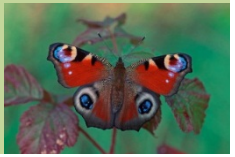
Introduction to distribution and abundance

Marc Kéry
Swiss Ornithological Institute
marc.kery@vogelwarte.ch

Occupancy Workshop
Patuxent
Nov 2013

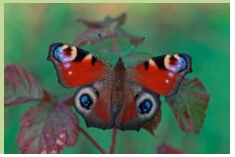
Contents

- Why do we care about distribution and abundance ?
- What is occurrence/distribution ?
- The ugly truth about all distribution & abundance data
- Kinds of data on distribution and abundance
- Value of hierarchical models for ecological data

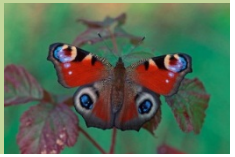


Contents

- **Why do we care about distribution and abundance ?**
- What is occurrence/distribution ?
- The ugly truth about all distribution & abundance data
- Kinds of data on distribution and abundance
- Value of hierarchical models for ecological data



Why do we care about distribution and abundance ?

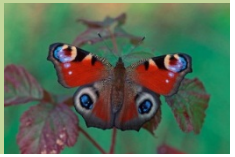


Why do we care about distribution and abundance ?

HUGE TOPICS !!!

- Population central concept in ecology and management
- Population size (N) first descriptor of the population
- Population size single-most important “thing” in ecology and management
- Distribution almost as important ...

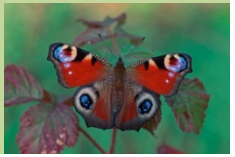
  .... or more (e.g., cheaper data, simpler models)
- “Ecology = Study of distribution and abundance” (e.g., textbook by Krebs)
- though see later about relationship between the two



Why do we care about distribution and abundance ?

HUGE TOPICS !!!

- Interest for distribution:
  - Species distribution modeling industry
  - Metapopulation ecology
  - Climate change
  - Disease ecology/disease surveillance
  - Invasive species
  - Niche ecology



Relationship between distribution and abundance

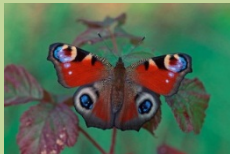
- Distribution and abundance sometimes treated as separate quantities
- e.g., zero-inflated Poisson modeling of occurrence and of abundance, given occurrence
- Model for occurrence/distribution at site *i*:

$$z_i \sim Bernoulli(\psi_i)$$
$$logit(\psi_i) = \alpha_0 + \alpha_1 * Xi + \ ...$$

- Model for abundance at site *i*, given occurrence (i.e., $z_i = 1$):

$$N_i \sim Poisson(z_i * \lambda_i)$$
$$log(\lambda_i) = \beta_0 + \beta_1 * Xi + \beta_2 * W_i$$

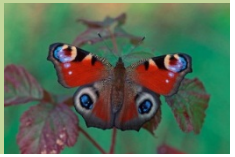
- That does not, or very rarely, make sense ! **Why ?**



Relationship between distribution and abundance

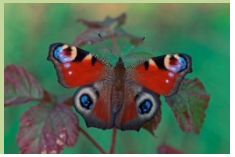
- **Because it's the same thing !**
- As if modeling height of people in two steps:
  - 1st model: probability of being $>Y_0$ cm (e.g., Y_0 = 150 cm)
  - 2nd model: truncated Normal for Y, given >150 cm

- Let's look at the relationship between occurrence, distribution and abundance in more detail ...



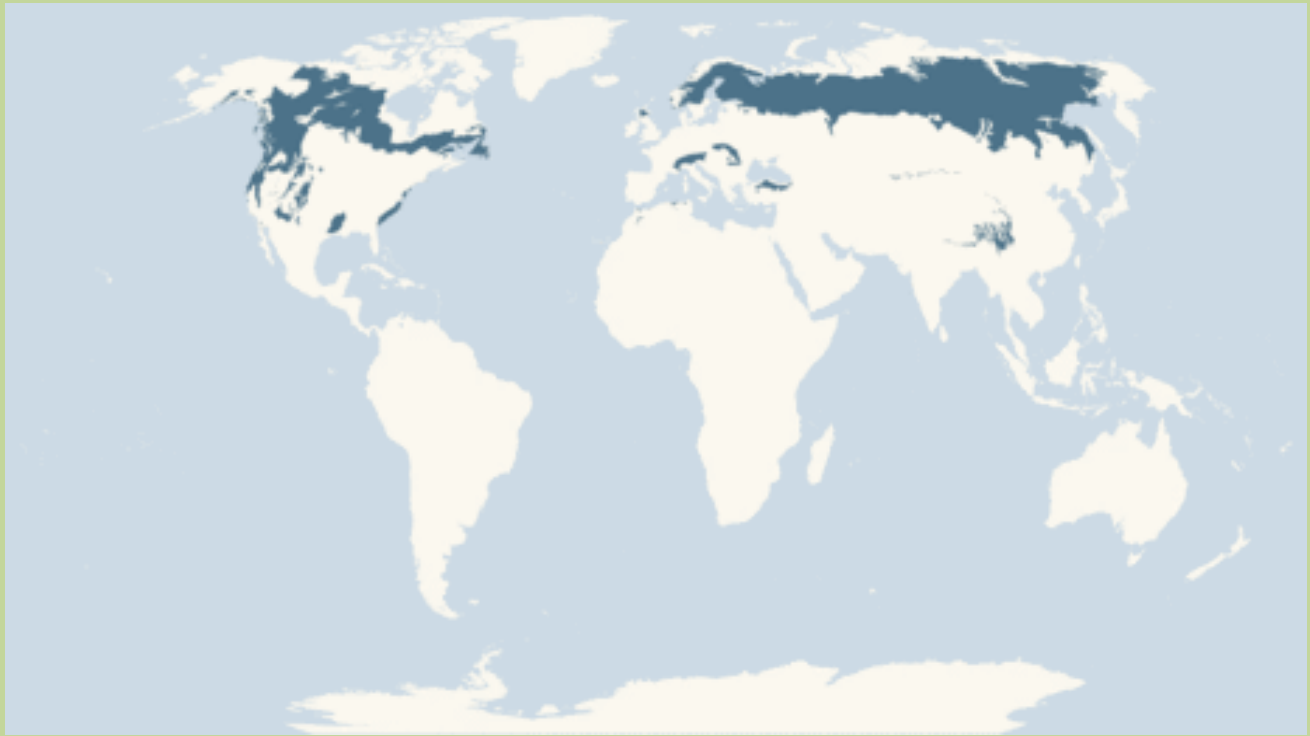
Contents

- Why do we care about distribution and abundance ?
- **What is occurrence/distribution ?**
- The ugly truth about all distribution & abundance data
- Kinds of data on distribution and abundance
- Value of hierarchical models for ecological data

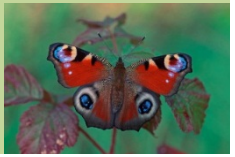


What is occurrence / (species) distribution ?

