Integrating citizen science data to estimate bird population trends

Philipp Boersch-Supan @pboesu

British Trust for Ornithology @_BTO

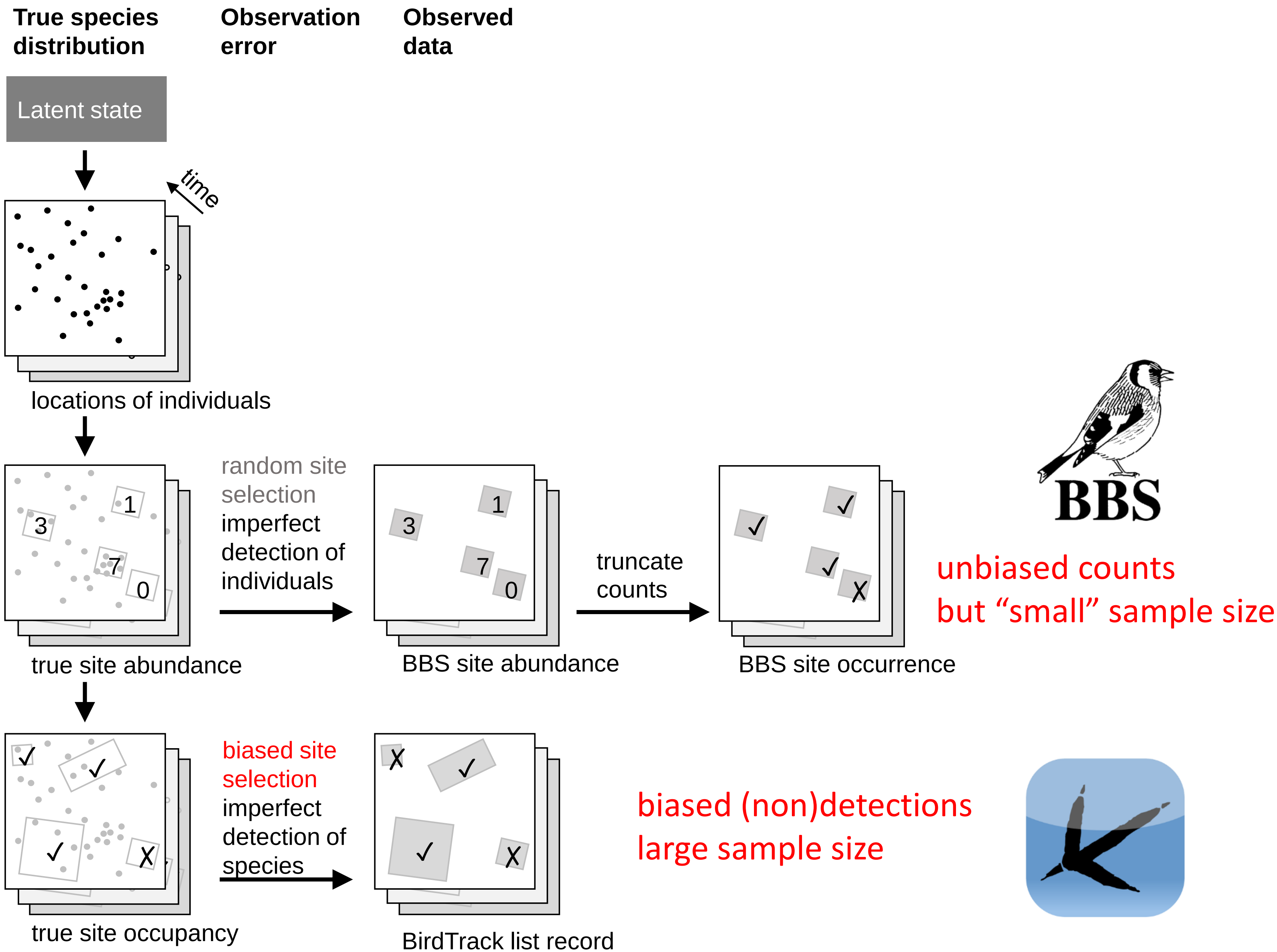
How do we count UK breeding birds?

- **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

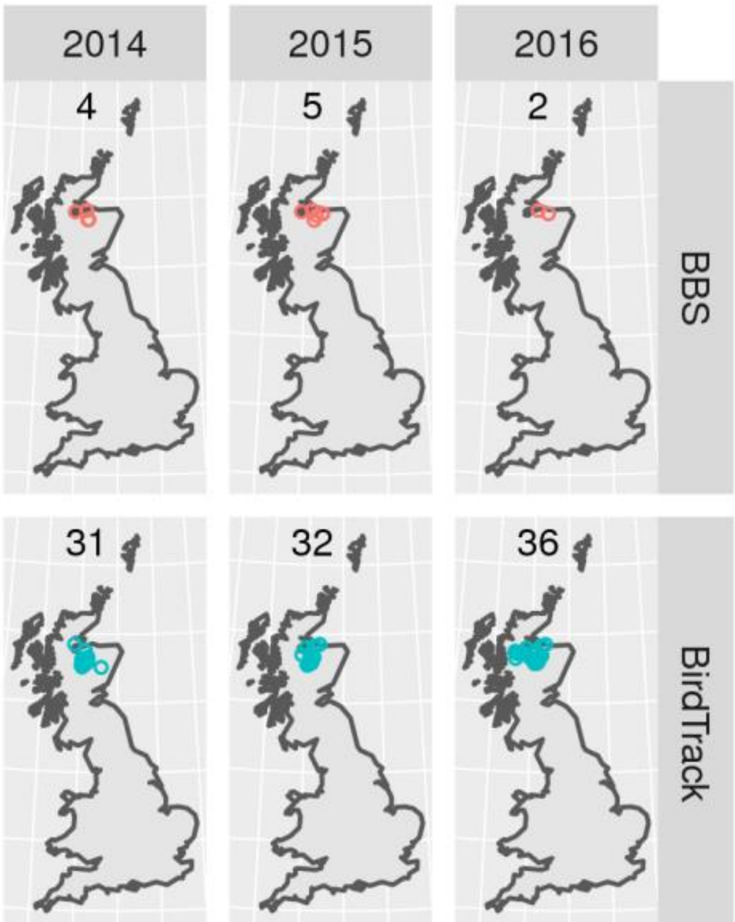
  → **Bigger data (currently ~15k sites, ~100k lists per year)**



