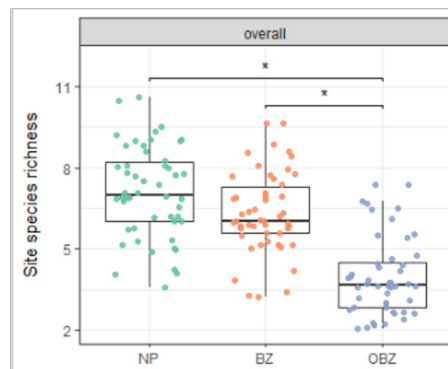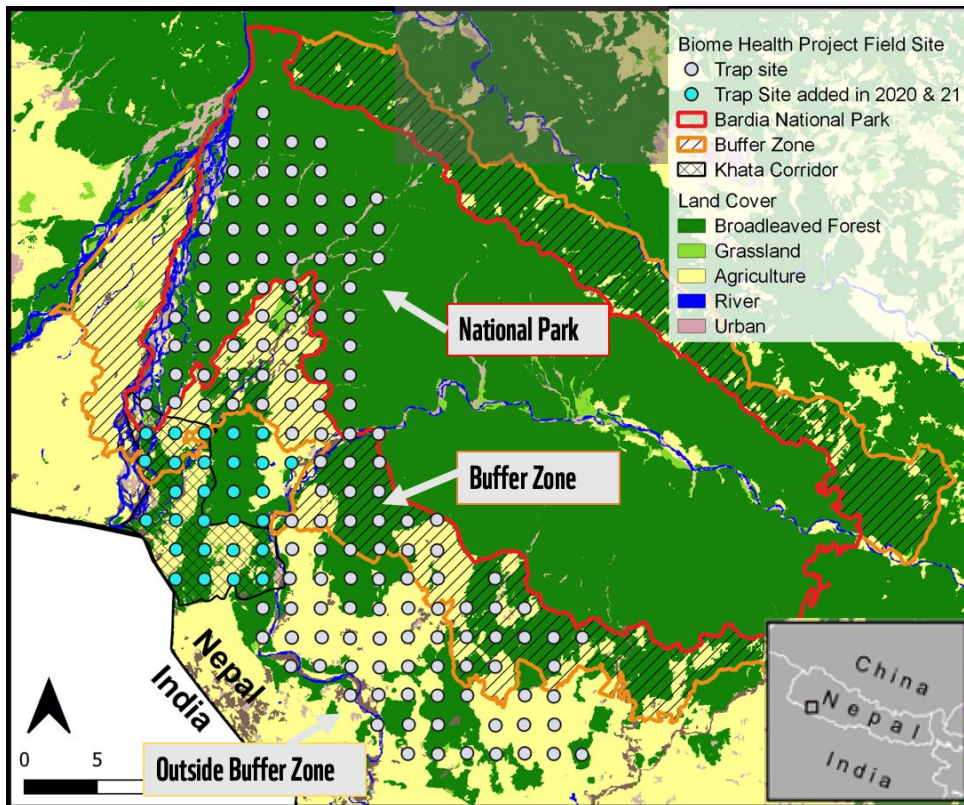# Can we infer ecological health from unlabelled data?

Peggy Bevan
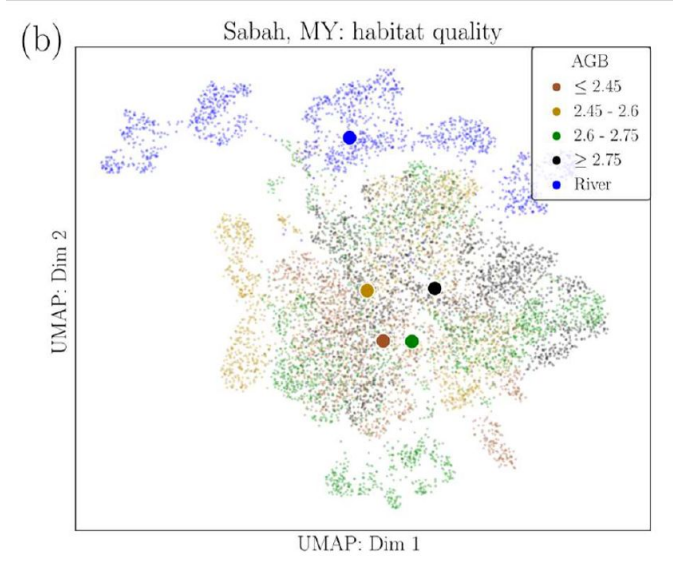University College London, Zoological Society of London

Background

Ferreira et al., *in prep*

Analysing passively collected data is a lengthy process...

