

Integrated models

Diana Bowler

German Centre for Integrative Biodiversity Research

29th August 2021

Many people are excited about integrated models!!!

Methods in Ecology and Evolution



ADVANCES IN MODELLING DEMOGRAPHIC PROCESSES | [Free Access](#)

The recent past and promising future for data integration methods to estimate species' distributions

David A. W. Miller , Krishna Pacifici, Jamie S. Sanderlin, Brian J. Reich

First published: 04 February 2019 | <https://doi.org/10.1111/2041-210X.13110> | Cited by: 4

This article has been contributed to by US Government employees and their work is in the public domain in the USA.

Trends in Ecology & Evolution

Volume 35, Issue 1, January 2020, Pages 56-67



Review

Data Integration for Large-Scale Models of Species Distributions

Nick J.B. Isaac ^{1, 2} , Marta A. Jarzyna ³, Petr Keil ^{4, 5}, Lea I. Damby ^{1, 2}, Philipp H. Boersch-Supan ^{6, 7}, Ella Browning ^{2, 8}, Stephen N. Freeman ¹, Nick Golding ⁹, Gurutzeta Guillera-Arroita ⁹, Peter A. Henrys ¹⁰, Susan Jarvis ¹⁰, José Lahoz-Monfort ⁹, Jörn Pagel ¹¹, Oliver L. Pescott ¹, Reto Schmucki ¹, Emily G. Simmonds ¹², Robert B. O'Hara ¹²



Search



Special Feature: Data Integration for Population Models

A practical guide for combining data to model species distributions

Robert J. Fletcher Jr. , Trevor J. Hefley, Ellen P. Robertson, Benjamin Zuckerberg ... [See all authors](#)

First published: 30 March 2019 | <https://doi.org/10.1002/ecy.2710> | Citations: 5

Corresponding Editor: Brian D. Inouye.

Editors' Note: Papers in this Special Feature are linked online in a virtual table of contents at: www.wiley.com/doi/10.1002/ecy.2710

Methods in Ecology and Evolution



Technological Advances at the Interface between Ecology and Statistics | [Free Access](#)

Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection

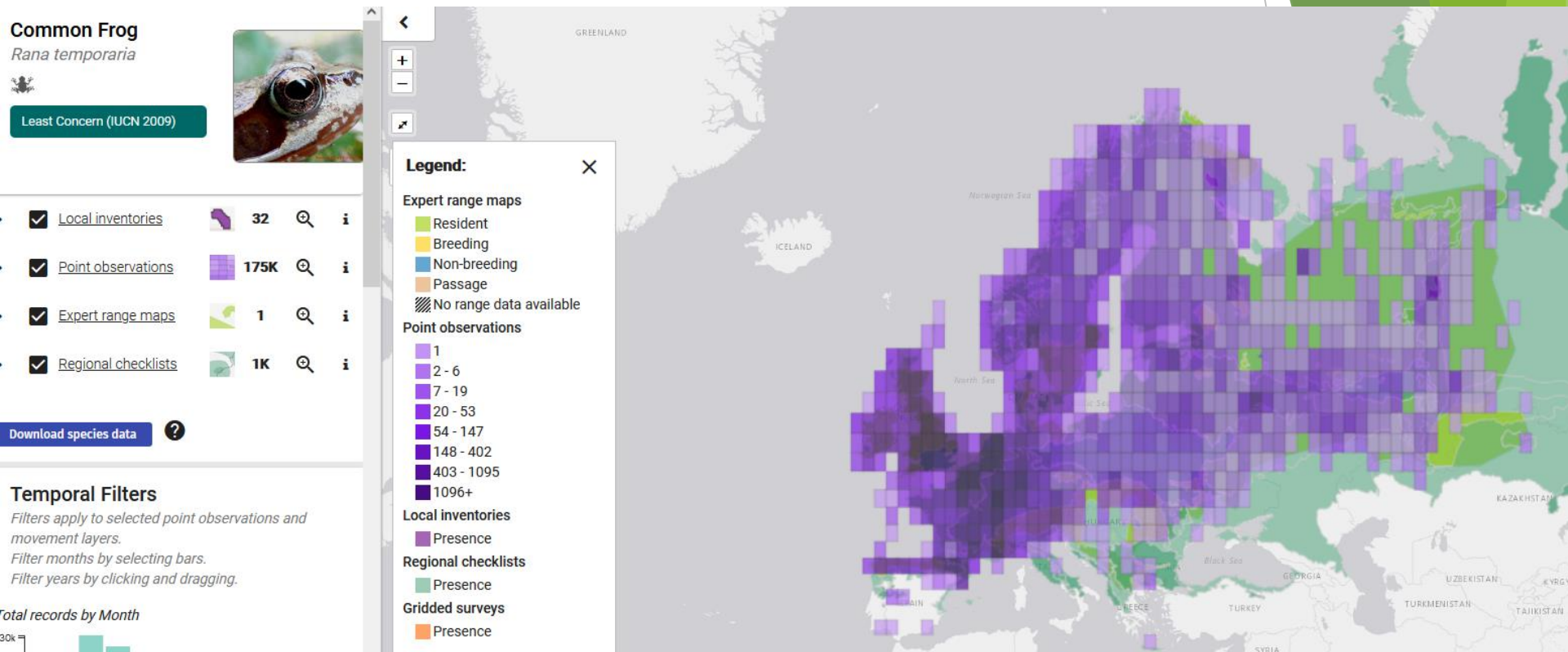
Vira Koshkina , Yan Wang, Ascelin Gordon, Robert M. Dorazio, Matt White, Lewi Stone