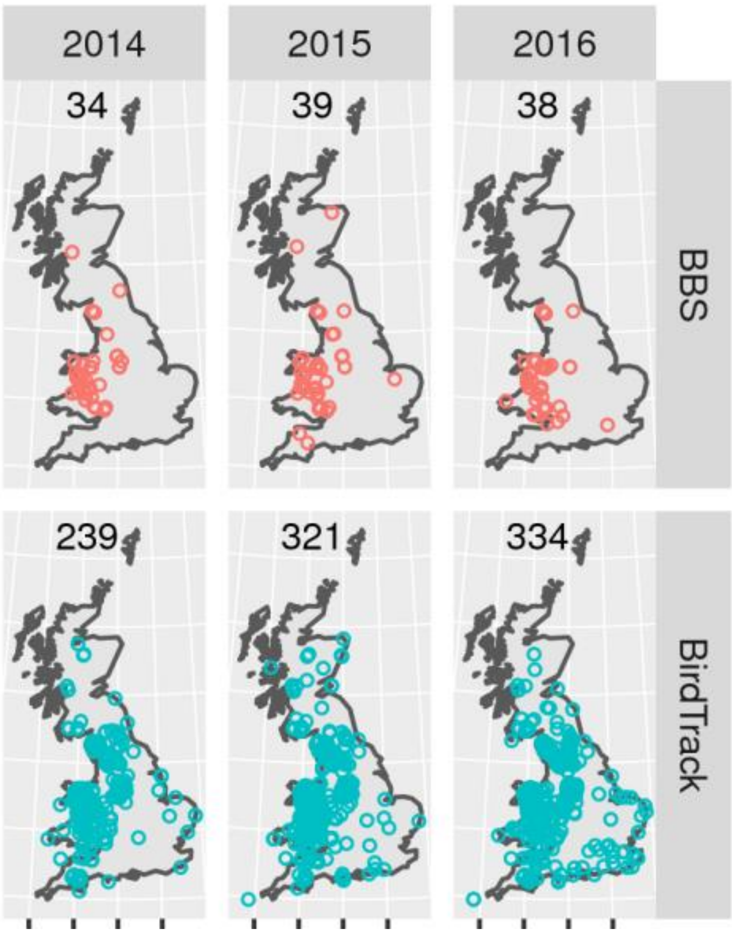
after Isaac et al. 2019 TREE

👍 BirdTrack has c. 10x more records across space

Crested Tit

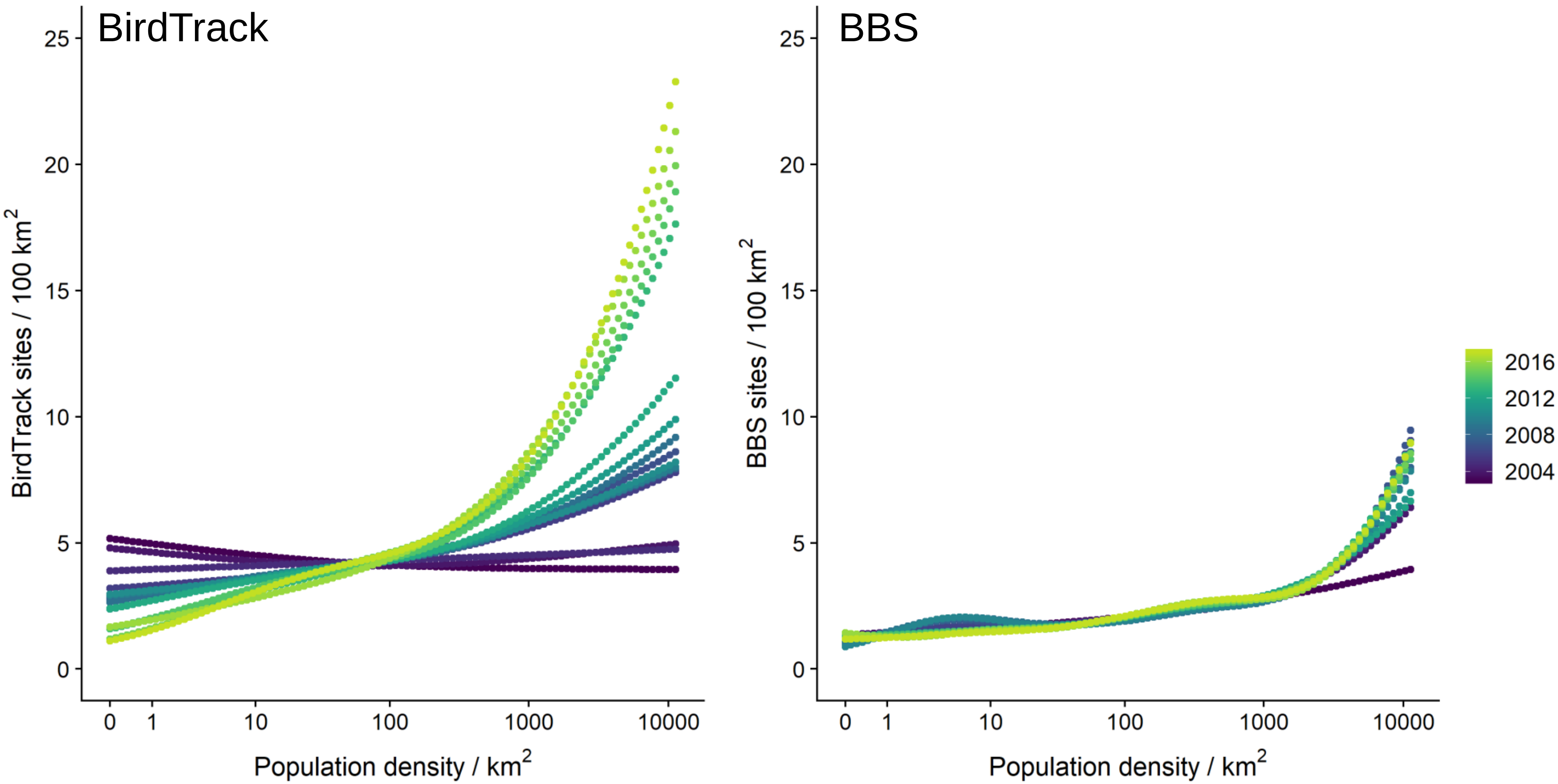


Pied Flycatcher



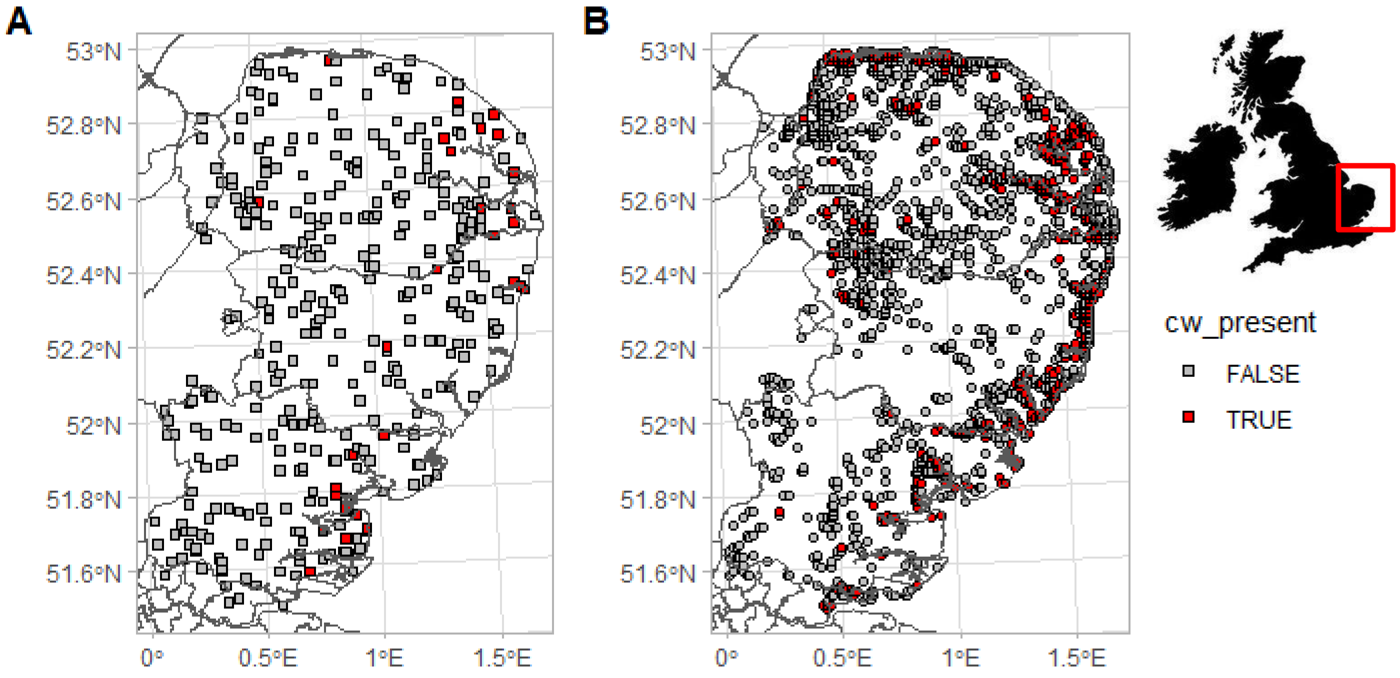
