Machine learning models in disease management

Chris Banks

Roslin Institute, University of Edinburgh, UK

AI for Next Generation Smart Animal Breeding

August 4th 2025, Roslin Institute

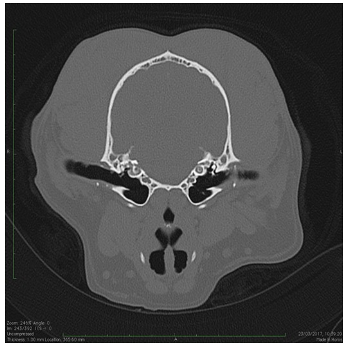
THE UNIVERSITY of EDINBURGH
The Royal (Dick) School
of Veterinary Studies

Biotechnology and
Biological Sciences
Research Council

Three projects. Three ML problems.

1

Deep Learning for diagnosis from CT scans
- Canine cranial CTs
- DL to to diagnose Otitis Media (middle ear fluid)
- *Small training set (~600 images)*



2

Machine Learning augmented diagnostic tests
- Skin test for bTB has problems with sensitivity
- ML to predict herd breakdown risk and augment test results
- *Complex correlated features, temporal evolution, missing data*

3

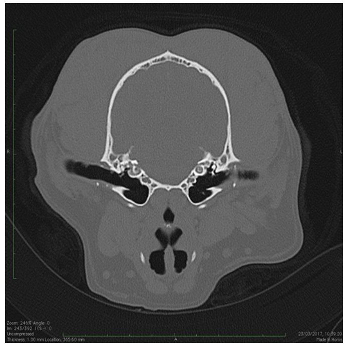
Impact of land management on disease transmission risk
- Predicting change in land use and affect on wildlife proximity to farms
- ML to predict wildlife species presence in land parcels
- *Predictions based on sparse and biased observations*

Three projects. Three ML problems.

1

Deep Learning for diagnosis from CT scans

- Canine cranial CTs
- DL to to diagnose Otitis Media (middle ear fluid)
- *Small training set (~600 images)*



2

Machine Learning augmented diagnostic tests

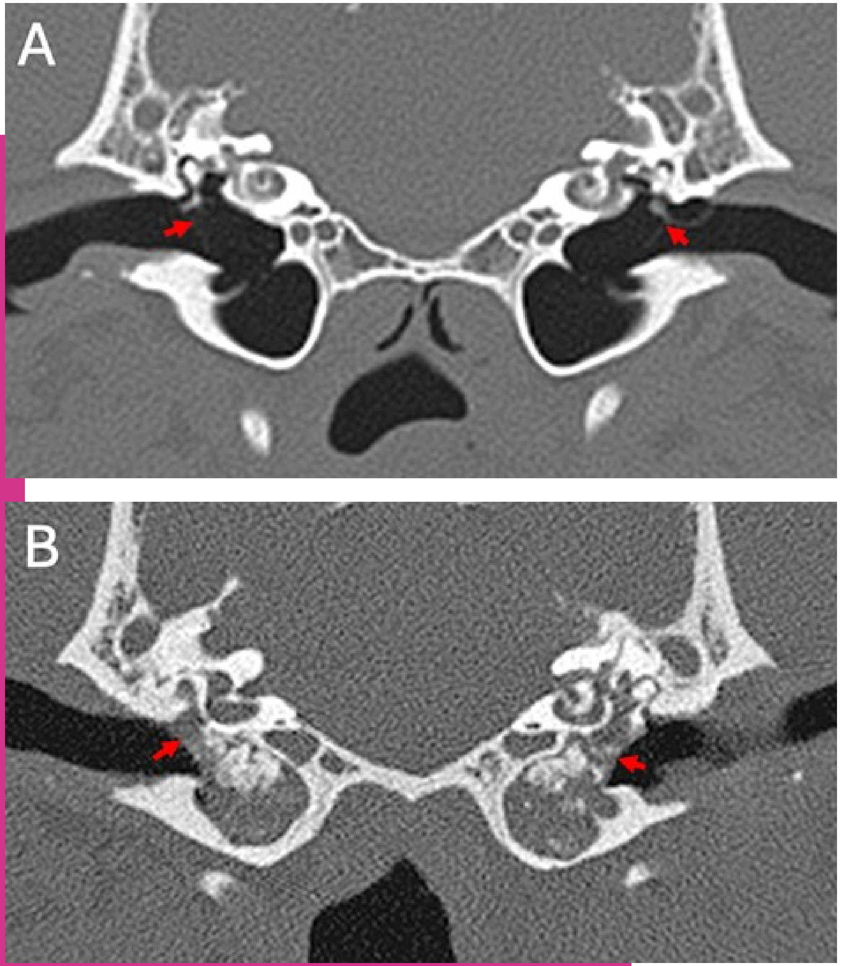
- Skin test for bTB has problems with sensitivity
- ML to predict herd breakdown risk and augment test results
- *Complex correlated features, temporal evolution.*

