



Introduction to hierarchical models

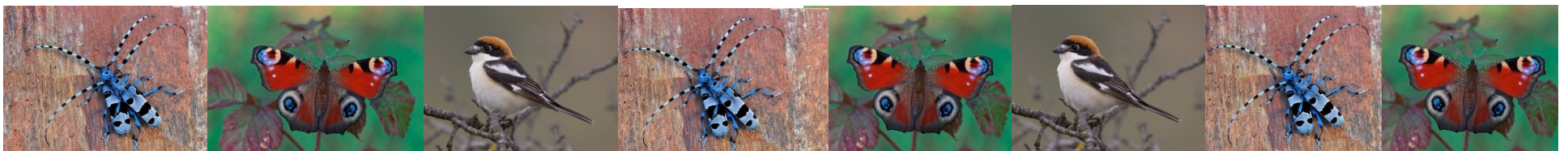
Marc Kéry

marc.kery@vogelwarte.ch

Aberdeen University

Scotland

25–29 June 2018



Outline of talk

- Variability, "random things"
- How to describe random "things" --- random variables
- Hierarchical models (HMs) ? --- Why are they useful/beautiful ?
- Components of HMs: GLM; what are they ?
- Heart of a GLM: linear model; how do we describe it ?
- What are random effects ? --- Why/when do we specify parameters as random variables ?
- Modeling measurement error



Variability

- Law of Life/Biology: EVERYTHING varies !
- e.g., the number of individuals at a site/at a time (population size), the number of species, whether an individual survives or dies, whether it reproduces or not, ...
- Variability causes uncertainty or (partial) unpredictability
- Unpredictable or chance element in ANYTHING that we might want to study or observe in the world
- Hence, great importance of understanding of statistical principles in science and also in everyday life



Variability

- e.g. the number of asp vipers (*Vipera aspis*) seen at 6 sites:
5, 1, 3, 0, 0, 2



Variability

- e.g. the number of asp vipers (*Vipera aspis*) seen at 6 sites:
5, 1, 3, 0, 0, 2
- want to describe (explain) these numbers, both in terms of mean/average patterns, but also in terms of variation/noise



Description of variability: random variable (r.v.)

- in statistics, **use probability to describe random variation**
- random variable (r.v.): chance process that produces some data (usually as *observed* outcome, though sometimes *hidden*)
- description of a r.v.: continuous: probability density function (pdf) --- discrete: probability mass function (pmf)
- pdf/pmf: "rule" that gives probability ("relative frequency") of observing any among an exhaustive set of all possible observations (i.e., sums/integrates to 1)
- pdf/pmf have key descriptors that govern specific form = parameters
- best known rv's/pdfs/pmfs: Normal distribution, Poisson distribution, Binomial/Bernoulli distribution



A Poisson random variable

- e.g., the number of asp vipers observed at a site

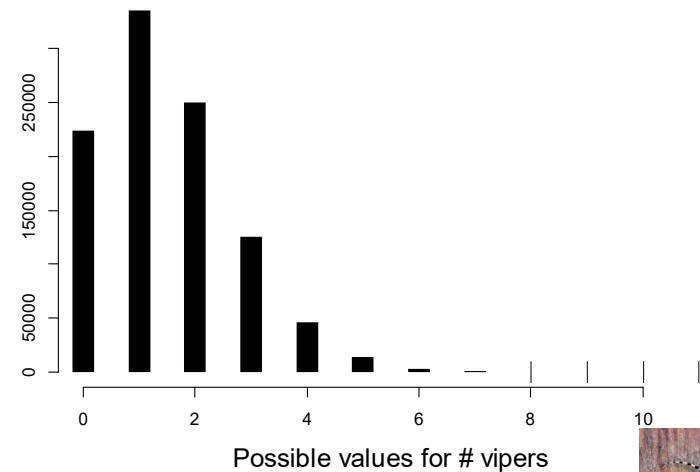
- pmf:
$$p(y \mid \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}$$

- "key descriptor": parameter λ
- two shorthands:

$$y \sim \text{Poisson}(\lambda)$$

$$\text{glm}(y \sim 1, \text{family} = \text{'poisson'})$$

- one possible picture (with $\lambda = 1.5$):



Parametric statistical models

- parametric statistical model: description of one or more random variables
- hence, can say statistical model = pmf/pdf
- statistical inference: use observed data from some putative random process (= pdf) to estimate parameters of the process that produced the data
- best known statistical models: generalized linear models (GLMs): pdf's for certain (very broad) class of random variables; see later



