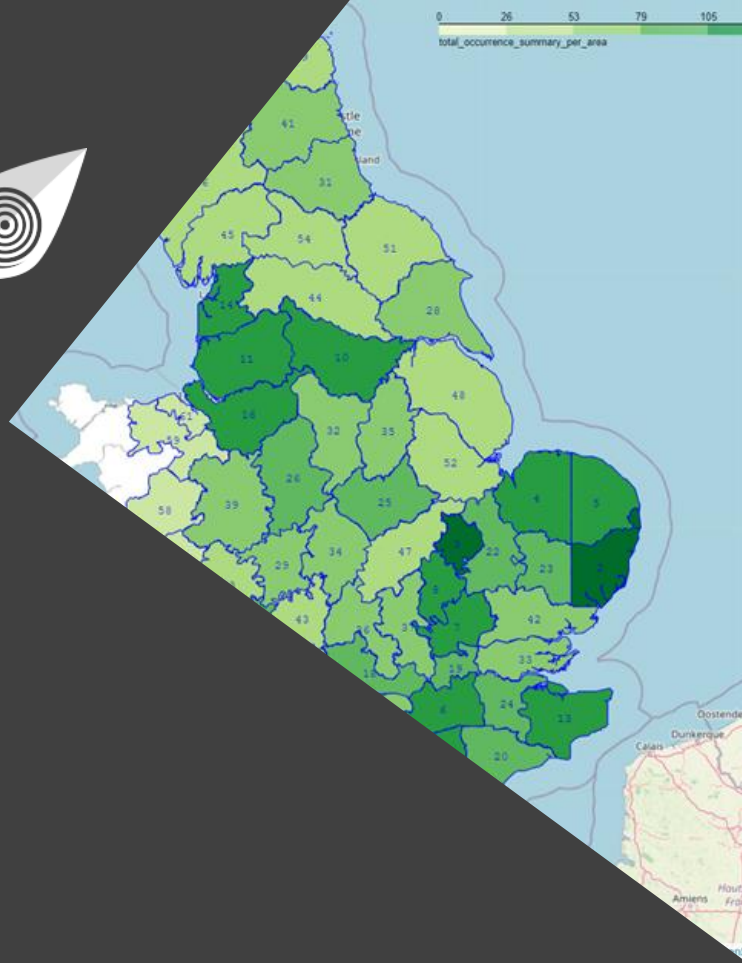
The Alan Turing Institute

NATURAL ENGLAND

NBN



Assessing NBN Atlas comprehensiveness and the impact of data providers: Insights from a Data Study Group

Simon Rolph

Data Scientist at UK Centre for Ecology & Hydrology

What is a data Study group?

