

Integrating citizen science data to estimate bird population trends

Philipp Boersch-Supan @pboesu

British Trust for Ornithology @_BTO

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)**

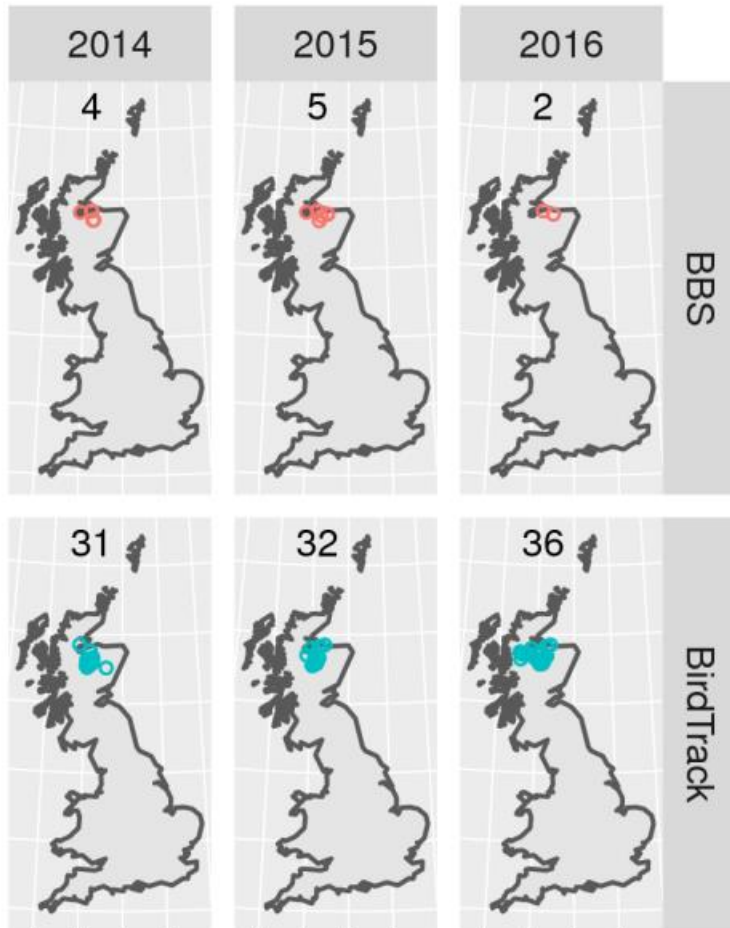
(but not big enough for many species of conservation concern)



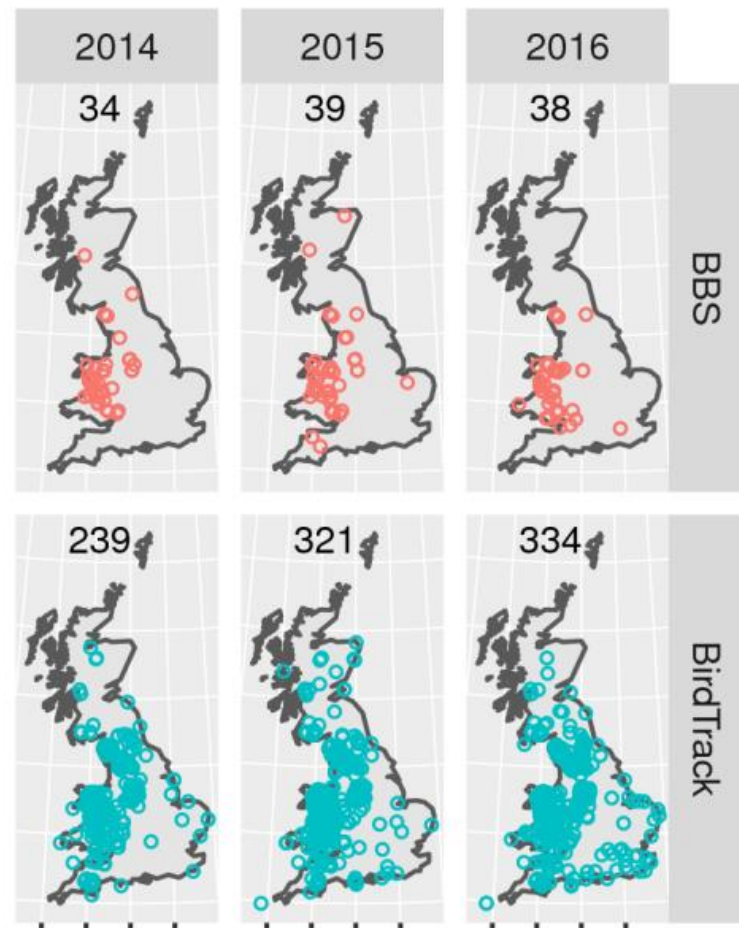
- **BirdTrack (and similar schemes)**
 - complete listing optional
 - counting optional
 - effort recording optional
 - sites self-selected**→ Bigger data (currently ~15k sites, ~100k lists per year)**