First published: 10 April 2017 | <https://doi.org/10.1111/2041-210X.12738> | Citations: 17

[Correction note: The article title was modified on 25 April 2017]



Multiple sources of information on species distributions



How can we make use of all this data?

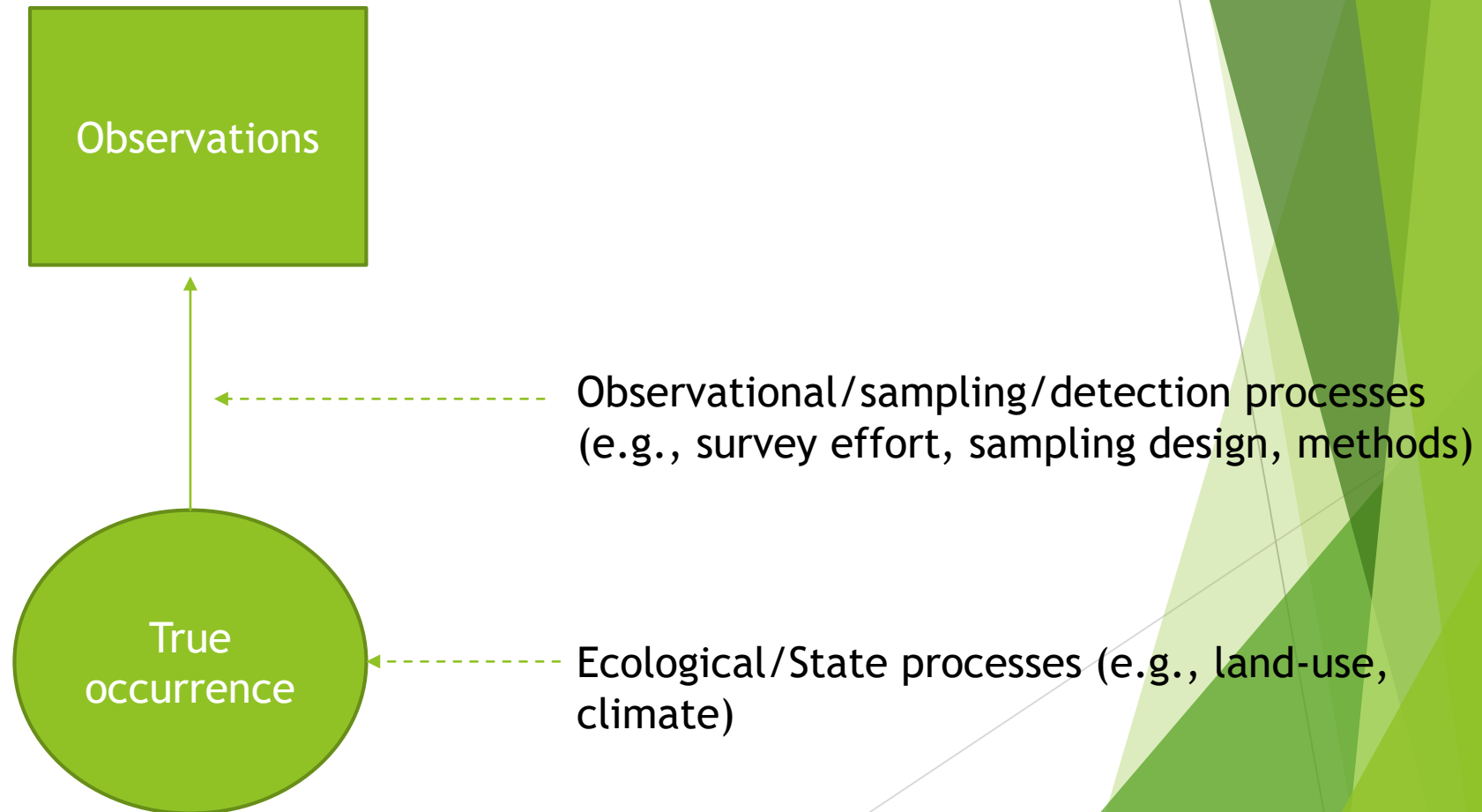
Principle

- ▶ Integrated models
 - ▶ Combine multiple datasets in the same analysis
 - ▶ Datasets are not simply merged!
 - ▶ We specify what information (i.e., parts of the linear predictor) are shared by each dataset
- ▶ Two main types:
 - ▶ Integrated distribution models
 - ▶ Integrated population models

