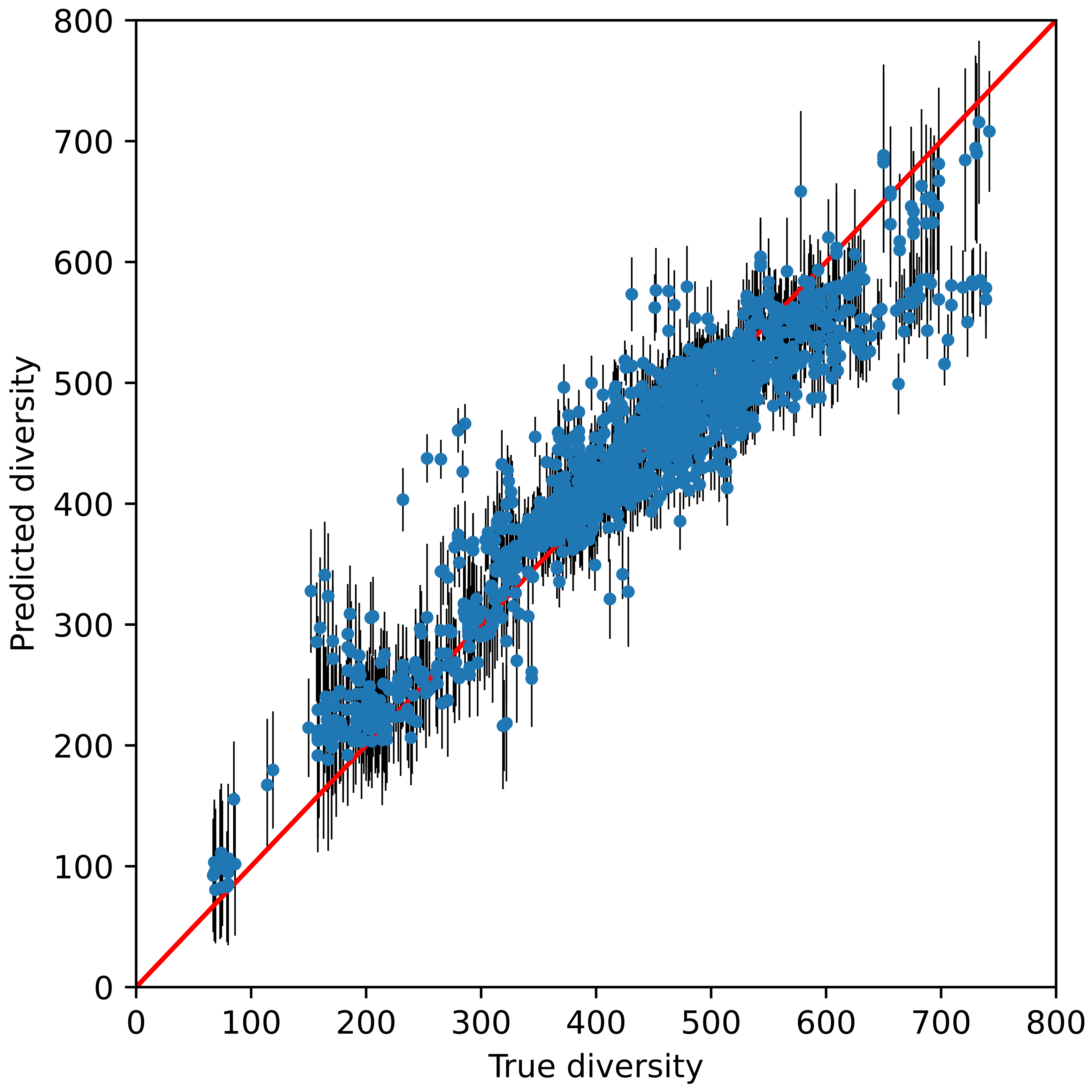


Andermann et al. (2022). ***Frontiers in Plant Science***

Quantifying uncertainty with MC dropout

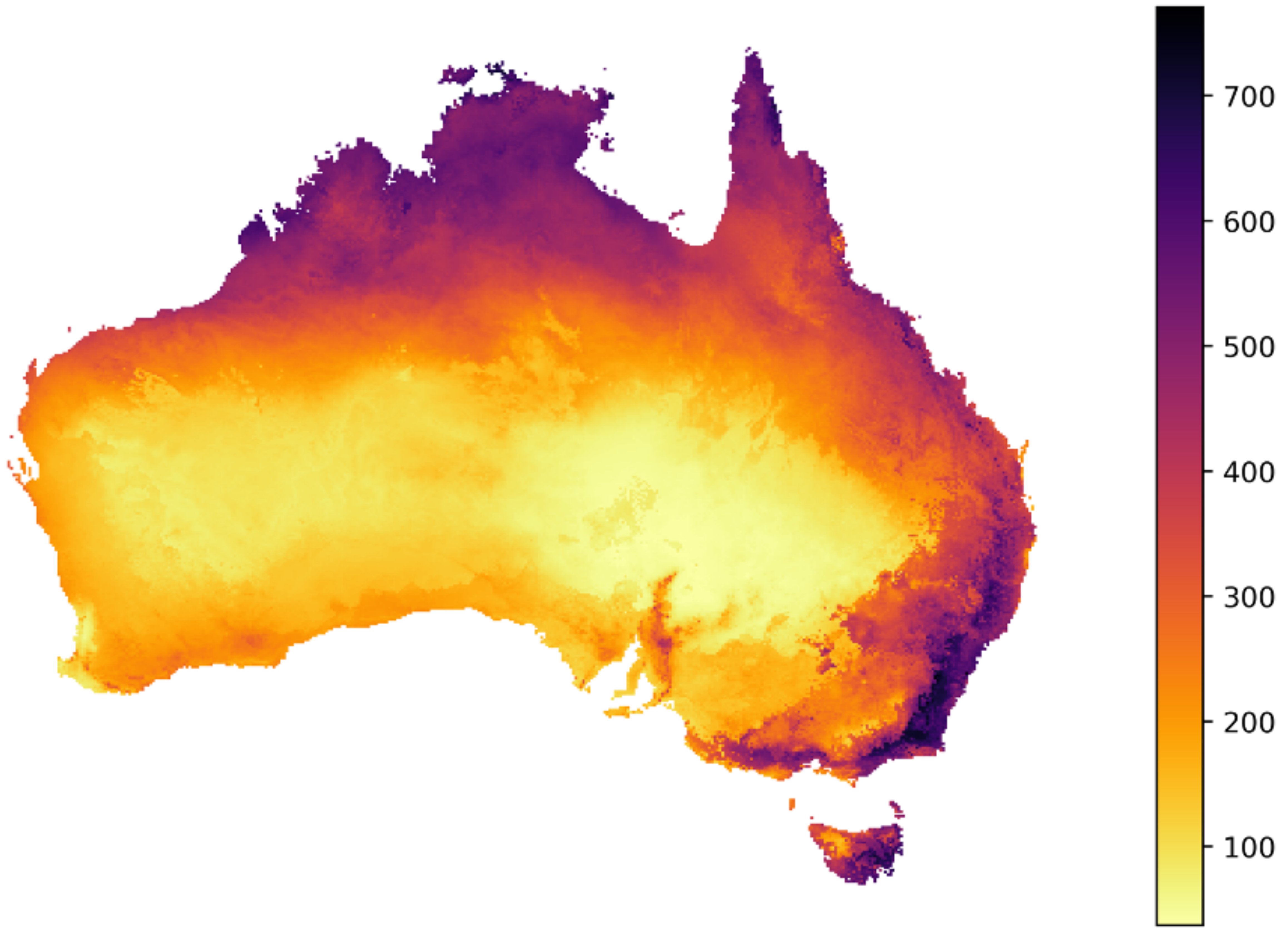


Test set predictions (single prediction)



MC dropout predictions (N=100)
estimating mean and standard deviation

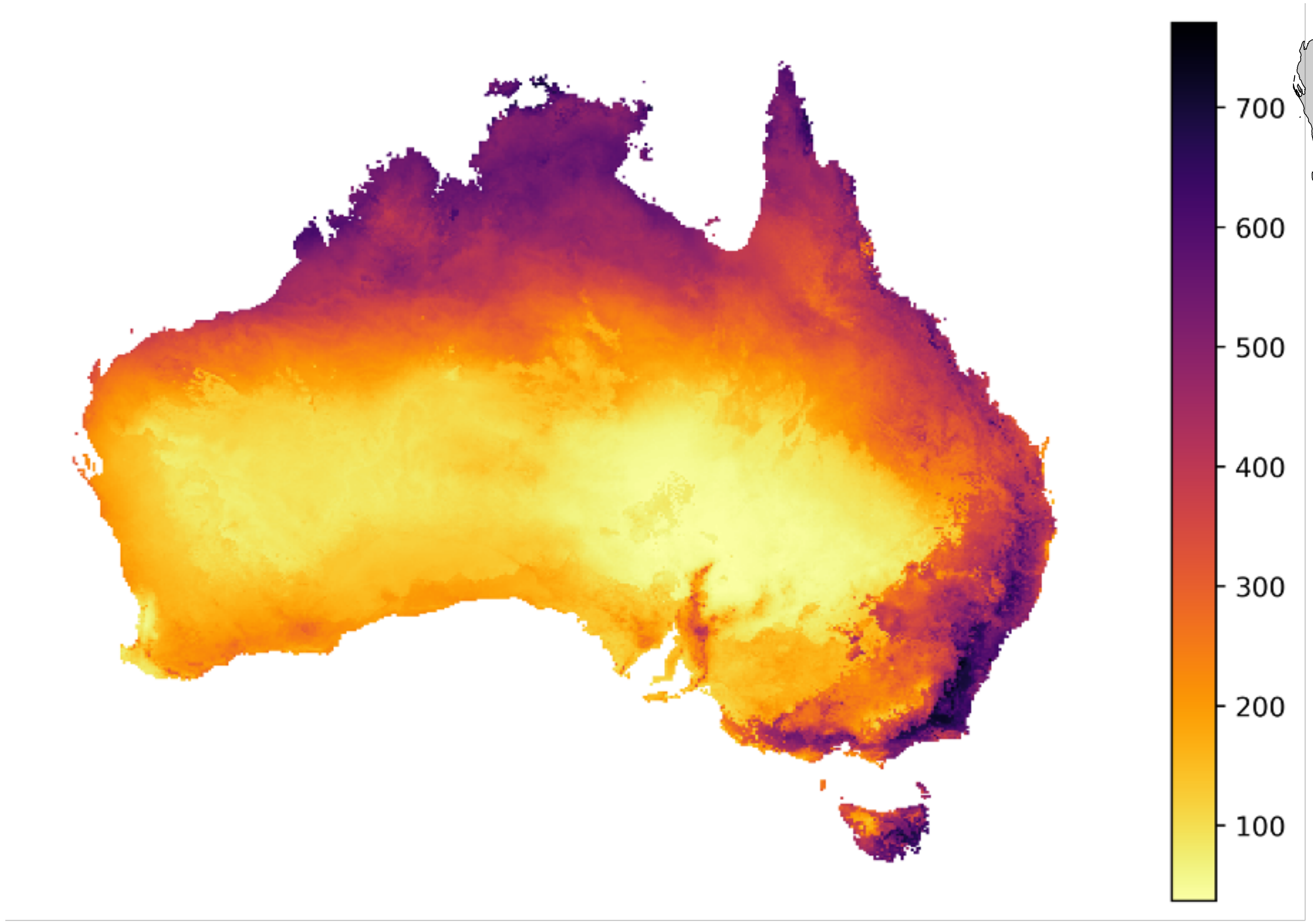
Quantifying uncertainty with MC dropout



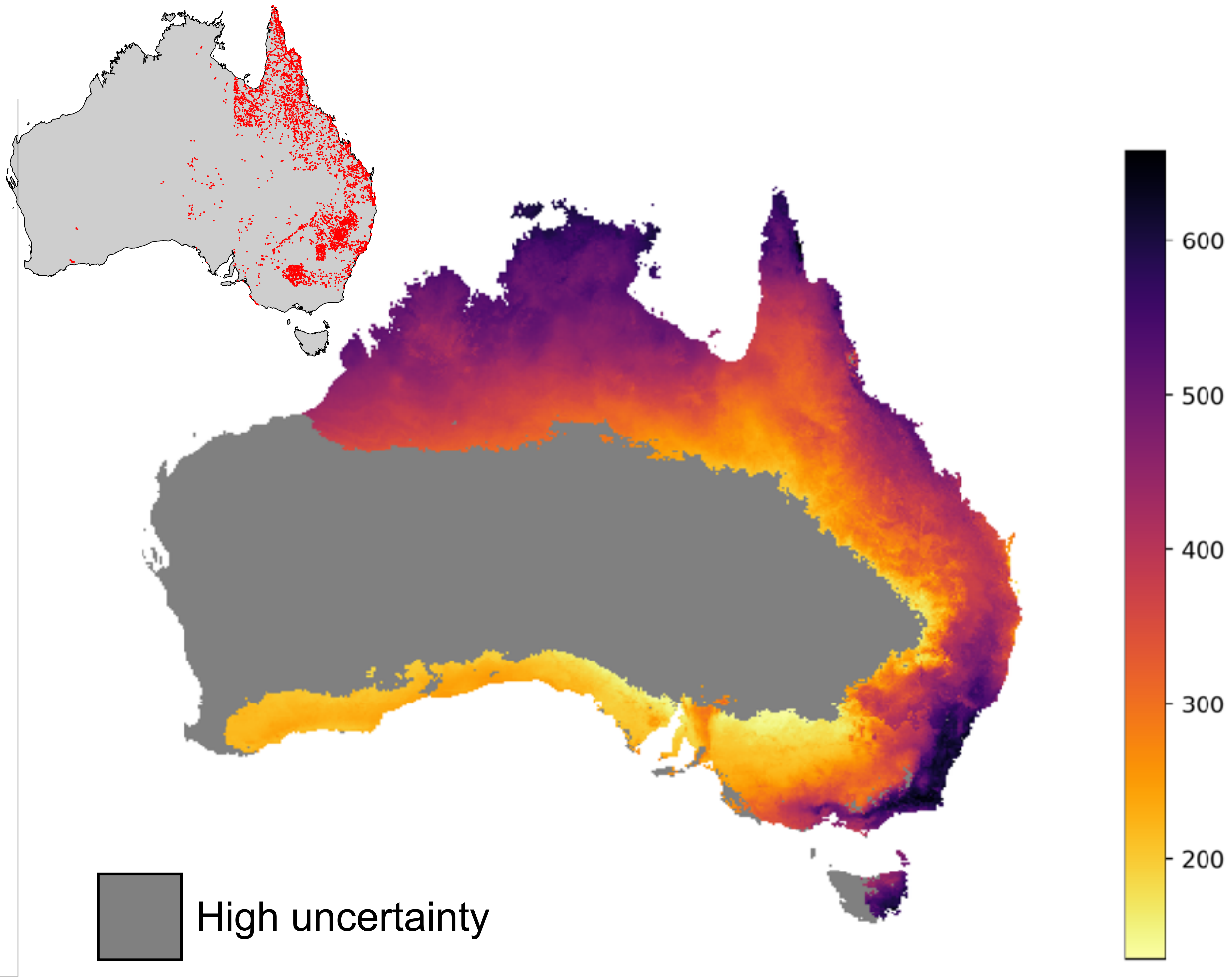
Point estimate (single prediction)

