BEeS
Heraklion
30 June - 3 July 2025

LifeWatch ERIC
hcmr ΕΛΚΕΘΕ
iMBBC ΙΘΑΒΒΥΚ

Session: Exploring boundaries of life hosting spaces: Habitat mapping

2 July 2025 | 08:30-10:30

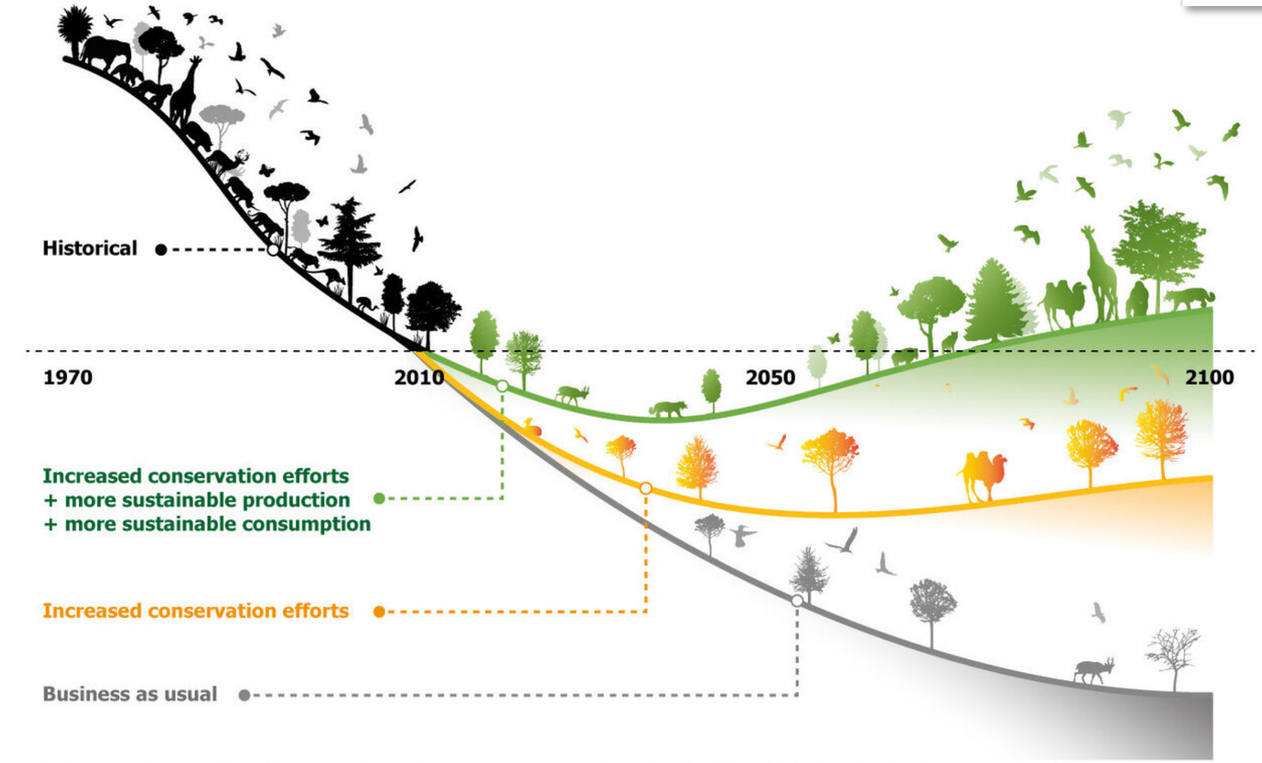
Predicting butterfly species presence from satellite imagery using soft contrastive learning

Thijs van der Plas (Wageningen University & Research)

2 July 2025

Biodiversity monitoring at scale

‘Crucial’ Cop15 deal includes target to protect 30% of nature on Earth by 2030



This artwork illustrates the main findings of the article, but does not intend to accurately represent its results (https://doi.org/10.1038/s41586-020-2705-y)

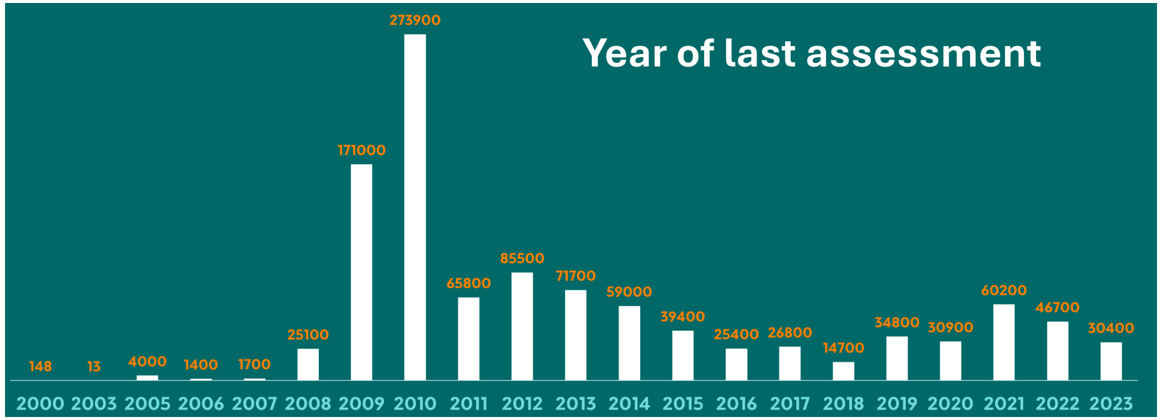
Leclère et al., 2020, *Nature*

Top: The Guardian, 19 Dec 2022
Bottom: Kunming-Montreal COP15 2022

Biodiversity monitoring at scale in the UK

In England, 66% of SSSIs [nature reserves] had not been assessed for >10 years (in 2023). Recent assessments show worse conditions.

Can we create scalable monitoring methods? -> ML and EO



Top: Wild Justice, 2023, *A sight for sore SSSIs*

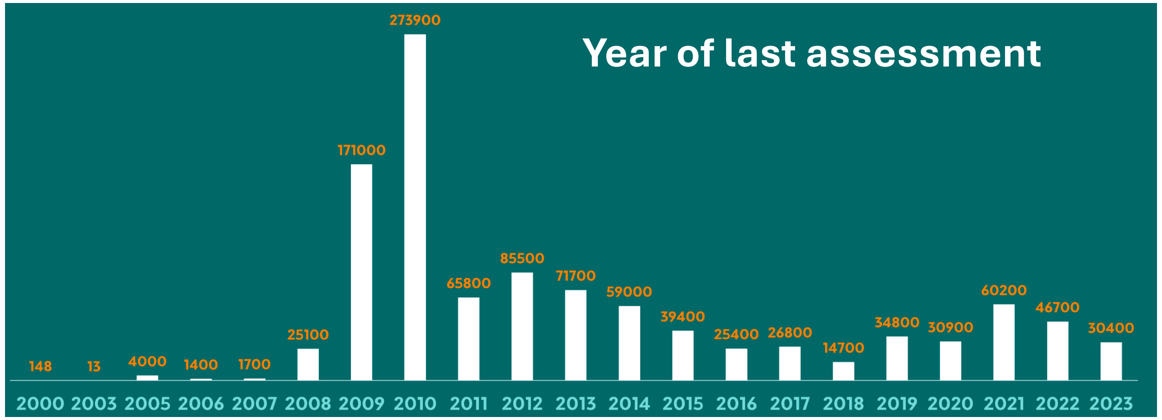
Biodiversity monitoring at scale in the UK

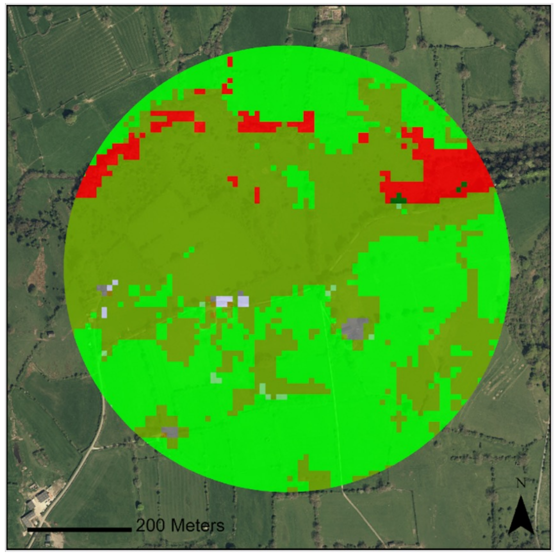
In England, 66% of SSSIs [nature reserves] had not been assessed for >10 years (in 2023). Recent assessments show worse conditions.