- Google “common crossbill distribution”
- We immediately understand what this means ...

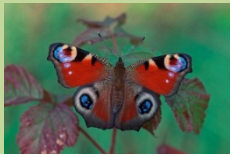


http://www.bbc.co.uk/nature/life/Common_Crossbill



What is occurrence / (species) distribution ?

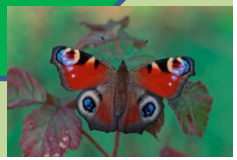
- Or do we really ??? What **exactly** is meant by that picture ?



What is occurrence / distribution ?

- What **exactly** is meant by such distribution maps ?

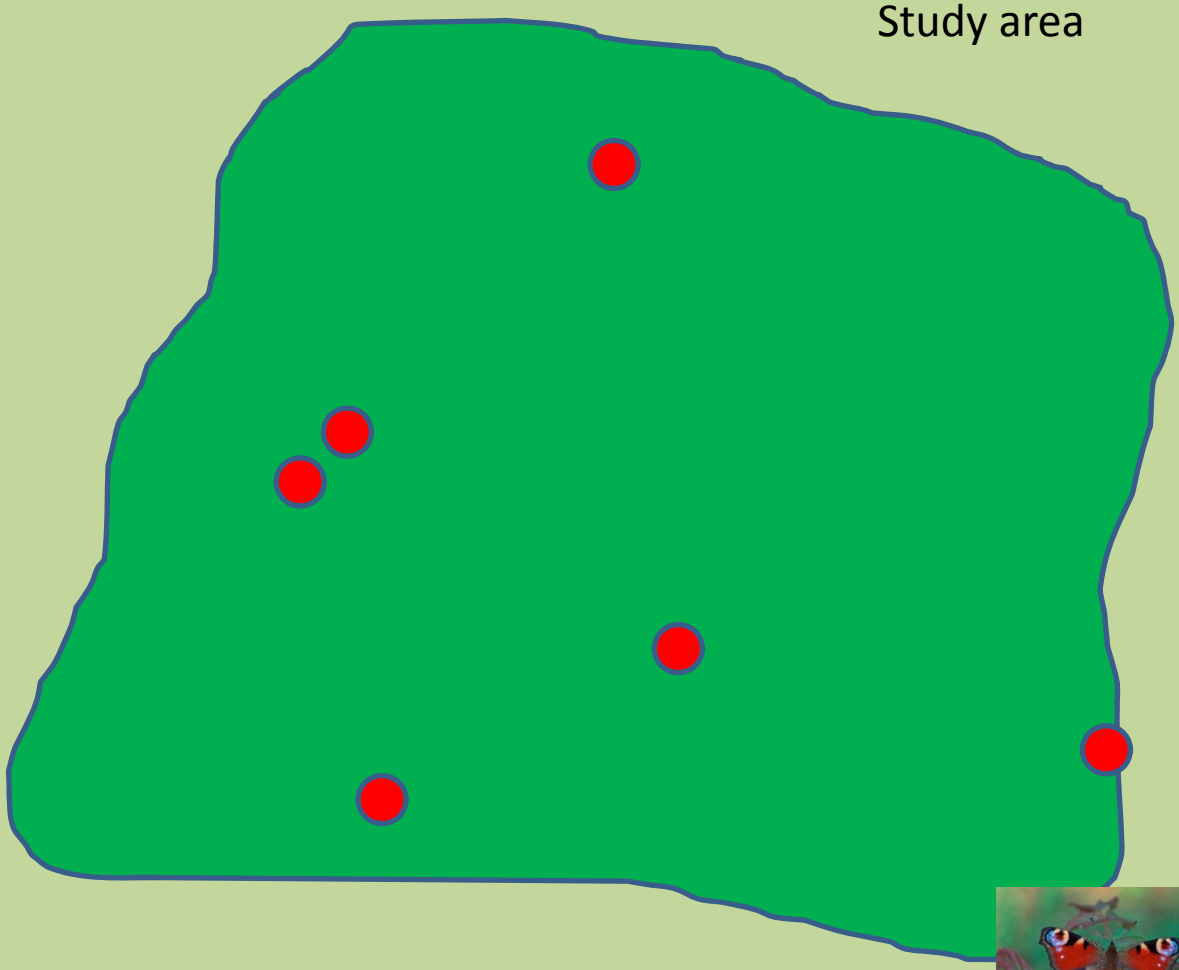
Male crossbill

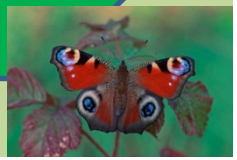


What is occurrence / distribution ?

- What **exactly** is meant by such distribution maps ?