3

Impact of land management on disease transmission risk

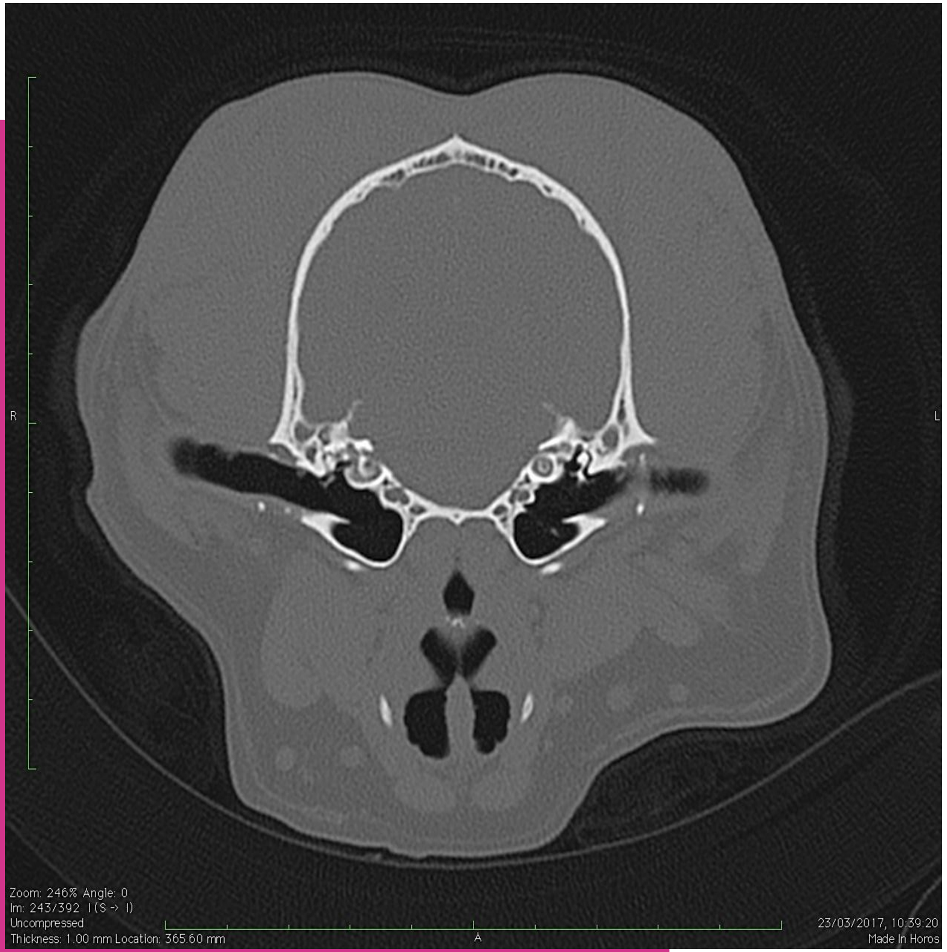
- Predicting change in land use and affect on wildlife proximity to farms
- ML to predict wildlife species presence in land parcels
- *Predictions based on sparse and biased observations*

Deep Learning diagnostics on CT scans



- We attempted to leverage existing image records in the Hospital for Small Animals.
- Started with a *relatively* simple task:
  ← Identifying middle ear fluid.
- Canine cranial CTs
  535 patients: 402 normal, 133 diseased.
- Tested a range of techniques for minimising the impact of a small training set

Deep Learning diagnostics on CT scans



- Convolutional Neural Networks with:
  - Data Augmentation
    - Static / Dynamic
  - Class Weighting
  - Oversampling
  - Pre-trained models
    - Feature extractor / Fine tuned

ROSLIN

Deep Learning diagnostics on CT scans

Model	DA	CW	OS	Accuracy	Sensitivity	Specificity	AUC
Baseline				77.78%	0.583	0.972	0.87
FT_01		Y	Y	75.00%	0.556	0.944	0.86
FT_02	St	Y	Y	80.56%	0.667	0.944	0.86
FT_03	Dy		Y	81.94%	0.667	0.972	0.89
FT_04	Dy	Y		79.17%	0.778	0.806	0.89
FT_05	**Dy**	**Y**	**Y**	**84.72%**	**0.722**	**0.972**	**0.88**
FE_01		Y	Y	73.61%	0.556	0.917	0.81
FE_02	St	Y	Y	75.00%	0.500	1.000	0.79
FE_03	Dy		Y	70.83%	0.444	0.972	0.76
FE_04	Dy	Y	Y	73.61%	0.472	1.000	0.81

Three projects. Three ML problems.

1

Deep Learning for diagnosis from CT scans
- Canine cranial CTs
- DL to to diagnose Otitis Media (middle ear fluid)
- *Small training set (~600 images)*

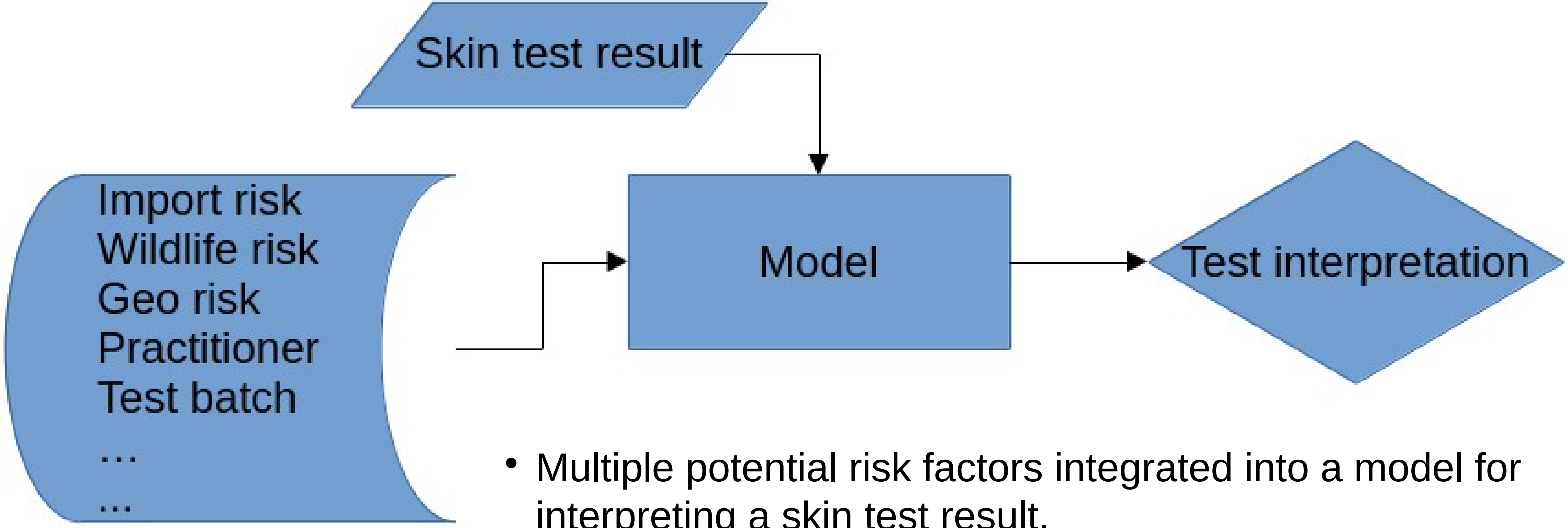
2

Machine Learning augmented diagnostic tests
- Skin test for bTB has problems with sensitivity
- ML to predict herd breakdown risk and augment test results
- *Complex correlated features, temporal evolution, missing data*

