Using machine learning with wild bird reporting data to produce risk maps of Highly Pathogenic Avian Influenza in Britain and determine possible biases in the wild bird reporting

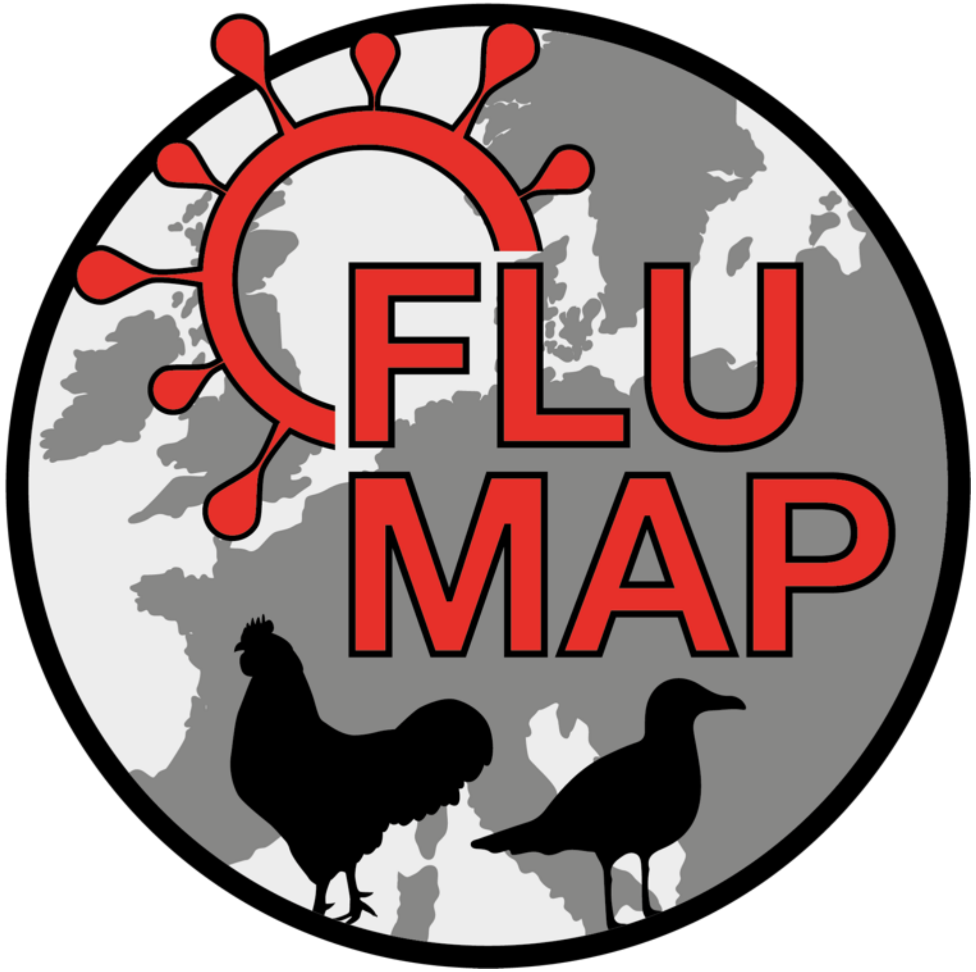
Aeron Sanchez[1], Alex Mastin[3], Stephen Vickers[4], Rowland R. Kao[1,2]

ROSLIN 1

THE UNIVERSITY of EDINBURGH
School of Physics
and Astronomy 2

Animal &
Plant Health
Agency 3

Royal
Veterinary
College
University of London 4



Animal &
Plant Health
Agency

Imperial College
London

THE Pirbright
INSTITUTE

UNIVERSITY OF LEEDS

RVC Royal
Veterinary
College
University of London

UNIVERSITY OF
CAMBRIDGE

ROSLIN

The University of
Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Funders

UKRI Biotechnology and
Biological Sciences
Research Council

defra Department for Environment
Food and Rural Affairs

Highly Pathogenic Avian Influenza – A Global Animal Pandemic

More than 50,000 wild birds in UK killed by avian flu - double previous estimates

H5N1 bird flu may lead to extinction of species as data revealed by Guardian shows worst losses in decades

National Trust rangers clear dead birds from one of the Farne Islands off Northumberland, where at least 6,000 birds have died from avian influenza. Photograph: Owen Humphreys/PA

Alarm as HPAI kills almost 9,000 sea creatures in Chile

Mark Clements: Over 8,887 sea creatures have been recorded dead on the Chilean coast so far this year due to infection with the highly pathogenic avian influenza (HPAI) virus.

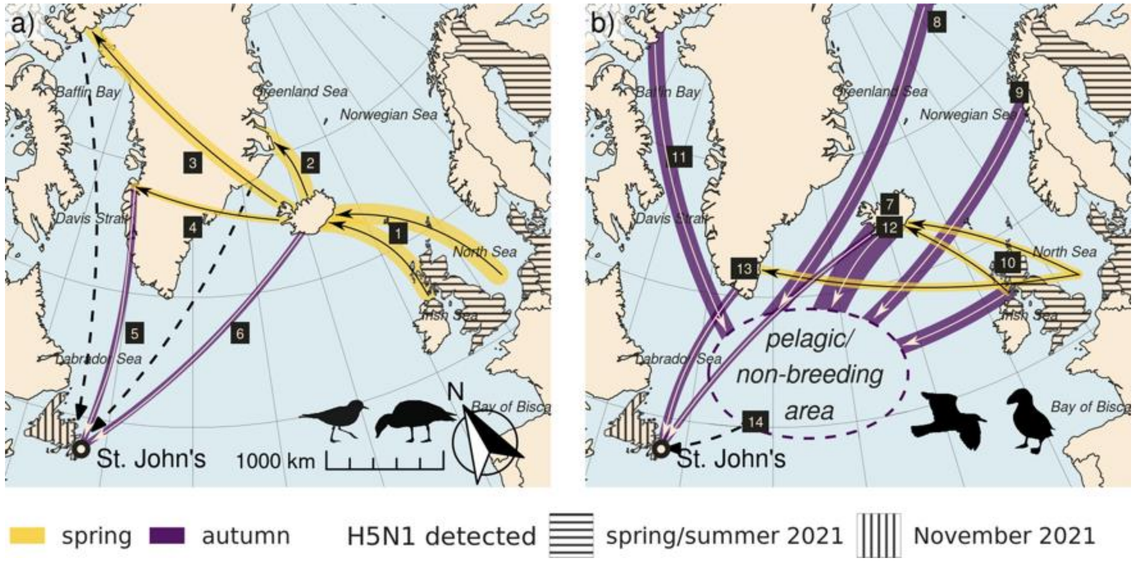
Mark Clements
May 30, 2023

The Guardian
Newspaper of the year

Bird flu: H5N1 virus in Brazil wild birds prompts animal health emergency

Health declaration to last 180 days, as world's biggest exporter of chicken meat detects virus for first time ever

Transatlantic spread of highly pathogenic avian influenza H5N1 by wild birds from Europe to North America in 2021



The UK's largest avian flu outbreak has left millions of birds dead - and scientists extremely concerned

NATIONAL GEOGRAPHIC

2022 has seen a deadly and highly pathogenic new strain of H5N1 bird flu make a cataclysmic impact on captive and wild bird populations. Conservationists and scientists are counting the cost.

BBC NEWS

Record avian flu outbreak sees 48m birds culled in UK and EU

More than 50,000 wild birds in UK killed by avian flu – double previous e

H5N1 bird flu may lead to extinction of species as data revea
Guardian shows worst losses in decades

Alarm as HPAI kills almost 9,000 sea

labanimal

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › lab animal › news & views › article

News & Views | Published: 14 August 2024

Infectious disease

The highly pathogenic H5N1 virus found in U.S. dairy cattle has some characteristics that could enhance infection and transmission among mammals

Fabien Filaire & Sander Herfst

Lab Animal (2024) | Cite this article

745 Accesses | **88** Altmetric | Metrics

A highly pathogenic H5N1 influenza virus is spreading in U.S. dairy cattle and has been transmitted to other species, including humans, probably through contaminated milk. Understanding how the virus spreads among cattle and its potential for mammalian adaptation and airborne transmission is crucial for effective outbreak control and public health safety.

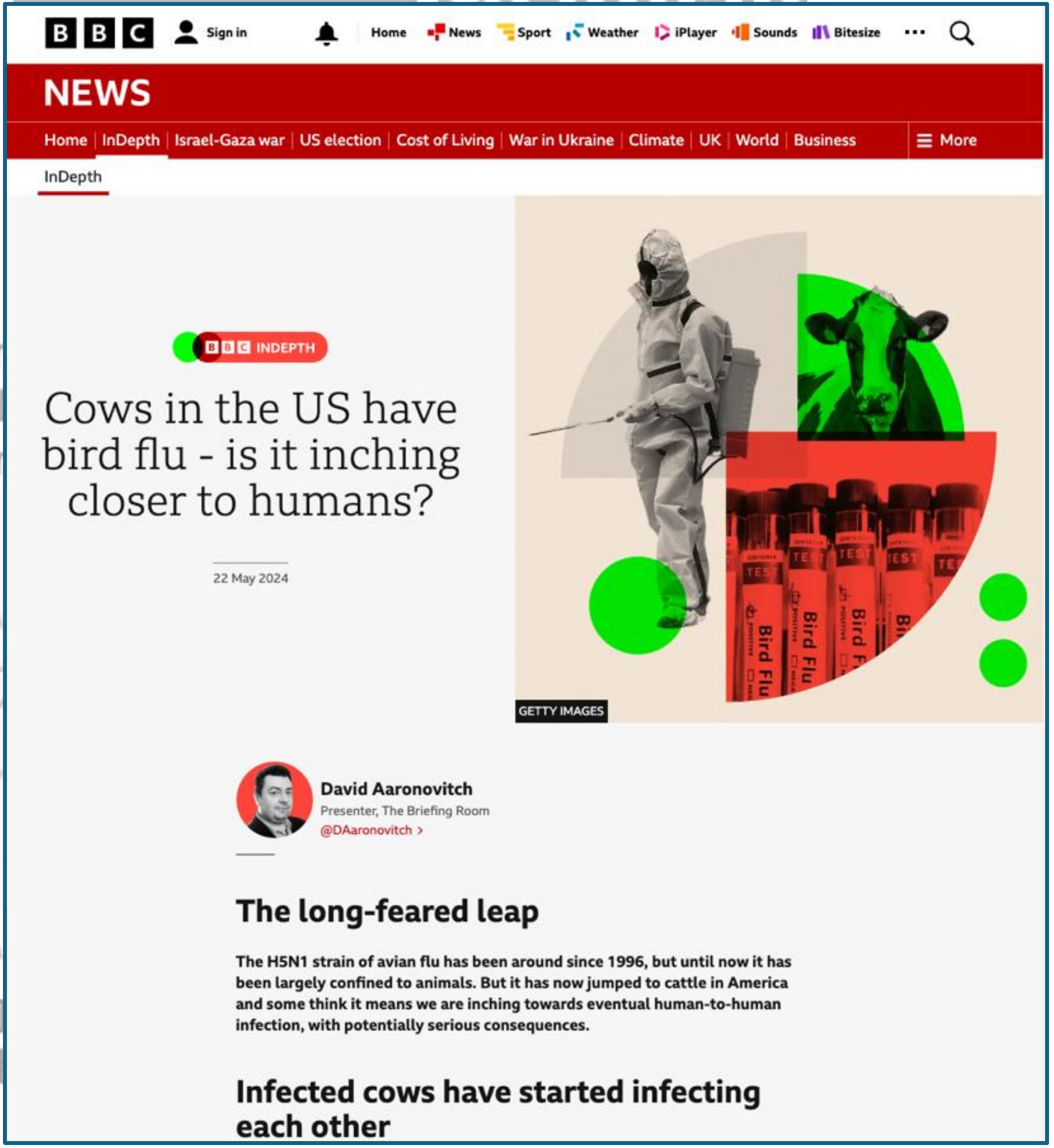
Deadly H5N1 avian influenza str
mainland Antarctica for the first

By environment reporter Peter de Kruijff

ABC Science | Science and Technology

Mon 26 Feb

The detection of avian influenza in a pair of deceased brown skua presents a risk to Antarctica's unique wildlife which could be devastated by the disease. *(Liam Quinn, Gentoo Penguin defends against a Brown Skua, CC BY-SA 2.0 DEED)*



Cows in the US have bird flu - is it inching closer to humans?