Site selection bias is non-stationary

BirdTrack spatial bias is *increasingly* urban

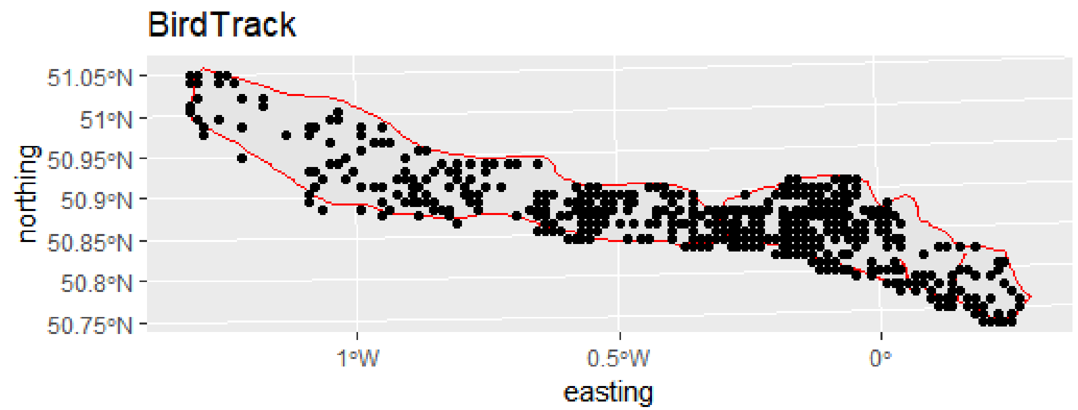


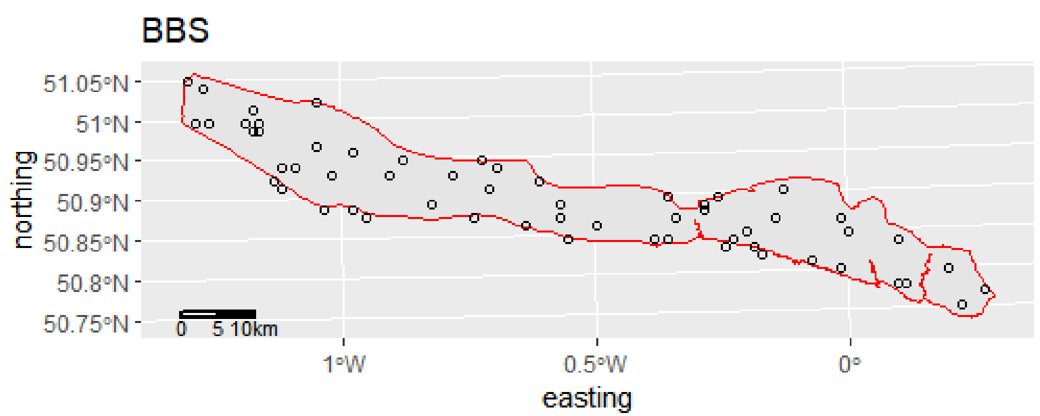
Boersch-Supan et al. in prep.