Standard deviation of predictions across **100 MC dropout predictions**

Quantifying uncertainty with MC dropout



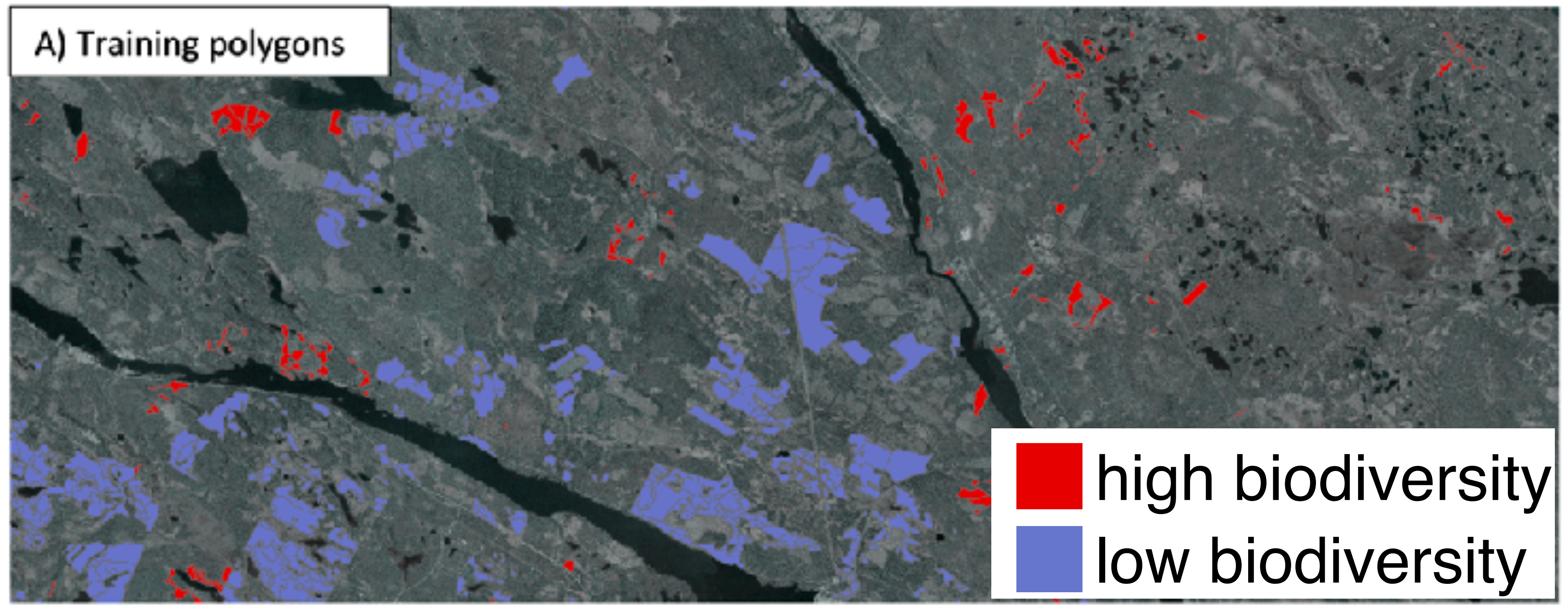
Point estimate (single prediction)



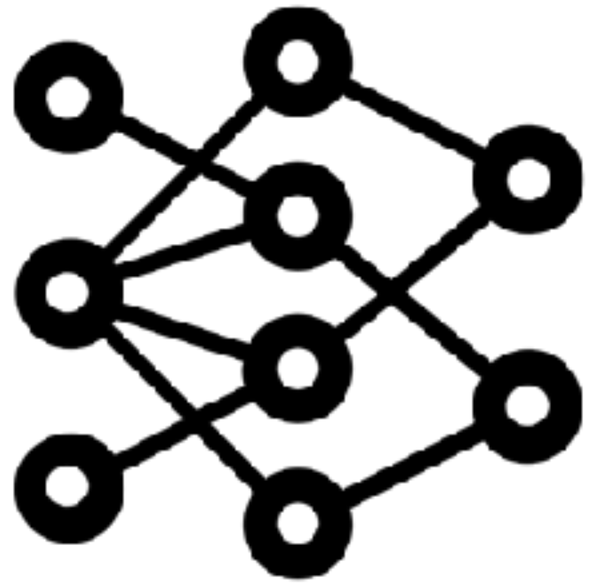
MC dropout predictions (N=100)
applying uncertainty cutoff (maximum standard deviation of 25%)

Resolving biodiversity from space

Resolving biodiversity from space



Annotated polygons/
biodiversity data

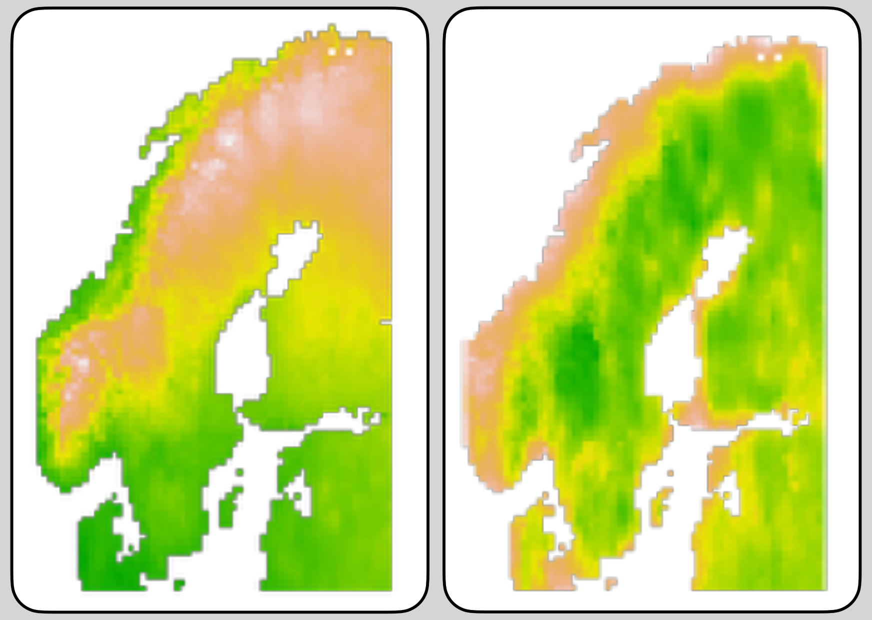


Convolutional Neural Network model

Multidimensional image data

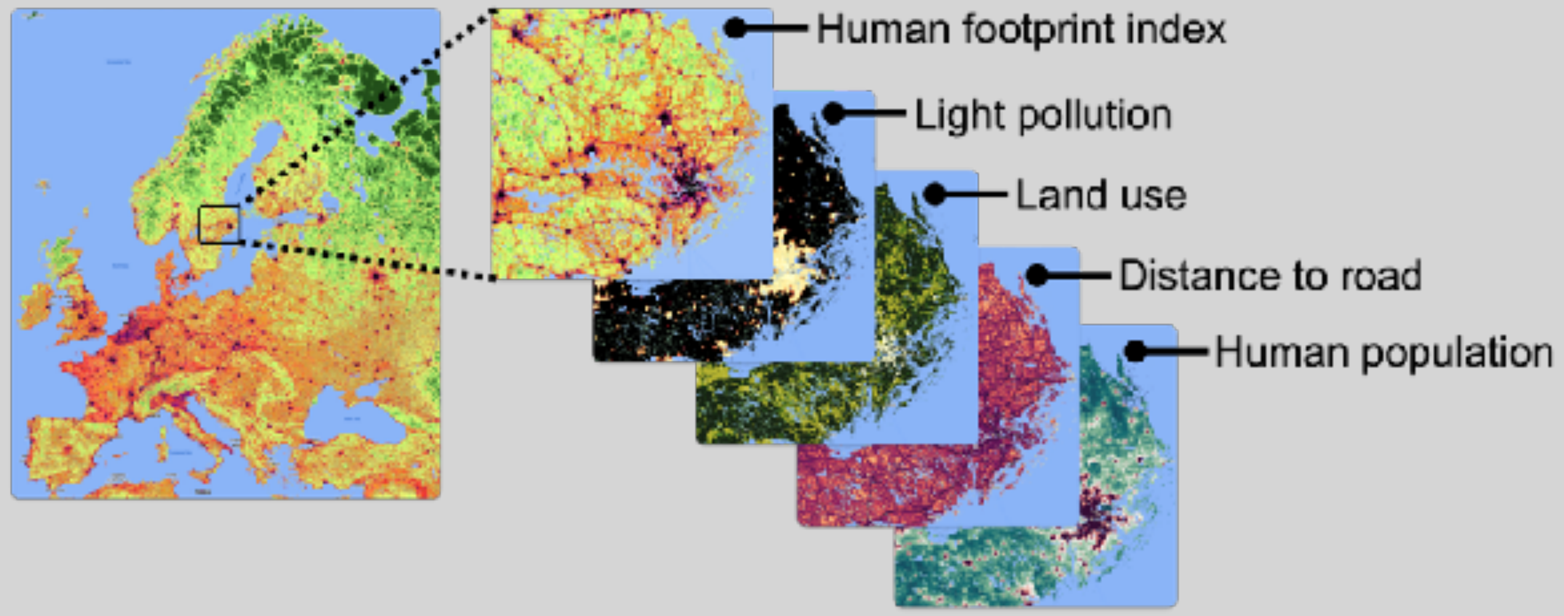
Environmental variables (*open access via API*)

Climate data



(available from WorldClim)

Human impact

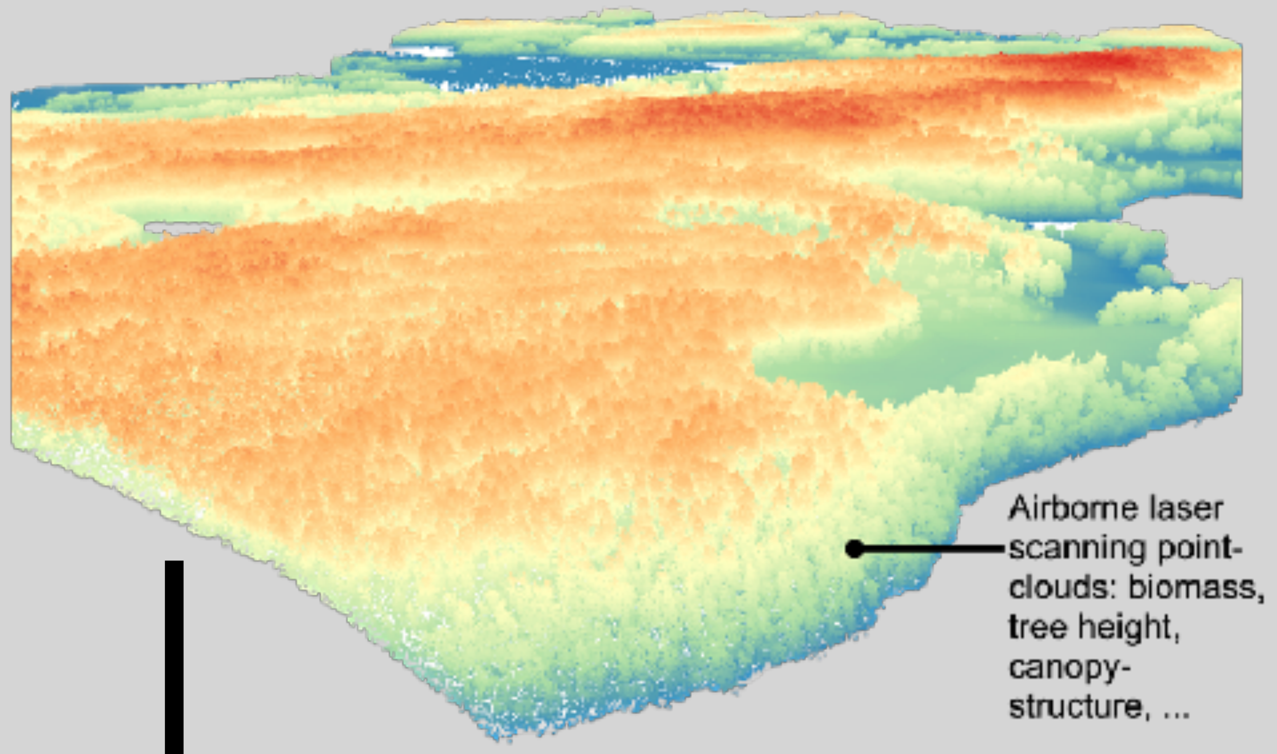


(available from Google Earth Engine)

Satellite and airborne photography

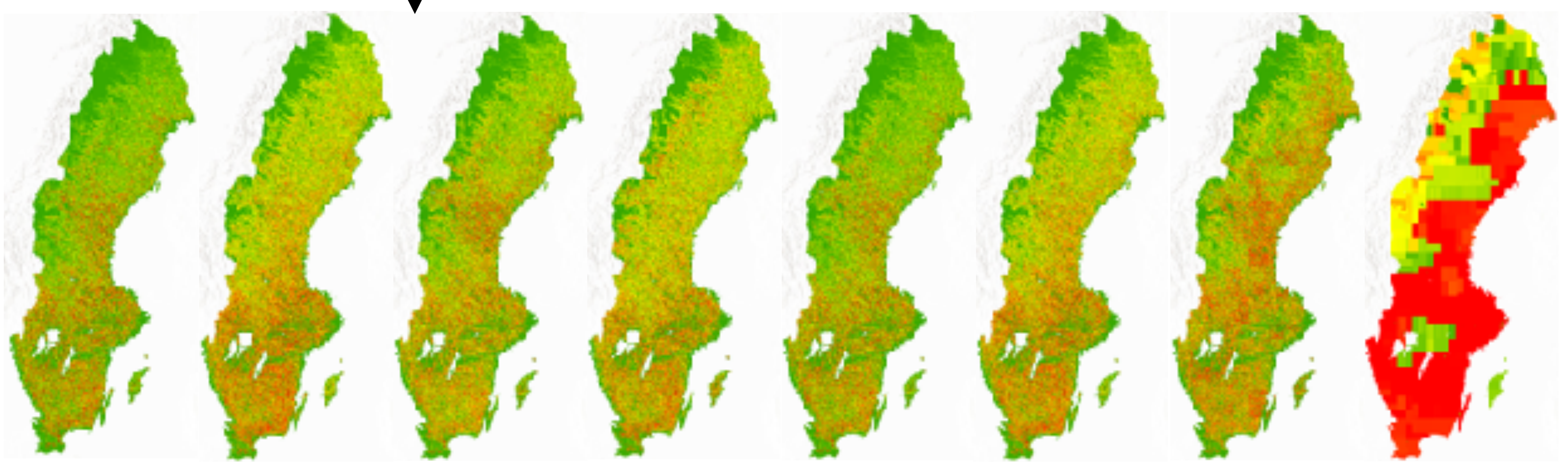
(available from Sentinel-2)

Biomass and Topographic data



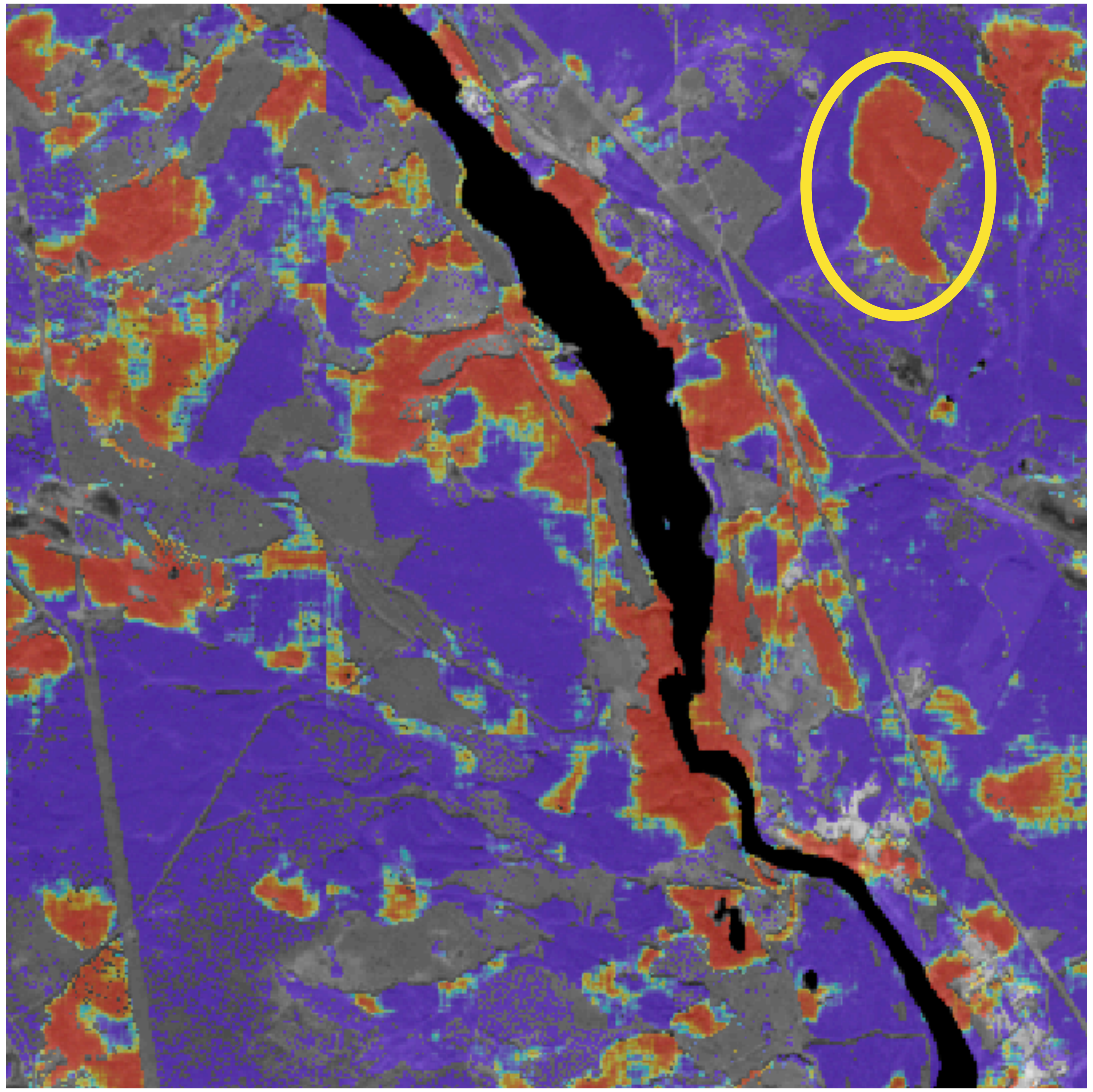
(available from Lantmäteriet)

+ many more ...



