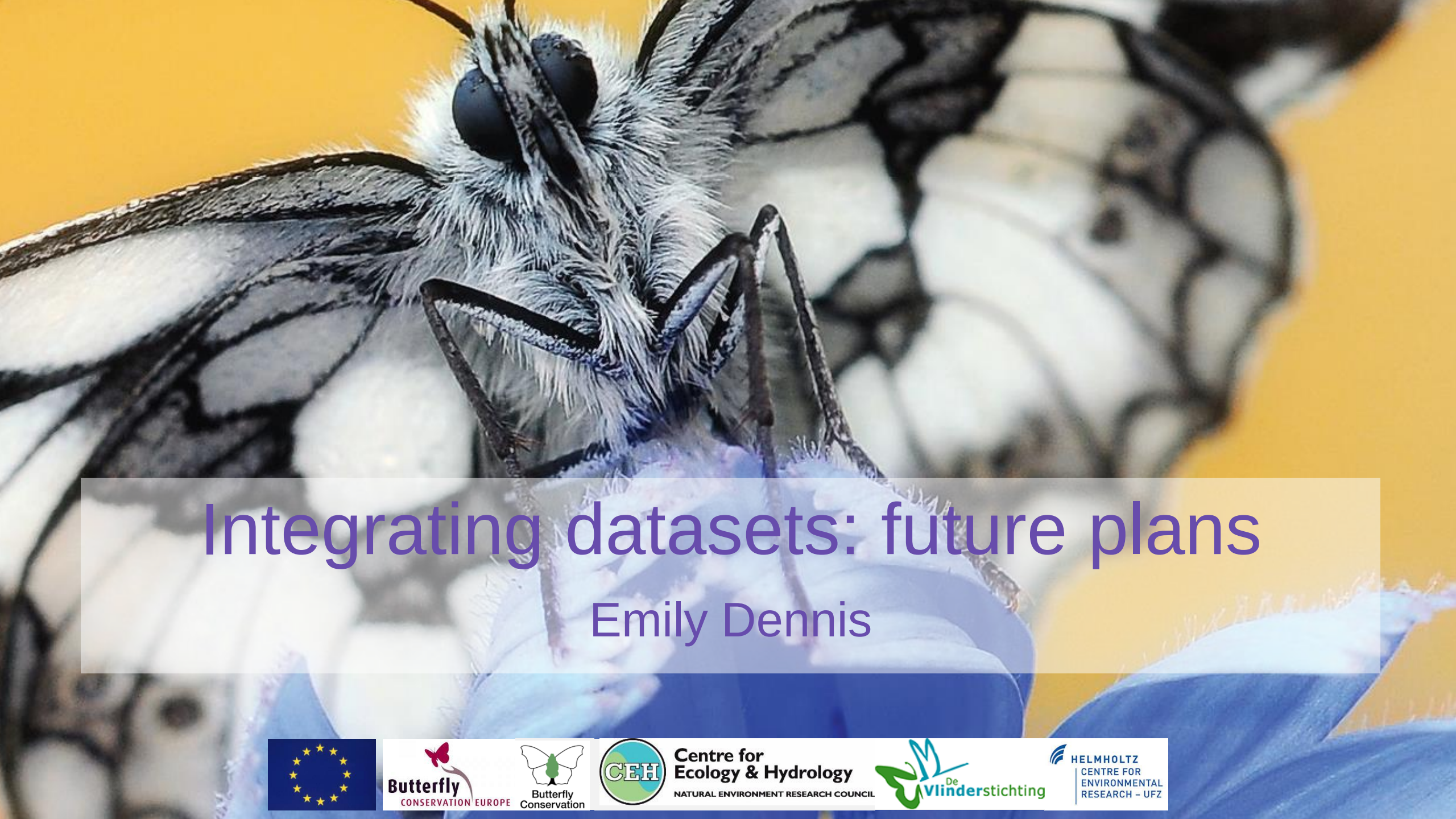


Integrating datasets: future plans

Emily Dennis

Butterfly Conservation Europe
Butterfly Conservation
CEH Centre for Ecology & Hydrology
Natural Environment Research Council
De Vlinderstichting
Helmholtz Centre for Environmental Research – UFZ

Ways of monitoring butterflies

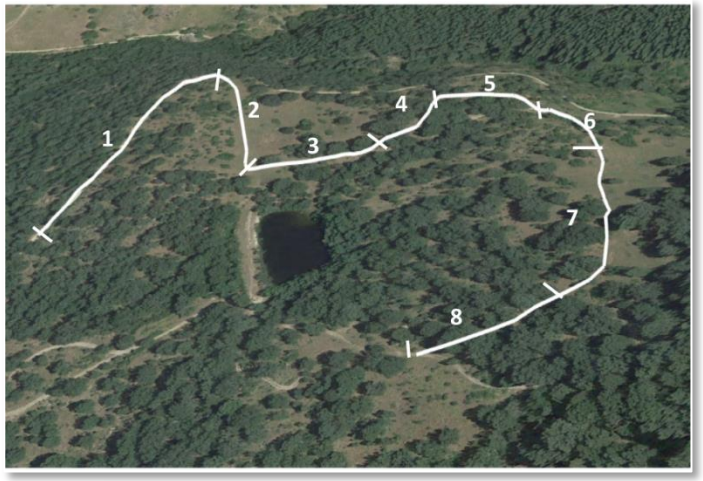
- Butterfly Transects (e.g. Butterfly Monitoring Schemes)
- Timed Counts
- Sightings data (e.g. Atlas maps, iNaturalist, Observation.org)

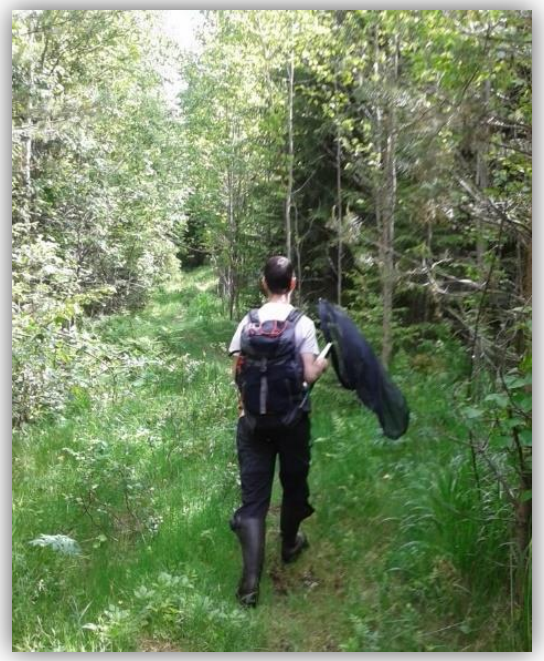
- How can we integrate datasets?
- Which data is best to collect?