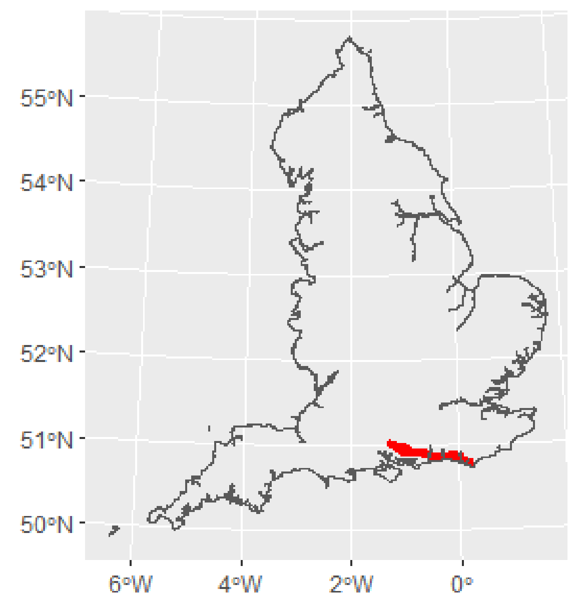
Integrated state-space model for *Cettia cetti*



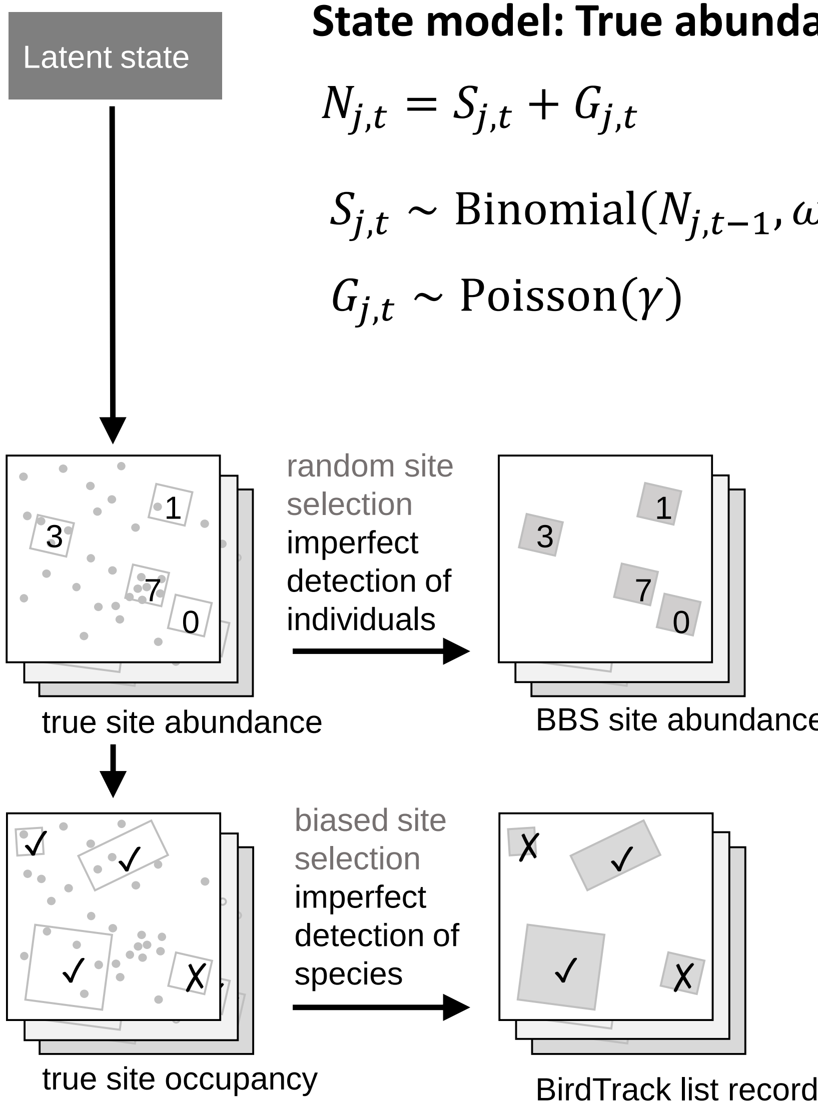
South Downs Corn Bunting *Emberiza calandra*







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



State model: True abundance = Survival + Recruitment

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

$$S_{j,t} \sim \mathrm{Binomial}(N_{j,t-1}, \omega) \qquad \mathrm{logit}(\omega_{j,t}) = \beta_{\omega} + \beta_{frost} x_{frost,j,t}$$

$$G_{j,t} \sim \mathrm{Poisson}(\gamma)$$

Count observations: N-mixture model

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

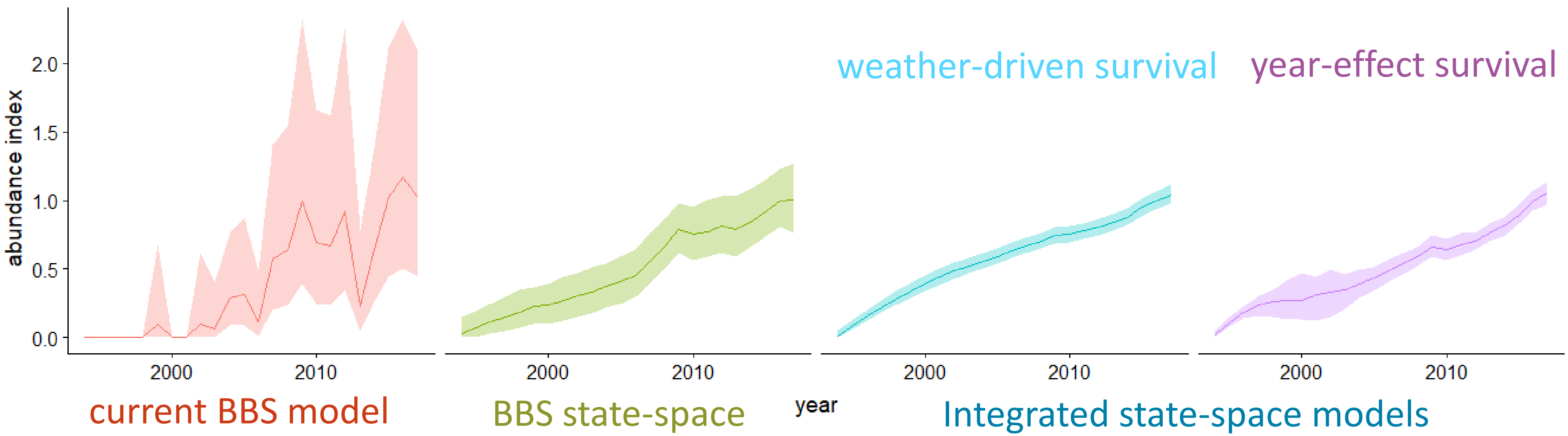
Species list observations:

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

$$\mathrm{logit}(p_{occ,j,t,k}) = \beta_p + \beta_{obs} x_{obs,j,t,k}$$

Boersch-Supan et al. in prep., Zipkin et al. 2018 Ecology

Initial results for *Cettia cetti*



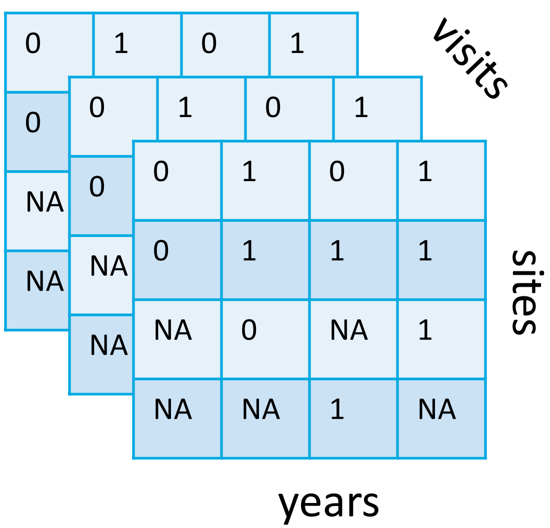
- big improvement in precision for **abundance trend**
- good agreement with data from Rare Breeding Birds Panel
- computational effort is high
  - too high for routine UK-scale applications
  - but country-level joint trends are in reach

Boersch-Supan et al. in prep

Implementation details

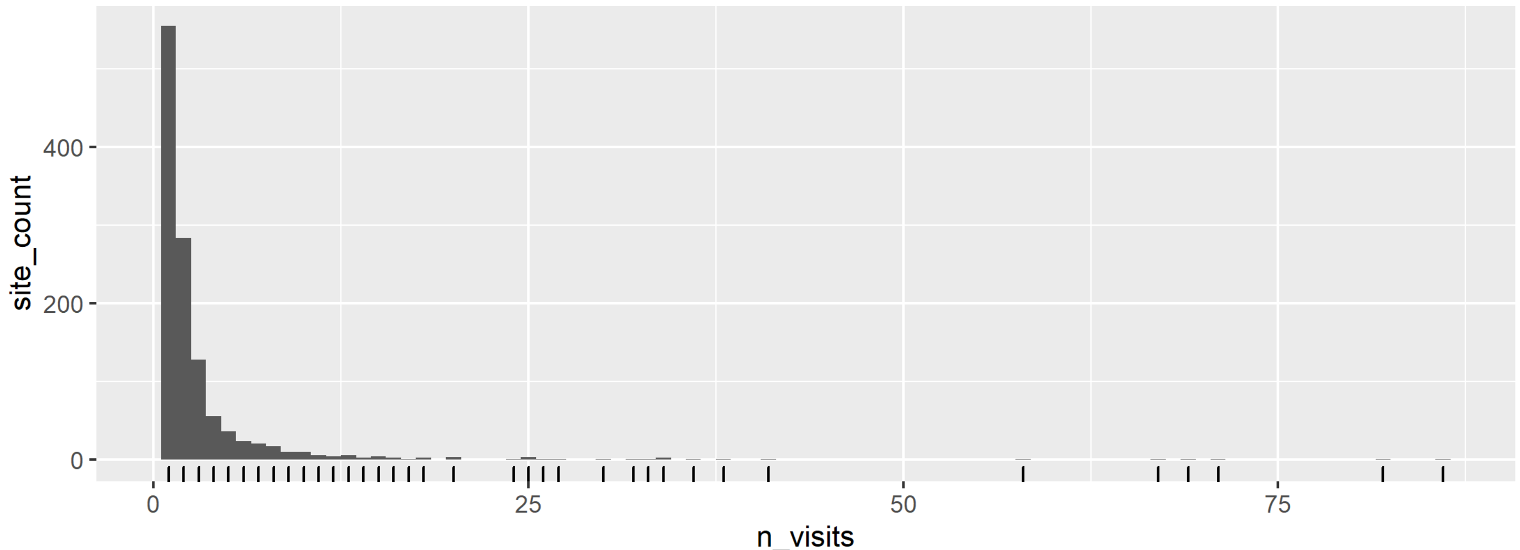
- Model is computationally costly because there’s a latent state variable for every site
- Standard observation model implementation make matters worse given the sparsity of the data

```
# 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
```



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*4 = 15236 array elements; c. 2.5% of i-j-k array