Male crossbill

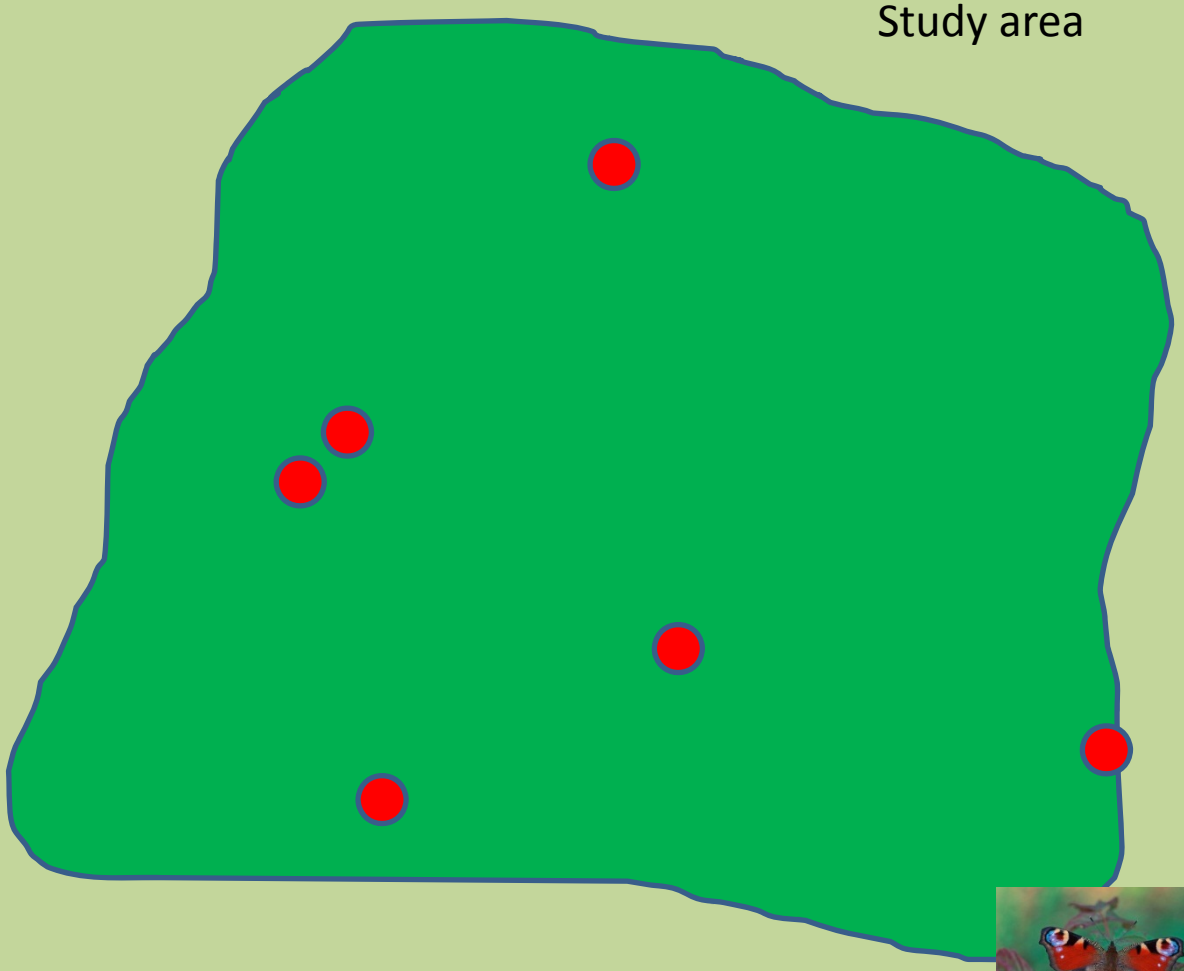


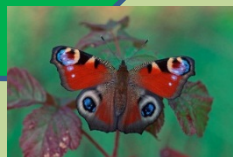


What is occurrence / distribution ?

- **Distribution:** “area where a species occurs”

$$\frac{Area\ with\ crossbill}{Total\ area}$$

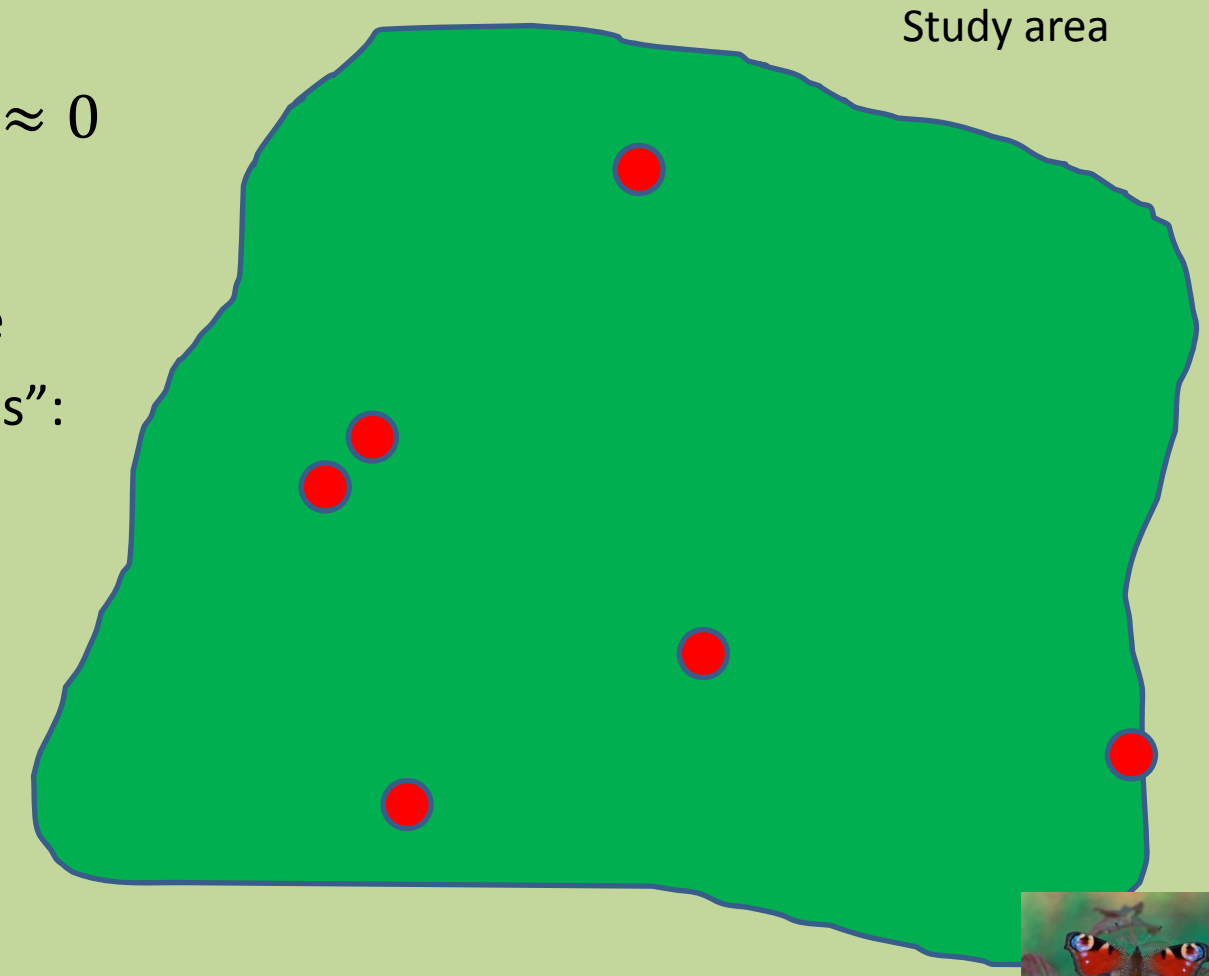




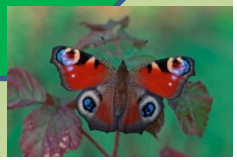
What is occurrence / distribution ?

- **Distribution:** “area where a species occurs”

