Statistical tools for the analysis of species' population trends

Diana Bowler

German Centre for Integrative Biodiversity Research

LTER-D 10th March 2020

sMon | Biodiversitätstrends in Deutschland

- Few large-scale standardized monitoring schemes
- But large-scale trends are important for conservation policy
- Despite the lack of standardized data, there are large of amount of opportunistic and semi-structure data
  - Natural history societies
  - Skilled natural historians
  - Casual citizen scientists
- How can we make use of the opportunistic data that is available?

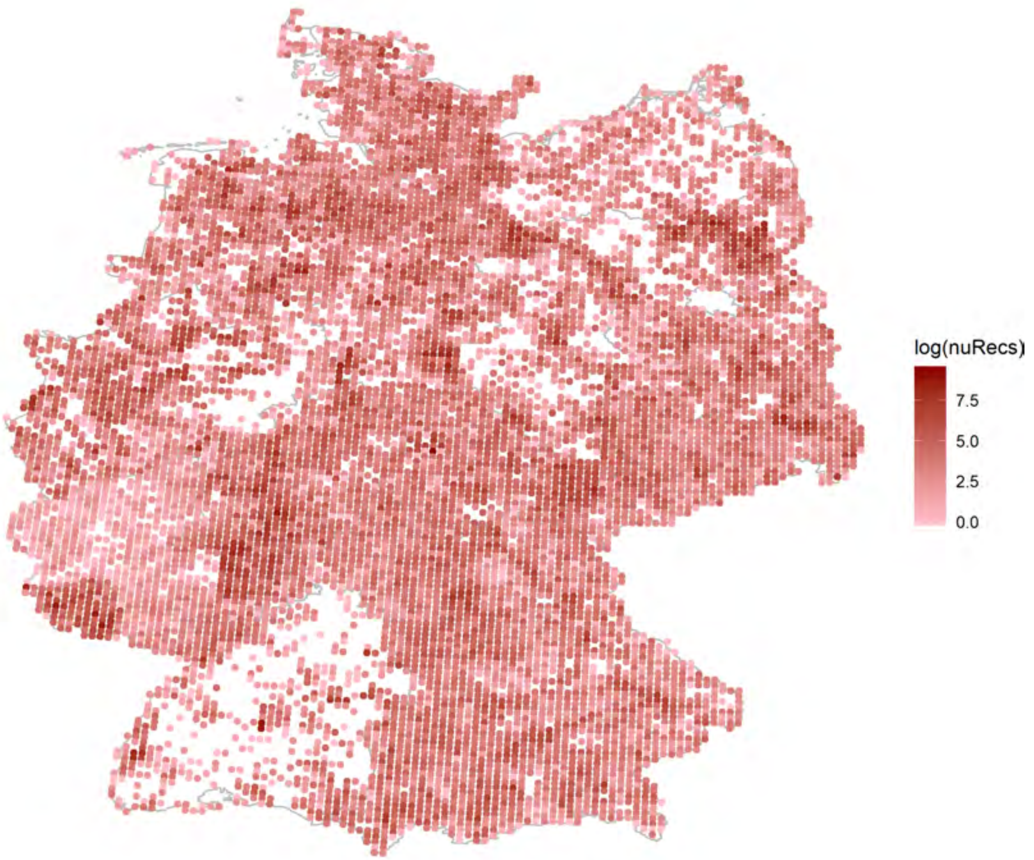
Making use of opportunistic data

Opportunistische Daten

Taxon	DATUM_von	DATUM_bis	Länge	Breite	Toleranz	Beobachter	X.METHODE.	Bodenfalle	Autokescher
Cicindela sylvatica L., 1758	30.09.1902		10.288095	53.31201	2500	298			
Cicindela sylvatica L., 1758	15.08.1905		9.911728	53.46148	1000	374			
Cicindela sylvatica L., 1758	27.06.1909		10.079441	53.35690	1000	256			
Cicindela sylvatica L., 1758	06.09.1909		10.826340	53.11999	5000	27			
Cicindela sylvatica L., 1758	15.05.1910		9.473562	53.59541	2500	299			
Cicindela sylvatica L., 1758	15.06.1910		9.473562	53.59541	2500	299			
Cicindela sylvatica L., 1758	15.07.1910		10.012665	53.25761	2500	256			
Cicindela sylvatica L., 1758	01.09.1912		9.876451	53.30072	2500	27			
Cicindela sylvatica L., 1758	17.04.1916		10.785422	53.84138	2500	177			
Cicindela sylvatica L., 1758	15.06.1925		10.744200	53.86969	1000	177			
Cicindela sylvatica L., 1758	01.07.1928		10.159950	53.64993	2500	310			
Cicindela sylvatica L., 1758	14.06.1932		9.857612	53.47696	2500	204			
Cicindela sylvatica L., 1758	15.08.1935		8.620148	53.86042	2500	204			
Cicindela sylvatica L., 1758	06.06.1936		9.757651	53.58347	1000	336			
Cicindela sylvatica L., 1758	15.08.1936		9.592767	52.86158	2500	204			
Cicindela sylvatica L., 1758	15.06.1944		9.744787	53.32170	1000	97			
Cicindela sylvatica L., 1758	15.07.1944		9.747169	53.32212	500	97			
Cicindela sylvatica L., 1758	01.07.1945		9.105606	53.69803	1000	2			
Cicindela sylvatica L., 1758	25.04.1946		9.714661	53.57773	2500	1			
Cicindela sylvatica L., 1758	27.04.1946		9.757651	53.58347	1000	97			
Cicindela sylvatica L., 1758	27.04.1946		9.674835	53.62077	2500	97			
Cicindela sylvatica L., 1758	27.04.1946		9.757651	53.58347	1000	97			
Cicindela sylvatica L., 1758	24.07.1946		9.912041	53.14848	1000	310			
Cicindela sylvatica L., 1758	20.07.1947		10.556746	52.96658	2500	2			
Cicindela sylvatica L., 1758	24.07.1947		10.680943	53.57375	2500	283			
Cicindela sylvatica L., 1758	14.08.1947		10.556746	52.96658	2500	2			
Cicindela sylvatica L., 1758	20.06.1948		9.956961	53.30883	2500	91			
Cicindela sylvatica L., 1758	16.07.1949		9.842377	53.45806	1000	191			
Cicindela sylvatica L., 1758	08.07.1950		9.842377	53.45806	1000	191			