High resolution biodiversity AI models

Model predictions



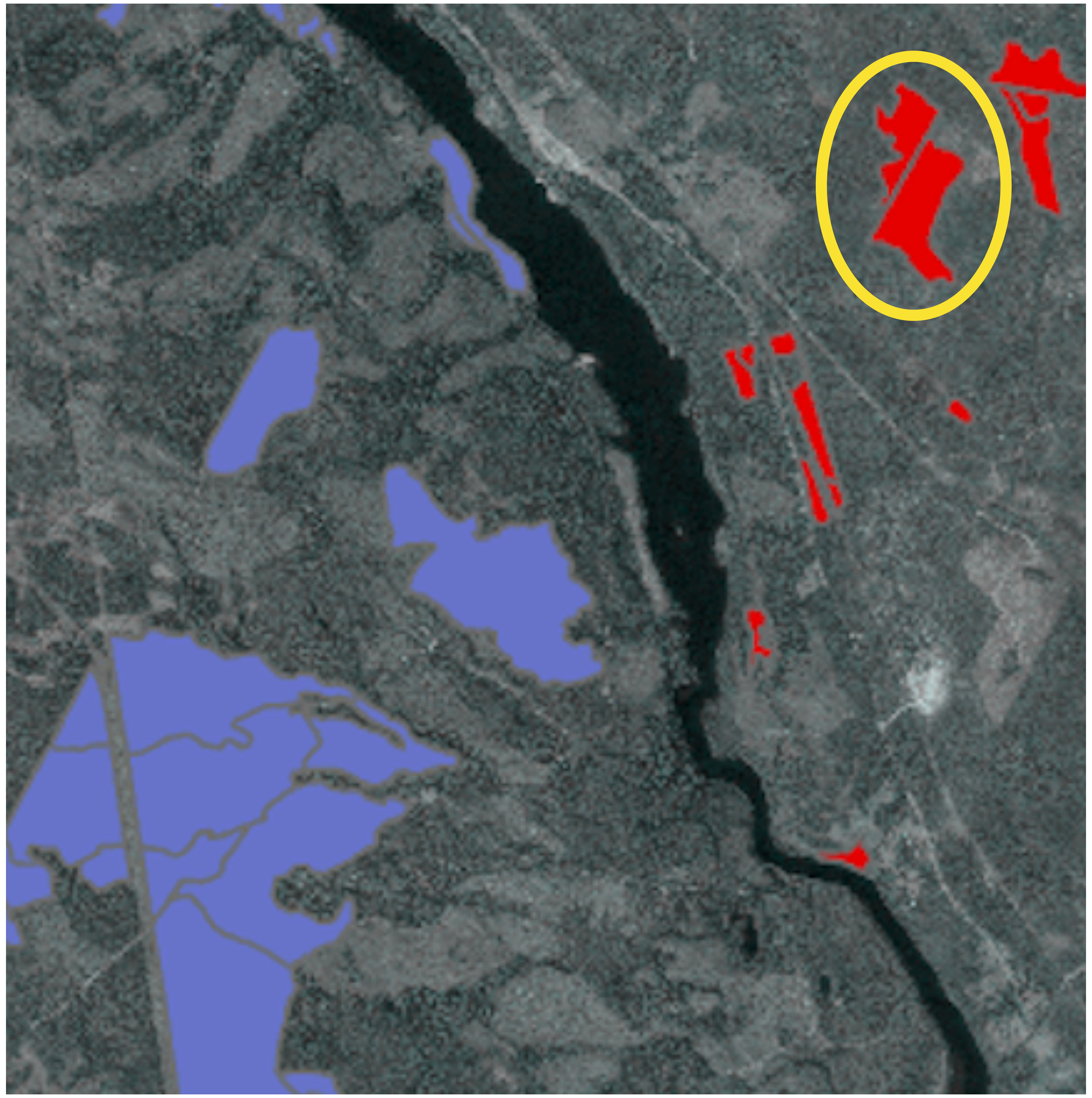
10m x 10m pixel resolution!

evaluate predictions

>93% *of high biodiversity predictions were correct* ***(precision)***

>90% *of high biodiversity pixels were retrieved* ***(recall)***

Test set of annotated forest polygons



Andermann et al., *in prep.*

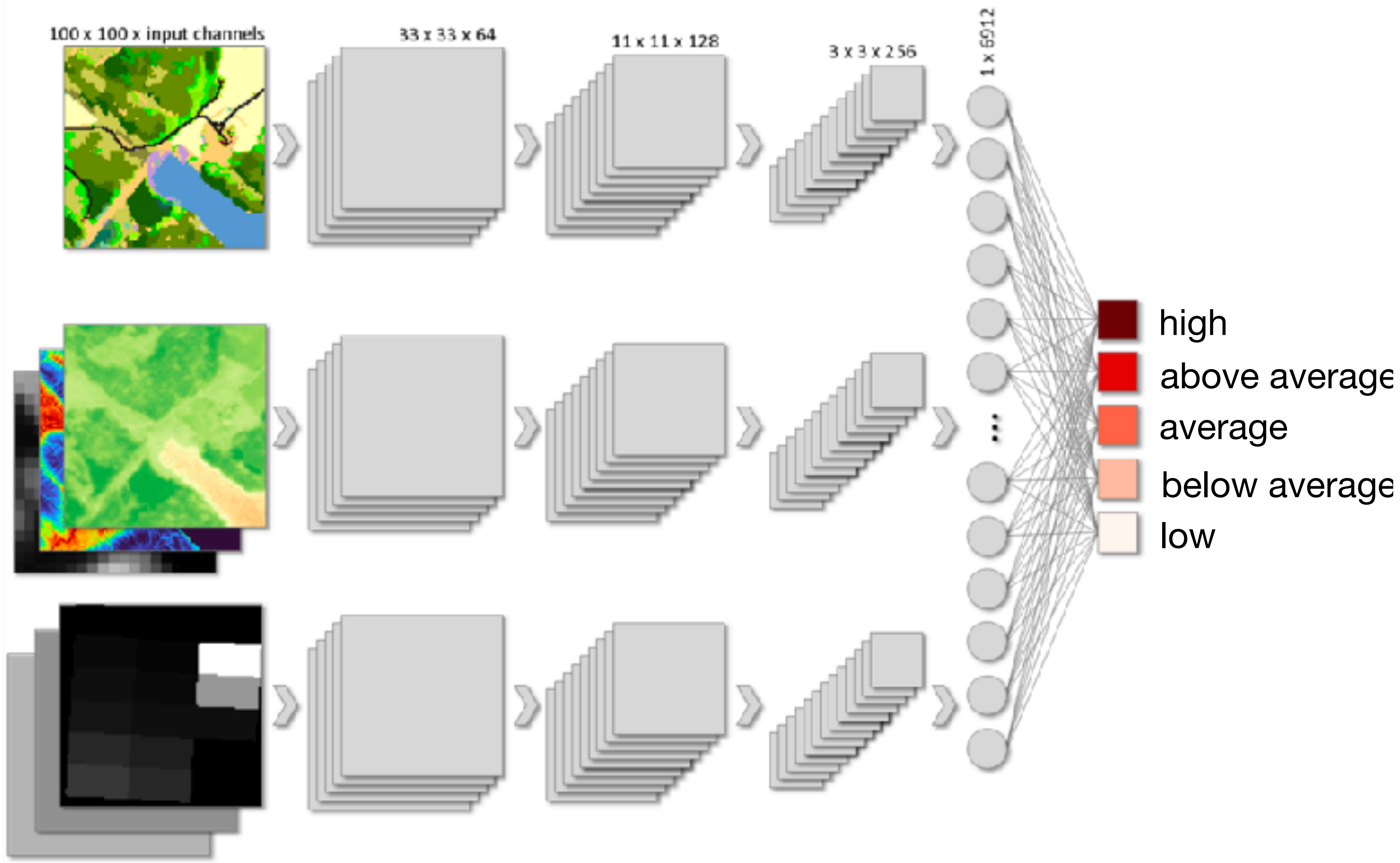
In collaboration with

SKOGSSTYRELSEN

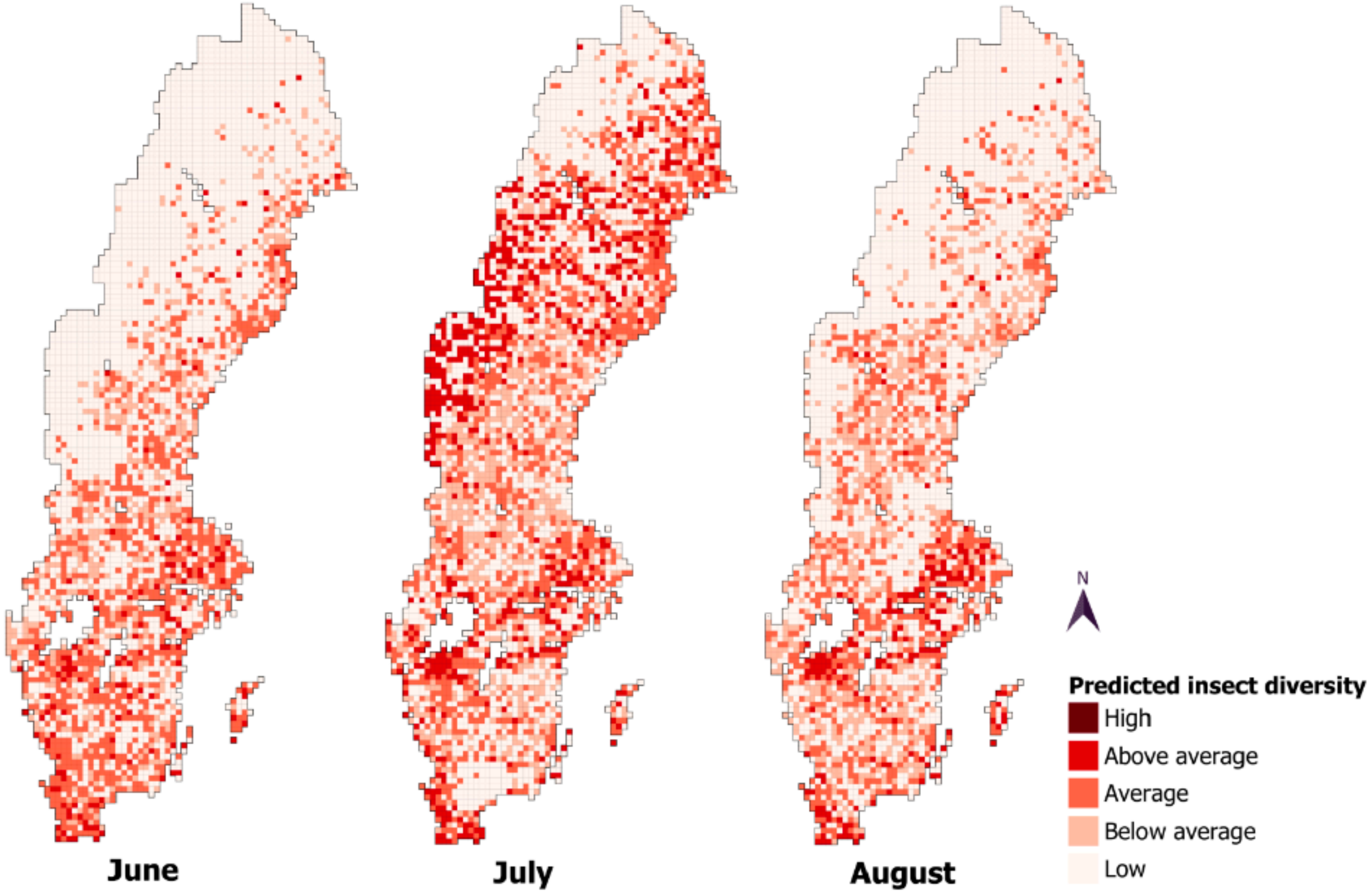
AI models trained on eDNA data

Adrian Baggström

- Neural Network models trained on insect diversity from environmental DNA (eDNA) data
- Collaboration with Insect Biome Atlas, which collected eDNA data over several years on weekly basis



Predicted insect diversity



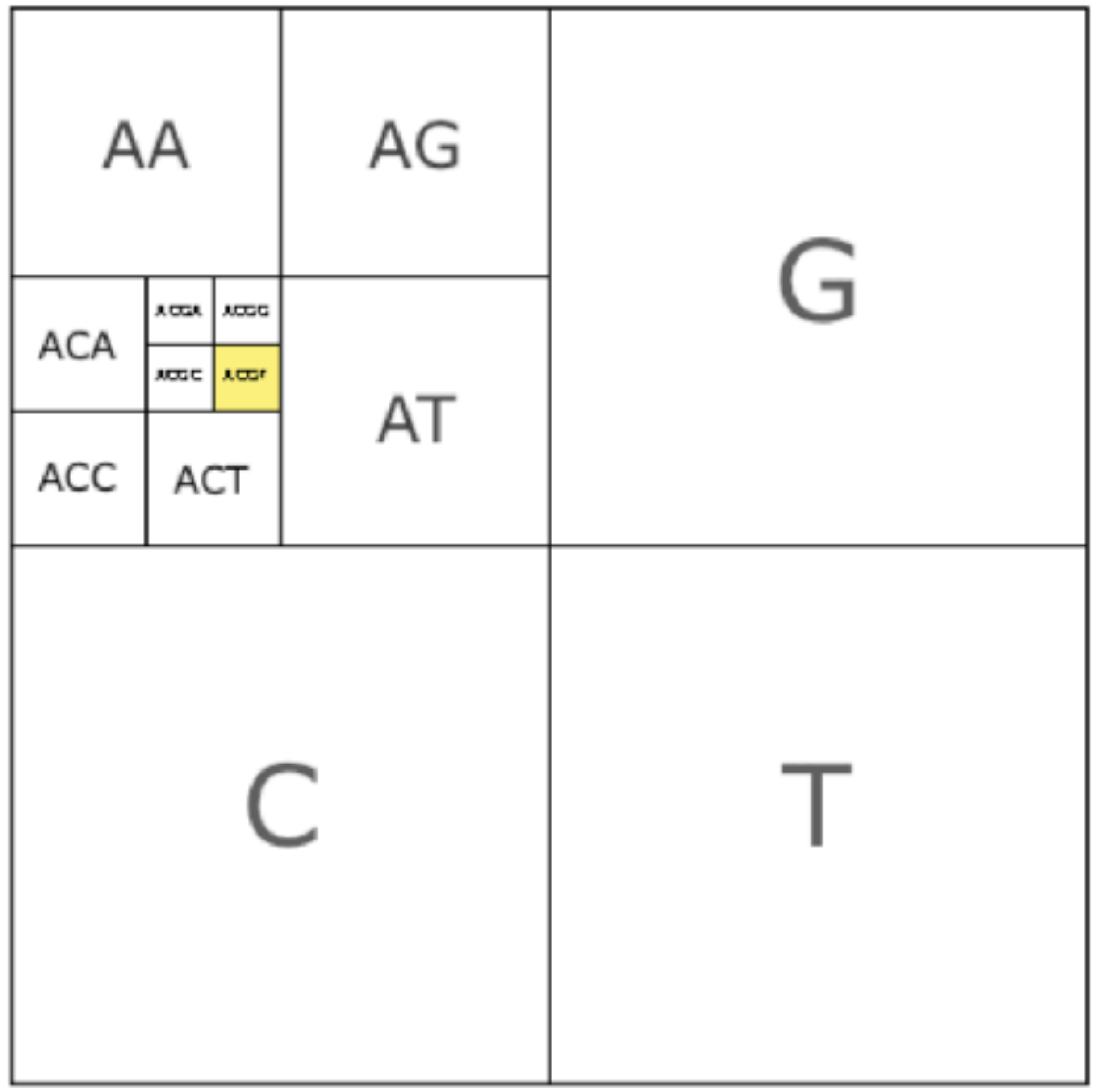
Baggström and Andermann, *in prep.*


"And now for something completely different."
Monty Python

Chaos game representation of DNA sequences

ACGTGGTGTAGATTTGGGATTGACCTACTAGTACGATTCA ...

