Big data in conservation

Monitoring bird populations through volunteer-collected observations

Philipp Boersch-Supan @pboesu

British Trust for Ornithology @_BTO

The British Trust for Ornithology

- We are an independent charitable research institute
- We count and ring UK birds with the help of over 40,000 volunteers
- We employ c. 100 staff
  - c. 50 scientists
  - c. 50 communicators, fundraisers, volunteer coordinators, administrators, software/database/web developers
- We don’t advocate but aim to provide impartial evidence to inform the public, opinion-formers and policy/decision-makers.

Image: DAVID TIPLING

About me

		Field	Lab	Maths	Stats	Coding
2005	Field Ornithologist					
2006-2008	UG in Chemistry					
2008-2009	MRes in Environmental Biology					
2009-2013	PhD in Marine Ecology					
2013-2014	Analyst at British Antarctic Survey					
2014-2018	Postdoc Quantitative Ecology					
2018-	Ecological Statistician					

What do I do as an **Ecological Statistician**?

I try to build models that

- lead to a better understanding of **biological processes**
- allow predictions outside the range of our data
- enable informed decisions for conservation actions

- make the best use of limited (and sometimes ‘messy’) data
- are statistically and computationally **tractable**
- use tools/workflows that are easy to reuse and share

Ecological models

@pboesu

Tactical/Phenomenological

Strategic/Mechanistic

Describe patterns without elucidating mechanism

Focus on process/mechanisms

Prediction within range of data

Explanation or understanding
Prediction outside range of data

Statistical models (regressions, etc.)

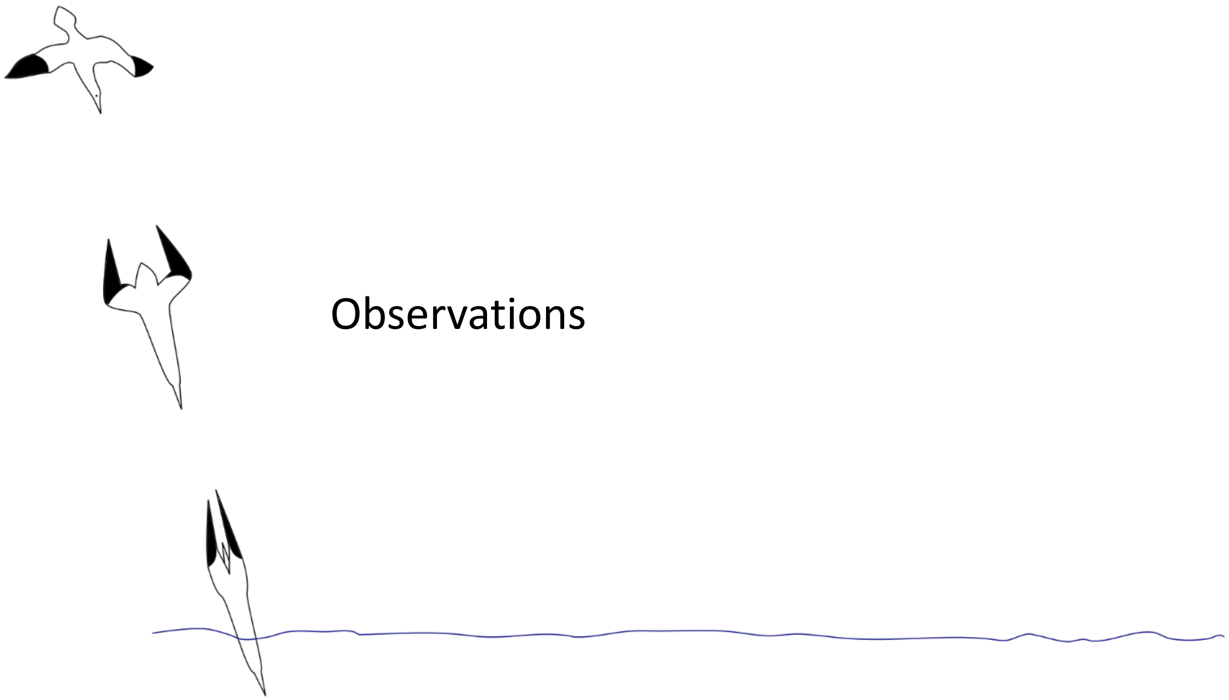
Math models e.g. ODEs, PDEs
Individual/Agent based Models

Historically largely separate communities within ecology

tendency to ignore ecological theory
& focus on "significance"

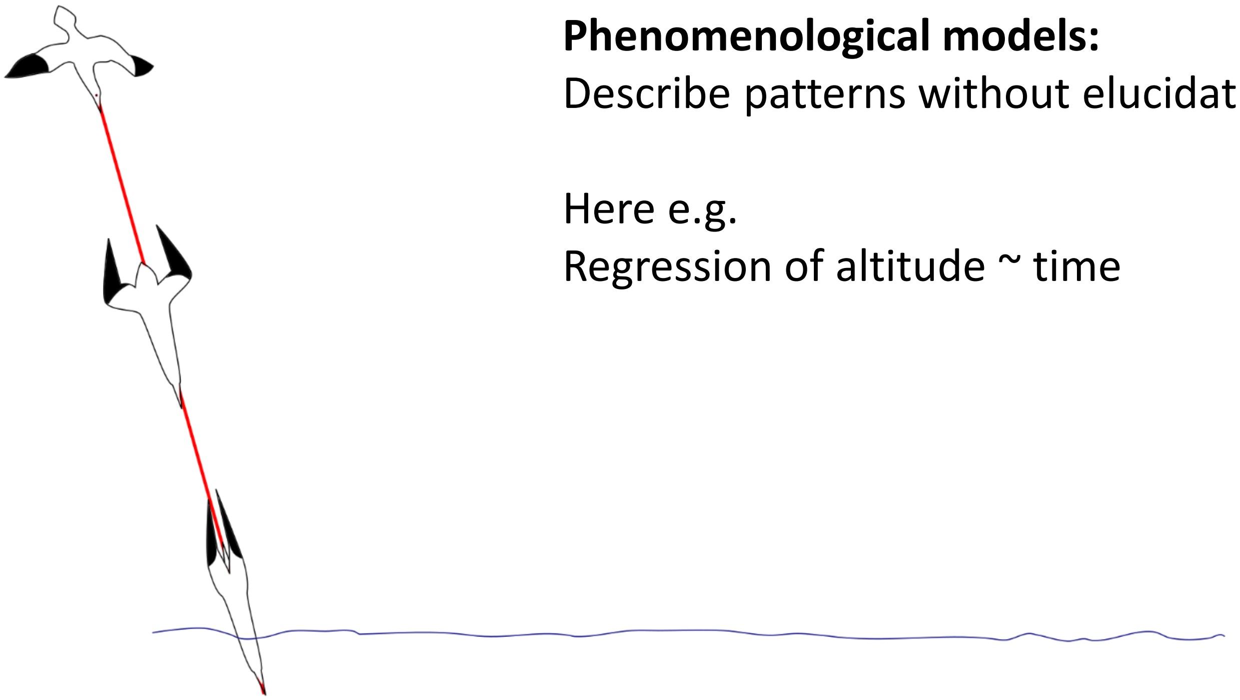
tendency to ignore ecological reality
& focus on mathematical properties

Inspired by Leah Johnson



Observations

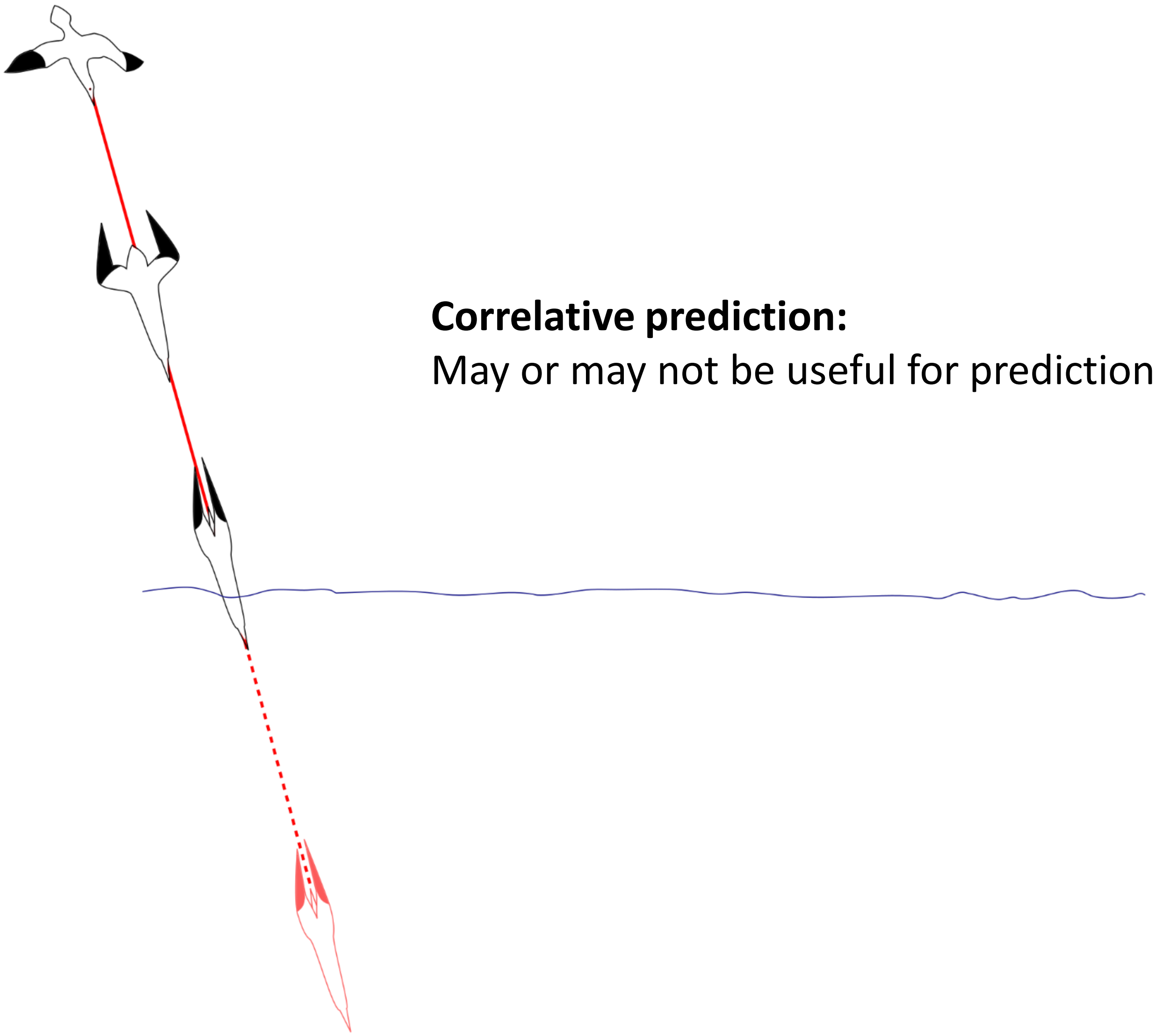
inspired by Maino et al. (2016) Curr Opp Insect Sci