Hierarchical model: sequence of random variables

- Hierarchical model (HM): sequence of random variables (rv's), observed (y) or unobserved (x)

$$\begin{array}{l} x \sim f(\omega) \\ y \sim g(x, \theta) \end{array}$$

- Factorization of joint distribution $[x, y]$ into marginal ($[x]$) and conditional distribution, $[y|x]$:

$$[x, y] = [x] [y|x]$$

- Estimands: parameters ω and θ ; latent variables x
- x : latent variables = unobserved rv's = random effects**
- Large class of models can be described as HMs



Hierarchical model: sequence of random variables

- HMs powerful manner of building statistical models
- **Typically reflect our thinking about mechanistic structure of process studied: science-based statistical modeling**
- Describe stochastic processes with multiple layers/levels of variability and uncertainty
- (Note: variability vs. uncertainty)
- HMs can be fitted using classical (i.e., likelihood) methods (e.g., R packages **lme4**, **unmarked**, **secr**, ...)
- Bayesian analysis often more straightforward
- BUGS language ideal for fitting HMs: indeed, BUGS software developed especially for HMs !



Models naturally described hierarchically in BUGS language



Why hierarchical models ?

- HMs make the fitting of complex models easier (especially bayesianly)
- ... foster mechanistic, cleaner thinking about a problem
- ... lead to mechanistic modeling
- ... are **science-based statistical models**
- ... foster synthetic understanding of large array of models
- ... enable honest accounting for all components of variability and uncertainty in a system
- ... ideal for rigorous modeling of measurement-error processes in ecological field data
- great for combining information ('integrated models', meta-analyses)



HMs foster mechanistic thinking about system and synthetic understanding of large class of models

- e.g., state-space model for time-series of observations (y)

Ecological process

$[y] =$ initial system state z_0 *

system state z_{t+1} given system state z_t *

observation y given system state z

Measurement-error process,
observation process

- y counts at 1 site:
state-space model (BPA chap. 5)
- y individual detection histories, multiple individuals:
JS model (BPA chap. 10)
- y site-detection histories, multiple sites:
dynamic occupancy model (BPA chap. 13)



HMs foster synthetic understanding of models

- **moreover ...**

[y] = initial system state z_0^*

system state z_{t+1} given system state z_t^*

observation y given system state z

- y counts, no dynamics:
random-effects models, GLMMs (BPA chap. 4)
- y counts, multiple sites:
Dail-Madsen Nmix model (BPA chap. 12)
- y individual detection histories, no dynamics:
closed-population model (BPA chap. 6)
- y individual detection histories, initial state not modelled:
CJS model (BPA chap. 7), multi-state models (BPA chap. 8)
- y site-detection histories or counts, multiple sites, but no dynamics:
static occupancy model (BPA chap. 13), static N-mixture model (BPA chap. 12)



vogelwarte.ch



Components of HMs: Generalized linear models (GLMs)

- HMs represent description of two or more random variables
- How do we describe structure in random variables ?
--- > typically as GLMs
- **Can say that HMs consist of a sequence of 2 or more GLMs !**
- **GLMs typically represent the modules in HMs**



Generalized linear models (GLMs)

- Outcome of a random variable y (response); explanatory variable(s) x
- Explain variability in mean response as a linear model:
 - > weighted sum of explanatory variables
- Linear model in terms of x applied to transformation g of mean response $E(y)$,
 - i.e., $g(E(y)) = \text{some linear model in } x$
- Three components of GLM:
 - (1) random part of response: statistical distribution
 - (2) link function g
 - (3) systematic part of response (linear predictor):
 - some linear function of x



Three most frequent GLMs

- Normal response:

Random part:

$$y \sim \text{Normal}(\mu, \sigma^2)$$

(typical) Link function:

none (“identity”)

Systematic part:

SLM (some linear model)

- Poisson response:

Random part:

$$y \sim \text{Poisson}(\lambda)$$

(typical) Link function:

log

Systematic part:

SLM

- Binomial response:

Random part:

$$y \sim \text{Binomial}(p, N) = N * \text{Bernoulli}(p)$$

(typical) Link function:

$$\text{logit} = \log(p / (1-p))$$

Systematic part:

SLM



Importance of linear models in GLMs and HMs

- statistical model as a description of a random variable:

response \sim deterministic part + stochastic part

e.g. $y_i = \alpha_i + \varepsilon_i$

with $\varepsilon_i \sim \text{Normal}(0, \sigma^2)$

- stochastic/random part: statistical distribution for ε_i
- deterministic/systematic part: non-linear or linear function α_i
- non-linear may be more mechanistic and better for extrapolation, but (much) more difficult to fit
- linear model by far easiest for deterministic part of response
- For hierarchical modeling, must understand GLM
- For GLM and HMs, must understand linear modeling
- see Chapter 3 in BPA; chap 3 in AHM1, also Kéry (2010)