3

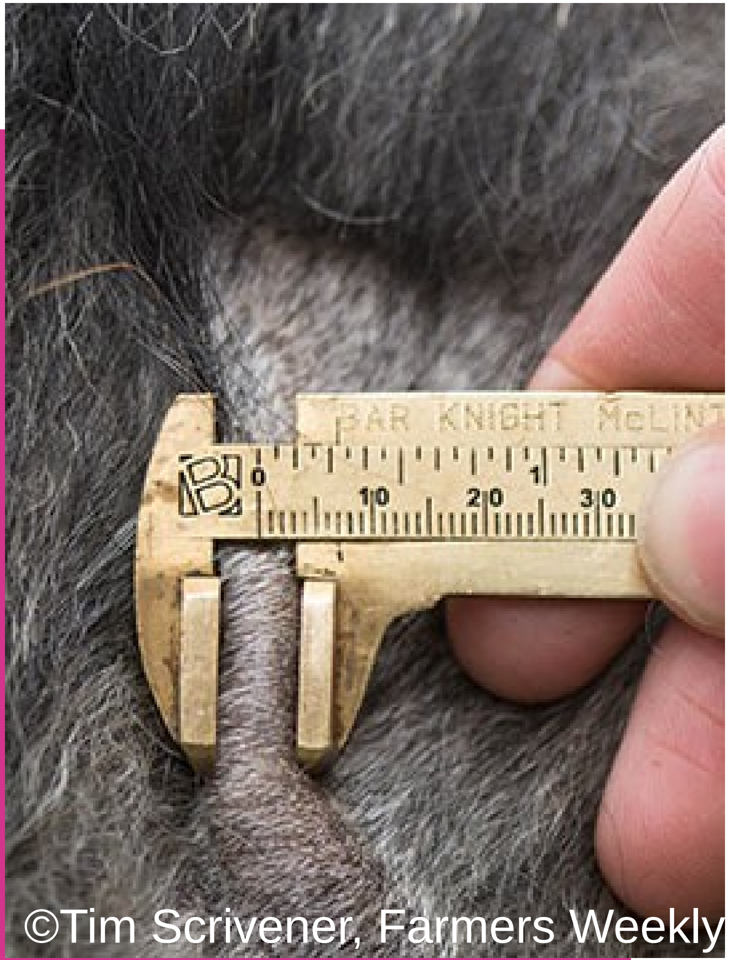
Impact of land management on disease transmission risk
- Predicting change in land use and affect on wildlife proximity to farms
- ML to predict wildlife species presence in land parcels
- *Predictions based on sparse and biased observations*

Machine learning for bTB diagnostics



- Multiple potential risk factors integrated into a model for interpreting a skin test result.
- **Predicts risk of breakdown from test/herd metadata. Used to augment test result.**
- Also gives us an indication of “most important” risk factors.

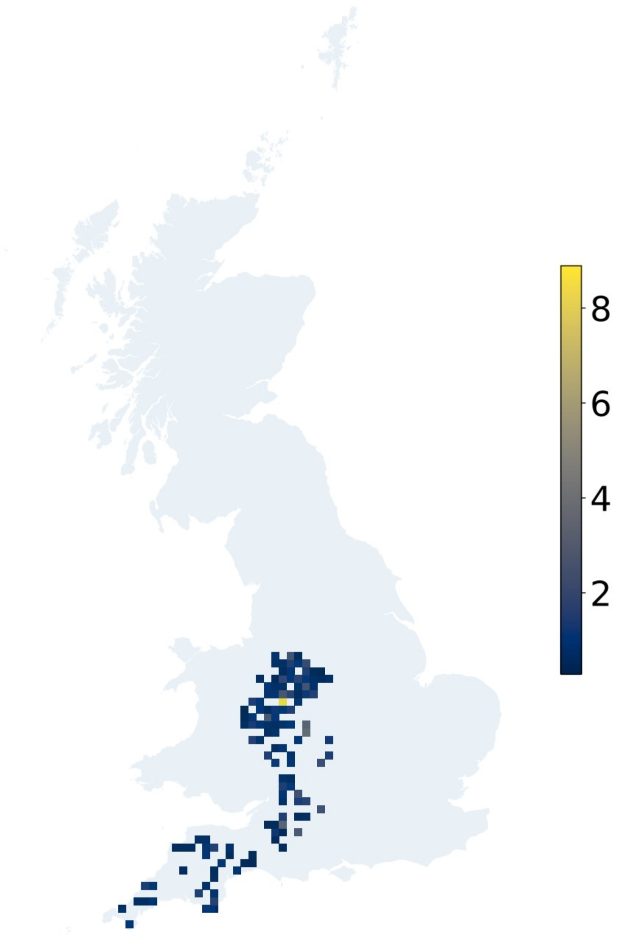
Machine learning for bTB diagnostics



- Problems to consider:
  - Temporal data set (time series of test records)
  - Missing data and left-censoring
  - Highly correlated features
- Means to mitigate these:
  - Temporal cross-validation
  - Histogram Gradient Boosting Trees
  - SHAP for feature importance

Machine learning for bTB diagnostics

% of tests early detected by area (in 2020)



- Results:
  - Herd-level sensitivity increased 5.2%-points
  - This means **240 extra breakdown herds** caught by the model in one year (2020).
  - Result is equivalent to a modelled increase in individual test sensitivity of 12%.
  - SHAP really only confirms known bTB risk factors, but also provides indication these change over time.
- Bonus result:
  - Change to focus on Specificity and we can **catch around 5200 false positives in 2020**.

Three projects. Three ML problems.

1

Deep Learning for diagnosis from CT scans

- Canine cranial CTs
- DL to to diagnose Otitis Media (middle ear fluid)
- *Small training set (~600 images)*

2

Machine Learning augmented diagnostic tests

- Skin test for bTB has problems with sensitivity
- ML to predict herd breakdown risk and augment test results
- *Complex correlated features, temporal evolution.*

3

Impact of land management on disease transmission risk

- Predicting change in land use and affect on wildlife proximity to farms
- ML to predict wildlife species presence in land parcels
- *Predictions based on sparse and biased observations*

Assessing the potential impact of environmental land management schemes on emergent infectious disease risks

Land use/ Species distribution

Existing land use and species distributions are taken as input.

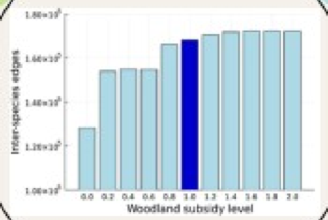
£

Economic model

Economic model of woodland subsidies predicts the conversion of existing land use into new woodland.