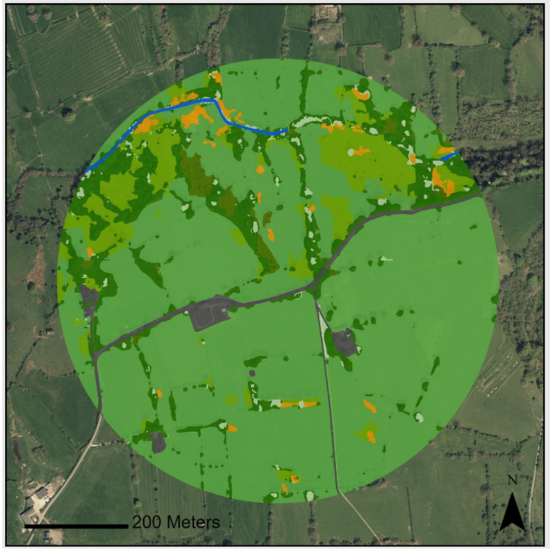
Can we create scalable monitoring methods? -> ML and EO

We improved UK National Park land cover map from 10 m to 12.5 cm resolution.

How can we map habitat *condition*?

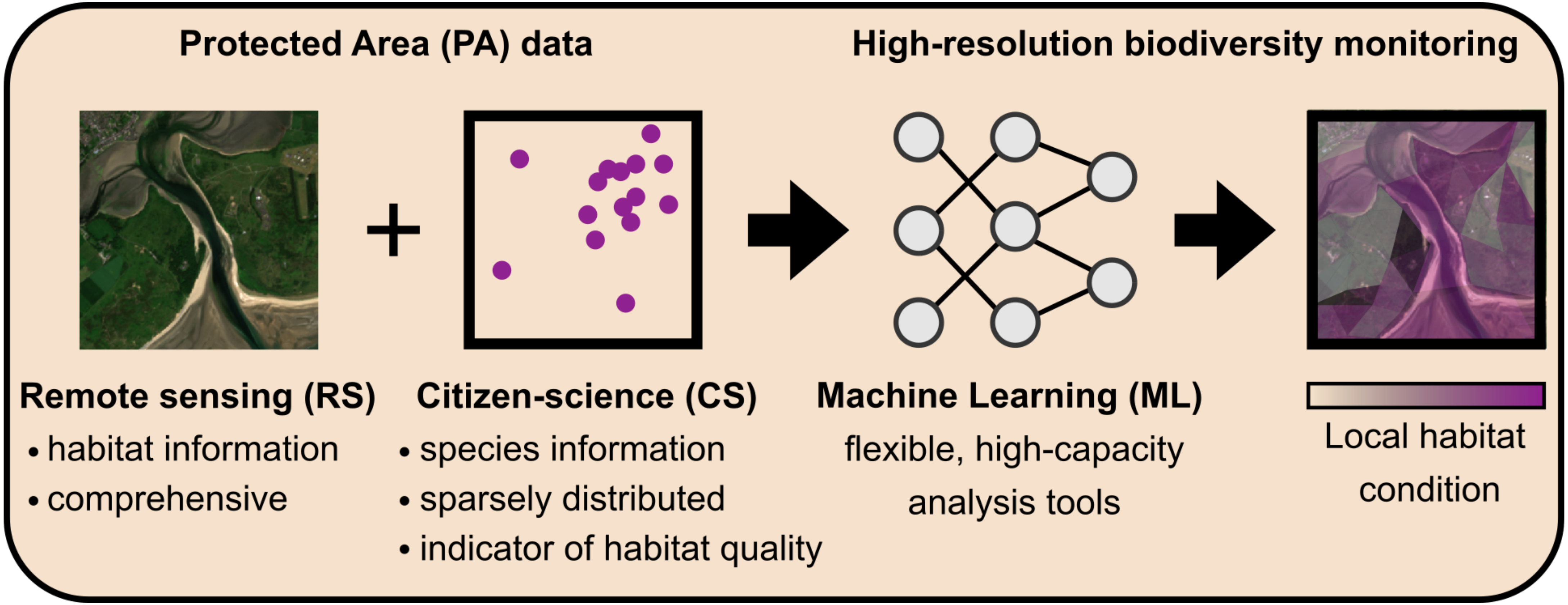






Top: Wild Justice, 2023, *A sight for sore SSSIs*
Bottom: Van der Plas et al., 2023, *Rem. Sens.*

Biodiversity monitoring at scale in the UK



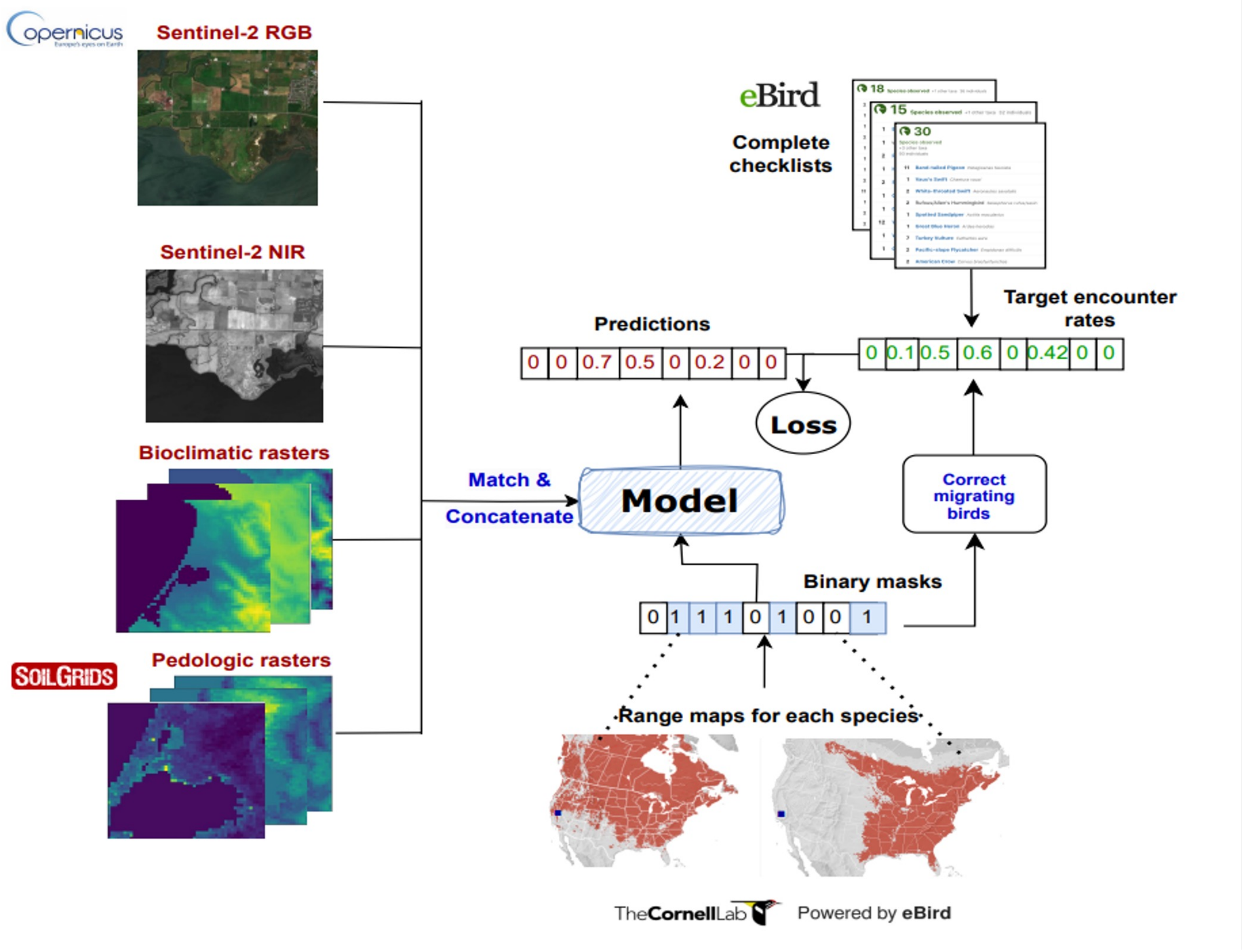
Van der Plas et al., 2025, *Ecol. Sol. Evid.*

