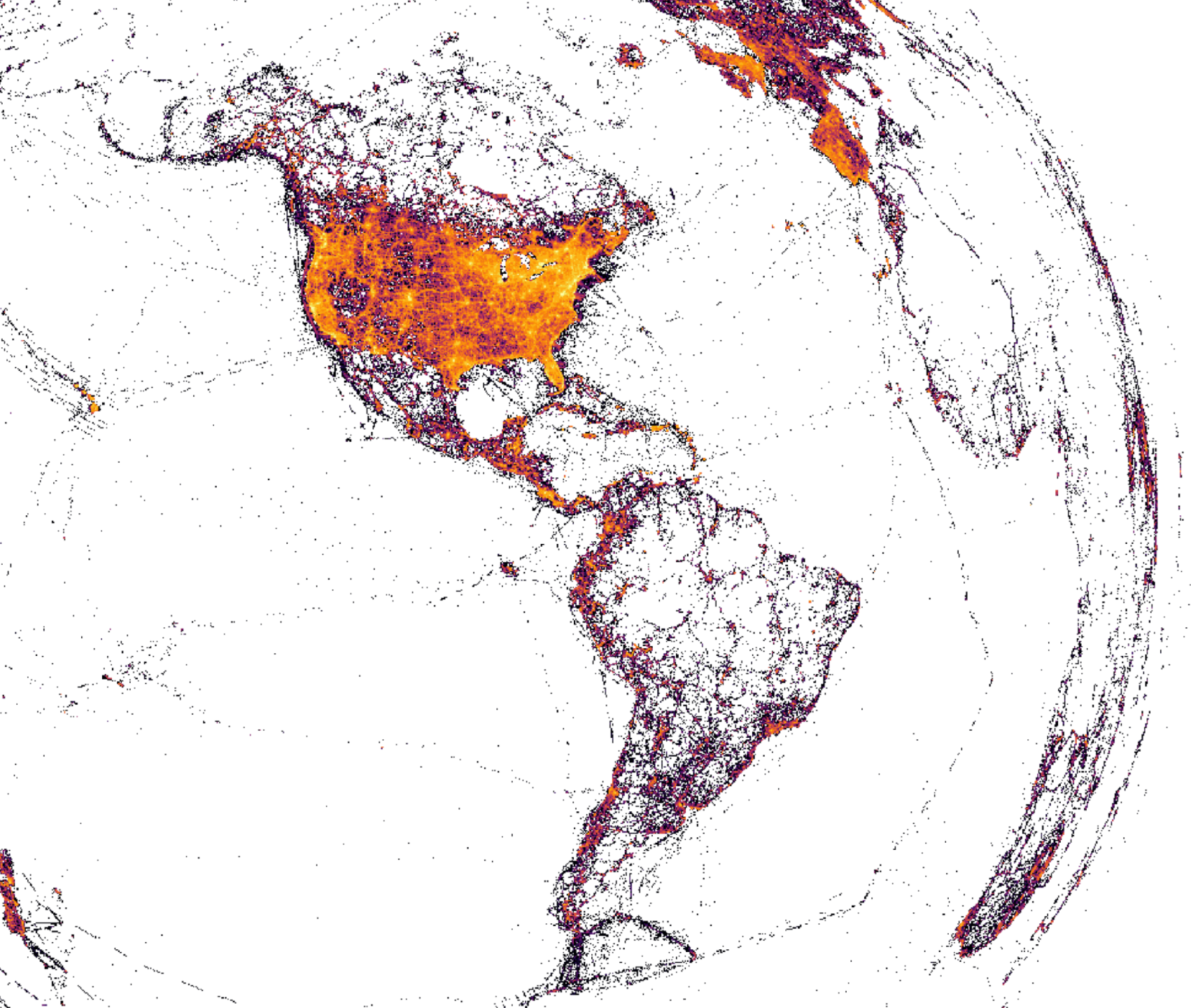
eBird Best Practices II

General principles for modeling

The Cornell Lab of Ornithology

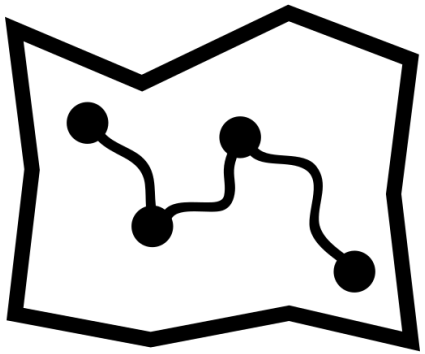


eBird

May 2019

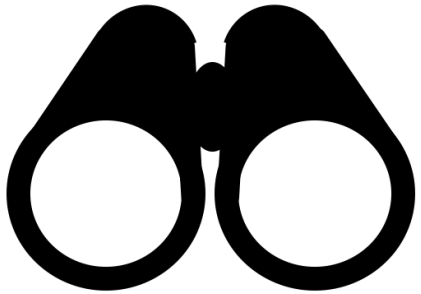
1.2 Million checklists
20 Million observations