22 May 2024

David Aaronovitch
Presenter, The Briefing Room
@DAaronovitch ›

The long-feared leap

The H5N1 strain of avian flu has been around since 1996, but until now it has been largely confined to animals. But it has now jumped to cattle in America and some think it means we are inching towards eventual human-to-human infection, with potentially serious consequences.

Infected cows have started infecting each other

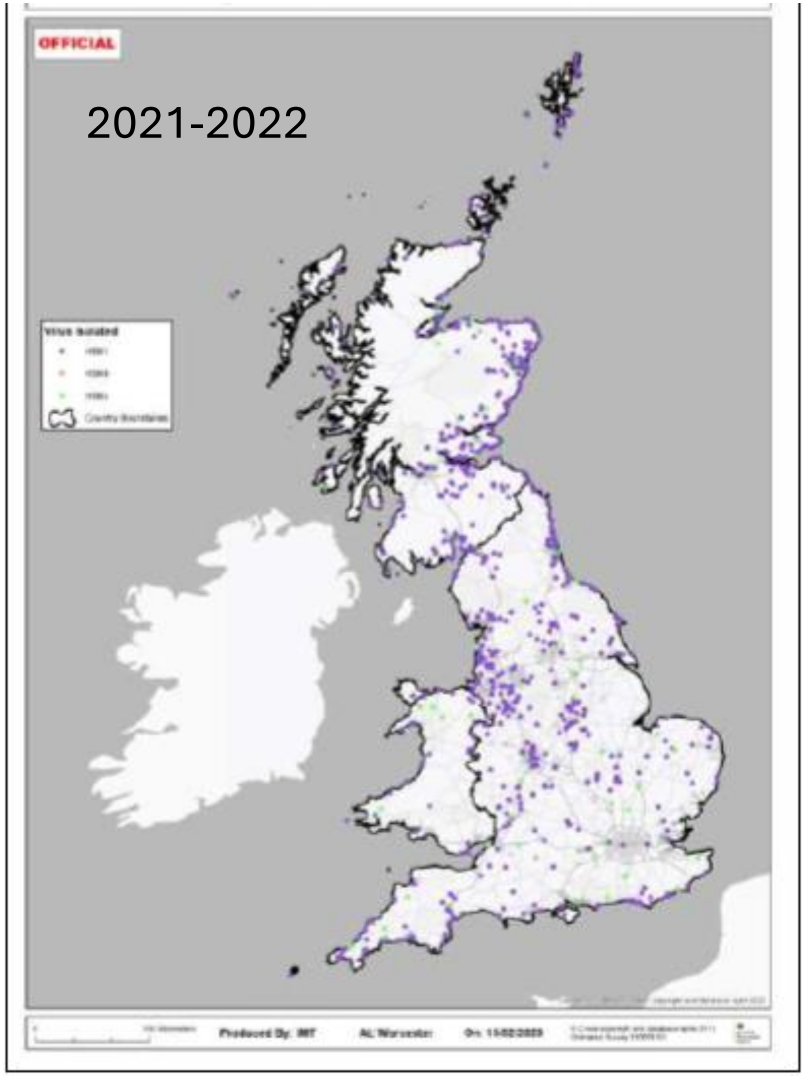
Novel emergence in cattle

Continued spread to new populations

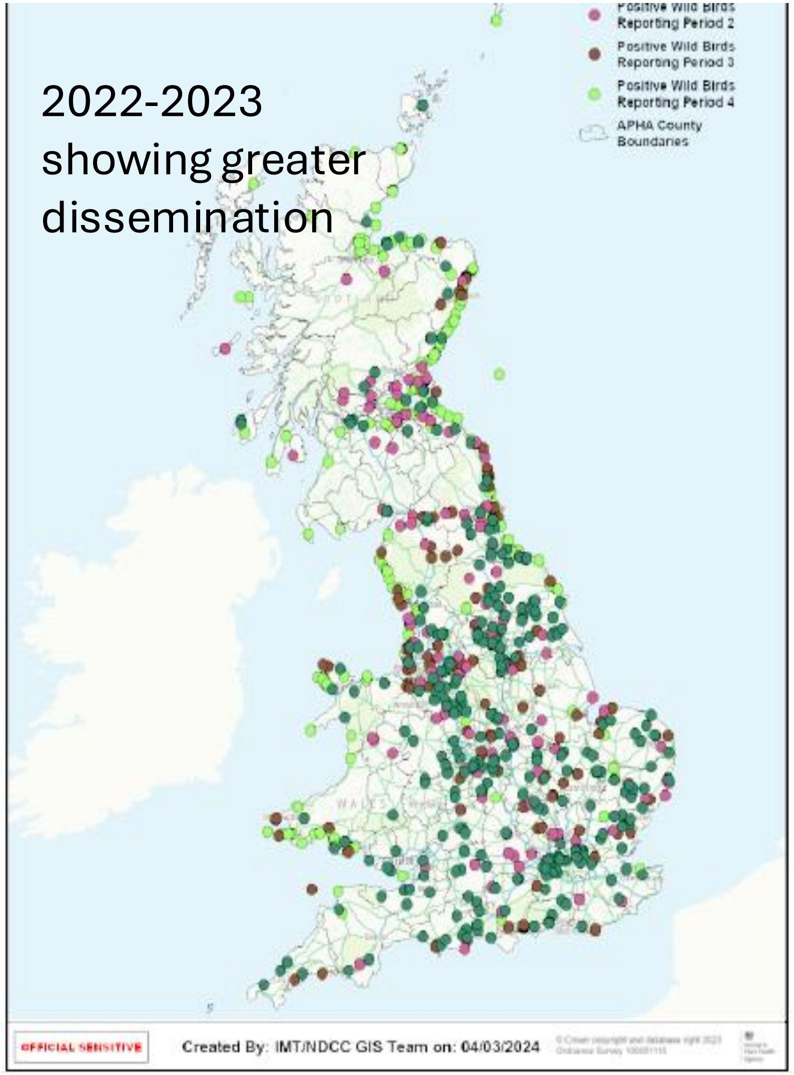
https://www.abc.net.au/news/science/2024-02-27/deadly-h5n1-avian-influenza-strain-detected-on-antarctica/103473276

https://www.bbc.co.uk/news/articles/clee5685w19o
https://www.nature.com/articles/s41684-024-01425-z

HPAI in Great Britain (H5N1)



2021-2022

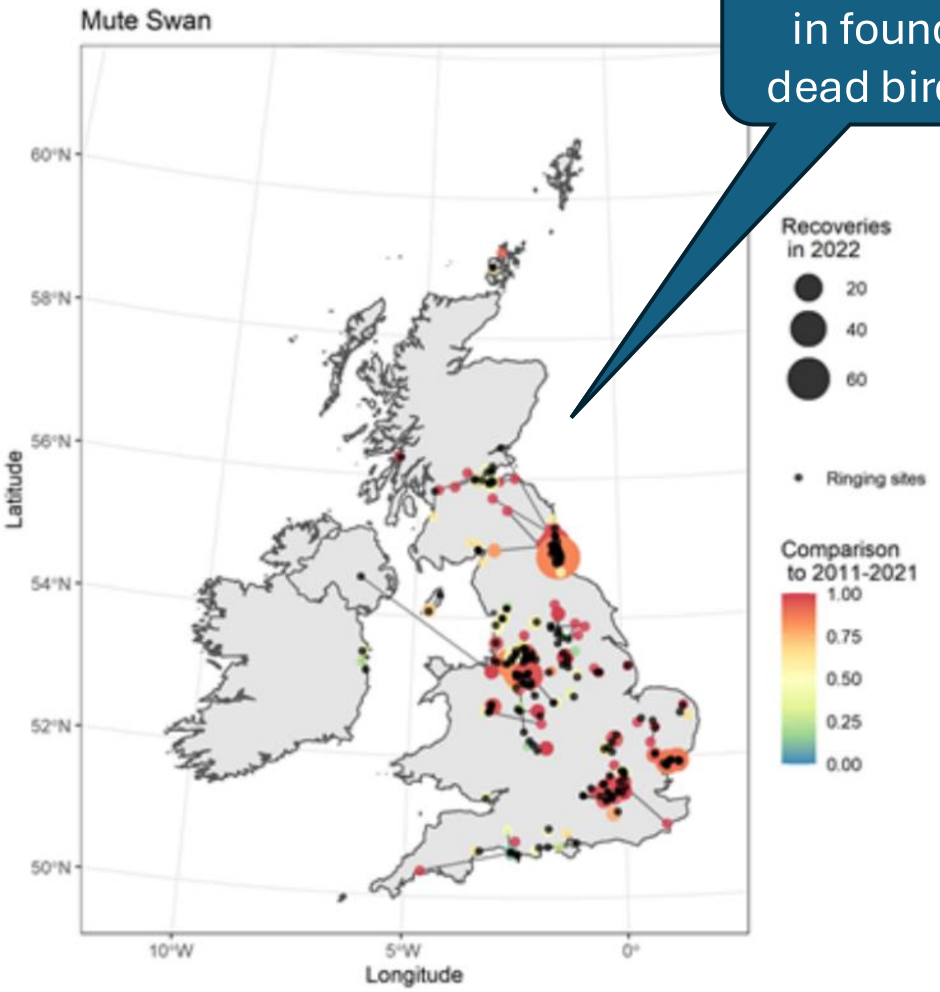


2022-2023 showing greater dissemination

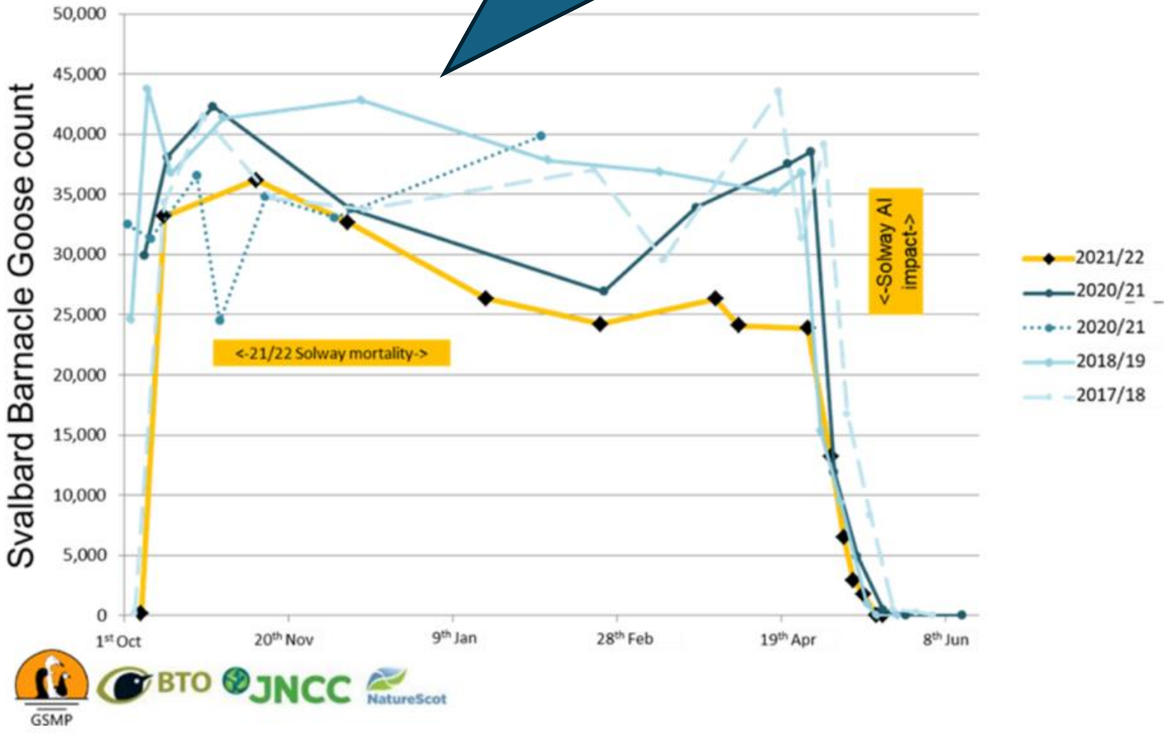
Source 1: October 2021 to September 2022: Highly Pathogenic Avian Influenza H5N1 outbreaks in Great Britain. The Animal and Plant Health Agency. [*https://www.gov.uk/government/publications/reports-relating-to-recent-cases-of-avian-influenza-bird-flu*](https://www.gov.uk/government/publications/reports-relating-to-recent-cases-of-avian-influenza-bird-flu)

