Uncertainty in Ecological Analysis

Noel Cressie

Director, Program in Spatial Statistics

and Environmental Sciences

The Ohio State University

(www.stat.ohio-state.edu/∼sses)

The Workshop

- Science
  - MBI Organizing Committee, 2005-6 Year of Ecology and Evolution
  - Workshop Organizing Committee: Kate Calder, Jim Clark, Noel Cressie (Chair), Jay Ver Hoef, Chris Wikle
  - You (Ohio State, USA Academic/Gvt, Canada, UK, France, Iceland, Brazil, Australia, New Zealand)

The Workshop, ctd.

- Finances/Visas/Facilities/Computing (MBI)
  - Avner Friedman, Director
  - Tony Nance, Assistant Director
  - Nikki, Matt, Michael, Becca, Stella

The Workshop, ctd.

- Finances/Visas/Facilities/Computing (MBI)
  - Avner Friedman, Director
  - Tony Nance, Assistant Director
  - Nikki, Matt, Michael, Becca, Stella
- Everything Else (SSES)
  - Terry, Kate, Noel

The Workshop, ctd.

- Discussion, Discussion, Discussion!
  - Shorter Talks (Speakers'/Chairs' responsibility)
  - 2 Discussants Per Session (3-5 mins. each)
  - Lots of Time for General Discussion
  - Ecological Study Discussion (6 separate groups)
  - Panel Session on Last Day
  - Healthy Debate Please!

The Workshop, ctd.

- Party, Party, Party!
  - Opening Reception This Evening (5 p.m., 7th floor of Math Tower)
  - SSES Sponsored Reception on Tuesday (6-8:30 p.m., OSU Wetland)
  - Workshop Dinner on Wednesday (Holiday Inn: 6-7 p.m., Cash Bar; 7 p.m., Dinner)
  - Lunch/Reception on Thursday (Holiday Inn: 12 noon)

Sources of Uncertainty

- Data, Data, Data!

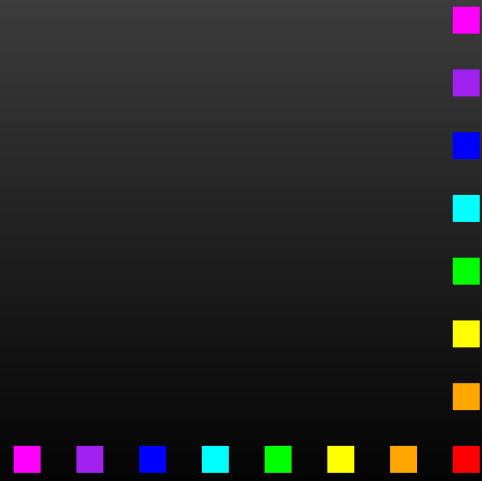
Sources of Uncertainty

- Data, Data, Data!
- Models
  - Stochastic or deterministic
  - Stochastic: Some components modeled as random; allows simple description of incomplete science knowledge
  - Parametric classes of models

Sources of Uncertainty, ctd.

- Parameters
  - Fixed (but unknown)?
  - Random?
  - We all want to make inference on unknown parameters, but how?

Topics to Avoid at Parties (But not at this Workshop!)



Topics to Avoid at Parties (But not at this Workshop!)

- Sex - MBI Year of Ecology and Evolution

Topics to Avoid at Parties (But not at this Workshop!)

- Sex - MBI Year of Ecology and Evolution
- Money - MBI's generosity has made this moot

Topics to Avoid at Parties (But not at this Workshop!)

- Sex - MBI Year of Ecology and Evolution
- Money - MBI's generosity has made this moot
- Politics - Impacts funding of research; panel/general discussion

Topics to Avoid at Parties (But not at this Workshop!)

- Sex - MBI Year of Ecology and Evolution
- Money - MBI's generosity has made this moot
- Politics - Impacts funding of research; panel/general discussion
- Model Parameters - Fixed or random?

Parameters: Fixed or Random?

- Philosophical - comes from experience with different types of problems and from beliefs!

Parameters: Fixed or Random?

- Philosophical - comes from experience with different types of problems and from beliefs!
- Tutorial-Style presentations this AM
  - Classical (Fixed Parameters): Byron Morgan
  - Bayesian (Random Parameters): Mark Berliner

Parameters: Fixed or Random?