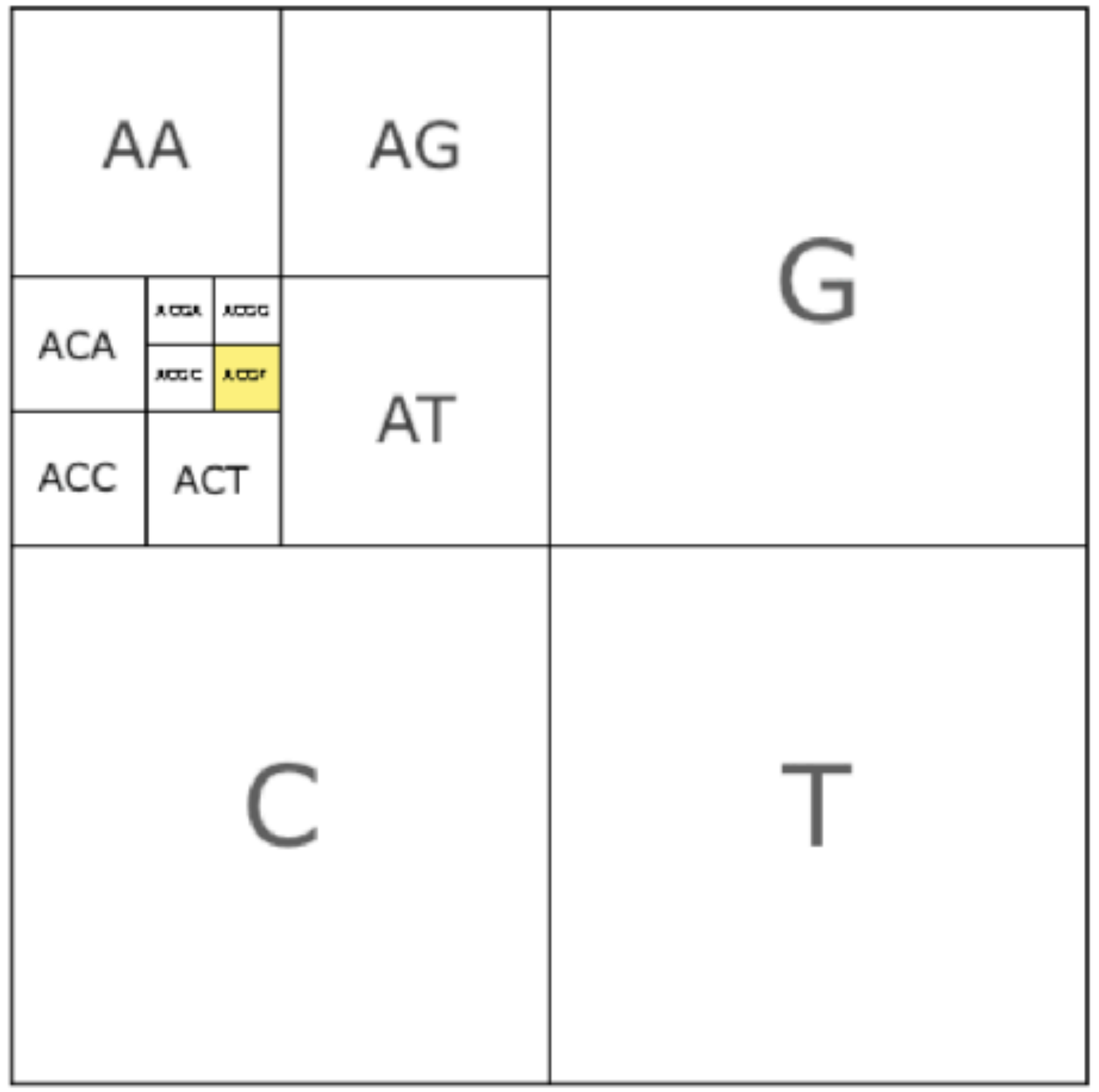
- Decide on kmer length, e.g. kmer=4



Chaos game representation of DNA sequences

ACGT GGTG TAGA TTTG GGAT TGAC CTAC TAGT ACGA TTCA ...

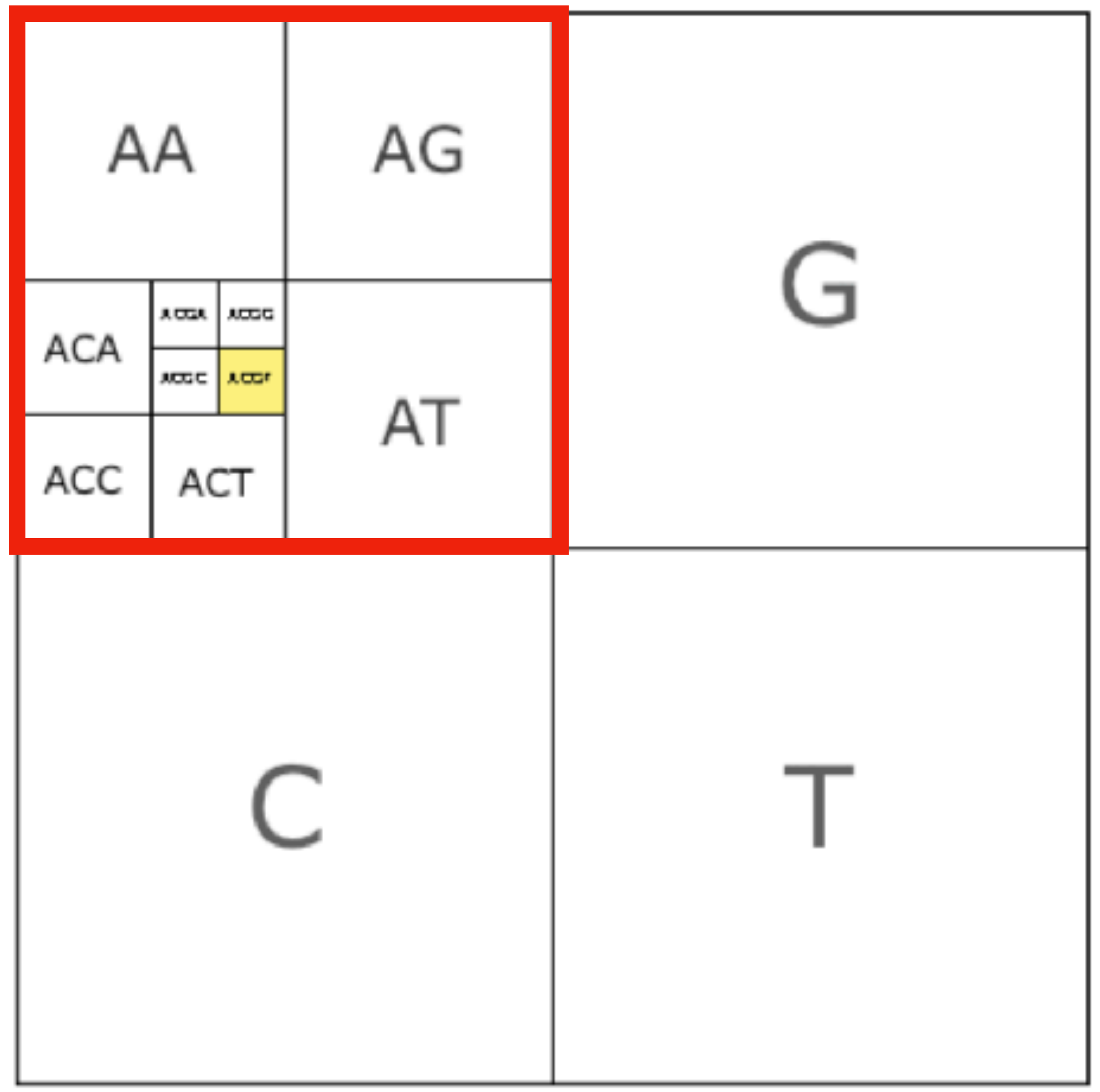
- Decide on kmer length, e.g. kmer=4
- Identify correct pixel, using chaos game scheme
- Count number of occurrences of each pixel across whole sequence



Chaos game representation of DNA sequences

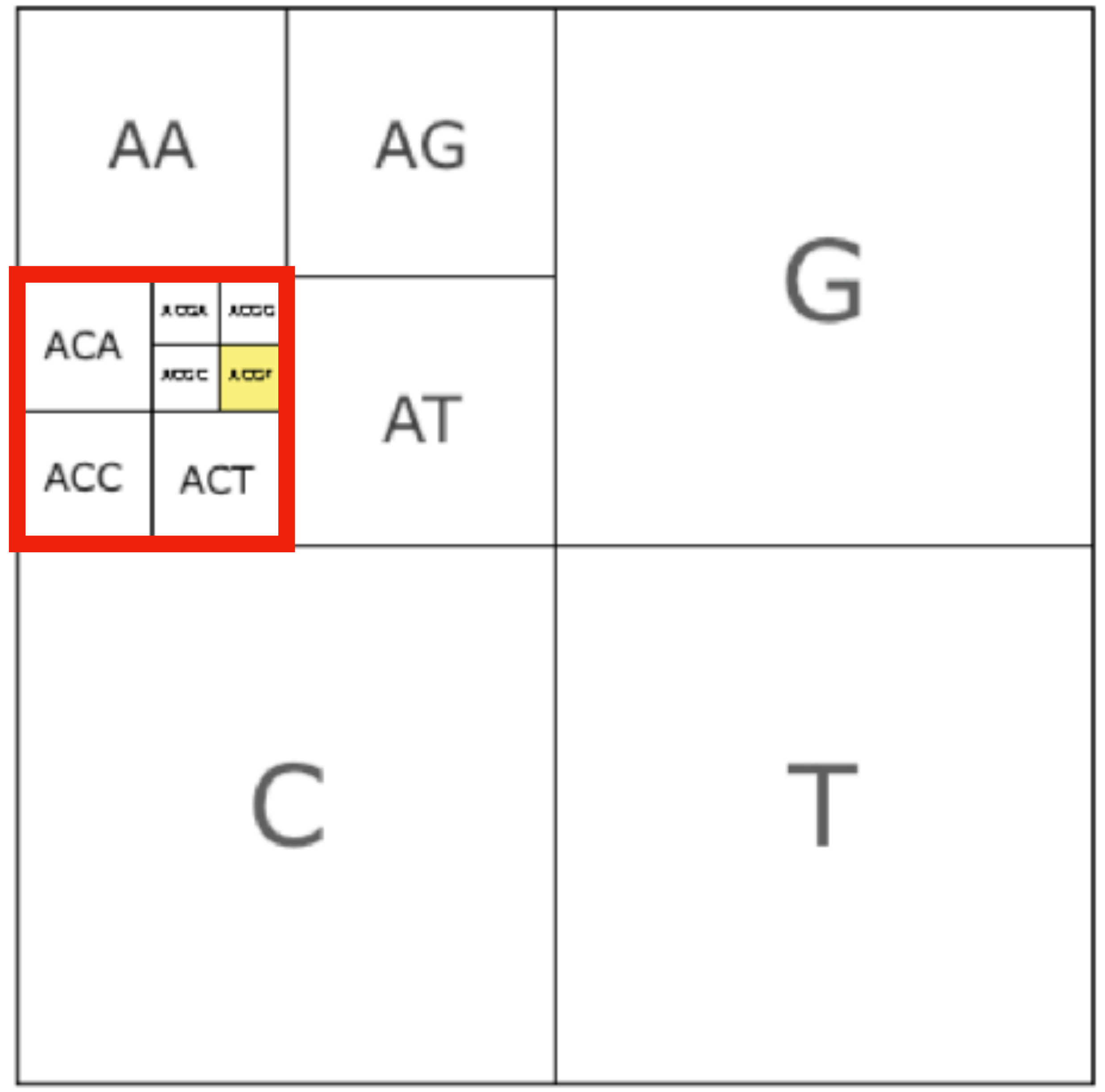
ACGT GGTG TAGA TTTG GGAT TGAC CTAC TAGT ACGA TTCA ...

- Decide on kmer length, e.g. kmer=4
- Identify correct pixel, using chaos game scheme
- Count number of occurrences of each pixel across whole sequence



Chaos game representation of DNA sequences

ACGT GGTG TAGA TTTG GGAT TGAC CTAC TAGT ACGA TTCA ...