New species distribution

Species distribution model predicts the new distribution of wildlife given the changes in land use and new woodland.



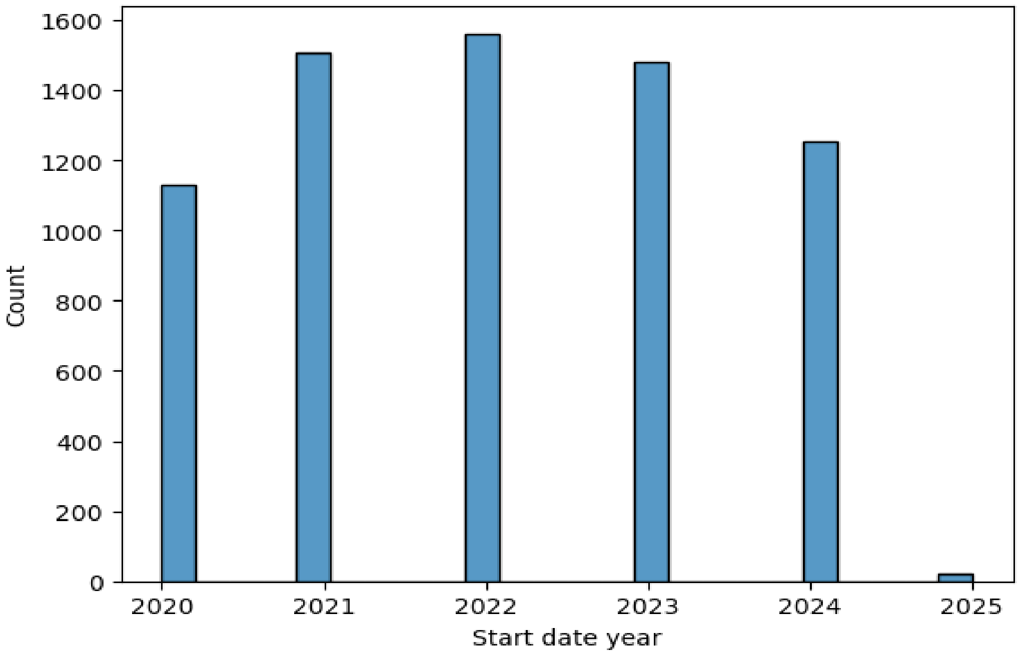
Wildlife/ livestock network

Proximity network models the relationship between land with wildlife and land with agricultural holdings.

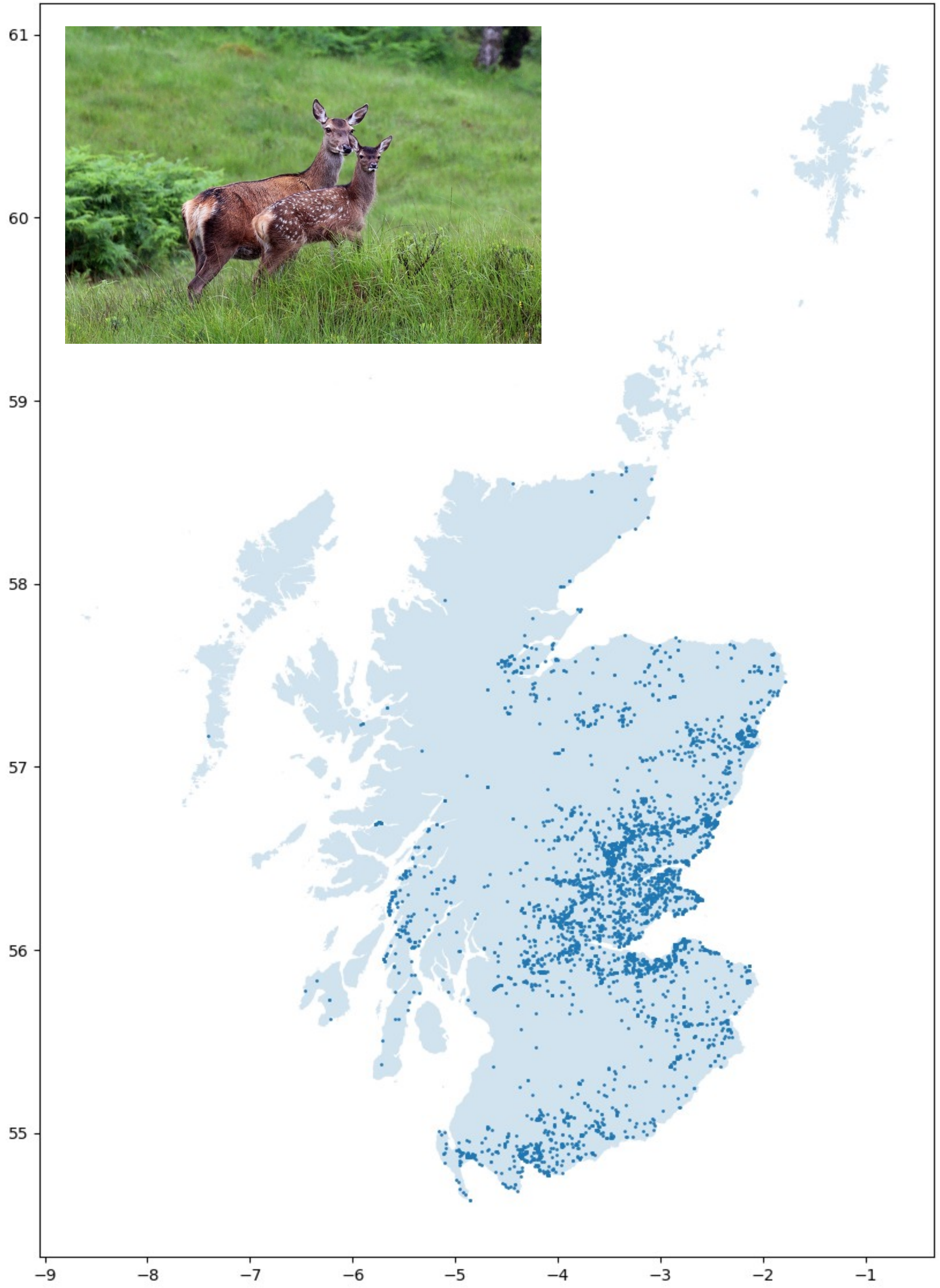
Disease transmission risk

Change in disease risk is estimated from the network before and after land use change.