Odonata data set

Over 1 mill occurrence records



Analyse-Methoden

Methods in Ecology and Evolution

Methods in Ecology and Evolution 2014, **5**, 1052–1060 doi: 10.1111/2041-21

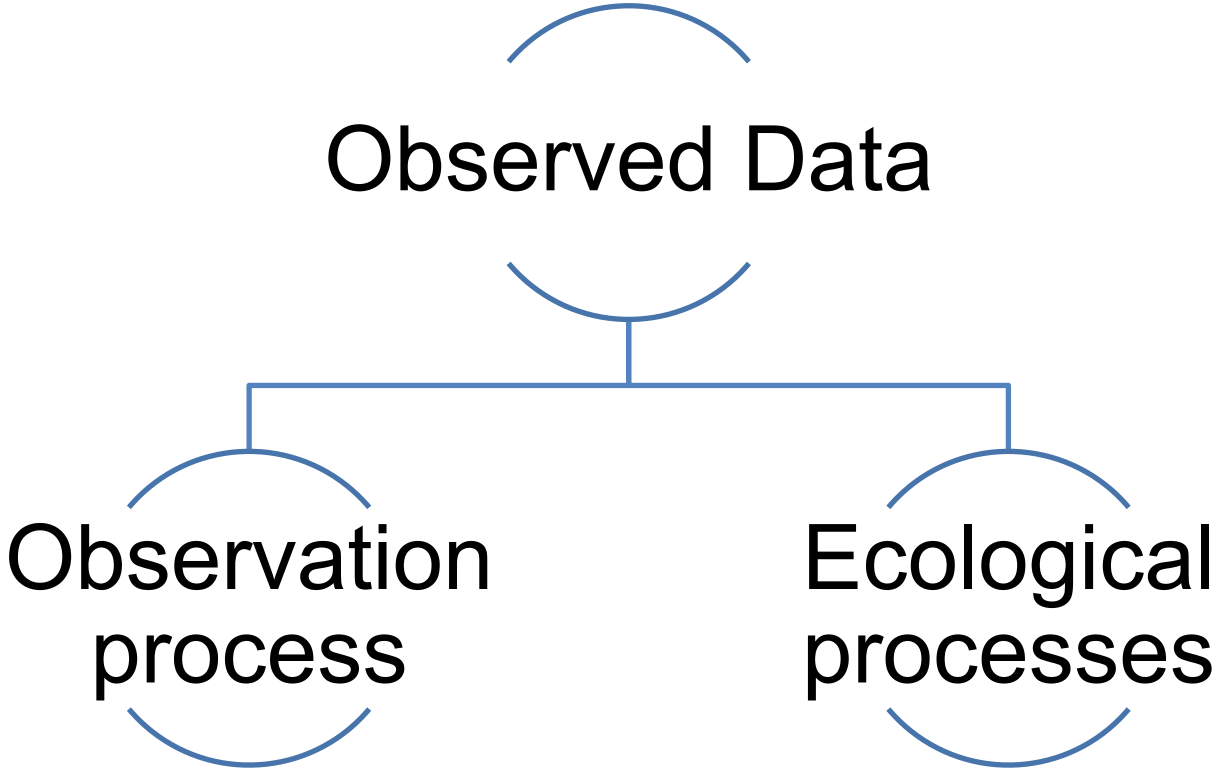
Statistics for citizen science: extracting signals of change from noisy ecological data

Nick J. B. Isaac[1]*, Arco J. van Strien[2], Tom A. August[1], Marnix P. de Zeeuw[2] and David B.

[1]*NERC Centre for Ecology & Hydrology, Crowmarsh Gifford, Maclean Building, Wallingford, OX10 8BB, UK; and* [2]*St Netherlands, PO Box 24500, 2490 HA The Hague, The Netherlands*

- Occupancy-Detection Models
- Considers:
  - Imperfect Detection
  - Sampling variation

Occupancy-Detection Model is a Hierarchical model



Core equation of occupancy-detection model

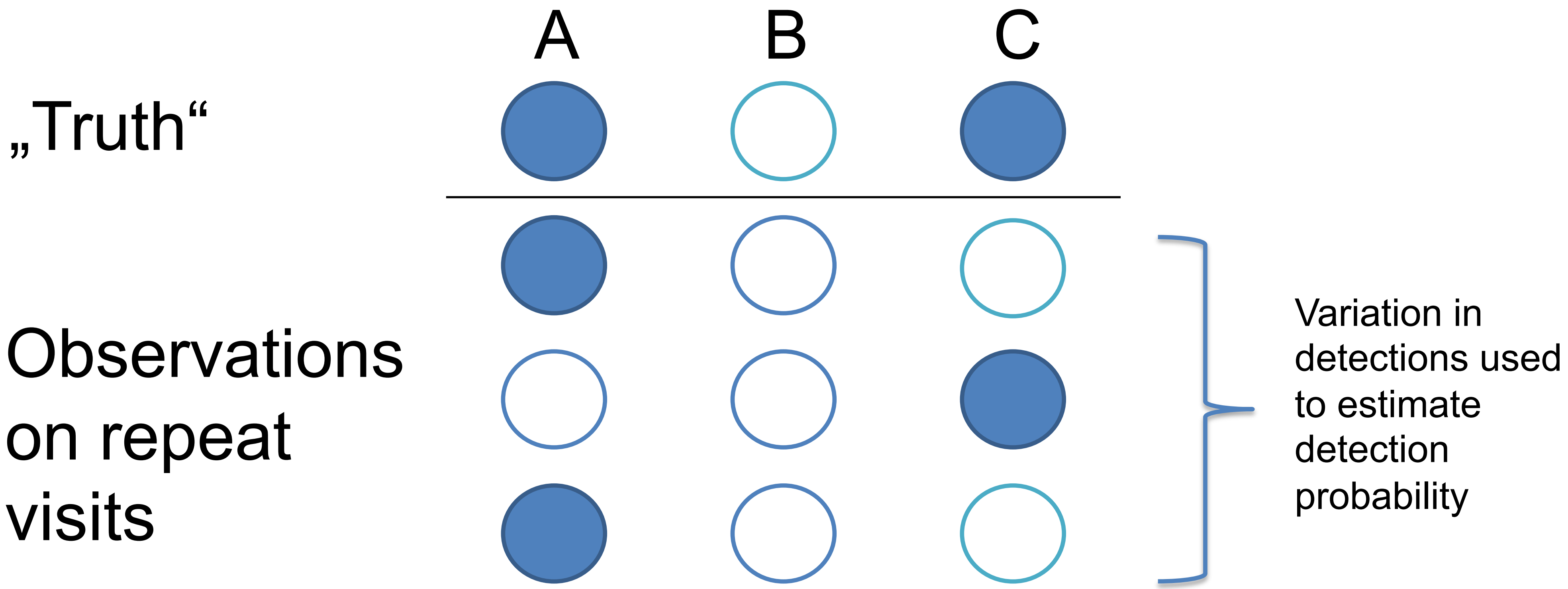
Reality

Detection
probability

Observations

Observation processes: estimation of detection probability

- Definition of a visit = same observer visits same site on same date
- Repeated visits to the same site within the flight period



Dynamic occupancy models

- Ecological model:

```
z[i,t] <- persist[i,t-1]*z[i,t-1] + colonize[i,t-1]*(1-z[i,t-1])
```

Dynamic occupancy models

- Ecological model:

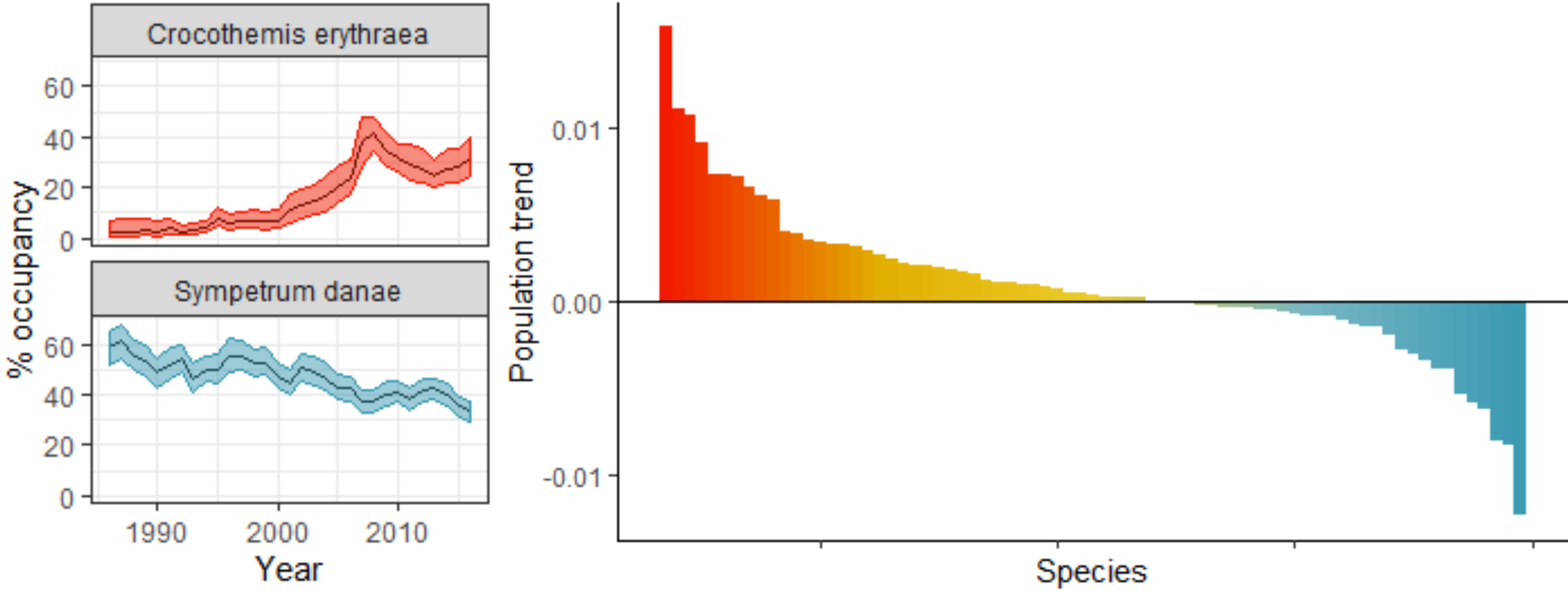
```
z[i,t] <- prob.persist[i,t-1]*z[i,t-1] + prob.colonize[i,t-1]*(1-z[i,t-1])
```

- Observation model:

  Probability of detection – varies by site, year, date and listlength (single list, log list length)

```
logit(p[j]) <-  mup[year[j]] + lp[craum[j]] +lp.r[raum[j]] +
                    phenol.s[craum[j]] * yday[j] +
                    phenol2.s[craum[j]] * pow(yday[j],2)+
                    effort.p * shortList[j] + single.p * singleList[j]
```