Structured surveys



Standardized observation process

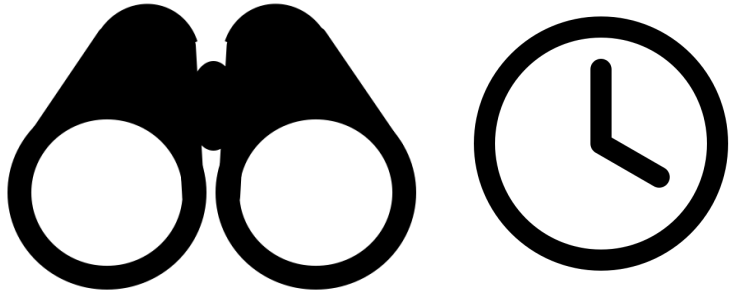
Unstructured/ opportunistic



Messy observation process

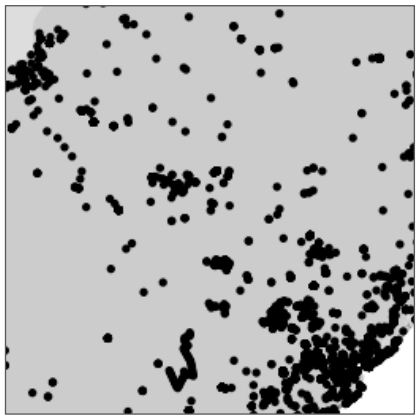
eBird

Semi-structured



Messy observation process **with effort information**

Challenges with eBird observation process



Spatial
bias

Variation
in protocol

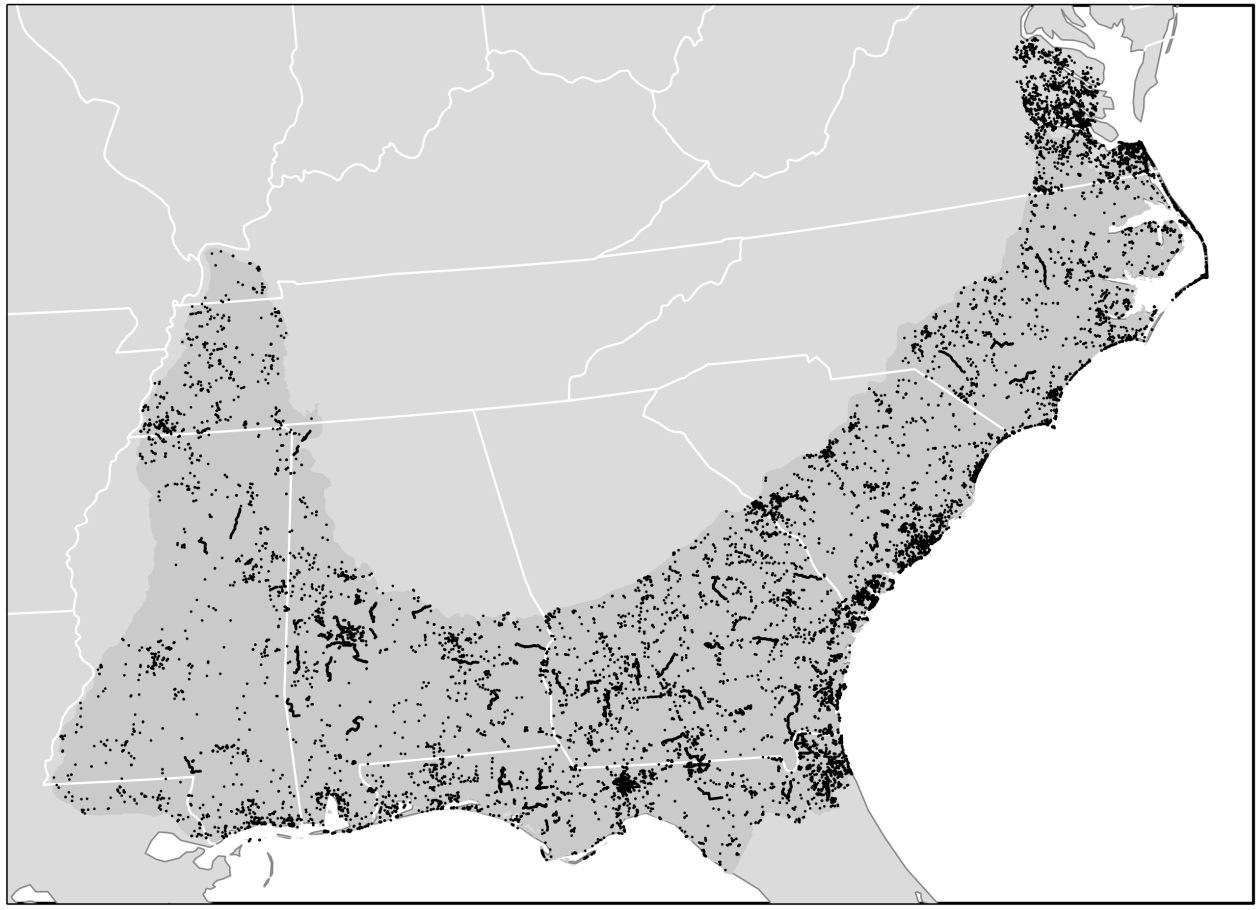