$$\frac{Area\ with\ crossbill}{Total\ area} \approx 0$$

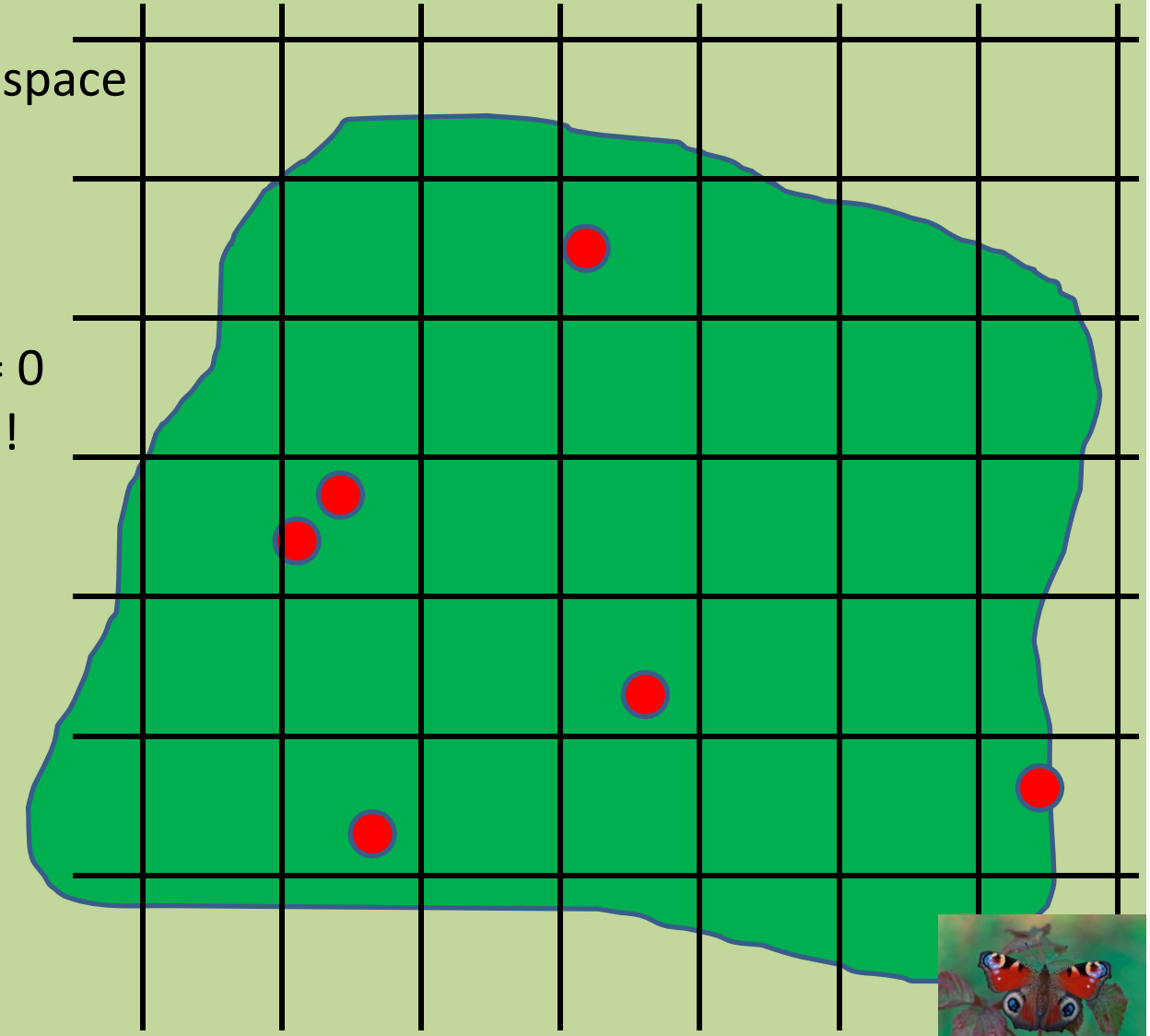


- Birds small & discrete
- Approximately “points”:
- Spatial point pattern: realization of spatial point process



What is occurrence / distribution ?

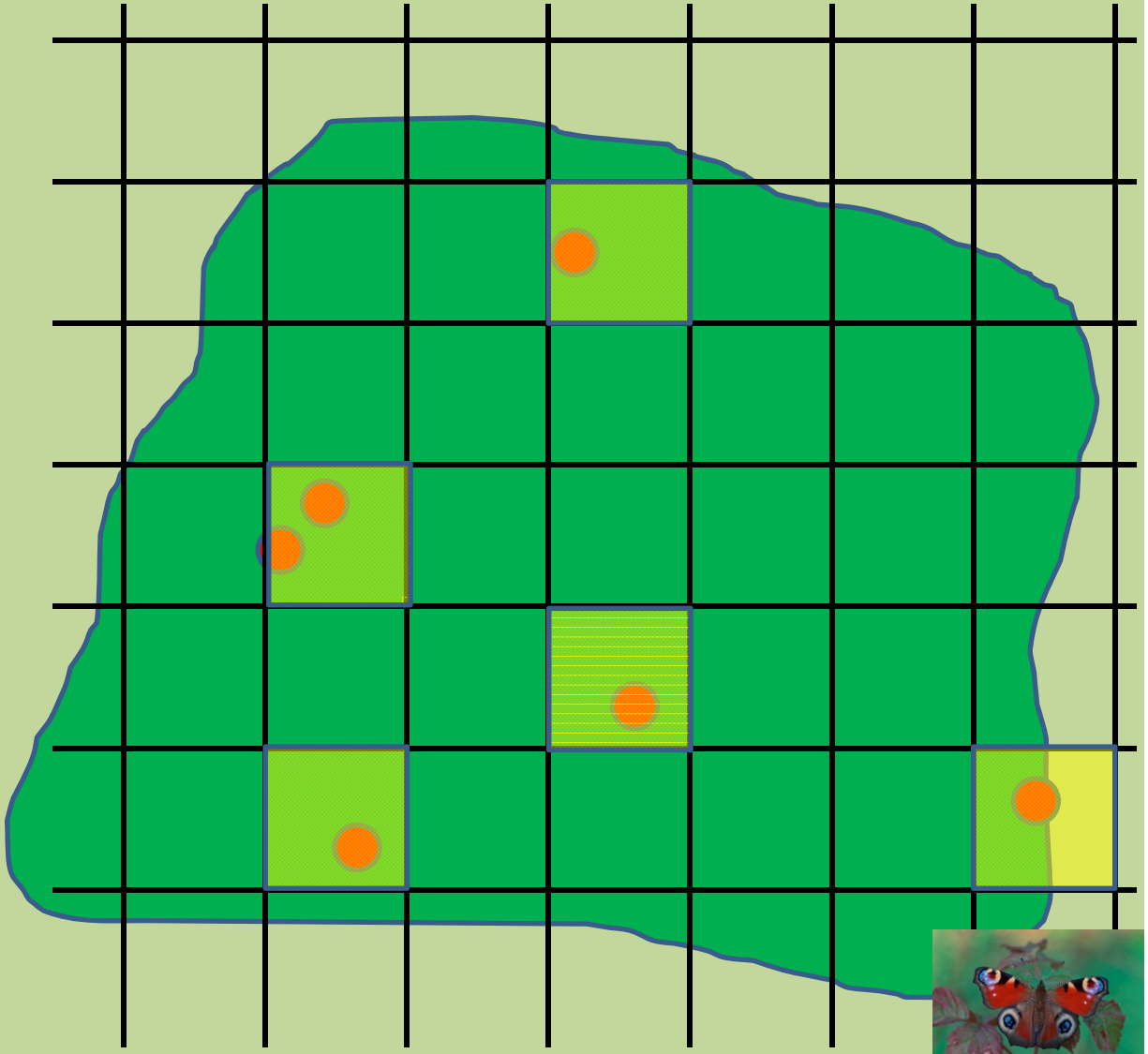
- In practice (*),
  most often discretize space
- e.g., impose grid
  and summarize
  spatial point process
- Otherwise bird area = 0
  (and distribution = 0) !