Population trends

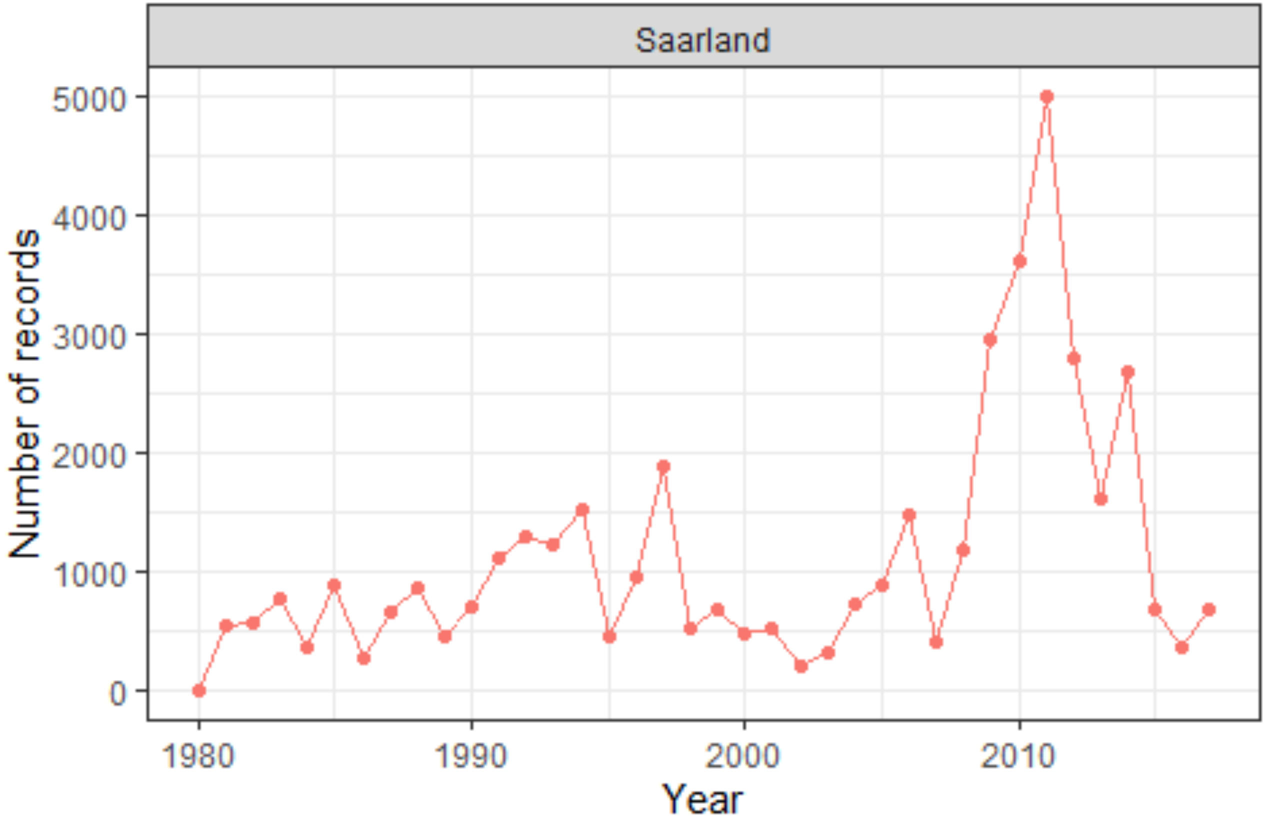


Checking robustness of occupancy models

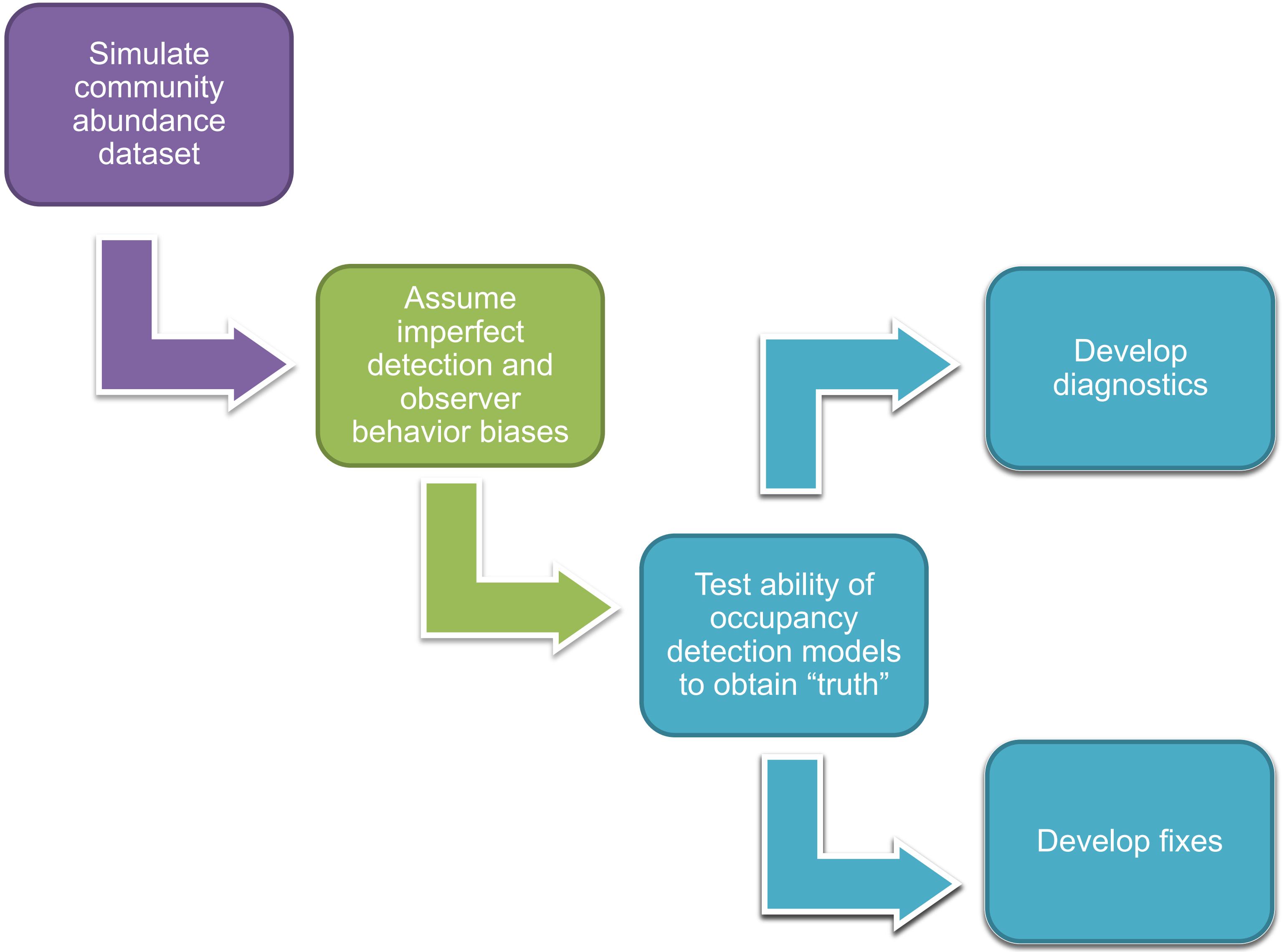
Simulation experiments to model citizen scientist behaviours….

- Assumption of OD model:

  No unmodeled heterogeneity in detection probabilities



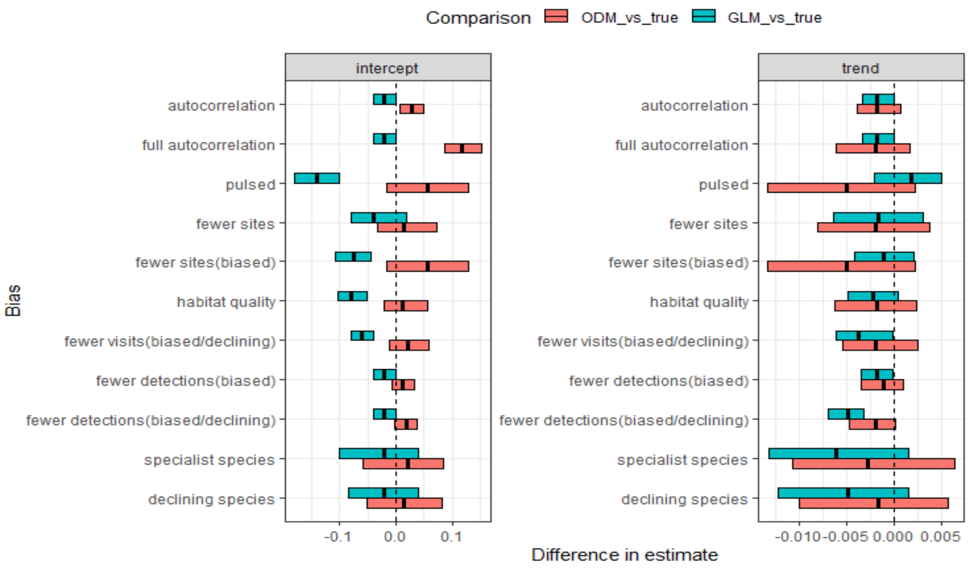
Simulations to test the robustness of occupancy detection models



CS behavior scenarios for the simulations

Serial dependence	If seen on last visit, less likely to be reported next time
	Once reported within a season, not reported again
Atlas schemes	Pulsed activity (more visits to each grid) in a given year
	Pulsed activity (increased number of grids visited – at random) in a given year
	Pulsed activity (more grids but extension into lower quality grid cells) in a given year
“Car park” effect	Lower quality sites less likely to be visited
	Lower quality sites are visited for short times (lower detection prob)
	Fewer visits/lower detection probability as sites declining in habitat quality
	Accessibility effects
Project type	Known change in project type (5 years – project 1, 5 years –project 2)
	Unknown change in project types/data mixes (e.g, GBIF type) New method innovation/new guide effects/binoculars
Observer species preferences	Some observers report rare species, Other observers report all species
	Observations of declining species are more likely to be reported
	Site more likely to be visited if focal species seen there previously

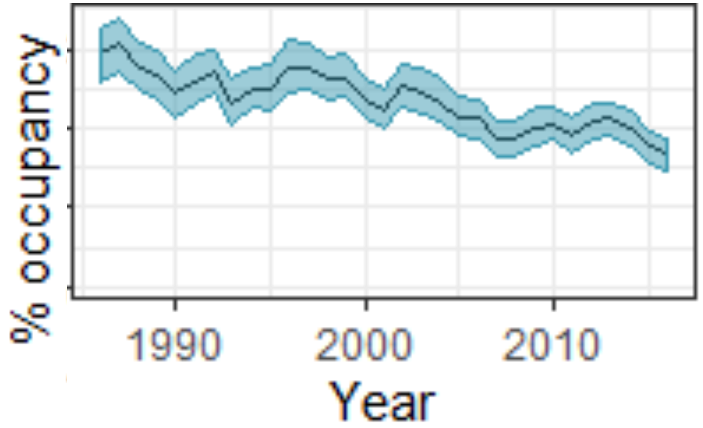
Prelim results



Impression so far:

Intercept affected by CS behaviors

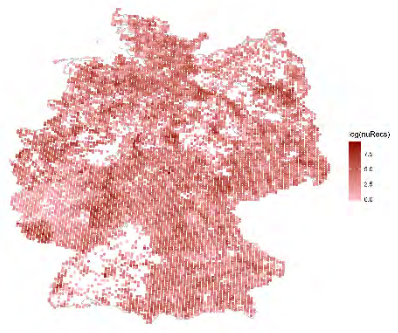
Slope (trend estimated) less affected



Extensions to the basic occupancy-detection model possible and already developed for capture-recapture e.g., “trap happy”

But what citizen scientist behaviours are even common?