👍 BirdTrack has c. 10x more records across space

Crested Tit



Pied Flycatcher

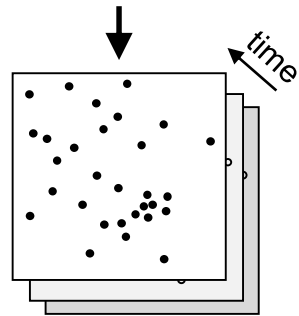


True species
distribution

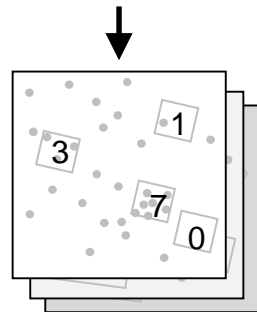
Observation
error

Observed
data

Latent state

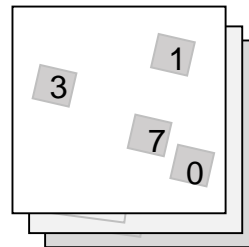


locations of individuals



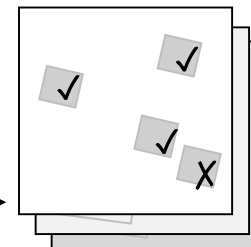
true site abundance

random site
selection
imperfect
detection of
individuals



BBS site abundance

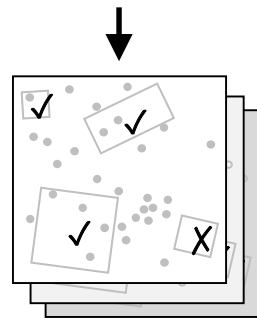
truncate
counts



BBS site occurrence

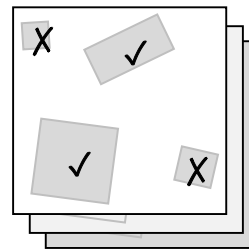


unbiased counts
but "small" sample size



true site occupancy

biased site
selection
imperfect
detection of
species



BirdTrack list record

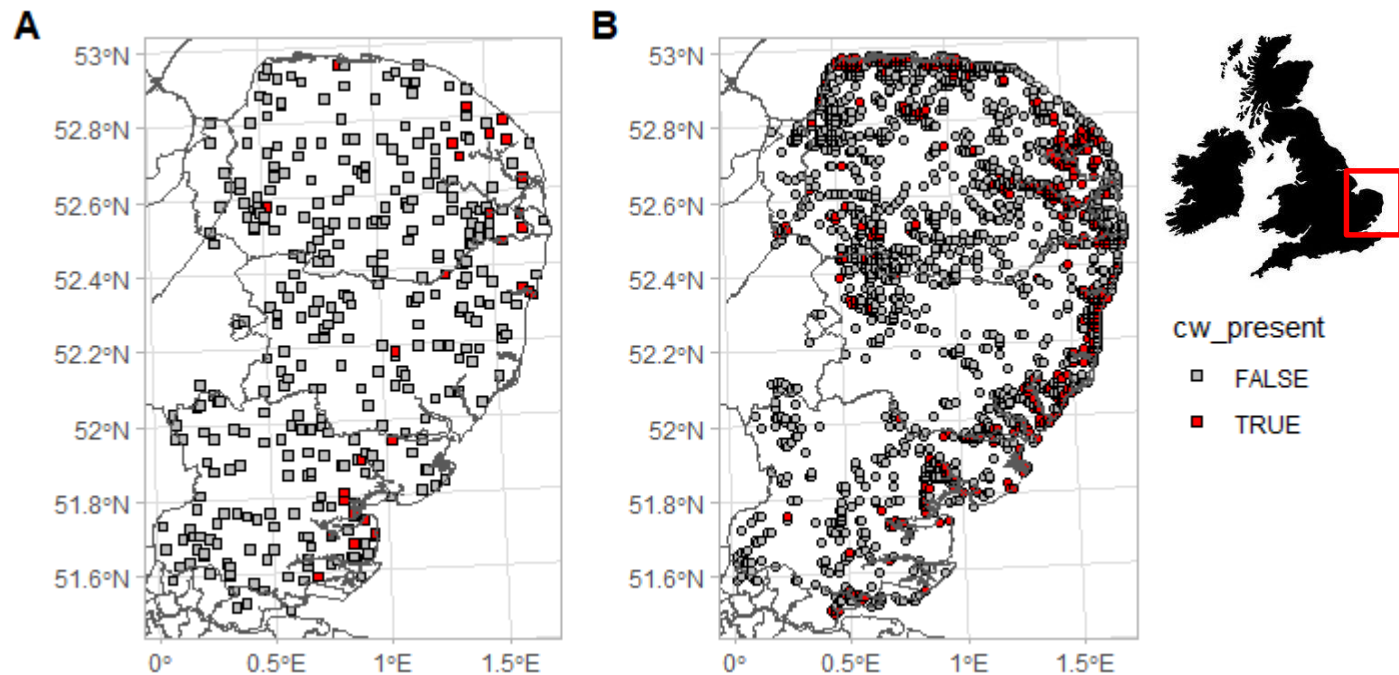
biased (non)detections
large sample size



Integrated state-space model for *Cettia cetti*



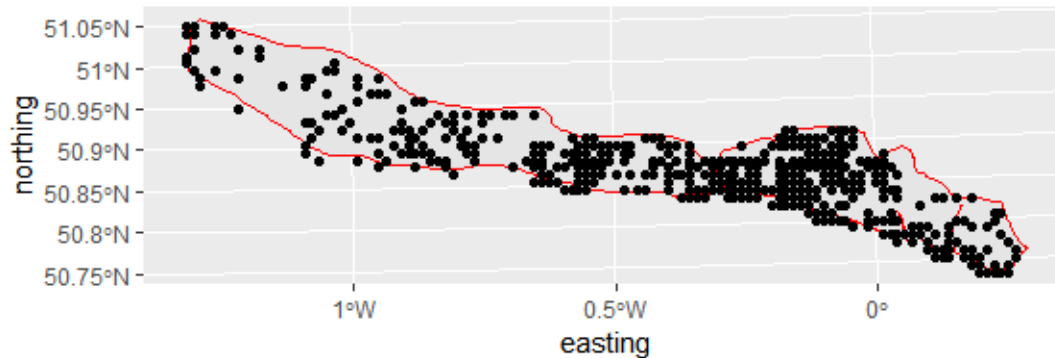
Neil Calbrade/BTO



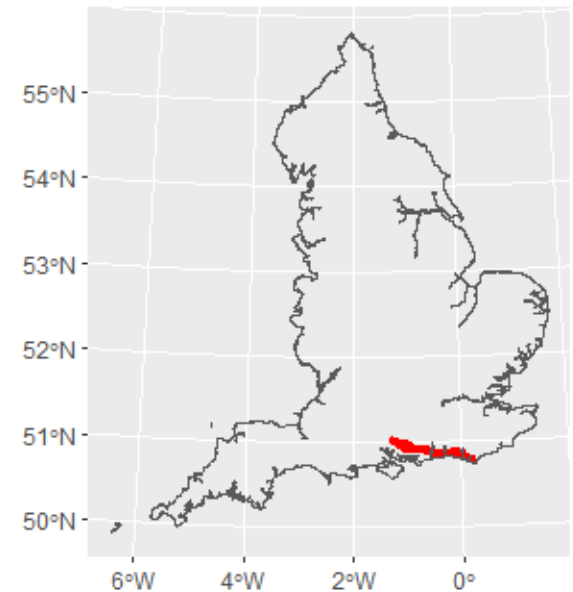
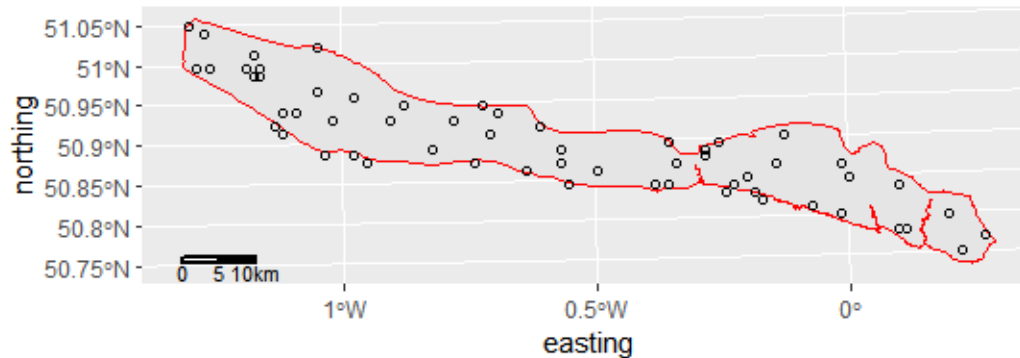
South Downs Corn Bunting *Emberiza calandra*