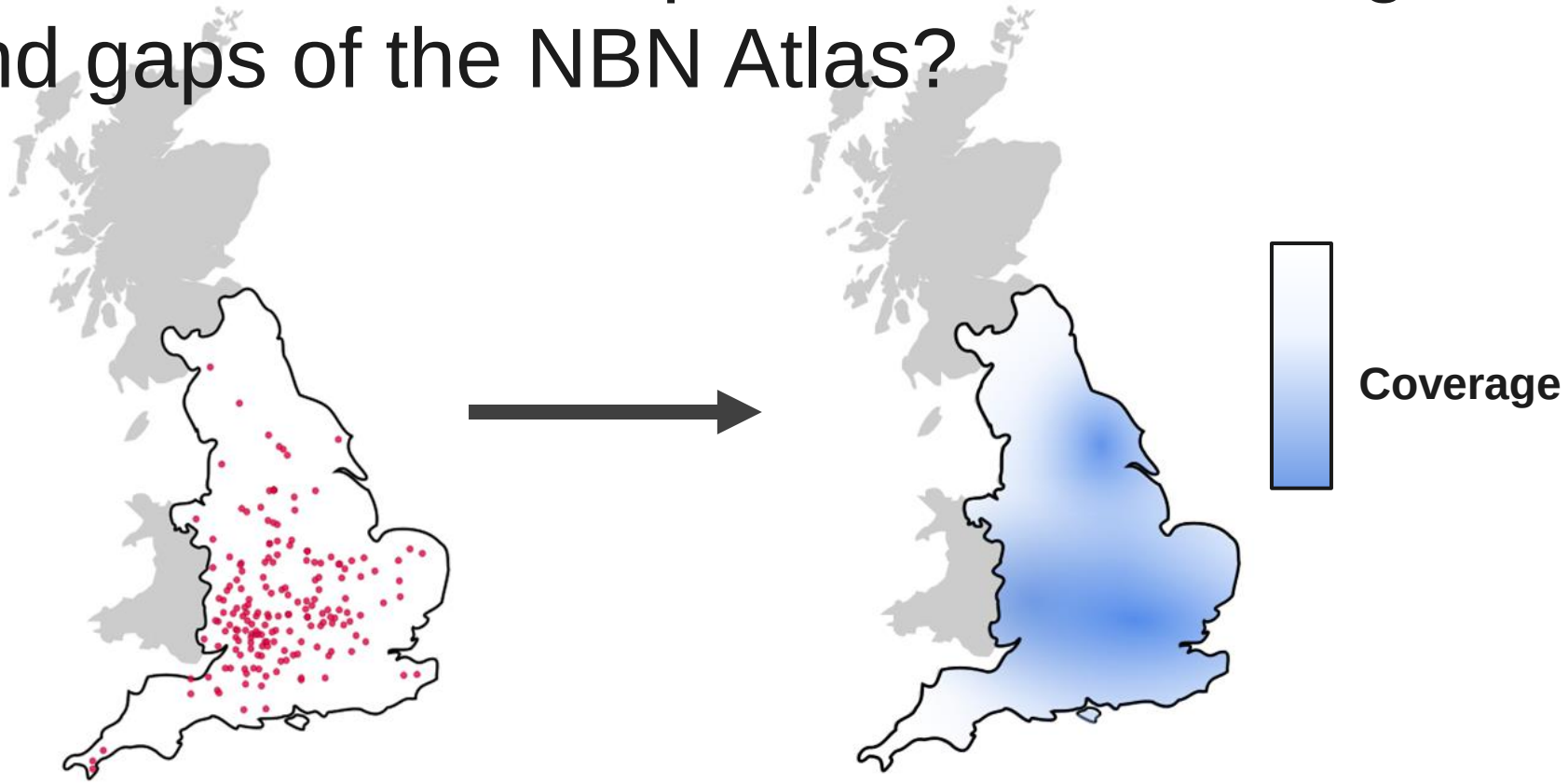
- Intensive **'collaborative hackathons'** which bring together organisations from industry, government, and the third sector, with talented multi-disciplinary researchers from academia

- Organisations act as Data Study Group **'Challenge Owners',** providing real-world problems to be tackled by small groups of researchers

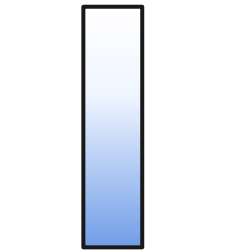
- Researchers and a PI (Dr Silvia Liverani) develop **data science solutions**, presenting their work at the end of the week.

What were our objectives?

How can we best represent the coverage and gaps of the NBN Atlas?



How can we identify species that we do not have comprehensive coverage for?



Coverage

Coverage

How can we highlight the value that data providers bring to the Atlas?