Species
bias

Variation
in observers

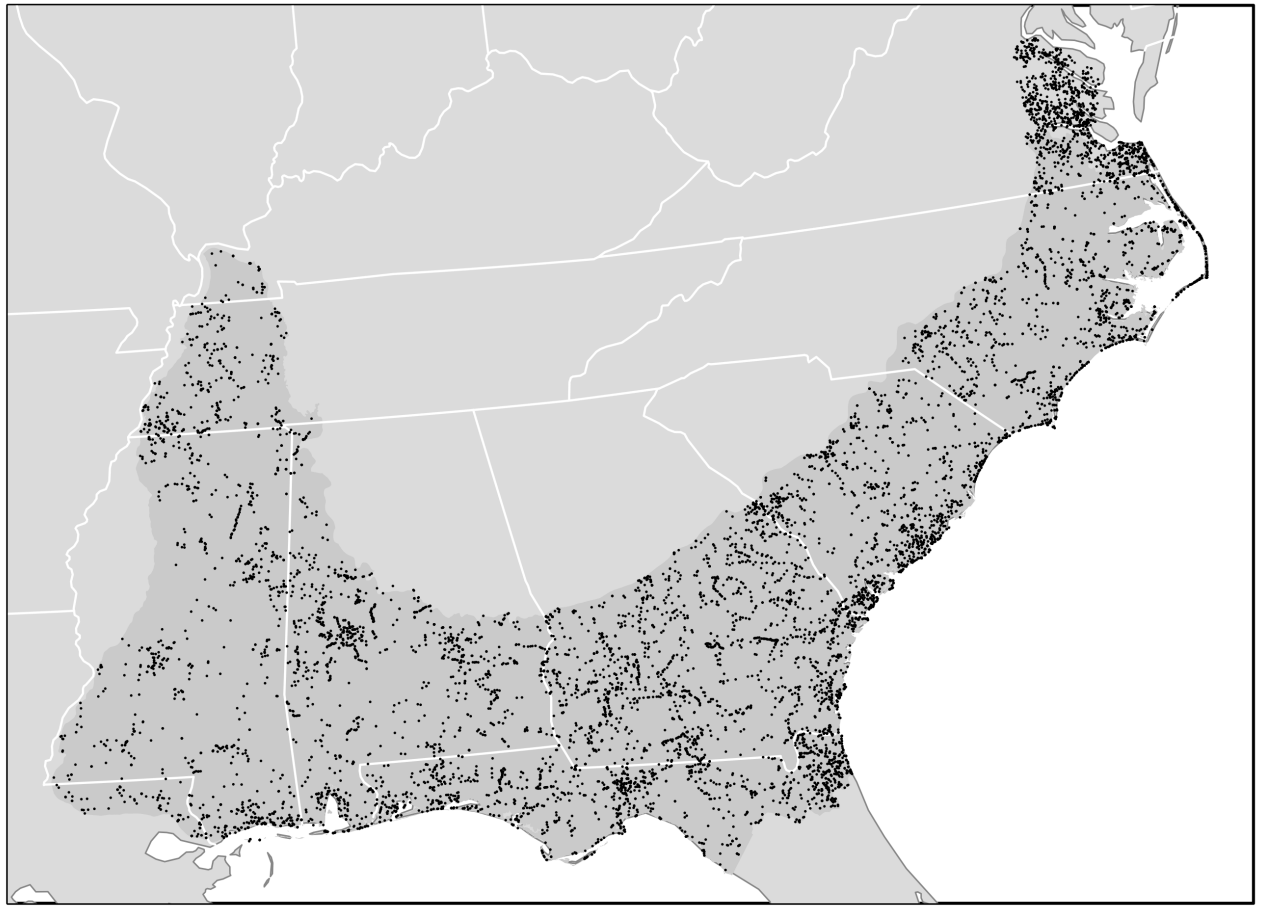
FILTERING

ANALYSIS

Spatial bias



June 2016 – all checklists



June 2016 – after spatial subsampling

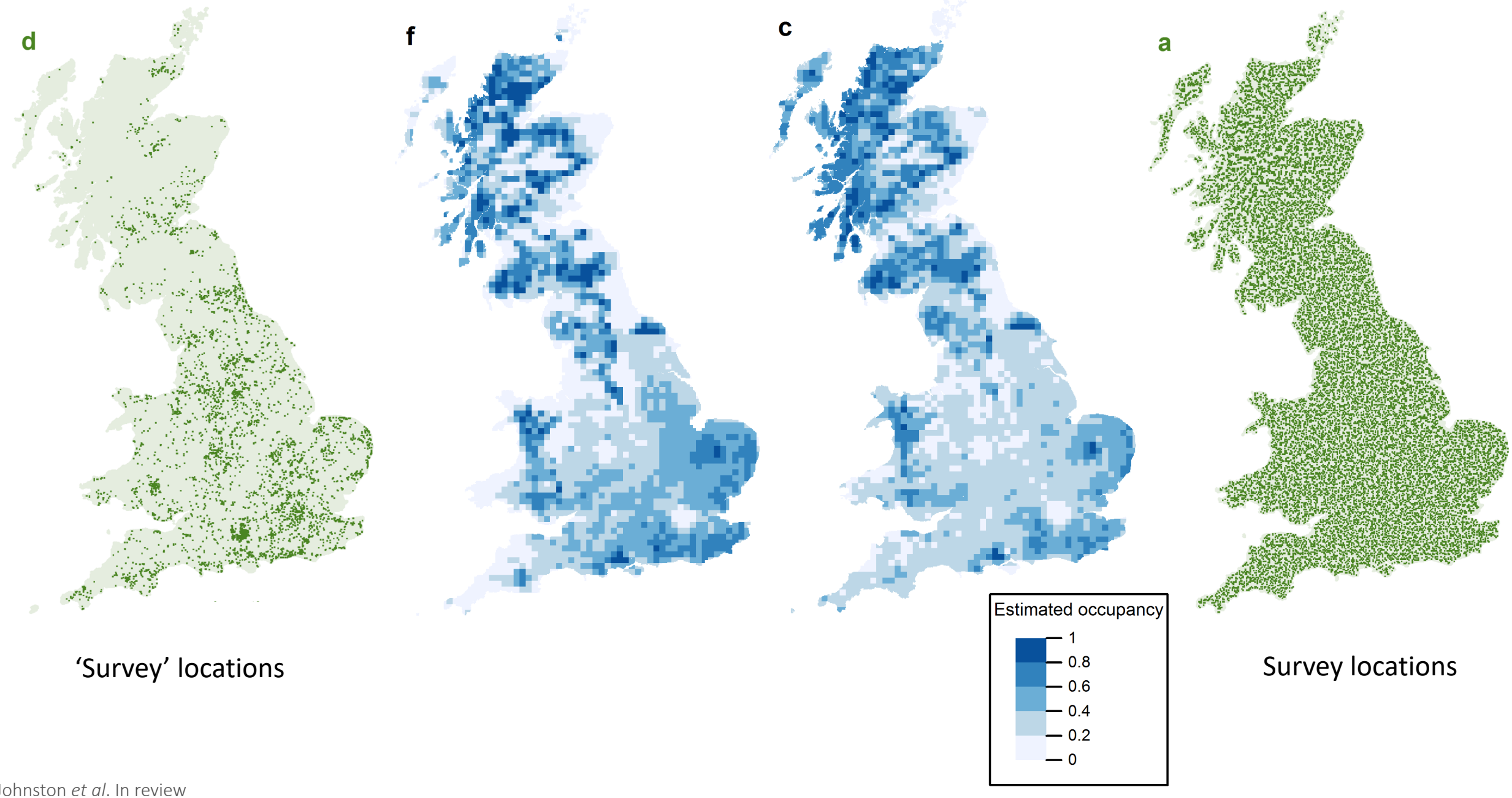
Large scale and fine-scale spatial bias

Fine-scale spatial bias

Johnston *et al*. 2019 *BioRXiv*

OPPORTUNISTIC DATA

STRUCTURED DATA