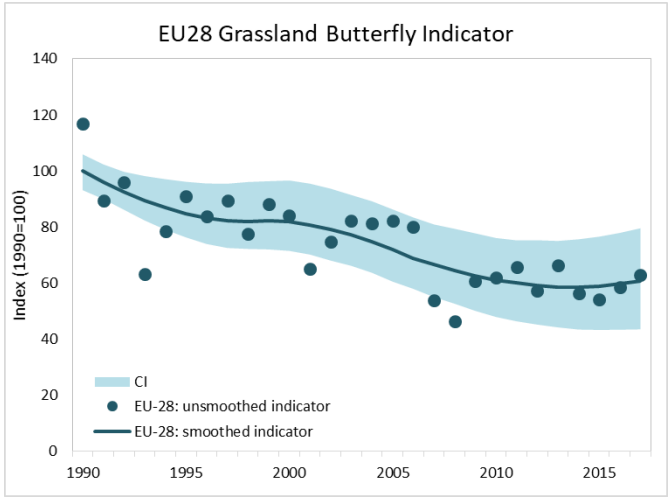
Approaches to butterfly monitoring

Transects

- The “gold” standard
- Standardized: regular commitment, fixed route
- Involves expert citizen science
- Robust results: Annual indices of relative abundance. Sensitive measure of change







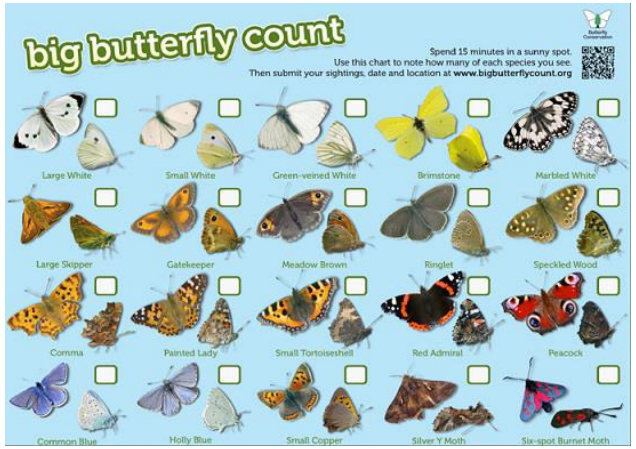
Approaches to butterfly monitoring

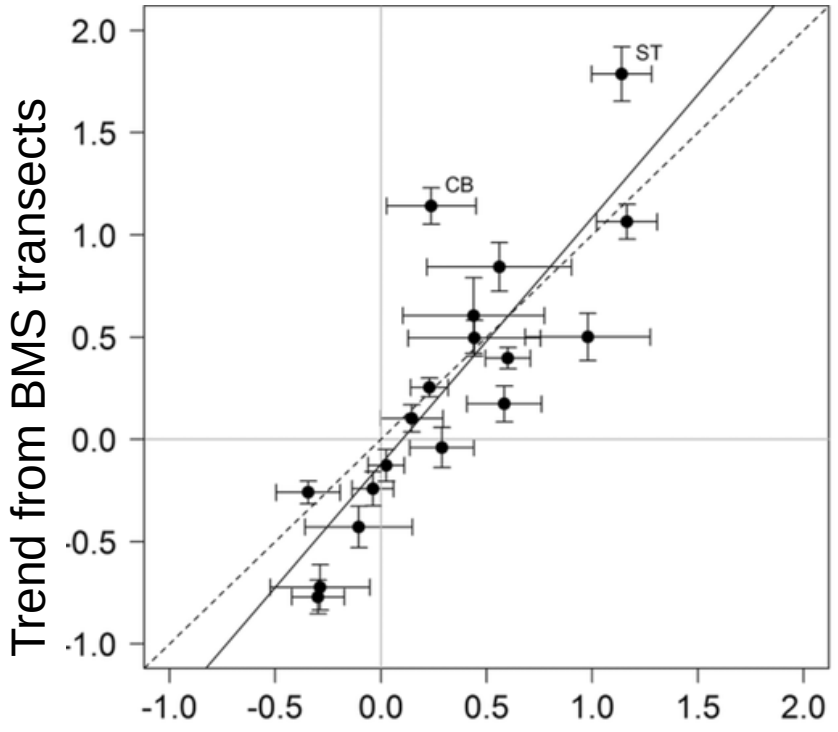
Timed counts

- Easy to do: suitable weather
- Standardised “reduced effort” method: time fixed, defined area
- Involves expert and non-expert citizen science
- Important for species less suited to transect counts (e.g. tree species, species requiring larval or egg counts)
- Also for remote areas and sites that are difficult to survey
- Public engagement benefits

Example: Big Butterfly Count in the UK

- Citizen-science survey launched in 2010
- Count 18 common species for 15 minutes during bright weather
- Large sample
  - Average ~47,600 participants per year 2013-2015
  - Most counts undertaken in gardens (65%)

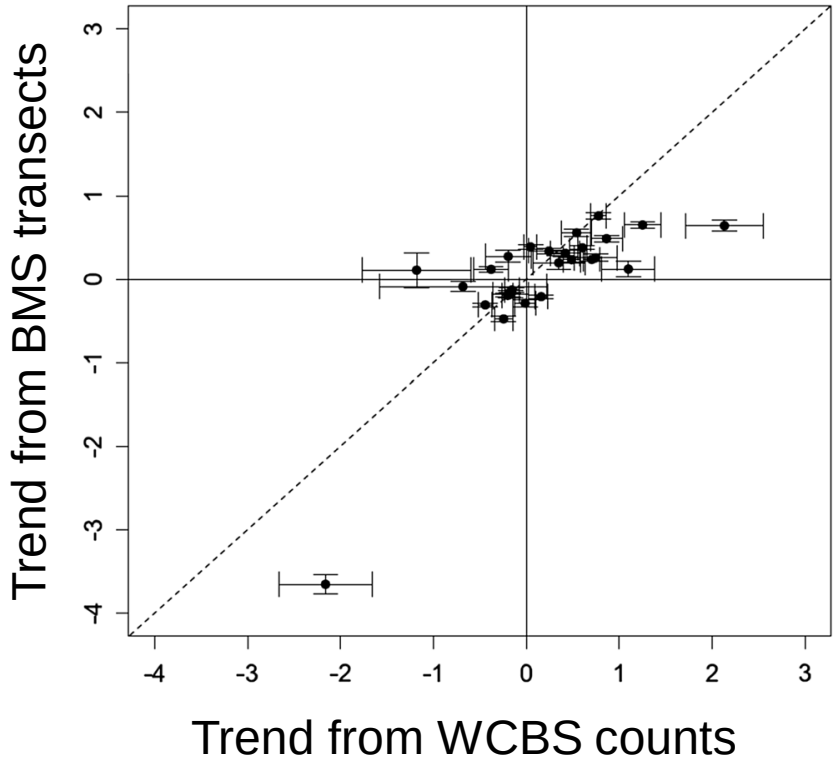




Dennis, EB *et al.* (2017) Using citizen science butterfly counts to predict species population trends. *Conservation biology* 31(6): 1350-1361