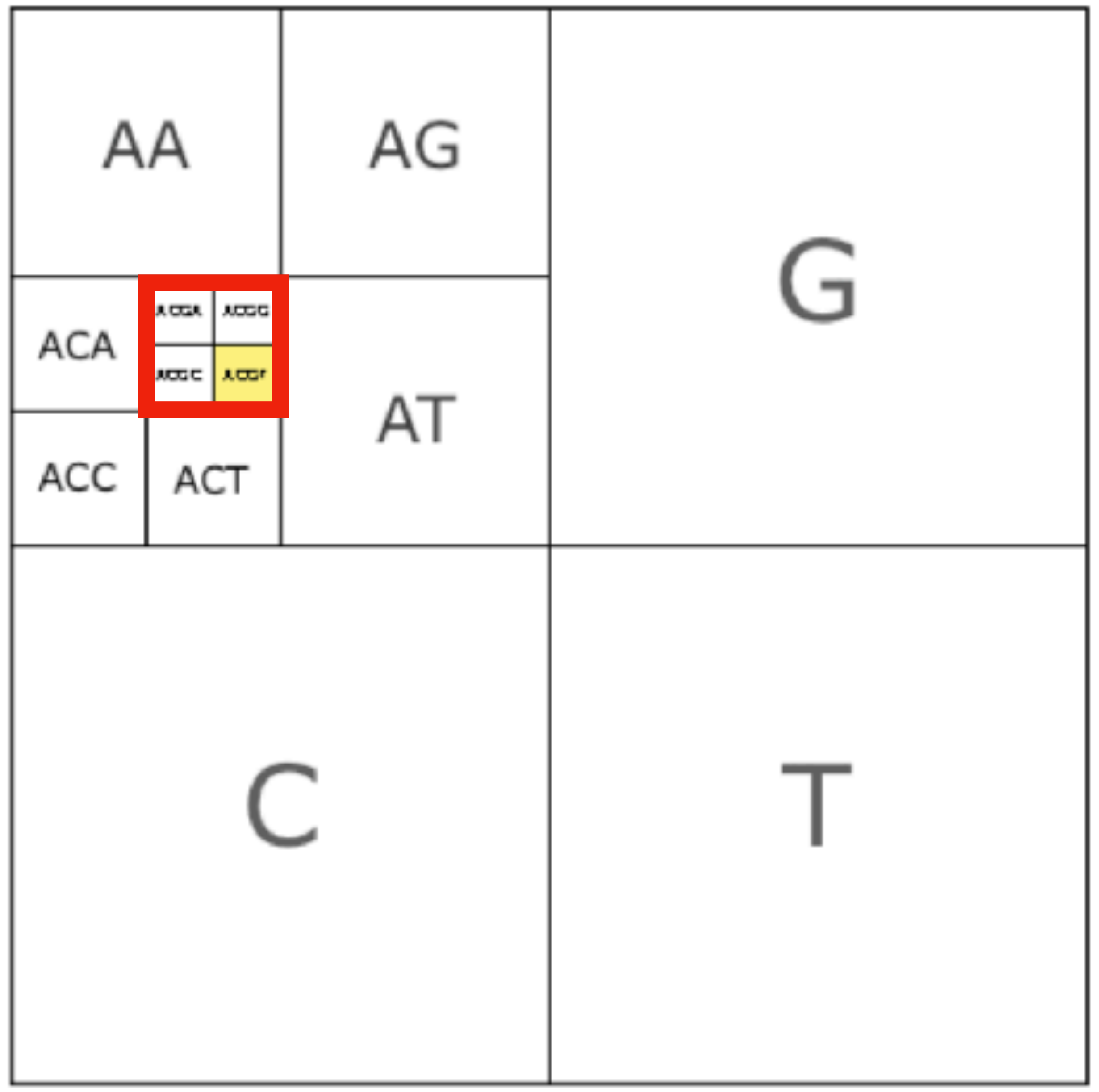


- Decide on kmer length, e.g. kmer=4
- Identify correct pixel, using chaos game scheme
- Count number of occurrences of each pixel across whole sequence

Chaos game representation of DNA sequences

ACGT GGTG TAGA TTTG GGAT TGAC CTAC TAGT ACGA TTCA ...

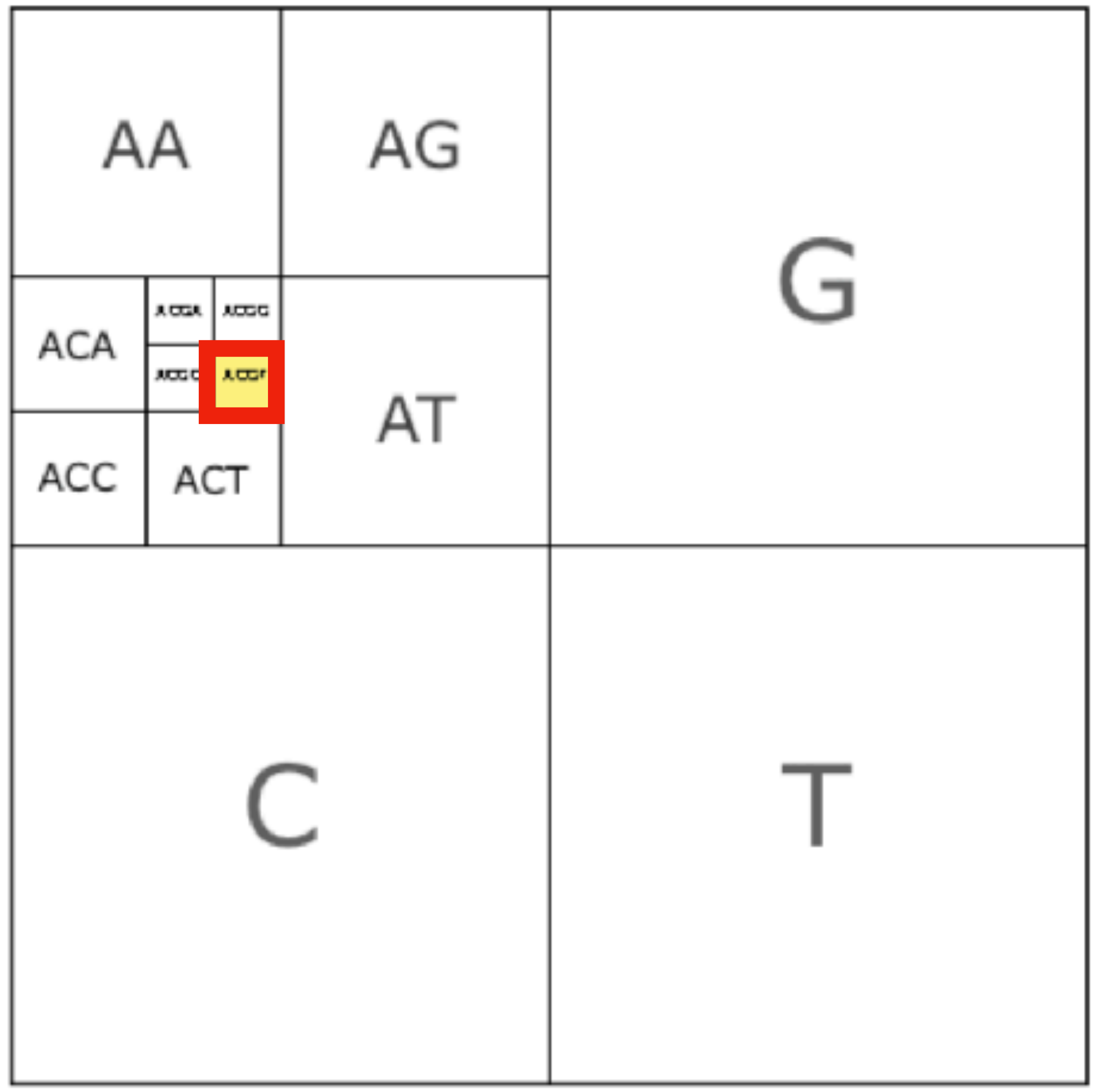
- Decide on kmer length, e.g. kmer=4
- Identify correct pixel, using chaos game scheme
- Count number of occurrences of each pixel across whole sequence



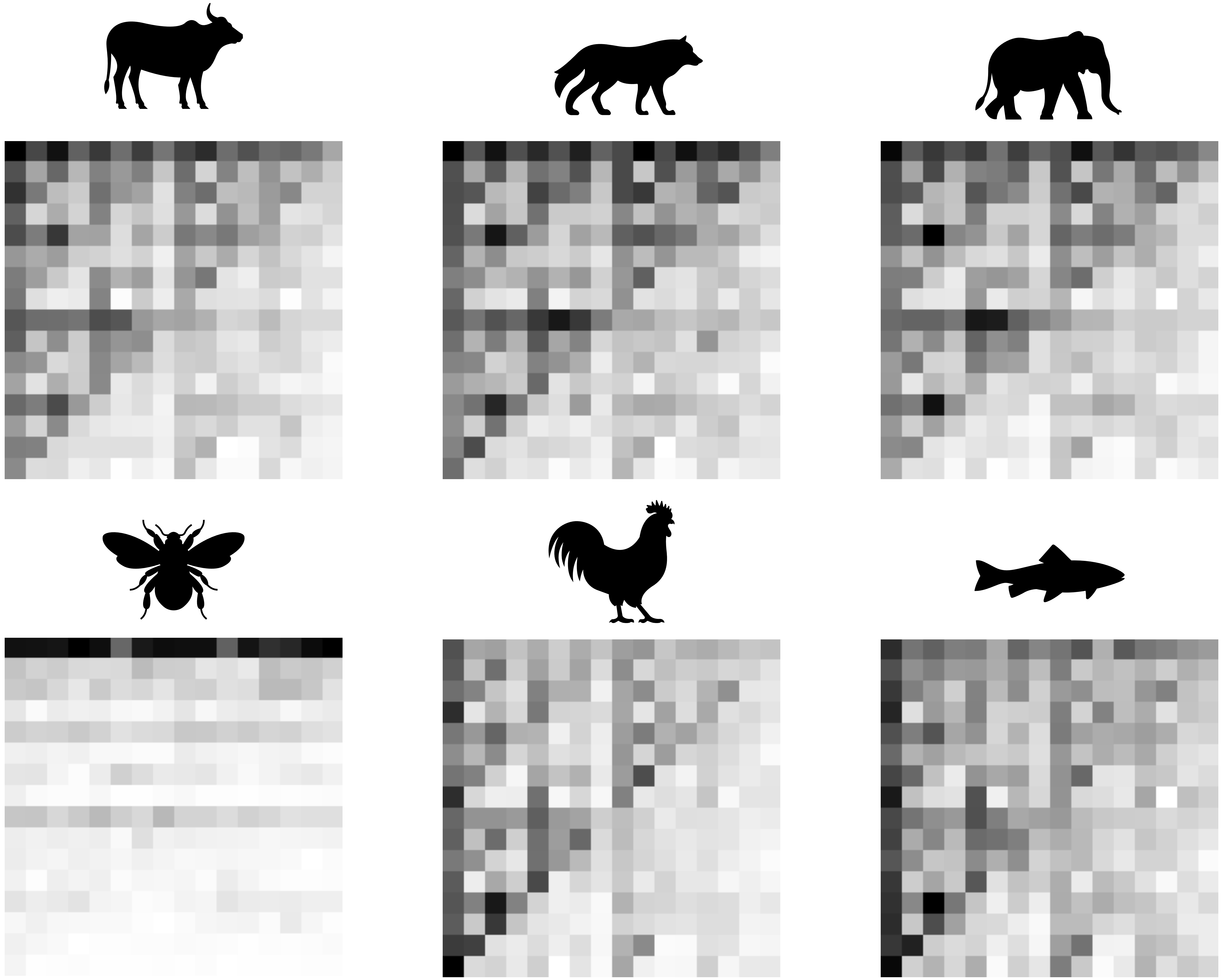
Chaos game representation of DNA sequences

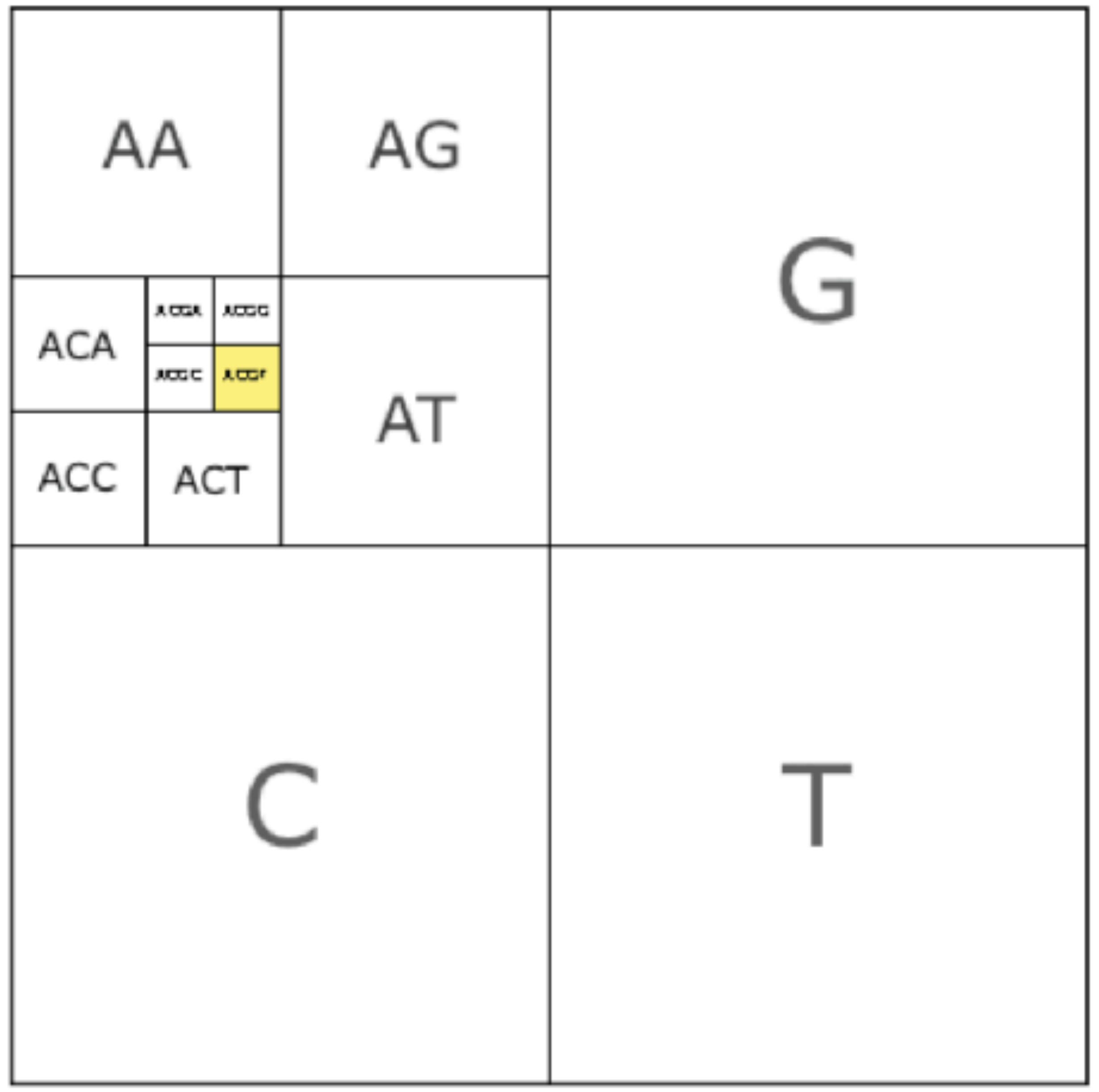
ACGT GGTG TAGA TTTG GGAT TGAC CTAC TAGT ACGA TTCA ...