Phenomenological models:
Describe patterns without elucidating mechanism

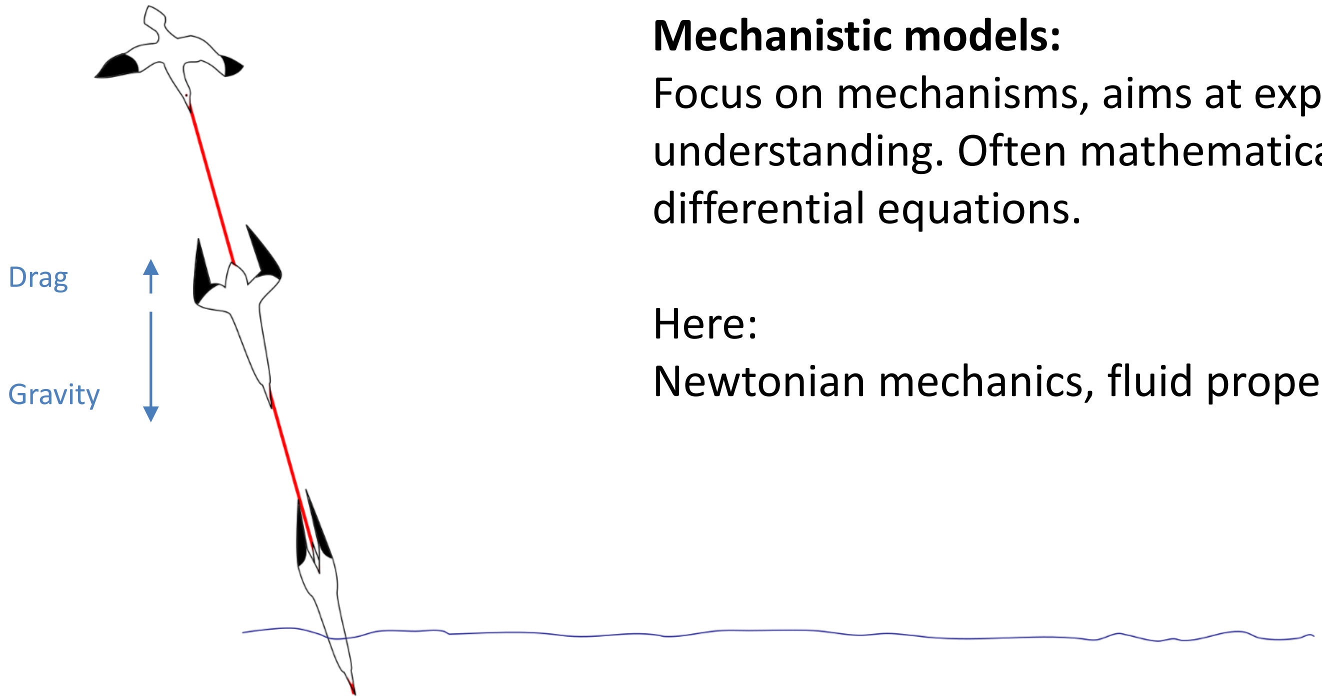
Here e.g.
Regression of altitude ~ time

inspired by Maino et al. (2016) Curr Opp Insect Sci



Correlative prediction:
May or may not be useful for prediction beyond data

inspired by Maino et al. (2016) Curr Opp Insect Sci

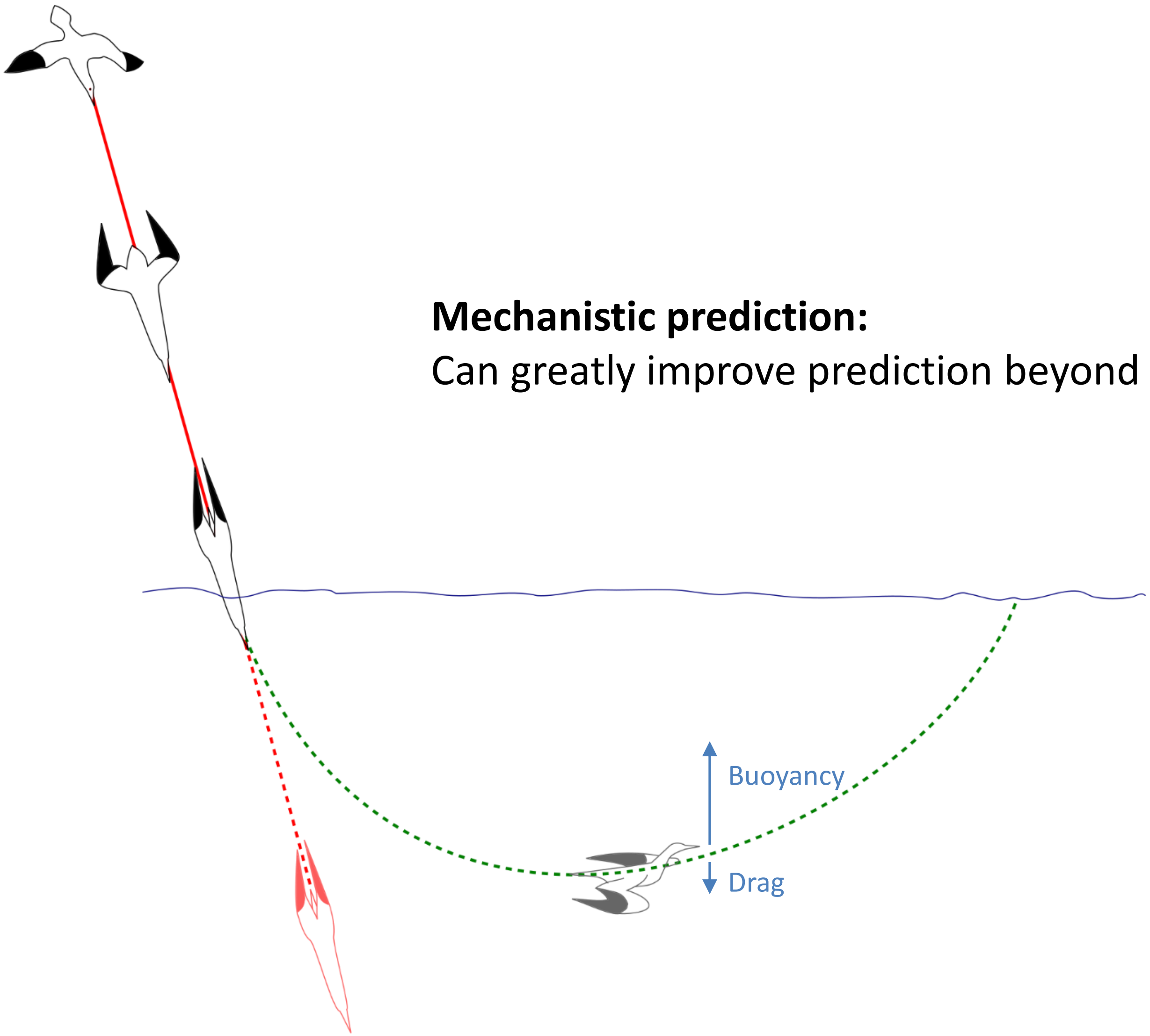


Mechanistic models:
Focus on mechanisms, aims at explanation or understanding. Often mathematical models like differential equations.

Here:
Newtonian mechanics, fluid properties

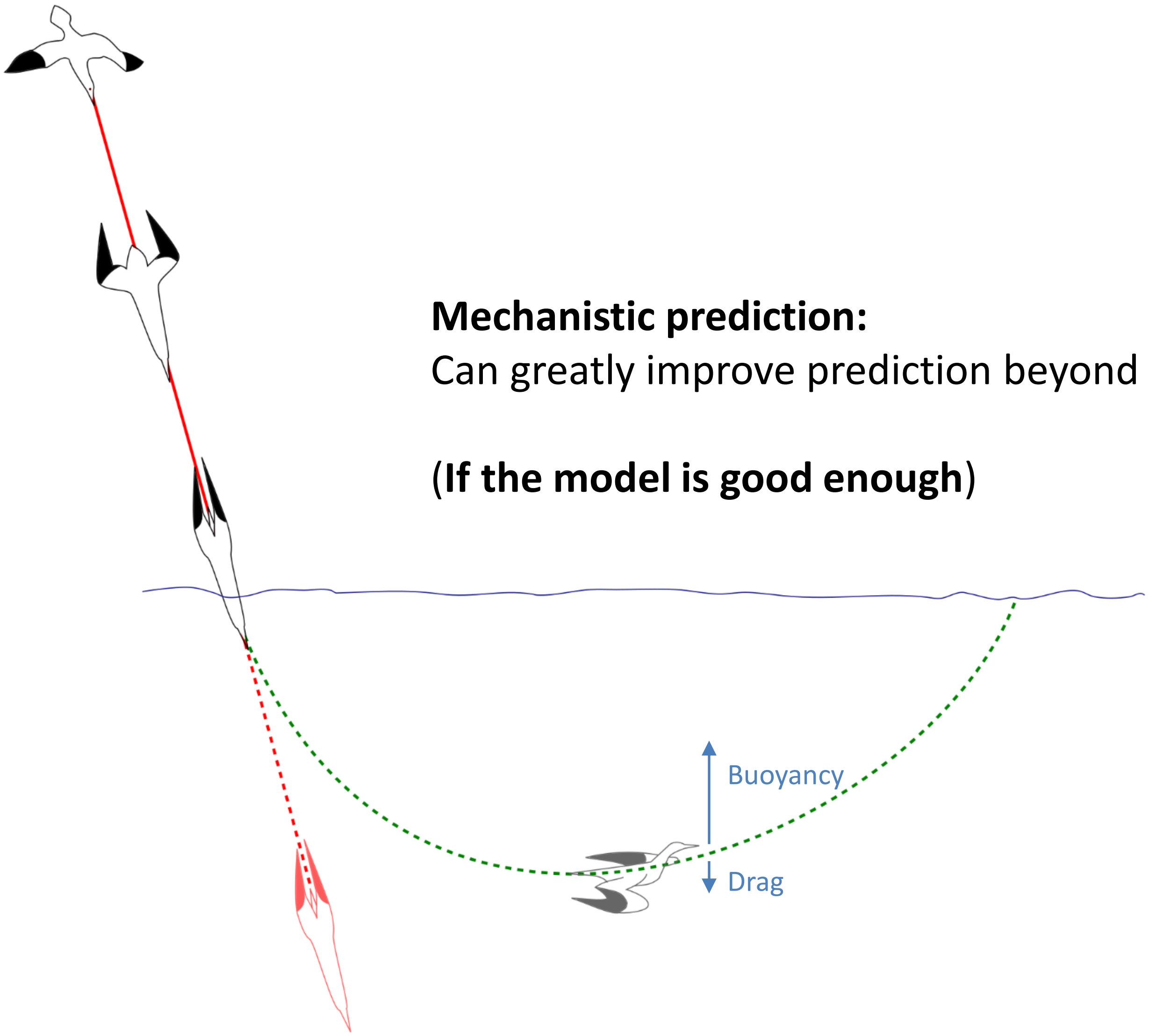
Other examples:
Lotka-Volterra predator-prey models
SIR model in epidemiology

inspired by Maino et al. (2016) Curr Opp Insect Sci



Mechanistic prediction:
Can greatly improve prediction beyond data

inspired by Maino et al. (2016) Curr Opp Insect Sci



Mechanistic prediction:
Can greatly improve prediction beyond data

(If the model is good enough)

inspired by Maino et al. (2016) Curr Opp Insect Sci

Tactical/Phenomenological | Strategic/Mechanistic

Describe patterns without elucidating mechanism

Prediction within range of data

Statistical models (regressions, etc.)

BBS

UK Breeding Bird Trends

Focus on process/mechanisms

Explanation or understanding
Prediction outside range of data

Math models e.g. ODEs, PDEs
Individual/Agent based Models

BTO Nest Record Scheme

BBS

Ringing Schemes

Integrated Population Models

BTO data collection schemes

- Demographic Monitoring
  - Ringing Schemes
  - Nest Records Scheme
- Population Monitoring
  - Breeding Bird Survey
  - Wetland Birds Survey
  - Garden Birdwatch
  - Bird Track
- Movement and Migration Monitoring
  - Species-specific tagging projects
- Other Surveys
  - Atlases (comprehensive distribution mapping in ~20 year cycles)
  - Species-specific surveys (e.g. Tawny Owl Surveys 2018)
  - …

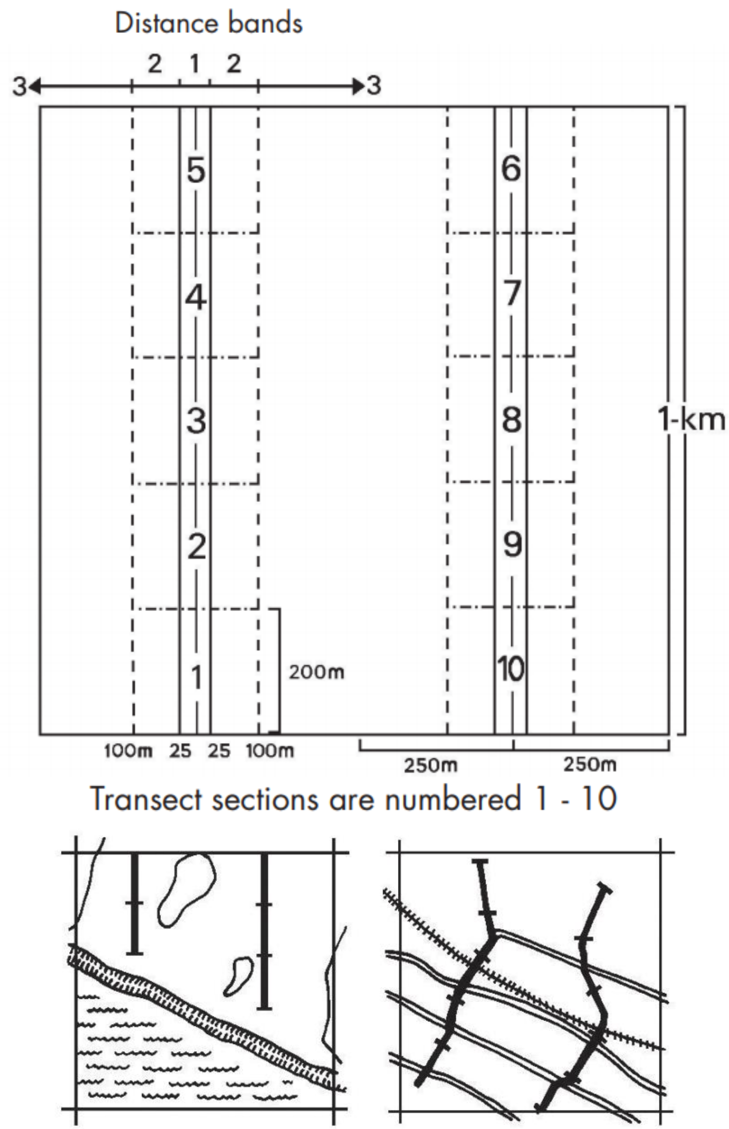
How do we count UK breeding birds?

- Breeding Bird Survey (BBS)
  - strict count protocol of all birds encountered
  - line transects, distance sampling, known survey effort
  - randomized site selection, high coverage (1.66% of UK!)

  → Big data (25 years, >4000 sites)

BBS Design

- Standardised annual surveys
- Randomly selected 1km squares
- Two visits per year, 4 weeks apart in April-June
- Two distance sampling transects
- All birds seen or heard are recorded.



BBS Trend Model

- BBS trends are based on a Poisson GLM

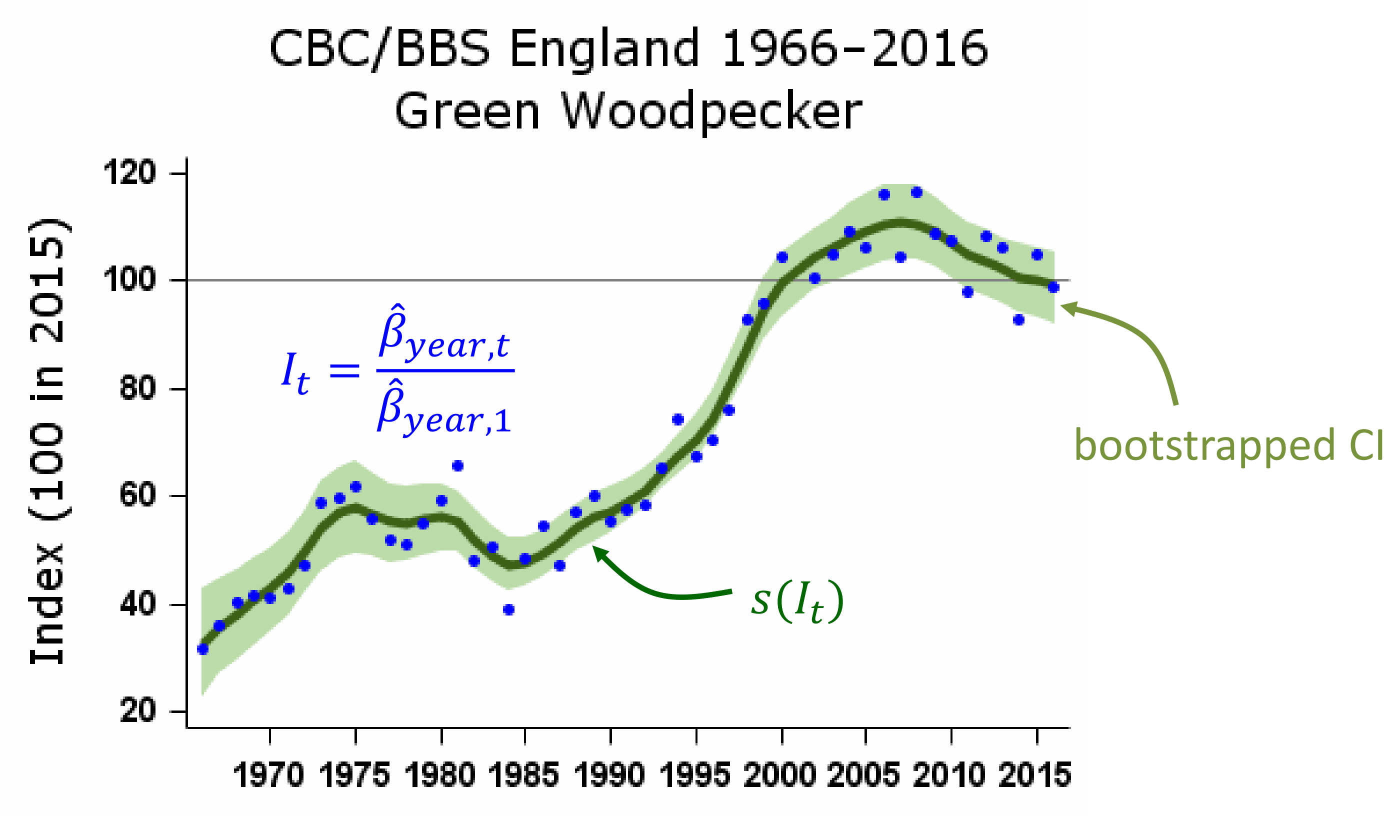
$$N_{max,it} \sim Poisson(\lambda_{it})$$
$$\log \lambda_{it} = \beta_{year,t} + \beta_{site,i} + \log N_{segments,i}$$