(*) spatial point patterns
can be modelled directly,
but may be difficult,
analyses/software not
well-known, data inadequate,
hard to acccommodate
imperfect detection ...

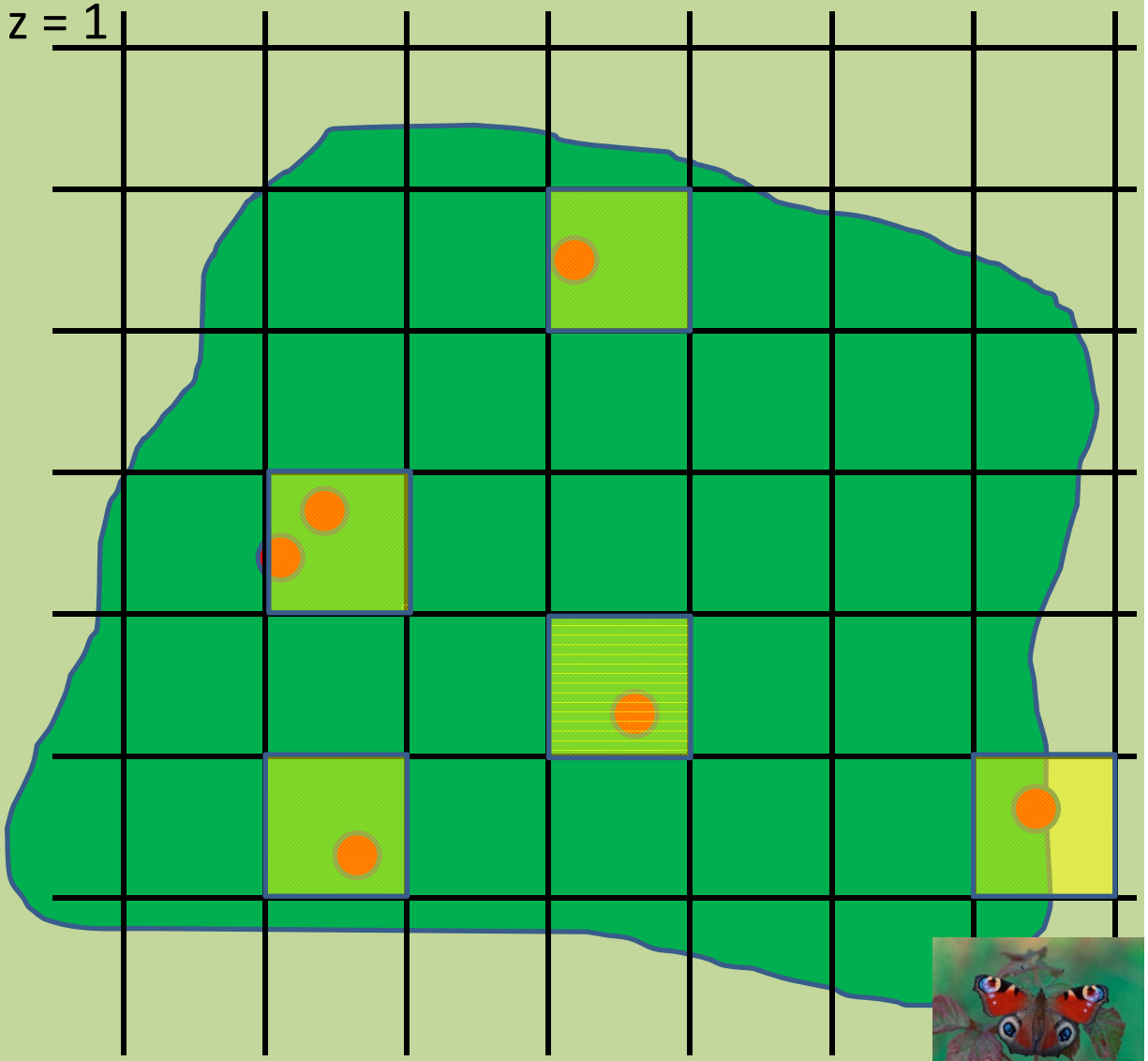
What is occurrence / distribution ?

- Occurrence (z):
  areal summary
  of pixel abundance N
- Distribution:
  pattern of z,
  sum of occurrences
  (5 pixels)



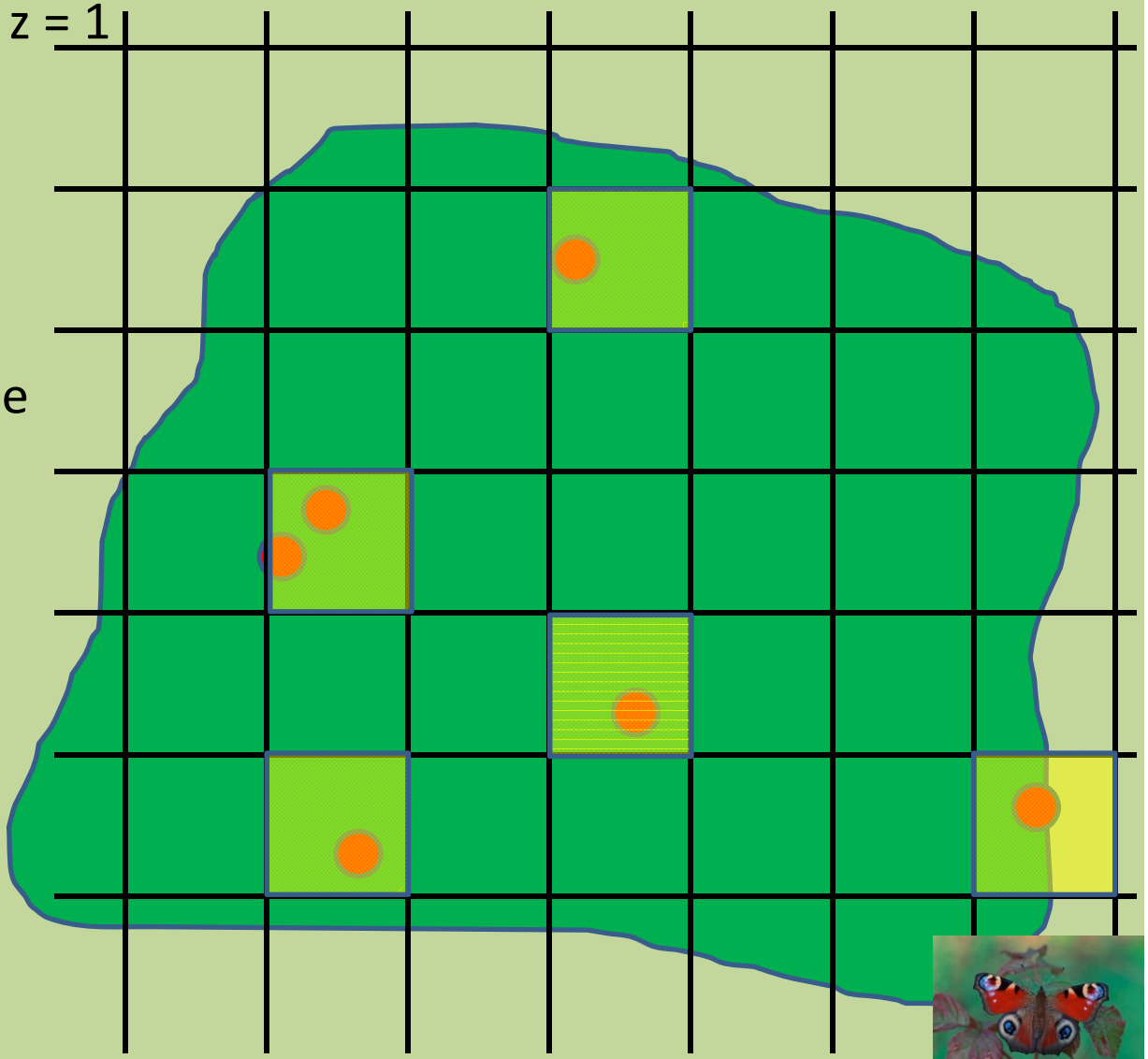
Mathematical relationship between distribution and abundance