BTO
British Trust for Ornithology

National Trust

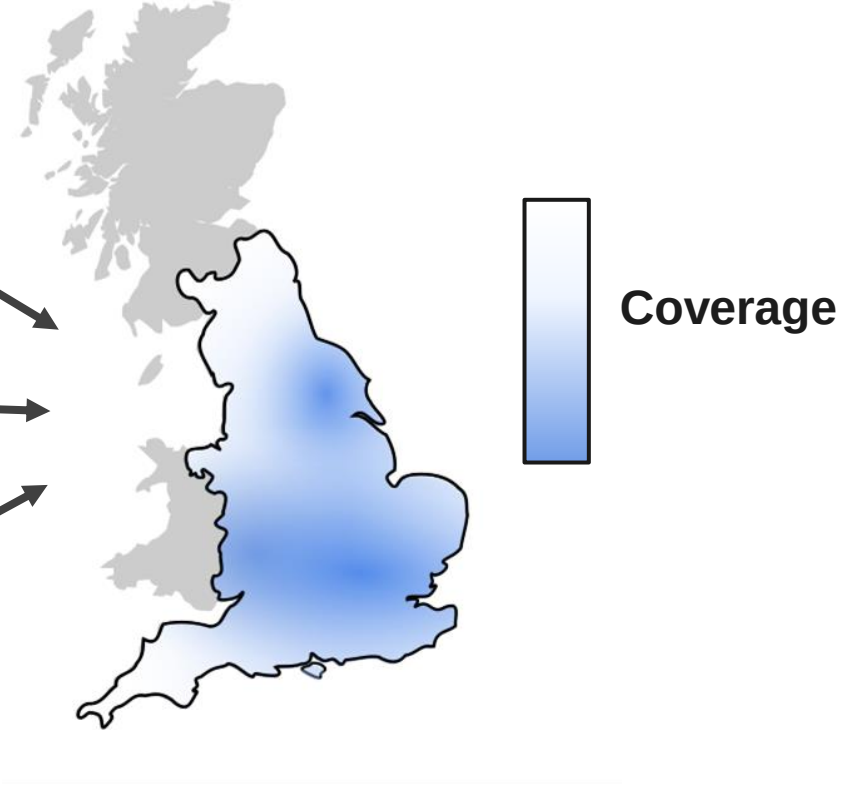
froglife

The Wildlife Trusts

WildFish.

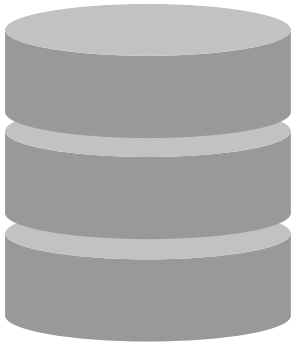
iSpot
ispotnature.org

seasearch



943 Priority species

1970 - 2020 in England



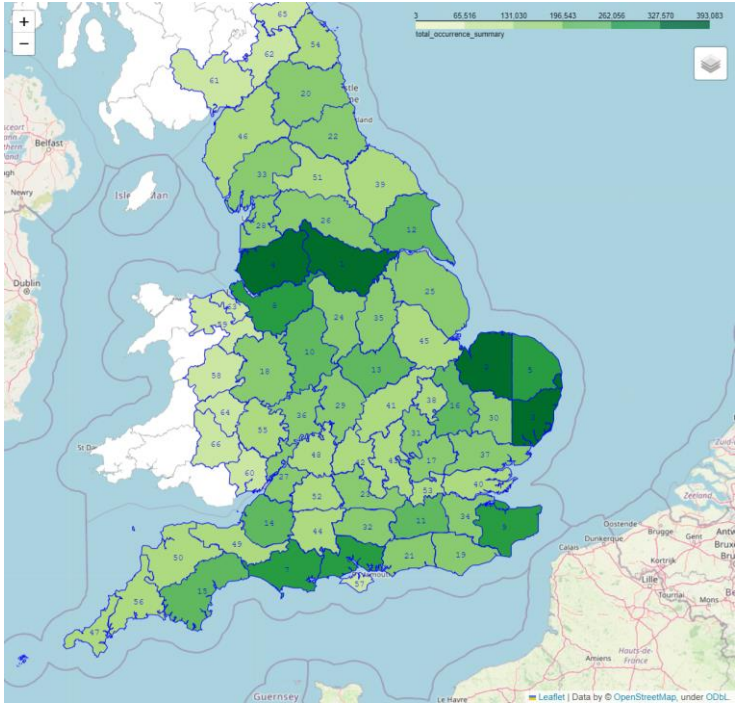
#	Name	Common Name	Taxon	Year	Location
1	*Arvicola amphibius*	Water vole	Mammal	2019	52.9388° N 1.1981° W
2	*Apamea anceps*	Large Nutmeg	Moth	1975	51.7548° N 1.2544° W
3	*Passer domesticus*	House Sparrow	Bird	2001	50.3759° N 4.1396° W