Relevance to citizen science

- ▶ (1) Modelling of unstandardized citizen science data can be helped by included even a small amount of standardized data
 - ▶ Integrations helps factor out bias, e.g., spatial bias, in citizen science data
- ▶ (2) There are usually trade-offs in data - standardized data is more limited and unstandardized data is more widespread
 - ▶ Integration combines the strengths of both datasets
 - ▶ Upscaling

Hierarchical model: separation of processes generating data

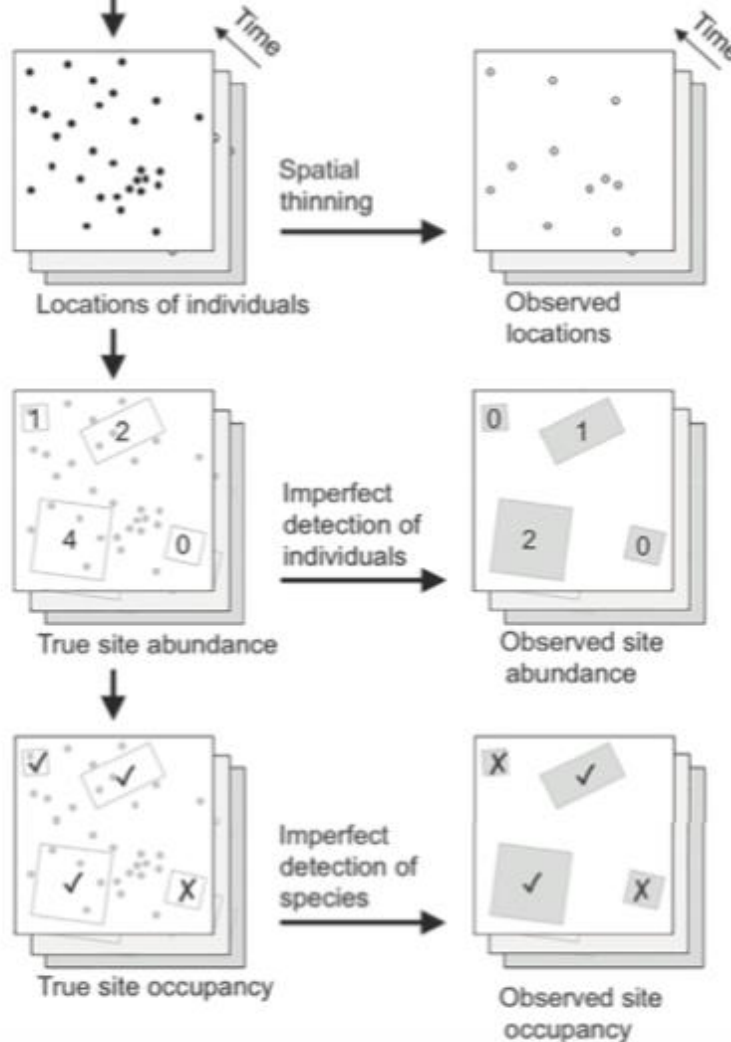


Connecting the information in different kinds of surveys

State

Observation

Ecological
processes are
constant

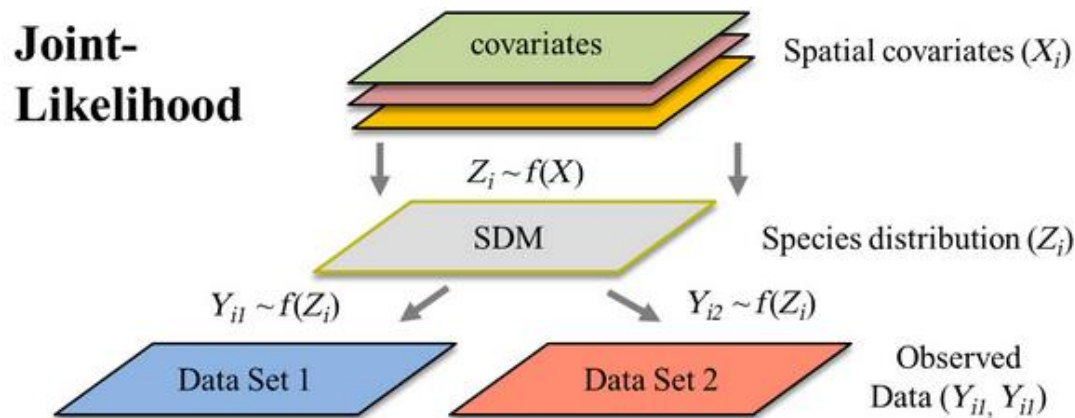


Only observational
processes change!!