- Decide on kmer length, e.g. kmer=4
- Identify correct pixel, using chaos game scheme
- Count number of occurrences of each pixel across whole sequence



Chaos game representation of DNA sequences



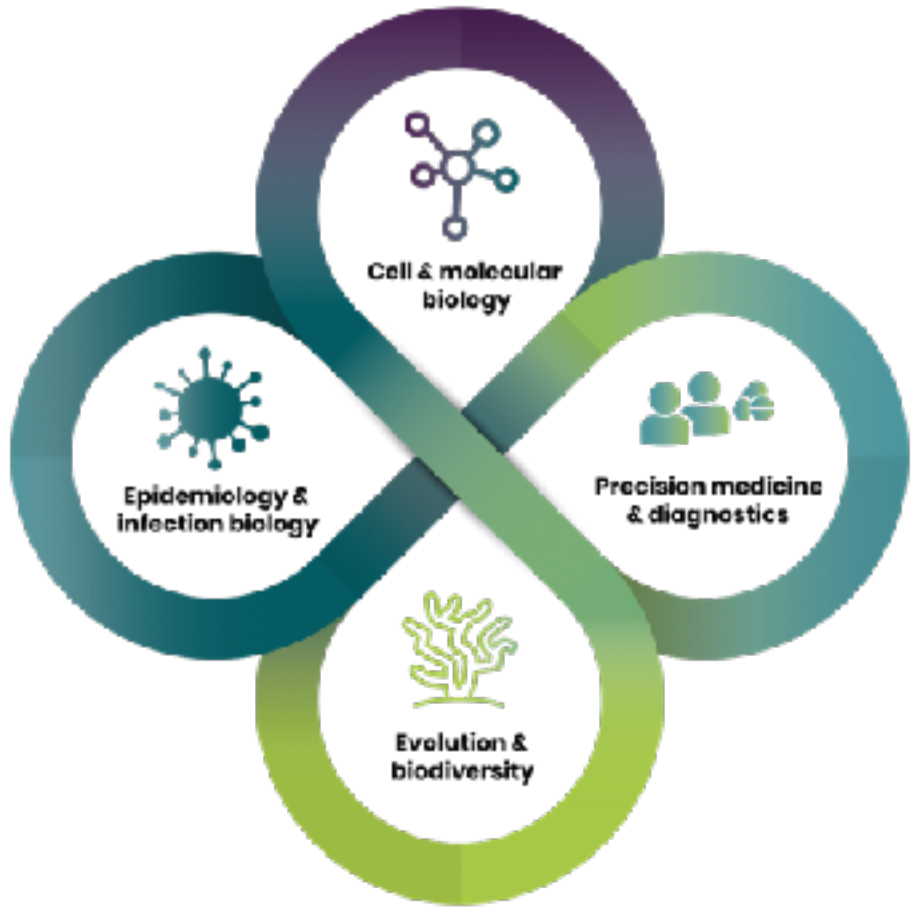


kmer = 4

UPPSALA UNIVERSITET

SciLifeLab

Biodiversity Data Lab (biodiversity.se)



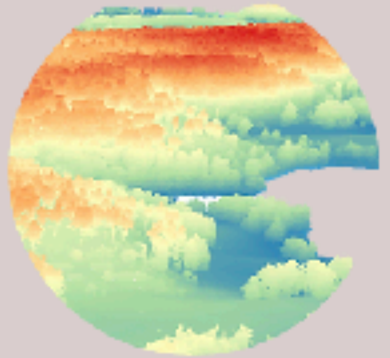
Data-Driven Life Science (DDLS)

What we do

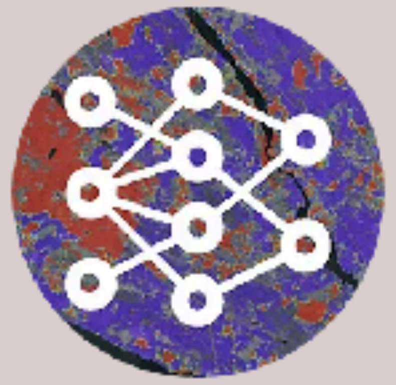
We are passionate about biodiversity and our mission is to combat the ongoing biodiversity crisis by providing and analyzing high-resolution biodiversity data. One focus is to improve field- and lab-work protocols to produce more accurate biodiversity assessments using environmental DNA. The other focus of our group is to develop computational methods that utilize these data to help us better understand and manage the spatial distribution of biodiversity.

Our research is interdisciplinary, including the fields of ecology, molecular biology, geomatics, bioinformatics, data science, and machine learning. If you have a background or interest in any of these fields and care about contributing with your work towards helping biodiversity, the Biodiversity Data Lab is the place for you!

Biodiversity data



Spatial data



Machine learning

A toolbox for neural network models

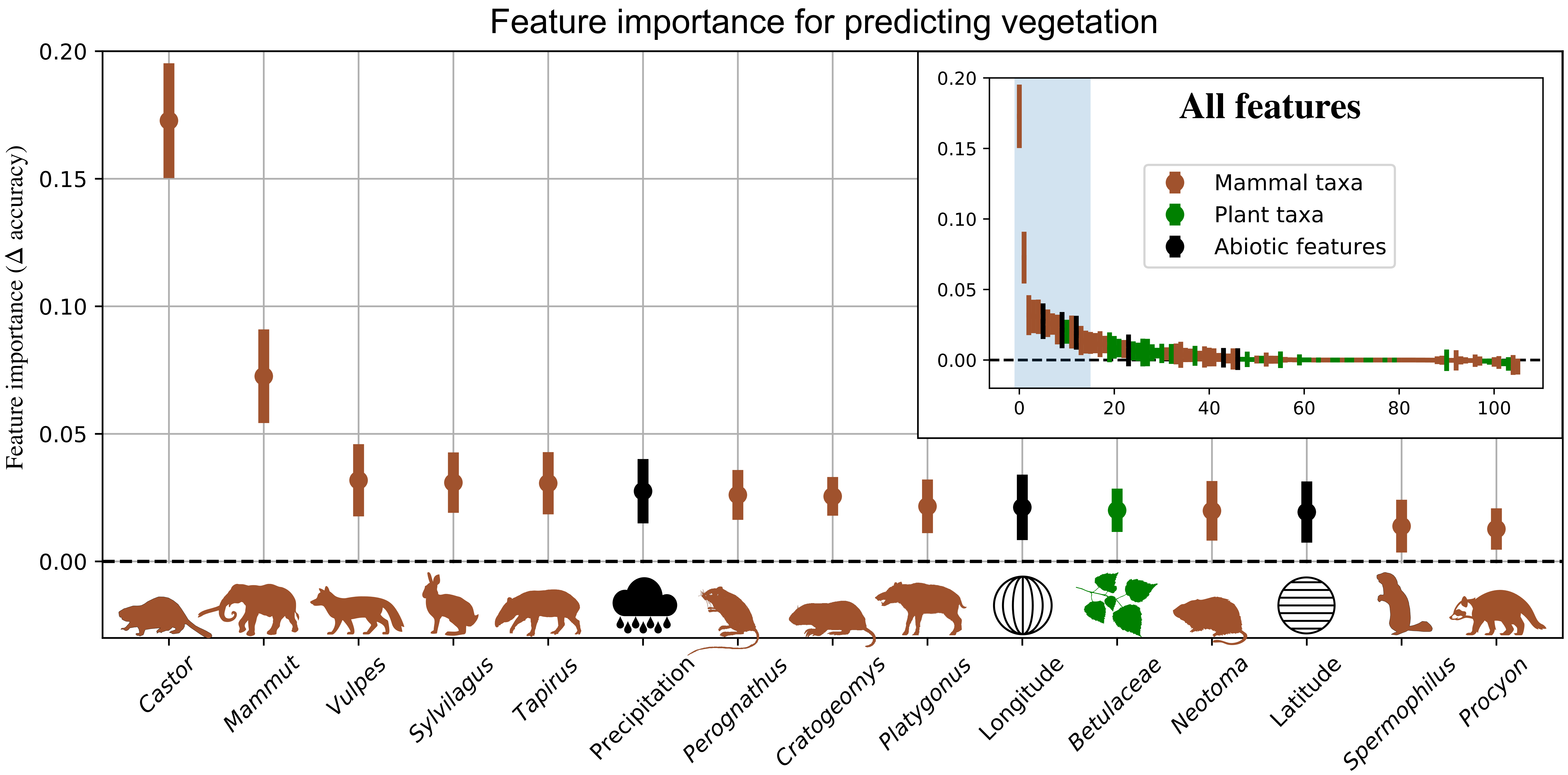
Permutation feature importance

lon	lat	elevation	hfp	bio_1	bio_2	bio_3	bio_4	bio_5	bio_6	bio_7	bio_8	bio_9	bio_10
136.5329	-28.4025	36	3.9215	21.8833	14.5667	45.8071	636.0436	37.9	6.1	31.8	17.4167	28.85	29.55
150.8236	-24.8218	373	5.2002	20.0292	14.5417	54.2599	459.8392	32	5.2	26.8	25.0667	13.7667	25.0667
136.2989	-25.2995	138	0	22.6333	15.6	46.4286	666.822	38.8	5.2	33.6	30.4333	16.0167	30.4333
133.7812	-25.9481	455	6.1402	20.9	14.2167	44.8475	640.3905	36.5	4.8	31.7	27.6833	14.6667	28.4667
141.1054	-17.6605	3	8.8415	27.225	11.6167	55.0553	313.2926	36.2	15.1	21.1	29.4333	24.0333	30.3167
141.7411	-31.0796	171	9.1679	19.3917	14.15	46.3934	616.0388	35.1	4.6	30.5	26.8667	11.6333	26.8667
134.3736	-14.1506	126	0.265	27.3167	13.25	57.8603	301.2424	37.8	14.9	22.9	29.1833	23	30.4833
133.6164	-22.2898	555	0.2611	22.8417	15.8	49.375	592.7203	37.5	5.5	32	29.3167	17.05	29.3167
133.8822	-14.6768	58	0.0152	27.5917	13.4333	56.6807	322.9962	38.3	14.6	23.7	29.4667	22.9833	30.95
141.817	-20.0843	124	3.2501	26.0458	14.7917	54.3811	448.4796	38.1	10.9	27.2	30.5	21.55	30.6
135.409	-15.5381	25	7.776	27.4042	13.425	56.6456	332.9173	38.1	14.4	23.7	30.4333	23.9167	30.9667
140.8475	-17.4635	3	7.8711	26.7333	11.4	54.0284	322.1331	35.6	14.5	21.1	29	23.4	29.9833
137.0116	-17.8959	262	0	26.0458	14.5417	54.6679	411.6401	37.9	11.3	26.6	29.9167	21.8833	30.2167
143.7243	-20.1088	455	3.25	23.8792	14.025	54.1506	432.4374	35.4	9.5	25.9	28.35	19.55	28.35
133.2699	-22.2836	601	0.2681	22.5167	14.5	47.8548	583.1666	36.3	6	30.3	28.85	14.7	28.85
136.5648	-25.7836	101	0	22.7417	15.3333	45.9082	670.7792	39	5.6	33.4	30.6667	16.1167	30.6667
135.4576	-26.4805	126	5.5066	22.2167	14.2	44.795	641.7531	37.8	6.1	31.7	21.9333	15.8833	29.8333
133.6333	-23.8357	606	1.2281	21.2542	14.775	46.4623	618.0742	36.4	4.6	31.8	27.5667	15.2667	28.3
135.83	-16.8239	192	0.2546	26.5917	14.7167	56.3857	382.9659	38.3	12.2	26.1	29.4333	21.2	30.55
150.0949	-23.6582	109	15.603	22.3083	12.7	52.9167	417.3011	32.9	8.9	24	26.85	16.6167	26.85
141.1789	-17.295	10	7.6072	26.9792	11.4417	55.008	299.4576	36	15.2	20.8	29.0167	23.9	30.0167
139.53	-34.32	94	2.3531	16.8875	13.6083	51.9402	471.2031	31.5	5.3	26.2	14	22.4167	22.7167
141.0037	-23.6464	123	0.7295	24.5708	14.9417	48.6699	597.4737	38.6	7.9	30.7	30.5333	16.5667	31.2
135.8615	-28.0871	79	3.2674	22.2167	13.7167	44.9727	623.6671	37.7	7.2	30.5	17.8	22.7	29.7333
135.5817	-26.3667	109	0.4948	22.3875	14.675	45.5745	643.2062	38.1	5.9	32.2	26.35	16	30.0167
134.3555	-14.177	188	0.2518	26.9542	13.175	57.7851	307.3897	37.4	14.6	22.8	28.8833	22.55	30.1667
133.4385	-25.7375	494	2.42	21.05	14.55	45.6113	637.9299	36.7	4.8	31.9	27.75	12.8667	28.5333
134.4207	-14.0479	58	8.7569	27.5667	13.6667	57.9096	298.6129	38.3	14.7	23.6	29.3	23.3	30.7167
142.4779	-20.0137	191	9.7238	25.525	14.5	54.5113	439.615	37.3	10.7	26.6	29.9167	21.1167	29.9833

Test set

- Predict test set with trained model to establish baseline accuracy
- Shuffle feature column in test set to break information content
- Predict accuracy with modified test set
- Difference in accuracy compared to baseline signifies importance of feature

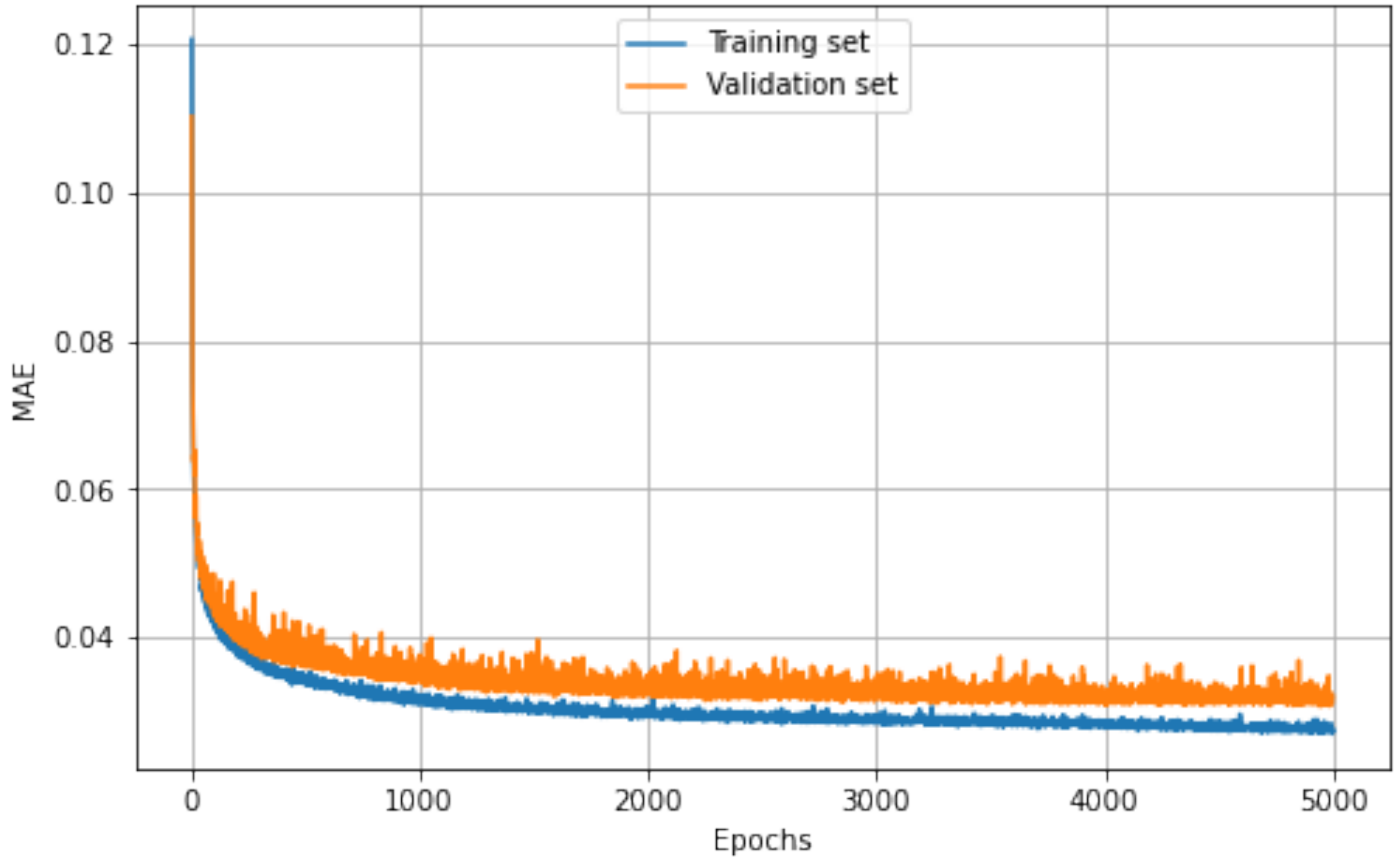
Feature importance



Andermann *et al.* The origin and evolution of open habitats in North America inferred by Bayesian deep learning models. ***Nature Communications***