Source 2: October 2022 to September 2023: Highly Pathogenic Avian Influenza H5N1 outbreaks in Great Britain. The Animal and Plant Health Agency. [*https://www.gov.uk/government/publications/reports-relating-to-recent-cases-of-avian-influenza-bird-flu*](https://www.gov.uk/government/publications/reports-relating-to-recent-cases-of-avian-influenza-bird-flu)

National impact on seabird populations



Increases in found dead birds



Decreased observations consistent with 1/3 population reduction

Pierce-Higgins et al. British Trust for Ornithology report 2023

UK Government commissioned report advised on four specific issues:

1. the host range of the current virus and the potential roles of non-avian hosts
2. the possibility of interventions to reduce the impact of HPAI on wild birds
3. the potential to supplement current approaches to control with vaccination
4. the potential to model the expected future trajectory of the outbreak

30 March 2023 — Independent report

Highly pathogenic avian influenza in Great Britain: evaluation and future actions

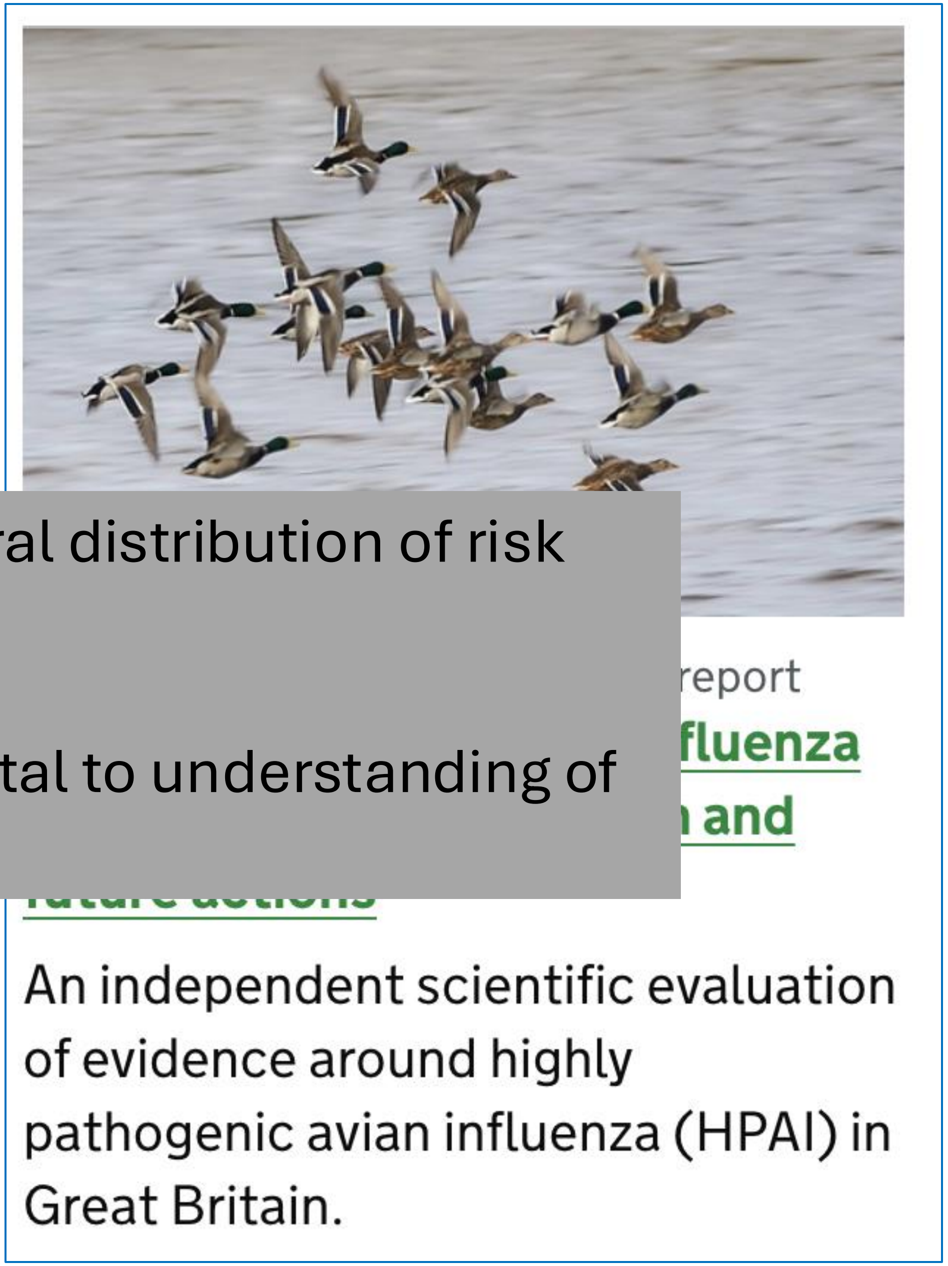
An independent scientific evaluation of evidence around highly pathogenic avian influenza (HPAI) in Great Britain.

UK Government commissioned report advised on four specific issues:

1. the host range of the current virus and the potential roles of non-avian hosts
2. the possib
   reduce th
3. the poten
   approaches to control with vaccination
4. the potential to model the expected future trajectory of the outbreak

Ascertainment of the spatio-temporal distribution of risk based on limited, biased data.

Characterizing circulation of virus vital to understanding of future risks



report

fluenza

and

future actions

An independent scientific evaluation of evidence around highly pathogenic avian influenza (HPAI) in Great Britain.

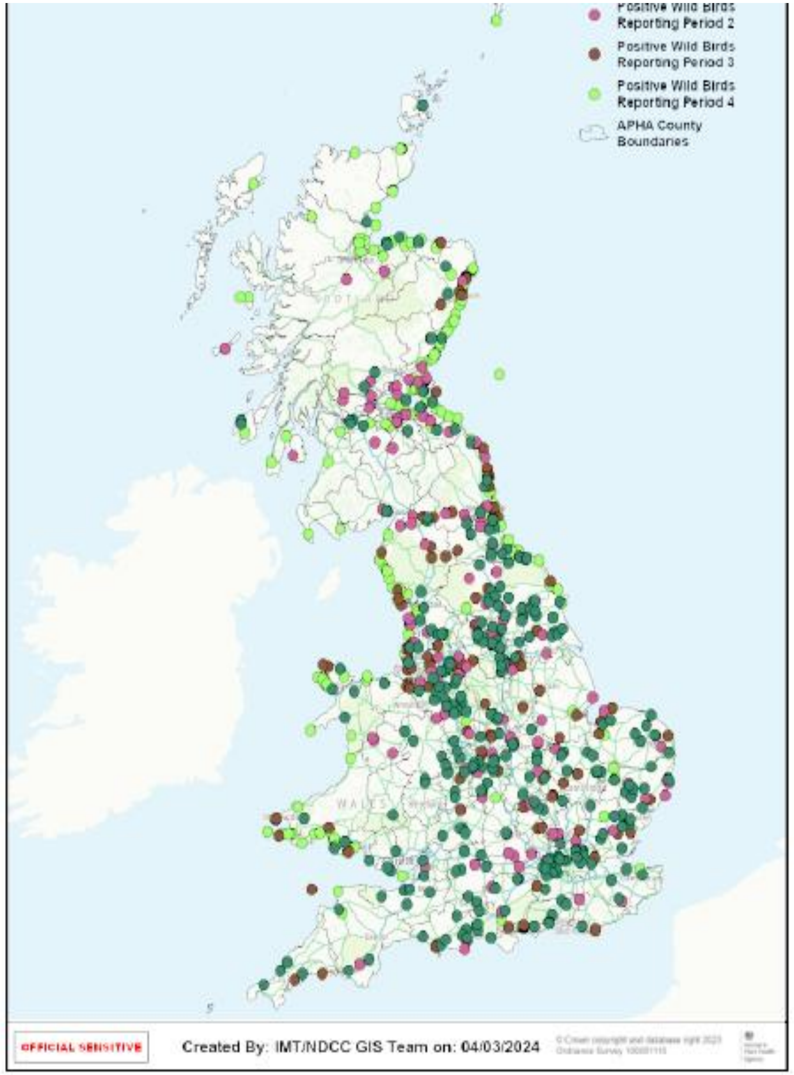
Wild bird infection reporting

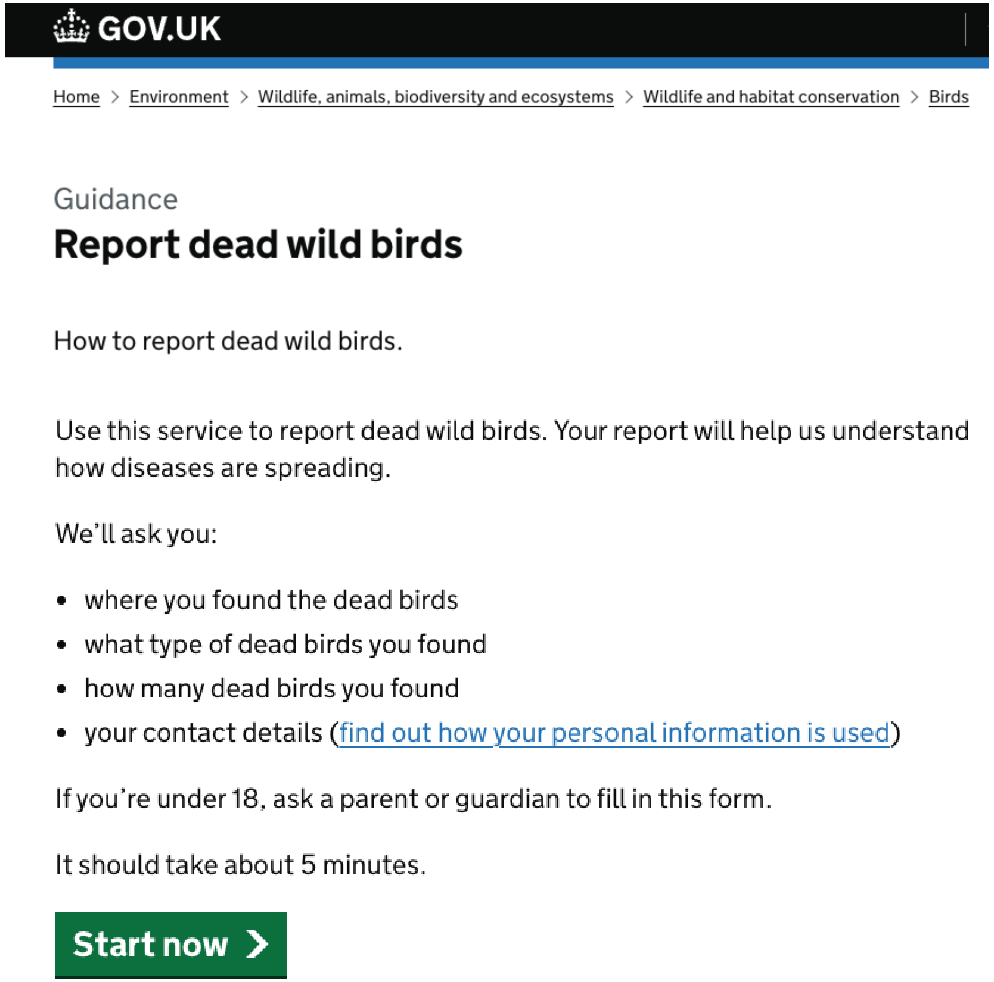
Passive surveillance syster

Previously via phone, now online

Species recorded as Goos Gull, Swan, Duck, Bird of Prey and Other. In this study, grouped into

- Landbirds
- Waterbirds
- Seabirds





Guidance

Report dead wild birds

How to report dead wild birds.