Predicting species presence from satellite data

Predict bird species in US using eBird and sentinel-2.

Aggregate ‘complete’ observations across visits to compute P(observed) per species.

ML task: predict P(observed) per species per location from remote sensing data.

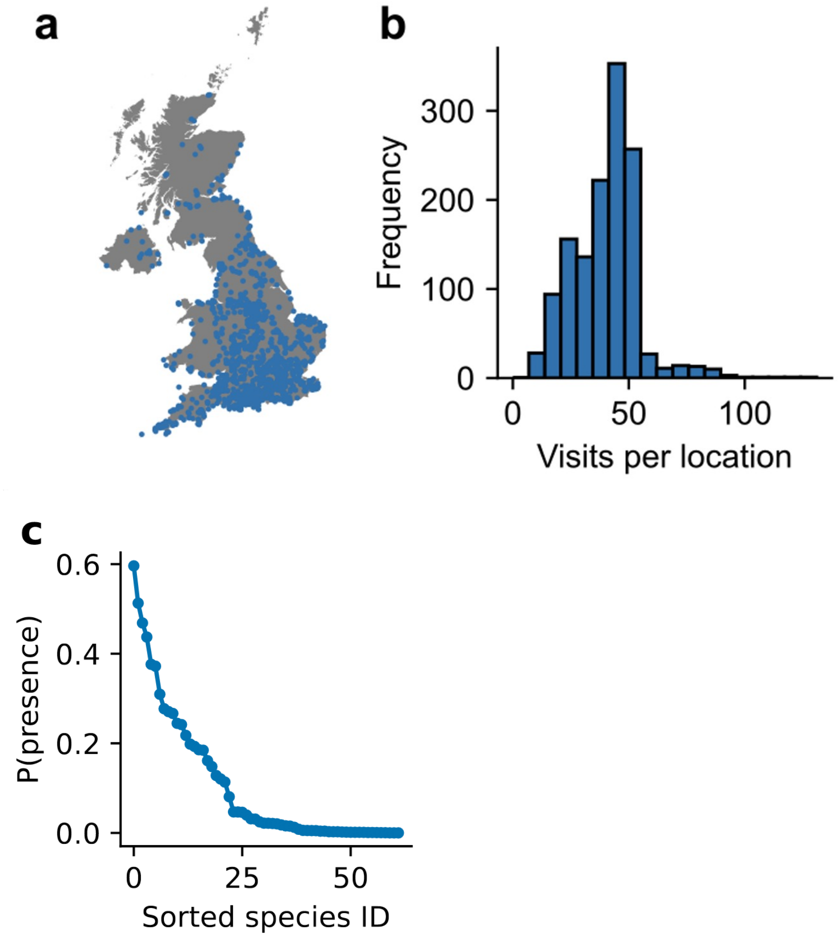


Teng et al., 2023, *NeurIPS Datasets and Benchmarks*

S2BMS data set

UK Butterfly Monitoring Scheme (UKBMS): highly structured citizen science project for recording butterflies.

We used 2018 and 2019, locations with > 200 obs: 1329 locations & 62 species remaining.



UKBMS UK Butterfly Monitoring Scheme

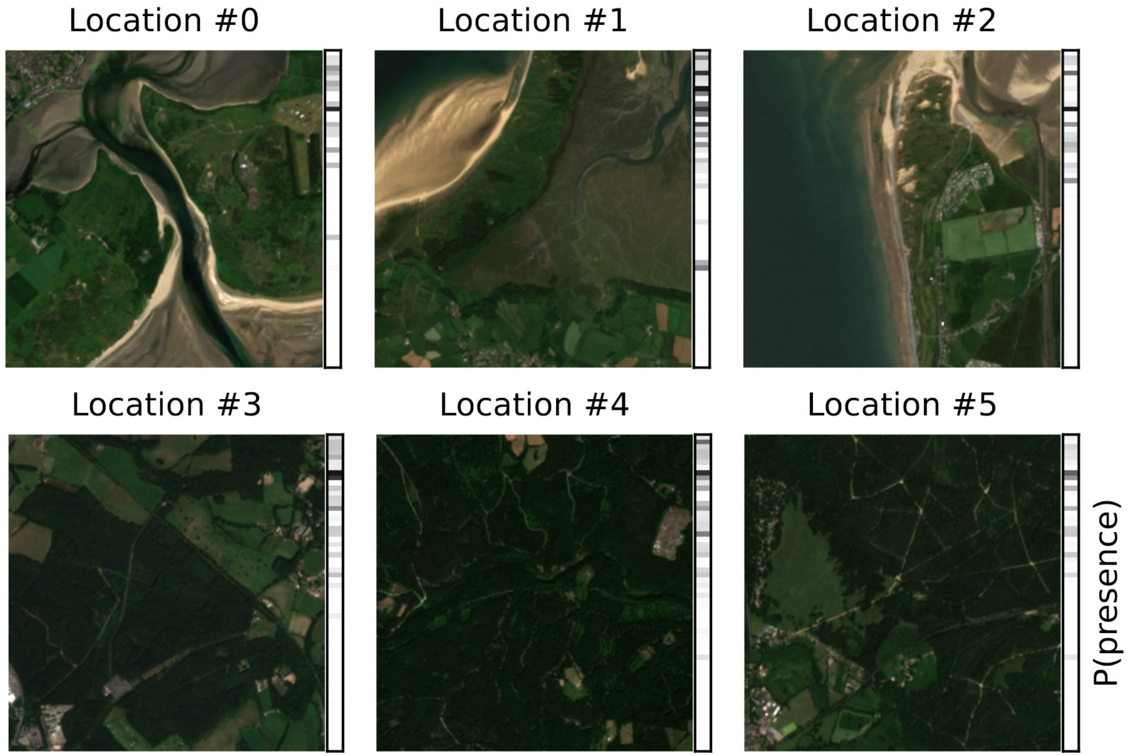
Van der Plas et al., 2025, *CVPRW proc.*

S2BMS data set

UK Butterfly Monitoring Scheme (BMS): highly structured citizen science project for recording butterflies.

We used 2018 and 2019, locations with > 200 obs: 1329 locations & 62 species remaining.

2.56 km x 2.56 km sentinel-2 images of 10 m RGB+IR bands. Cloud-free between June-Aug.