Labelling 50,000 images = 3 months with 3 people
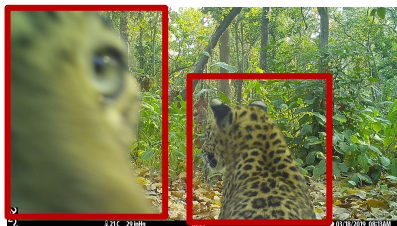
My dataset = Nearly 3.5 million images!

Predicted: chital 92.62%
Actual: goat
Image Id: 2019 BZ36 041132 - box4

Sethi *et al.,* 2020

...But can we infer habitat health from unlabeled image & acoustic data?

Plotting feature vectors (embeddings)
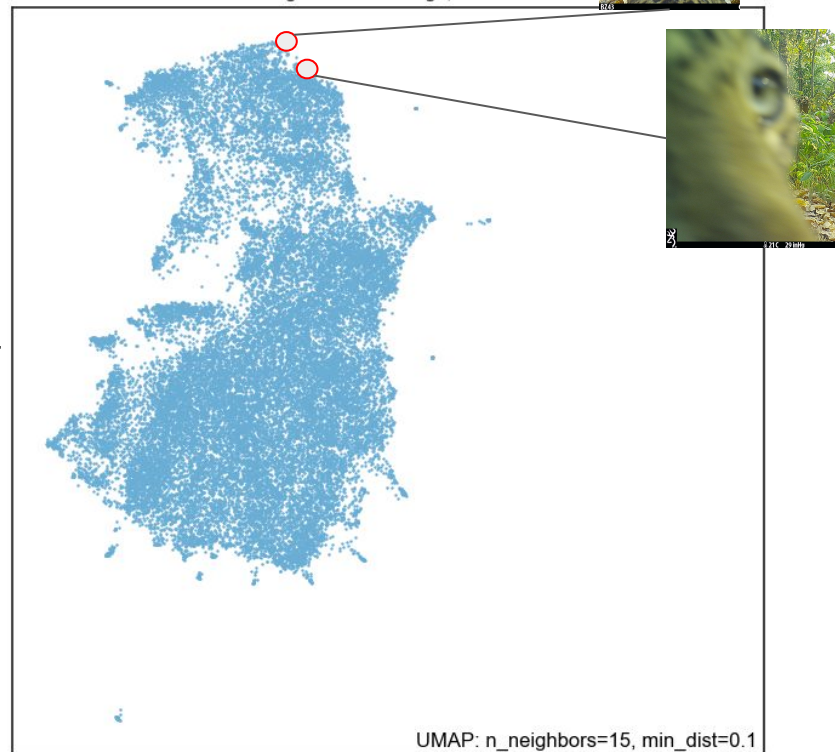
Input Image

Cropped to
animal

Run through
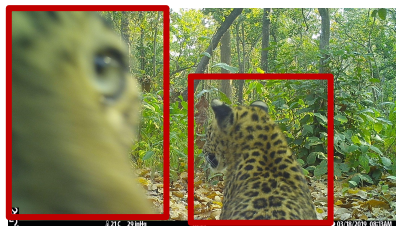Pre-trained CNN, with
final layer removed