Use this service to report dead wild birds. Your report will help us understand how diseases are spreading.

We'll ask you:

- where you found the dead birds
- what type of dead birds you found
- how many dead birds you found
- your contact details (find out how your personal information is used)

If you're under 18, ask a parent or guardian to fill in this form.

It should take about 5 minutes.

https://www.gov.uk/guidance/report-dead-wild-birds

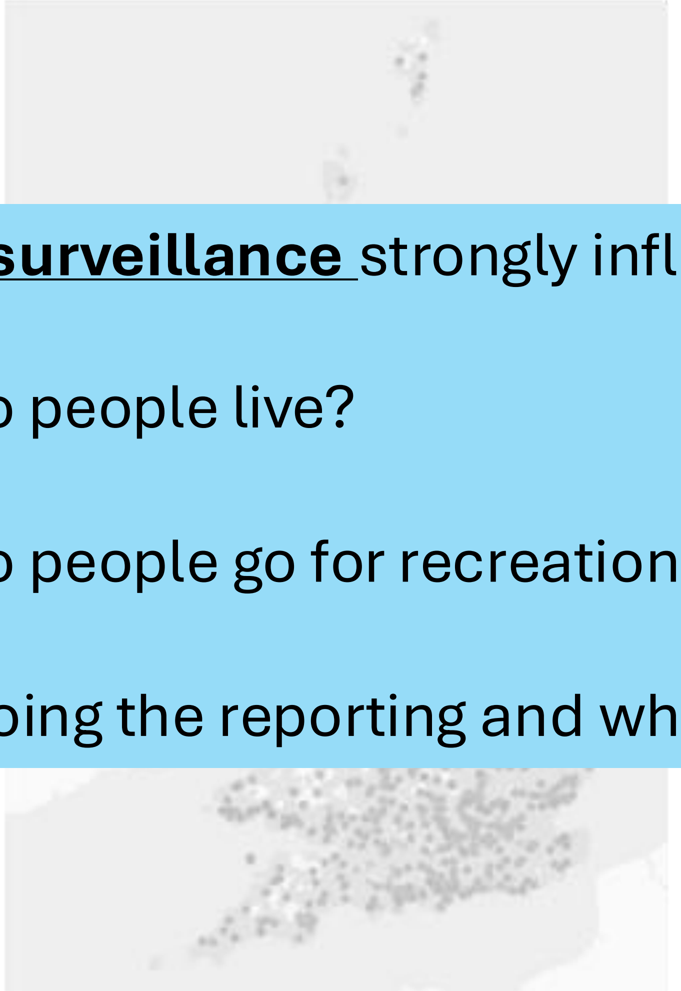
Wild bird infection reporting

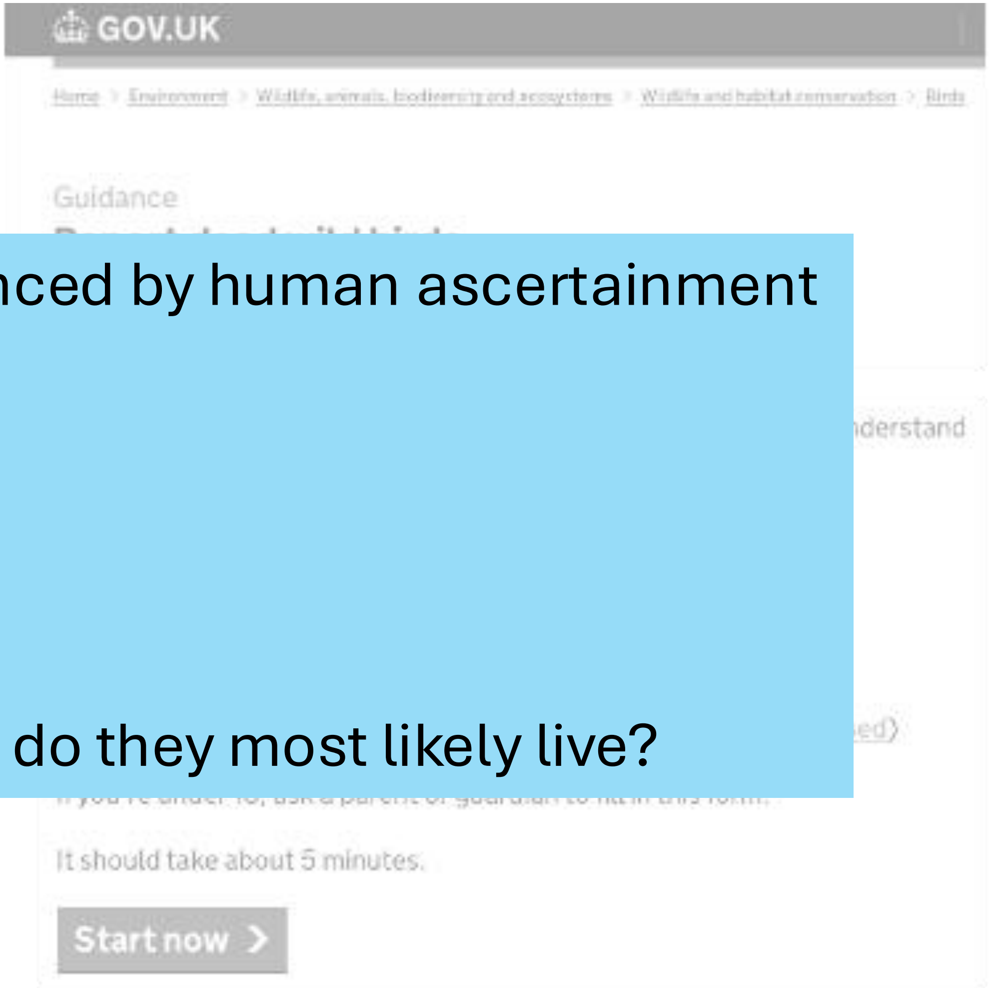
Passive surveillance system

Previously via ph
online

Species recorde
Gull, Swan, Duck
Prey and Other.
study, grouped i

- Landbirds
- Waterbirds
- Seabirds





It should take about 5 minutes.

Passive surveillance strongly influenced by human ascertainment

Where do people live?

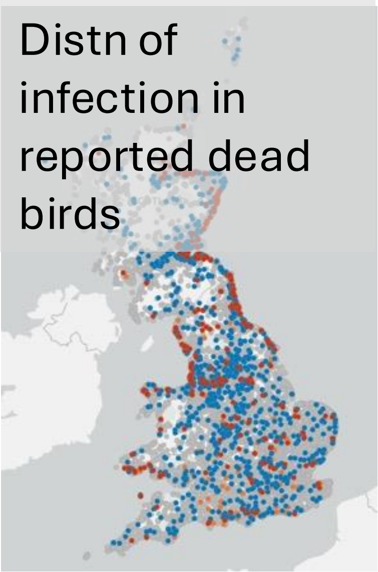
Where do people go for recreation?

Who is doing the reporting and where do they most likely live?

https://www.gov.uk/guidance/report-dead-wild-birds

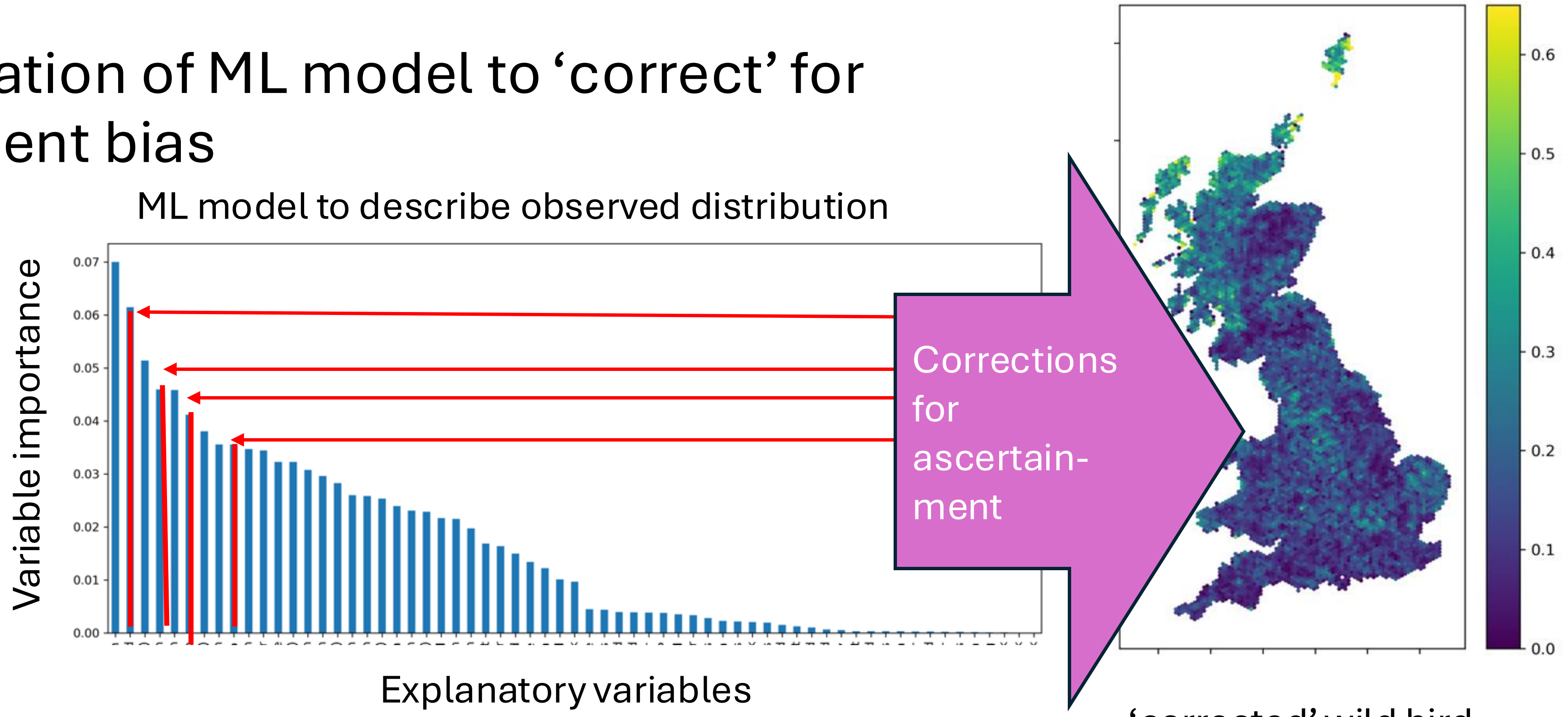
Defining wild bird and poultry contact networks (Roslin/Aeron Sanchez)

- Implementation of ML model to ‘correct’ for ascertainment bias



Distn of infection in reported dead birds

ML model to describe observed distribution



Explanatory variables

Corrections for ascertain-ment

‘corrected’ wild bird infection risk map

ROSLIN

THE UNIVERSITY of EDINBURGH
School of Physics and Astronomy

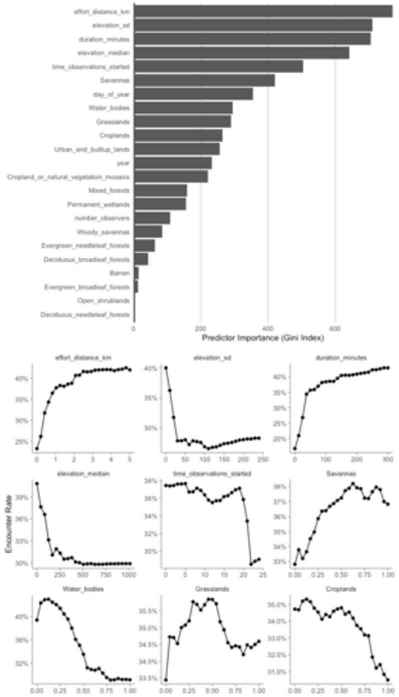
Estimates of wild bird densities

General public observe and count birds

eBird

Data submitted to eBird

eBird combine observations with measures of habitat, weather, effort, etc.