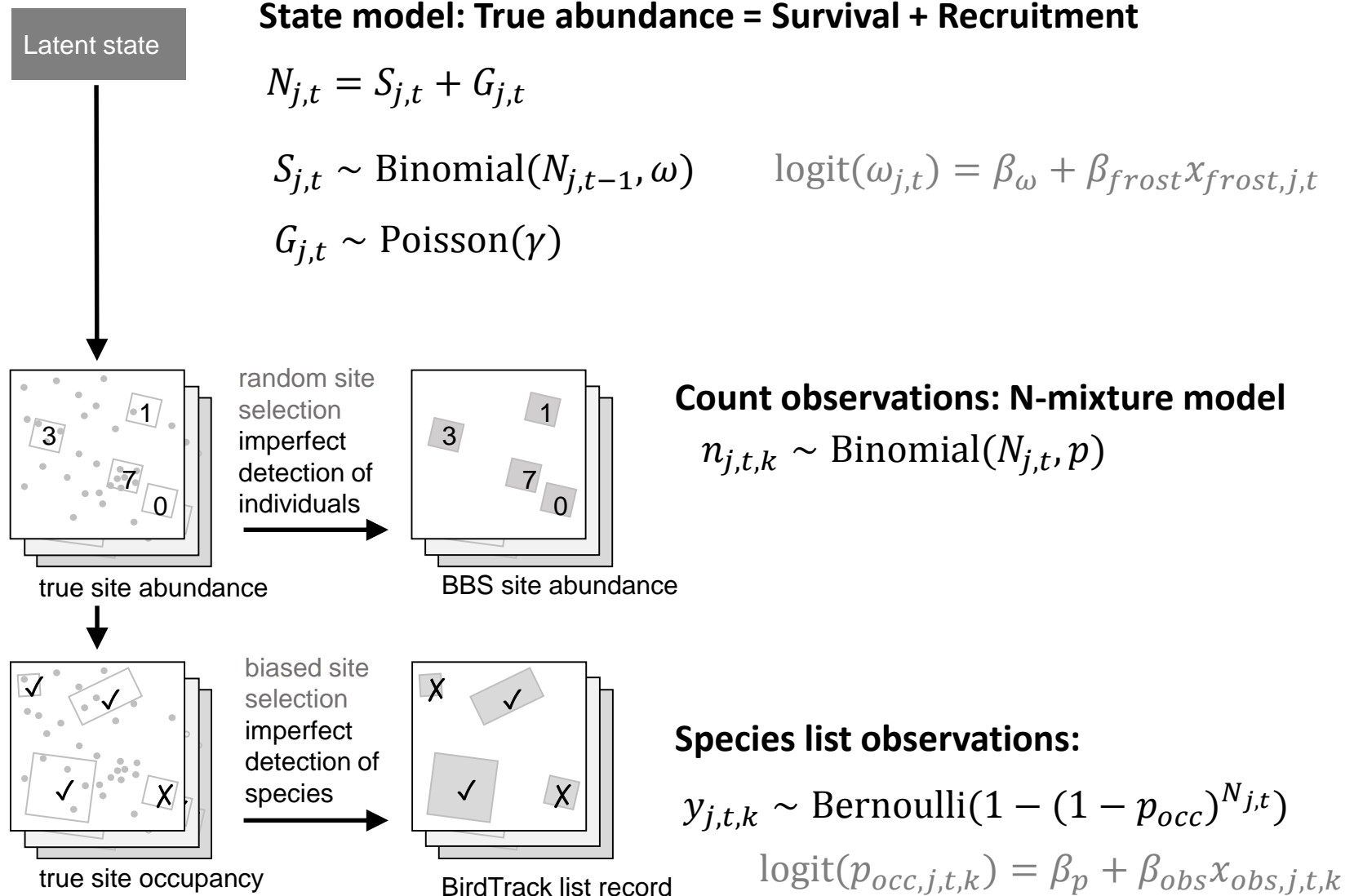
BirdTrack



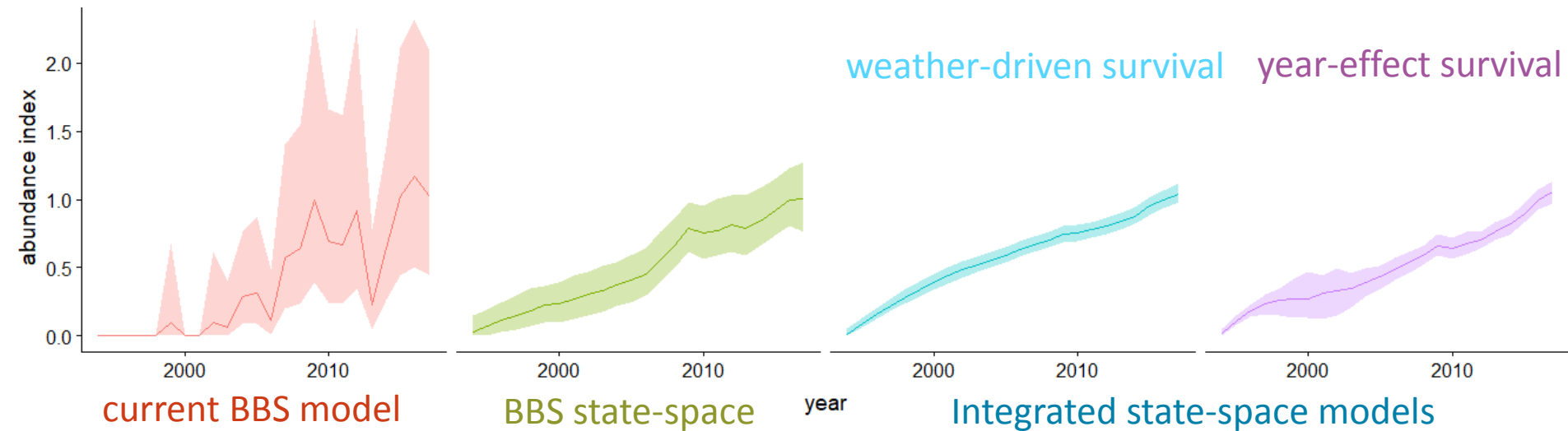
BBS



Integrated model following Zipkin et al. 2018; Parameter estimation via MCMC



Initial results for *Cettia cetti*

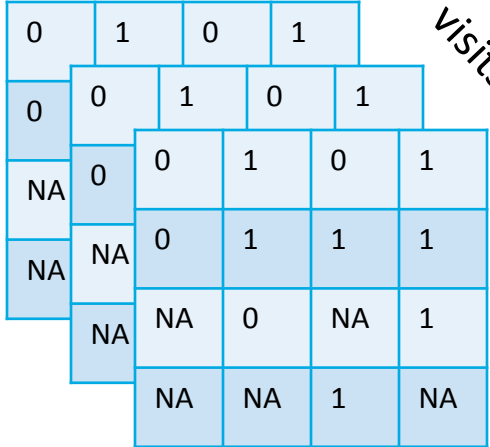


- big improvement in precision for **abundance trend**
- good agreement with data from Rare Breeding Birds Panel
- computational effort is high
 - too high for routine UK-scale applications
 - but country-level joint trends are in reach

Implementation details

- Model is computationally costly because there's a latent state variable for every site
- Standard observation model implementation make matters worse given the sparsity of the data

```
# i-j-t array-form Observation model (Kery & Schaub book, Zipkin paper, ...)
for (i in 1:nsite){
  for (j in 1:nrep){
    for (t in 1:nyear){
      muy[i,j,t] <- z[i,t]*p[t]
      y[i,j,t] ~ dbern(muy[i,j,t])
    } #t
  } #j
} #i
```

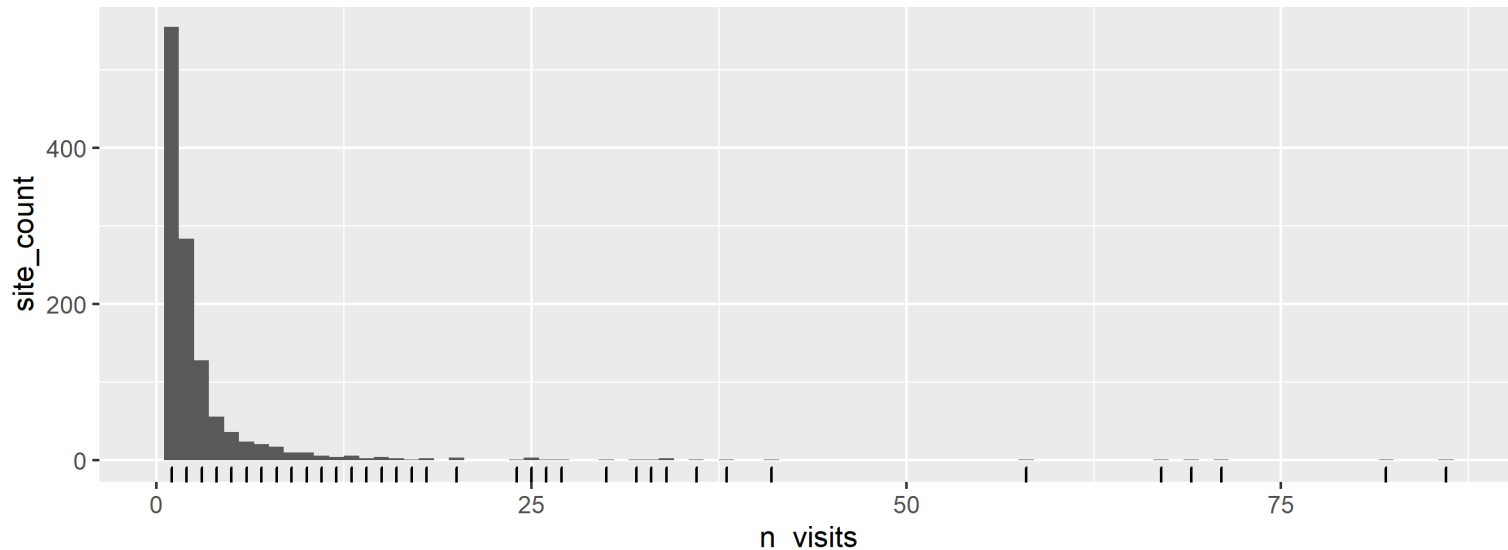


visits	years	1	2	3	4	5
1	0	1	0	1		
2	0	0	1	0	1	
3	NA	0	0	1	0	1
4	NA	NA	0	1	1	1
5	NA	NA	NA	0	NA	1
6		NA	NA	1	NA	

years

South Down case study BirdTrack data

- Process model: 432 sites * 17 years = 7344 parameters
- Data: 3809 lists after initial quality control



- $\max(n_visits) = 86$
- Using all the data yields observation array of size 631,584
 - 99.4% sparse, i.e. 627,775 additional parameters
- Additional filtering ameliorates situation slightly but dense arrays and default NA imputation not a sensible route

Implementation details

- More efficient to use dense table of observations indexed by a site-year-visit triplet

```
# "long-form" Observation model
# year-site expectations
for (i in 1:nsite){
  for (t in 1:nyear){
    p.site[i,t] <- 1-pow( (1-p.occ),N[i,t] )
  } #t
}#i

#likelihood - loop over i-j-t triplets
for (k in 1:nOccObs) {
  OccObs[k] ~ dbern(p.site[OccSite[k],OccYear[k]])
}#k
```