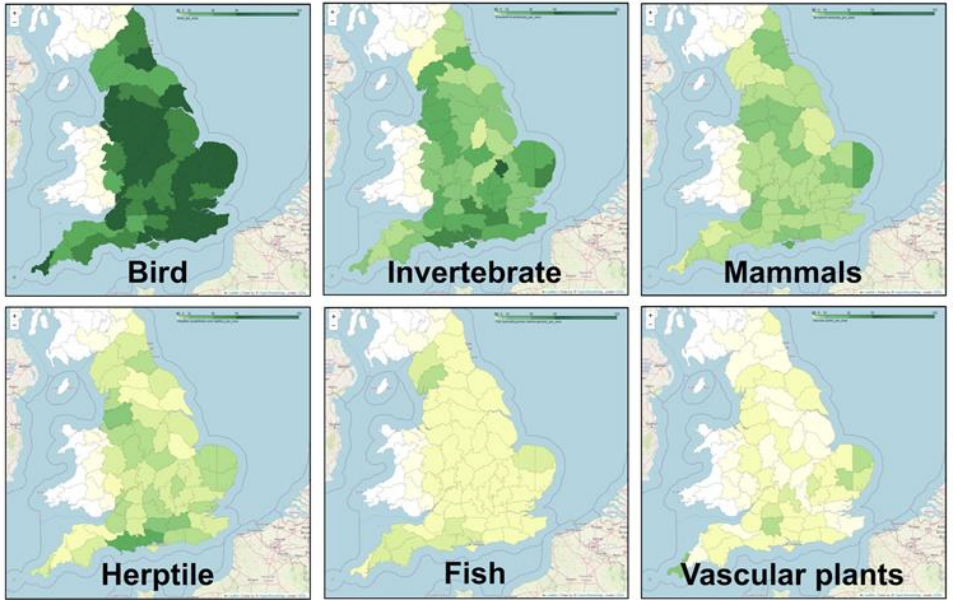
⋮

10,319,464 Observations

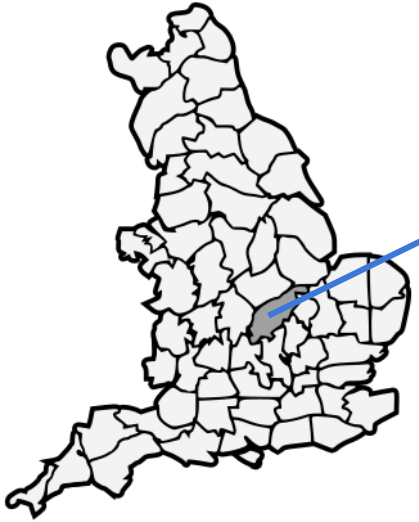
How did we approach the challenge?

Getting to grips with the data: geographical and taxonomic biases for priority species