‘Survey’ locations

Survey locations

Johnston *et al*. In review

Species bias

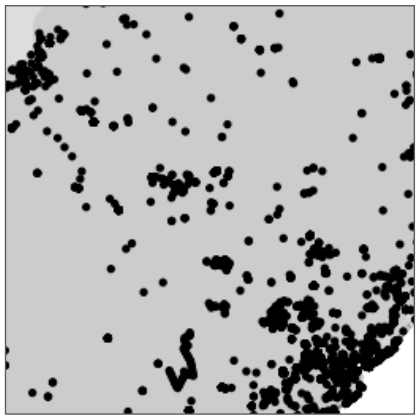







Schuetz & Johnston 2019 *PNAS*

Challenges with eBird observation process



Spatial
bias

Variation
in protocol

Species
bias

Variation
in observers

FILTERING

ANALYSIS

Variation in protocol

Checklist duration

Distance travelled

Time of day

Number of observers

Johnston *et al*. 2019 *BioRXiv*

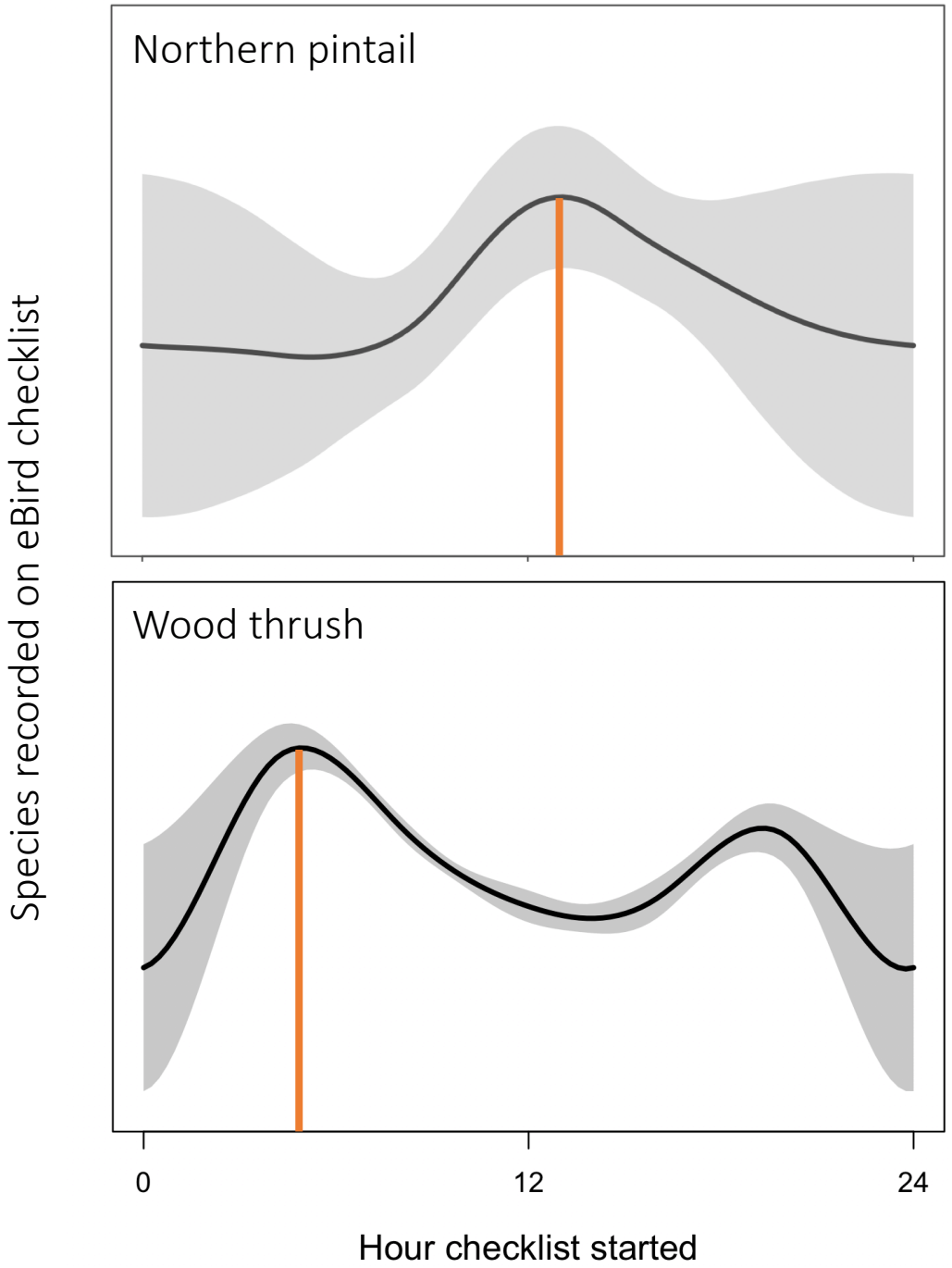
Variation in protocol

Checklist duration

Distance travelled

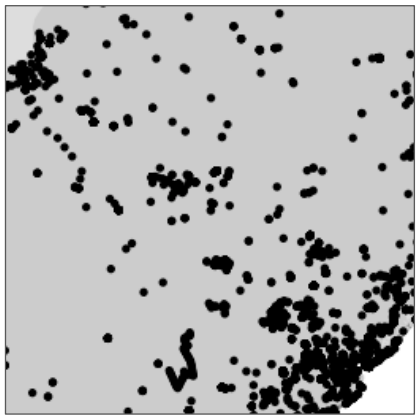
Time of day

Number of observers



Johnston *et al*. 2019 *BioRXiv*

Challenges with eBird observation process



Spatial
bias

Variation
in protocol

Species
bias

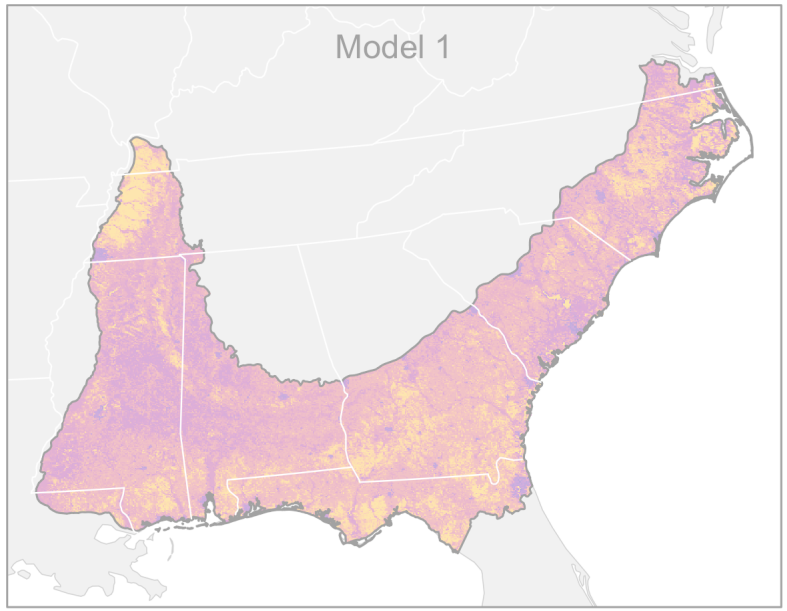