ResNet50

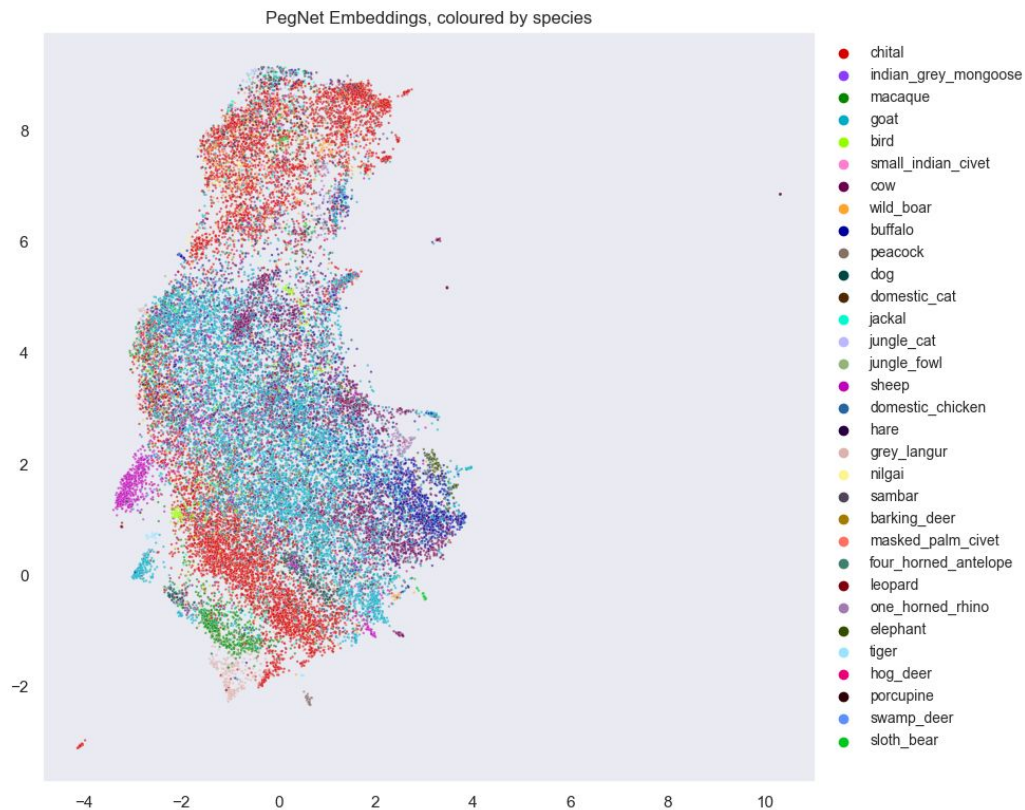Embedding   [0.7, 0.3, 0.88….0.9] (x2048)

PegNet Embeddings, UMAP

UMAP: n_neighbors=15, min_dist=0.1

Plotting feature vectors (embeddings)

PegNet

[0.7, 0.3, 0.88….0.9] (2048)

PegNet Embeddings, coloured by species

- chital
- indian_grey_mongoose
- macaque
- goat
- bird
- small_indian_civet
- cow
- wild_boar
- buffalo
- peacock
- dog
- domestic_cat
- jackal
- jungle_cat
- jungle_fowl
- sheep
- domestic_chicken
- hare
- grey_langur
- nilgai
- sambar
- barking_deer
- masked_palm_civet
- four_horned_antelope
- leopard
- one_horned_rhino
- elephant
- tiger
- hog_deer
- porcupine
- swamp_deer
- sloth_bear