- All NBN Atlas observations for each deer species in Scotland from >2020.

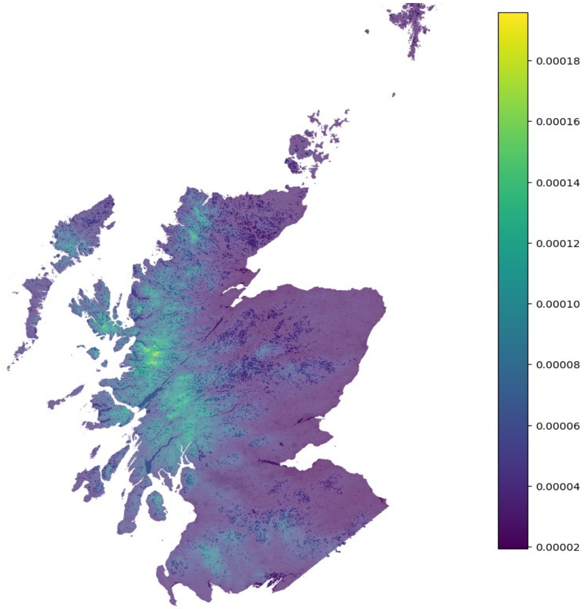


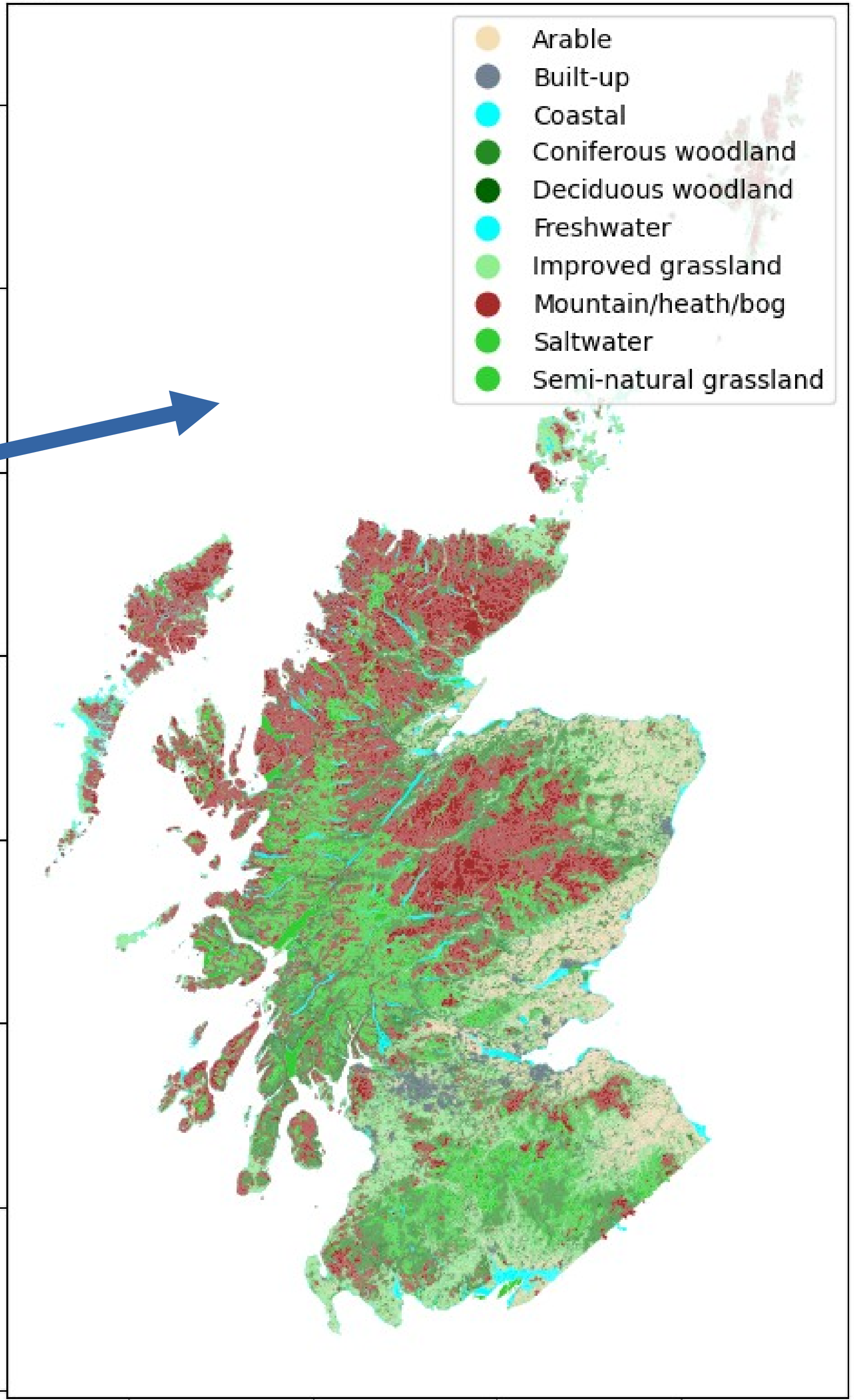
E.g.
Roe Deer
<----------->



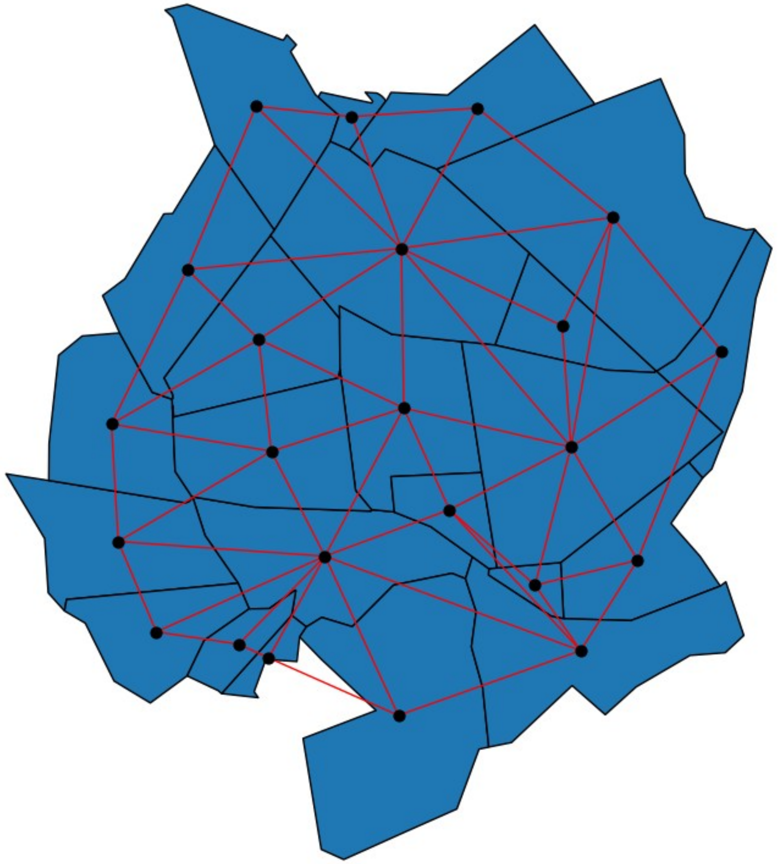
ROSLIN

- Presence observations mapped to Land Parcels