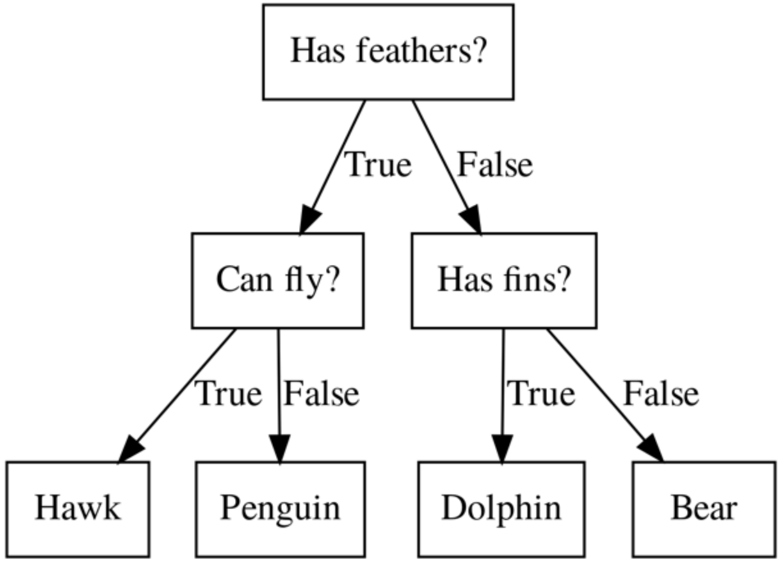
eBird run sophisticated predicted relative abundance modelling

Used by S. Vickers (RVC but now Edinburgh) to estimate wild bird population densities in GB

Machine Learning tools to identify epidemiological risks

- Binary Decision Trees

What set of questions allow me to most efficiently identify the right suspect?



https://amueller.github.io/aml/02-supervised-learning/08-decision-trees.html

(similar methods for continuous variables)

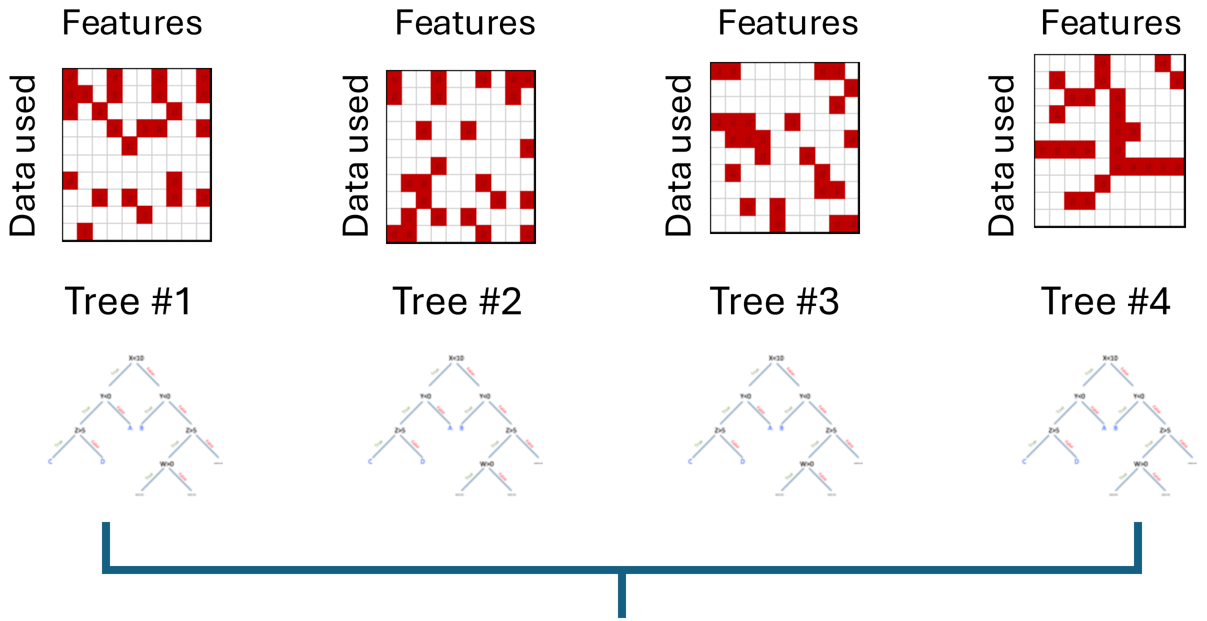
Correlation between variables

- Simple Decision Trees are highly susceptible to OVERFITTING
  - Very precise description of training dataset
  - Very poor prediction on any new data

  - Increasingly problematic with small datasets
  - Highly correlated variables
  - Noisy data

Randomly selected features (until pre-set minimum points per leaf)

Randomly chosen training data subsets (66%)



"VOTE"
Take aggregate of answers, rather than the answer of the aggregate

Controls for overfitting and variable correlations

Assessing univariate roles

- Assess contribution of individual variables to model fit
- “permutation importance” – assess reduction in model fit after randomizing each variable in turn
- Partial dependence/Accumulated local effects – assess univariate impact on outcomes
- Shapley values/Shapley importance ensures ‘fair’ distribution of contribution to all contributors (efficient, symmetric, additive and with only contributors benefiting)

Risk factors included in the model

- Date
- Location
- Species Group
- Wild bird populations for species grouping and date
- Human populations
- Distance to rivers and lakes
- Distance to roads and walking paths
- Distance to towns
- Driving distance and times to towns
- Land usage
- Park or green Space
- Reported on a bank holiday
- Distance to eBird hotspot and eBird hotspots within a set radius
- Number of poultry farms and animals within set distances.

Data sources:

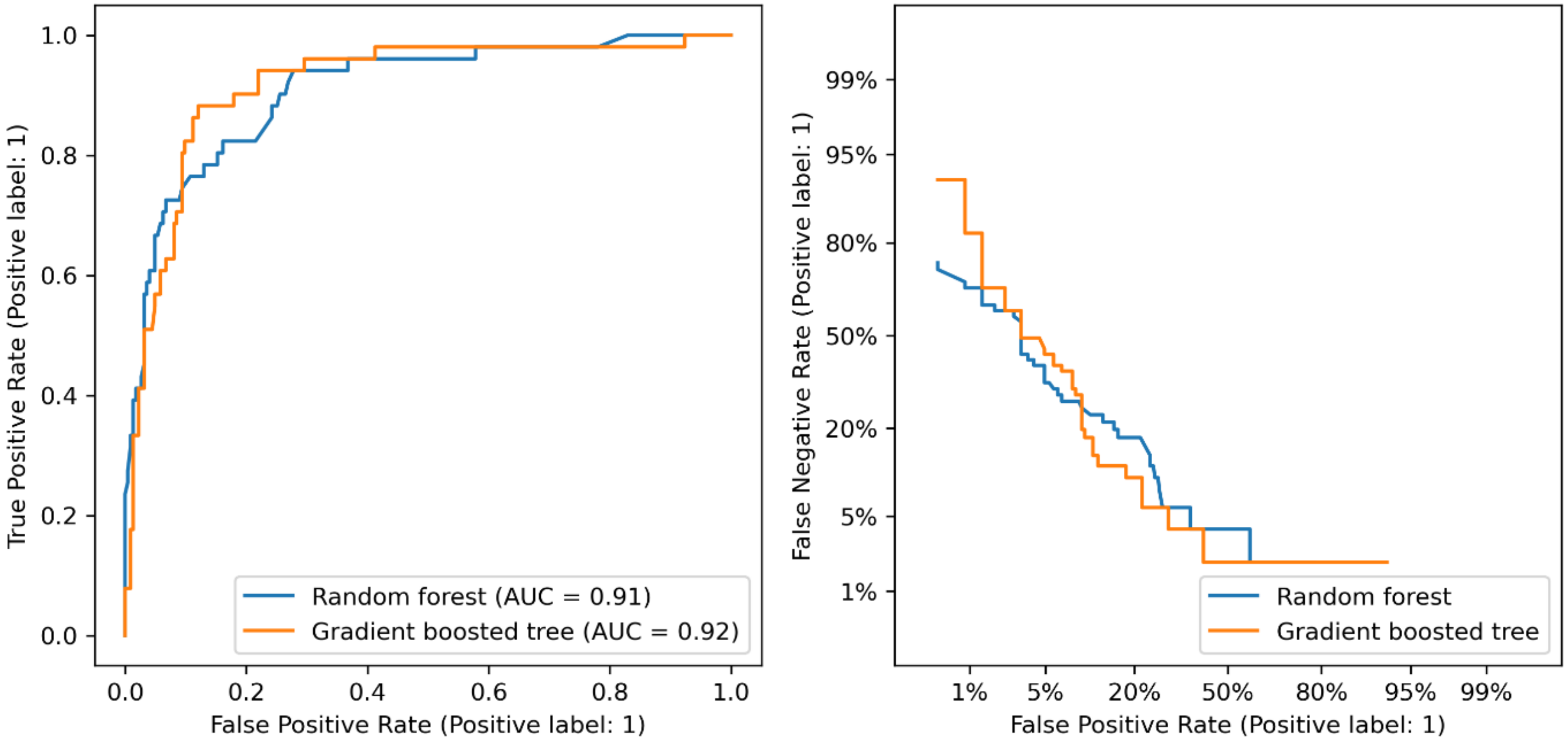
Density estimates based on
e-Bird (Vickers, RVC)

UK Census data

Training and testing ML models

- Consider multiple approaches: random forest models, GLMs, SVMs and gradient boosted trees
- Models trained to predict binary outcome: **will a tested wild bird that result as positive/negative for H5N1 (w/o specifying lab test sensitivity/specificity)**.
- For testing and analysing our models we used 5-Fold cross validation (run 4 times).
- Area under the ROC curve or AUC as a measure of the model performance.
- Random forest models and gradient boosted decision trees were the best performing (GBT .
- For the rest of the investigation, we will use boosted decision tree models
  - Decision threshold at 0.5 this gave an accuracy of 0.879 [0.858,0.894], sensitivity of 0.678[0.584,0.726] and specificity of 0.938 [0.905,0.957].

Assessing model fit

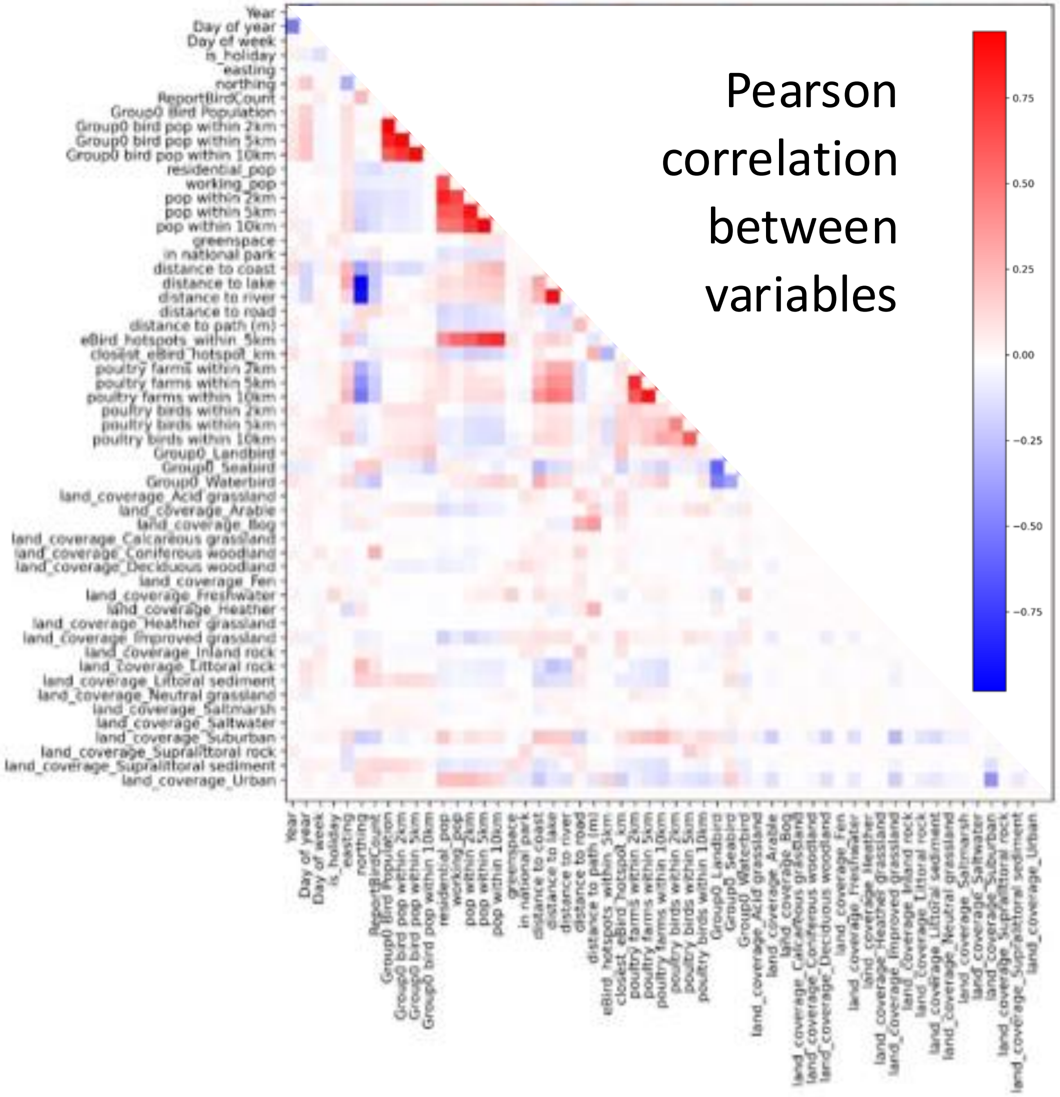


Prediction depends on the chosen decision threshold used on the probability of the output prediction.