Identifying gaps with reasonable comparisons

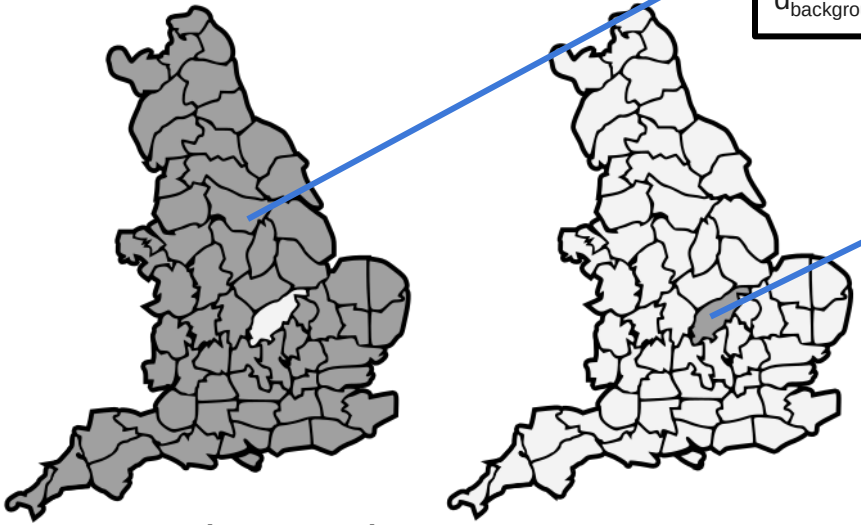


32 observations over 2,364 km²

d_{target} = 0.0135

Target

Identifying gaps with reasonable comparisons



Background

Target

5330 observations over 130,279 km²

$d_{background}$ = 0.0409

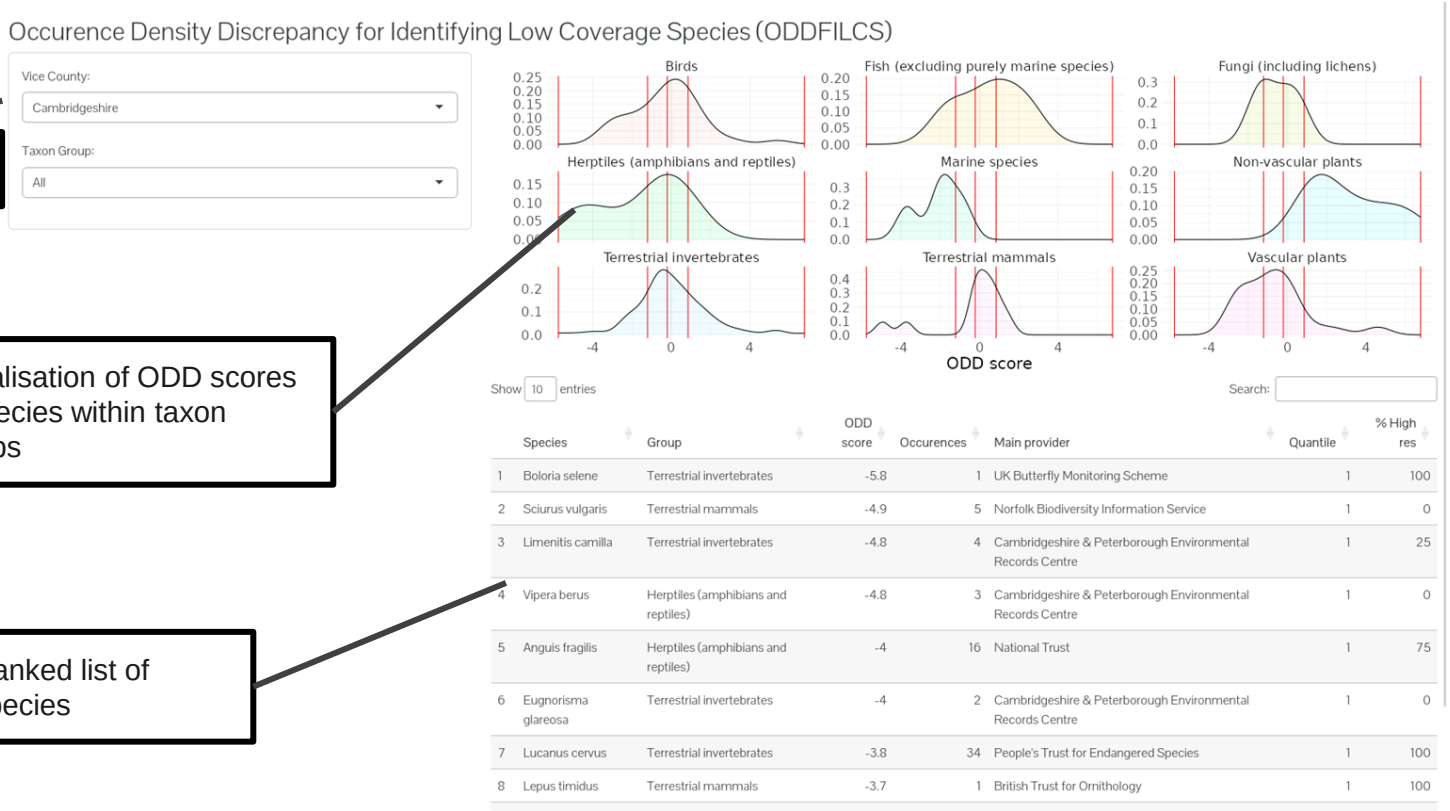