- Observations are weighted to account for uneven observer availability across the UK
- weights = inverse sampling probability within stratum

- Relative abundance indices $I_t = \frac{\widehat{\beta}_{year,t}}{\widehat{\beta}_{year,1}}$
- I_t are smoothed post-hoc using a thin-plate spline with 0.3*years df

Fewster et al. (2000) Ecology

BBS Trends



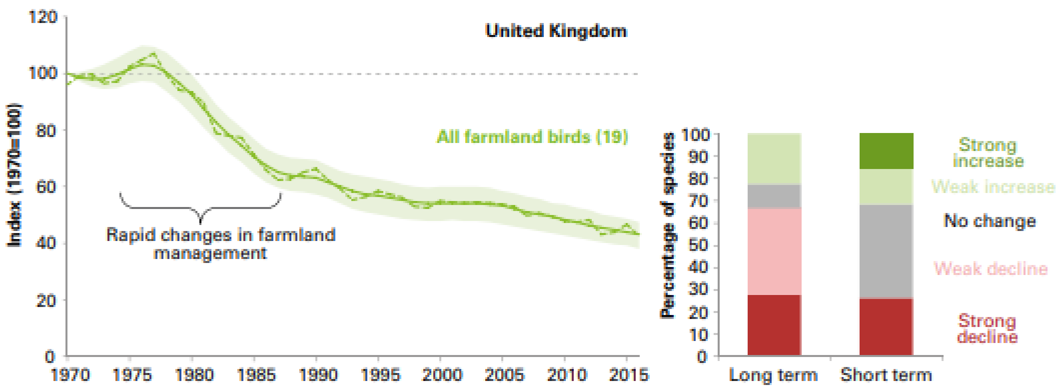
- BBS index: average count relative to a reference year (here 2015) $I_t = \frac{\widehat{\beta}_{year,t}}{\widehat{\beta}_{year,2015}}$
- Index values are smoothed to highlight longer term (3-5 year) patterns

Simple model + big(ish) data ≠ trivial fitting

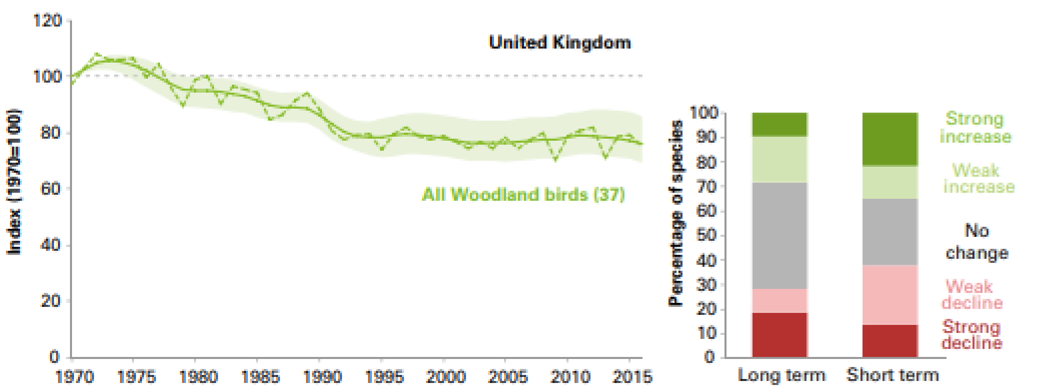
- As the survey lifetime and site coverage has increased, so has the size of the ($n \times p$) design matrix
- Time complexity of fitting a GLM scales with p^3
- Example: Wren 1994-2016
    - 58110 records across 6004 distinct squares:
    - (6004 sites + 23 years) * 58110 records * 8 bytes = 2.6 GB to store a single copy of the design matrix
    - Fitting the GLM naively in R takes c. 15 GB of memory and several hours
- The bootstrap requires every model to be fitted 200 times
- Running the annual trends currently takes 3-4 weeks

Uses of BBS data

Breeding farmland birds in the UK, 1970 to 2016.



Breeding woodland birds in the UK, 1970 to 2016.



State of the Birds 2017 – RSPB/BTO/WWT/DAERA/JNCC/NE/NRW

- Multi-species trend indicators are widely used in policy and communication

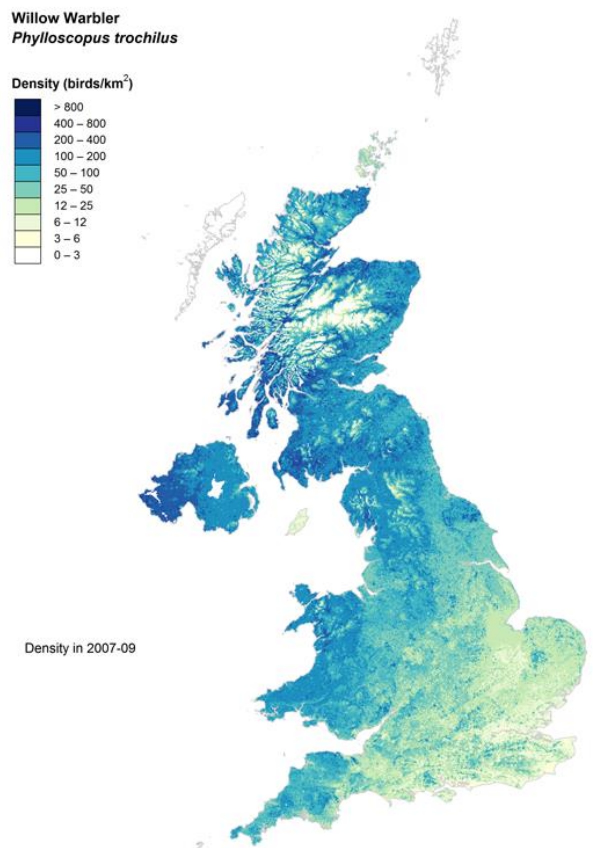
- Geometric mean of species abundance indices

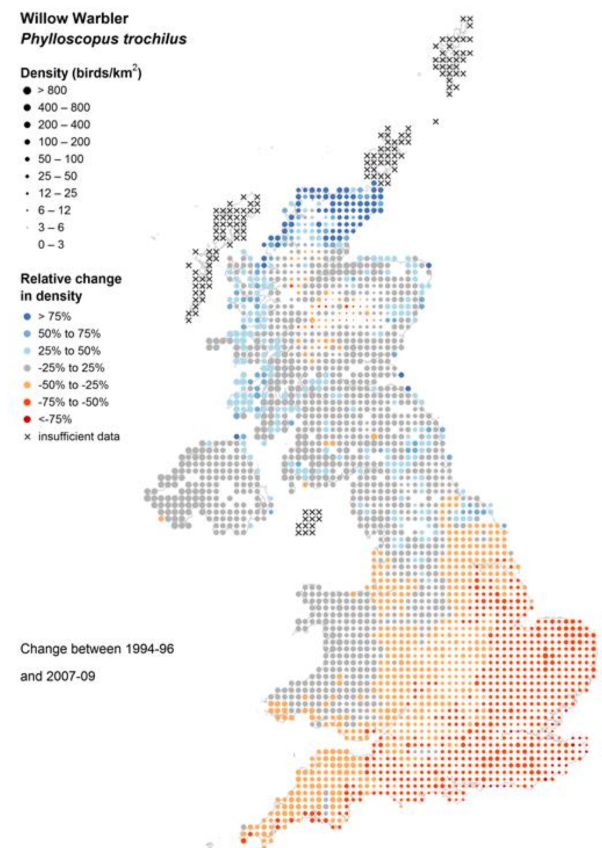
$$\sqrt[n]{I_{t,1}I_{t,2}\dots I_{t,n}}$$

Uses of BBS data

- Density and trend maps using GAMs and habitat data
- Computational tractability requires coarser spatial/temporal scale

$$\log \lambda_{it} = s(lat_i, lon_i) \quad + \boldsymbol{\beta}_{habitat,i} + \log N_{segments,i}$$





Massimino et al. 2015 Ecol Indic, Massimino et al. 2018 Biol Cons

Constant Effort Ringing - CES


Image: DAVID TIPLING

- BTO issues ringers permits and oversees training in the UK
- Qualified ringers are free to ring selected sites
- CES is a site-specific structured scheme
  - Sites are volunteer-selected
  - mainly reedbed, scrub, woodland
  - regular habitat management
  - 12 annual capture events, 10 days apart May-August
  - same nets in same positions
  - no lures, bait, etc.
  - all captured birds are processed, focus on 24 songbird species
- Other schemes exist to target non-passerines

Ringing Scheme Objectives

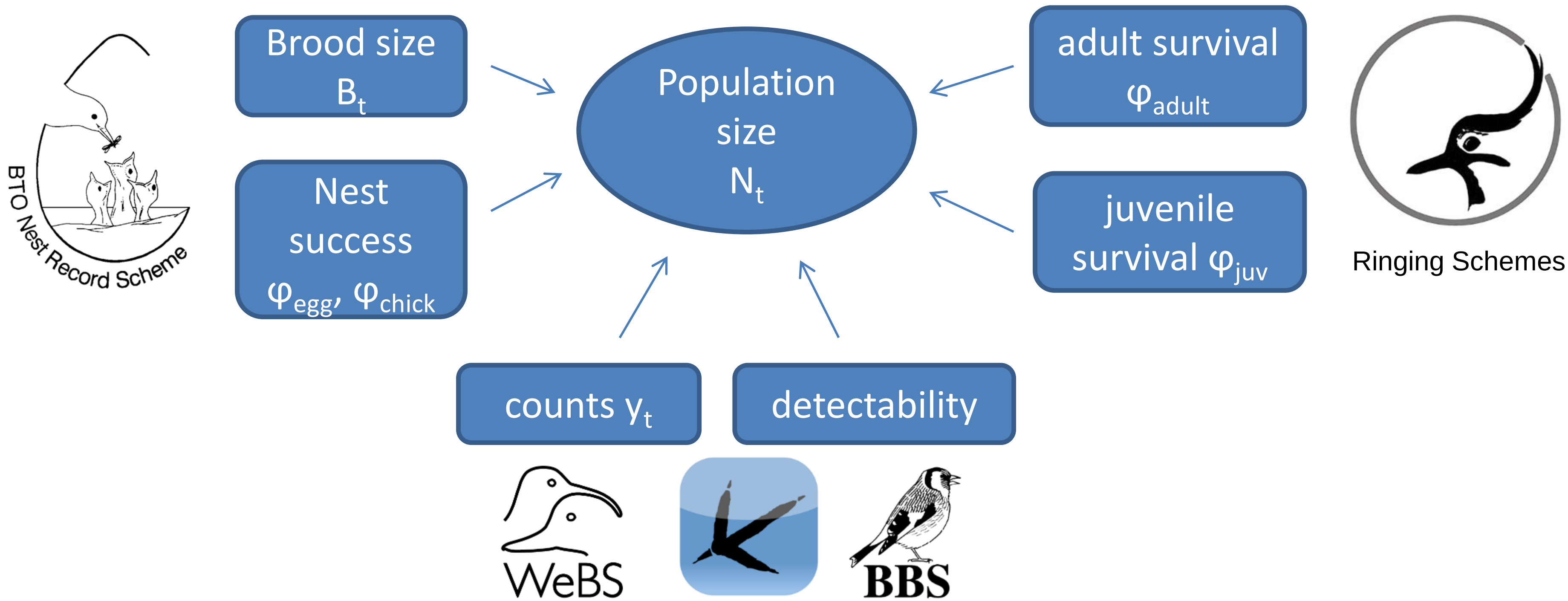
- Measure change in local breeding population size
  - From adult capture rates
- Measure changes in productivity
  - From juvenile to adult ratios
- Measure survival and recruitment
  - From recaptures across seasons

Nest Records Scheme

- Volunteer selected sites, species
- Multiple visits during the breeding season to record
  - Nest status
  - Numbers of eggs
  - Numbers of chicks
  - Fledging success

Integrated Population Monitoring

- **Mechanistic population dynamics** models that integrate different data sources to explicitly account for demographic processes



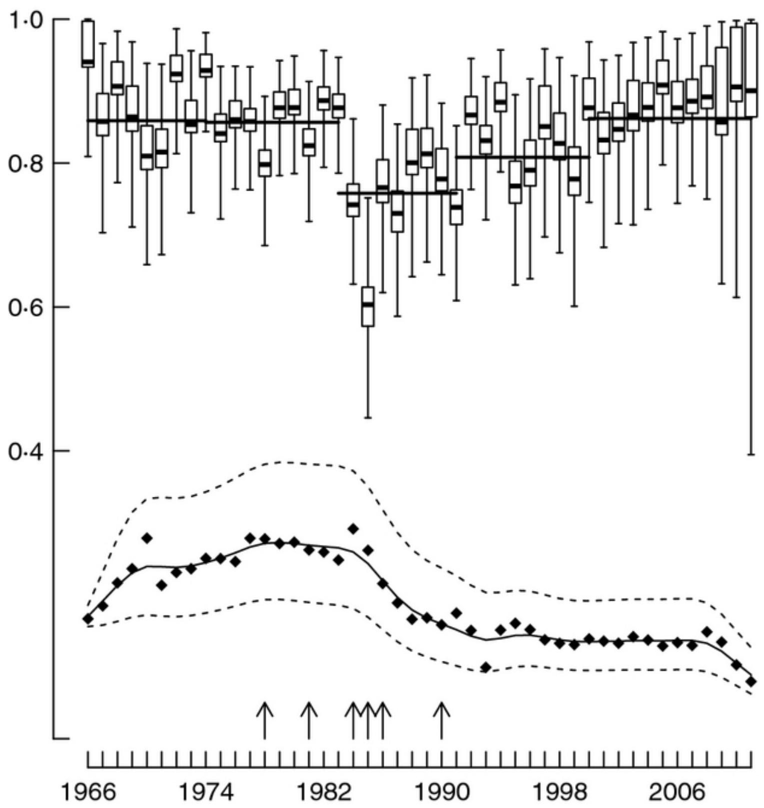
State/process model $N_{t+1} = \underbrace{N_t \rho_t (B_t \varphi_{egg,t} \varphi_{chick,t} \varphi_{juv,t})}_{\text{recruitment}} + \underbrace{N_t \rho_t (\varphi_{adult,t})}_{\text{survival}}$

Observation model, e.g. $y_t \sim \mathcal{F}(N_t, \sigma)$

Robinson et al. 2014, Methods Ecol Evol

Integrated Population Models

- The fitted model provides a **biological explanation** for the shape of the population trajectory