Comparing UKBMS and WCBS

- Wider countryside butterfly survey – stratified random sample of sites (1-km squares)
- 2-4 visits per year
- UKBMS biased towards semi-natural habitat and protected sites
- Widespread species respond similarly across the landscape

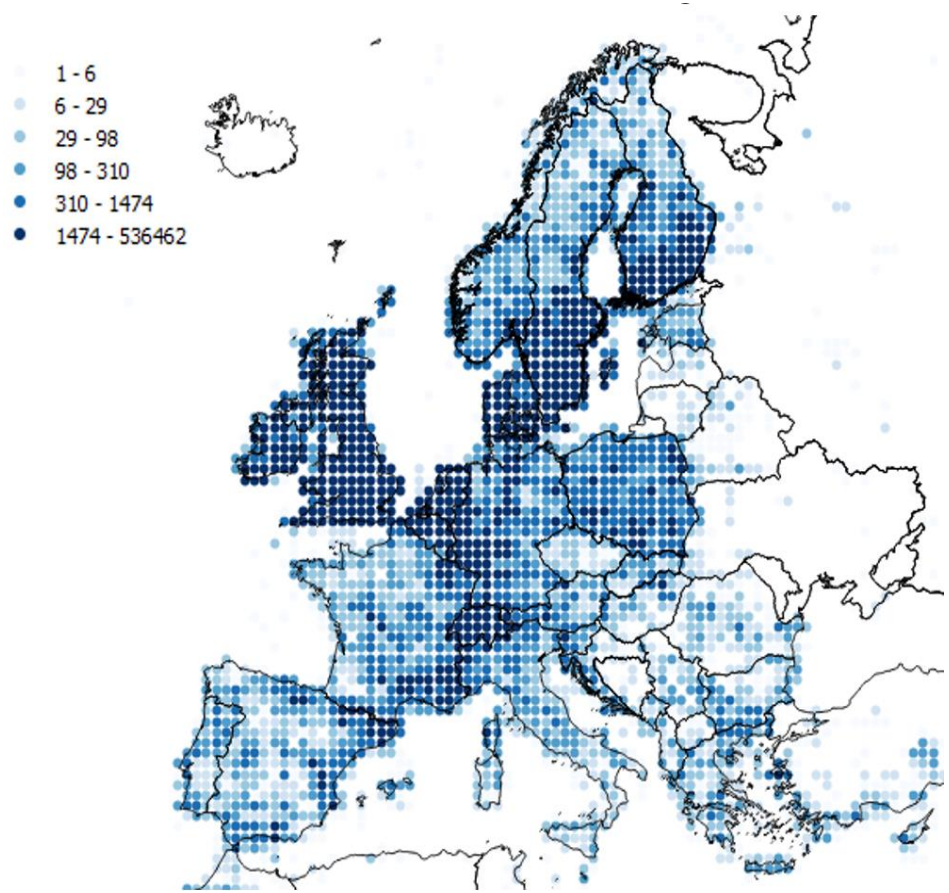


Roy, DB *et al.* (2015) Comparison of trends in butterfly populations between monitoring schemes. *Journal of Insect Conservation 19*(2): 313-324

Approaches to butterfly monitoring

Sightings (opportunistic data)

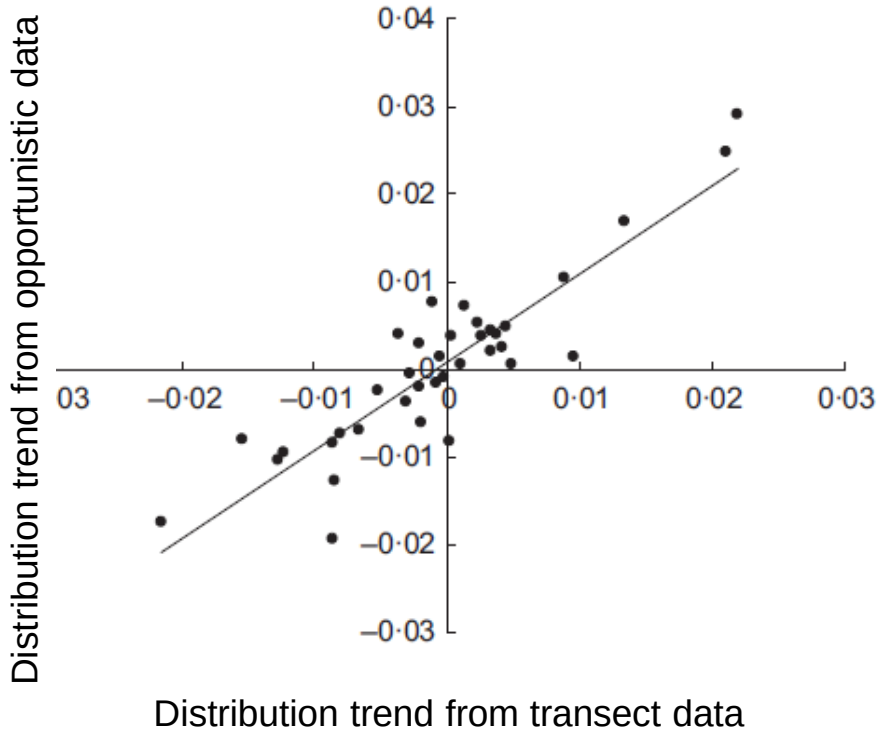
- Easy to do: almost no rules and conditions
- Can involve citizen science
- Typically offer high quantity and spatial coverage



1.3 million ‘accessible’ records, from GBIF, iNaturalist and Observation.org

Trends from opportunistic data

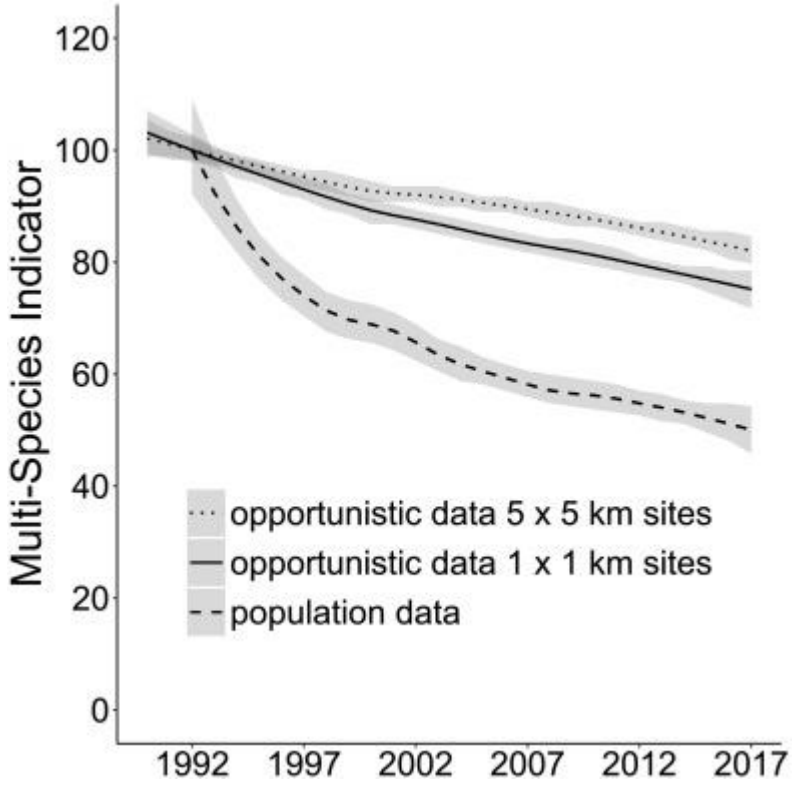
- Sources of bias
  - Change in effort over time
  - Incomplete and selective recording
  - Geographical bias
- Occupancy models account for imperfect detection
- Similar distribution trends from opportunistic and transect data