- Land parcel land use type
- CHESS-Met data (mean >2020)



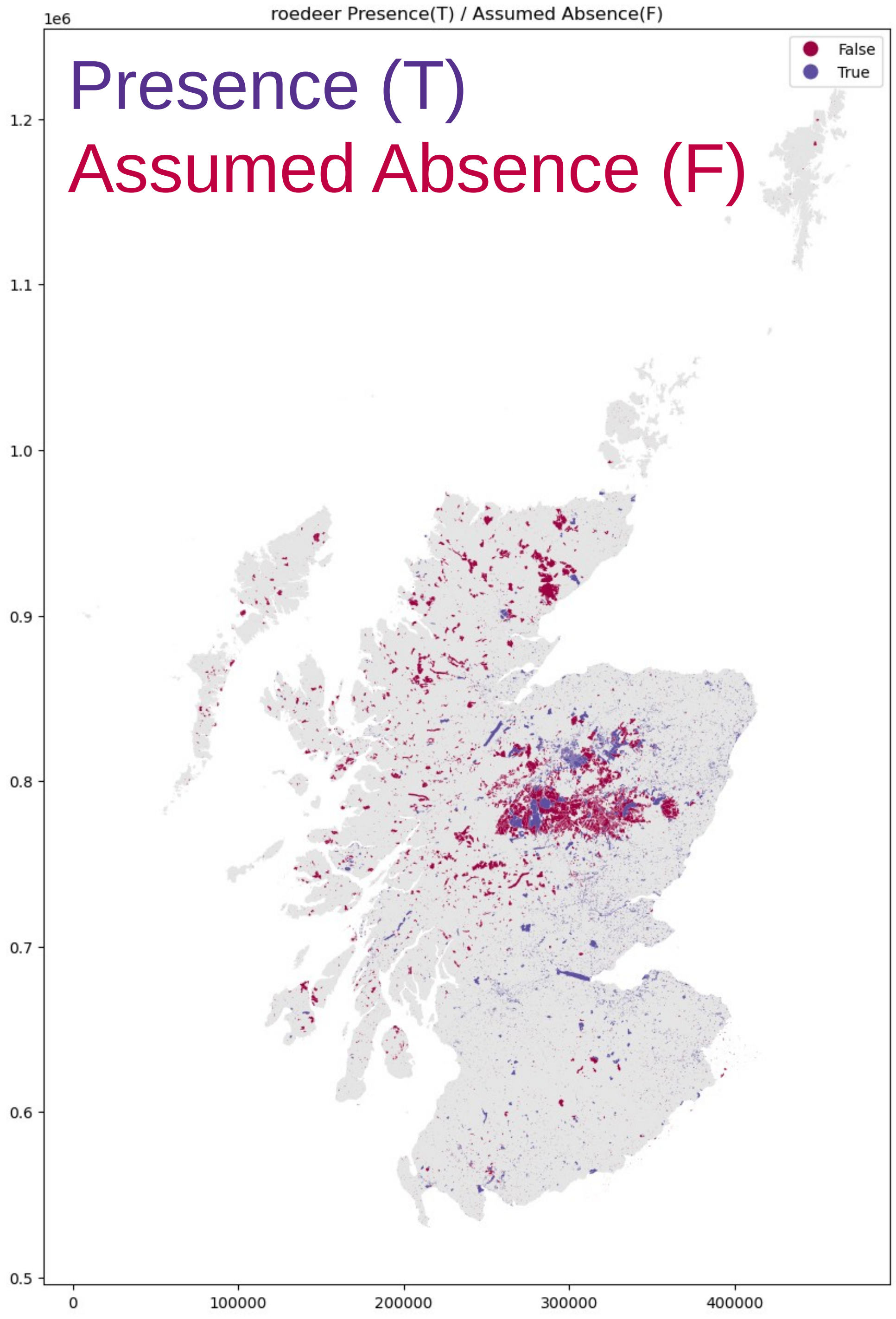


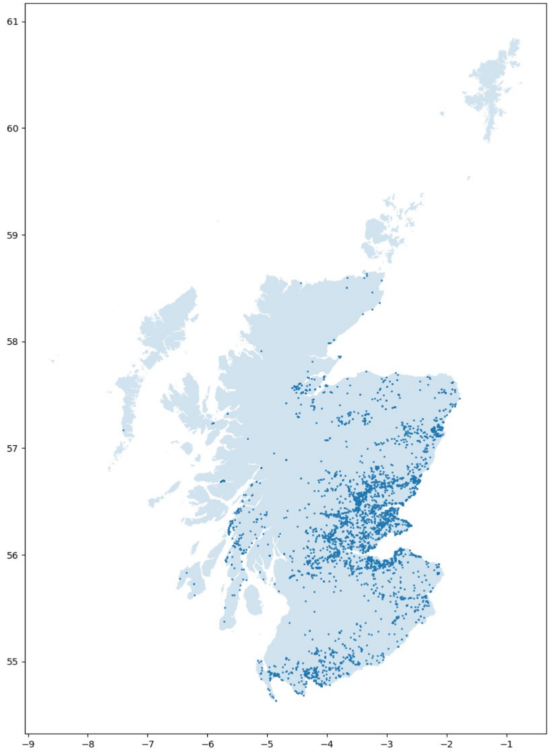
- Build network of land parcel adjacency
- Neighbouring parcel:
  - Land class
  - Observations
  - Presence of woodland