← Adult survival estimates

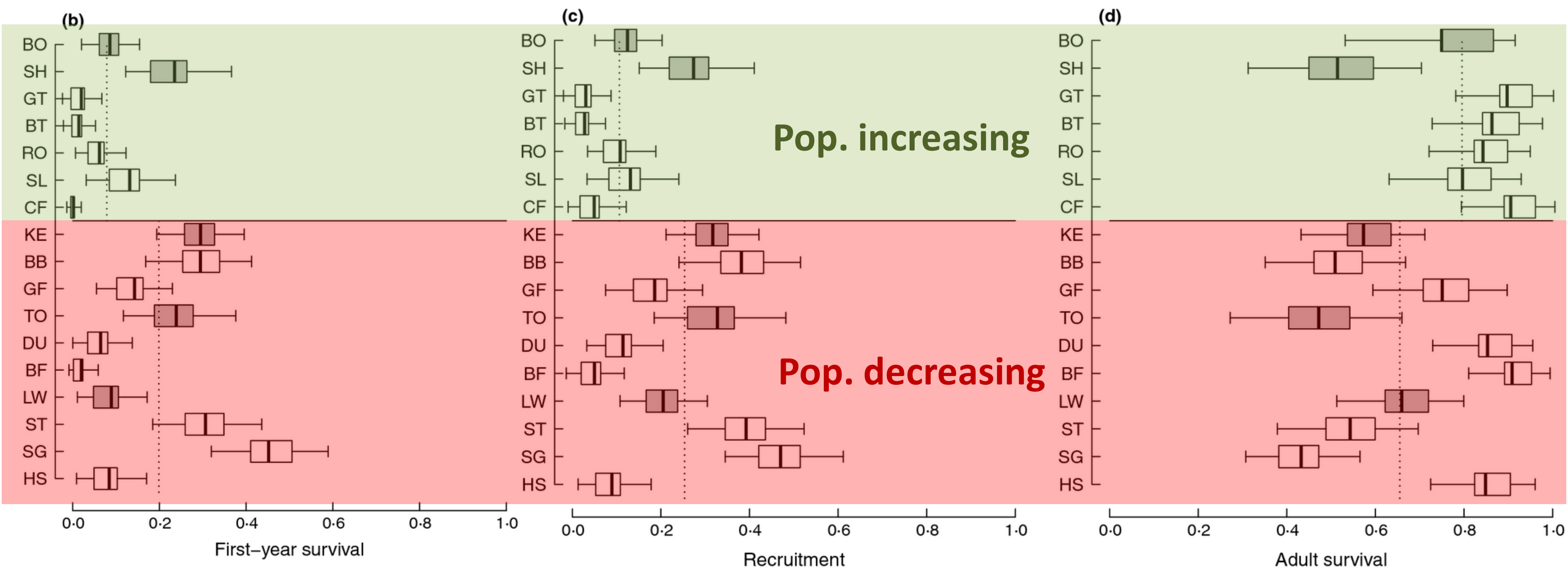
← Population trend (not to scale)

arrows indicate hard winters

Robinson et al. 2014, Methods Ecol Evol

Integrated Population Models

- **Biological explanation** can aid in better target management actions



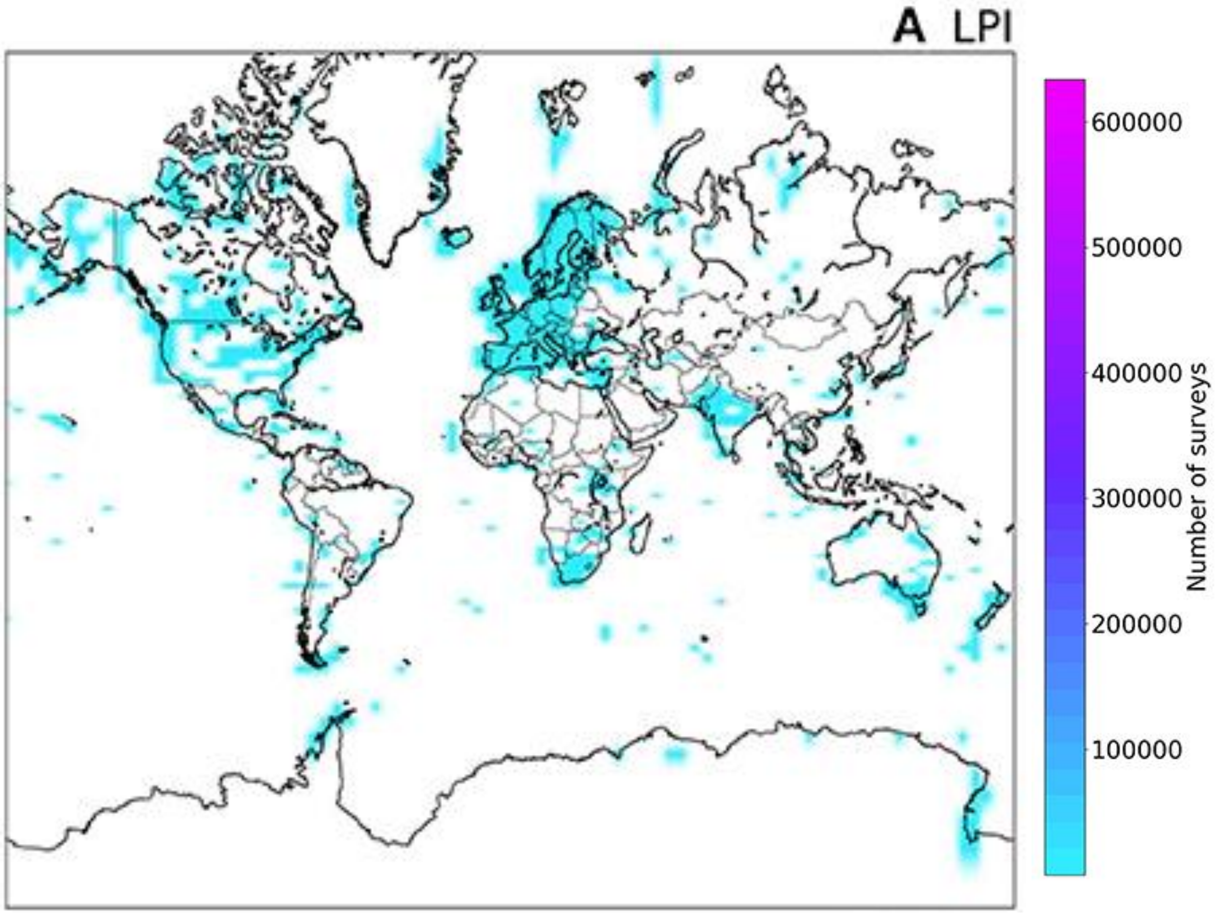
x-axis gives relative contribution to population growth rate of each species

Robinson et al. 2014, Methods Ecol Evol

Limitations of BBS Trends and IPMs

- Designed for inferences at UK/country level
  - Many end-users would like small area trends to inform management.
- Many species of conservation concern are not abundant or widespread enough
  - BBS designed to capture abundant and widespread species (~120 species trends reported)
  - IPMs are even more data hungry. (~20 spp.)

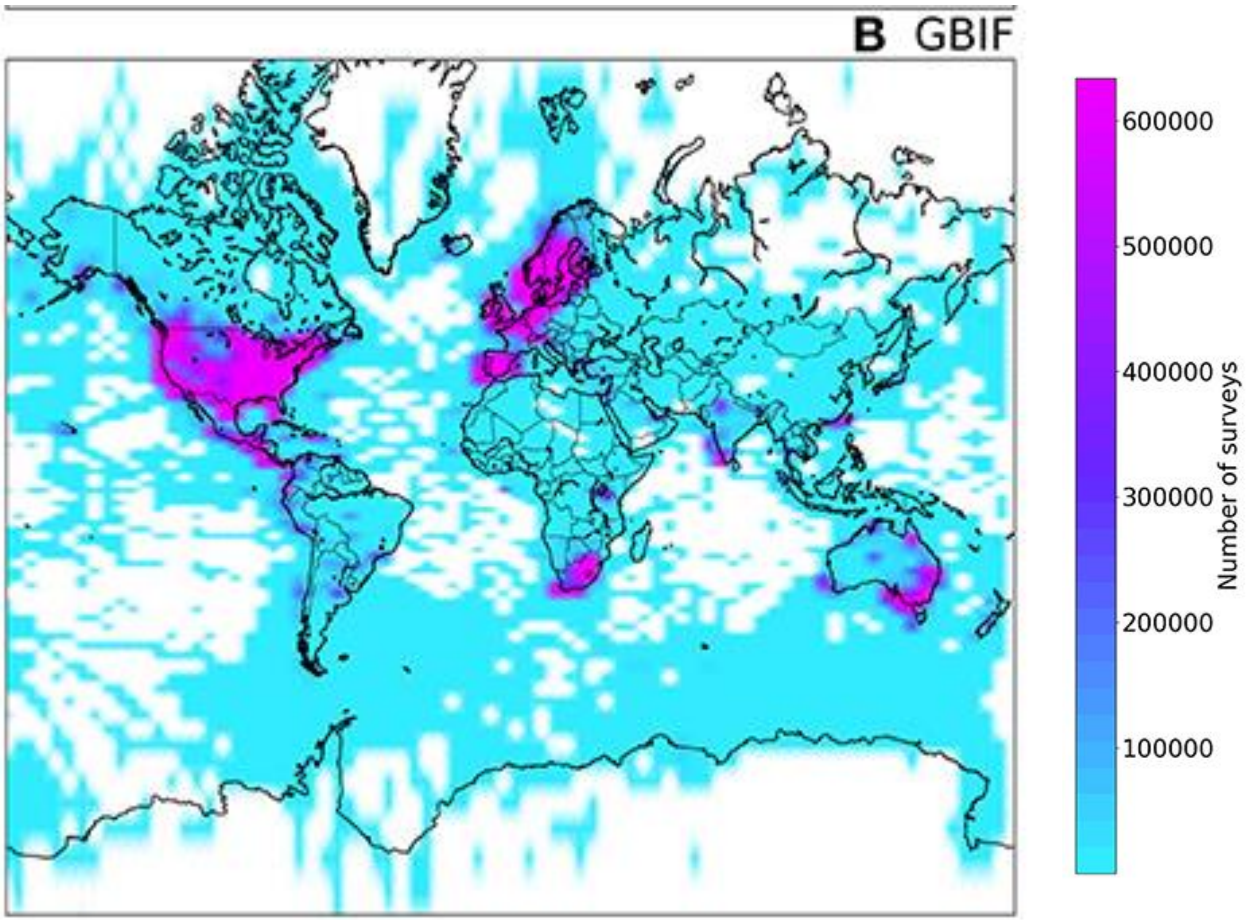
Structured biodiversity monitoring globally



Living Planet Index bird data 1970-2018: 9,230 time-series using consistent monitoring

Bayraktarov et al. 2019 Front Ecol Evol https://doi.org/10.3389/fevo.2018.00239

Unstructured biodiversity records globally



GBIF bird data 1970-2018: 537,424,092 surveys at 16,800,224 distinct survey locations

1000 times more data = 1000 times more information?

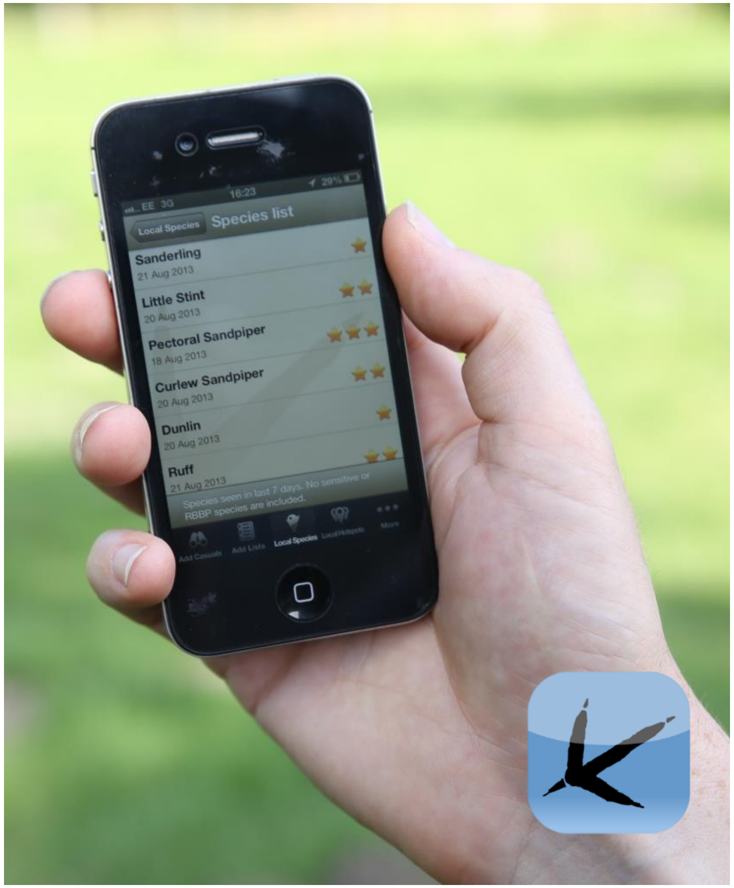
Bayraktarov et al. 2019 Front Ecol Evol https://doi.org/10.3389/fevo.2018.00239

Case study: BirdTrack

- (Web)App for personal bird record keeping (cf. eBird, iNaturalist)

- No fixed observation protocol
  - sites self-selected
  - complete listing optional
  - counting optional
  - effort recording optional