Van Strien, AJ *et al.* (2013) Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology* 50: 1450-1458

Opportunistic trends can be less sensitive

MSI for butterflies in the Netherlands

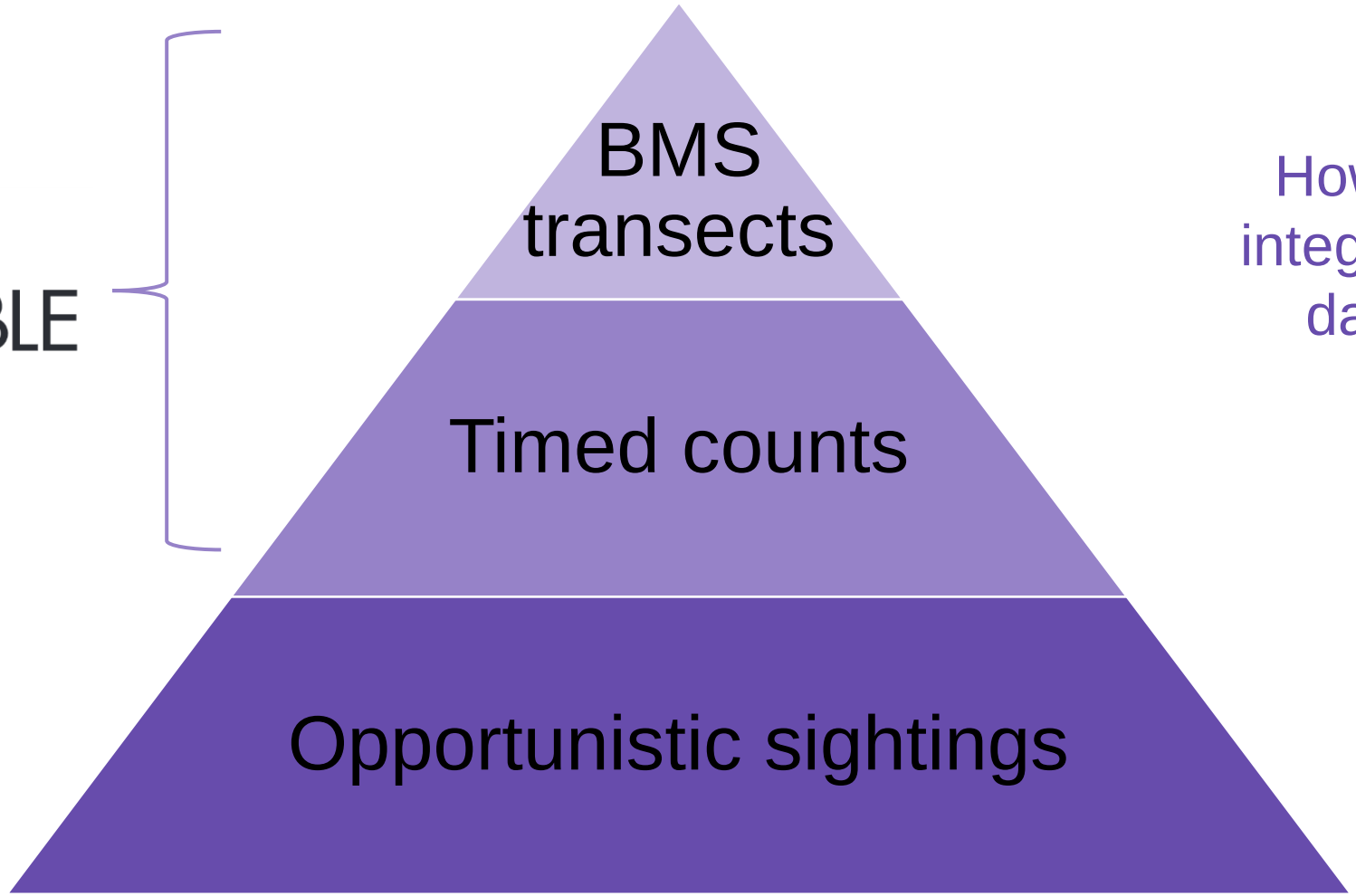


Van Strien, AJ *et al.* (2019) Over a century of data reveal more than 80% decline in butterflies in the Netherlands. *Biological Conservation* 234: 116-122

Summary of data sources

	BMS Transects	Timed Counts	Opportunistic sightings
Complete list of species sampled	✓	✓	?
Repeated throughout the season	✓	?	?
Repeat samples from the same location	✓	?	?
Sampling effort recorded	✓	✓	✗
Abundance (count) data	✓	✓	?
ABLE tools	✓	✓	✗

Summary of data sources



How can we integrate these datasets?

Combining transects and timed counts

- Requires suitable standardization of the timed counts to produce indices of abundance
- With the aim to combine with transect data
- Timed counts are included with UKBMS for producing trends e.g. for High Brown Fritillary *Argynnis adippe*
- Need for replication?
- Future task to bring in timed counts to European indicators
- Methods and approaches depend on the end goal

The bigger challenge

- Transect and timed count data can be included with sightings data for producing maps and estimating distribution change
- But can we more formally combine transects and timed counts with opportunistic sightings data?
- Models that integrate different types of population data have been around for a while
- Combining distribution data is a newer challenge

What has been done already?