(L) Receiver Operating Characteristic (ROC) plot of true positive rate (sensitivity) vs. false positive rate (1-specificity)

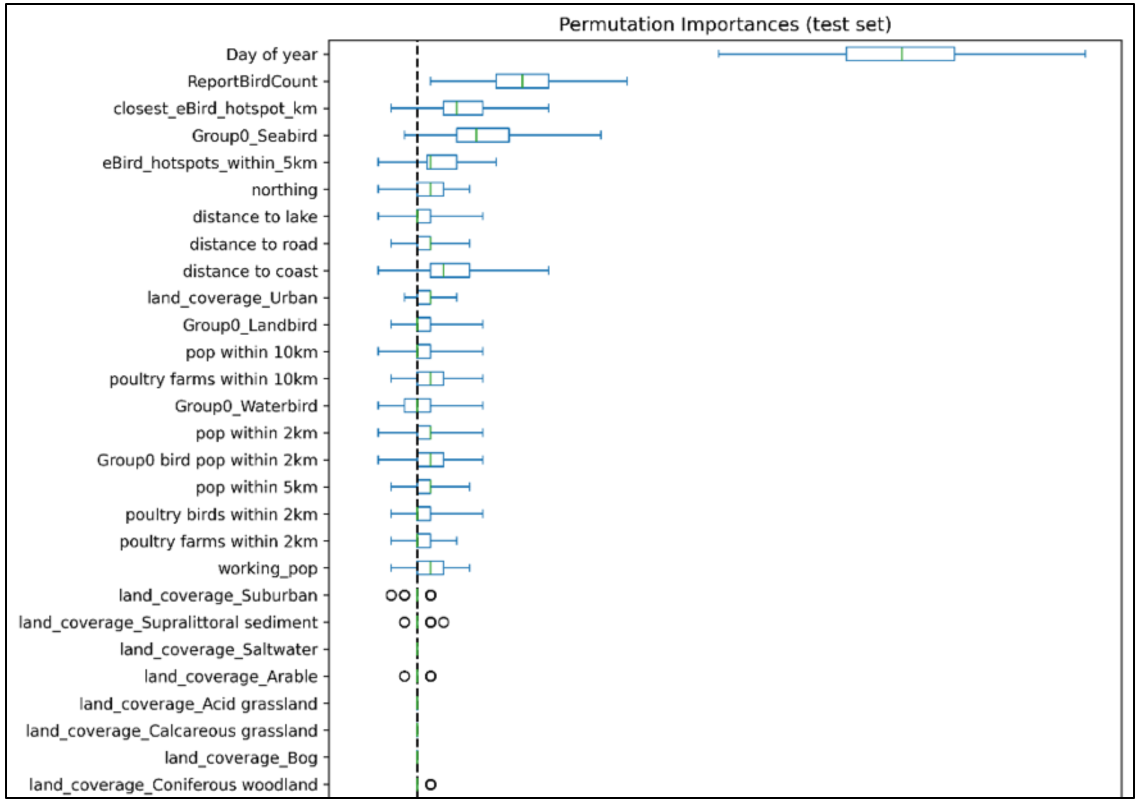
(R) Detection error tradeoff (DET) curve Shows false negative rate (1-sensitivity) against false positive rate(1-specificity).

Model tested for overfitting using 5 fold cross-validation



Pearson correlation between variables

Highly correlated variables can “mask” the importance of individual variables.



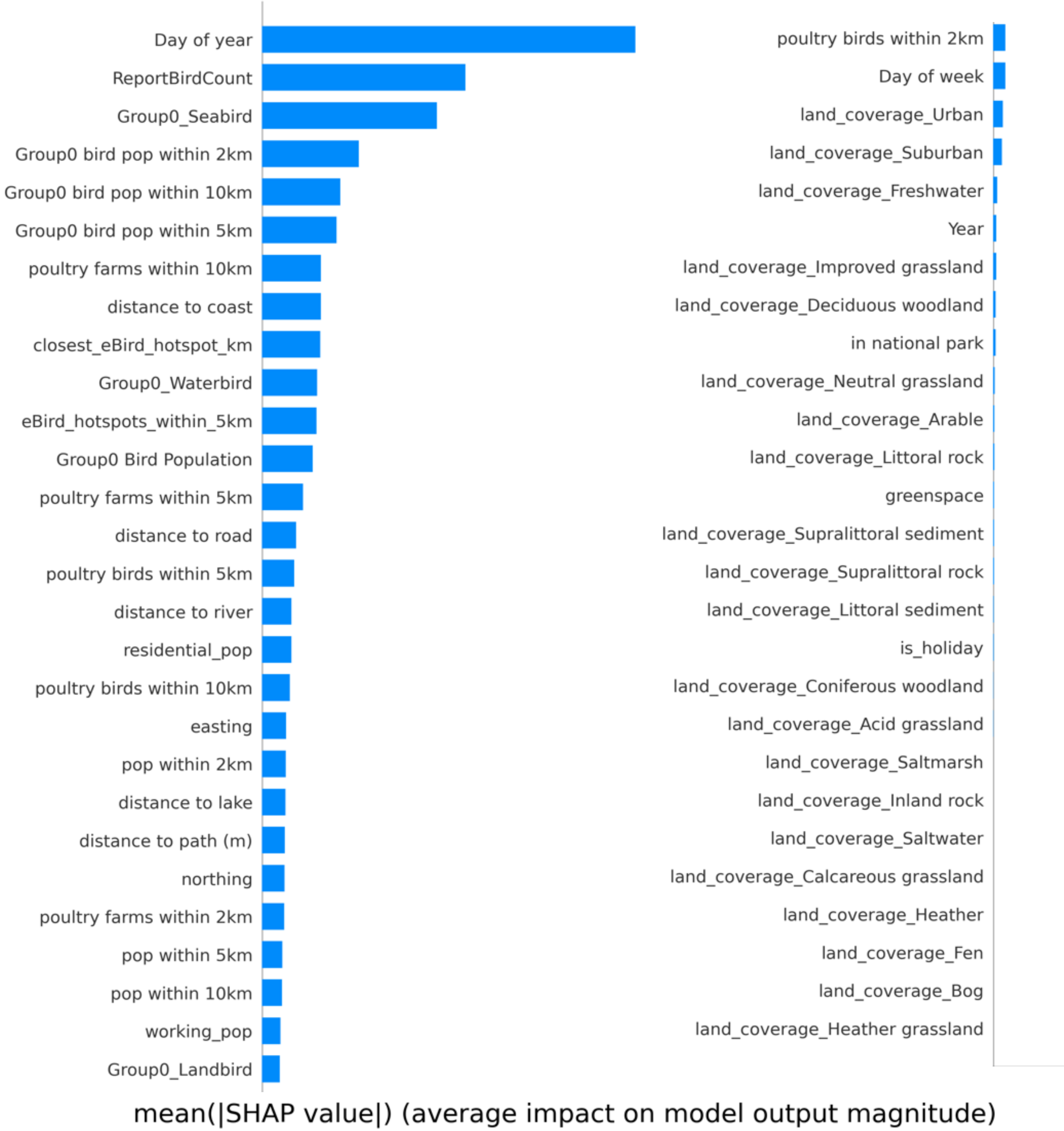
Shapley Feature Importance

Permutation feature importance shows many ”unimportant” variables

BUT model has high accuracy (R2 ~ 0.xx).

likely due to the correlations between a number of the input features

Shapely contributions or SHAP plots we see a higher number of features contributing to the model predictions.



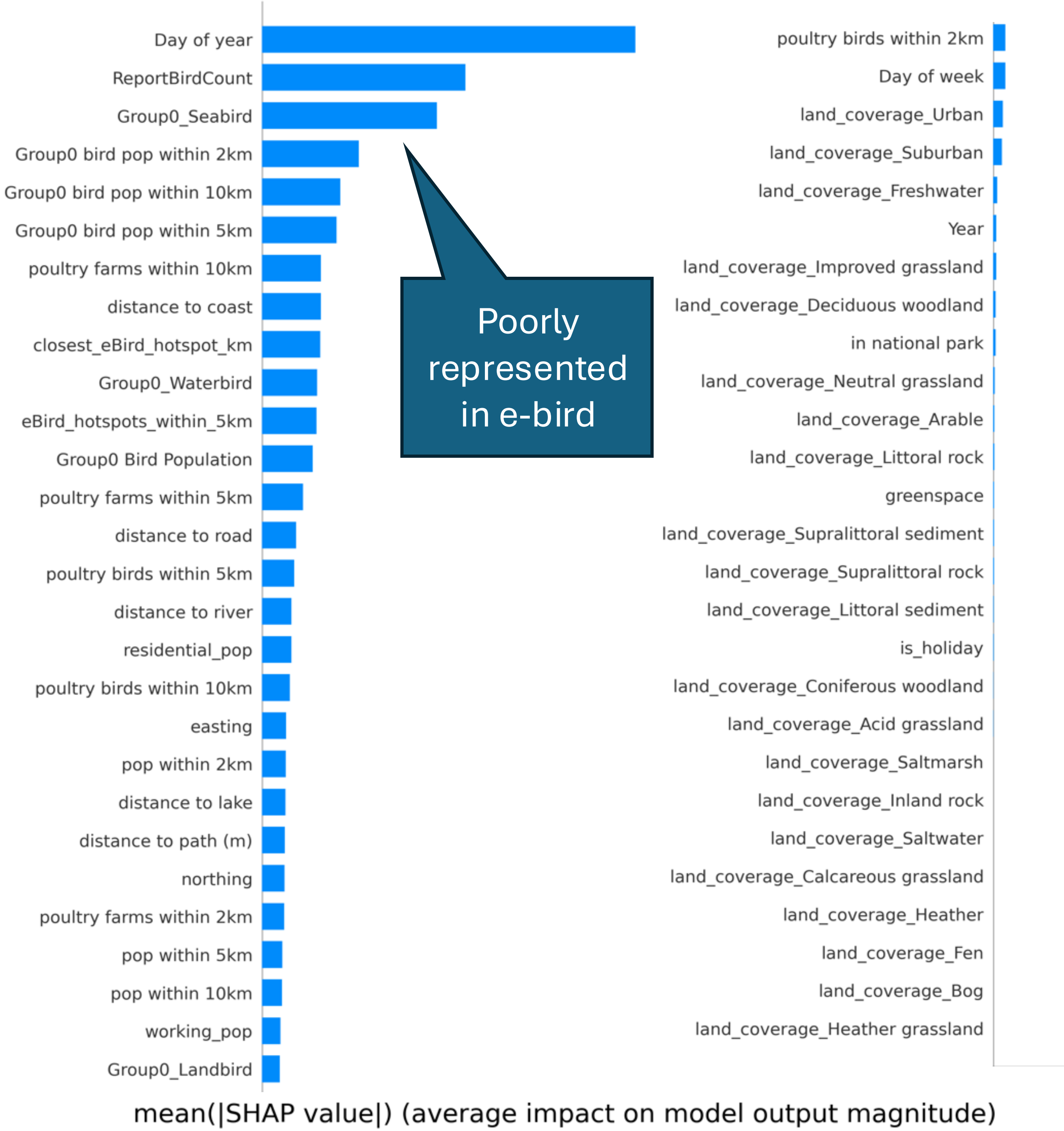
Shapley Feature Importance

Permutation feature importance shows many ”unimportant” variables

BUT model has high accuracy (R2 ~ 0.xx).