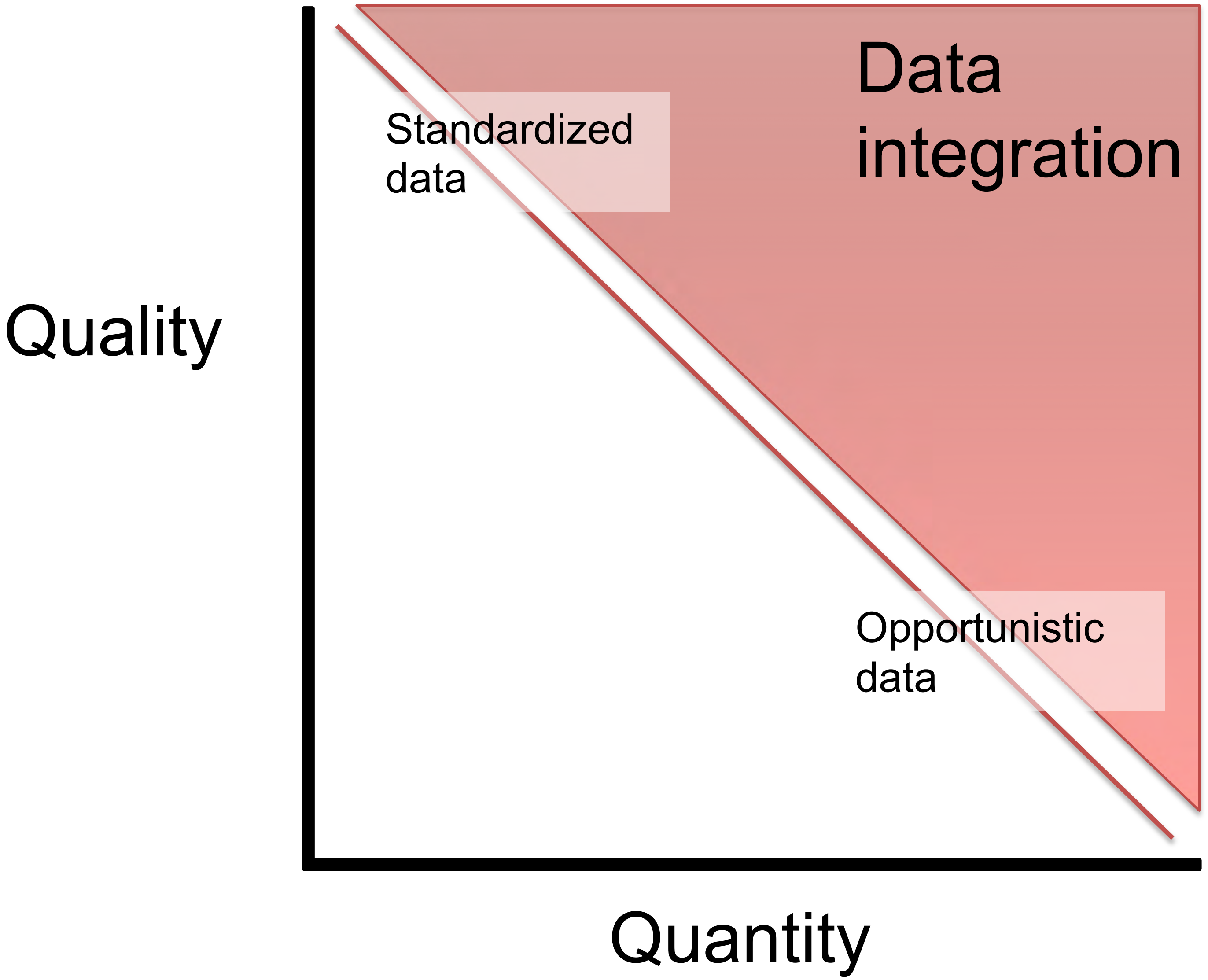
- Analysis of spatial bias in opportunistic data in Germany (dragonflies) underway
  - What biases are most common?



- Questionnaire being developed with GESIS
  - Ask questions to better understand decision-making of citizen scientists
    - What they report?
    - Where and when they sample?
    - How long they spend surveying?

Extending the data and modelling framework to combine different data types

Trade-offs in data



Integrated model as the way forward?

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

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

esa

Search

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

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/go/ecologyjournal

Trends in Ecology & Evolution

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

CellPress

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. Dambly [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]

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Technological Advances at the Interface between Ecology and Statistics | Free Access

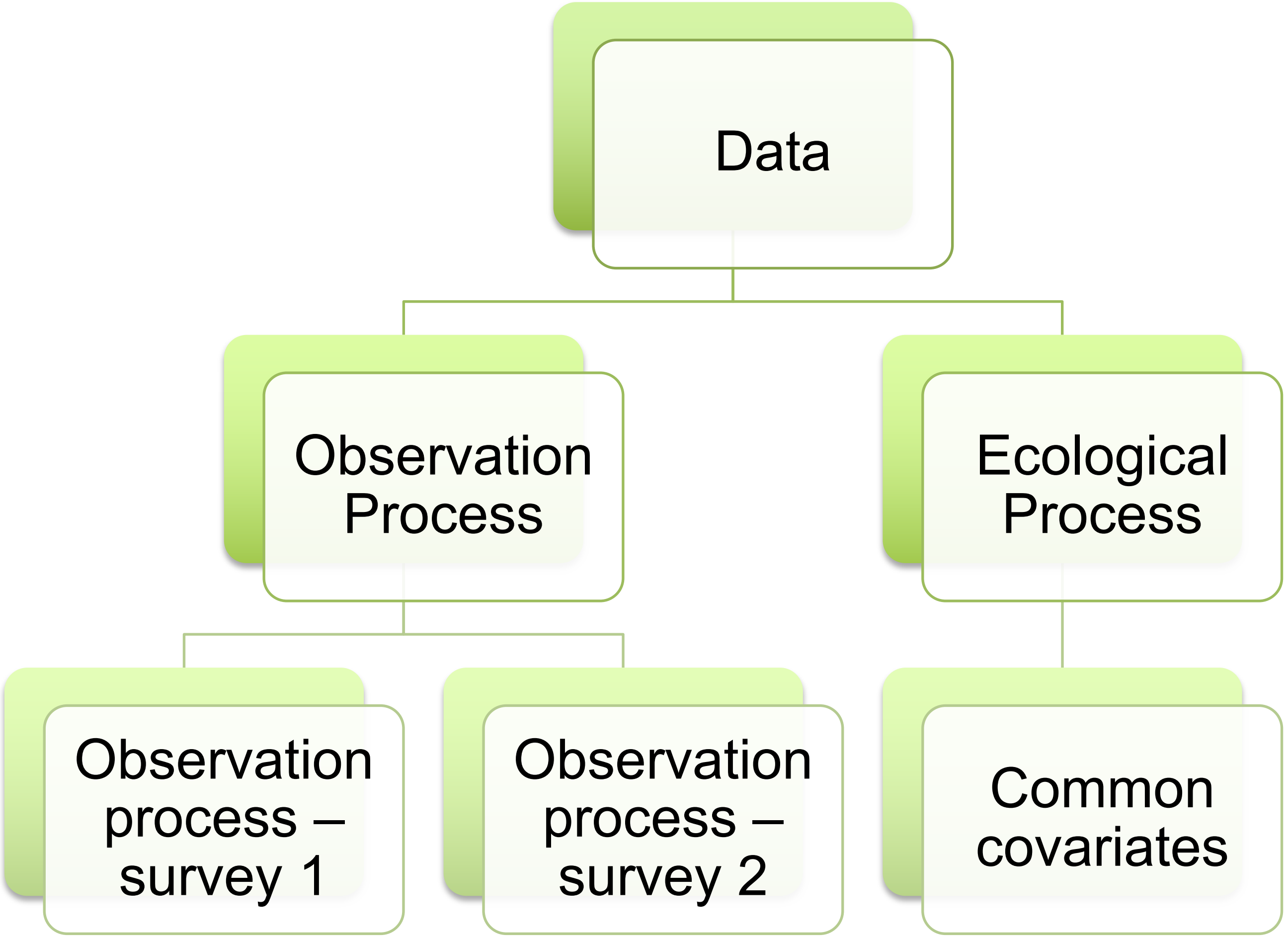
Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection

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]

Hierarchical models:



Example 1: Eld’s deer in Myanmar

SCIENTIFIC REPORTS

Article | Open Access | Published: 23 May 2019

Integrating data from different survey types for population monitoring of an endangered species: the case of the Eld’s deer

Diana E. Bowler, Erlend B. Nilsen, Richard Bischof, Robert B. O’Hara, Thin Thin Yu, Tun Oo, Myint Aung & John D. C. Linnell

Norwegian Institute for Nature Research

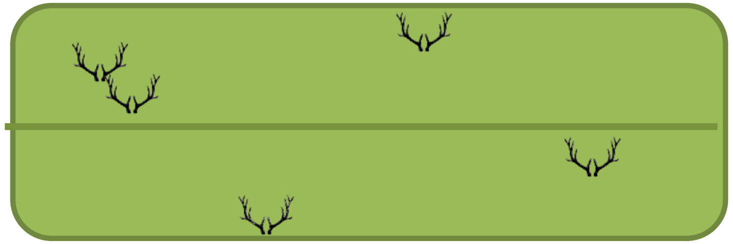
Eld's deer monitoring

Eld's deer Monitoring

Camera trapping

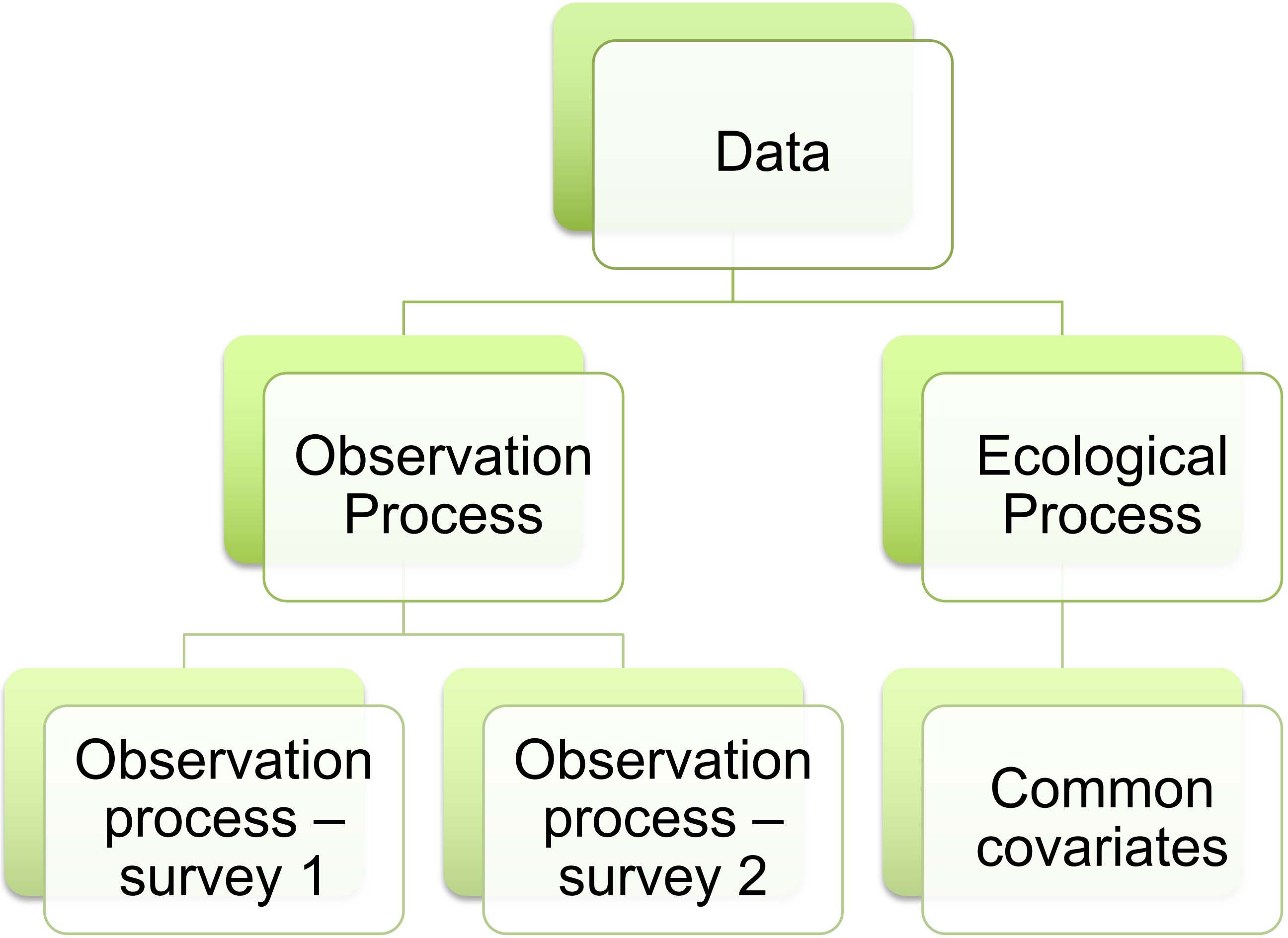
Line transect

Camera trapping at 160 stations in 2014 and 206



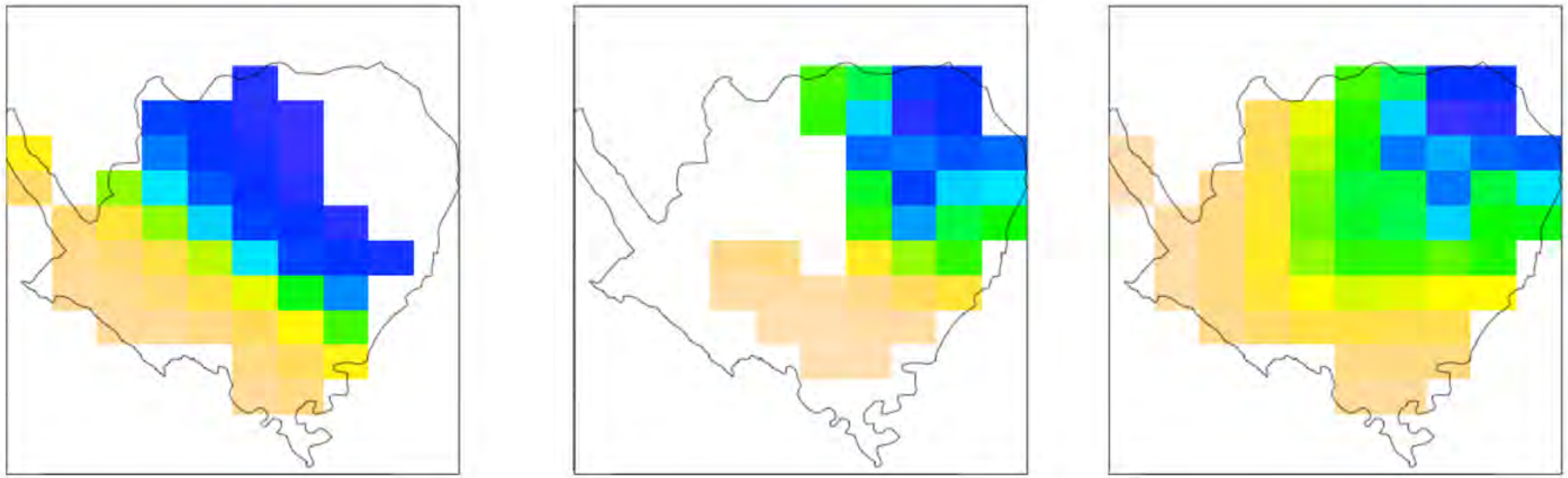
Distance-sampling along 25 line transects (1998 onwards)