Types of descriptions of linear model in GLM

- words: e.g., “A and x act additively”
- graphs: e.g., lines and bars
- specific labels: e.g., ANOVA, ANCOVA, t-test, regression
- algebra: e.g., $y_i \sim Normal(\alpha_j + \beta * x_i, \sigma^2)$
- matrix algebra: e.g., $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$
- system of equations
- R language: e.g.,
$$\text{lm}(y \sim A + x)$$
- BUGS language: e.g.,
$$y[i] \sim \text{dnorm}(\mu[i], \tau)$$
$$\mu[i] \leftarrow \alpha[A[i]] + \beta * x[i]$$
- alternative, equivalent descriptions: parameterisations



Specification of linear models in BUGS language

- almost exactly as we do in algebra !
- one of the main stumbling blocks of becoming a BUGS modeler at first
- **HOWEVER:** wonderful side effect that finally *enforces* an understanding of linear models fit by issuing R commands such as `lm(y ~ A*x)`



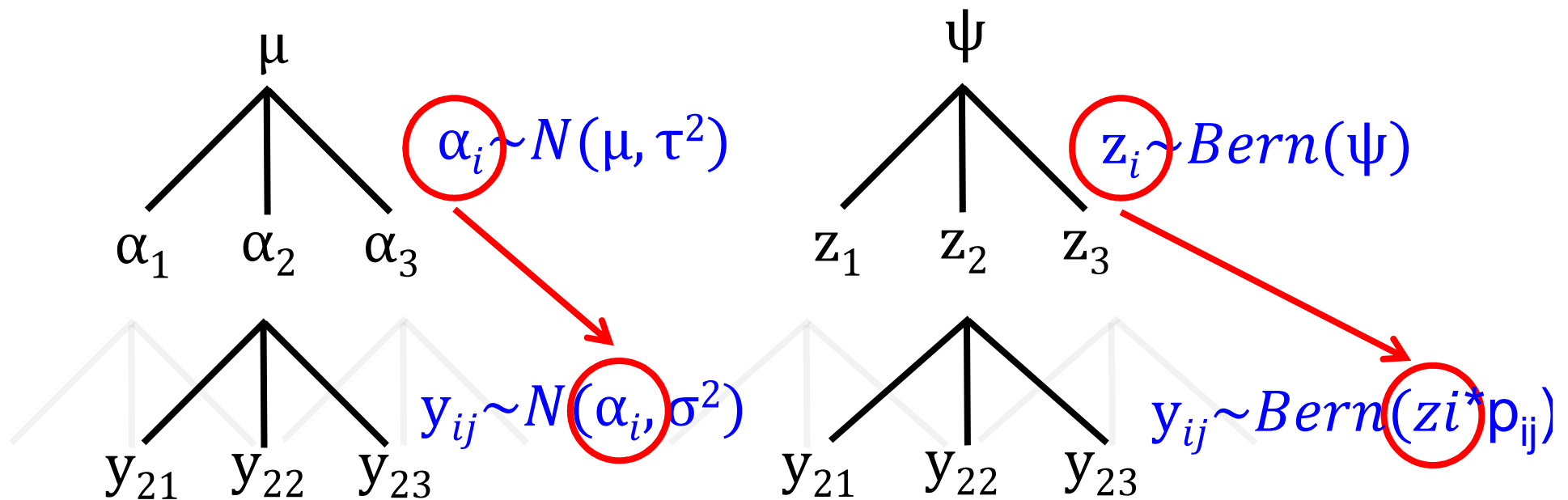
Some important linear models in glm/lm and BUGS