- Occurrence/Presence: z = 1
- Presence z = 1 if N > 0
- Absence z = 0 if N = 0



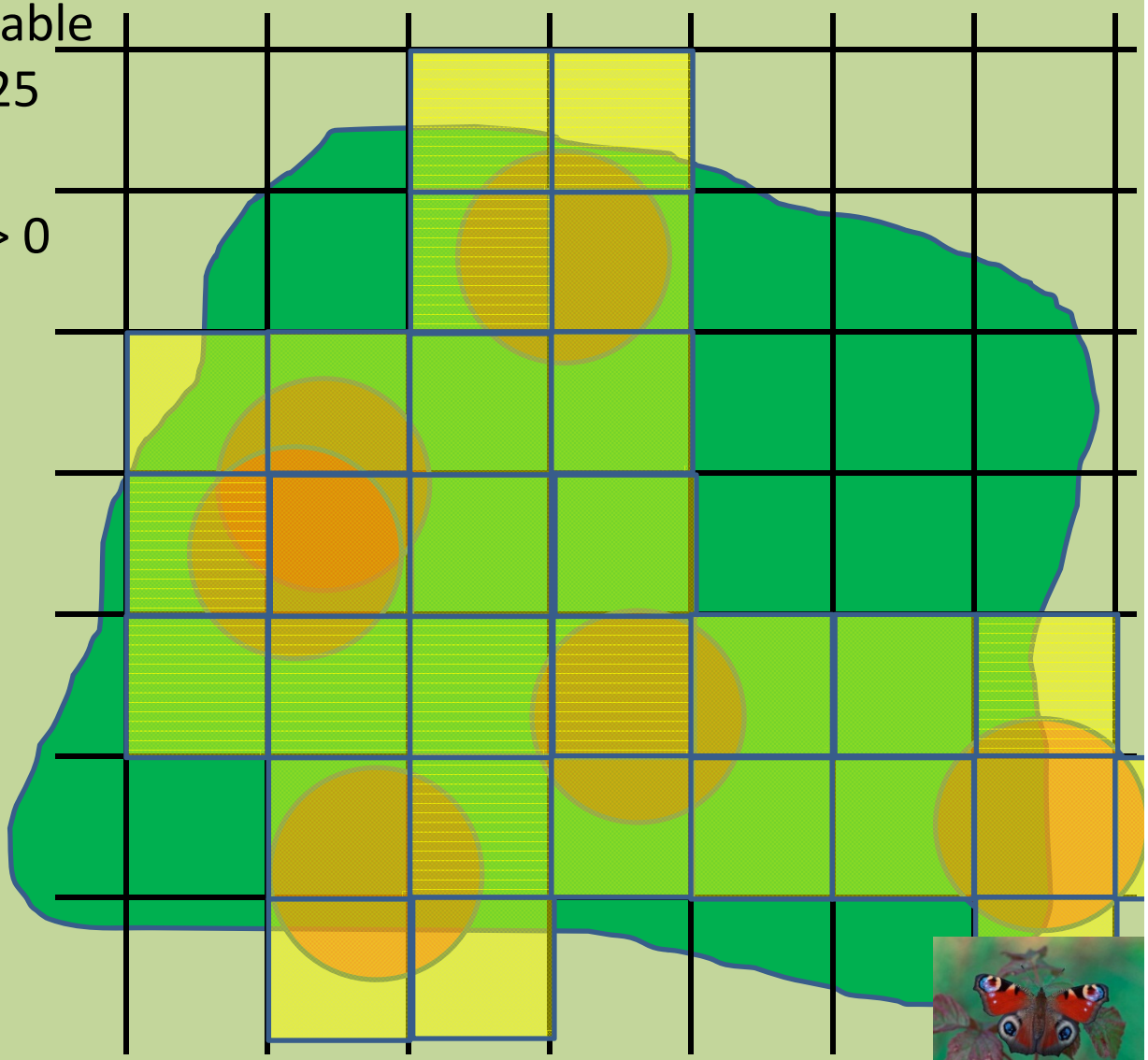
Mathematical relationship between distribution and abundance

- Occurrence/Presence: z = 1
- Presence z = 1 if N > 0
- Absence z = 0 if N = 0
- Distribution = information-reduced summary of abundance for some areal unit
- **AND:**
  Both distribution & abundance areal-summary of spatial point process