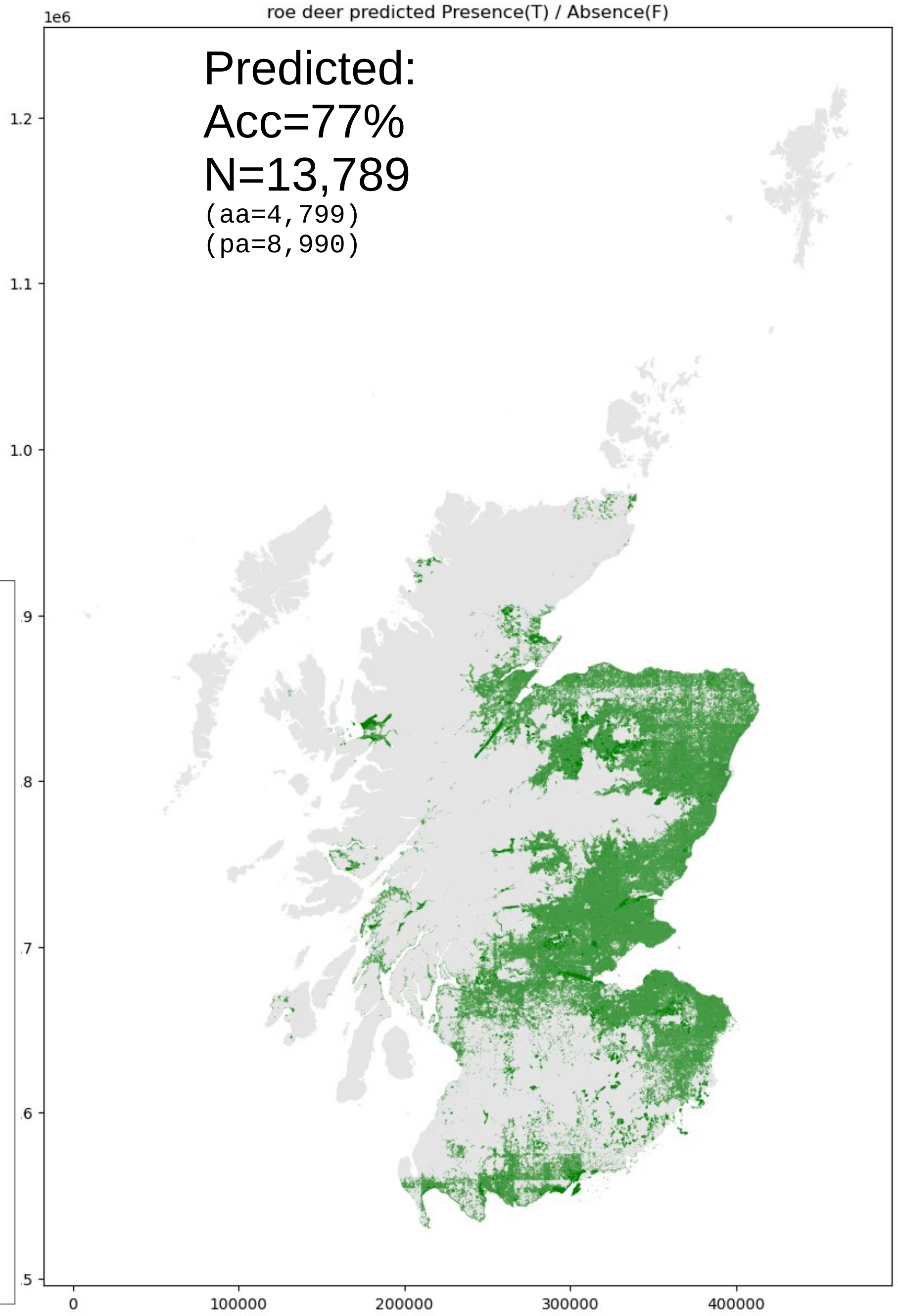
- Problem?
  - **Presence only** from observations
  - **“Assumed-Absence”** from no observation, but presence of other species.
  - |Absence sample| = |Presence|
  - Uniform Random Pseudo-Absence where insufficient A-As.

- Histogram Gradient Boosting Tree model
- Hyper-parameter randomised search with 10-fold CV.