Van der Plas et al., 2025, *CVPRW proc.*

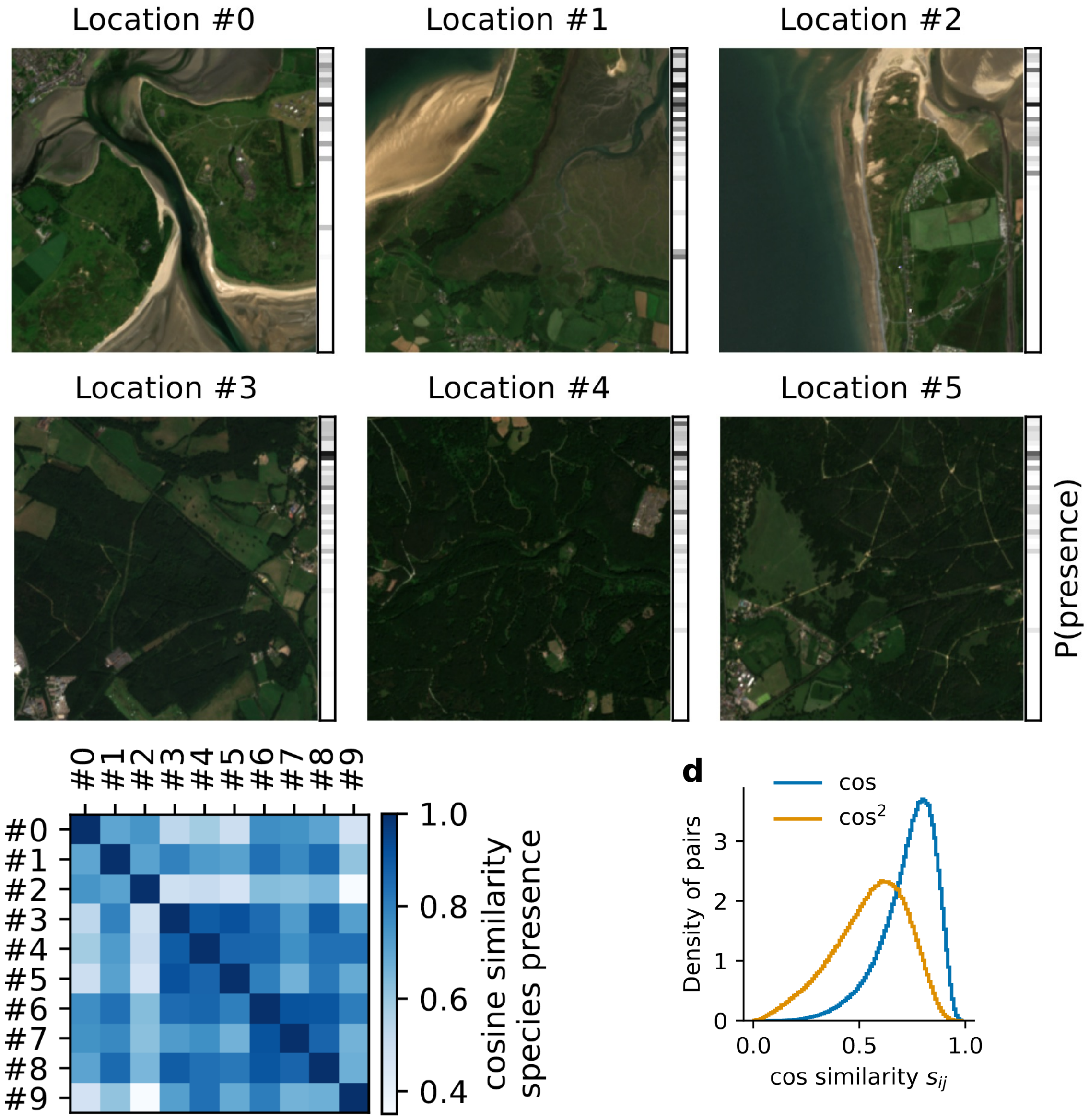
S2BMS data set

UK Butterfly Monitoring Scheme (BMS): highly structured citizen science project for recording butterflies.

We used 2018 and 2019, locations with > 200 obs: 1329 locations & 62 species remaining.

2.56 km x 2.56 km sentinel-2 images of 10 m RGB+IR bands. Cloud-free between June-Aug.

Compute species similarity between locations.



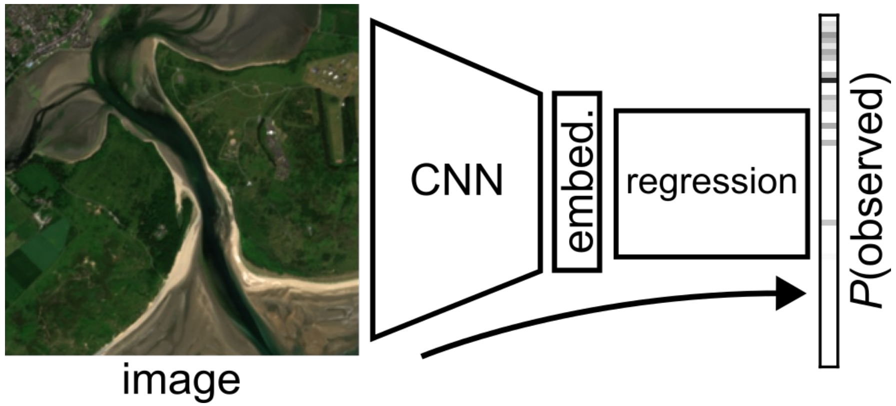
Van der Plas et al., 2025, *CVPRW proc.*

Predicting with CNN

~ 1000 'training' data points.

Convolutional NN (Resnet) encodes image to embedding.

Data augmentation: crop, flip, rotate.



Van der Plas et al., 2025, *CVPRW proc.*

Predicting with CNN

~ 1000 ‘training’ data points.

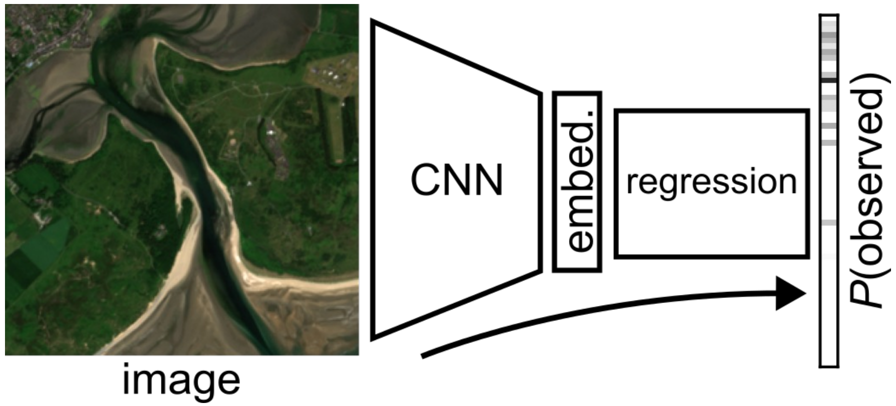
Convolutional NN (Resnet) encodes image to embedding.

Data augmentation: crop, flip, rotate.

Few, but high-quality data. How can we extract more information?