Variation
in observers

FILTERING

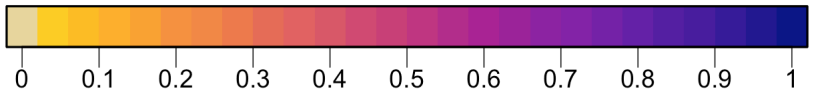
ANALYSIS

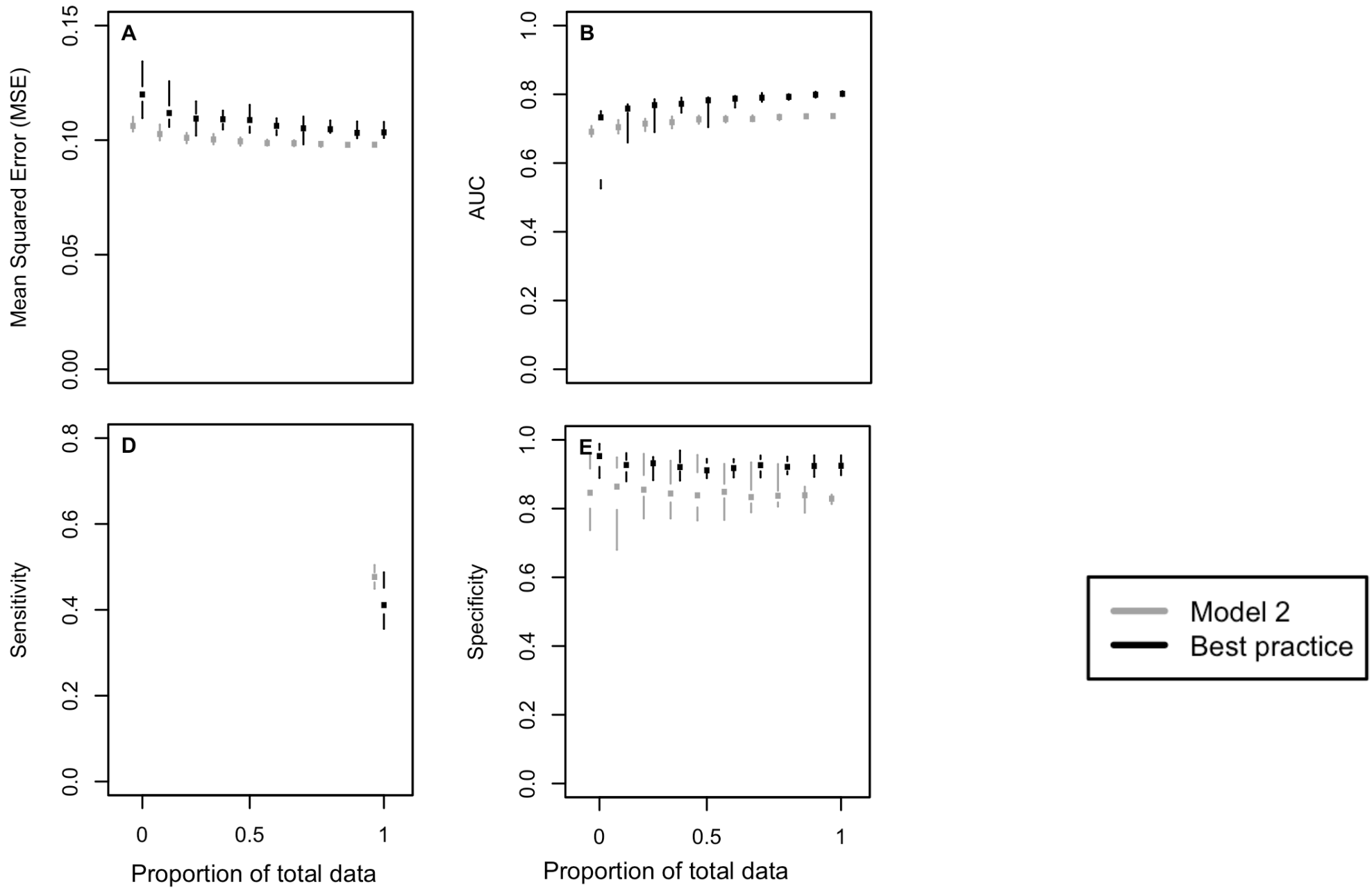
Advice for modelling eBird data

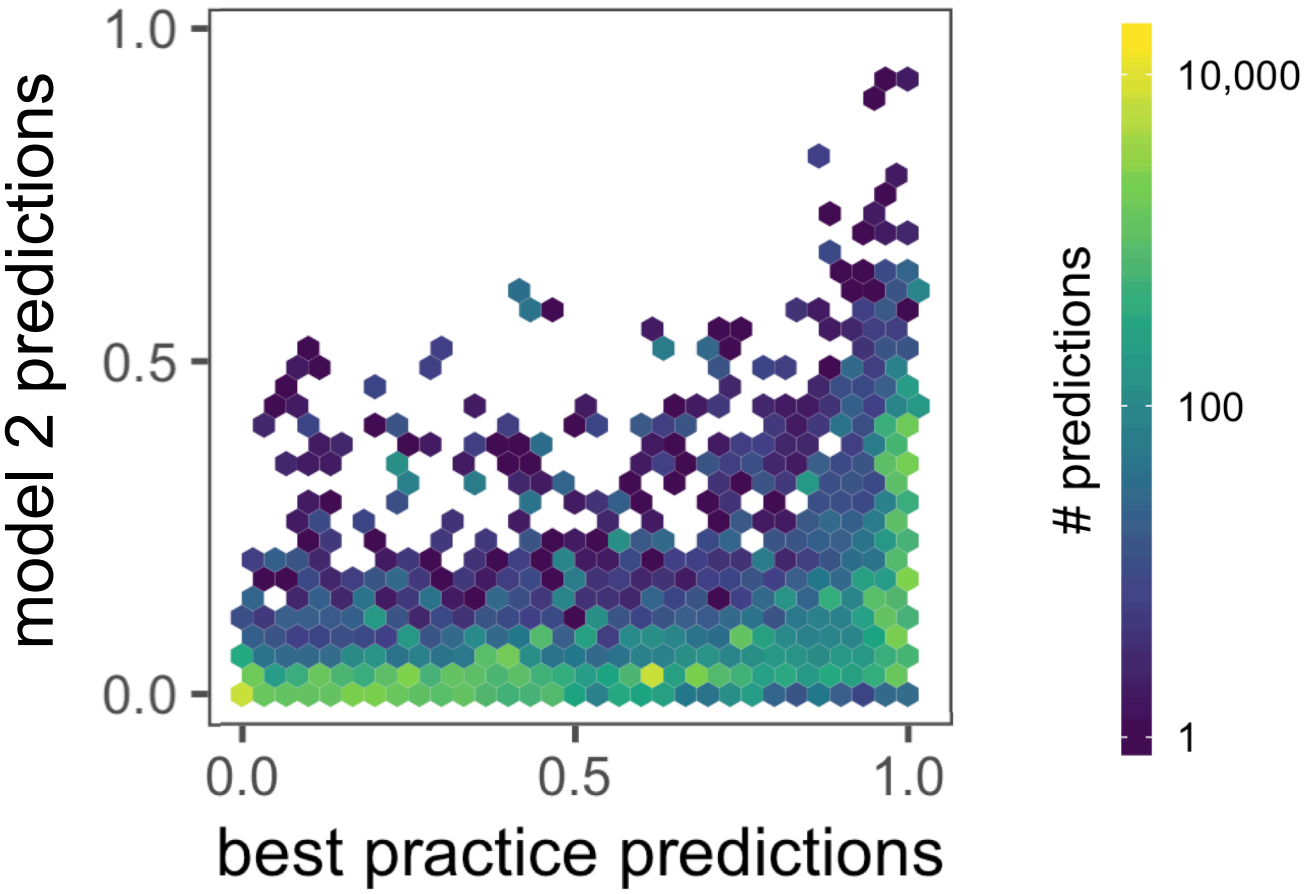
Model 1	
Model 2	Use non-detections (absences)
Model 3	Use complete lists
Model 4	Spatial subsampling
Model 5	Filter effort variables
Model 6	Include effort variables as covariates