- Conventional modelling is built around one data type
  - Structured data e.g. transects
  - Larger quantities of unstructured data
- Integrated modelling aims to combine datasets whilst
  - retaining the strengths of each
  - describing the differences in the datasets

Trends in Ecology & Evolution

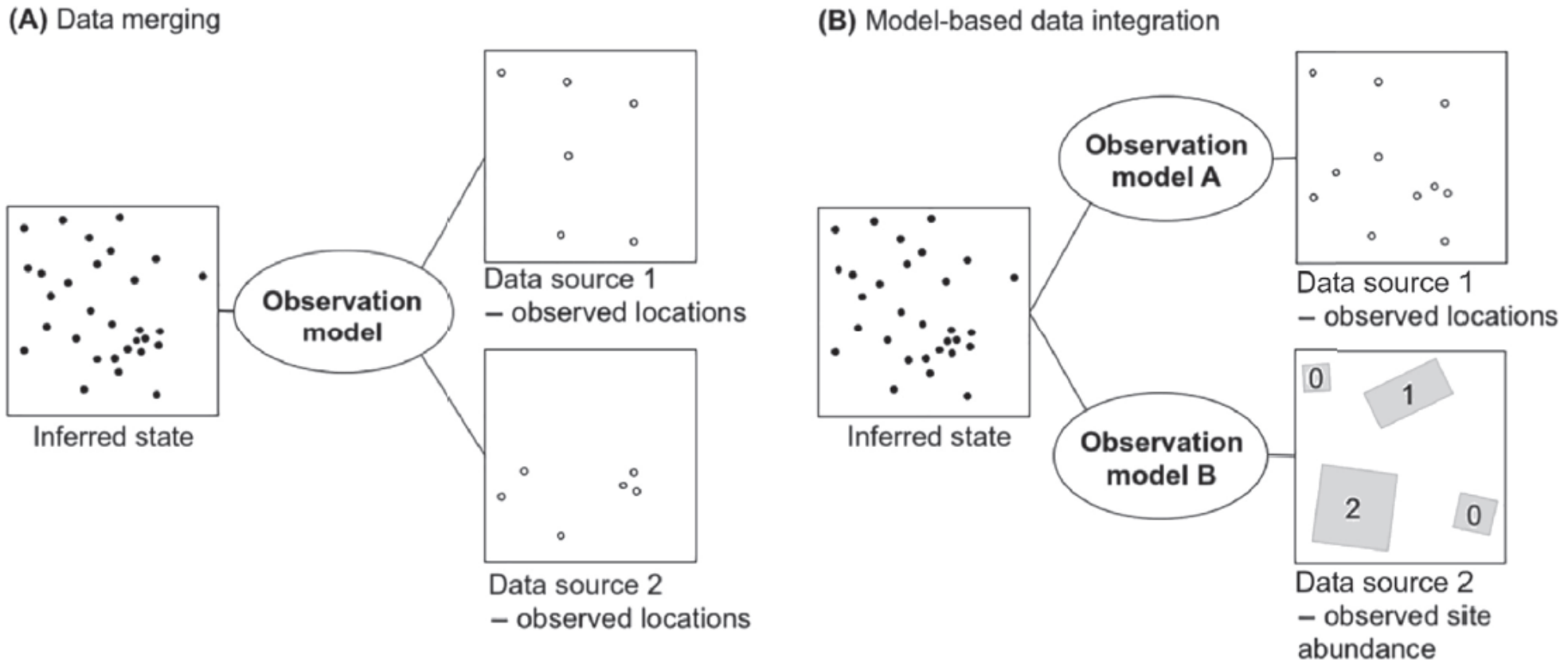
CellPress REVIEWS

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] and Robert B. O'Hara[12]

Data merging vs integration



Isaac, NJB *et al.* (2019) Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution*. In Press.

A butterfly example

Methods in Ecology and Evolution

British Ecological Society

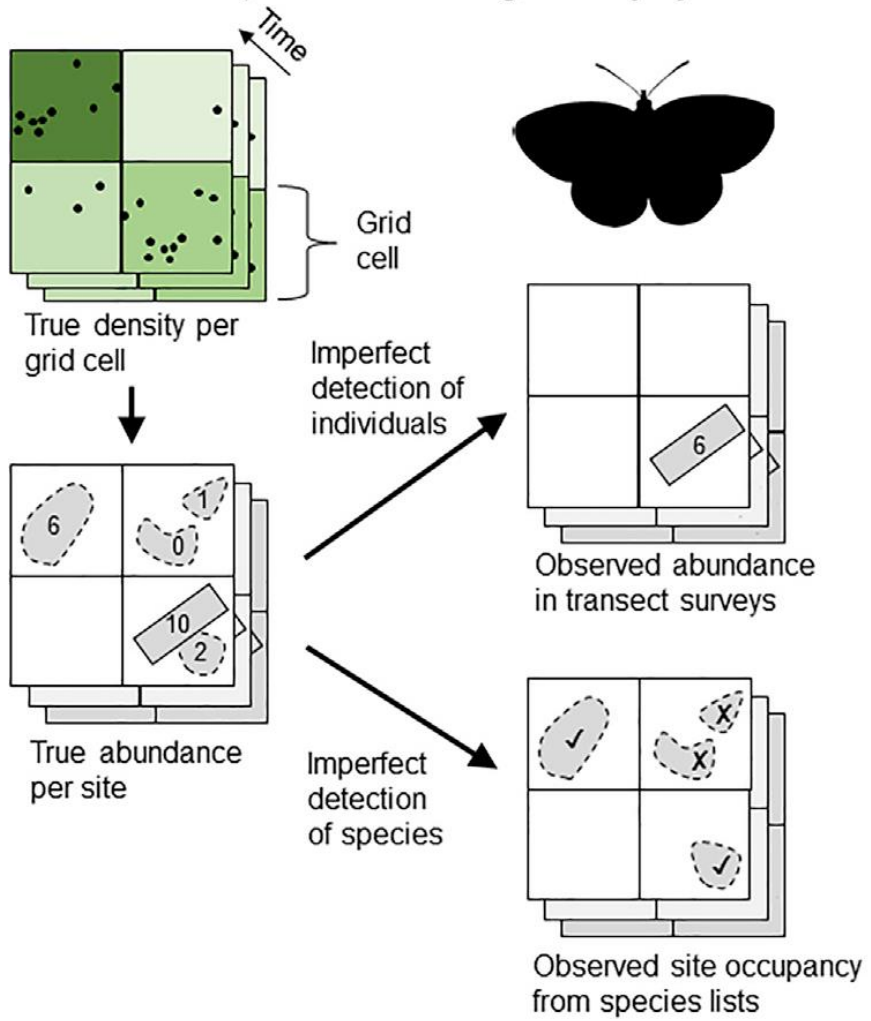
Methods in Ecology and Evolution 2014 doi: 10.1111/2041-210X.12221

Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records

Jörn Pagel[1,2*], Barbara J. Anderson[3,4], Robert B. O'Hara[5], Wolfgang Cramer[6], Richard Fox[7], Florian Jeltsch[1], David B. Roy[8], Chris D. Thomas[4] and Frank M. Schurr[2,9]