Isaac et al. 2020

Integration using hierarchical models

- ▶ “Joint-likelihood” approach
 - ▶ Currently most common
 - ▶ Assume datasets share the same ecological/state process
 - ▶ Allow datasets to differ in observation processes



Miller et al.
2019

Examples

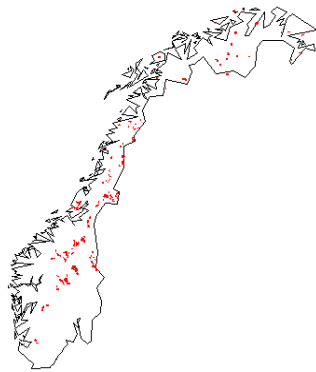
- ▶ Structured + unstructured data
- ▶ Abundance + occurrence data

Abundance + occurrence

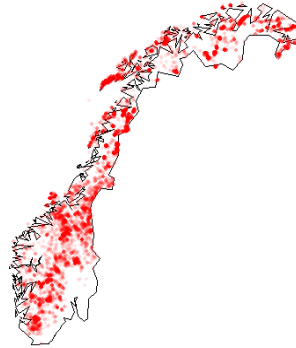
Willow ptarmigan in Norway



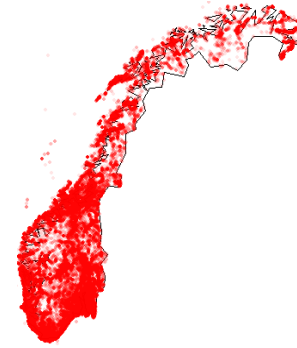