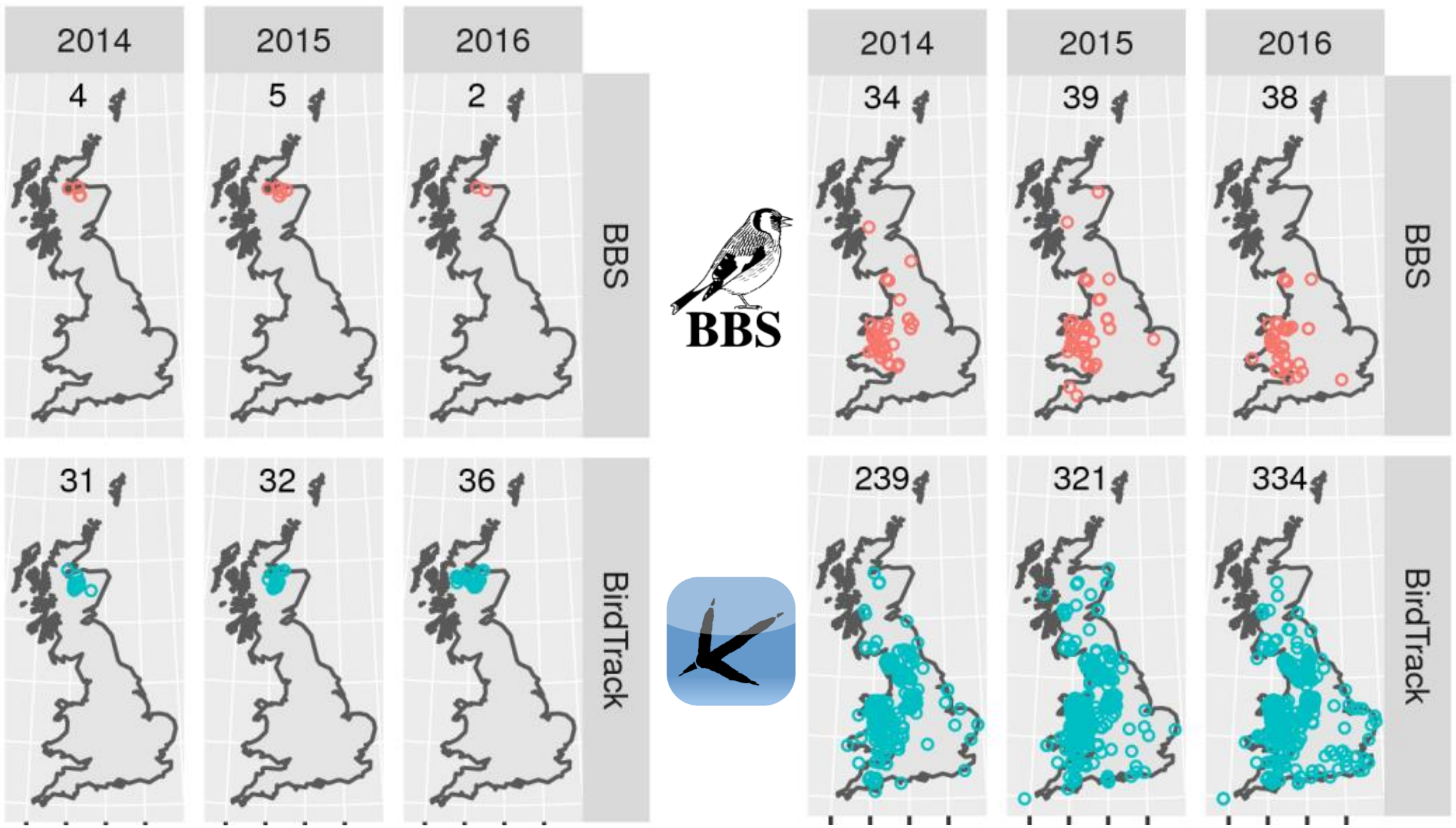
→ **Big data (currently ~15k sites, ~100k lists, ~7 million records per year)**



BirdTrack has c. 10x more records across space

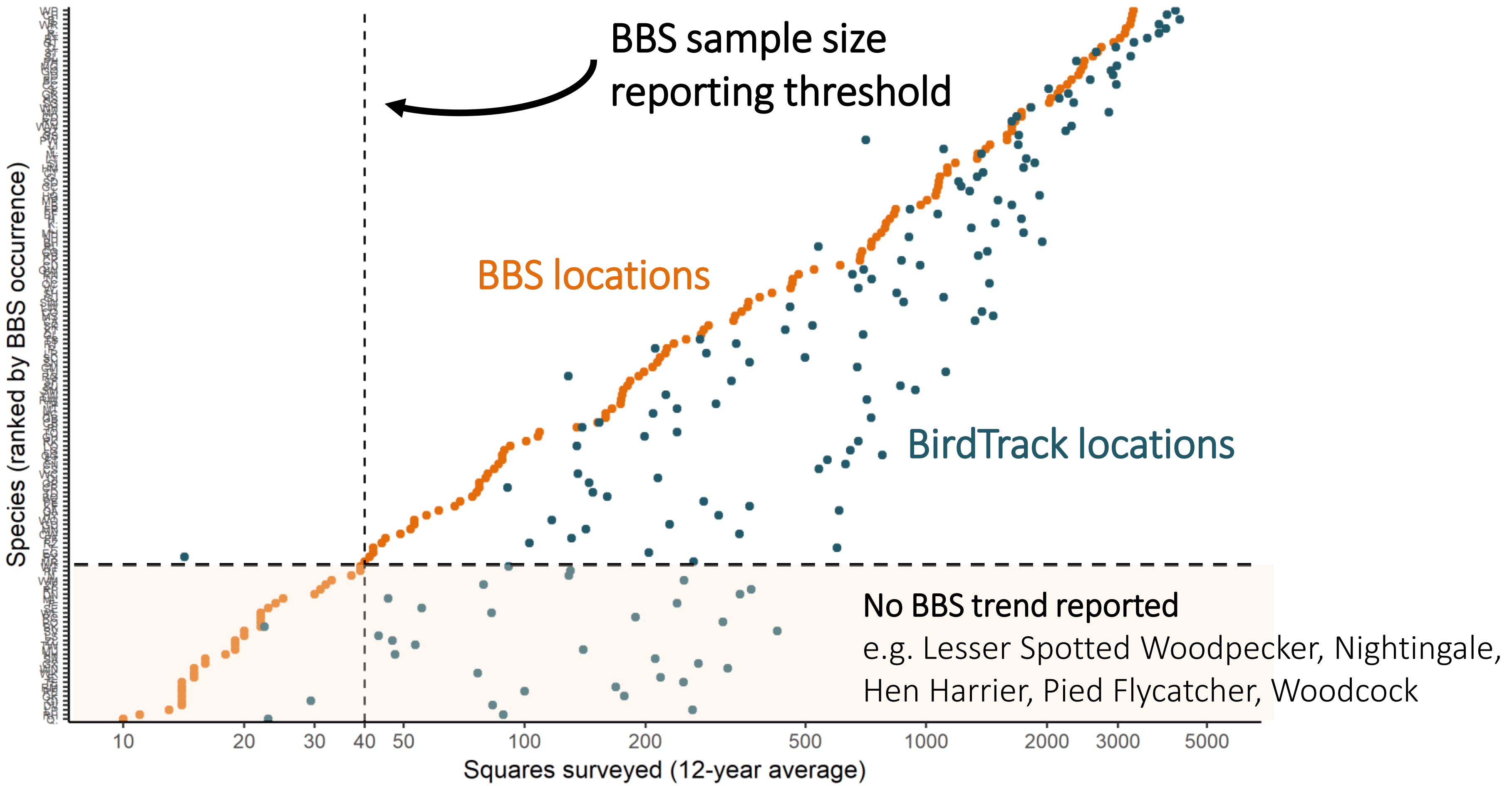
Crested Tit

Pied Flycatcher



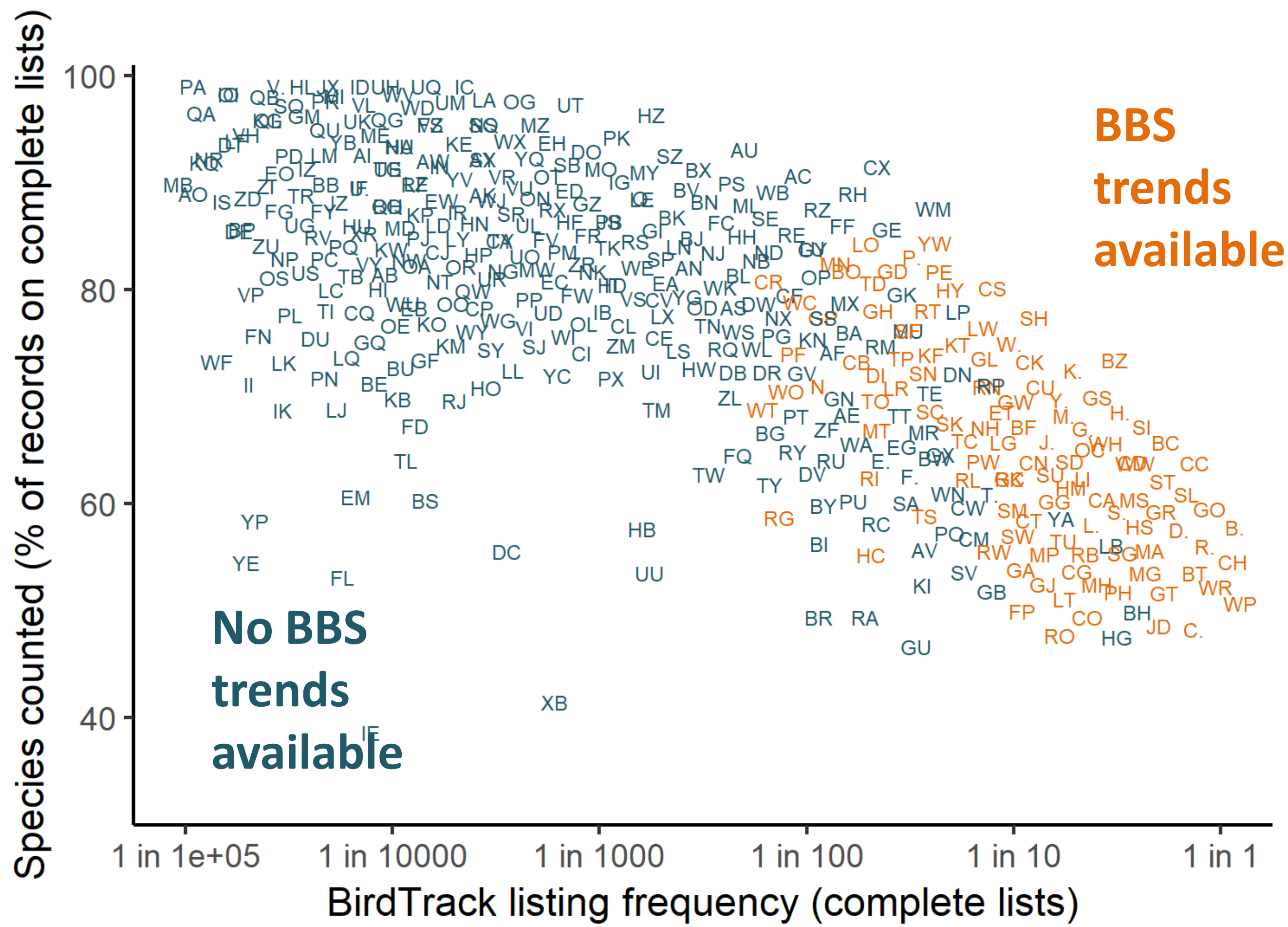
BirdTrack has c. 10x more records across space

BirdTrack captures species in many more places than the BBS



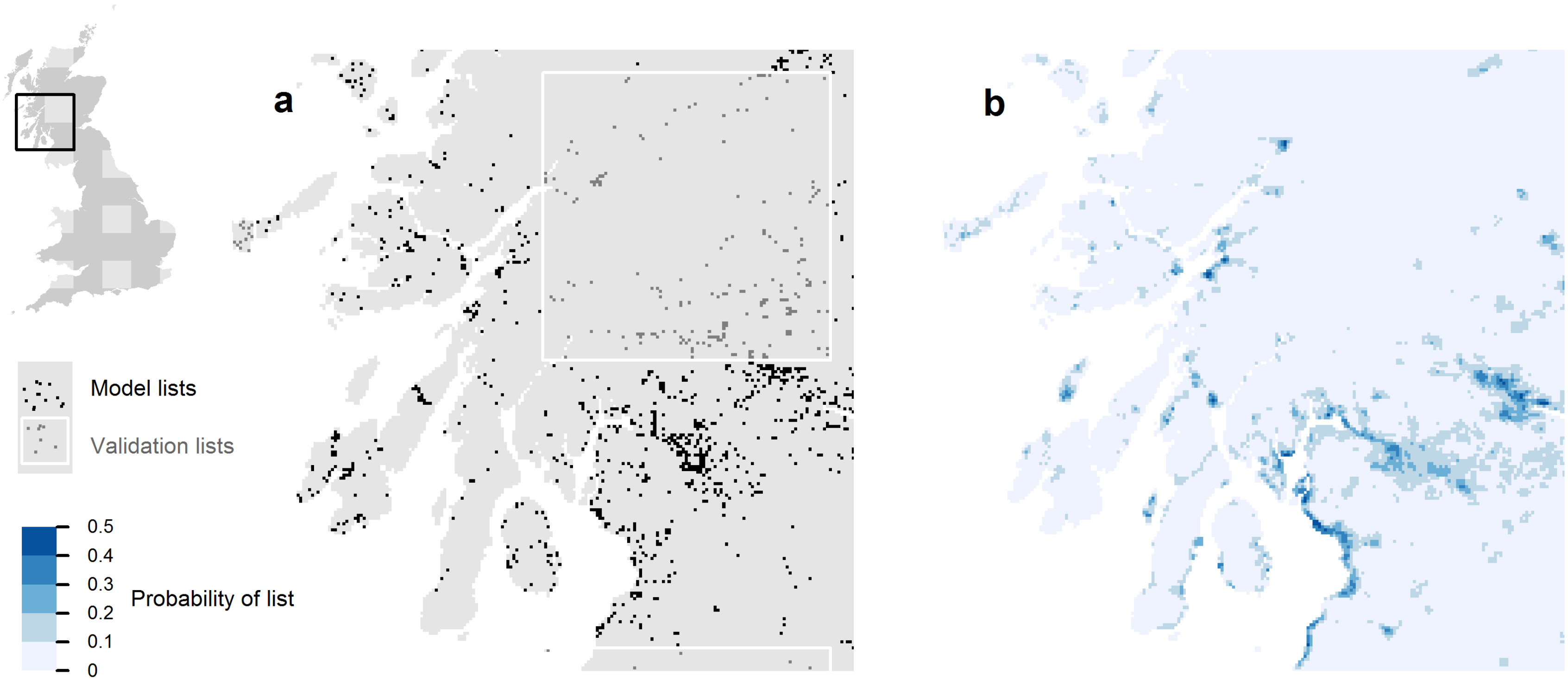
BirdTrack users don't count all species

- Common species are less likely to be counted by BirdTrack users
- Existing counts are difficult to relate to area sampled