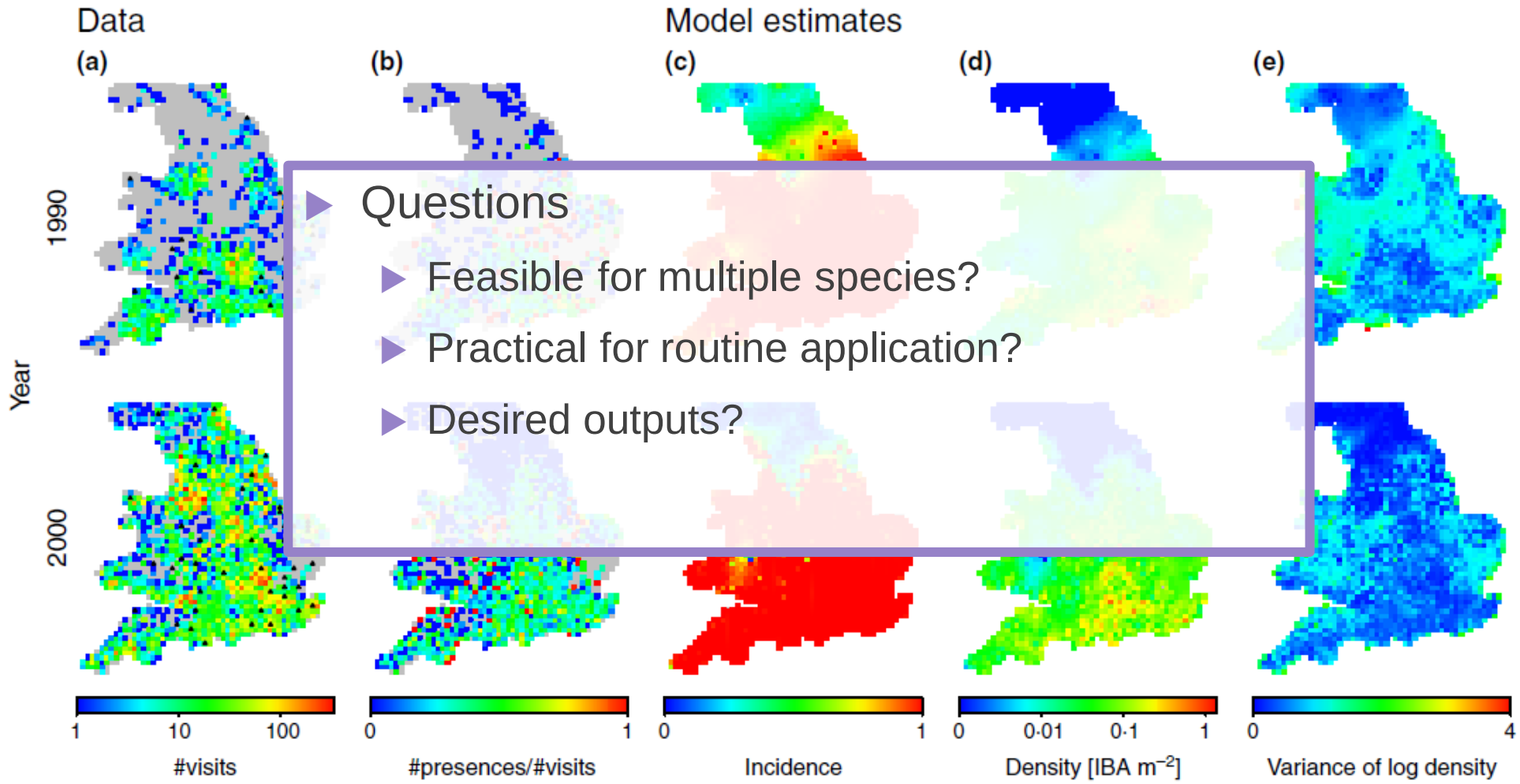
- Model that links spatial population density to transect counts and sightings data
- Estimate detectability-density relationship
- Based on a 10km grid

Schematic representation of Pagel *et al.* [42]



Isaac, NJB *et al.* (2019) Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution*. In Press.

Example for Gatekeeper *Pyronia tithonus*

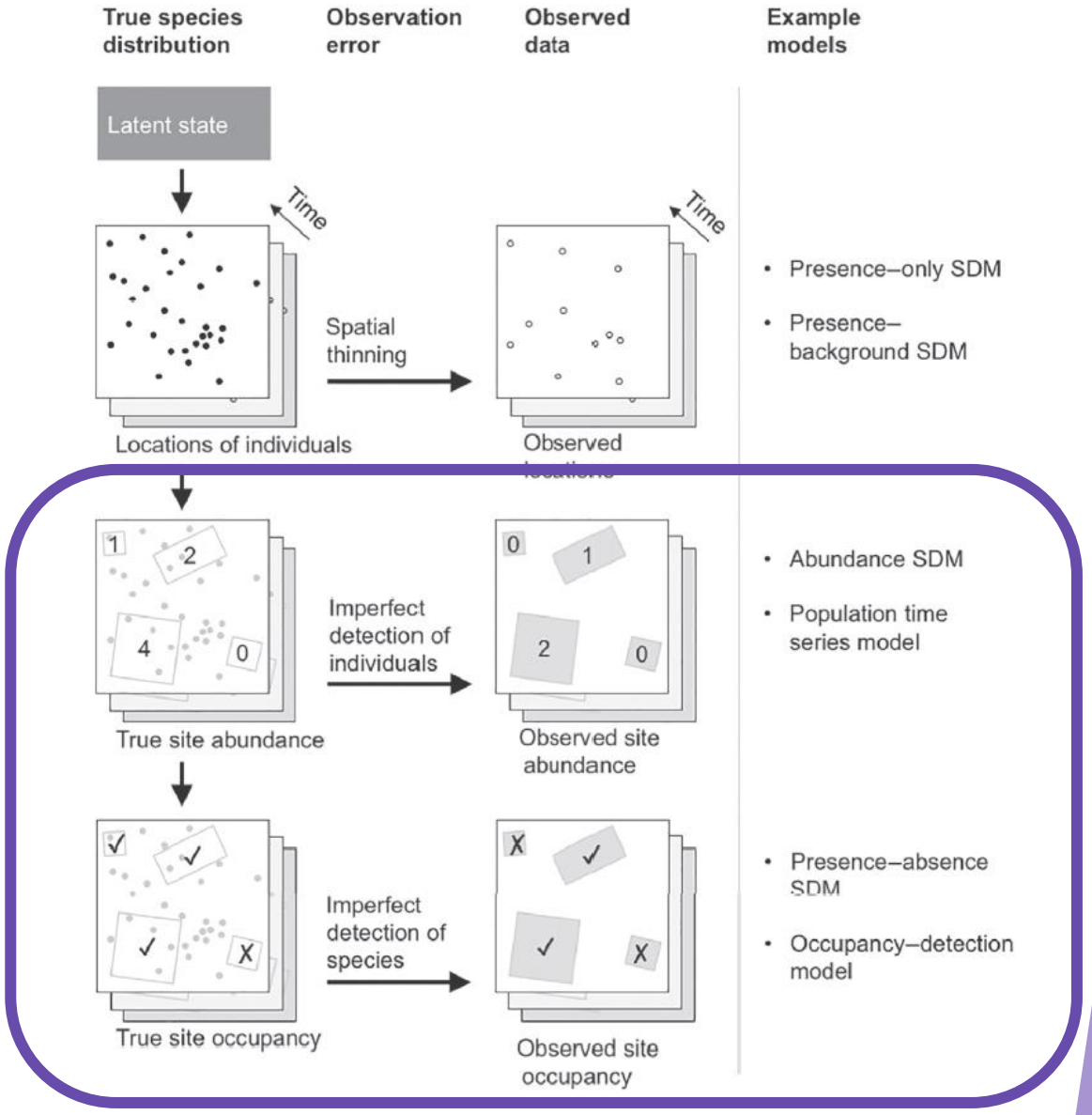


- Questions
  - Feasible for multiple species?
  - Practical for routine application?
  - Desired outputs?