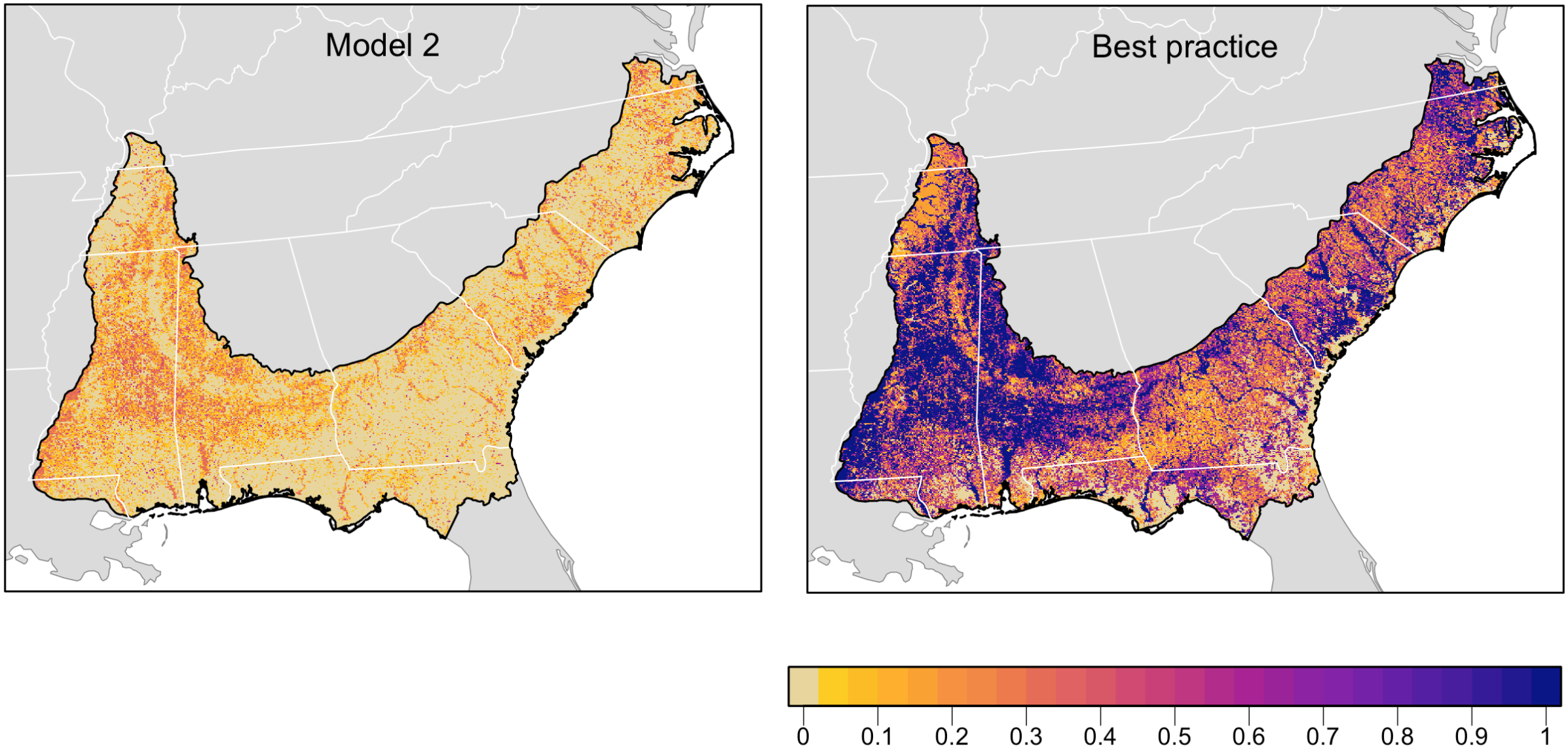


Encounter rate




A
B
D
E
Mean Squared Error (MSE)
AUC
Sensitivity
Specificity
0.00
0.05
0.10
0.15
0.0
0.2
0.4
0.6
0.8
1.0
0
0.5
1
Proportion of total data
Proportion of total data
Model 2
Best practice


model 2 predictions
1.0
0.5
0.0
0.0
0.5
1.0
best practice predictions
predictions
10,000
100
1


Model 2
Best practice
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Advice for modelling eBird data

Model 1	
Model 2	Use non-detections (absences)
Model 3	Use complete lists
Model 4	Spatial subsampling
Model 5	Filter effort variables
Model 6	Include effort variables as covariates

Johnston *et al*. 2019 *BioRXiv*

Variation in protocol

	Filters
Checklist duration	5 hours
Stationary/traveling	Remove incidental
Distance travelled	8km / 5 miles
Time of day	All times
Number of observers	10

eBird Best Practices II

General principles for modeling

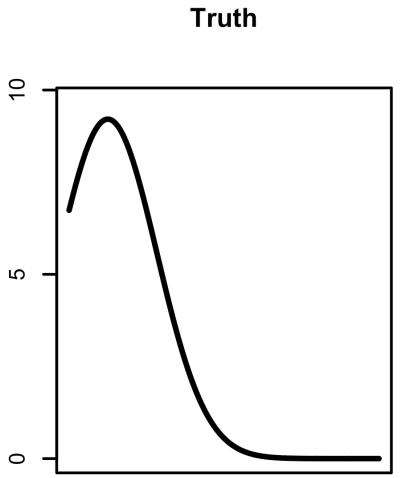
The Cornell Lab of Ornithology

What do we want to know?

What can we estimate?

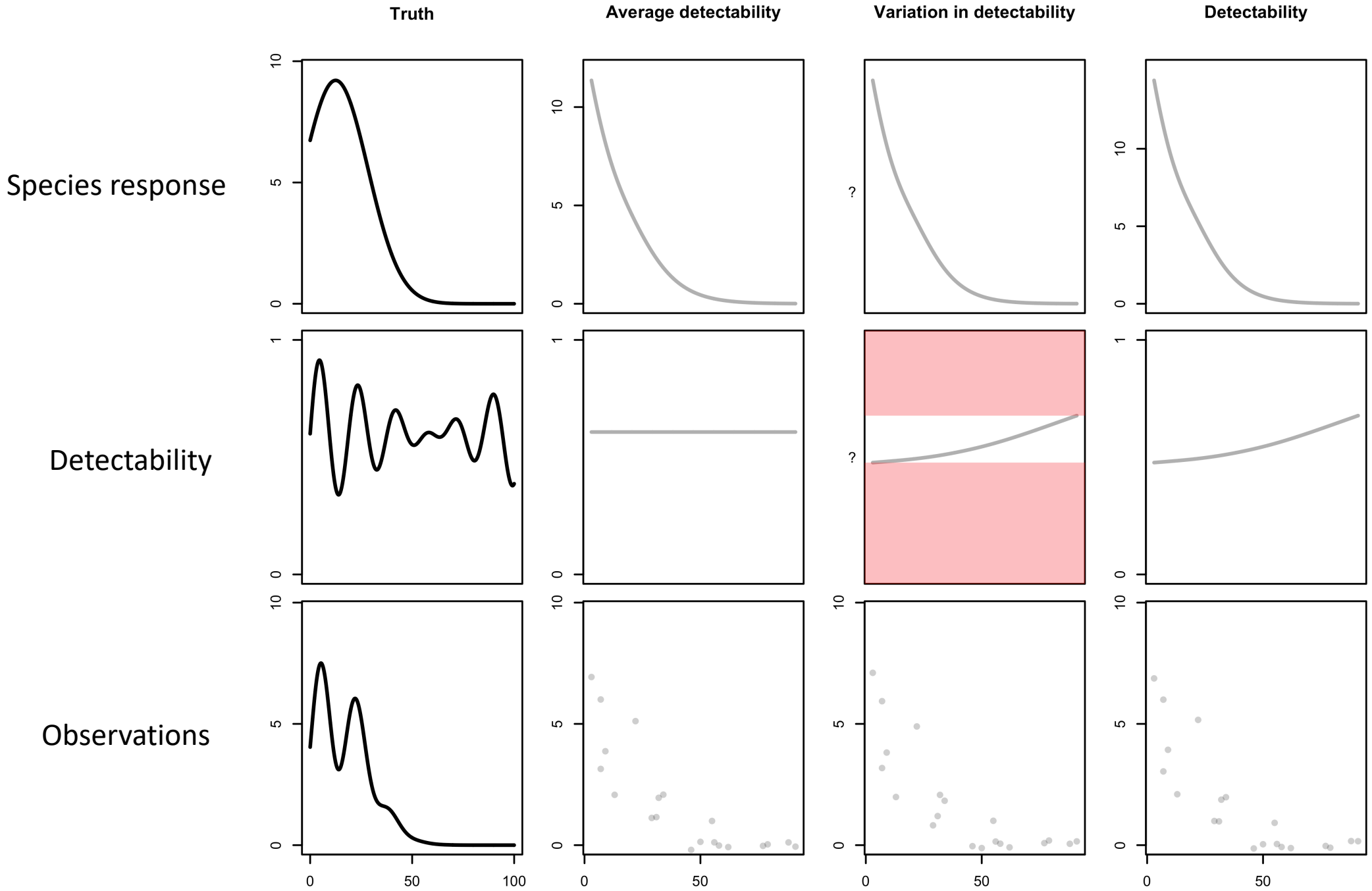
What are the gaps?