	P/A	year	site	Visit
1	0	1	1	1
2	1	1	1	2
3	0	1	2	1
4	1	2	1	1
...
k	0	21	5	1

- South Downs Case Study:
- $3809 \times 4 = 15236$ array elements; c. 2.5% of i-j-k array

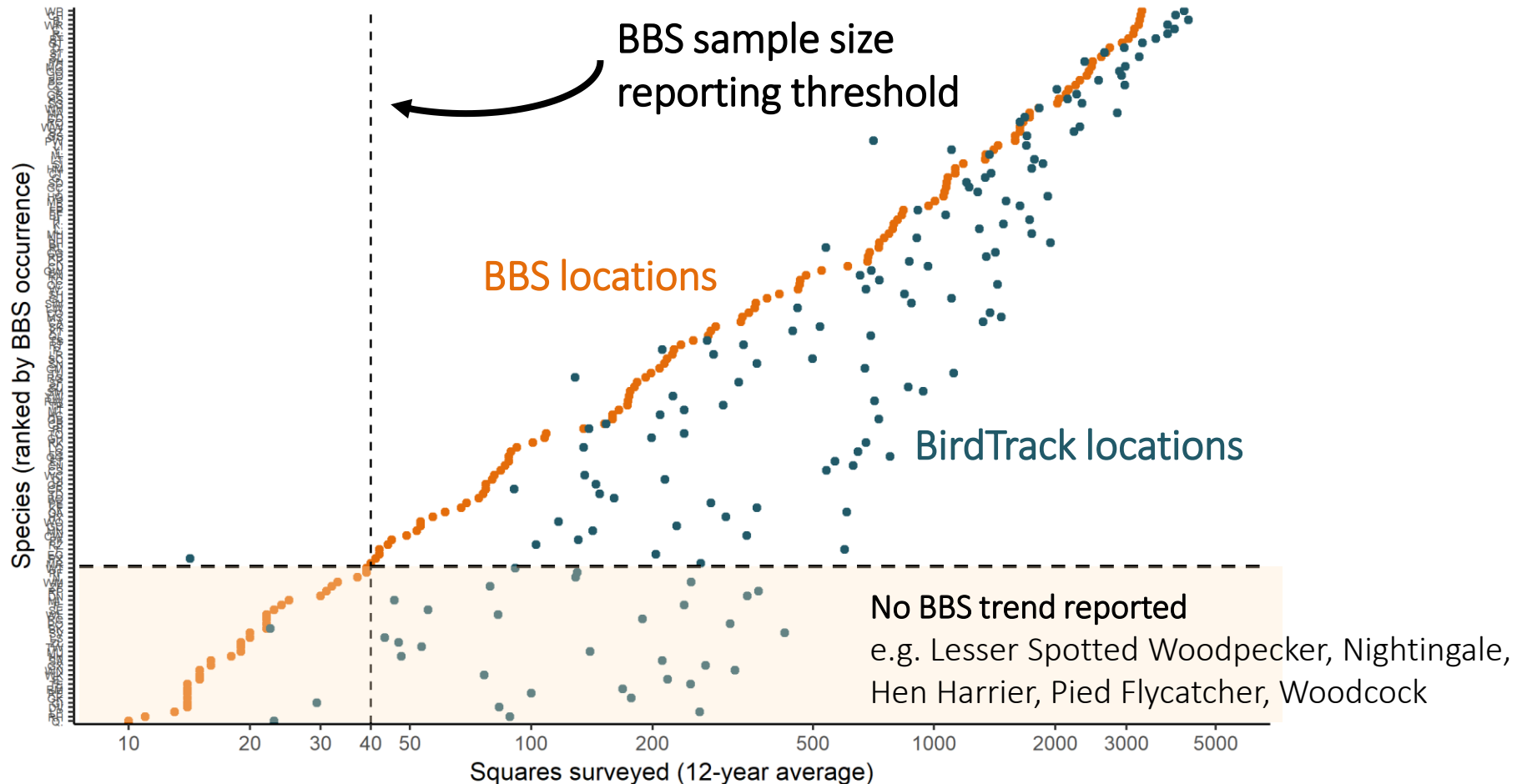


ACKNOWLEDGEMENTS

This analysis was only possible with the dedication of the thousands of **BBS volunteers** and **BirdTrack users**, the **contributors to the BirdTrack Appeal**, and the **survey partners**. Computations were conducted on **NERC's JASMIN** data analysis platform.