Currently *N* data points. If we could learn from similarities between locations: N^2

L	Model	Top-10 [%]	Top-5 [%]	MSE [1e-02]
-	Mean rate	67.3	58.7	1.39
1	ImageNet	68.9 ± 0.8	60.9 ± 1.0	1.33 ± 0.05
1	SeCo	69.1 ± 0.5	61.8 ± 0.9	1.34 ± 0.04
2	ImageNet	**69.7 ± 0.5**	61.5 ± 0.7	1.24 ± 0.03
2	SeCo	69.5 ± 0.4	**62.4 ± 0.8**	1.22 ± 0.04
3	ImageNet	69.0 ± 0.5	61.8 ± 0.7	1.24 ± 0.03
3	SeCo	69.6 ± 0.2	62.4 ± 1.0	**1.21 ± 0.04**

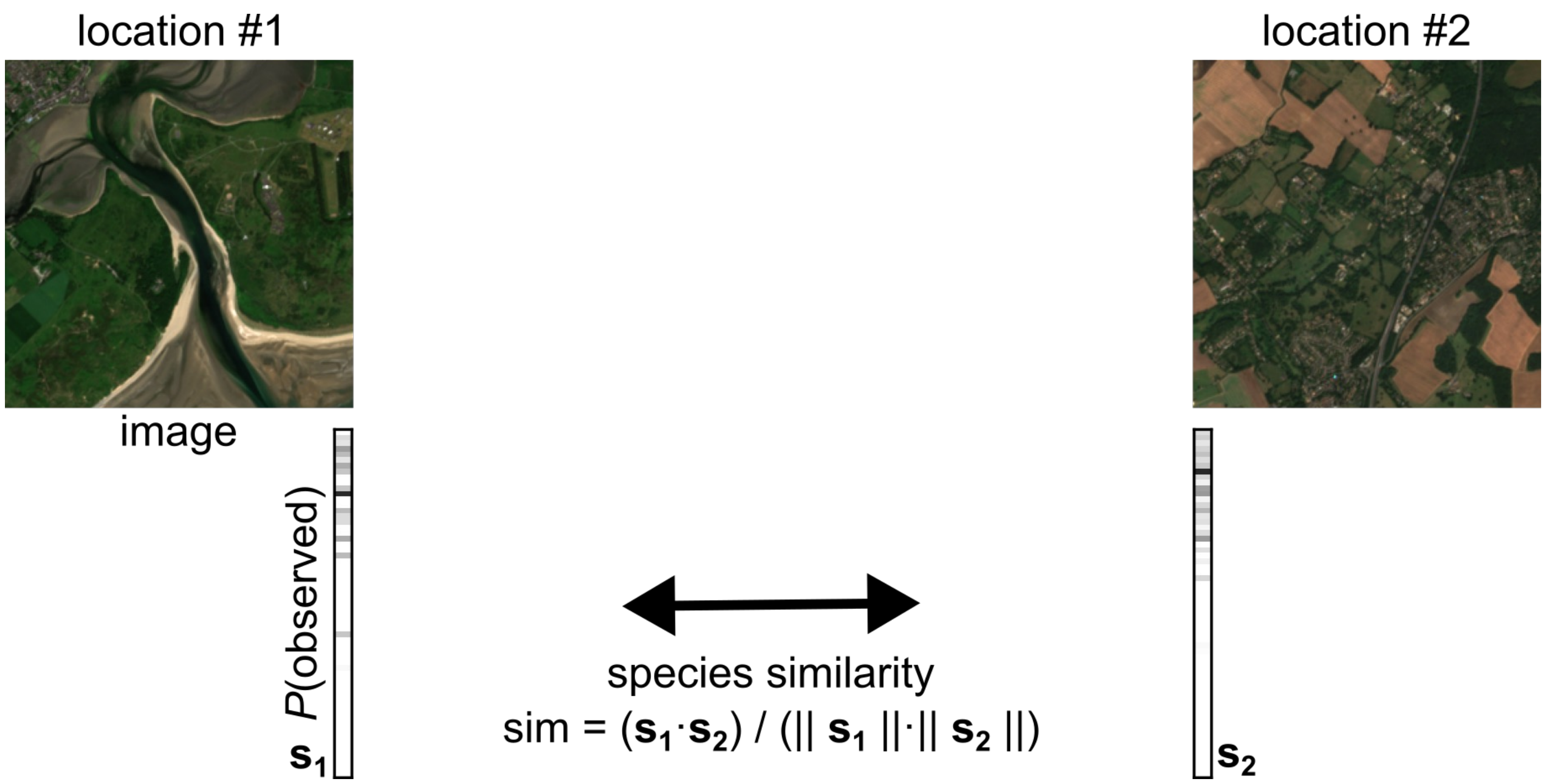


Van der Plas et al., 2025, *CVPRW proc.*

PECL: Paired Embeddings Contrastive Learning

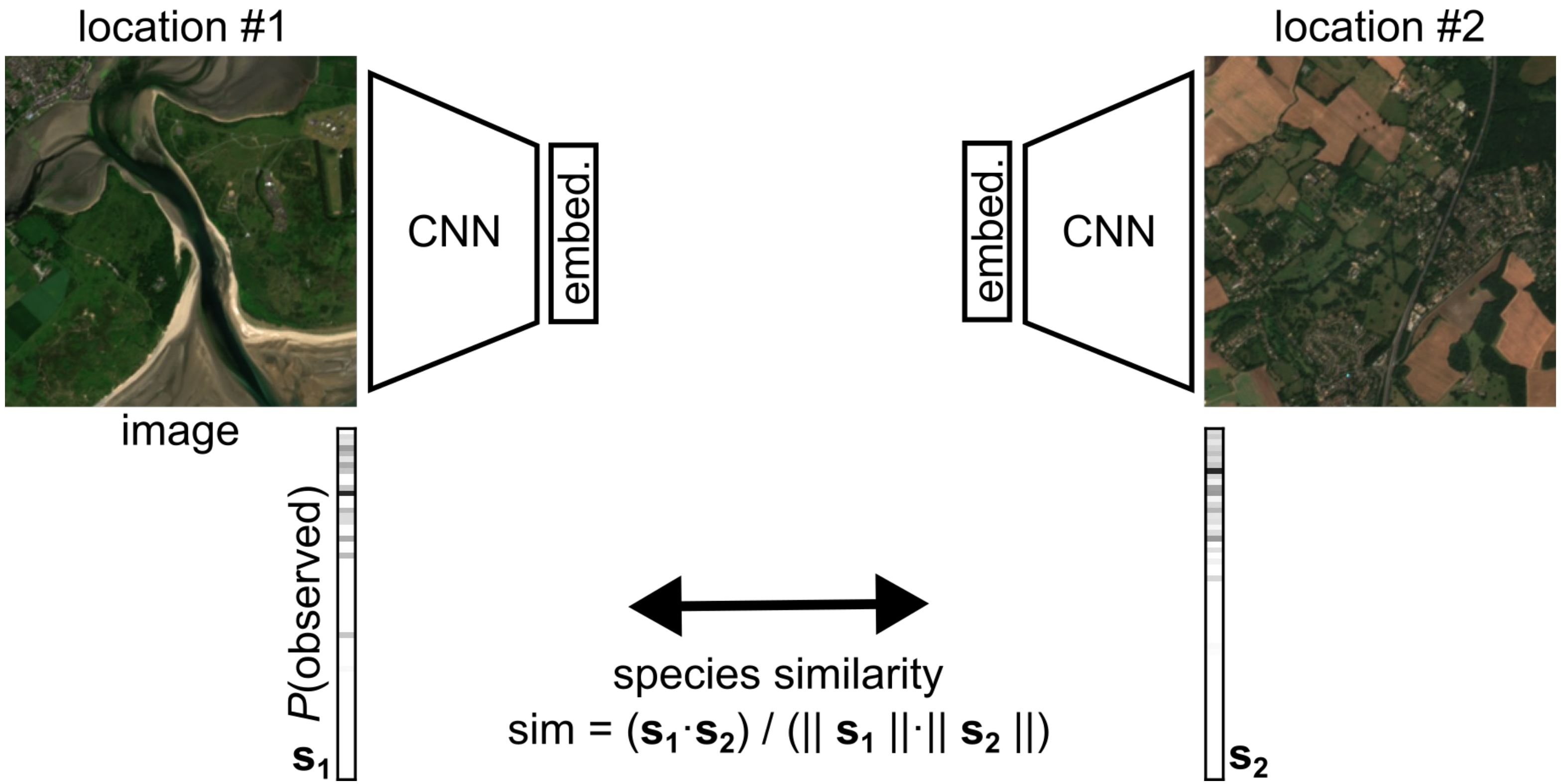
Van der Plas et al., 2025, *CVPRW proc.*