BirdTrack sites are not randomly selected

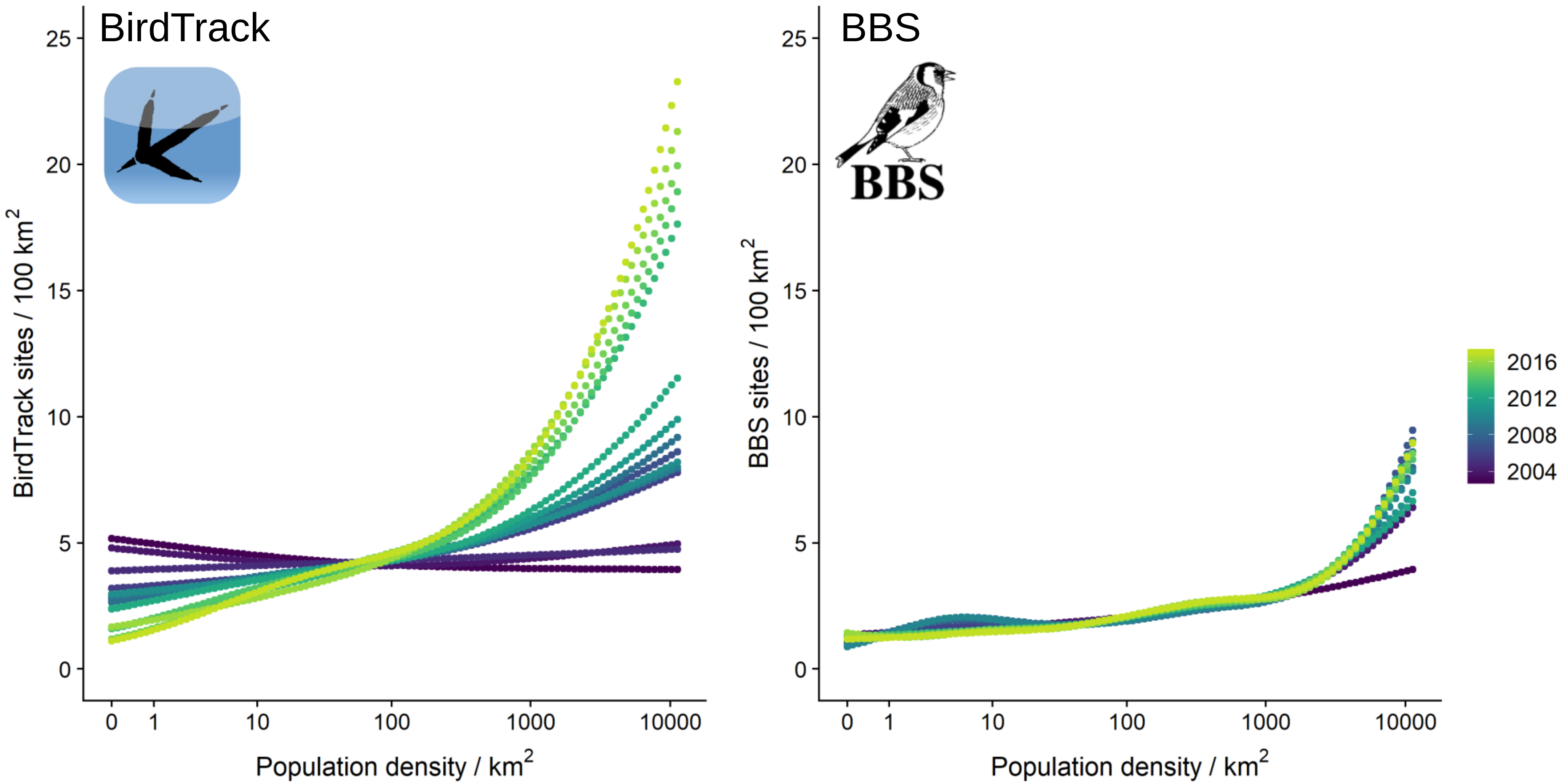
BirdTrack sites are biased towards urban areas, coasts, reserves



Johnston et al. in press Ecological Modelling

Site selection bias is non-stationary

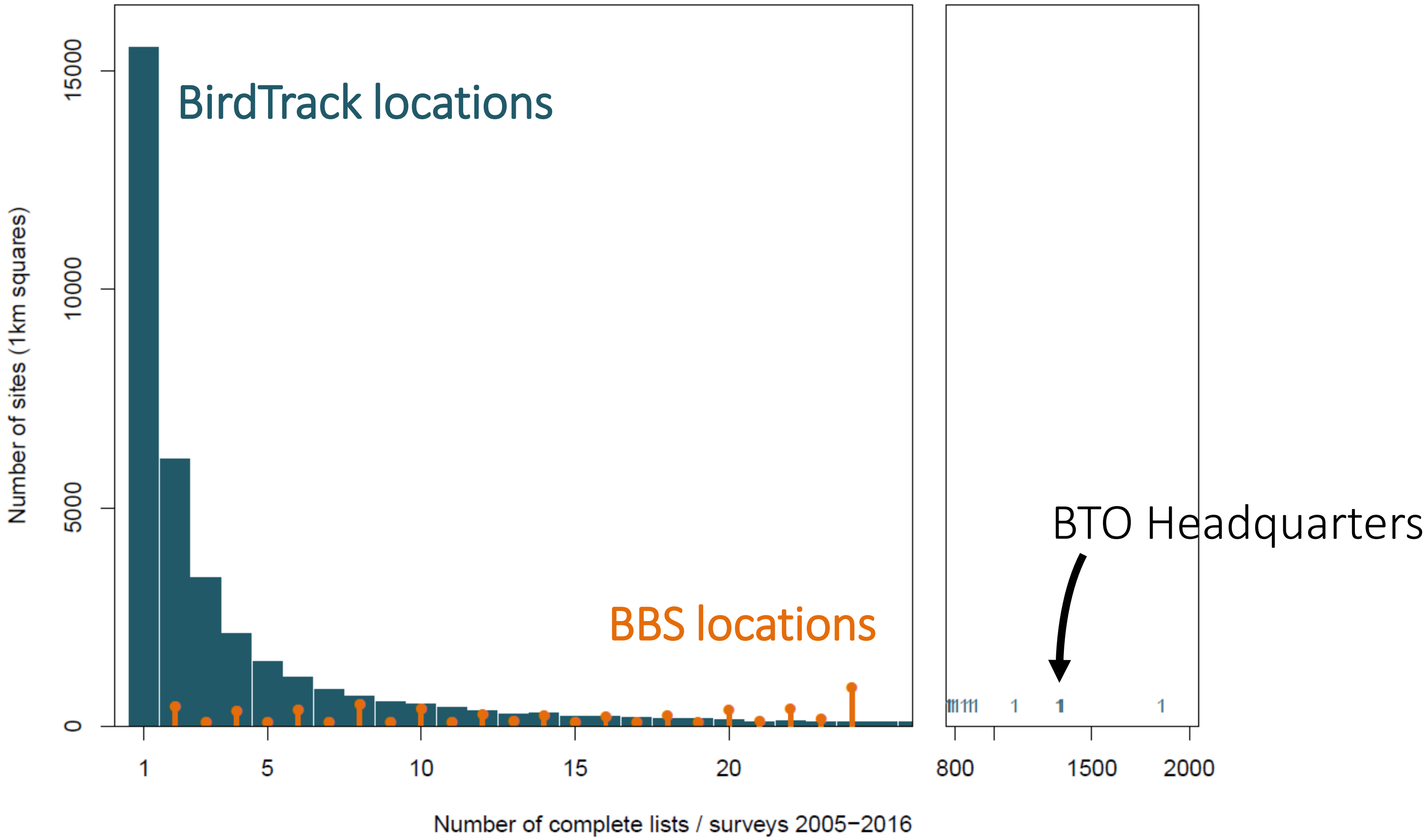
BirdTrack spatial bias is *increasingly* urban



Boersch-Supan et al. in prep.

Most BirdTrack sites are one-offs

Most BirdTrack sites are rarely revisited, a few **very** often

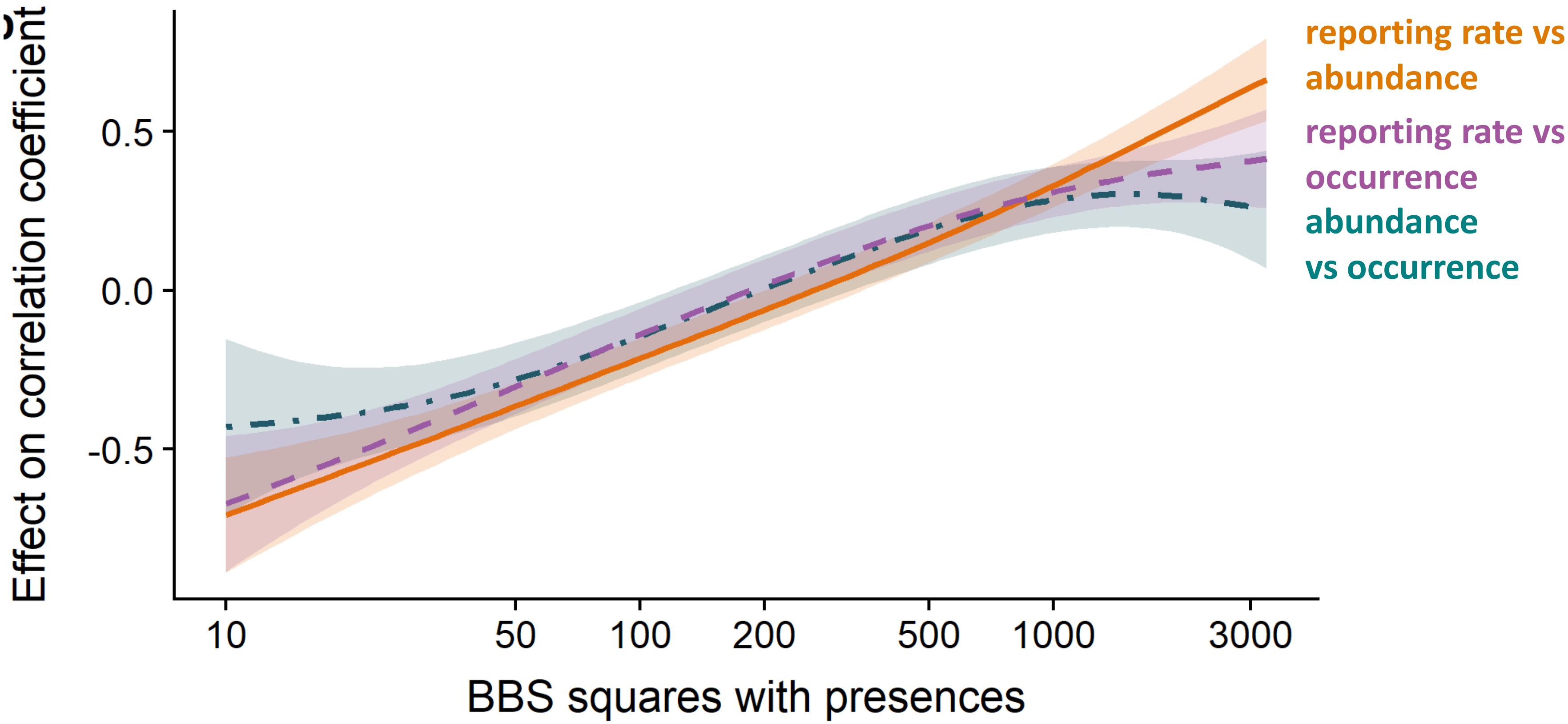


Can “big data” overcome BBS limitations?

Can we use unstructured bird listing data from recreational birdwatchers to generate trends for species which are poorly covered by the BBS?

We fitted trend models to BirdTrack detection/non-detection data, accounting for observer effort but not spatial bias.