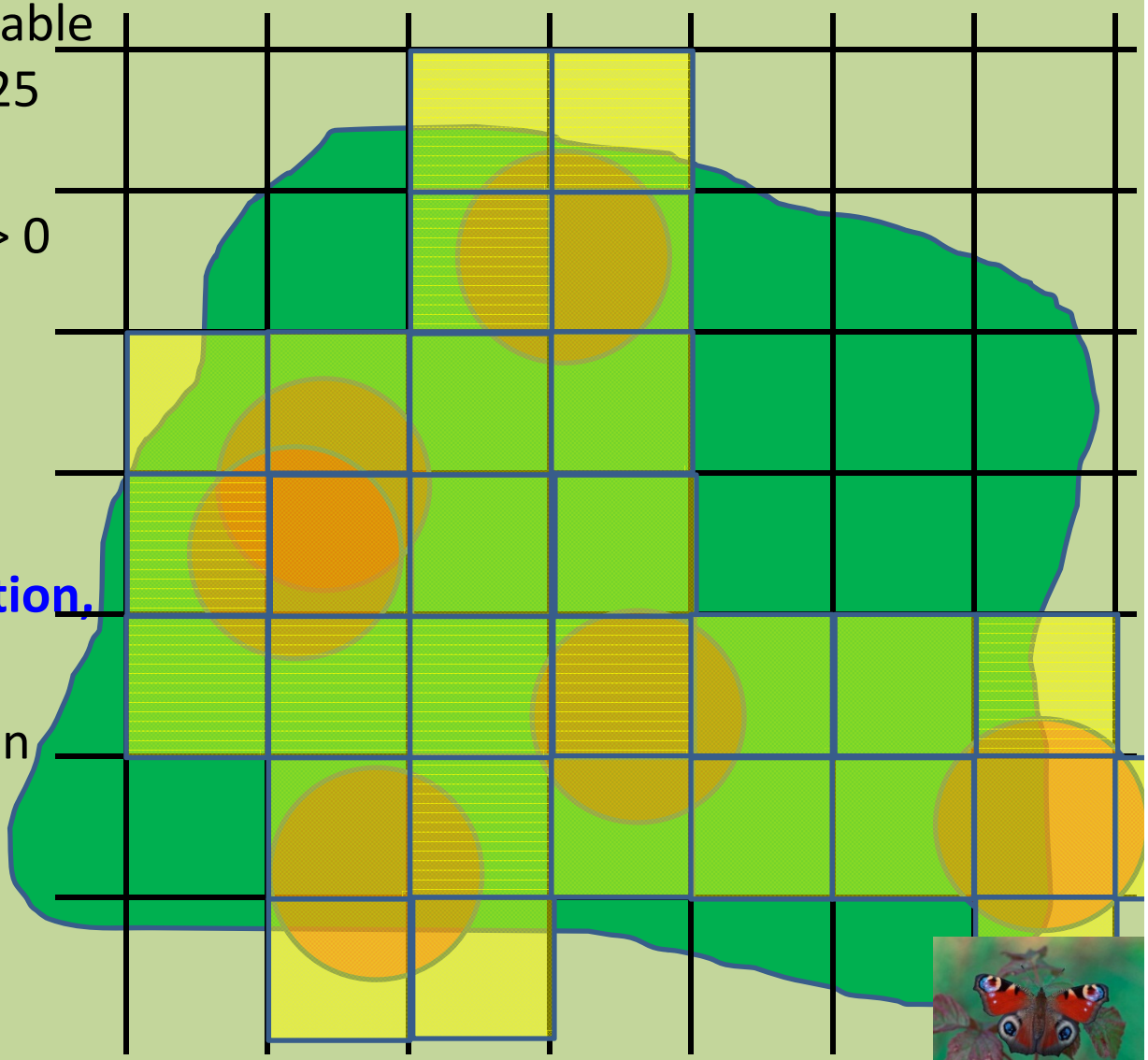
And what if there is movement (home-range) ?

- **New:** N is random variable
  E(N) fractional, e.g. 0.25
- Same:
  Presence z = 1 if E(N) > 0
- Same: Distribution = information-reduced summary of expected abundance



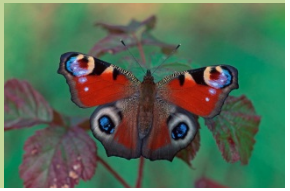
And what if there is movement (home-range) ?

- **New:** N is random variable E(N) fractional, e.g. 0.25
- Same: Presence z = 1 if E(N) > 0
- Same: Distribution = information-reduced summary of expected abundance
- **New: greater distribution, "use" vs. "presence"**
- Same: Both distribution & abundance areal-summary of SPP



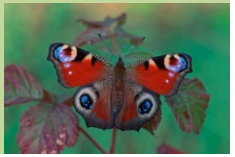
Section summary: what is occurrence/distribution ?

- Individual animals and plants small & discrete units
  (not continuous like sea surface temperature)
- Spatial point patterns !
- Fundamentally, abundance and species distribution both
  areal summary of spatial point patterns (= realisations of spatial point process)
- Occurrence (and distribution) is information-poor summary of abundance: $z = I(N > 0)$
- Continuous maps imply discrete spatial units,
  within which point pattern is summarized
- "Distribution" meaningless without:
  (i) definition of occurrence
  (ii) definition of spatial unit (pixel size)
  (iii) spatial scale of organism movement (if any)
  [(iv) temporal scale]



Contents

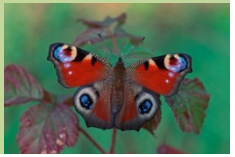
- Why do we care about distribution and abundance ?
- What is occurrence/distribution ?
- **The ugly truth about all distribution & abundance data**
- Kinds of data on distribution and abundance
- Value of hierarchical models for ecological data