32 observations over 2,364 km²

d_{target} = 0.0135

ODD score = $\log(d_{target}/d_{background})$

= -0.481

Occurrence density discrepancy

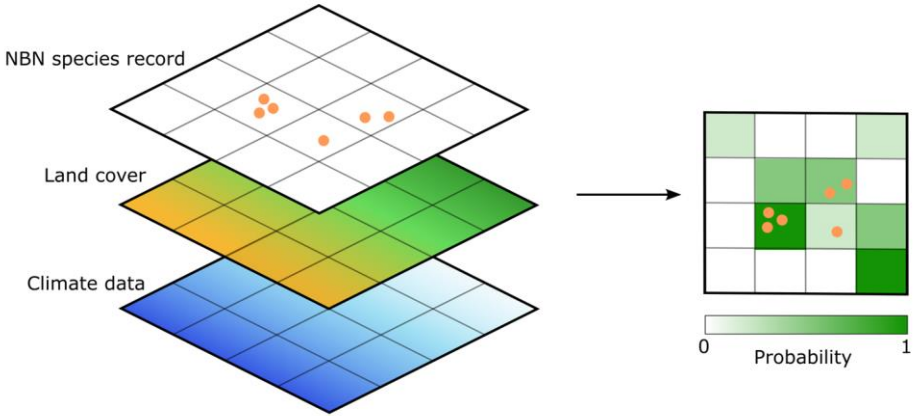


User controls

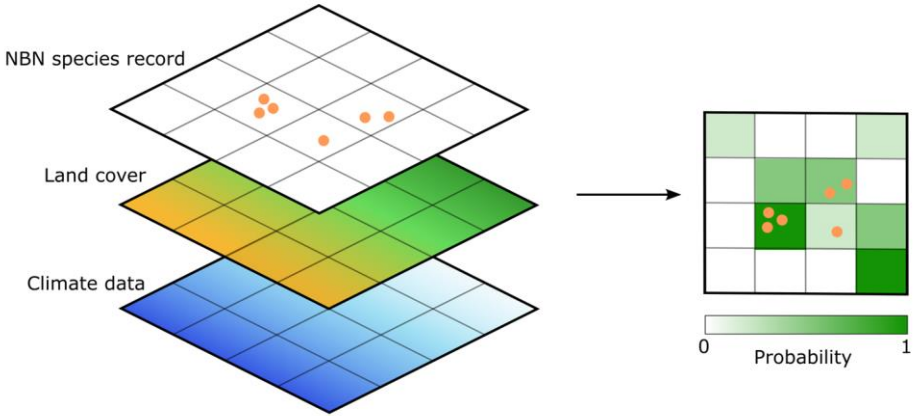
Visualisation of ODD scores of species within taxon groups