likely due to the correlations between a number of the input features

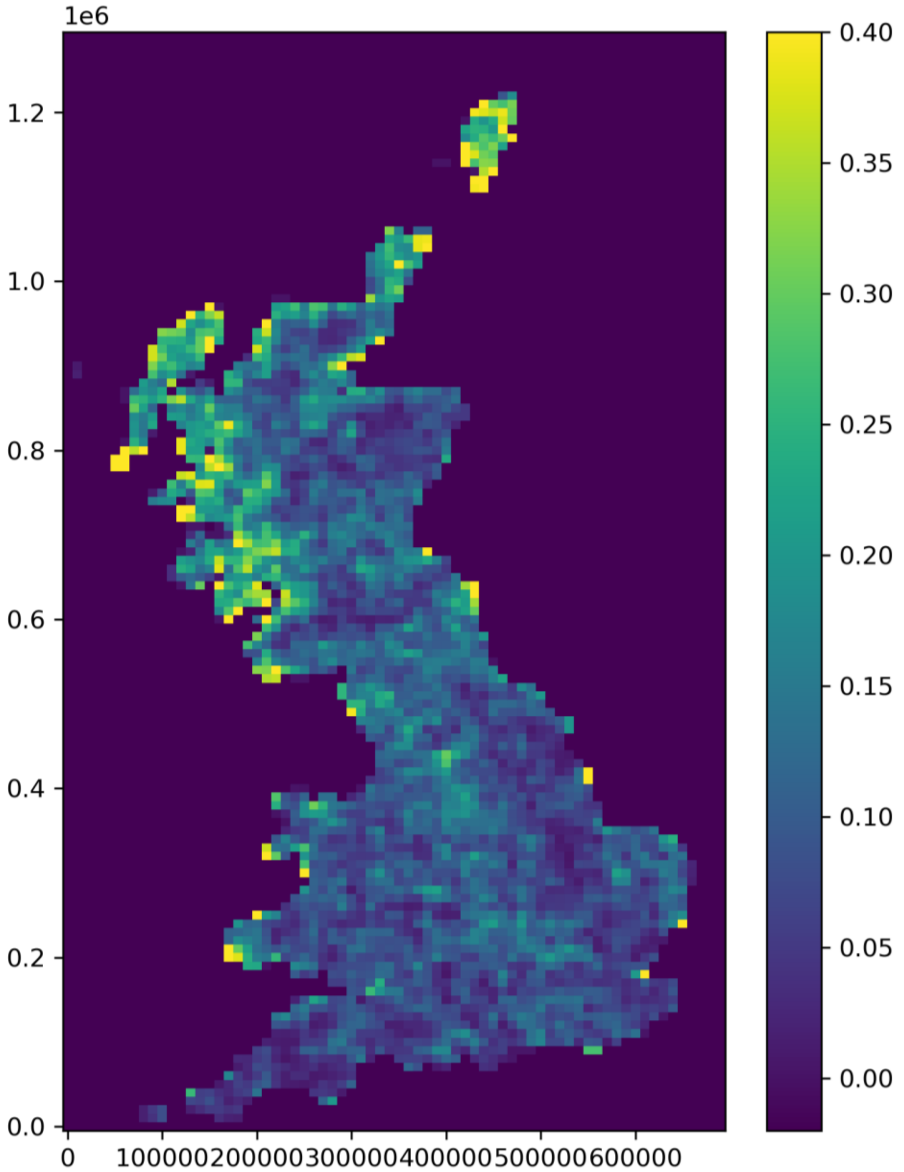
Shapely contributions or SHAP plots we see a higher number of features contributing to the model predictions.



Poorly represented in e-bird

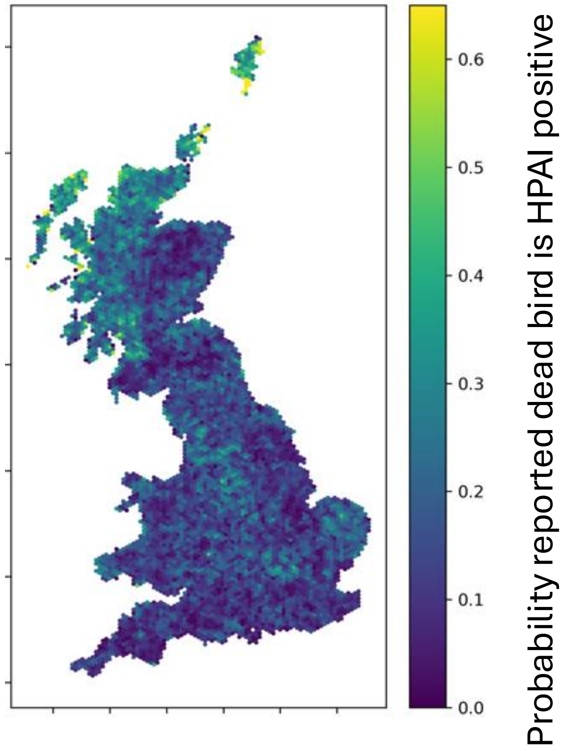
Generating risk maps

- models trained on the entire data set without a test set
- 50,000 mock data points generated with a random location date and species.
- Each mock data point passed through the final model; probability of testing positive calculated for each point.

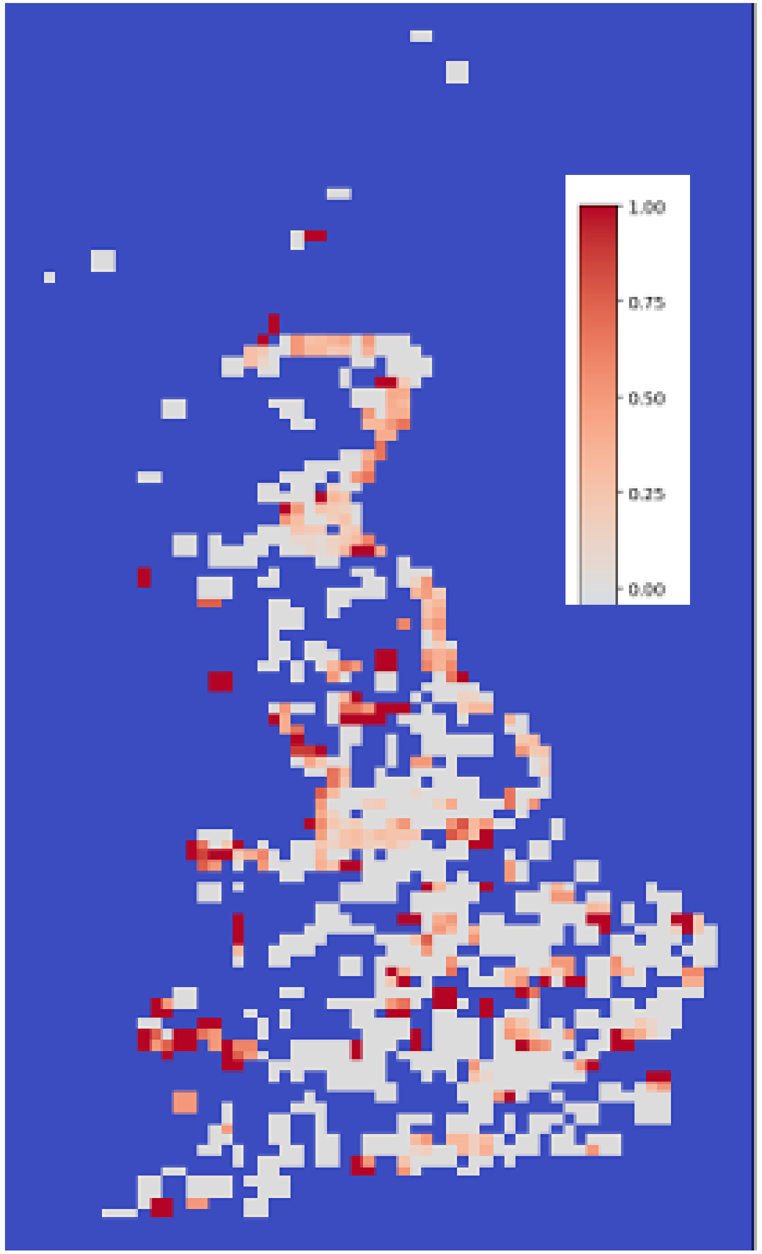


Scatter plot of all the mock points where the colour scale represents the model prediction probability of an that bird testing positive.

Ascertainment corrected risk map



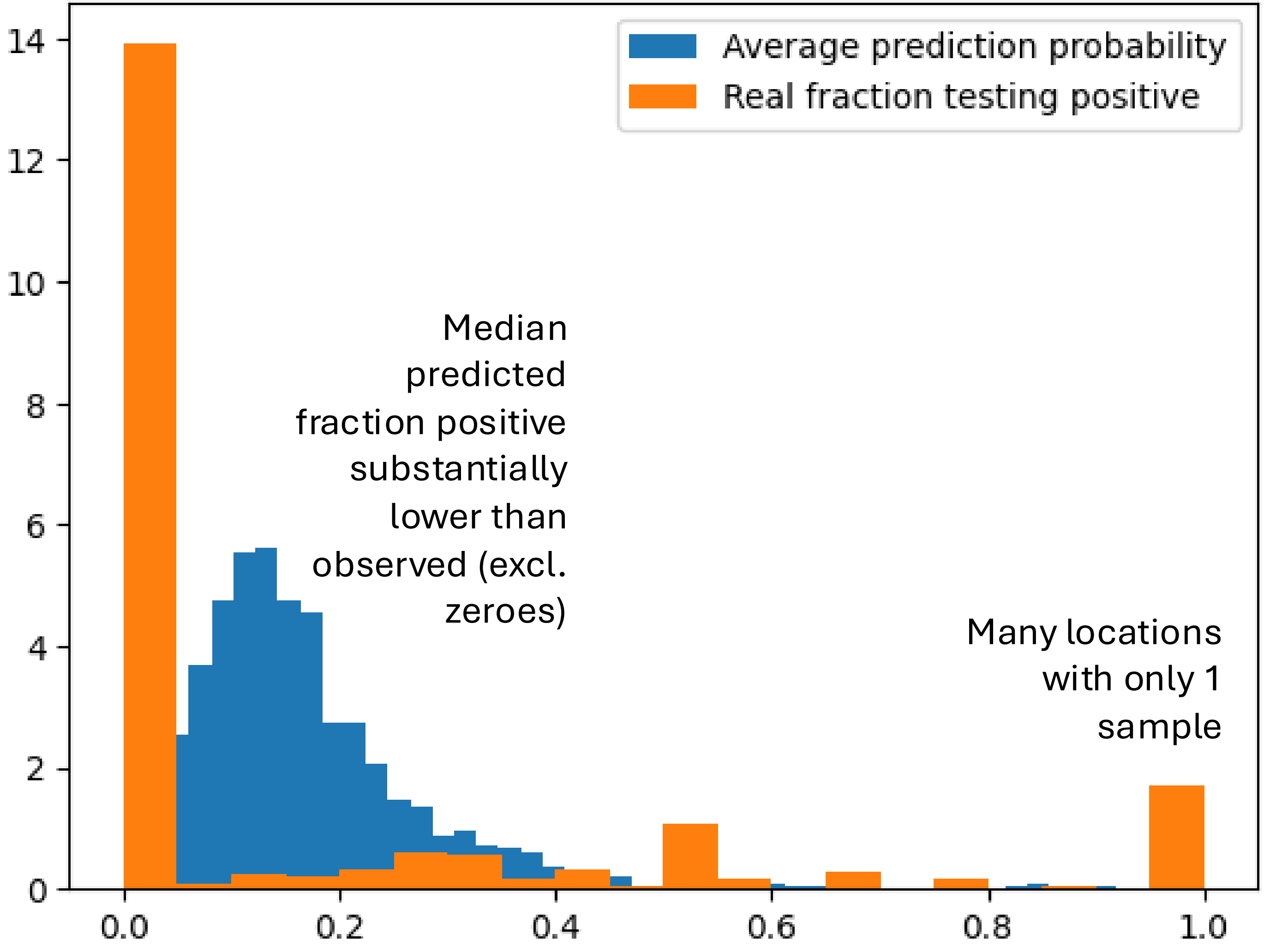
Where is under-reporting particularly evident?