Combining abundance and occurrence data



**Standardized abundance
survey data along line
transects**

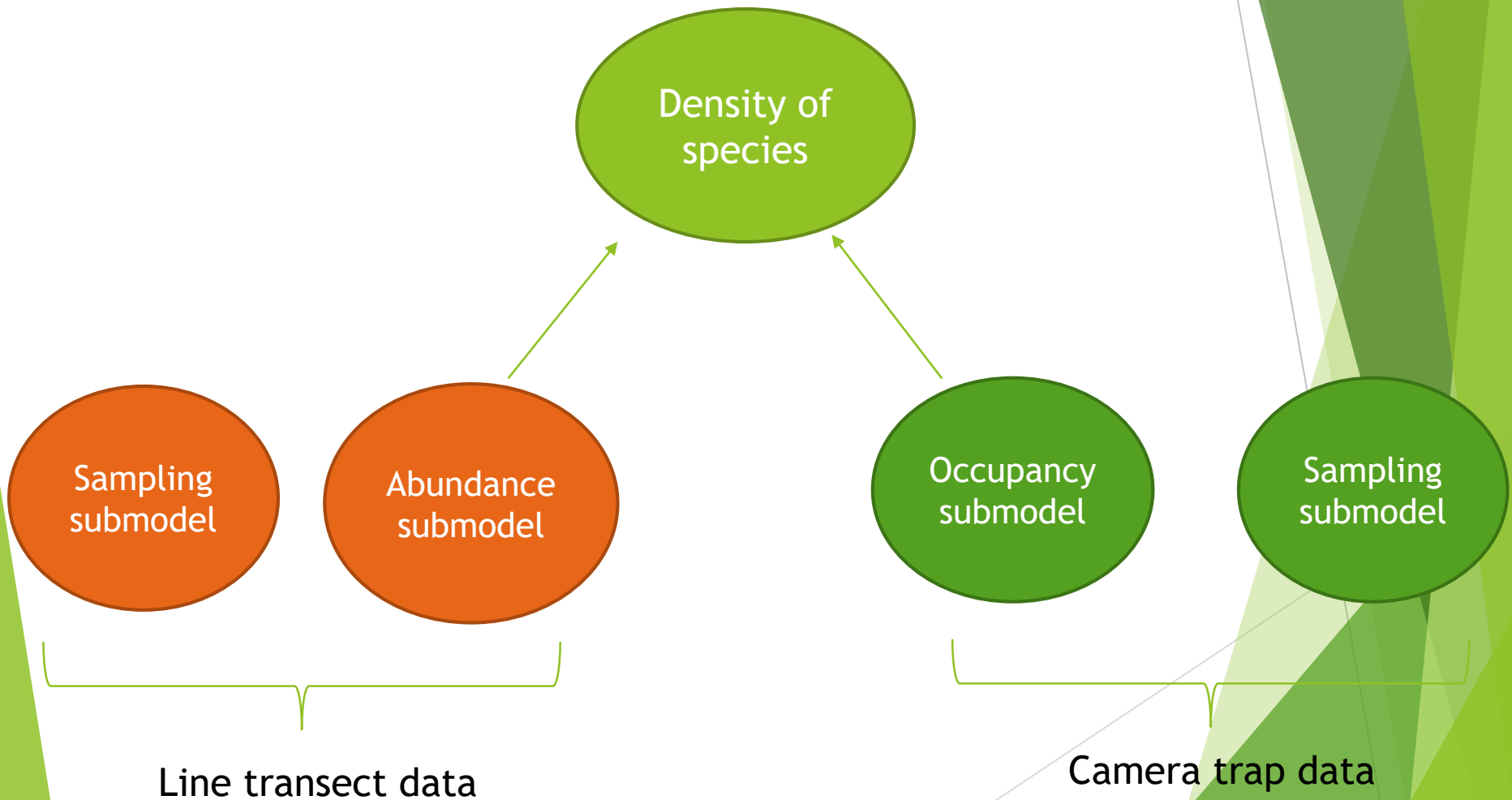


**Citizen science opportunistic
presence data**



**Citizen science total sampling
(absence data)**

Example: Integrated distribution model



Relationship between occurrence and abundance

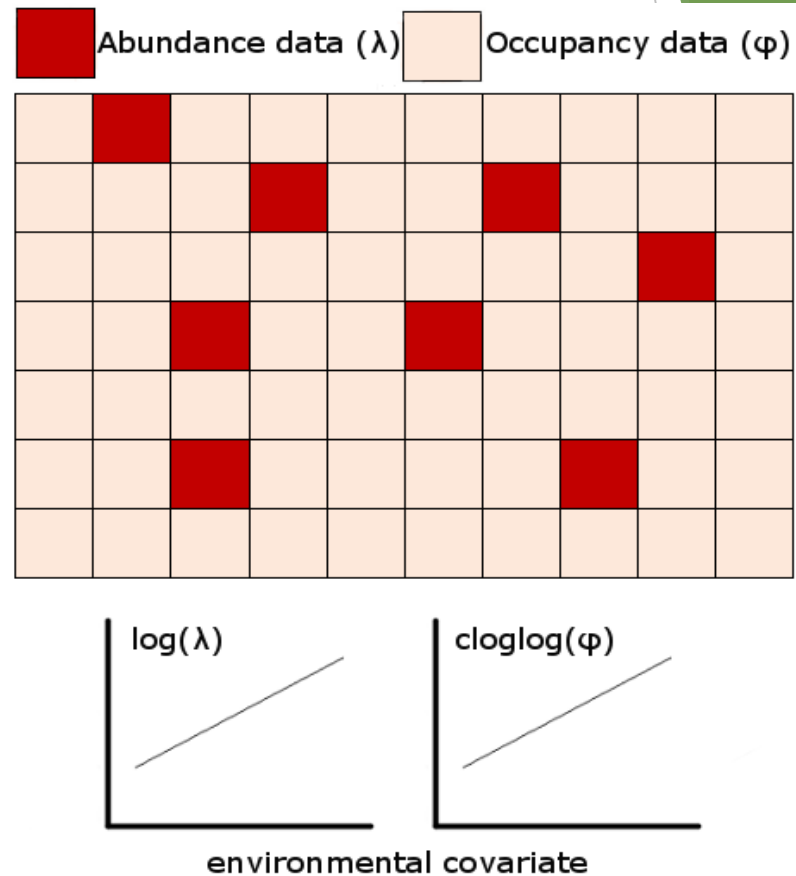
φ - Occurrence probability

λ is mean abundance

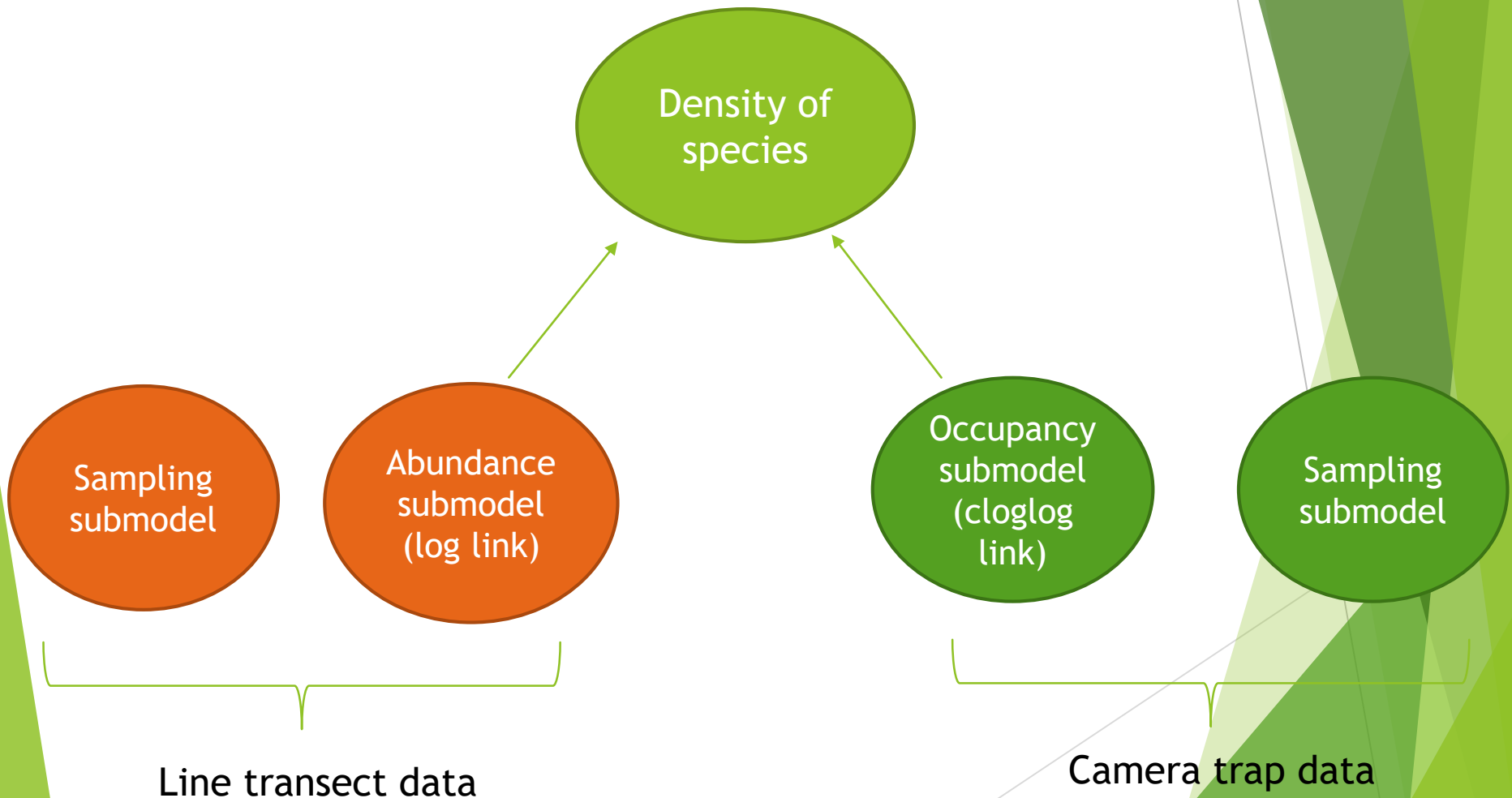