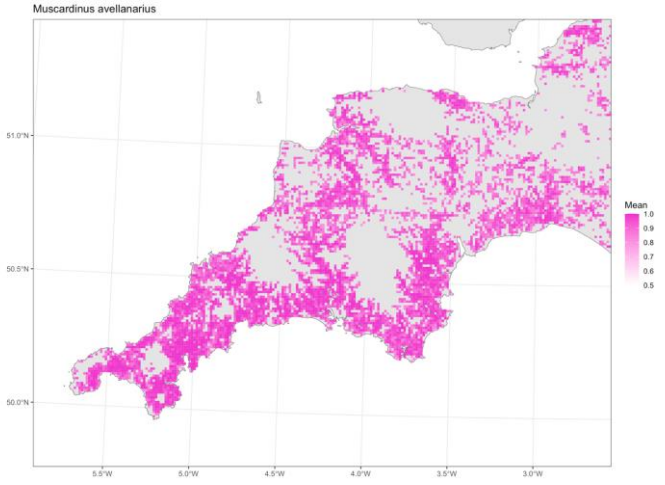
Ranked list of species

Environment-based gap identification



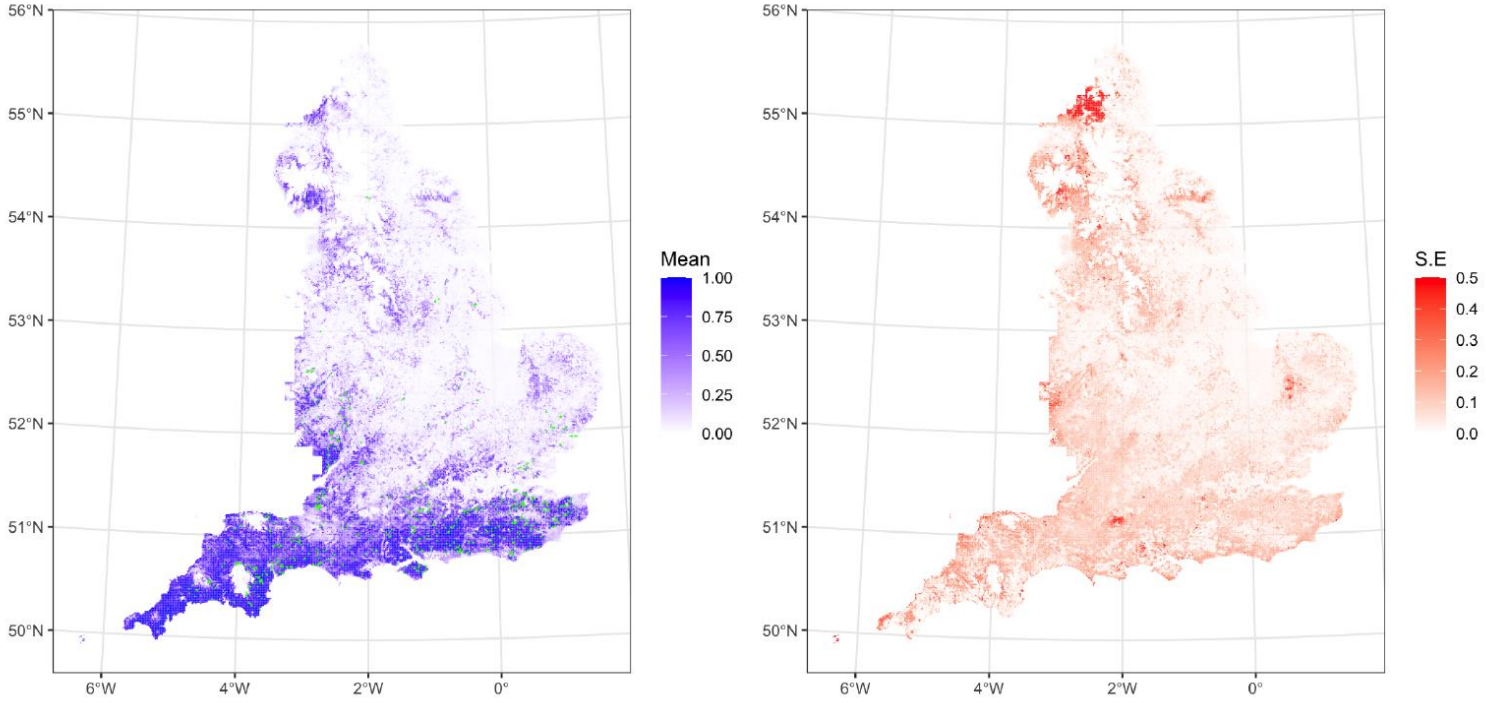
Environment-based gap identification





- ➔ Species specific
- ➔ Able to highlight areas of interest

Model-based adaptive sampling

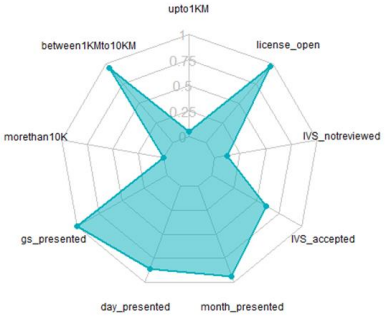


Maps showing the 2018 probability of presence (purple), standard error (red) and actual records (green) for *Muscardinus avellanarius*

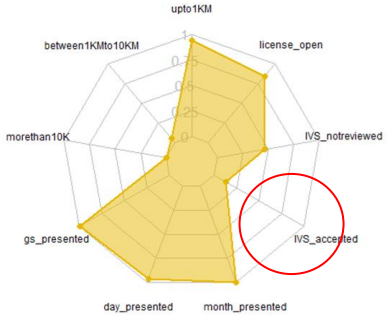
We found 4 groups of data providers

Spot place precision:	0-1 Km 1 Km to 10 Km More than 10 Km Not Provided
Date Information:	Day_provided Month_Provided
Data verification:	Accepted Unconfirmed Not Reviewed
Data usability:	Open Shared

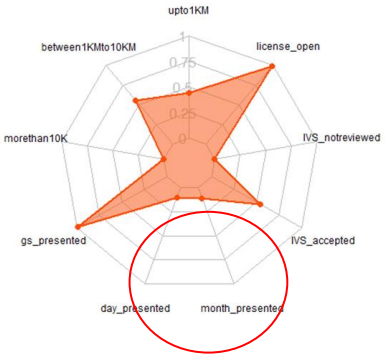
The regulars (30)



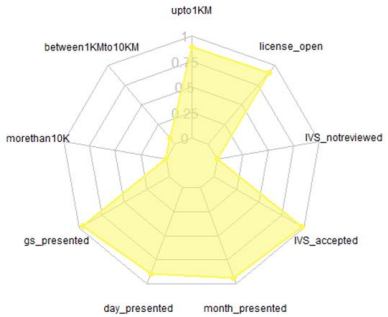
Unverified identification (17)



Incomplete date info (8)



Champions! (67)