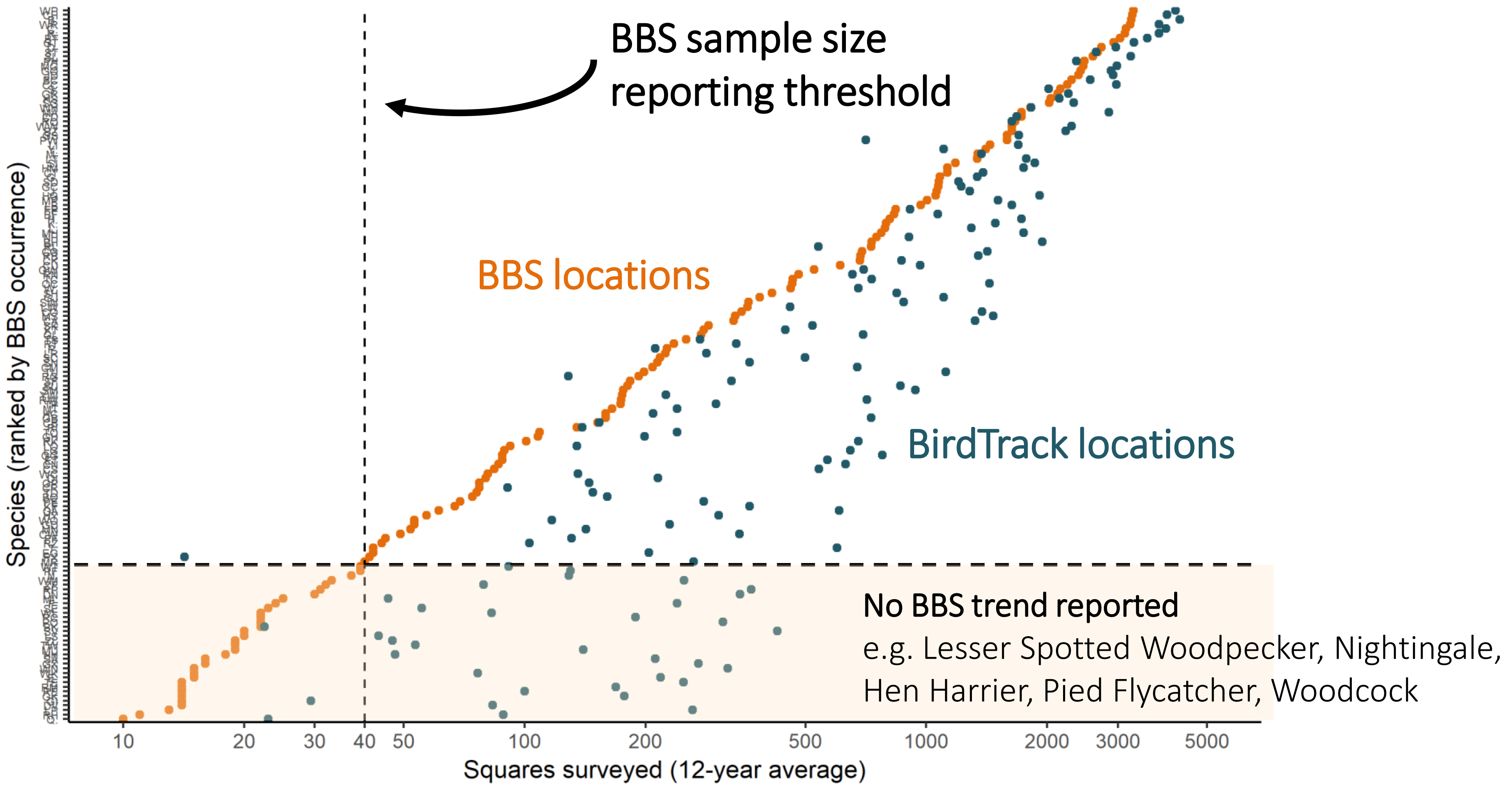
Tipling / BTO

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.

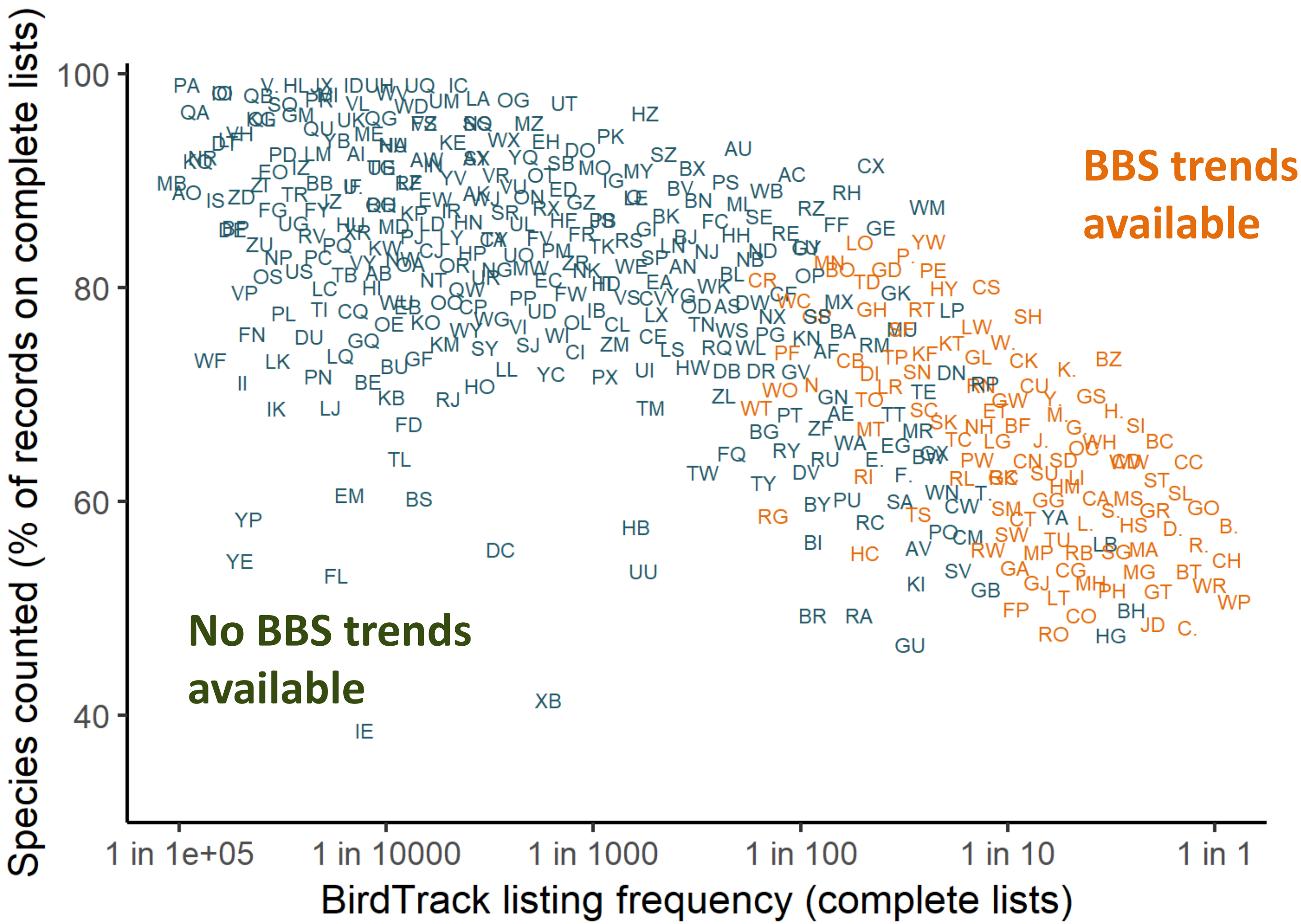
BirdTrack has c. 10x more records across space