$$\varphi = 1 - \exp(-\lambda)$$

$$-\log(1 - \varphi) = \lambda$$

$$\underbrace{\log(-\log(1 - \varphi))}_{\text{cloglog}} = \underbrace{\log(\lambda)}_{\text{log}}$$

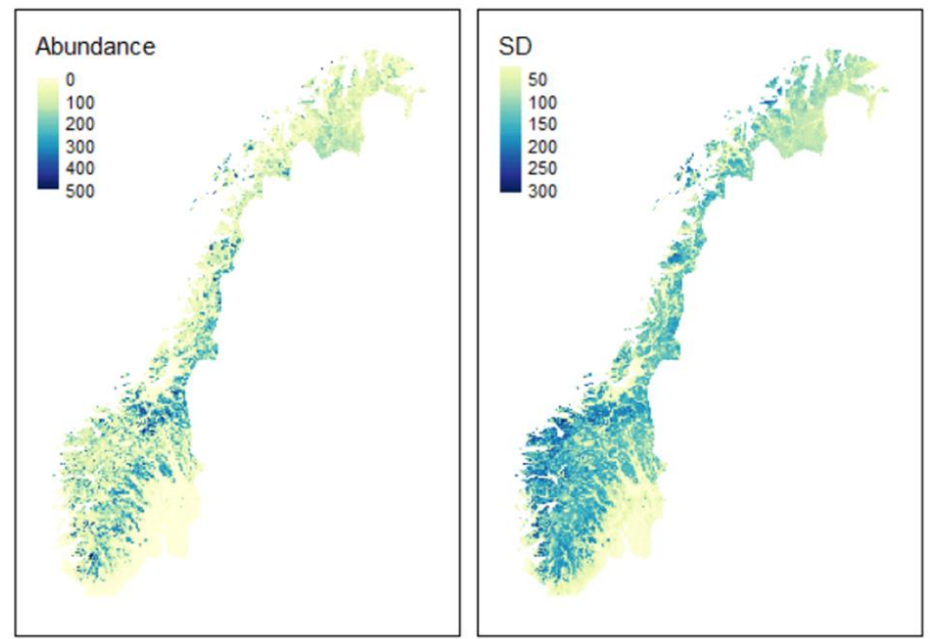


Example: Integrated distribution model



Integrated predictions

- ▶ Hierarchical model combining both data types
- ▶ Predictions of total abundance in Norway
- ▶ c. 1,000,000 individuals



The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern, layered effect on the right side of the slide.