PECL: Paired Embeddings Contrastive Learning



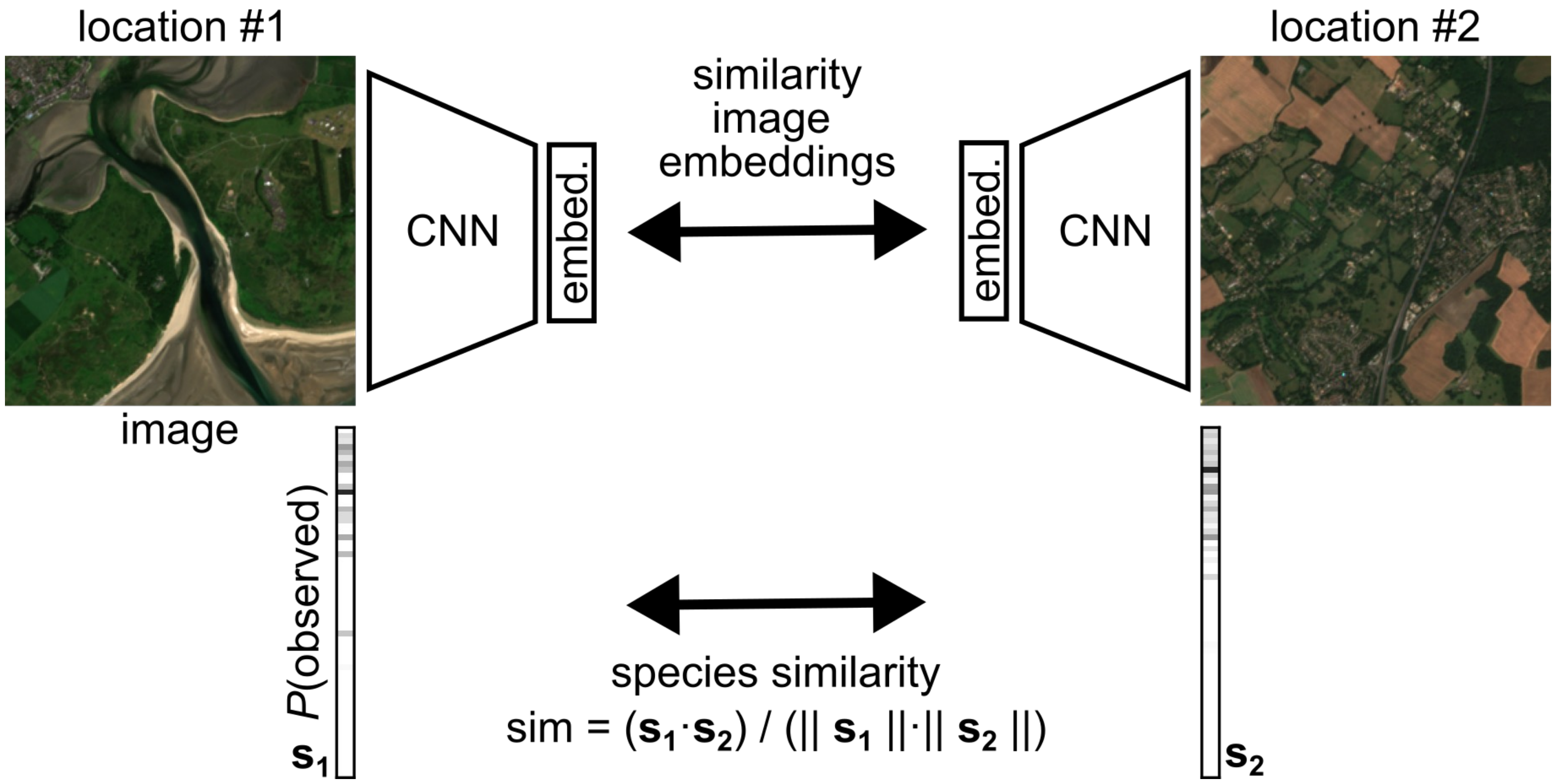
Van der Plas et al., 2025, *CVPRW proc.*

PECL: Paired Embeddings Contrastive Learning



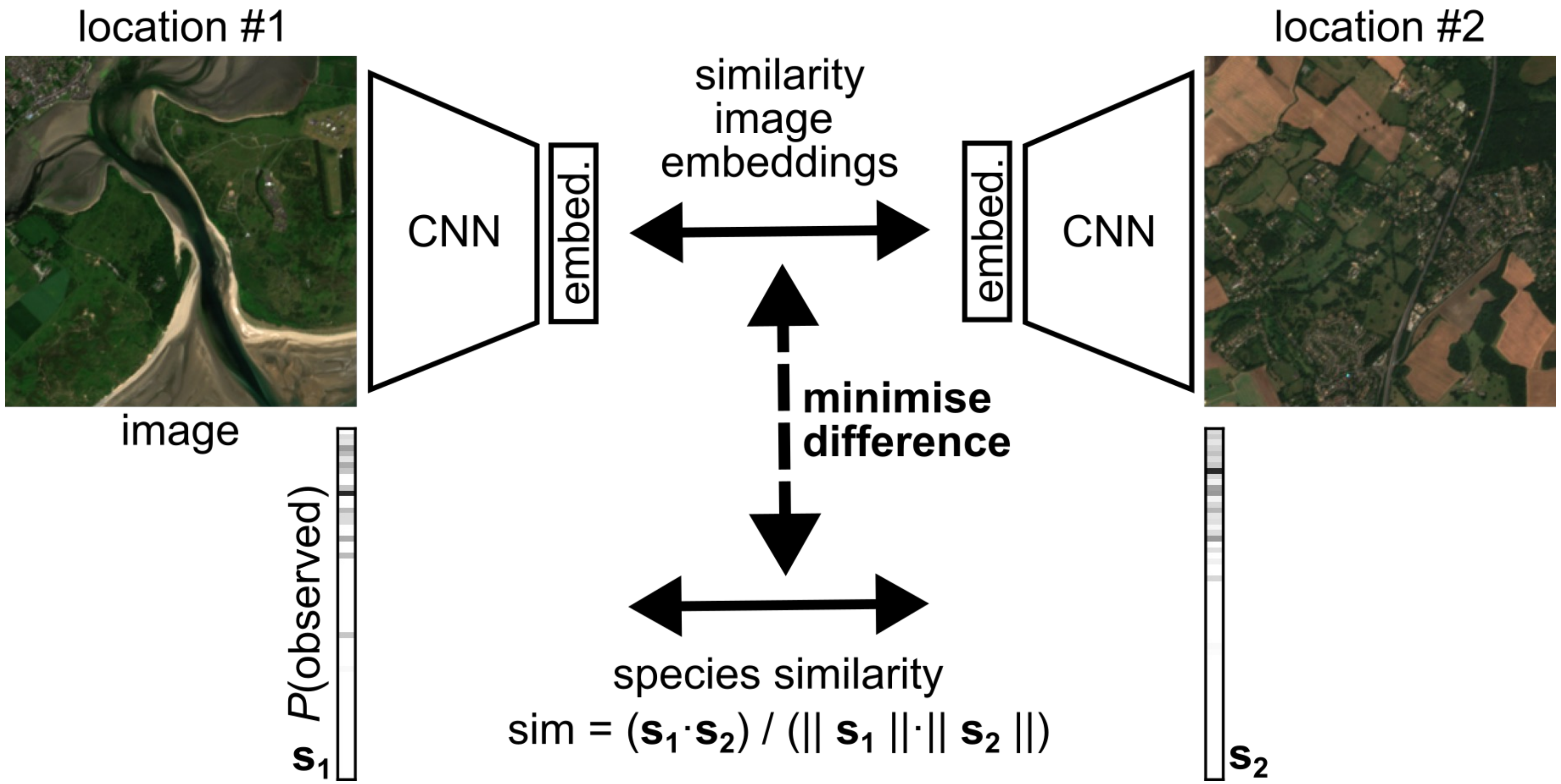
Van der Plas et al., 2025, *CVPRW proc.*