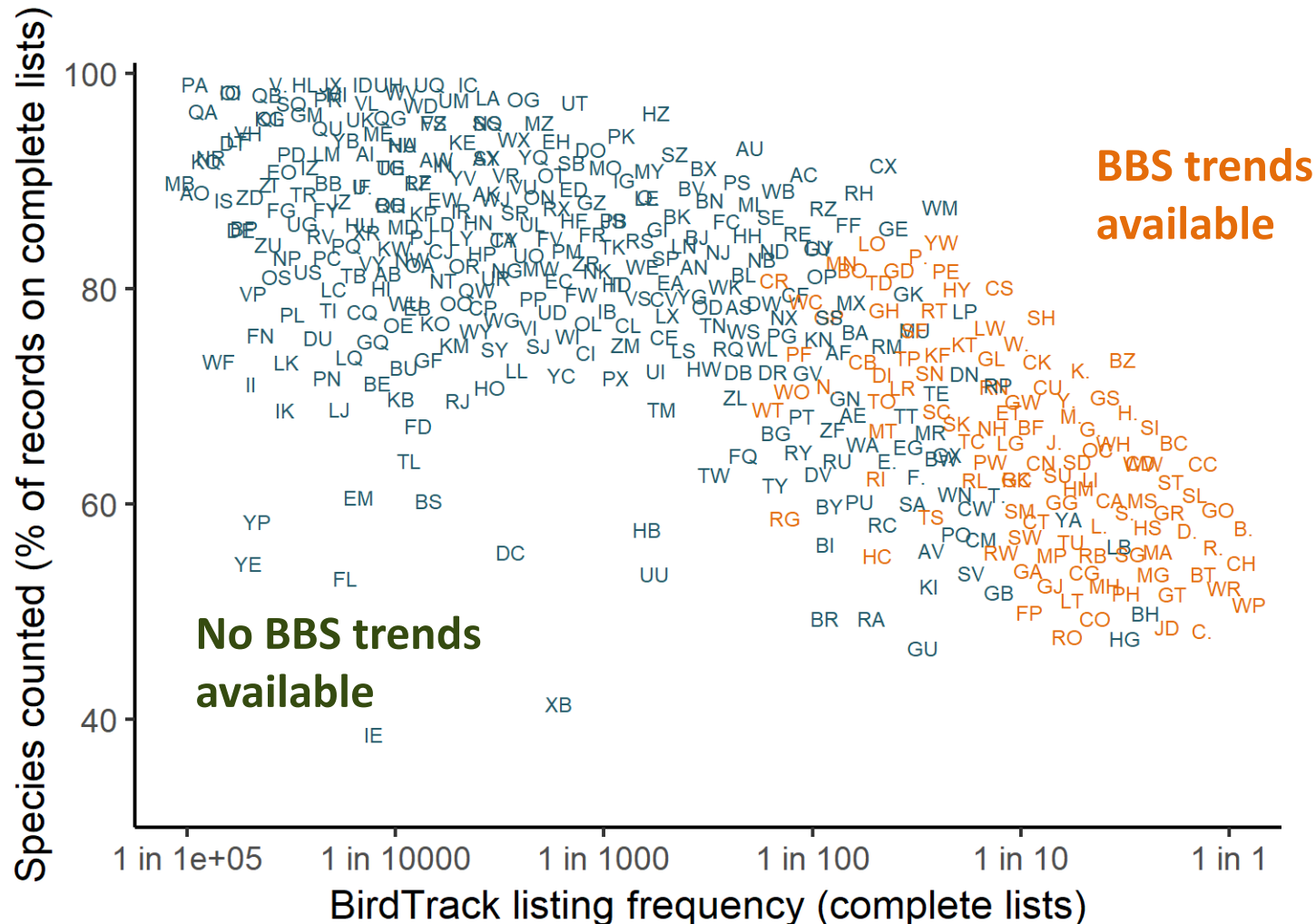
👍 BirdTrack has c. 10x more records across space