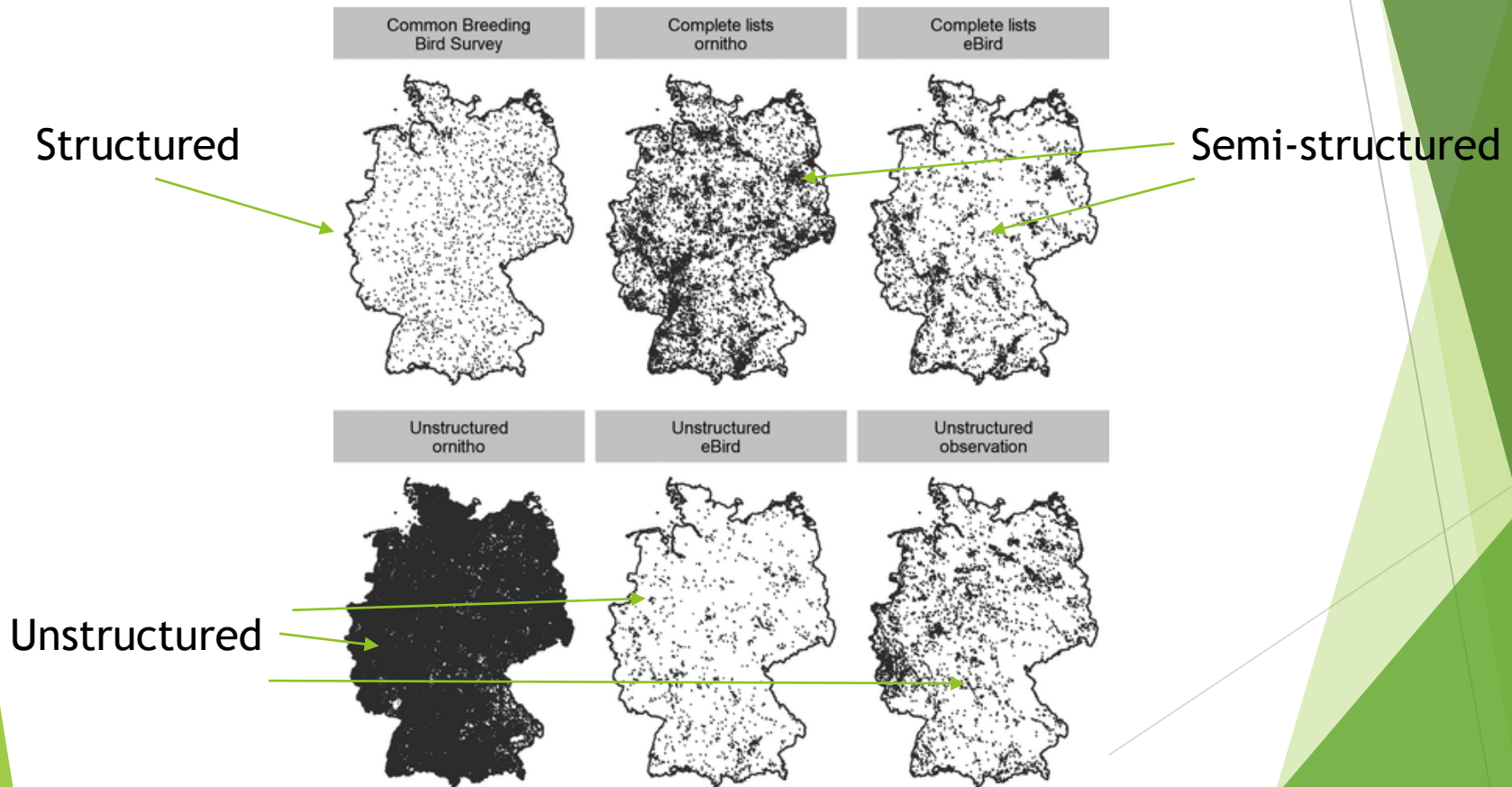
Structured + unstructured data

Integration of bird data in Germany

BIODIVERSITY RESEARCH | [Open Access](#) |  

Model-based integration of citizen science data from disparate sources increases the precision of bird population trends

Lionel R. Hertzog  Claudia Frank, Sebastian Klimek, Norbert Röder, Hannah G. S. Böhner, Johannes Kamp