Pagel *et al* 2014

Upcoming approaches

- Poisson process models
  - Statistical description of how points are distributed in space
  - Rather than grid-based
  - General framework to encompass different data types



Isaac, NJB *et al.* (2019) Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution*. In Press.

Outstanding questions

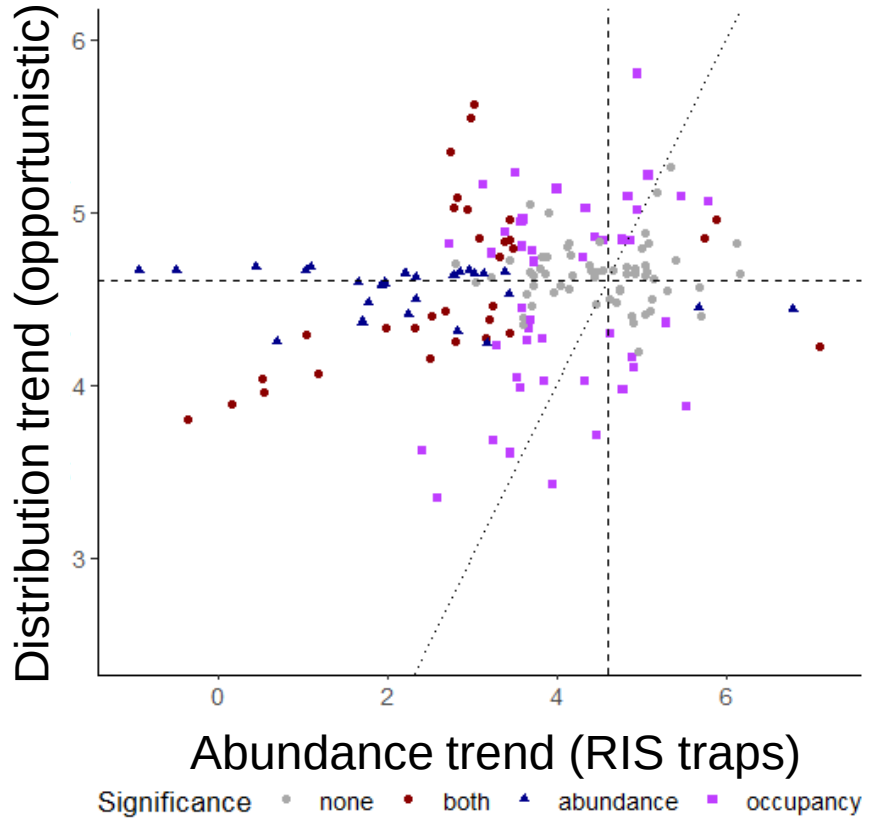
- When should complex integrated models be used?
- What are the potential benefits?
  - Optimizing use of data and effort
  - Increased spatial coverage
  - Improved precision
- How feasible are integrated models in practice?
  - Computationally demanding
  - How to validate and test these models
  - Considering assumptions and biases

Relevance for multi-species indicators

- Indicators can be based upon
  - Transect data/abundance only
  - Separately for abundance and occupancy e.g. UK priority species indicators
  - Combine abundance and occupancy
    - Occupancy as a proxy for abundance
- Assumptions
  - Assume equivalence across the two type of data
  - Is a 10% change in abundance equivalent to a 10% change in distribution?
  - What is the indicator measuring?

Example: Scottish trends for moths

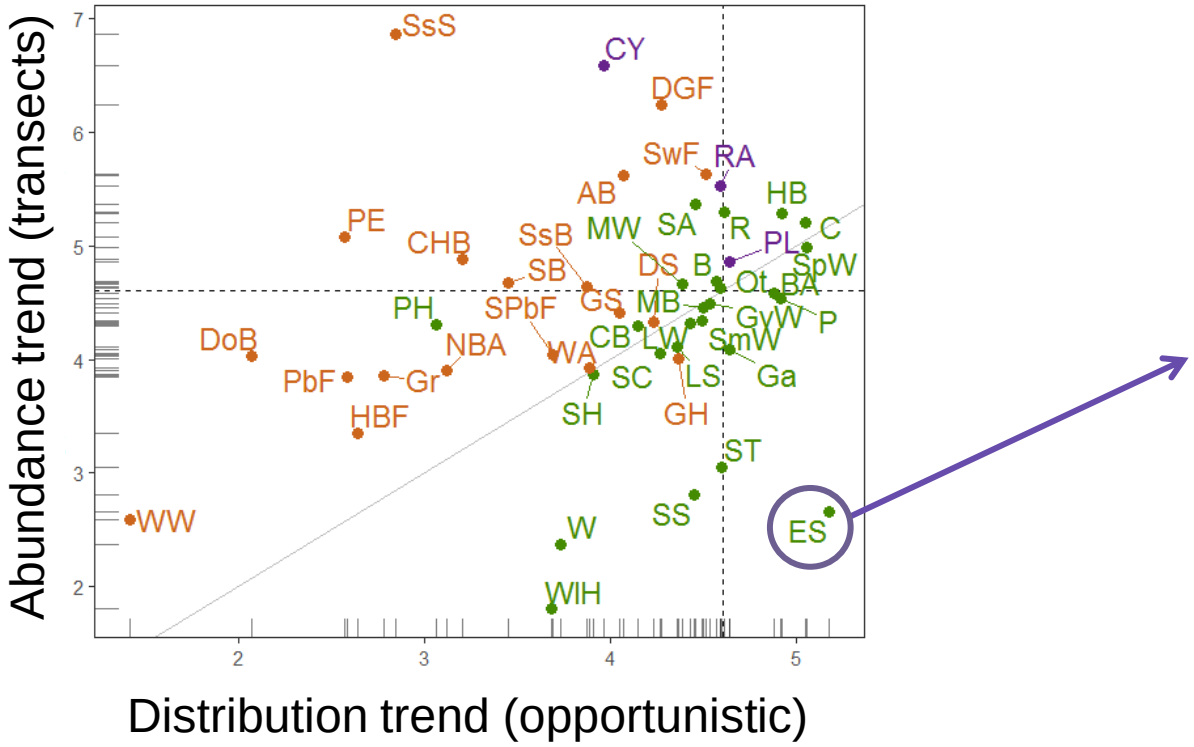
- 164 species trends for 1990-2014
- Abundance trend: -46%
- Distribution trend: +15.8%
- 13 species had significance trends of opposite signs