BirdTrack captures species in many more places than the BBS



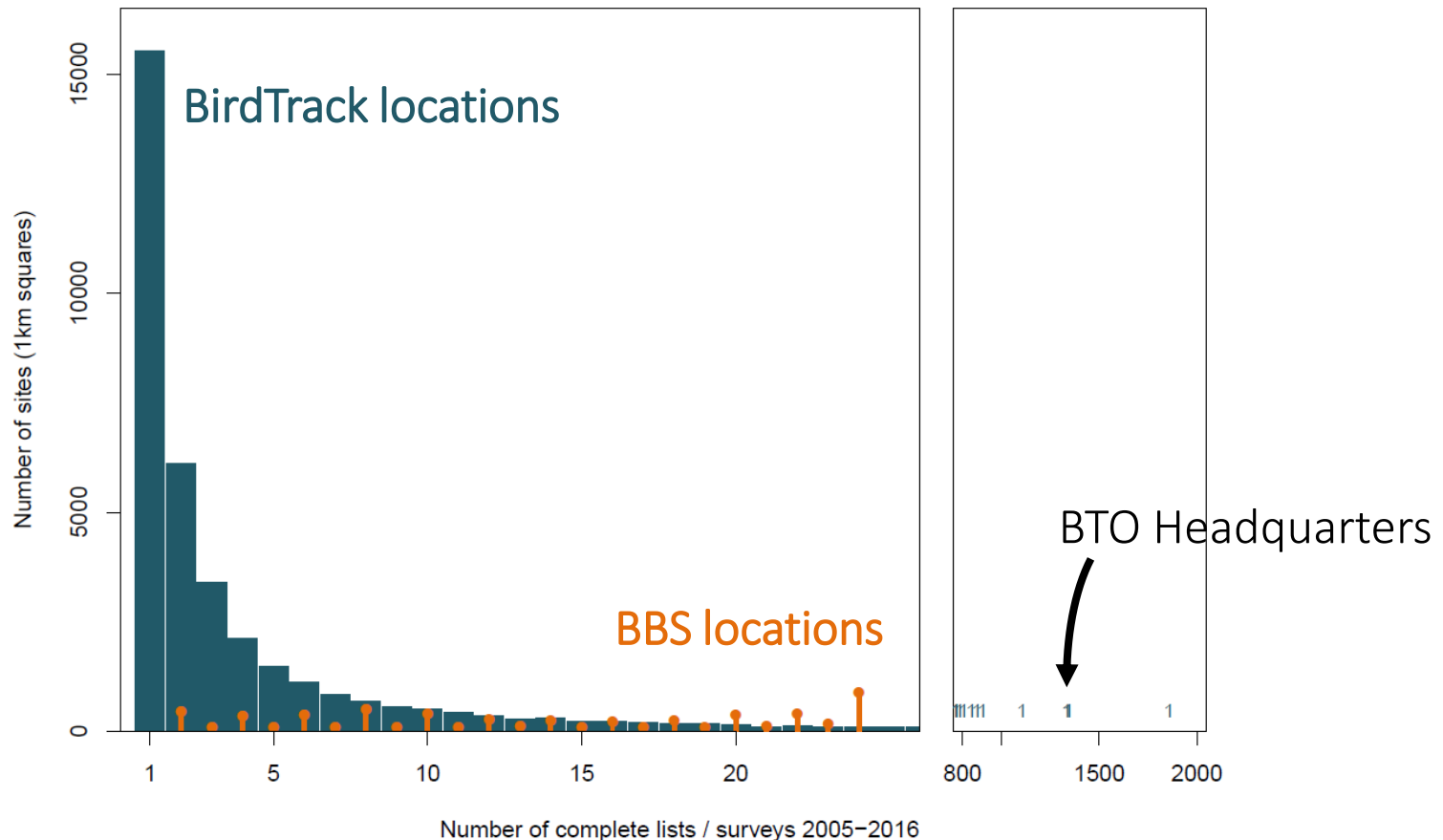
BirdTrack users don't count all species

Common species are less likely to be counted by BirdTrack users



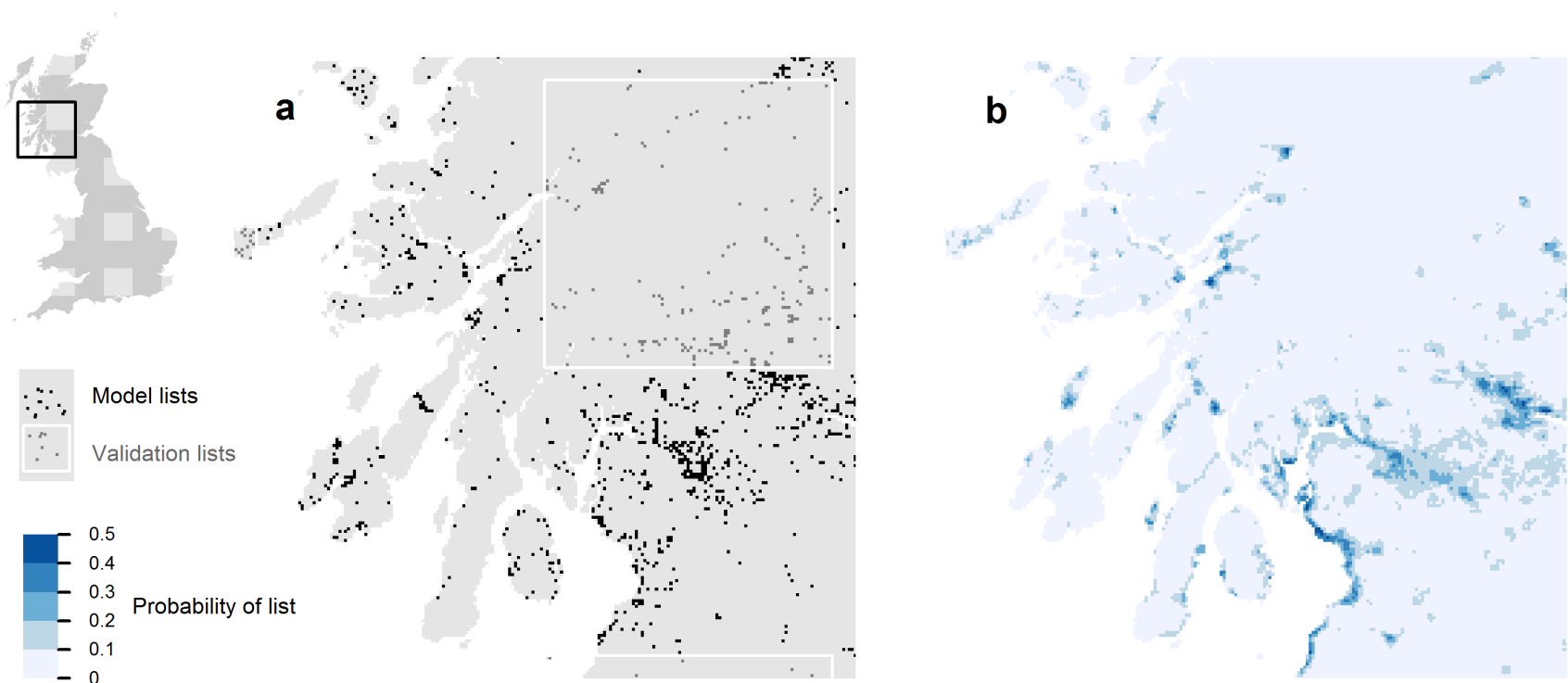