Embeddings reveal paths to classification

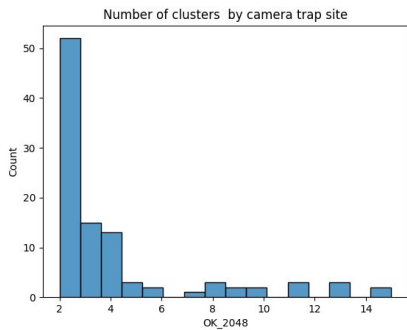PegNet = ResNet50,
Trained on ImageNet

Time of day

Protected area

Does the distribution of embeddings in feature space correspond to species richness?
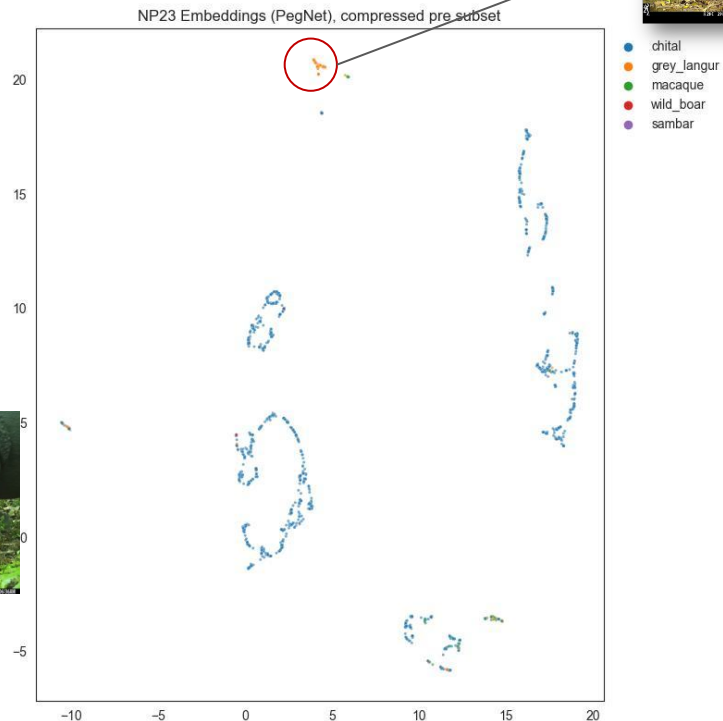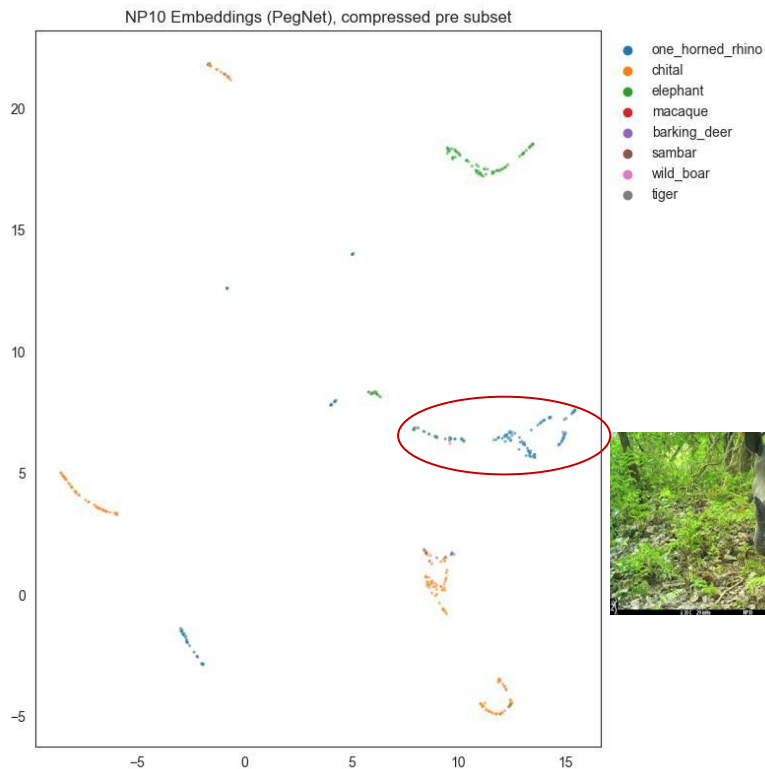
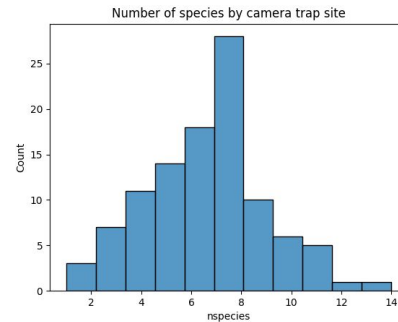# K-means and HDBSCAN clustering both under-fitting for species richness

Ground-truth reminder

K-means
with
Silhouette
Index

HDBSCAN
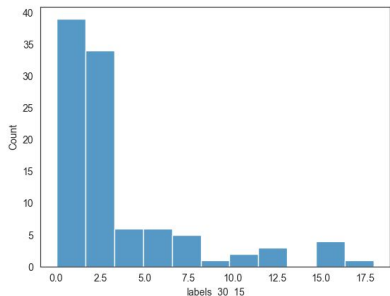
Clustering method	Model	
	Trained on ImageNet	Trained on Masai Mara CT images
K-Means	0.30	0.35
HDBSCAN	0.49	0.48
K-Means, day time only	0.28	0.38
HDBSCAN, day time only	0.45	0.46

Next steps...

- Maybe clustering algorithms are struggling to find clusters in data because they are not there.
- Could a more supervised approach work?

Lessons learnt...

- Completely unsupervised approaches are very ambitious
- AI can't solve everything!
- Be patient, start simple!!

Questions?
🐦 @PeggyBevan2