- Philosophical - comes from experience with different types of problems and from beliefs!
- Tutorial-Style presentations this AM
  - Classical (Fixed Parameters): Byron Morgan
  - Bayesian (Random Parameters): Mark Berliner
- Group sessions on Mon, Tue, Wed
  - Ecological Study: Jay Ver Hoef
  - 6 Groups
  - Group Leaders summarize results (5 min. ea.) at Workshop Dinner on Wed.

Ecology

- The study of organisms and how they interact with their surroundings

Ecology

- The study of organisms and how they interact with their surroundings
- It helps to keep the word “environment” to mean “the surroundings”. (Ecology and environment often confused in public arena)

Ecology

- The study of organisms and how they interact with their surroundings
- It helps to keep the word “environment” to mean “the surroundings”. (Ecology and environment often confused in public arena)
- Political statement! Why does the study of *Homo sapiens* receive enormous funding support, as if it can exist independently of other organisms?

Ecology

- The study of organisms and how they interact with their surroundings
- It helps to keep the word “environment” to mean “the surroundings”. (Ecology and environment often confused in public arena)
- Political statement! Why does the study of *Homo sapiens* receive enormous funding support, as if it can exist independently of other organisms?
- Earth could be viewed as the organism and we only have one replicate (to get it right)!

Ecology, ctd.

- Ecology is inherently *spatial* and *temporal*

Ecology, ctd.

- Ecology is inherently *spatial* and *temporal*
- Sometimes one or other is aggregated

Ecology, ctd.

- Ecology is inherently *spatial* and *temporal*
- Sometimes one or other is aggregated
- Temporal - Dynamic models, expressing current states in terms of past states, allow explanation (causation) and forecasting

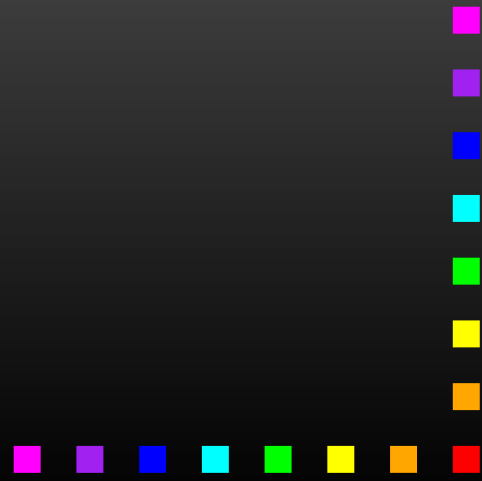
Ecology, ctd.

- Ecology is inherently *spatial* and *temporal*
- Sometimes one or other is aggregated
- Temporal - Dynamic models, expressing current states in terms of past states, allow explanation (causation) and forecasting
- Spatial - Dynamics can be spatially inhomogeneous! In other words, “why” can depend on “where”. (MBI Spatial Ecology Workshop was last month)

Spatial Support

Uncertainty due to lack of knowledge. Some sources:

- Science (incomplete or incorrect models)



Spatial Support

Uncertainty due to lack of knowledge. Some sources:

- Science (incomplete or incorrect models)
- Incomplete data (missing variables and missing observations on variables)

Spatial Support

Uncertainty due to lack of knowledge. Some sources:

- Science (incomplete or incorrect models)
- Incomplete data (missing variables and missing observations on variables)
- Measurement (noise)