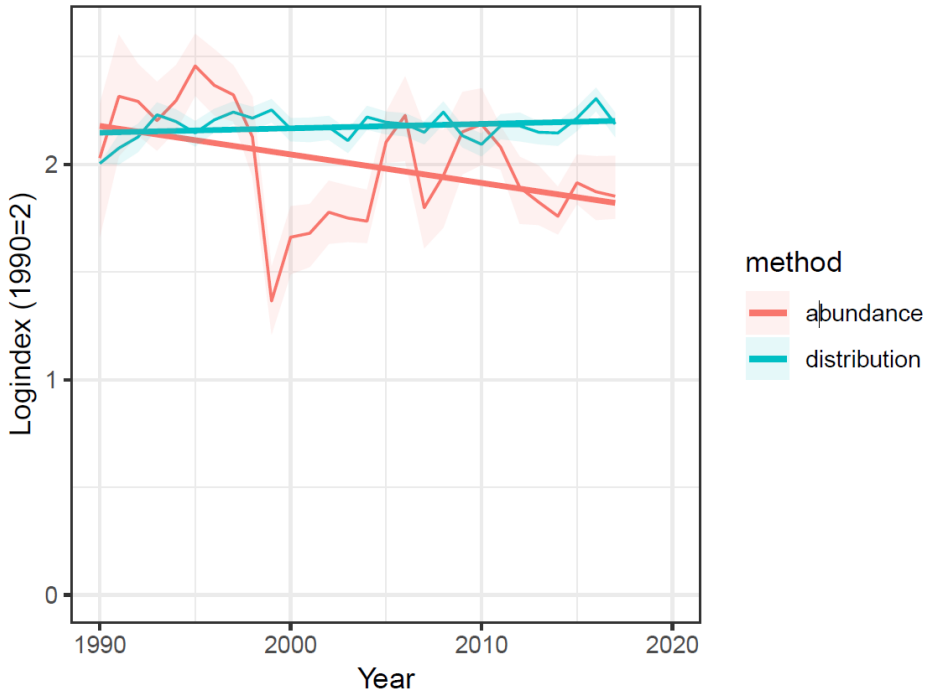
Dennis, EB *et al.* (2019) Trends and indicators for quantifying moth abundance and occupancy in Scotland. *Journal of Insect Conservation* *23*(2): 369-380

Example: comparing butterfly trends

UK butterfly trends

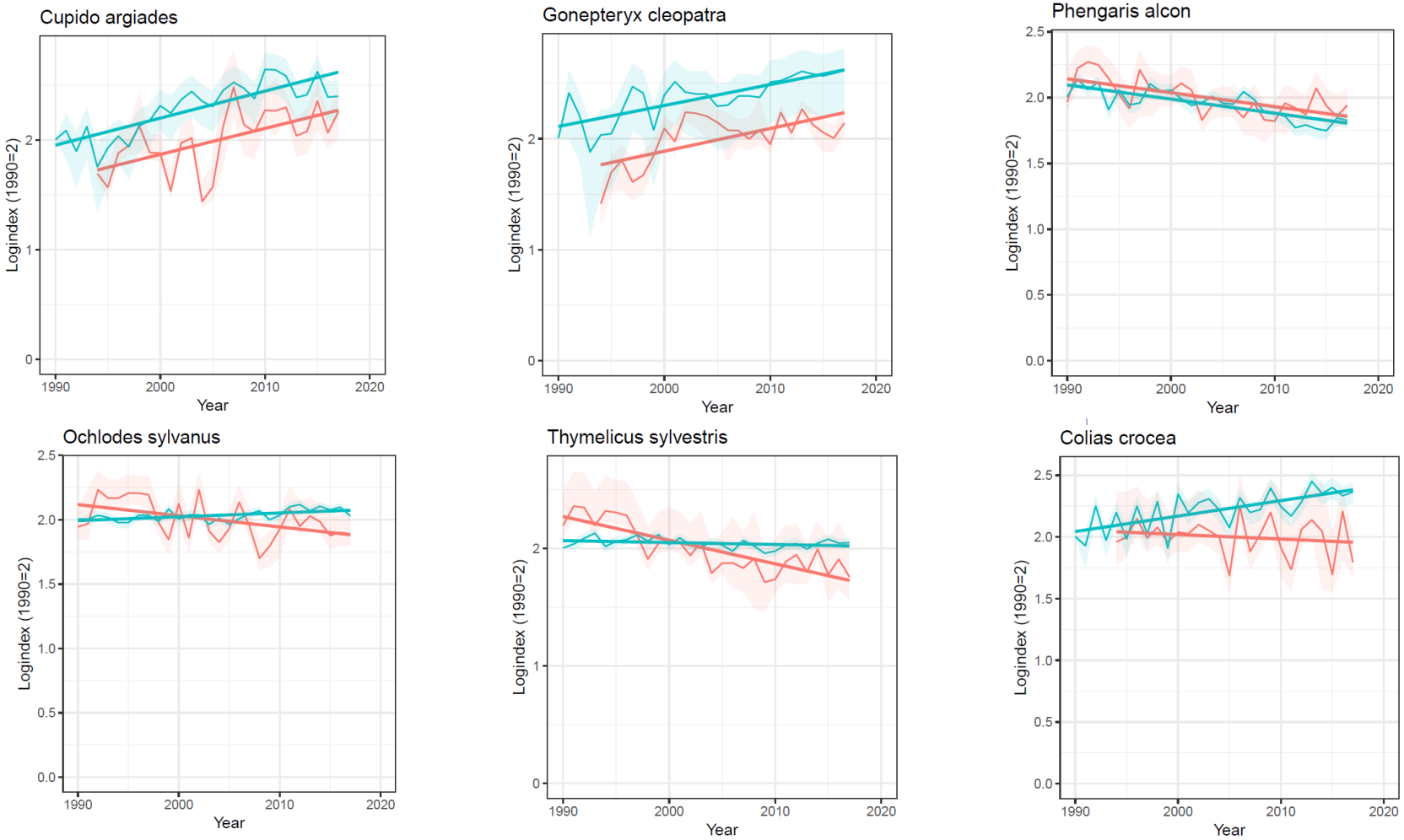


EU trends for Thymelicus lineola



Dennis, EB *et al.* (2019) Functional data analysis of multi-species abundance and occupancy data sets. *Ecological Indicators* 104: 156-165

EU butterfly trends



For discussion

- Opportunistic sightings data
  - Benefits for atlases and mapping
  - Cannot assume equivalence with abundance
  - Integrating with count data remains a challenge
- Prioritize transect data where feasible
- But timed count data offer a good alternative

- We want to bring together data from multiple sources to get best assessment of species' status – nice in theory but difficult in practice