Hierarchical models:



Model predictions

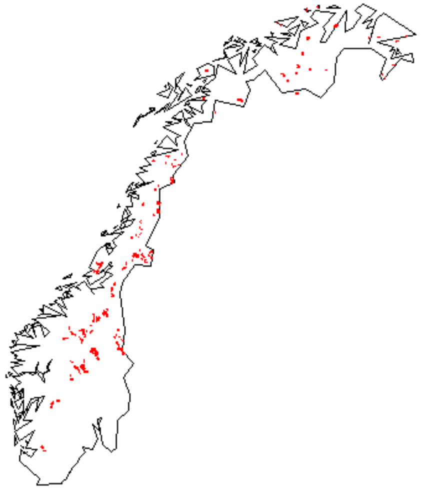
Predictions based on spatial spline (to region covered by each survey)

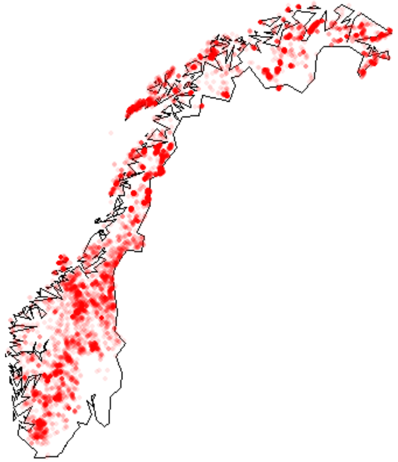


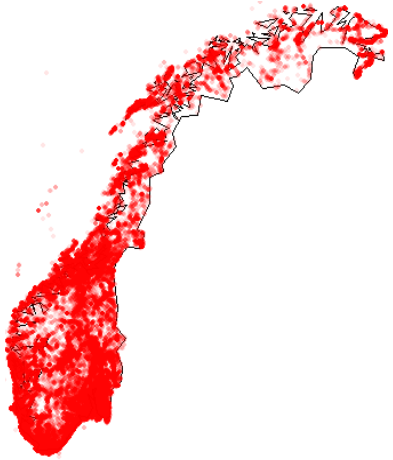
Example 2: Willow ptarmigan in Norway

NINA
Norwegian Institute for Nature Research

Combining abundance and presence data





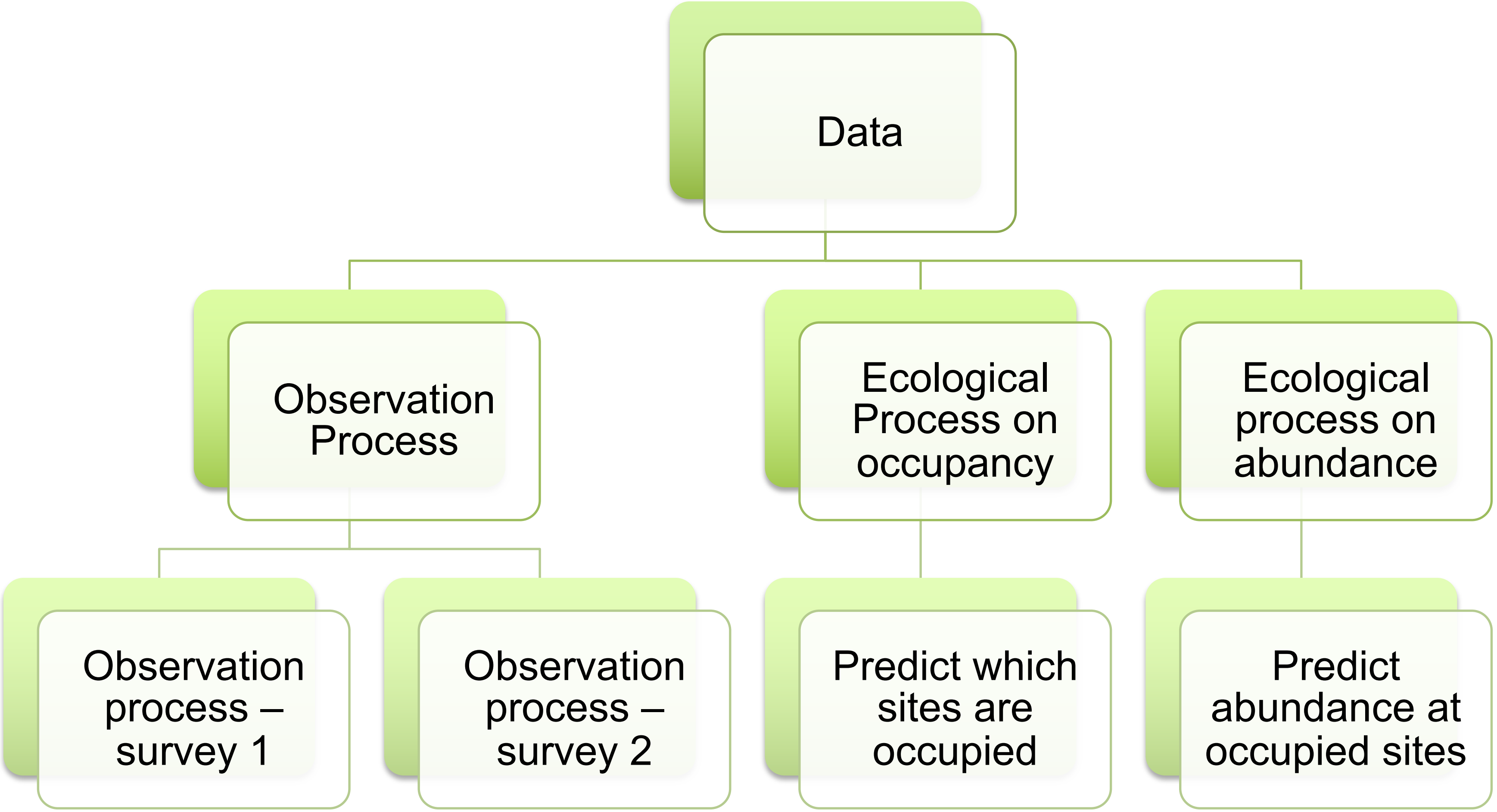


Standardized abundance survey data along line transects

Citizen science opportunistic presence data

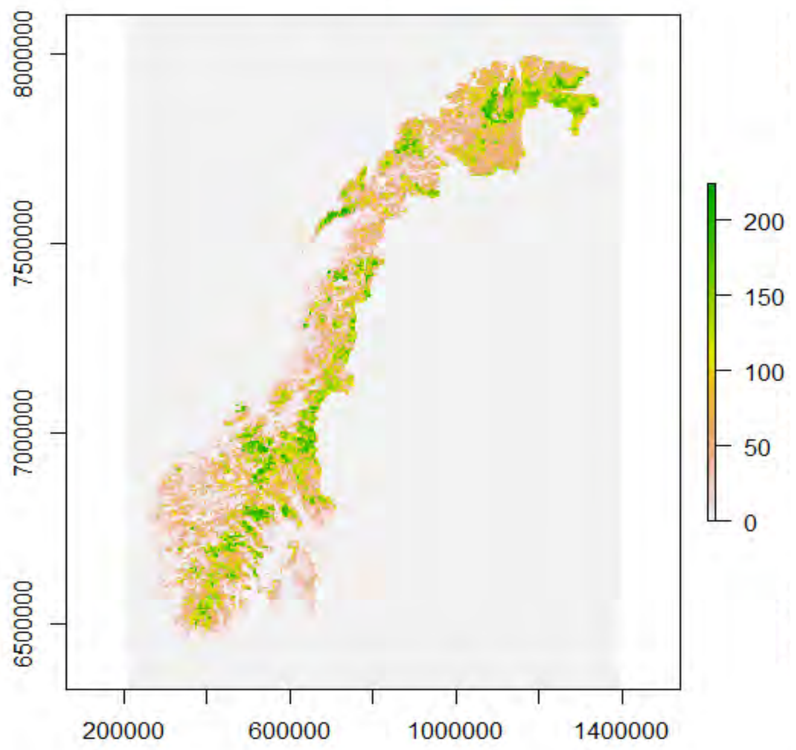
Citizen science total sampling (absence data)

Hierarchical models:

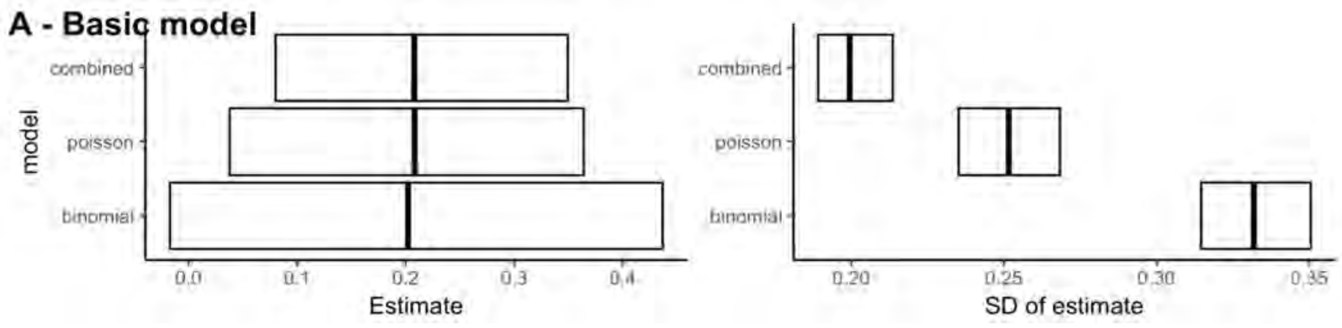


Combining abundance and presence data

- Hierarchical model combining both data types
- Predictions of total abundance in Norway
- c. 900,000 individuals



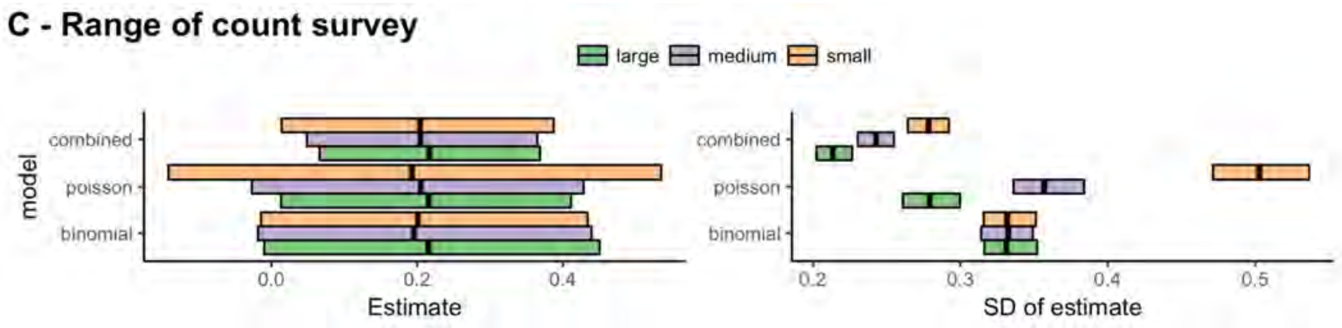
Simulation Experiments: Why and when is combining data useful?



Population estimates are narrower with data integration

Benefit decreases as the amount of high-quality data increases

Why and when is combining data useful?



A 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)

Why and when is combining data useful?


Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2014)

MACROECOLOGICAL METHODS

Accounting for imperfect detection and survey bias in statistical analysis of presence-only data