PECL: Paired Embeddings Contrastive Learning



Van der Plas et al., 2025, *CVPRW proc.*

PECL: Paired Embeddings Contrastive Learning



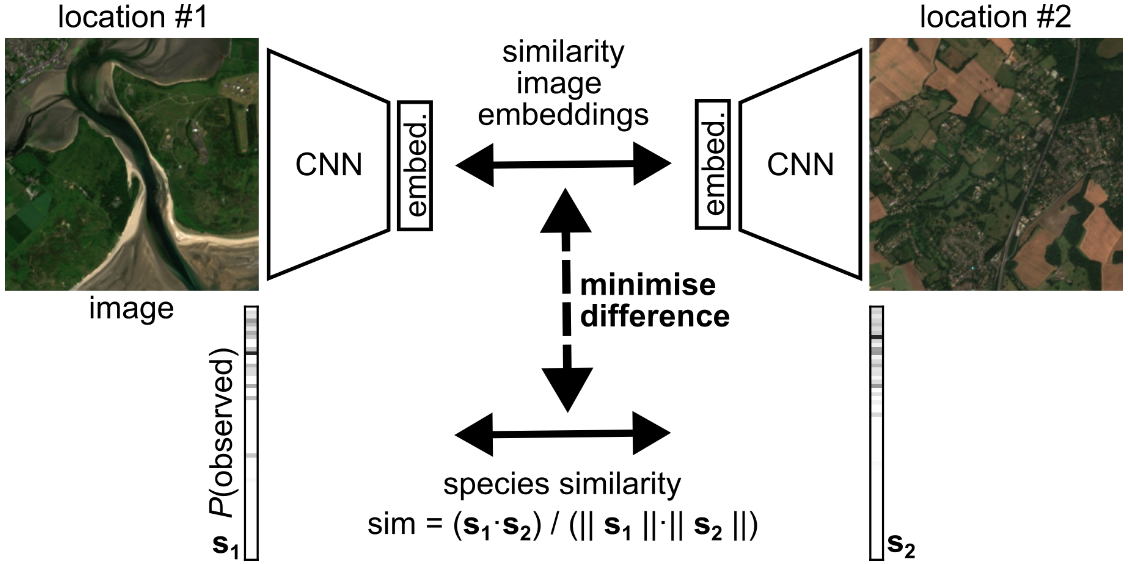
Van der Plas et al., 2025, *CVPRW proc.*

Predicting with CNN + contrastive regularisation

Pre-training or regularisation?

Challenge: more hyperparameters to optimise.

PECL improves top-10 accuracy by 35%.



Model	Top-10 [%]	Top-5 [%]	MSE [1e-02]
Mean rate	67.3	58.7	1.39
Unfrozen 3-layer SeCo	68.2 ± 0.2	61.8 ± 0.8	1.32 ± 0.03
Frozen 3-layer SeCo	69.6 ± 0.2	62.4 ± 1.0	1.21 ± 0.04
Frozen 3-layer SeCo + CR (best grid search)	70.0 ± 0.2	62.3 ± 0.6	1.21 ± 0.03
Frozen 3-layer SeCo + CR (best random search)	$\mathbf{70.4 \pm 0.2}$	$\mathbf{63.0 \pm 0.9}$	$\mathbf{1.20 \pm 0.03}$

Van der Plas et al., 2025, *CVPRW proc.*

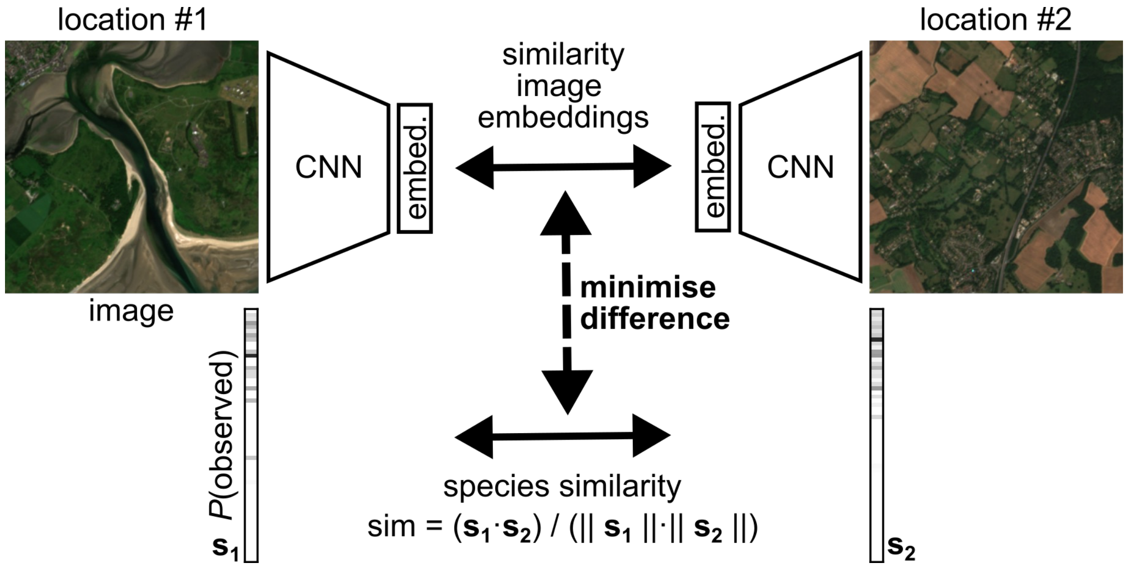
Predicting with CNN + contrastive regularisation

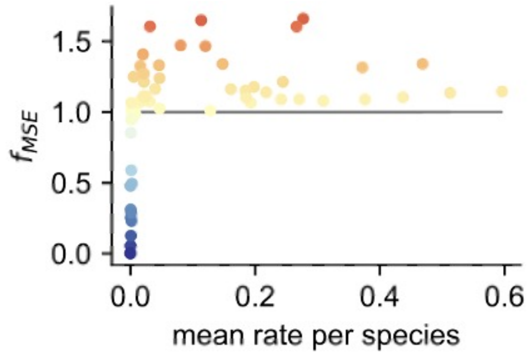
Pre-training or regularisation?

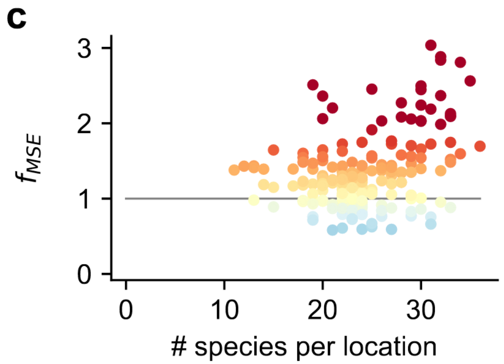
Challenge: more hyperparameters to optimise.

PECL improves top-10 accuracy by 35%.