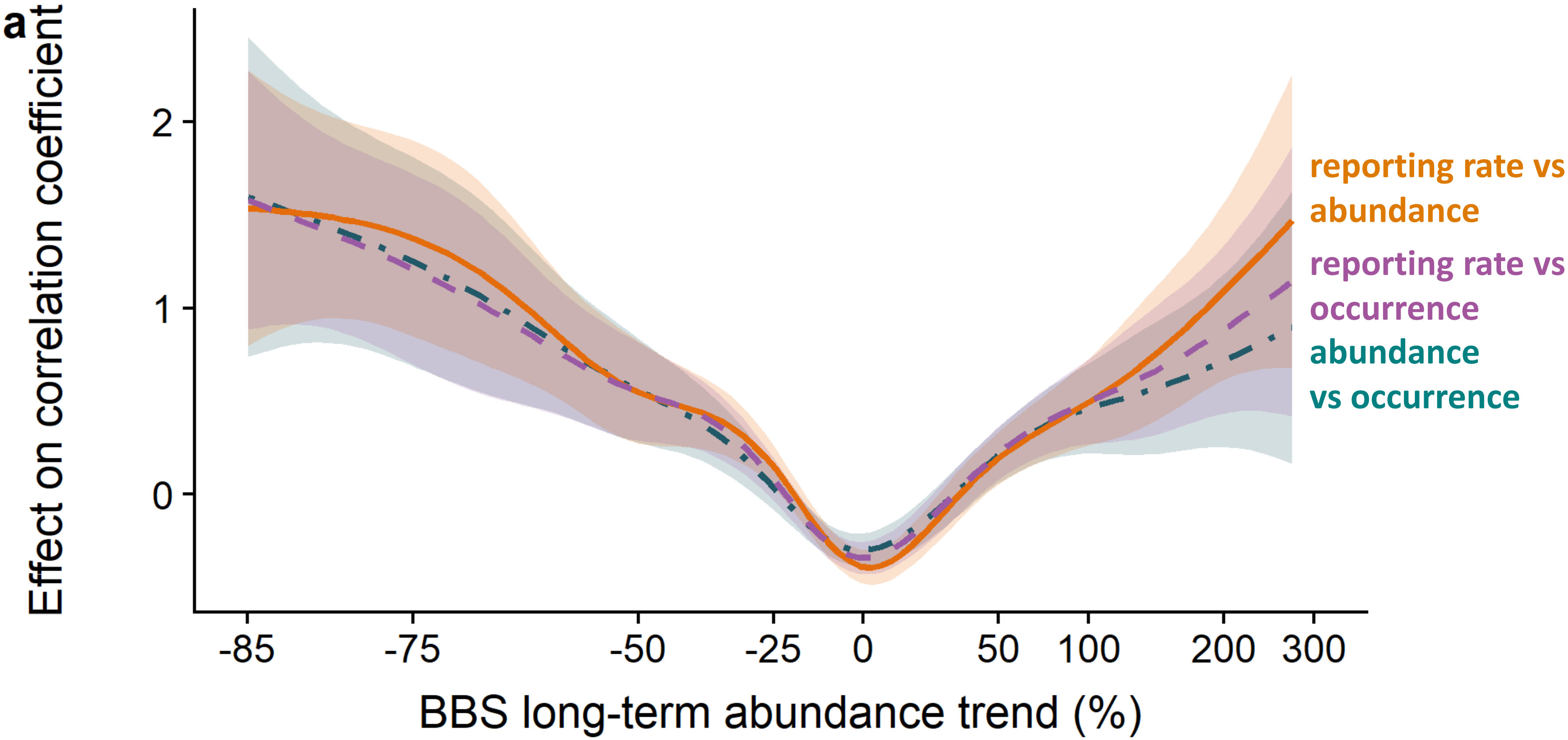
Comparing BBS and BirdTrack trends



Agreement between BBS and BirdTrack better for commoner species

Boersch-Supan et al. 2019 Biological Conservation

Comparing BBS and BirdTrack trends



Agreement among trends is strongest for strongly trending species

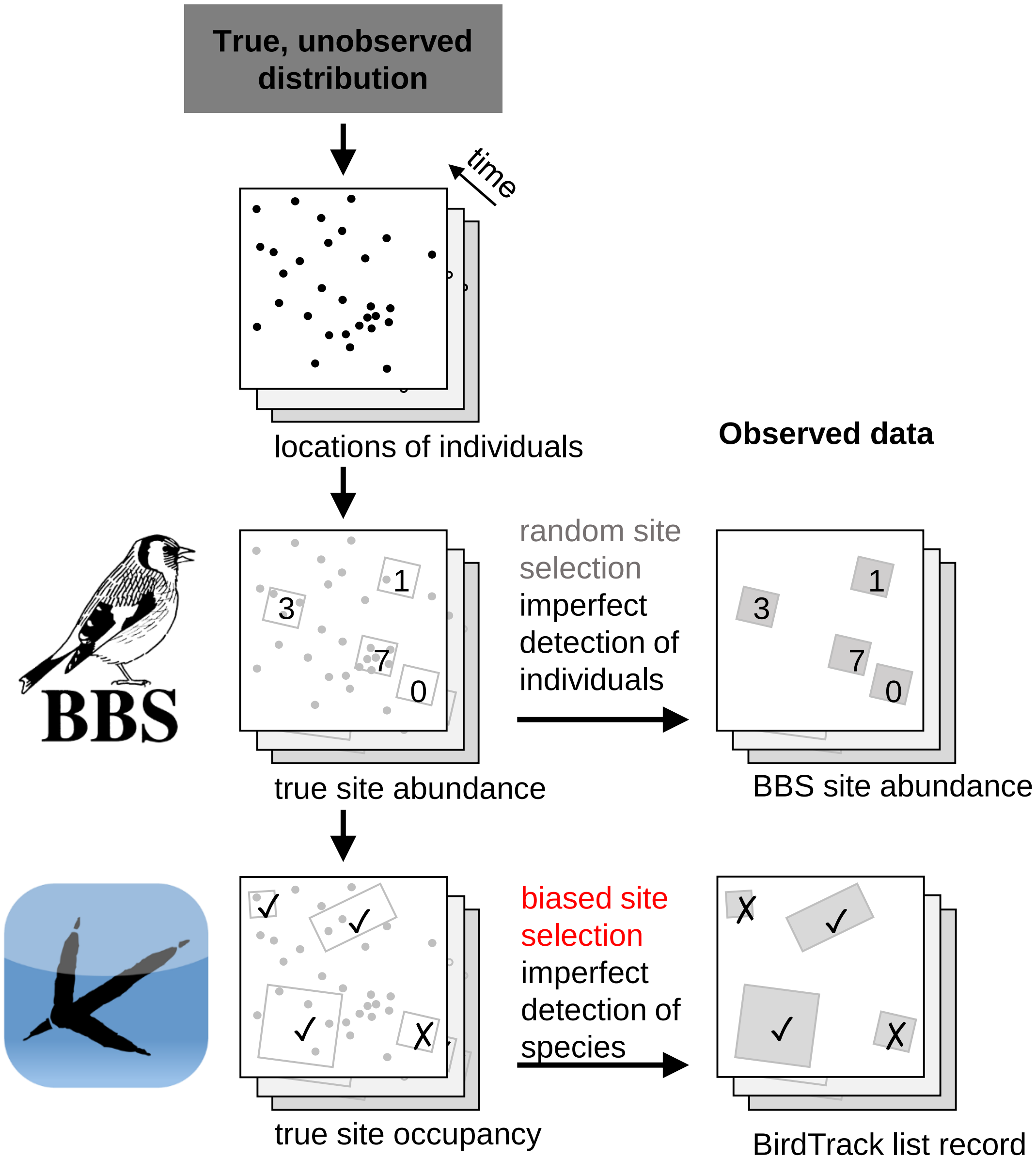
Boersch-Supan et al. 2019 Biological Conservation

Can “big data” overcome BBS limitations?

Big data from BirdTrack alone don’t offer a robust solution for poorly monitored species.

Can we do better by using **both** datasets?

Integrating structured and unstructured data sets



count data
unbiased sampling
standardized effort
but “small” sample size

presence/absence data
biased sampling
variable effort
large(r) sample size

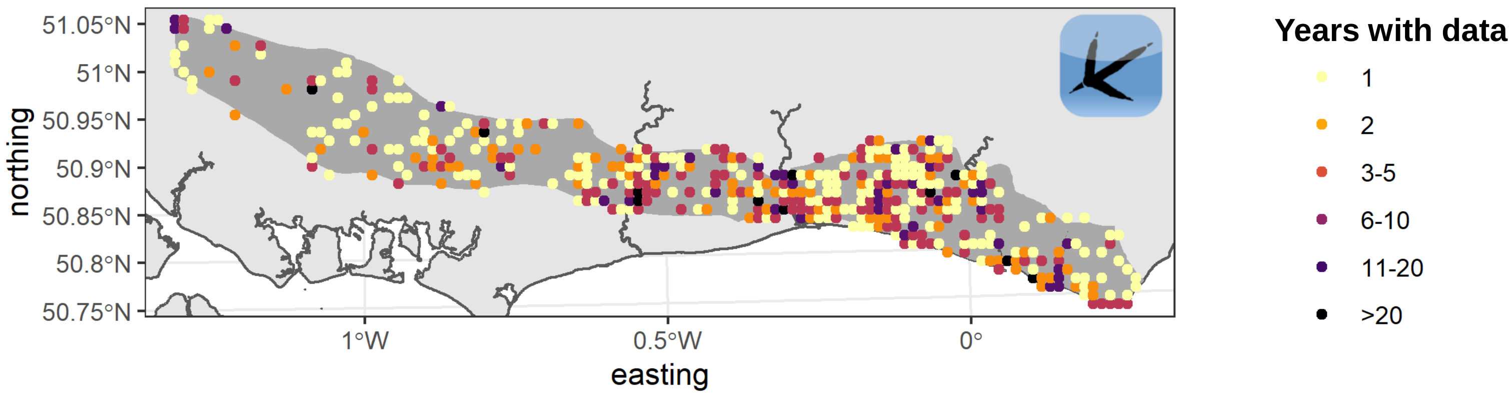
after Isaac et al. 2019 Trends Ecol Evol
https://doi.org/10.1016/j.tree.2019.08.006

Case study: Corn Bunting in the South Downs

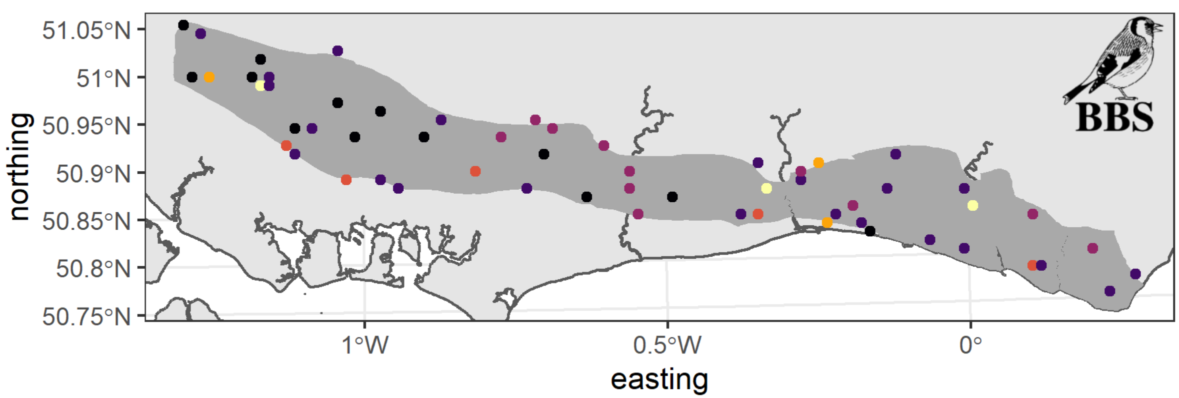
Can we use data integration to get trends for small areas?

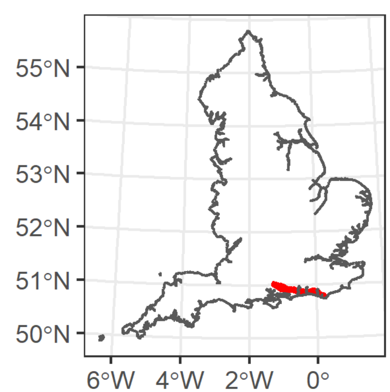
South Downs National Character Area

BirdTrack 2002-2018

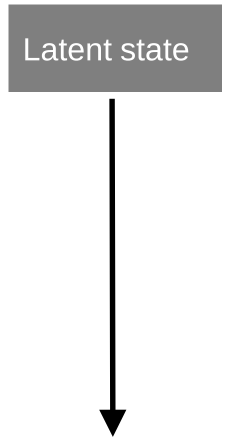


BBS 1994-2018





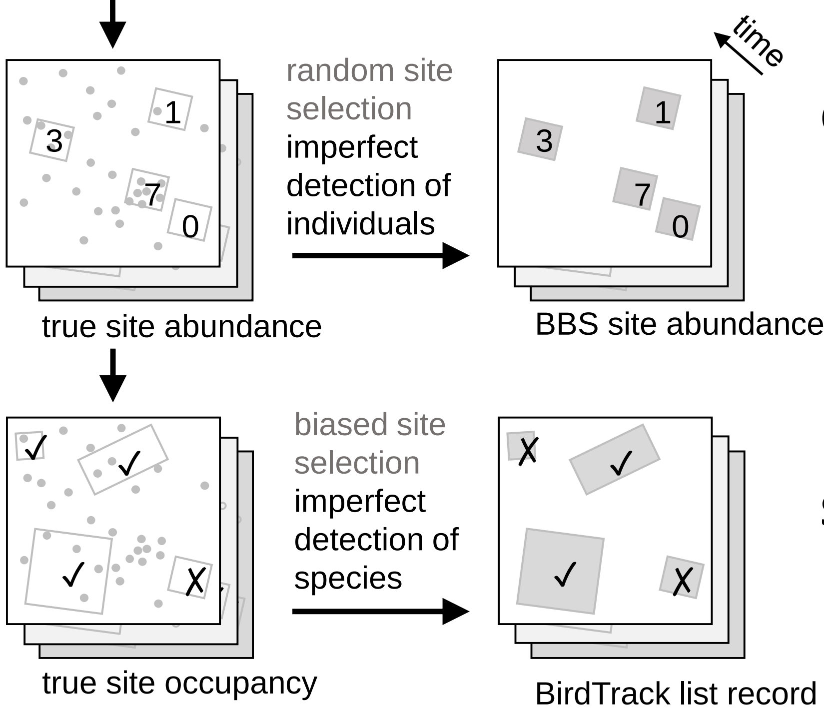
Integrated abundance model



State model: True abundance = Survival + Recruitment

$$N_{j,t} = S_{j,t} + G_{j,t}$$

$$\left.\begin{aligned} S_{j,t} &\sim \mathrm{Binomial}(N_{j,t-1}, \omega_{habitat,t}) \\ G_{j,t} &\sim \mathrm{Poisson}(\gamma_{habitat,t}) \end{aligned}\right\} \text{Annual rates modelled as habitat specific random effect}$$



Count observations: N-mixture model

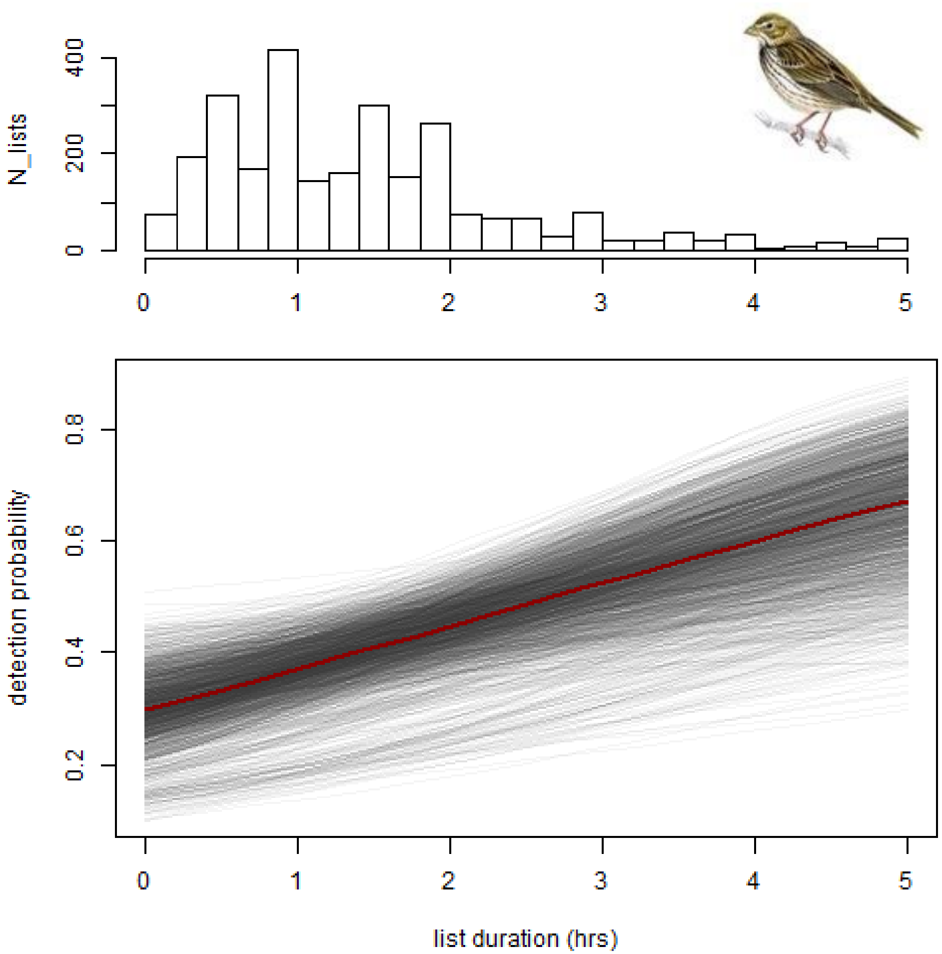
$$n_{j,t,k} \sim \mathrm{Binomial}(N_{j,t}, p)$$

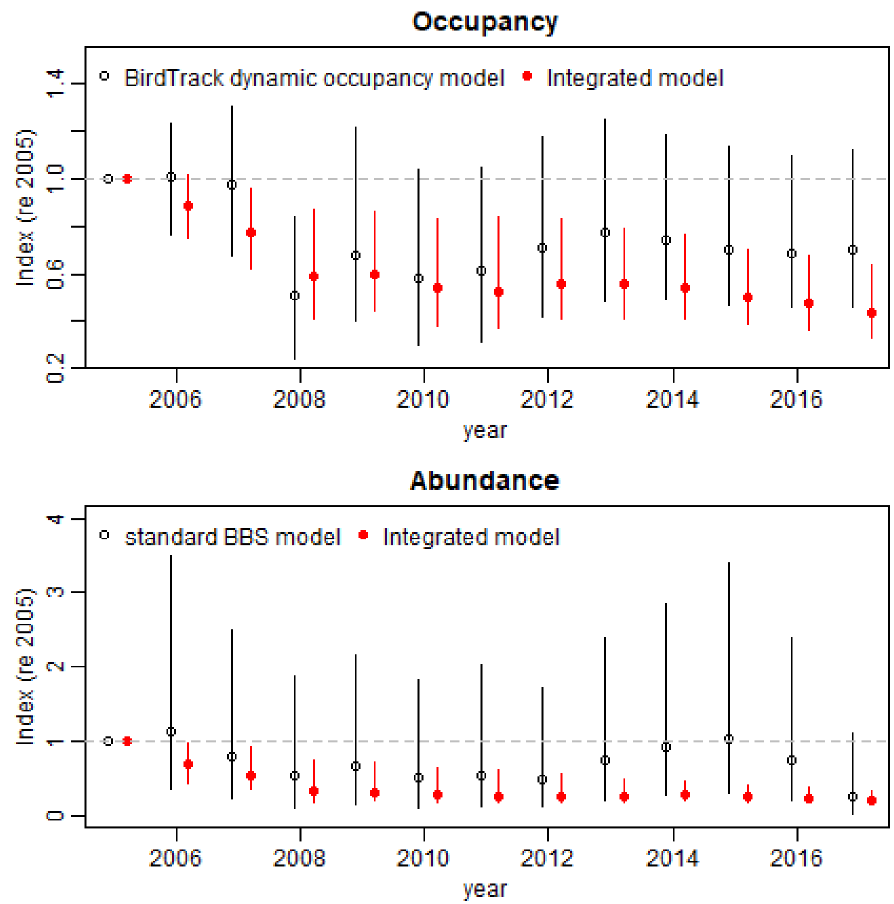
Species list observations:

$$y_{j,t,k} \sim \mathrm{Bernoulli}(1 - (1 - p_{occ})^{N_{j,t}})$$

Boersch-Supan et al. in prep., Zipkin et al. 2017 Ecology
https://doi.org/10.1002/ecy.1831