Robert M. Dorazio*

Standardized data can help factor out the bias in unstandardized data

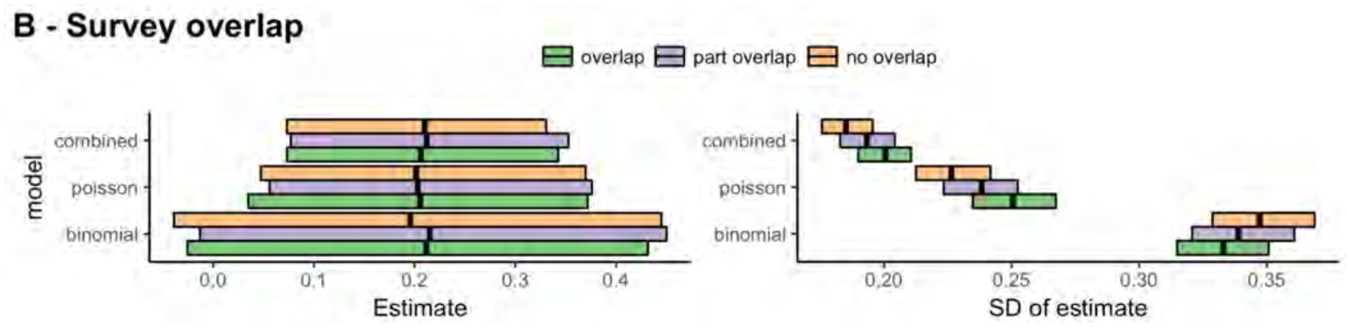
Dorazio, 2014: "Using mathematical proof and simulation-based comparisons, I demonstrate that biases induced by errors in detection or biased selection of survey locations can be reduced or eliminated by using the hierarchical model to analyse presence-only data in conjunction with counts observed in planned surveys"

Outlook

- We can get sensible results from careful analysis of opportunistic data
- In many scenarios of data availability, there can be a benefit to data integration
- Main cost is the time spent figuring out the best way!
- Might data integration models be a tool to upscale LTER data or other local standardized data?

Thank You !

Why and when is combining data useful?



Surveys don't need to spatially overlap for there to be a benefit to integration

If surveys are far apart – have to think about whether safe to assume same ecological processes at play

Combining data: Relationship between occurrence and abundance

φ is Occurrence probability

λ is Expected abundance

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

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

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