BirdTrack captures species in many more places than the BBS



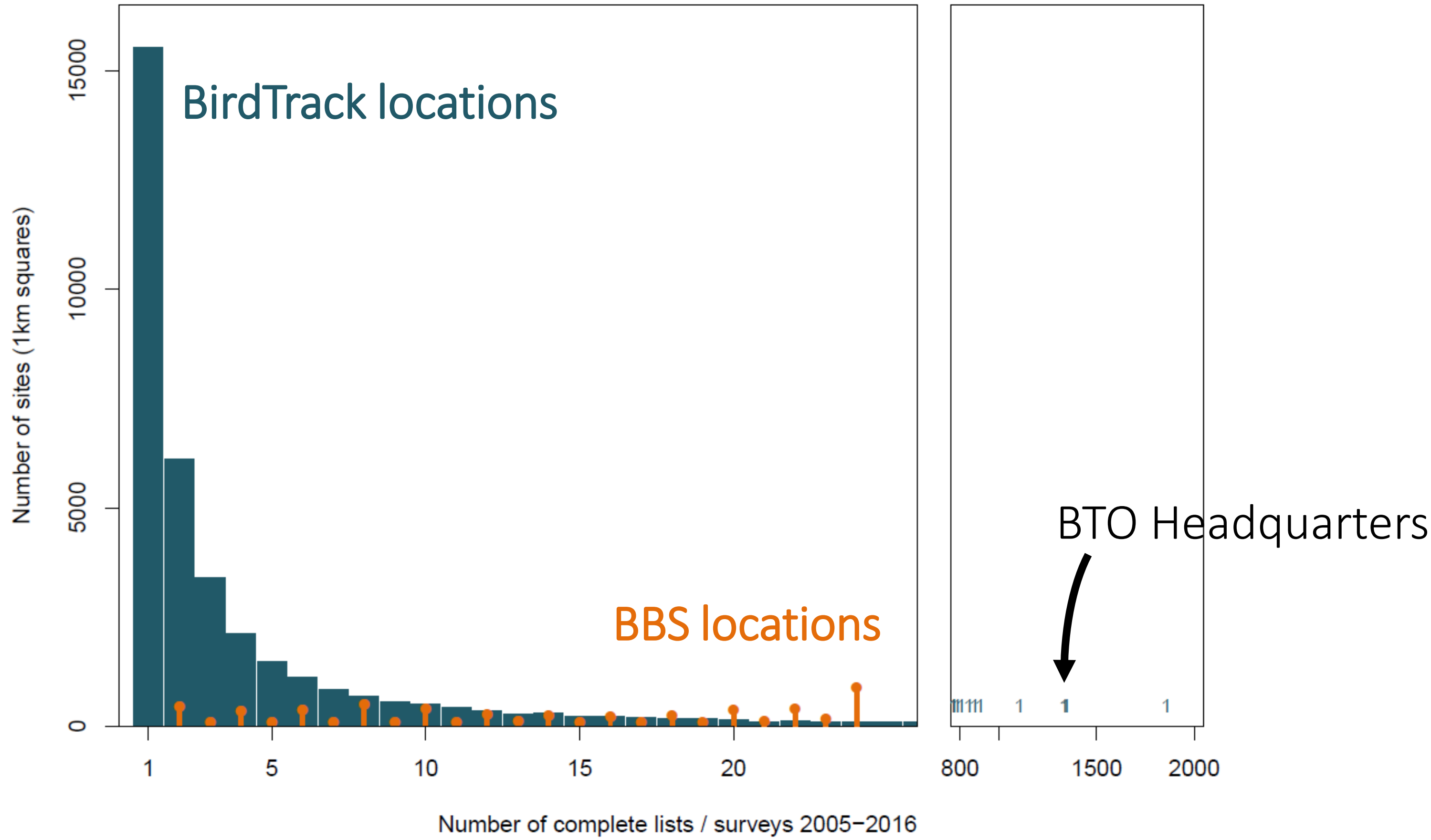
BirdTrack users don’t count all species

Common species are less likely to be counted by BirdTrack users



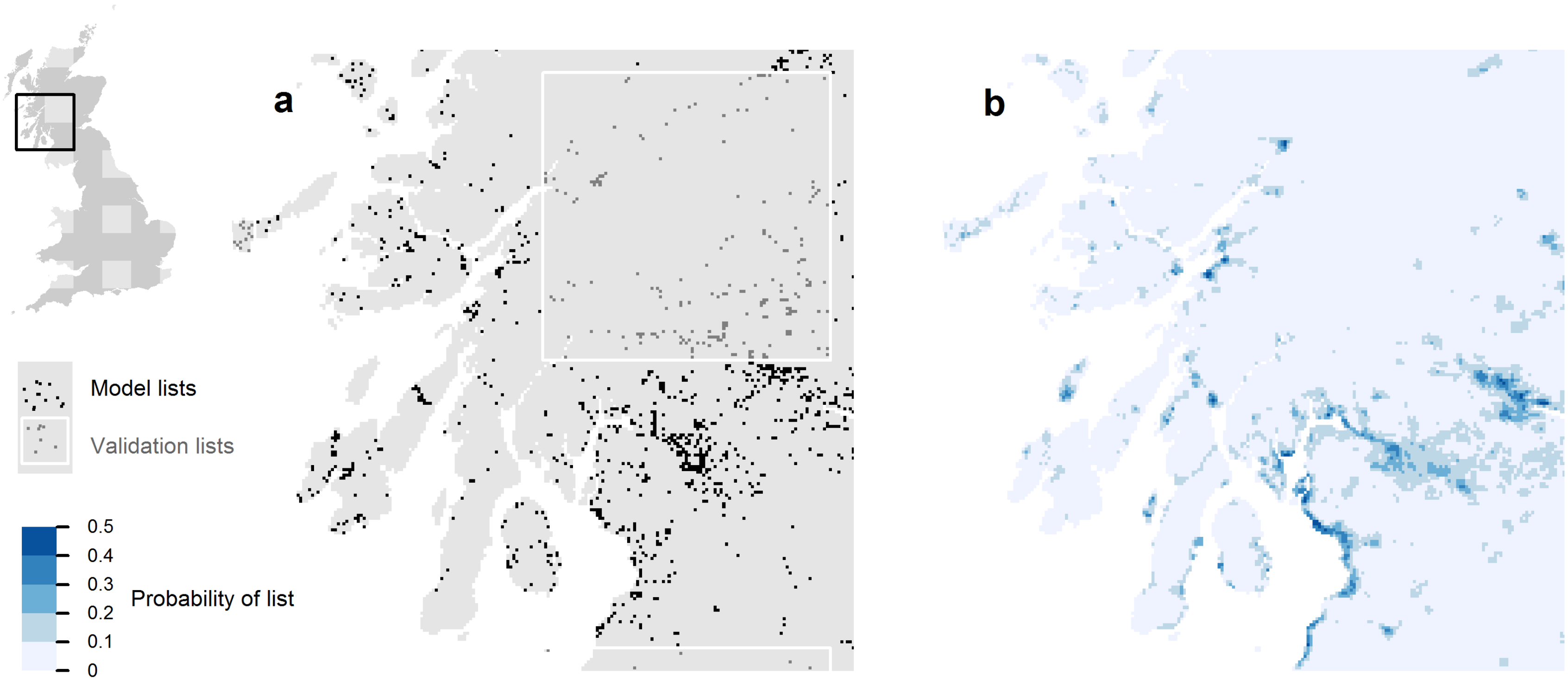
Most BirdTrack sites are one-offs

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



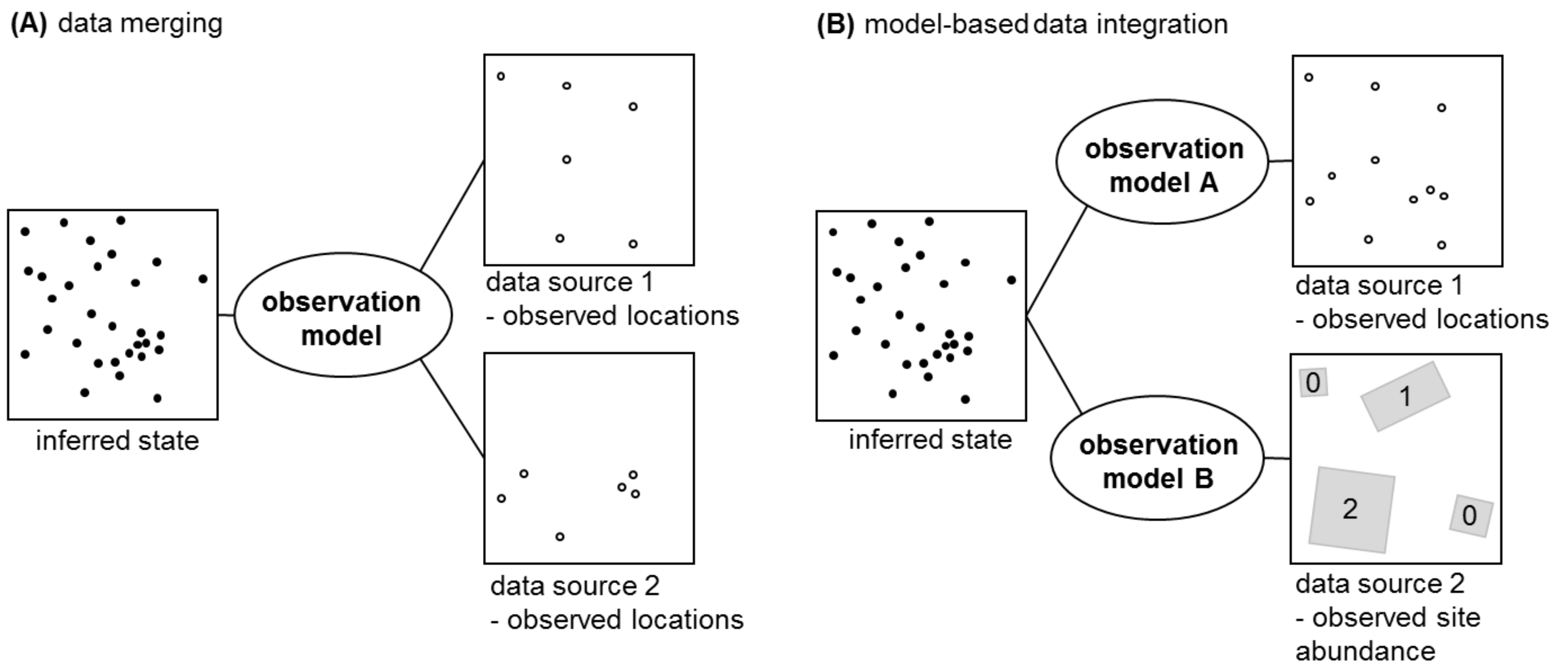
BirdTrack sites are not randomly selected

BirdTrack sites are biased towards urban areas, coasts, reserves



Johnston et al. in review

How do we integrate these data?



Isaacs et al. in revision at TREE, see also Miller et al. 2019 MEE, Zipkin et al 2019 Ecology, Pacifici et al. 2019 Ecology

Data

- **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_{lists} = 321,901; $n_{locations}$ = 22568
- **Analysis timeframe:** 2005 – 2016
- **Species set:** 141 species that are reasonably covered by BBS

Boersch-Supan et al. submitted