Areas in red represented areas with at least one reported dead bird where ascertainment is particularly low

Comparison of real test locations to predicted

Removal of ascertainment biased variables suggests there may be a much broader circulation of virus than observed but most have lower test fraction than observed



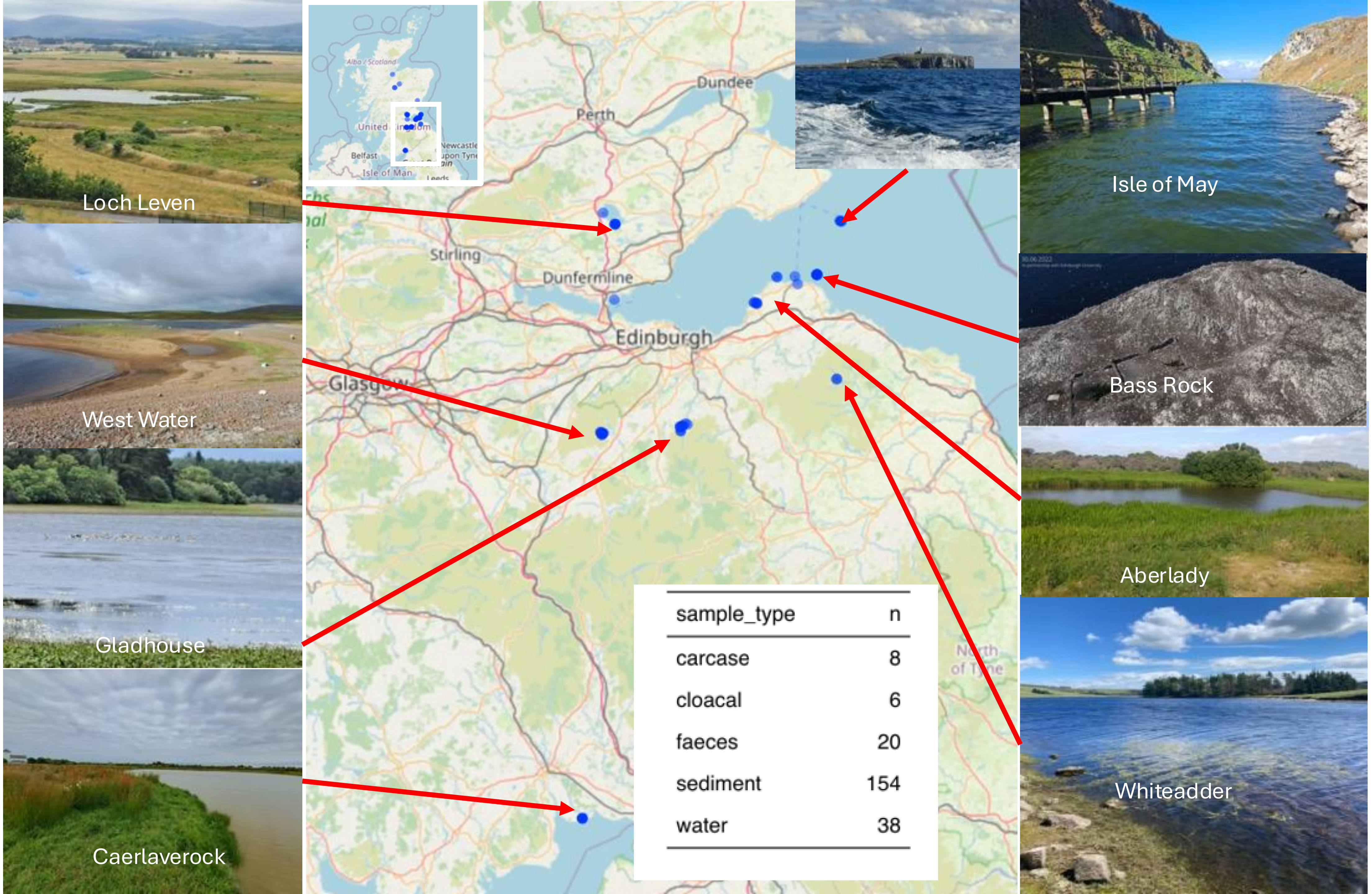
Median predicted fraction positive substantially lower than observed (excl. zeroes)

Many locations with only 1 sample

Ongoing work by collaborators uses/tests this model

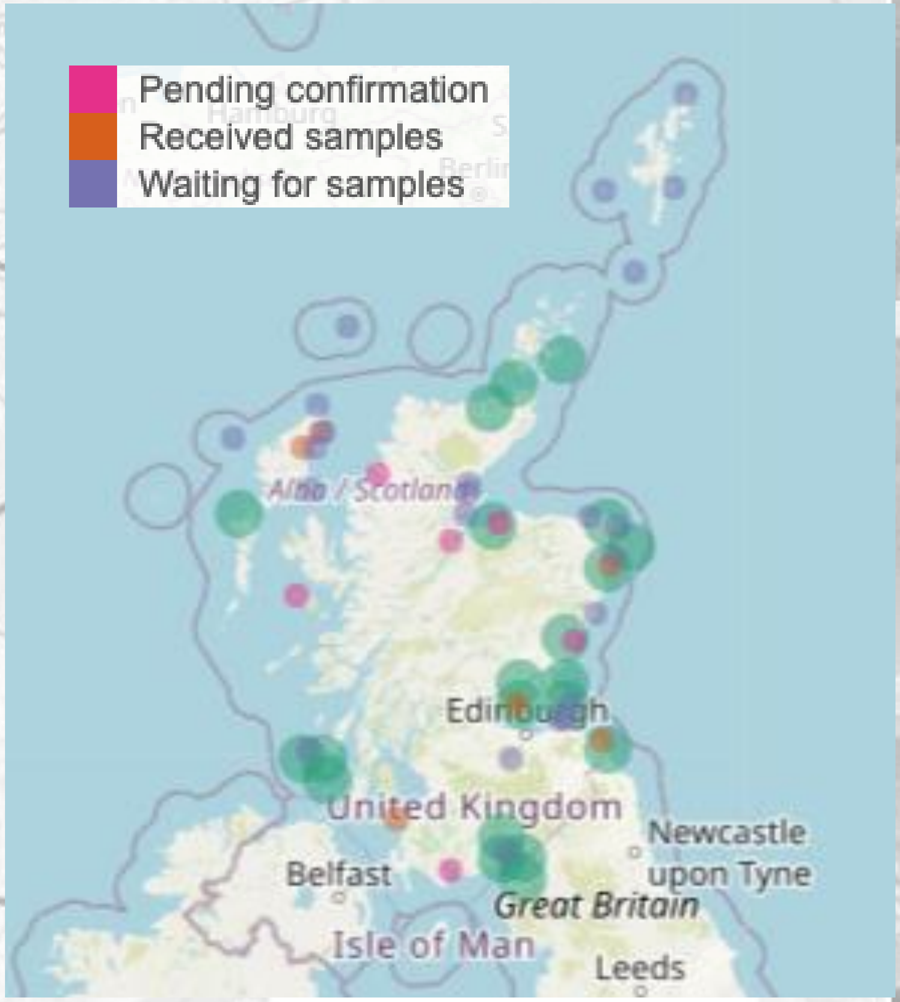
- Pilot: 7 main sites plus 5 one-off sample sites from the Highlands and also collections from Bass Rock and Fedra
        - (may increase –Loch Fleet, Wigton Bay)
    - Pond/lake sediments (5 sites/4 samples per site)
    - Faecal/environmental samples (mainly geese/gannets)
- Repeat sampling 3 times (summer, start of winter arrivals later winter) – currently on 2nd round
- rt-PCR for M-gene – RNA extractions underway (delayed by procurement issues at Edinburgh) (n>200)
- MinION seq of any +'ves
    - Does this approach work?
    - Can we detect AIV in different seasons?

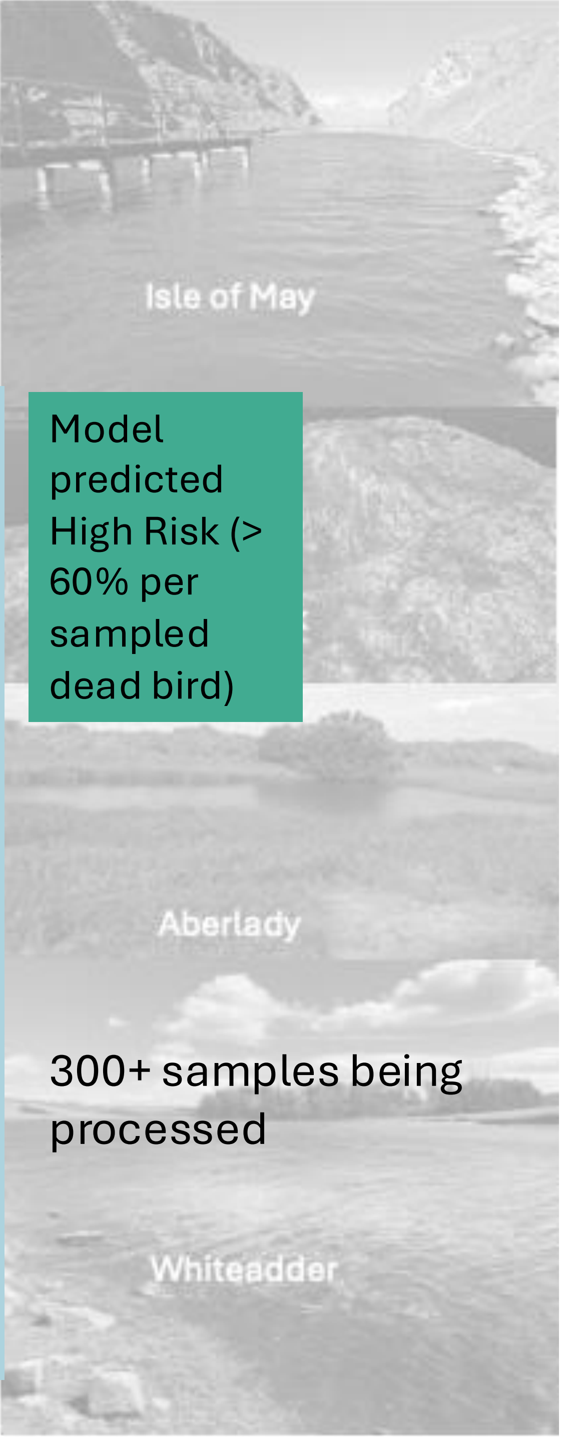
From work of Lina Gonzalez Gordon & Mark Bronsvoort, Roslin Institute



sample_type	n
carcase	8
cloacal	6
faeces	20
sediment	154
water	38

Ongoing work by collaborators uses/tests this model





Model predicted High Risk (> 60% per sampled dead bird)

300+ samples being processed

Conclusions

1. ML models describe surveillance data to high accuracy
2. Removal of human observation factors creates a “prediction landscape” based solely on natural factors
3. Ascertainment corrected map shows evidence of much wider spread dissemination but **no evidence of areas of high infection** (BUT may be beyond limits of ML models if strong unobserved nonlinearities
4. e-bird data coarse grained (seagulls in particular)
5. **Does not account for pattern vs. process** issues
6. Currently being used to guide field work/active surveillance

Thank you for listening!

Animal & Plant Health Agency

Imperial College London

The Pirbright Institute

UNIVERSITY OF LEEDS

VERSITY OF
MBRIDGE

ROSLIN

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Funders

UKRI Biotechnology and Biological Sciences Research Council

defra Department for Environment Food and Rural Affairs