The ugly truth about all distribution & abundance data

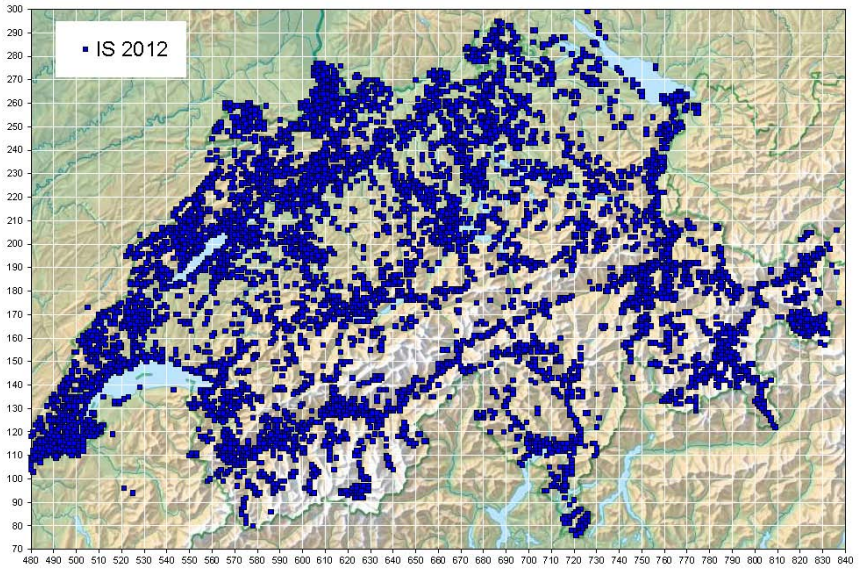
First complication: Spatial sampling

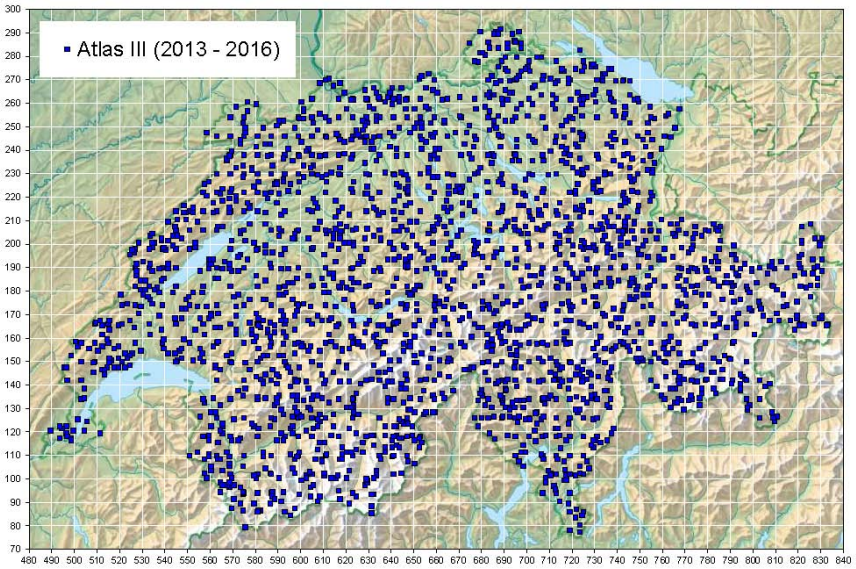
- Don't observe it in the entire domain about which you want to learn something, or, spatial heterogeneity in effort
- (OK, can predict ...)
- probability to get observation at a site may be related to the measurement (e.g., sites with some values of N or z or of covariates X may be more or less likely to be sample)
  - -> missing not at random (MNAR); non-ignorable missing value mechanism in statistics !
- Has to be modelled to avoid bias
- **Spatial sampling (i.e., *where* measurements are made) can be very important in ecology and management**



The ugly truth about all distribution & abundance data

- e.g., Swiss citizen-science program IS and designed Atlas III
- In principle, extrapolation based on:
  - random sampling
  - covariate relationships
  - Neighbourhood relationships





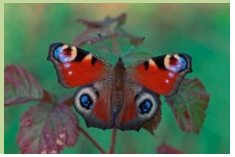
The ugly truth about all distribution & abundance data

Second complication: Measurement errors

- Two kinds for counts and occurrence data:
  - "Individuals" overlooked when present:

    -> false negative errors
  - "Individuals" misidentified/counted doubly:

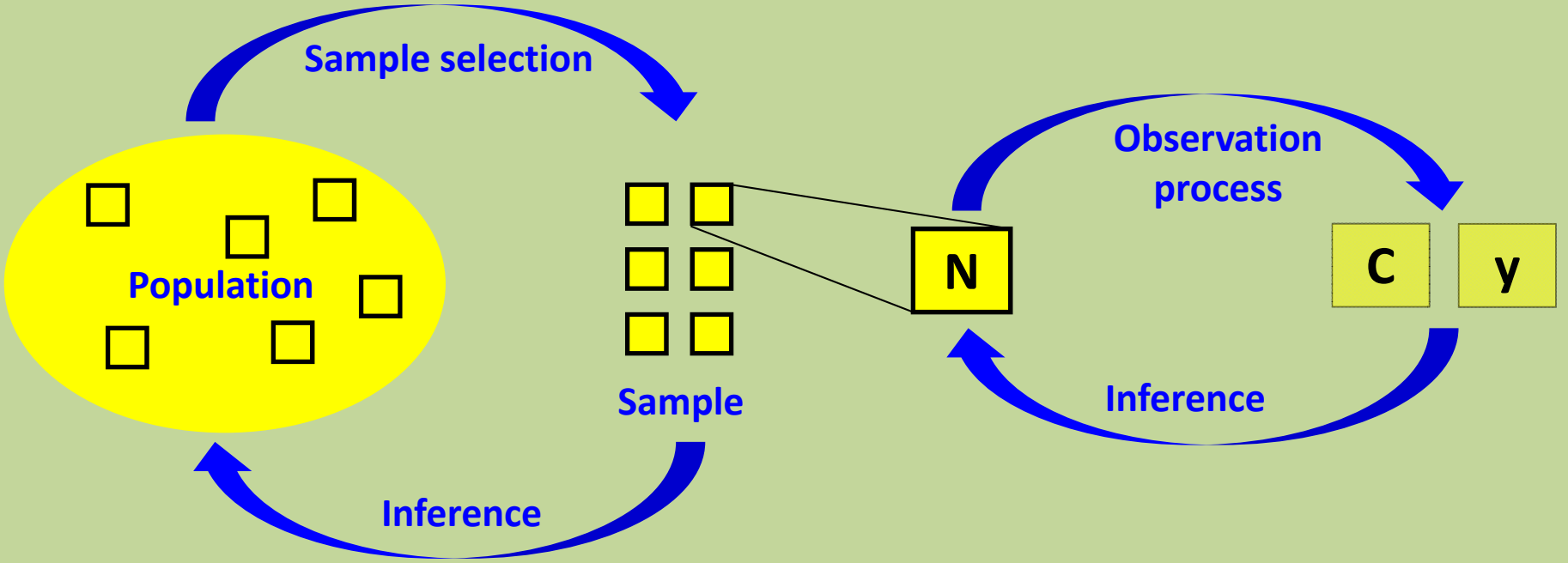
    -> false positive errors
- induces non-ignorable missing values (MNAR)
- Biased inference unless missingness process modelled
- Missingness = measurement process = observation process
- **Measurement error processes can be important for all analyses in ecology and management**

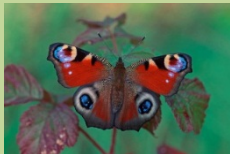


The ugly truth about all distribution & abundance data

Spatial sampling process

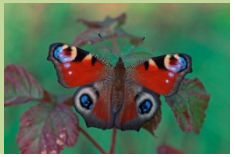
Measurement error process





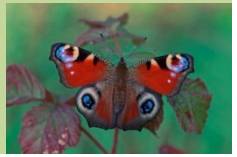
Contents

- Why do we care about distribution and abundance ?
- What is occurrence/distribution ?
- The ugly truth about all distribution & abundance data
- **Kinds of data on distribution and abundance**
- Value of hierarchical models for ecological data



Kinds of data on distribution and abundance

- Presence-only data:
  - invent pseudo-zeroes and do logistic regression (bad)
  - use background sample and do MaxEnt or Maxlike (less bad)
  - treat as spatial point pattern and do SPP modeling,
    e.g., R package spatstat (difficult: assumptions, statistics),
    also powerful INLA quasi-R-package
- Presence/absence: variant of logistic regression: GLM, GAM, BRT
- Presence/absence data replicated in time: site-occ. modeling
- Count data, with or without replication in time: Poisson GLM or Nmix modeling
- **Note:** presence/absence is misleading term,
  used for convenience only



Value of hierarchical modeling for ecological modeling

- Hierarchical models natural framework for ...
- ... *thinking* about processes underlying all ecological field data; ecological and measurement
- ... *describing*, in a model, these processes
- ... that is, for **mechanistic modeling** in ecology
- and, to *analysing* ecological data on distribution and abundance, while accounting for spatial sampling and (non-ignorable) measurement error processes