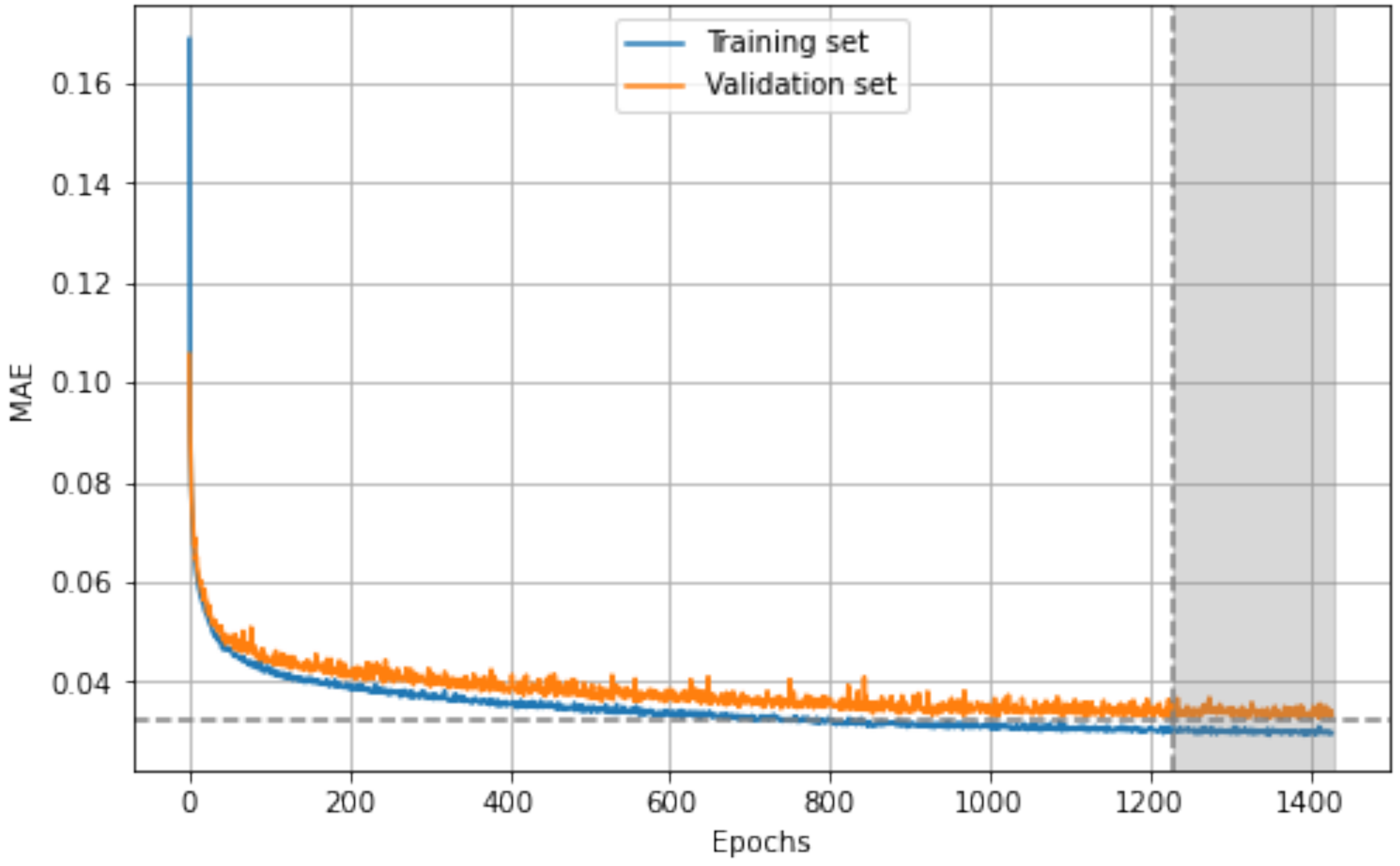
Fit model to data (= training)

```
# Run model training and store training history
history = model.fit(train_features,
                          train_labels,
                          epochs=5000,
                          validation_data=(validation_features, validation_labels),
                          verbose=1,
                          batch_size=40)
```



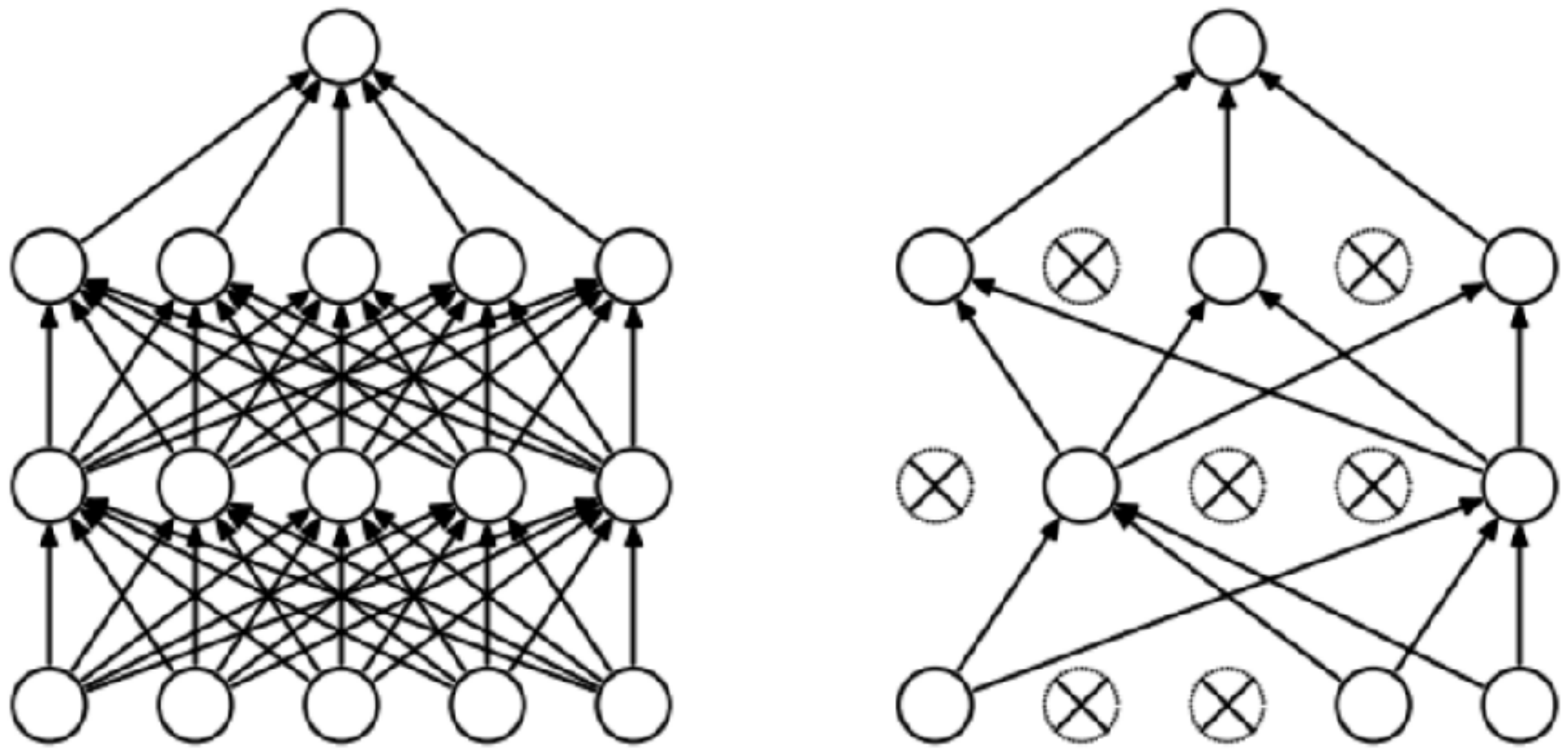
Early stopping

```
# Define early stop of training based on validation set
early_stop = tf.keras.callbacks.EarlyStopping(monitor='val_mae',patience=200,restore_best_weights=True)
# Run model training and store training history
history = model.fit(train_features,
                          train_labels,
                          epochs=5000,
                          validation_data=(validation_features, validation_labels),
                          verbose=1,
                          callbacks=[early_stop],
                          batch_size=40)
```



Dropout

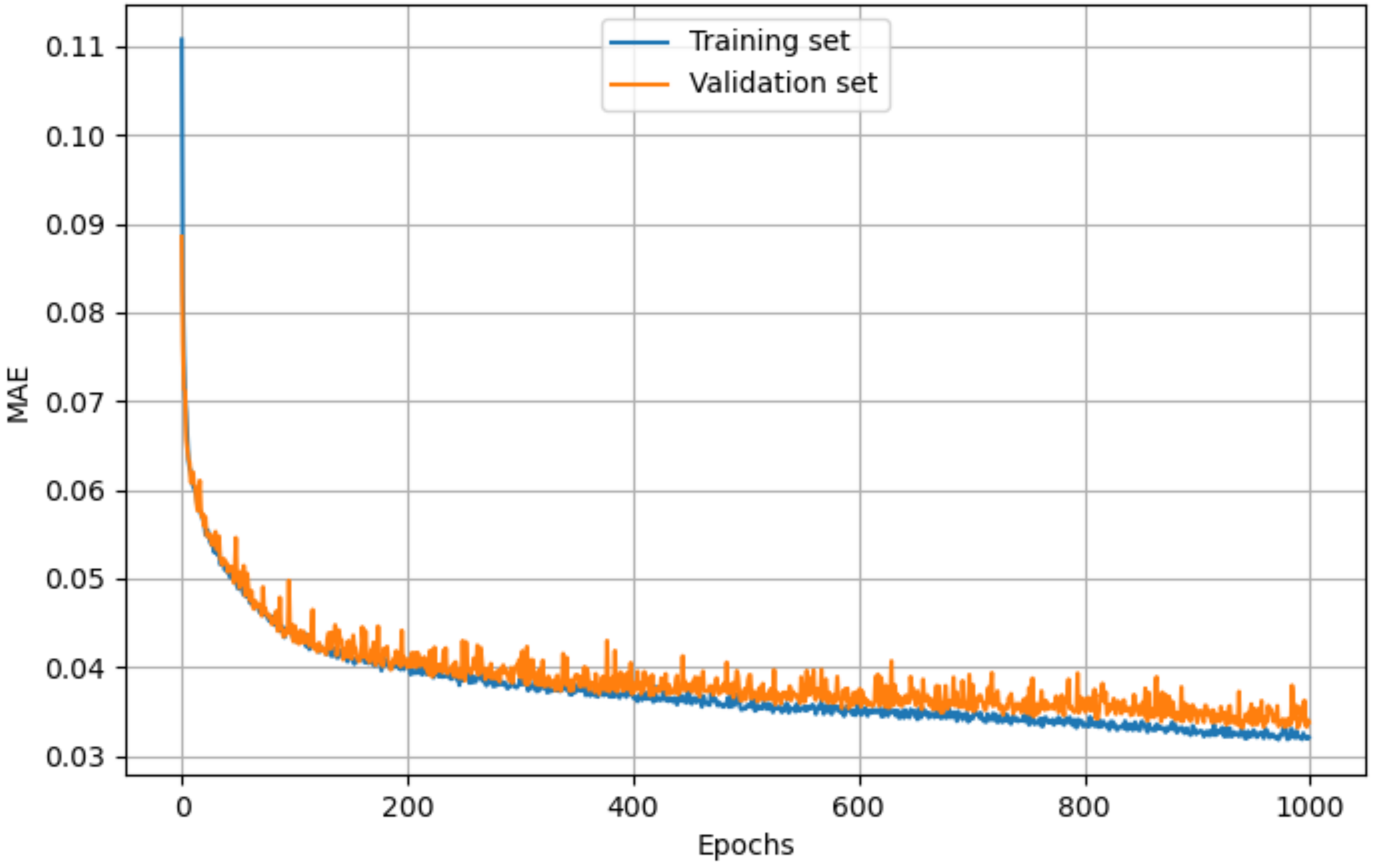
- Regularization technique
- Accuracy increase and less overfitting
- Every neuron (except output neurons) has a probability of being "dropped out"
- Parameter: dropout rate (between 10-50%)



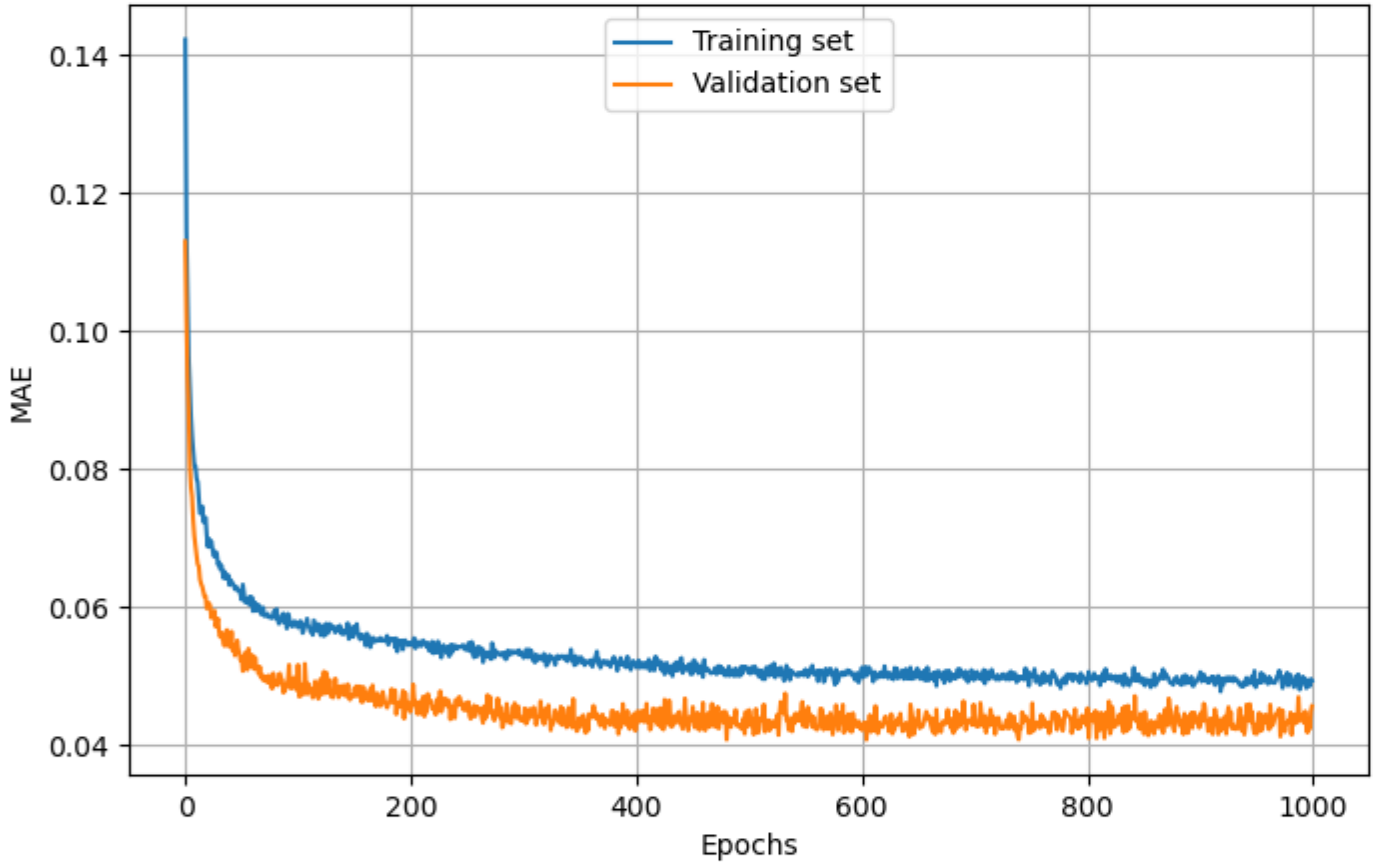
(a) Standard Neural Net (b) After applying dropout.

```
architecture = []
# Input layer
architecture.append(tf.keras.layers.Flatten(input_shape=[4]))
# 1st hidden layer
architecture.append(tf.keras.layers.Dense(5, activation='relu',use_bias=False))
architecture.append(tf.keras.layers.Dropout(0.2))
# Output layer
architecture.append(tf.keras.layers.Dense(1, activation='softplus',use_bias=False))
```