Species response



Detectability

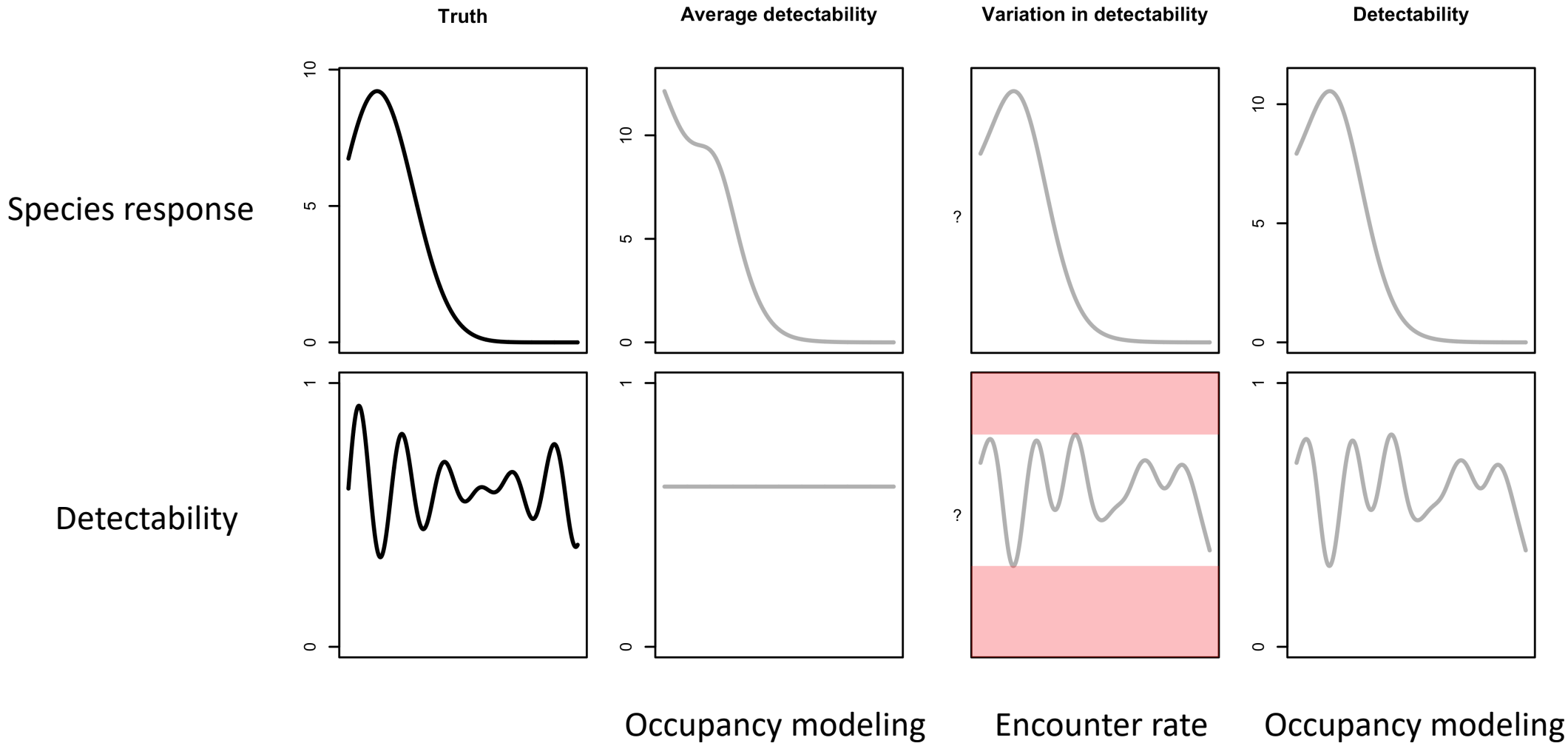
Observations


Truth
Average detectability
Variation in detectability
Detectability
Species response
Detectability
Observations
?
?
0
5
10
1
0
50
100

What do we want to know?

What can we estimate?

What are the gaps?


Truth
Average detectability
Variation in detectability
Detectability
Species response
Detectability
10
5
0
1
?
Occupancy modeling
Encounter rate
Occupancy modeling
Relative abundance

Simpler model structure Parametric relationships		Complex model structure Black box
Easier to fit Simpler interpretation More controlled structure	ADVANTAGES	More realistic complexity Closer fit to reality
Inference Explanation Process	BEST FOR	Prediction Description Pattern

Simpler model structure
Parametric relationships

Complex model structure
Black box

Occupancy modeling

GAMs

Random forest

Inference
Explanation
Process

Prediction
Description
Pattern

Encounter rate

Occupancy

Relative abundance

eBird Best Practices II

General principles for modeling

The Cornell Lab of Ornithology