Most BirdTrack sites are one-offs

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



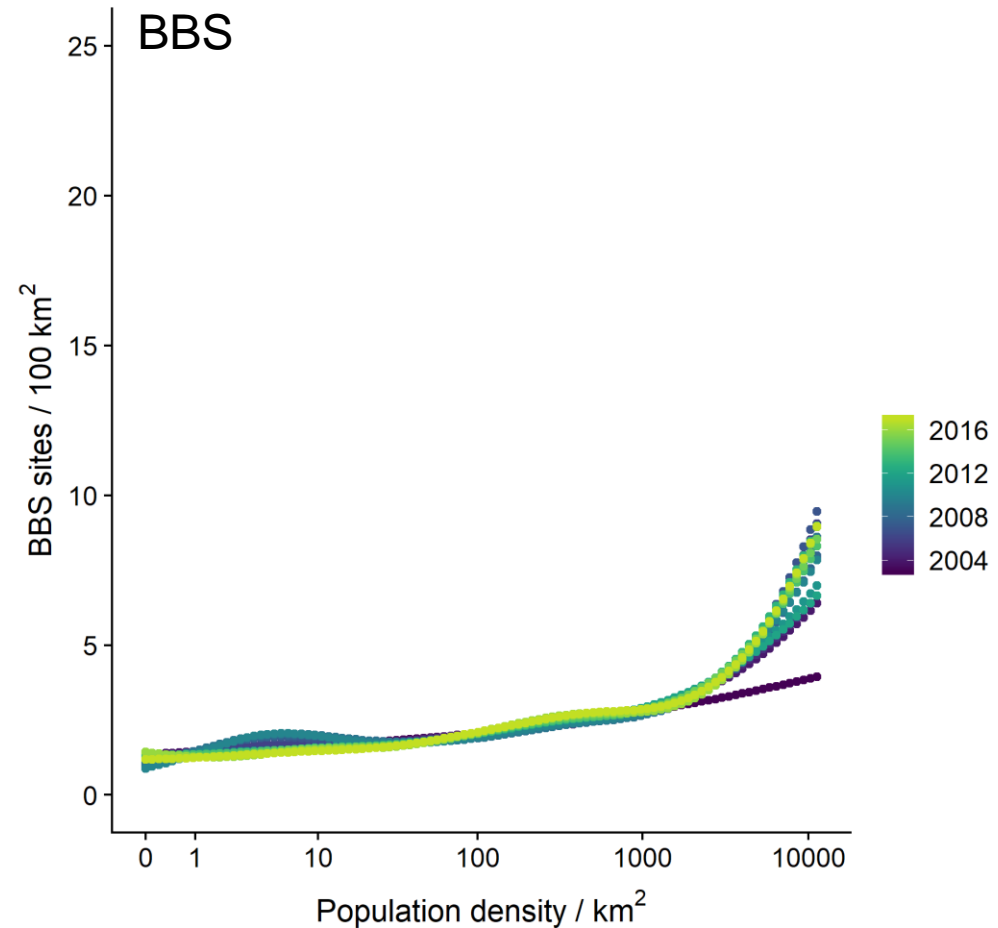
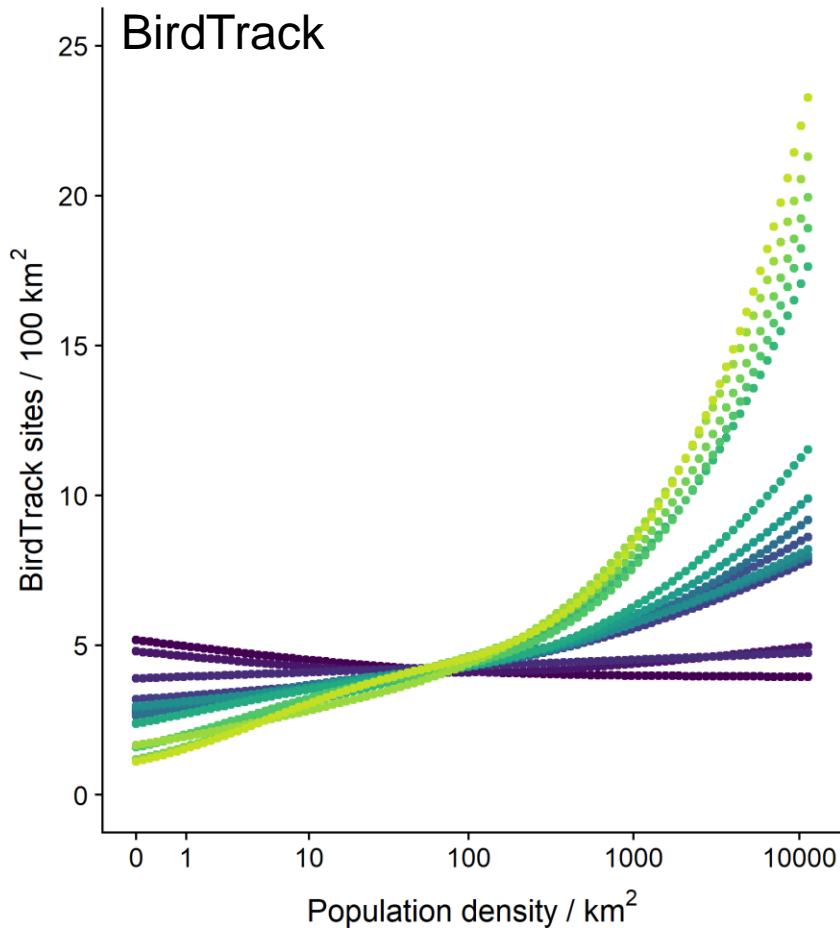
BirdTrack sites are not randomly selected