Dropout



Regular NN model

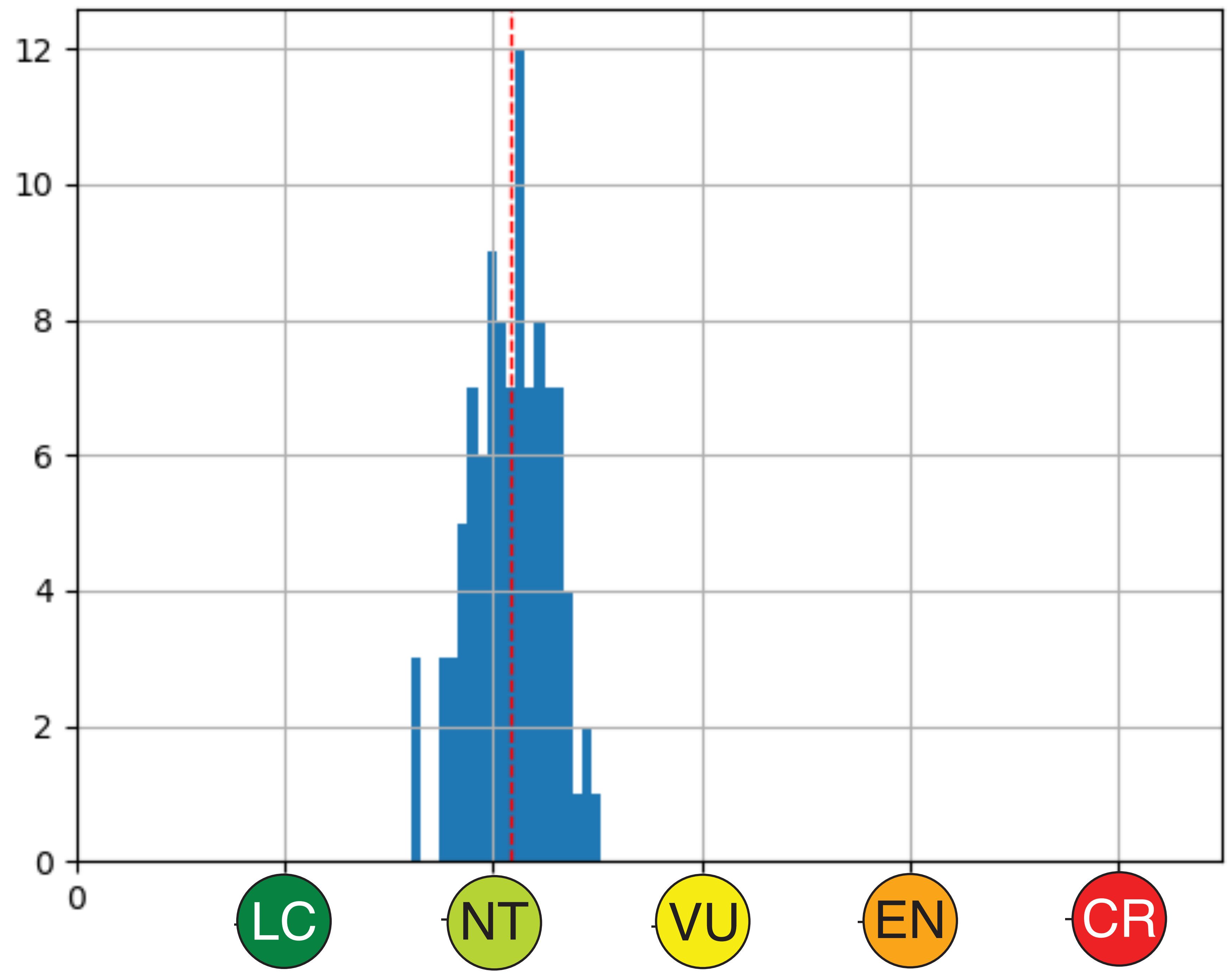


NN model with dropout

Monte Carlo (MC) Dropout

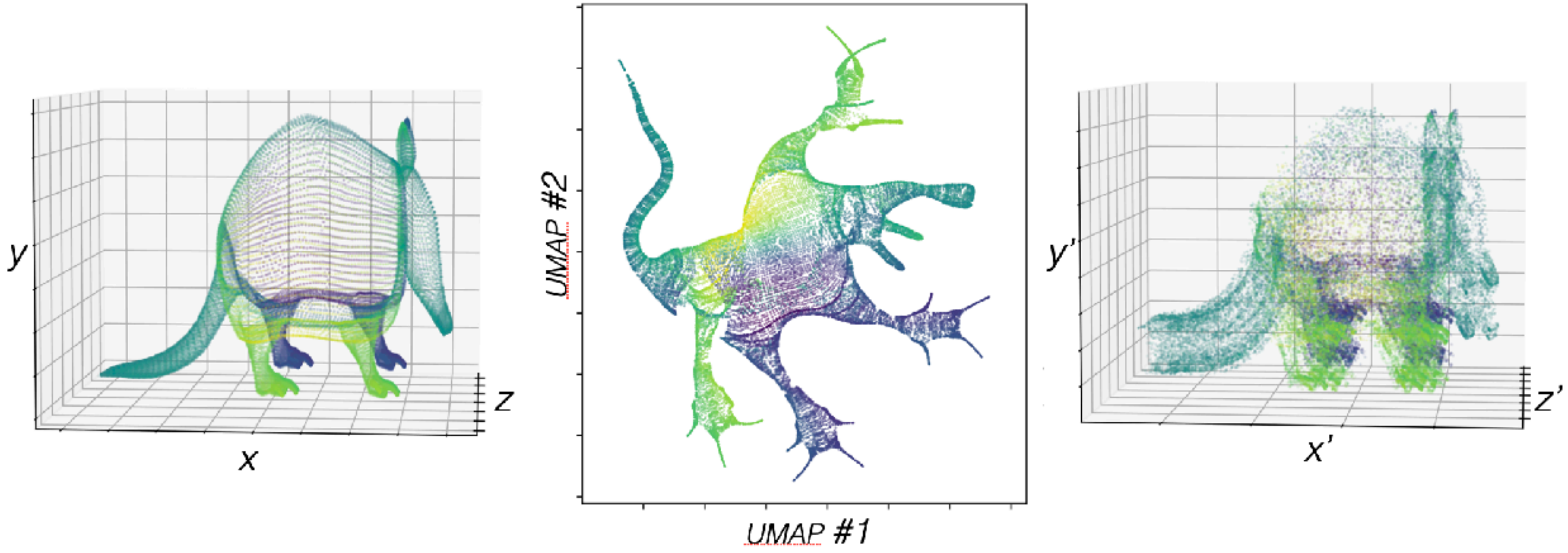
- Method to quantify uncertainty
- Can be applied to any model that was trained with dropout
- Dropout remains active also for predictions, introduces stochasticity
- Repeat predictions e.g. 100 times, quantify uncertainty

100 MC dropout predictions for one test instance




```
mc_dropout_pred = np.stack([model(test_features,training=True) for i in np.arange(100)])
mc_dropout_mean = mc_dropout_pred.mean(axis=0)
mc_dropout_std = mc_dropout_pred.std(axis=0)
```

Dimensionality reduction with UMAP



Purposes:

- Visualize data structure that can't be viewed in original form
- Clustering and outlier detection
- Represent data in fewer dimensions
- Reduce collinearity between axes

Figure provided by Daniele Silvestro

Dimensionality reduction with UMAP

BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) (×100)

BIO4 = Temperature Seasonality (standard deviation ×100)

BIO5 = Max Temperature of Warmest Month

BIO6 = Min Temperature of Coldest Month

BIO7 = Temperature Annual Range (BIO5-BIO6)

BIO8 = Mean Temperature of Wettest Quarter

BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month

BIO15 = Precipitation Seasonality (Coefficient of Variation)

BIO16 = Precipitation of Wettest Quarter

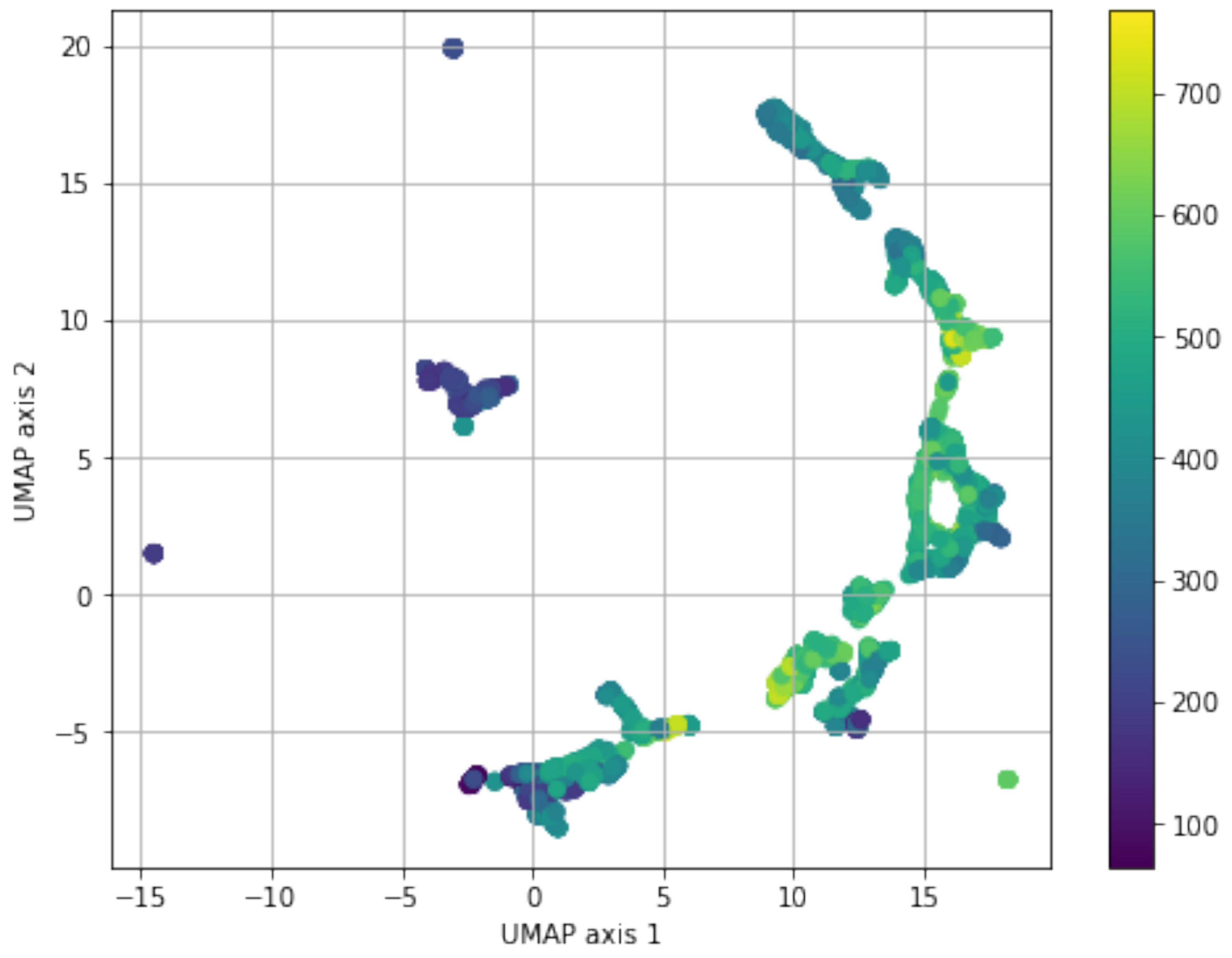
BIO17 = Precipitation of Driest Quarter

BIO18 = Precipitation of Warmest Quarter

BIO19 = Precipitation of Coldest Quarter

BioClim data - highly correlated

```
import umap.umap_ as umap

reducer = umap.UMAP(n_neighbors=250, min_dist=0)
umap_obj = reducer.fit(biome_features)
biome_features_transformed = reducer.transform(biome_features)
```

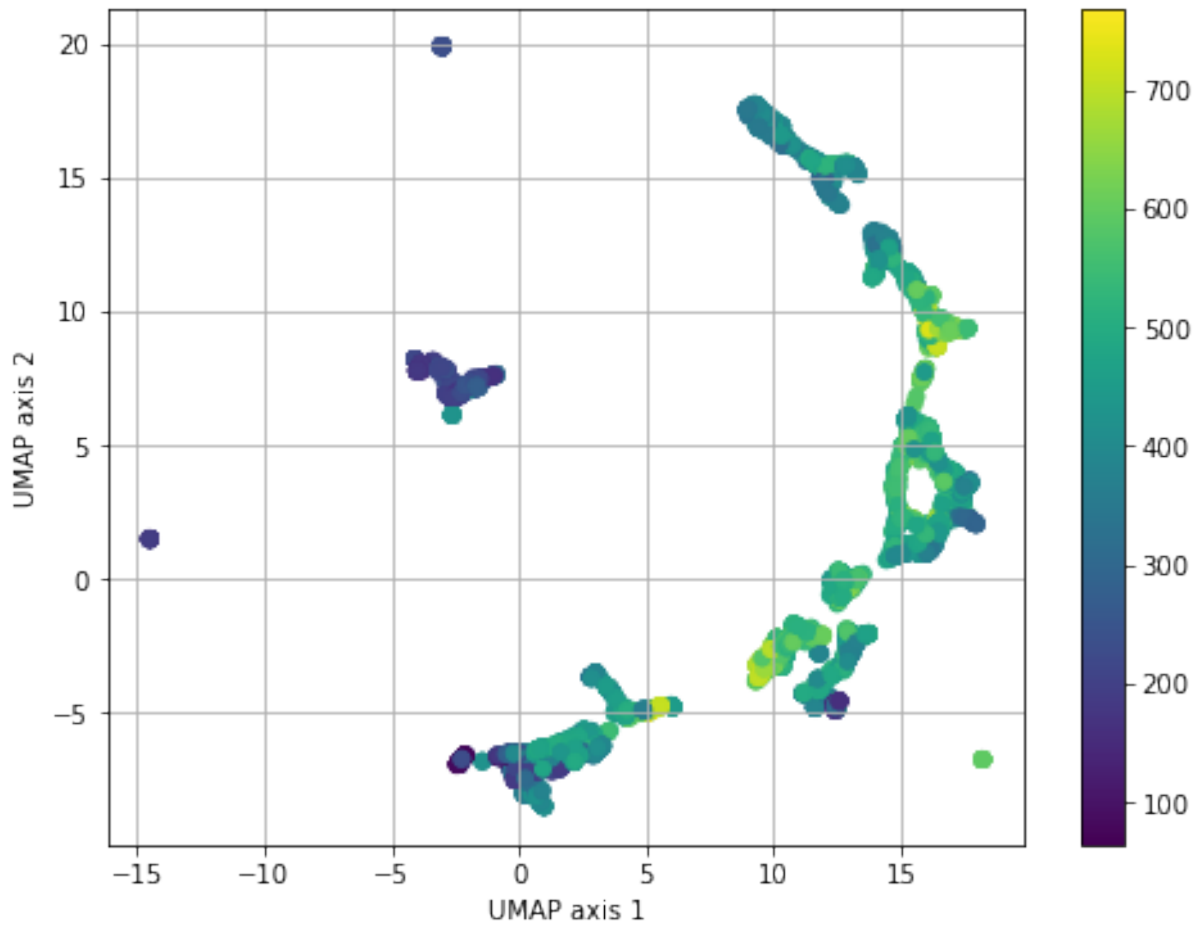


- From 19 to 2 dimensions
- Points colored by species diversity at each site
- UMAP transformed BioClim variables seem to carry some informativeness

Dimensionality reduction with UMAP

UMAP

```
import umap.umap_ as umap

reducer = umap.UMAP(n_neighbors=250, min_dist=0)
umap_obj = reducer.fit(biome_features)
biome_features_transformed = reducer.transform(biome_features)
```



PCA

```
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
pca.fit(biome_features)
biome_features_pca_transformed = pca.transform(biome_features)
```