Case study: Corn Buntings in the South Downs





Integrated model accounts for heterogeneous observation effort in BirdTrack data

Precision of integrated trend estimates is better than using either dataset alone

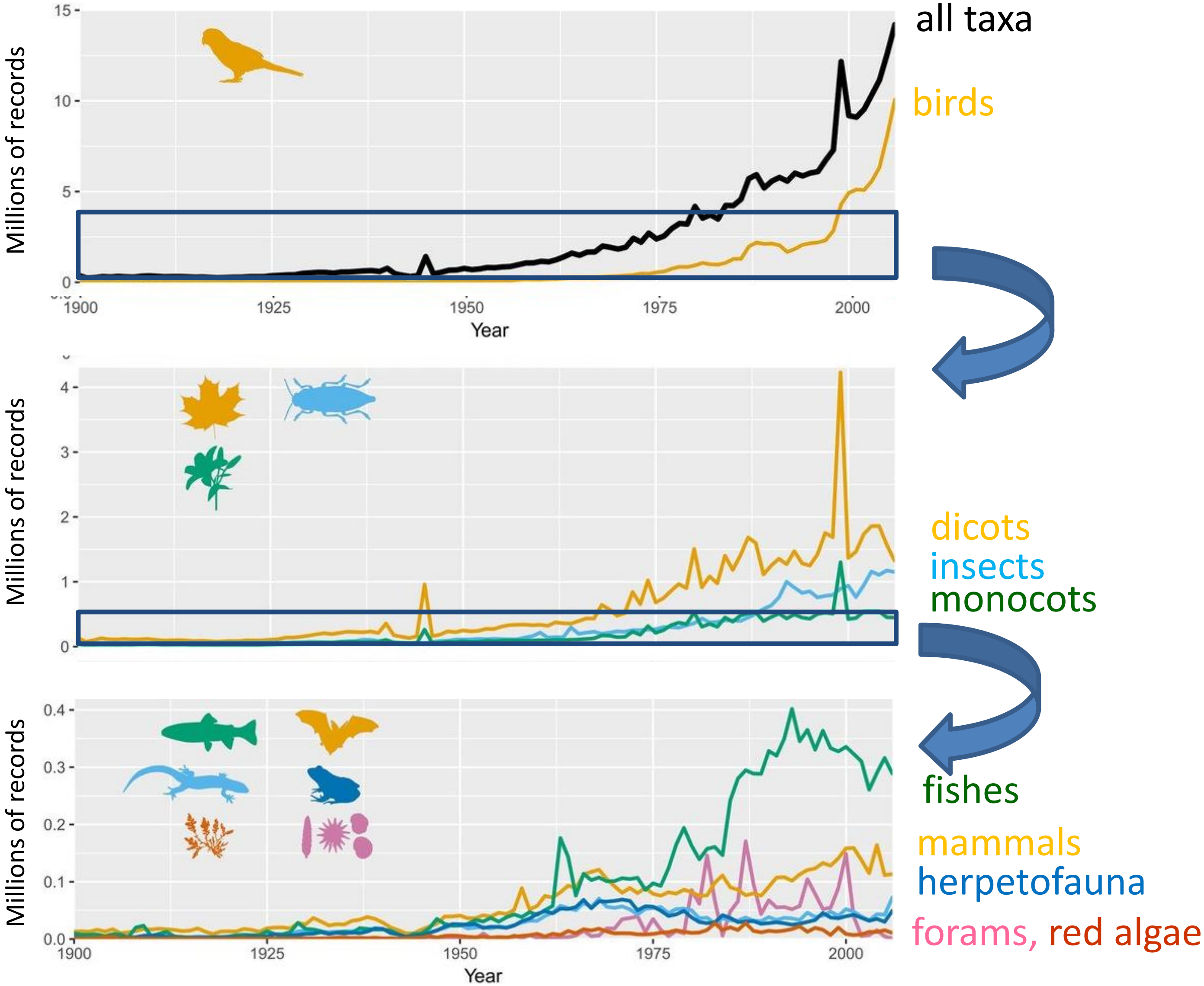
Integrated trend models

- Opportunities of integrated modelling
  - leverage strengths of both structured and unstructured data
  - great potential to improve precision of regional bird trends (~1000-10,000 km^2)

- Challenges of integrated modelling
  - no simple, one-size-fits-all approach: models require customization for each application
  - Some species and/or areas will be better suited than others

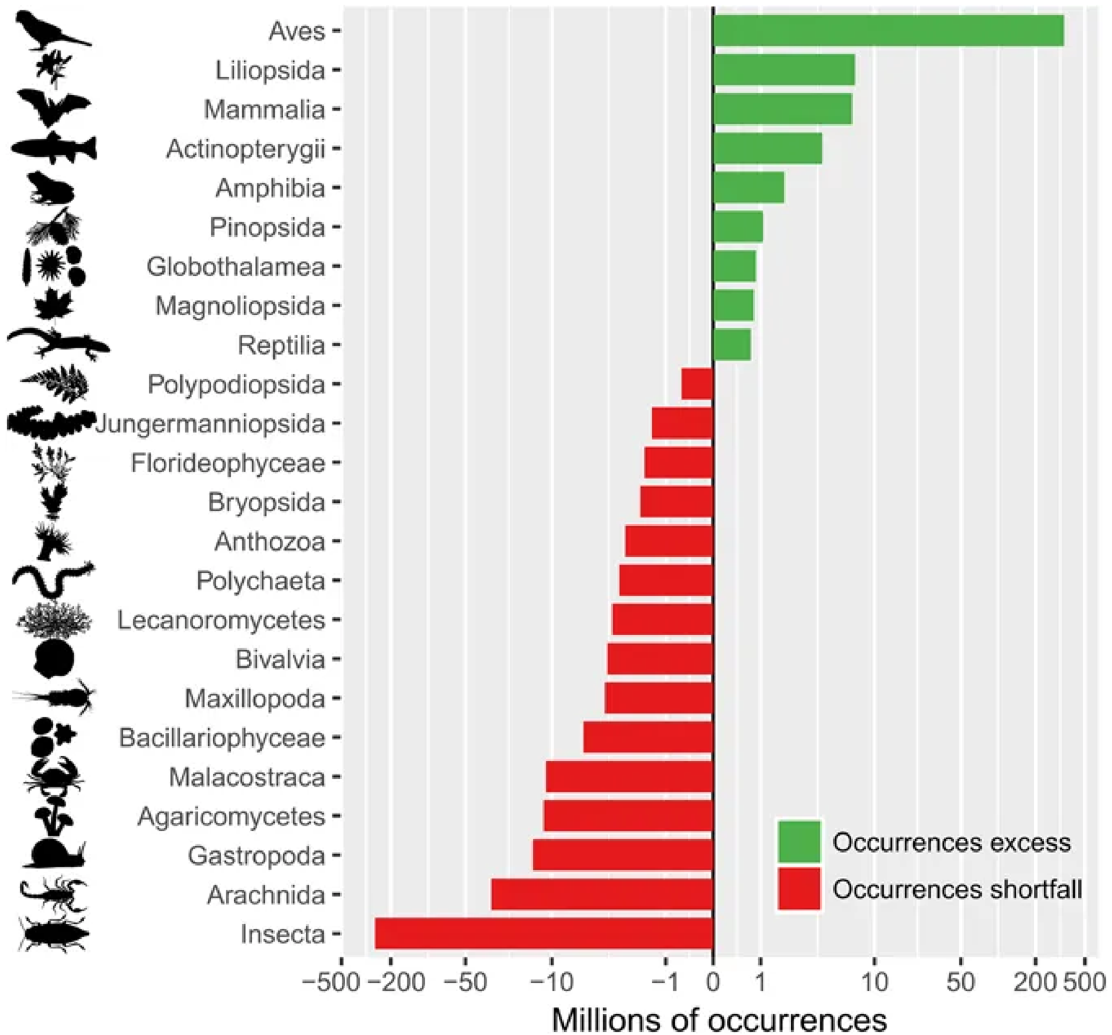
- Implications for scheme design
  - encourage recording of effort
  - encourage complete list recording
  - can we encourage recording in "boring areas"?

Gaps and sample sizes revisited



Troudet et al 2017 Scientific Reports https://doi.org/10.1038/s41598-017-09084-6

Gaps and sample sizes revisited



Troudet et al 2017 Scientific Reports https://doi.org/10.1038/s41598-017-09084-6

General Conclusions

- Thanks to an enormous volunteer effort bird populations in the UK are among the best monitored globally
- National scale wildlife population assessment in practice involves balancing optimal survey designs and cutting edge statistics with limited resources (volunteers, time, software, hardware)
  - coverage vs. costs / quality vs. quantity trade-offs
- Rich datasets can be a technical challenge, but they provide exciting opportunities for methodological developments
- Most taxa and most places remain poorly monitored.
  - Opportunistic records and data integration can alleviate survey gaps, but “big data” are not a silver bullet solution.

Tipling / BTO

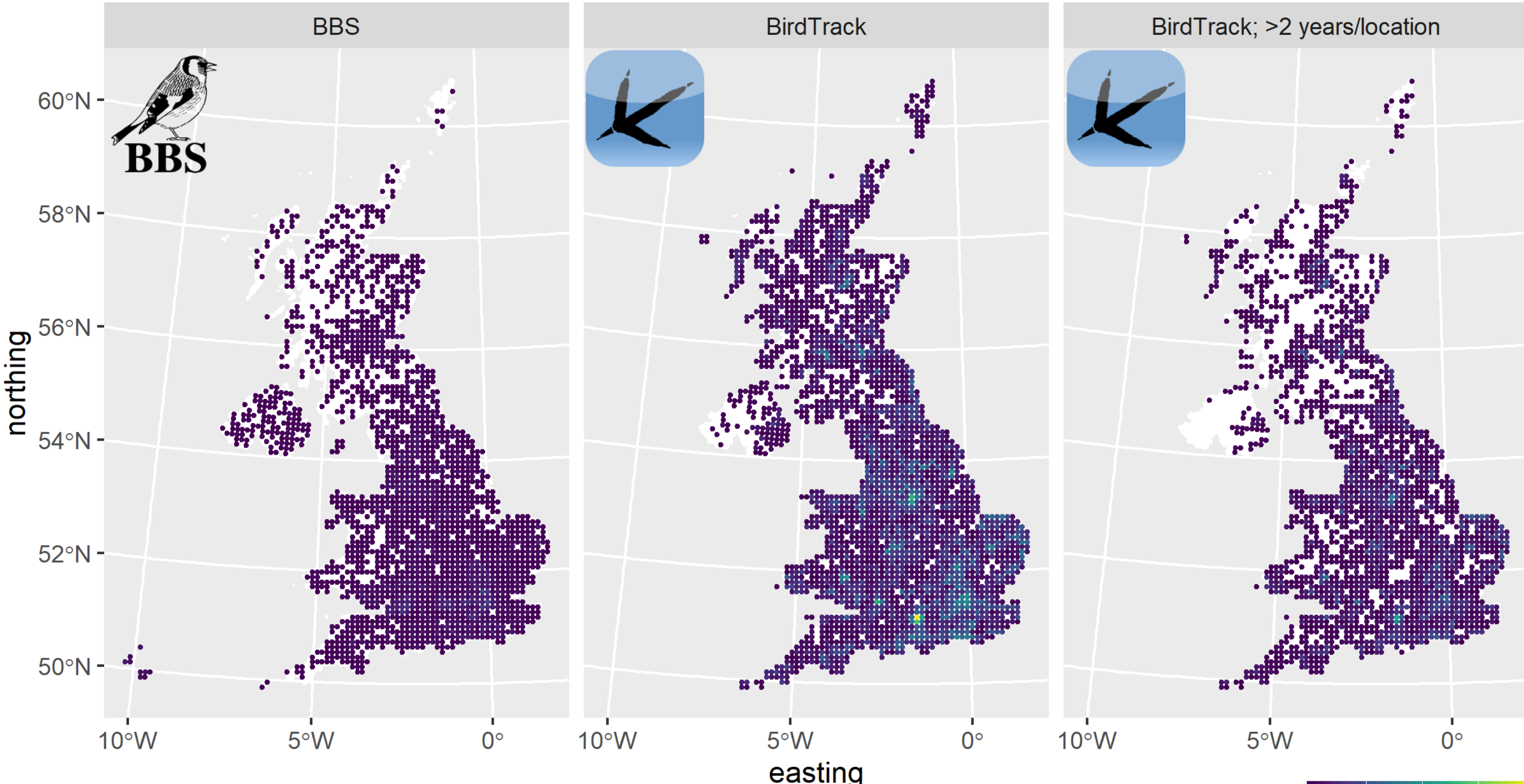
ACKNOWLEDGEMENTS

Bird monitoring in the UK is only possible because of the dedication of **thousands of volunteers** and the financial support from the **survey partners JNCC and RSPB** and from **charitable contributions to the BTO**.

Further reading

- Ecological Modelling Approaches:
  - Mouquet et al. 2015 J Appl Ecol https://doi.org/10.1111/1365-2664.12482
- Citizen science biodiversity data:
  - Amano et al. 2016 BioScience https://doi.org/10.1093/biosci/biw022
  - Bayraktarov et al 2019 Frontiers Ecol Evol https://doi.org/10.3389/fevo.2018.00239
  - Kelling et al. 2018 BioScience https://doi.org/10.1093/biosci/biz010
- Integrated models:
  - Robinson et al. 2014 MEE https://doi.org/10.1111/2041-210X.12204
  - Isaac et al. 2019 TREE https://doi.org/10.1016/j.tree.2019.08.006
- UK bird trends and other biodiversity assessments:
  - BirdTrends: bto.org/birdtrends
  - State of Nature: rspb.org.uk/stateofnature

Gaps and sample sizes



Gaps in both schemes: Uplands (Scotland, central Wales)

Gaps in BirdTrack: Rural Northern Ireland, areas w/ high-intensity agriculture

Boersch-Supan et al. 2019 Biological Conservation