BirdTrack sites are biased towards urban areas, coasts, reserves



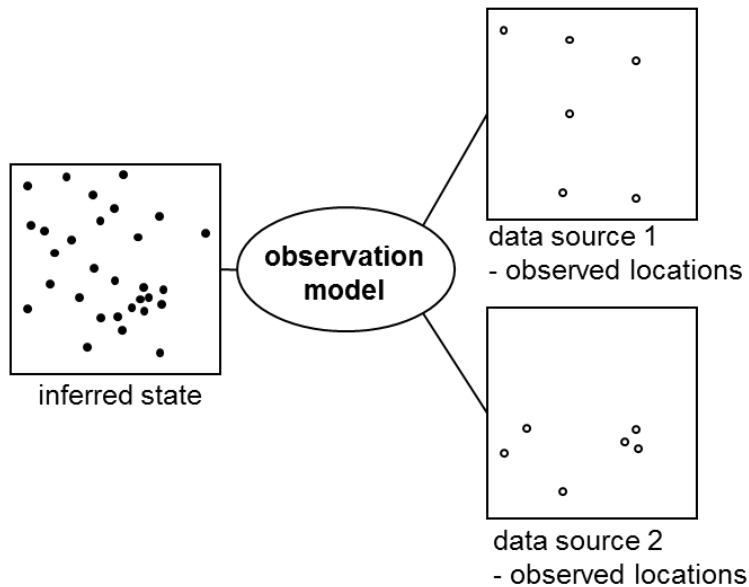
👎 Site selection bias is non-stationary

BirdTrack spatial bias is *increasingly* urban

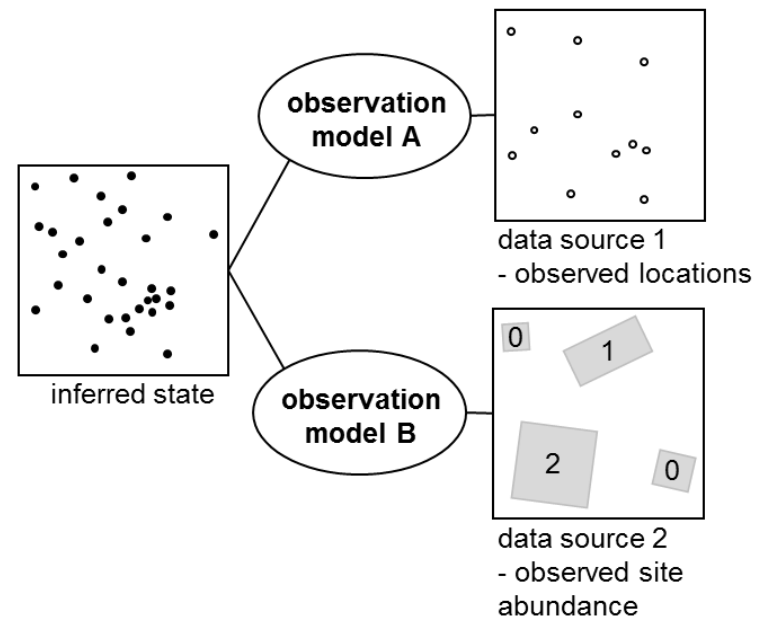


How do we integrate these data?

(A) data merging



(B) model-based data integration



- **BBS**
 - Abundance trends: maximum annual count per site
 - Occurrence trends: detection-nondetection based on max. count
- **BirdTrack**
 - species detection-nondetection
 - complete, timed lists with a 1km grid reference
 - recorded in BBS survey window (April – June)
 - Locations with <2 lists removed for trend calculation
 - $n_{\text{lists}} = 321,901$; $n_{\text{locations}} = 22568$
- **Analysis timeframe:** 2005 – 2016
- **Species set:** 141 species that are reasonably covered by BBS