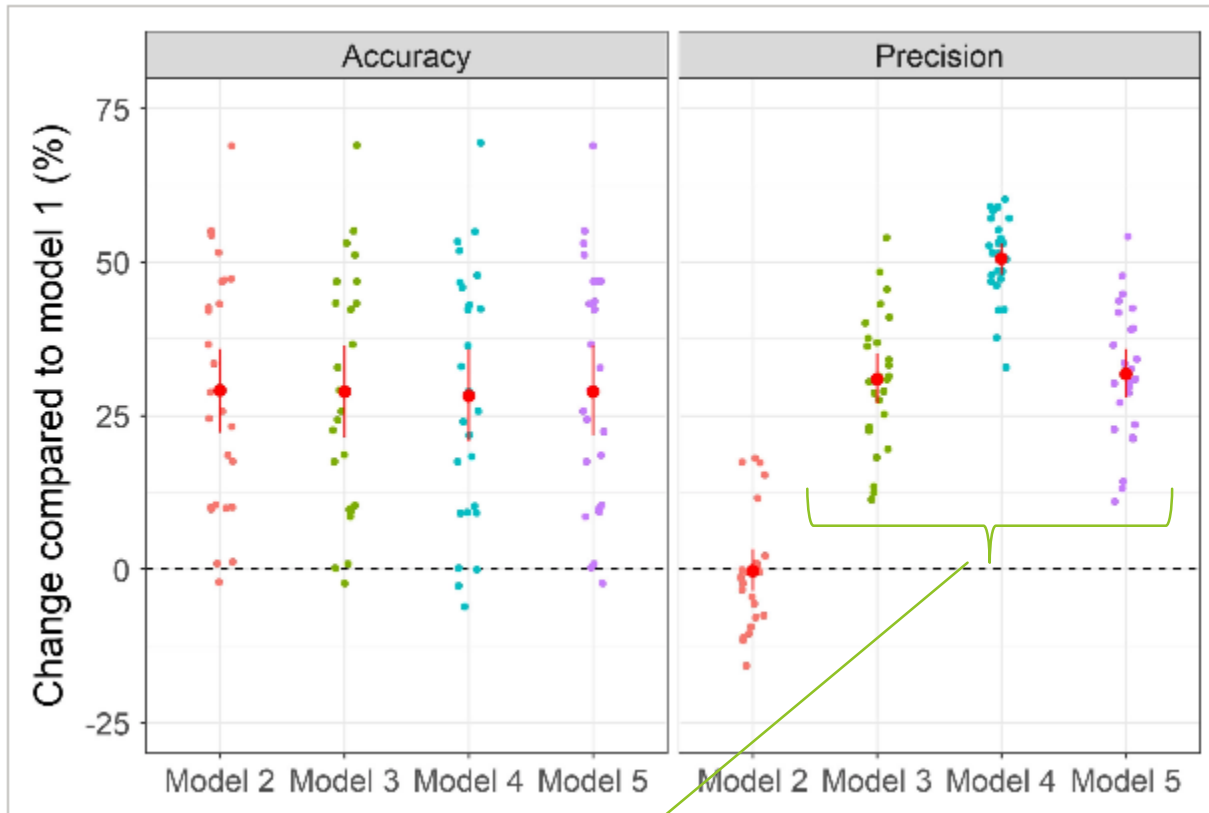
Model

Dataset	CBBS	Ornitho		eBird		Observation
Dataset type	Structured	Semi-structured (complete checklists)	Unstructured	Semi-structured (complete checklists)	Unstructured	Unstructured
Model 1 (TRIM)	x					
Model 2	x					
Model 3	x	x				
Model 4	x	x	x			
Model 5	x	x		x		
Model 6	x	x	x	x	x	x

In the integrated models (i.e., models with multiple datasets)

- State process: Yearly changes were jointly modelled across the datasets
- Detection/Sampling process: Intercepts were allowed to differ across the datasets. Some models included an offset term to account for sampling effort variation.

Results



- Increased precision (i.e., reduced uncertainty) by model integration

Simulation Experiments: Why and when is combining data useful?

Population estimates are narrower with data integration

Benefit decreases as the amount of high-quality data increases

i.e., if we have lots of abundance data, adding in occurrence data probably does not affect much

If we have lots of standardized data, there is probably less benefit to including unstructured data.

Most of the time in ecology, we don't have lots of abundance data or standardized data

Why and when is combining data useful?

A major benefit of integration is greater sampling of the environment range



**Standardized
abundance survey
data along line
transects**

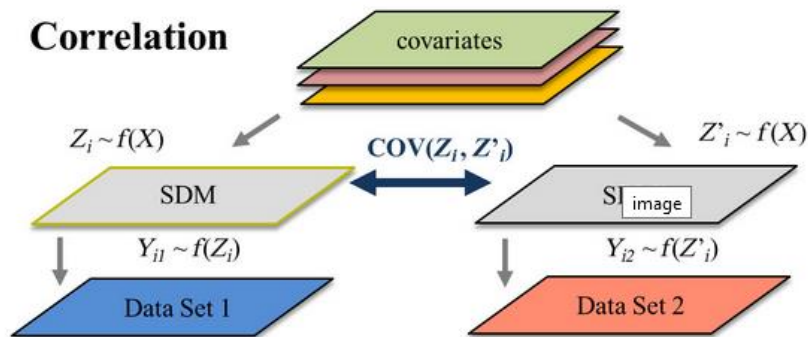


**Citizen science
opportunistic
presence data**



**Citizen science
total sampling
(absence data)**

Integrated models: alternatives to joint-likelihood



Miller et al. 2019