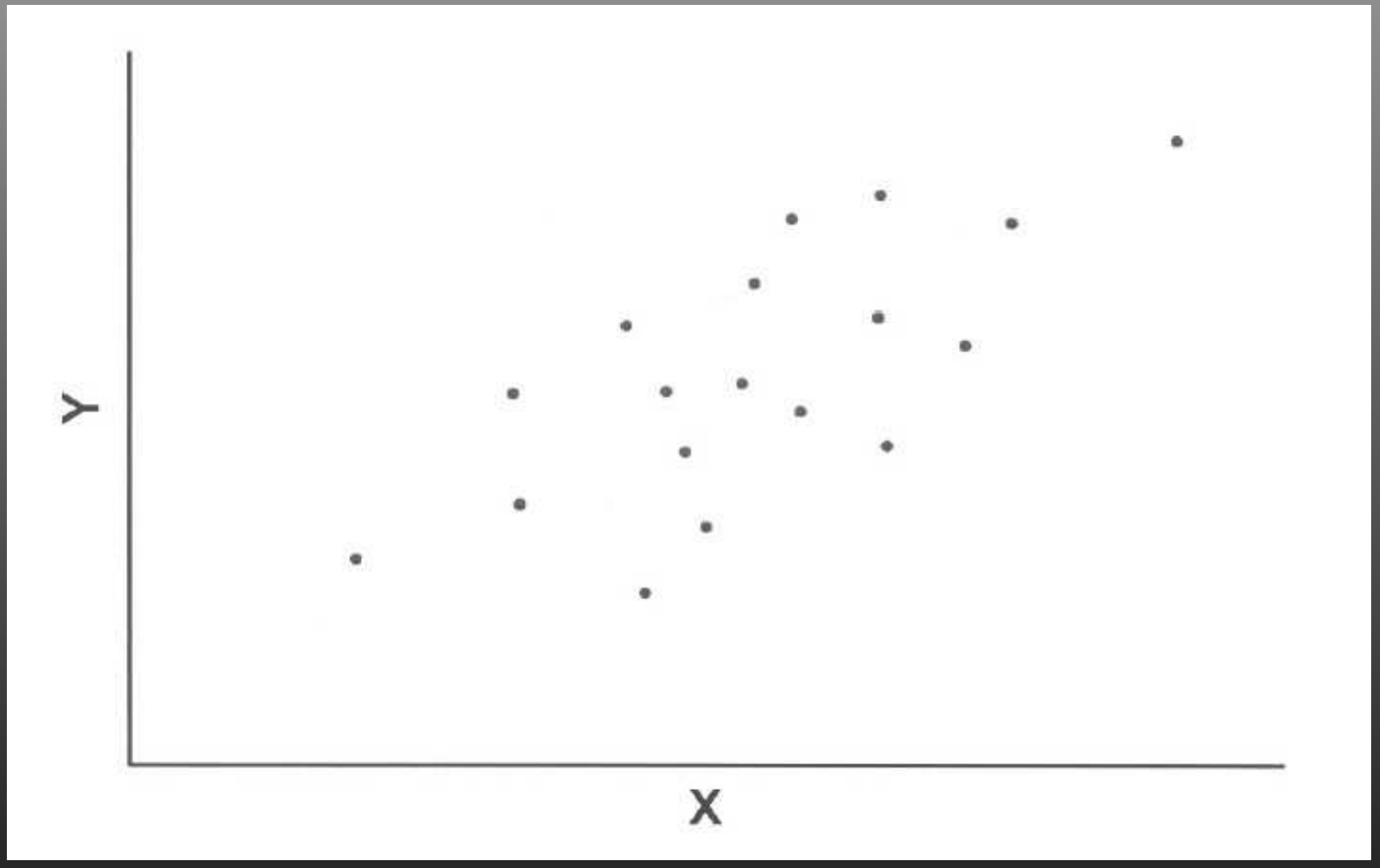
Spatial Support

Uncertainty due to lack of knowledge. Some sources:

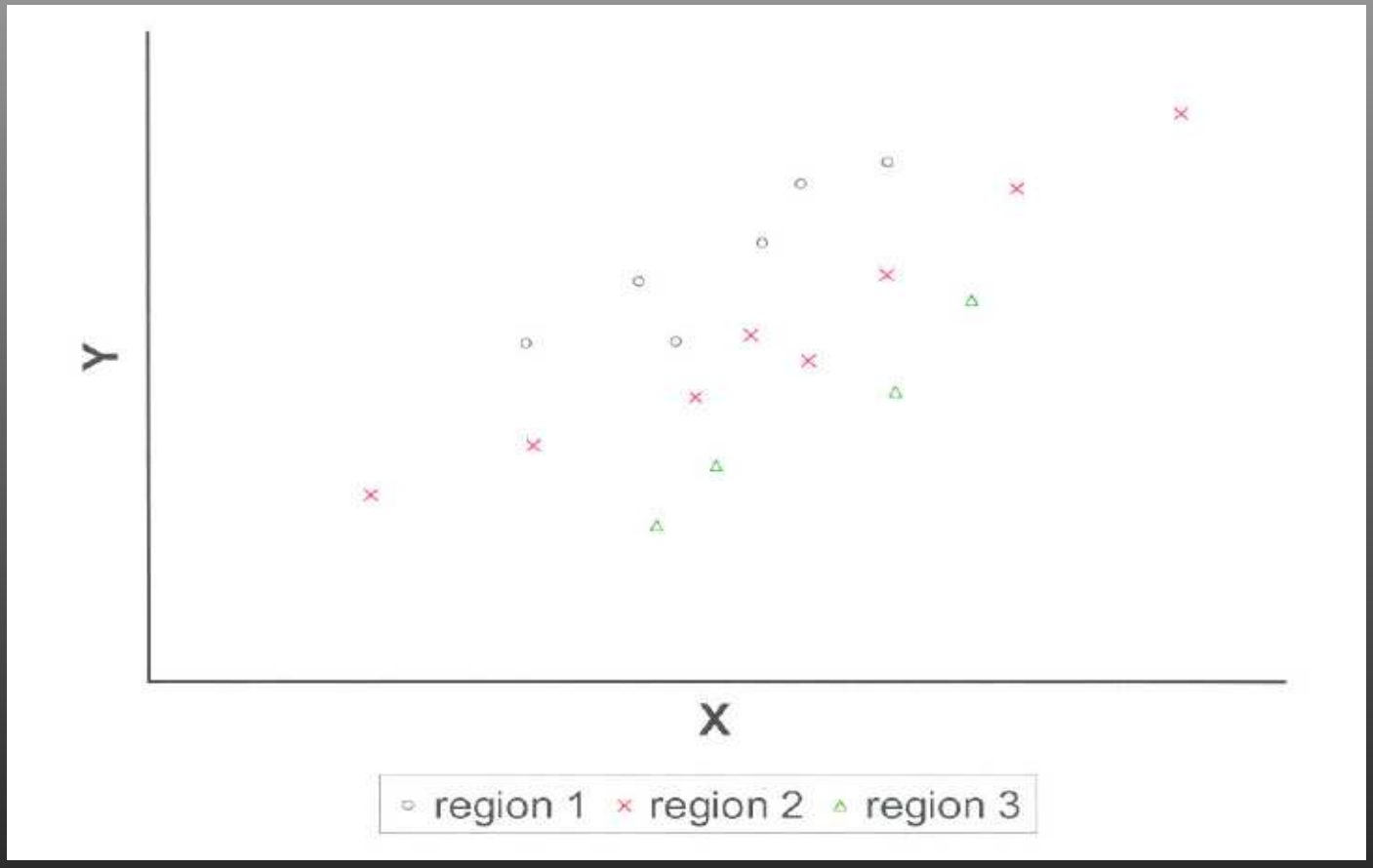
- Science (incomplete or incorrect models)
- Incomplete data (missing variables and missing observations on variables)
- Measurement (noise)
- Spatial uncertainty

Spatial Support, ctd.