$f_{MSE} = MSE_{baseline} / MSE_{model}$







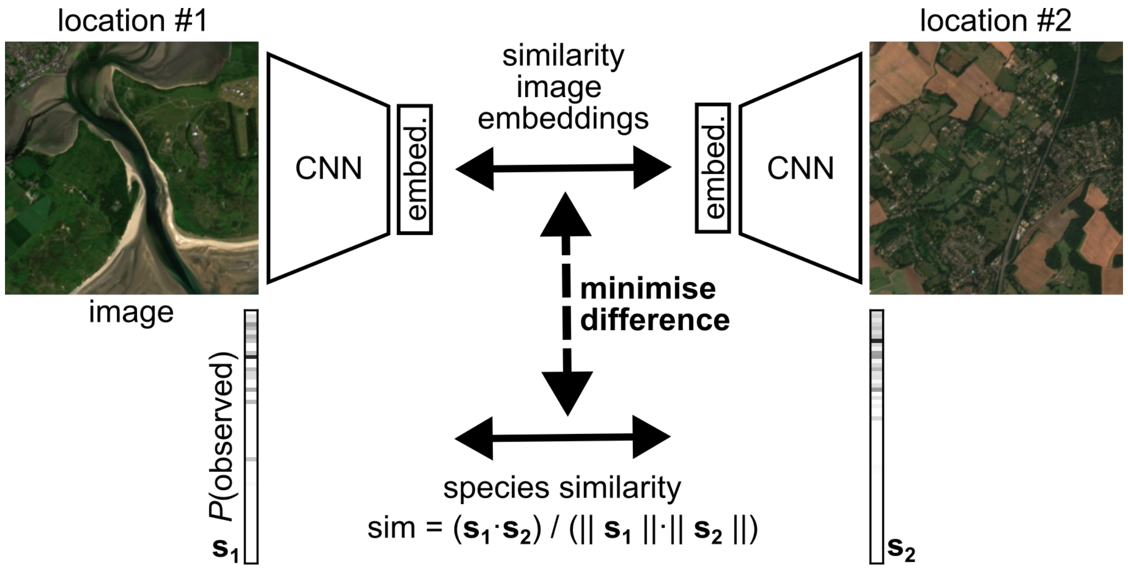
Van der Plas et al., 2025, *CVPRW proc.*

Outlook

Use bigger SatBird data set to **pre-train** using contrastive learning?

Integrate more **data modalities** (sensors, long-term monitoring, ..).

Derive comprehensive metrics of **habitat condition** (indicator species, XAI, ..).



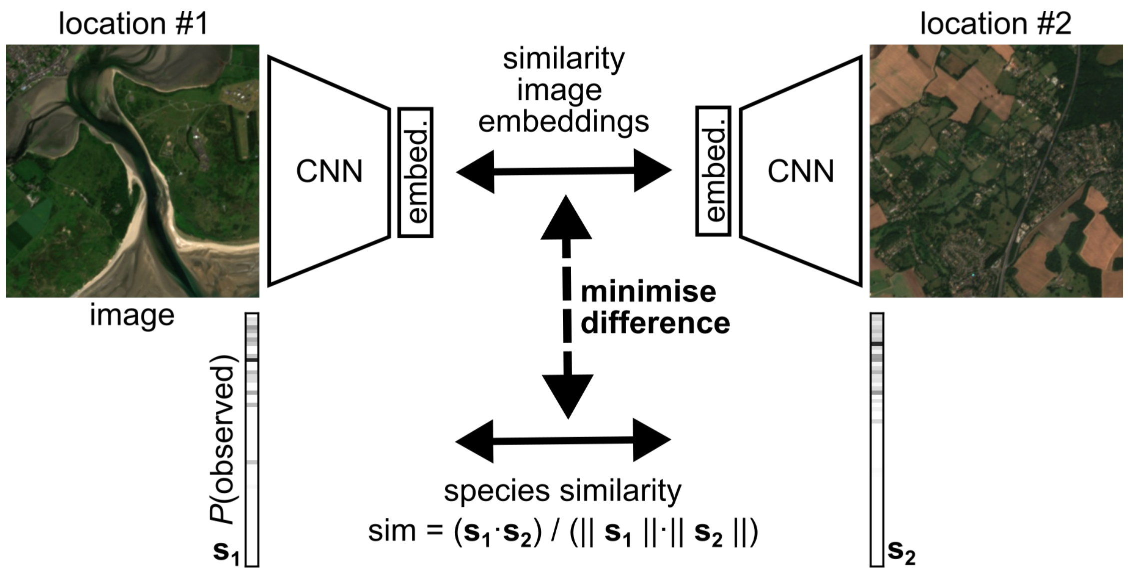
Outlook

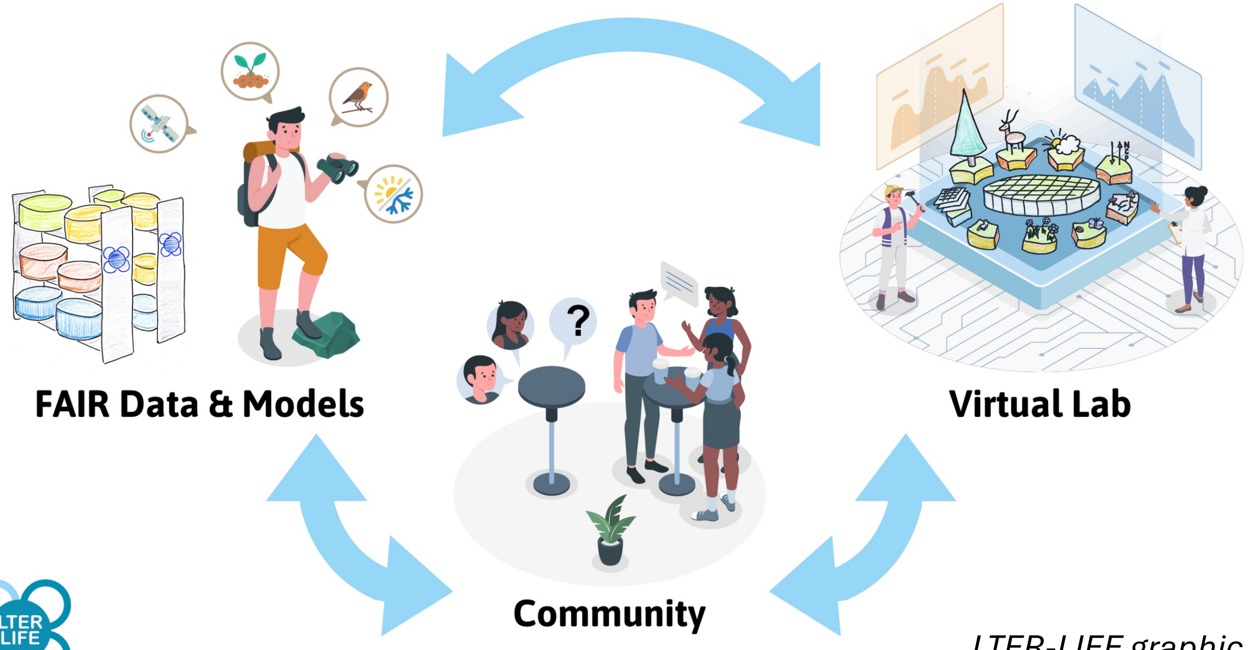
Use larger SatBird data to **pre-train** using contrastive learning?

Integrate more **data modalities** (sensors, long-term monitoring, ..).

Derive comprehensive metrics of **habitat condition** (indicator species, XAI, ..).

Feed into **ecosystem digital twins** for scenario forecasting.





LTER-LIFE graphic

Acknowledgements

- David Alexander (PDNPA)
- Simon Geikie (PDNPA)
- Daniel Simms (Cranfield)
- Matt Fry (Turing/CEH)
- **Michael Pocock (CEH)**
- Stephen Law (UCL)
- Ioannis Athanasiadis (WUR)

High-resolution land cover mapping:
Van der Plas, Geikie, Alexander*, Simms*, 2023, *Remote Sensing,* **15**, 5277.

ML for monitoring PAs perspective:
Van der Plas, Alexander, Pocock, 2025, *Ecol Solut Evid.*, **6**, e70040.

Predicting butterfly species with contrastive learning:
Van der Plas, Law, Pocock, 2025, *CVPR workshop proceedings,* ***PAGES***

thijs.vanderplas@wur.nl
vdplasthijs.github.io