Helping data providers understand their data

Prototype of a data provider dashboard

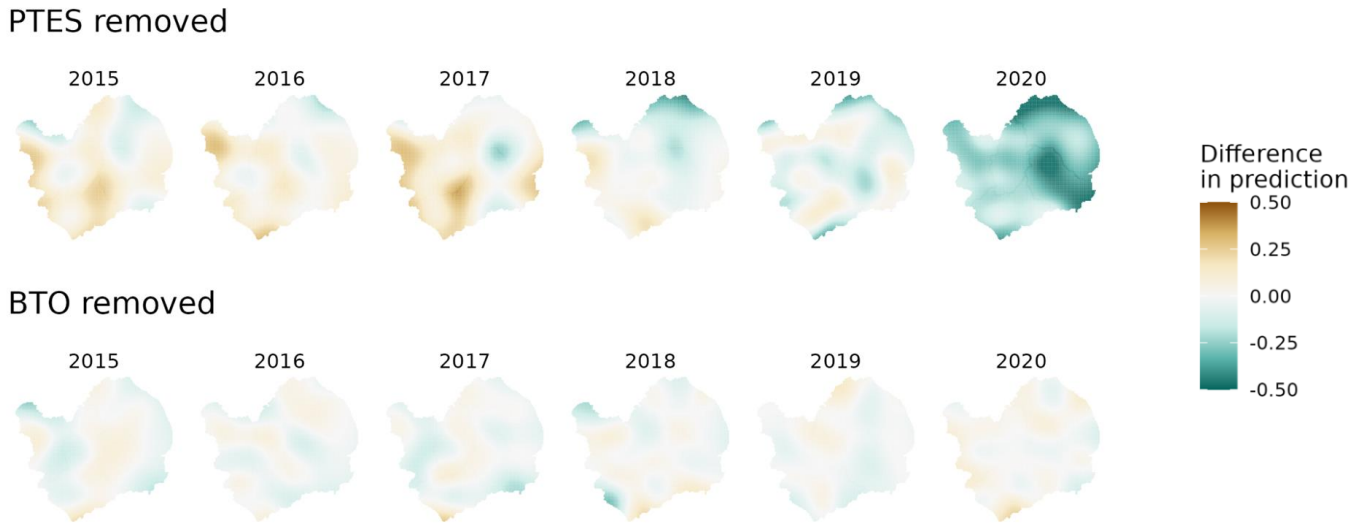
Map of records

Records over time

Visualising the impact of data providers

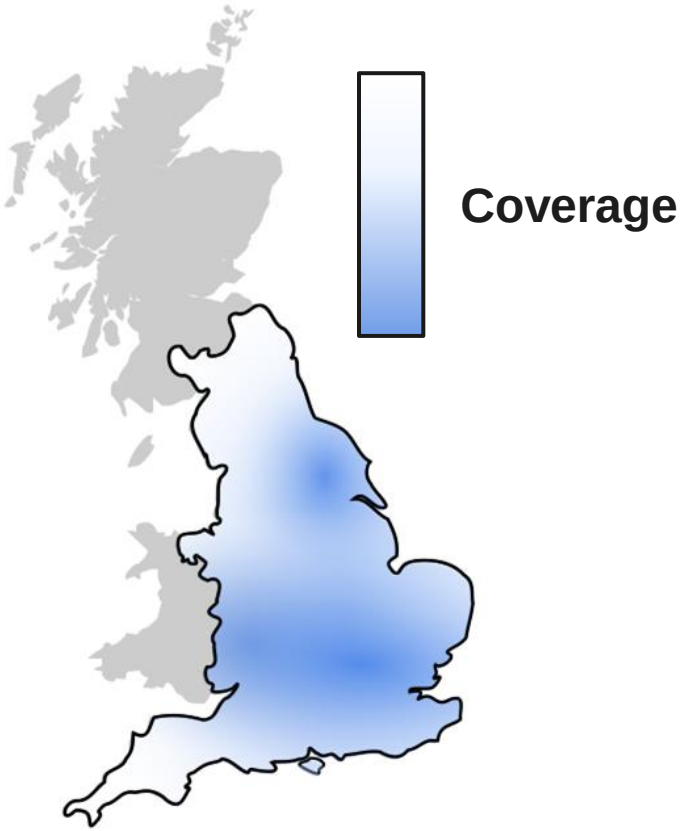
Can show influence of individual data providers by running species distribution models with and without their data

Example of hedgehog distribution in East Midlands:



Main conclusions

– Gaps are multifaceted and complex
– Reasonable comparisons and prioritisation help identify gaps
– Combining with environmental is important for identifying gaps with models



Reflections on a Data Study Group

- Time limited: successful ideas generation for future work, but many avenues unexplored
- Raised the profile of NBN Trust and the Atlas amongst data scientists
- Great learning experience for participants

Report is online now

Search “*NBN Turing*” to find the report

– Thanks for Natural England, the Alan Turing Institute and the NBN Trust for enabling this work