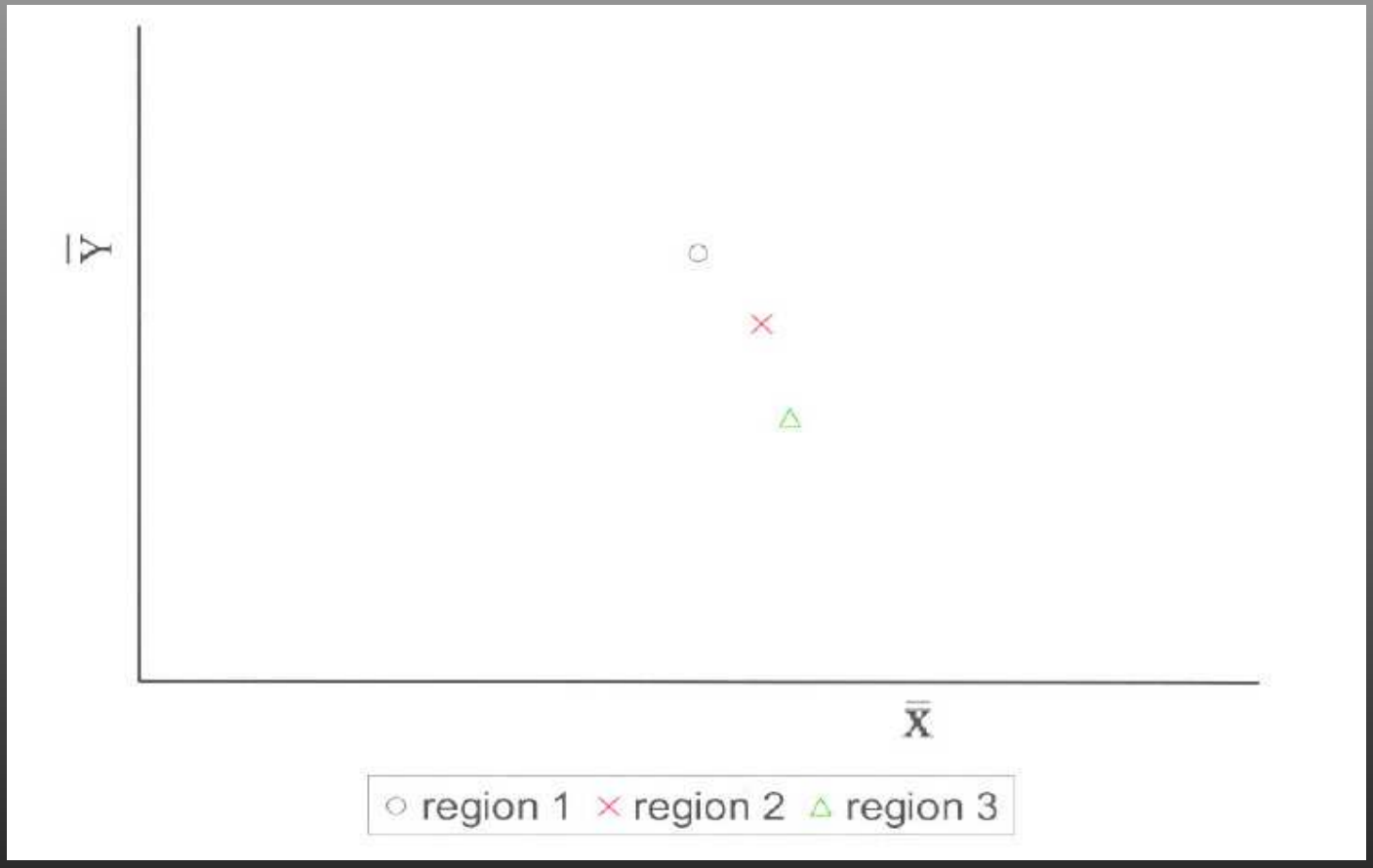
- Spatial uncertainty
  - Object view versus field view
  - Location error
  - Aggregation (COS, MAUP, ecological bias)
    This is, in effect, Simpson's Paradox in a spatial setting



$$Y = \beta_0^* + \beta_1^* X + \varepsilon\,; \quad \widehat{\beta}_1^* > 0$$



$$Y = \beta_0^* + \beta_1^* X + \varepsilon\,; \quad \widehat{\beta}_1^* > 0$$



$$\overline{Y} = \beta_0^* + \beta_1^* \overline{X} + \overline{\varepsilon}\,; \quad \widehat{\beta}_1^* < 0$$

Ecological Inference

Deducing *individual* behavior from *aggregate* data

Robinson (1950) *American Sociological Review*:

1930 data from lower 48 states:

$$\begin{aligned} Z &= \% \text{ foreign born in a state} \\ Y &= \% \text{ literate in a state} \\ &\Rightarrow \text{ corr.} = 0.53 \end{aligned}$$

Based on sampling, Robinson found individual corr.$= -0.11$

Thus, corr. value of 0.53 represents an *ecological bias*

Approach

- Use statistical modeling to avoid ecological fallacy

Approach

- Use statistical modeling to avoid ecological fallacy
- Fit to data that are aggregated (counts) using model built at point level

Approach

- Use statistical modeling to avoid ecological fallacy
- Fit to data that are aggregated (counts) using model built at point level
- Have to have a good point-level model!

(Aggregated) Disease Counts

$Y(A_1), \ldots, Y(A_n)$ modeled hierarchically:

$$Y(A_i) \mid p_i \sim \text{Poi}(N_i p_i) \,, \quad \text{indep.}$$

$$\log p_i | x(A_i) = \log(E_i/N_i) + \beta_0 + \beta_1 x(A_i) + \delta_i \,,$$

where $x(A_i)$ is an exposure variable

δ_i is a (spatial) error term.