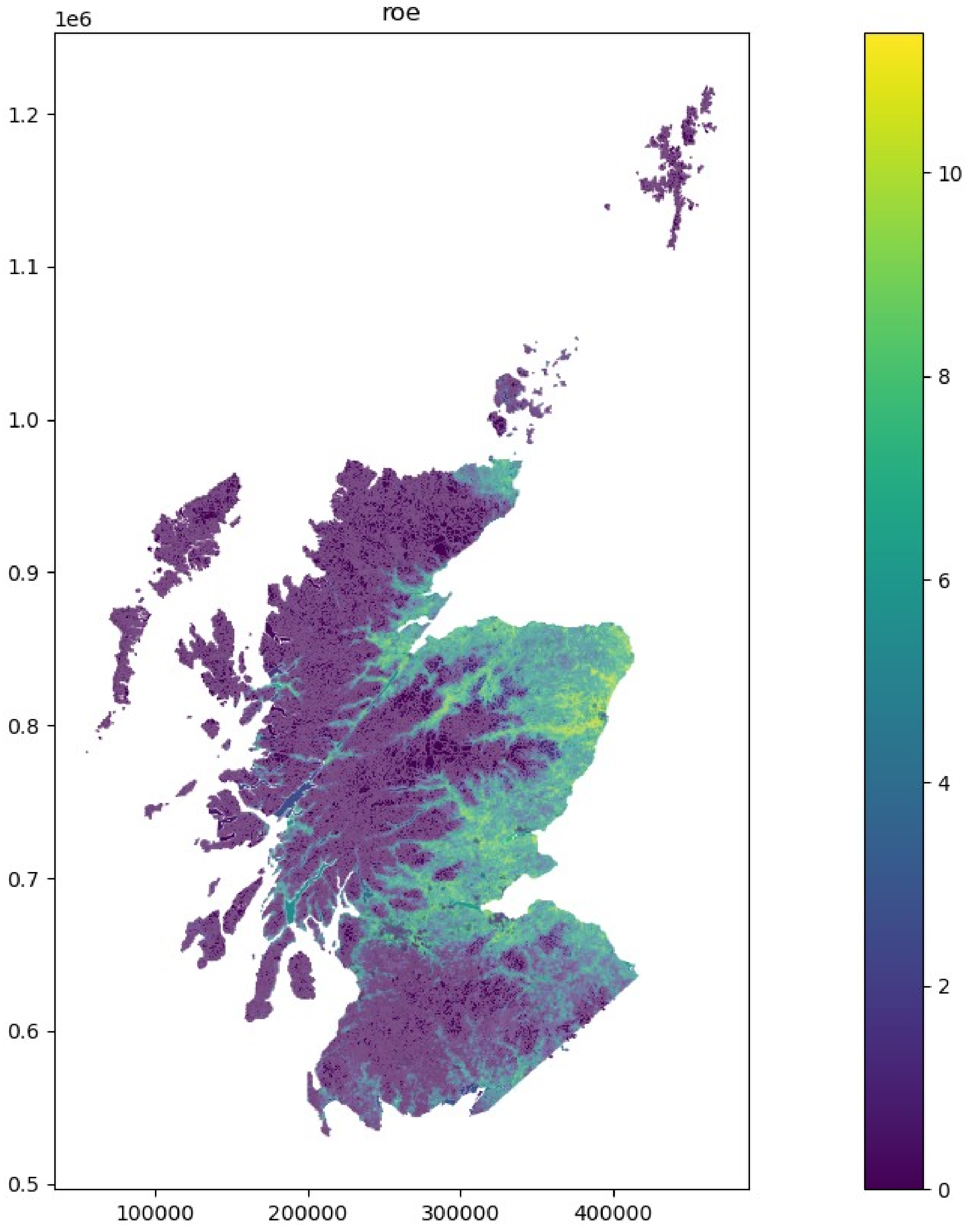




61
60
59
58
57
56
55
−9
−8
−7
−6
−5
−4
−3
−2
−1



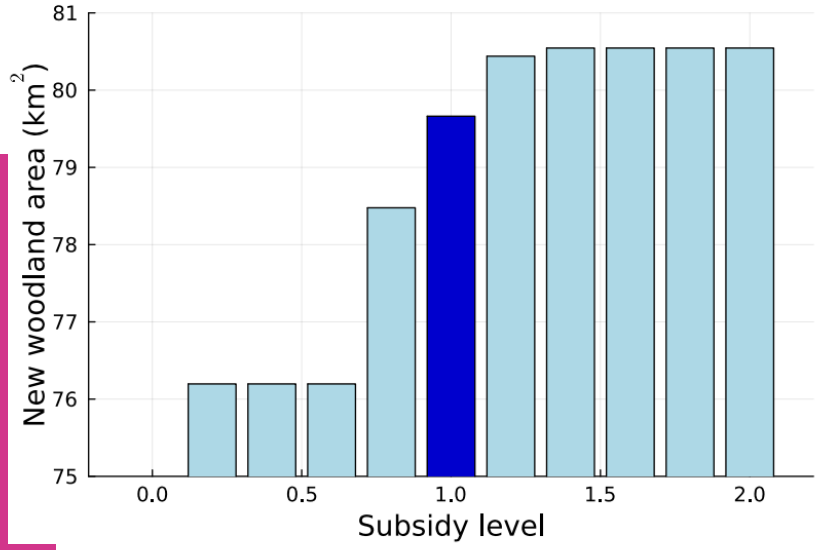
roe deer predicted Presence(T) / Absence(F)
Predicted:
Acc=77%
N=13,789
(aa=4,799)
(pa=8,990)
1e6
1.2
1.1
1.0
9
8
7
6
5
0
100000
200000
300000
400000



roe
1e6
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
100000
200000
300000
400000
10
8
6
4
2
0

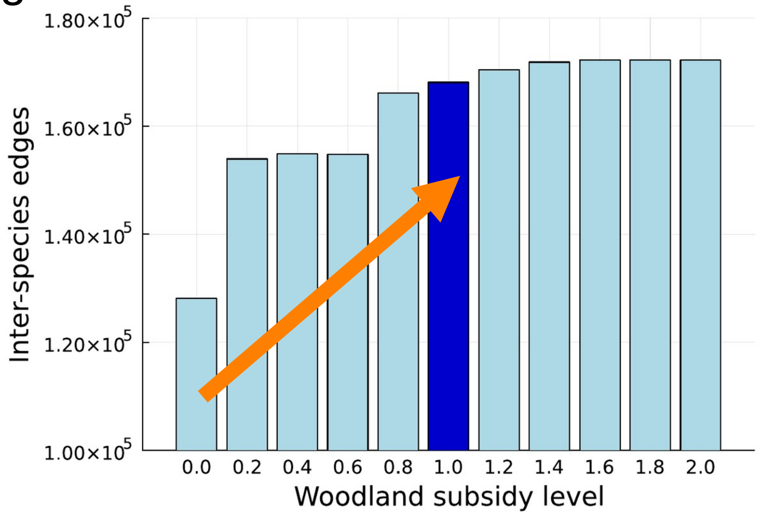
How can land management decisions affect disease?

- Area of **woodland created only varies by around 3%** between the low and high subsidy scenarios.
- But can allow up to a 57% increase in deer population.
- Overall **26%--35% increase in contact risk** between cattle and deer, depending on the level of subsidy provided.



Modest increase in woodland area

Significant increase in deer/cattle contact



- Small image set training:
  - Augmentation, weighting, oversampling, and fully fine tuned pre-trained models.
- Left(/right)-censoring, missing, or categorical:
  - HGBT solves in a natural way.
- Highly correlated features:
  - Use SHAP for feature importance.
- Species psudo-absence:
  - Improved by other species observations.

ROSLIN

BBSRC
bioscience for the future

UK Research and Innovation

Department for Environment Food & Rural Affairs

Animal & Plant Health Agency

UK FARMCARE

Thanks!

- University of Edinburgh:
  - Roslin Institute:
    - **Rowland Kao**
    - **Aeron Sanchez**
  - R(D)SVS:
    - **Tobias Schwarz**
    - **Richard Mellanby**
  - School of Informatics:
    - **Oisin Mac Aodha**
    - **Zhixuan Zhao**
- University of Glasgow:
  - **Nick Hanley**
  - **Katherine Simpson**
- UK Farmcare:
  - **Kate Bowen**
  - **Vicki Stewart**
- APHA:
  - **Graham Smith**
  - **Oliver Tearne**