Finally, $\{x(A_i)\}$ is potentially a spatial process.

(Aggregated) Disease Counts, ctd.

Inference on $\boldsymbol{\beta} = (\beta_0, \beta_1)$ might not reflect *disease/exposure relationship*, which occurs at the individual level. We see the possibility of the *ecological fallacy*.

Let $\boldsymbol{\beta}^*$ denote *individual-level coefficients* involved in disease/exposure relationship.

(Individual-Level) Disease Mechanism

$$Y_{ij} = \begin{cases} 1 & j\text{-th individual in area } A_i \text{ has disease} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \Pr(Y_{ij} = 1) &= \int \Pr(Y_{ij} = 1 | x_{ij}, \boldsymbol{\beta}^*) f(x_{ij} | \phi_i) dx_{ij} \\ &\equiv p(\boldsymbol{\beta}^*, \phi_i) \,, \end{aligned}$$

where $f(\cdot|\phi_i)$ is the *within-area distribution* of exposure within A_i.

Disease Mechanism (cont.)

$\Rightarrow Y_{ij}|\boldsymbol{\beta}^*, \phi_i \sim$ Bernoulli $(p(\boldsymbol{\beta}^*, \phi_i))$

$\Rightarrow Y(A_i)|\boldsymbol{\beta}^*, \phi_i \sim \text{Bin}(N_i, p(\boldsymbol{\beta}^*, \phi_i))$

$\Rightarrow Y(A_i) \dot{\sim} \text{Poi}(N_i p(\boldsymbol{\beta}^*, \phi_i))$

That is,

$$Y(A_i) \dot{\sim} \text{Poi}(\lambda(A_i; \boldsymbol{\psi}))$$

Loglinear Individual-Level Model

$$
\begin{aligned}
\Pr(Y_{ij} &= 1|x_{ij}, \boldsymbol{\beta}^*) = \exp(\beta^*_{0,i} + \beta^*_1 x_{ij}) \\
\Rightarrow p(\boldsymbol{\beta}^*, \phi_i) &= \exp(\beta^*_{0,i}) \int \exp(\beta^*_1 x_{ij}) f(x_{ij}|\phi_i) dx_{ij} \\
&\neq \exp(\beta^*_{0,i}) \exp(\beta^*_1 \int x_{ij} f(x_{ij}|\phi_i) dx_{ij}) \\
&\equiv \exp(\beta^*_{0,i} + \beta^*_1 x(A_i)) \,,
\end{aligned}
$$

Clearly, the use of sample-survey estimated covariate $\{\widehat{x}(A_i)\}$ results in an *ecological bias*.

Likelihood based on case counts

- Data are counts

$$(Y(A_1), \ldots, Y(A_n))$$

- Same intensity and Poisson assumptions
- Likelihood of ψ:

$$\ell(\boldsymbol{\psi}) = \prod_{i=1}^{n} \frac{e^{-\lambda(A_i;\boldsymbol{\psi})}(\lambda(A_i;\boldsymbol{\psi}))^{Y(A_i)}}{Y(A_i)!}$$

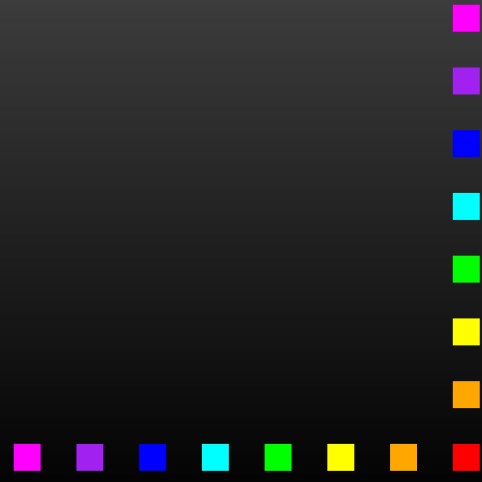
Likelihood based on case counts, ctd.

Parameters $\psi = (\beta_1^*, \{\beta_{0,i}^*\}, \{\phi_i\})$ are defined at individual level but inference is based on aggregated data (counts),

$$(Y(A_1), \ldots, Y(A_n))$$

Ecological Analysis

This workshop is about



Ecological Analysis

This workshop is about
... trying to be certain about the ecological problem's uncertainties ...