No.	Model in R	Model in algebra	Model in BUGS	Traditional name of technique based on that linear model	Number of coefficients	Meaning
1	1	α	alpha	"Model of the mean"	1	Constant term (intercept) only
2	pop	$\mu + \alpha_j$	mu + alpha[pop[i]] alpha[1] <- 0	One-way ANOVA	3	This is the default linear model parameterisation of a factor in R, which has an intercept (= the value for the first population) and two constants that are the differences between the values of population 2 and 1 and 3 and 1. In BUGS, the first level of the vector alpha must be manually set to zero to avoid overparameterisation. In R this is done automatically.
3	pop-1	α_j	alpha[pop[i]]	One-way ANOVA	3	This is a simple reparameterisation of model 2. There are three constants, one for each population (called the level of the factor). Called a t-test if factor has only two levels. In R this model is specified by "subtraction" of the intercept.
4	body	$\alpha + \beta * x_i$	alpha + beta * body[i]	Simple linear regression	2	An intercept plus a slope, common to all three populations (i.e., no effect of pop)
5	body-1	$\beta * x_i$	beta * body[i]	Simple linear regression through the origin	1	"Subtract the intercept": because body is a continuous covariate, this is NOT a mere reparameterisation of model 4. Regression through the origin; not usually a meaningful model.
6	pop+body	$\alpha_j + \beta * x_i$	alpha[pop[i]] + beta * body[i]	Main-effects ANCOVA	4	One separate intercept for each population and a common slope
7	pop*body	$\alpha_j + \beta_j * x_i$	alpha[pop[i]] + beta[pop[i]] * body[i]	Interaction-effects ANCOVA	6	Three separate intercepts and three separate slopes. That is, fully separate regression of wing on body for each population.



What are random effects ?

- **Hierarchical models: Dependent sequence of random variables (observed & unobserved)**
- e.g., randomised block ANOVA, „hierarchical logistic regression“



- Many other (similar) names: state-space, random-, mixed-effects, latent variables, unobserved components, **mixed models** ...



What are random effects ?

- **Random effects:**
Realized values of unobserved random variables in a HM !
- two or more latent variables/parameters that are given a (prior) distribution; **that's all !**
- prior distribution assumed Normal in all classical GLMMs
- **BUT:** can be non-normal, e.g., Poisson (discrete) or even Bernoulli (discrete and binary): cf. Nmix and site-occ. models
- can estimate parameters of prior distribution (also called mixing distribution), e.g., τ^2 or ψ , **and** realized values α and z !
- parameter, prior/mixing distribution, hyperparameters, hyperpriors, ...

$$\alpha_i \sim N(\mu, \tau^2)$$
$$y_{ij} \sim N(\alpha_i, \sigma^2)$$

$$z_i \sim \text{Bern}(\psi)$$
$$y_{ij} \sim \text{Bern}(z_i * p_{ij})$$



What are random effects ?

- **Random effects:**
Realized values of unobserved random variables in a HM !
- replication of study would (usually) yield different effects
- can estimate values of “new” draws from same prior distribution
- in analysis, treatment as fixed or random is your judgment; cannot be “tested”
- in Bayes, effects said exchangeable: “similar but not identical”
-> don’t treat “dissimilar” effects as random (e.g., captive and wild populations)
- (also: bad estimates when few levels, e.g. only 3–5)
- (HMs computationally more expensive)



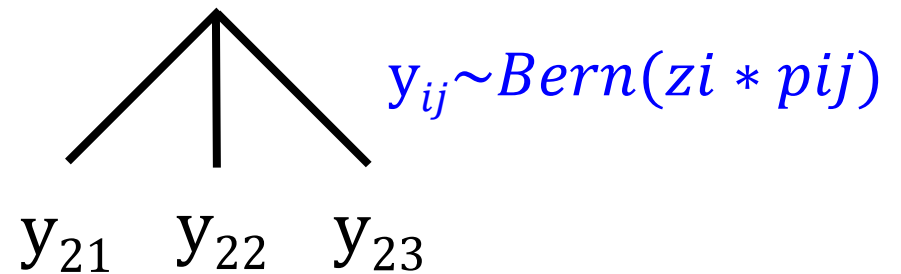
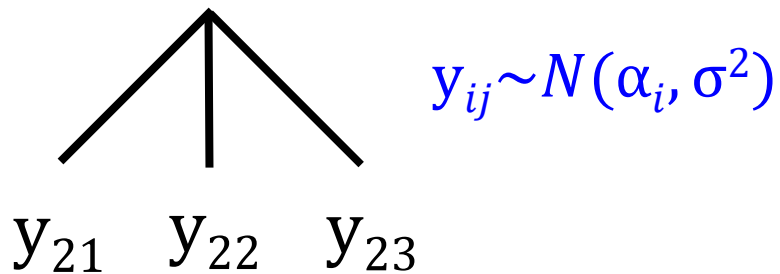
Motivations for treating as random a set of params

- extend scope of inference
- assess/partition system variability
- want to model variability among parameters
- proper assessment of uncertainty
- model random effects, e.g., as a GLM, correlations
- spatial and temporal correlations
- avoid pseudo-replication
- borrow strength/improve individual random effects estimates/shrinkage
- combine information (meta-analysis)
- others



Modeling of measurement errors in HMs

- Last level of HM before:
measurement error = observation process



- Qualitatively very different:

(1) cancel out

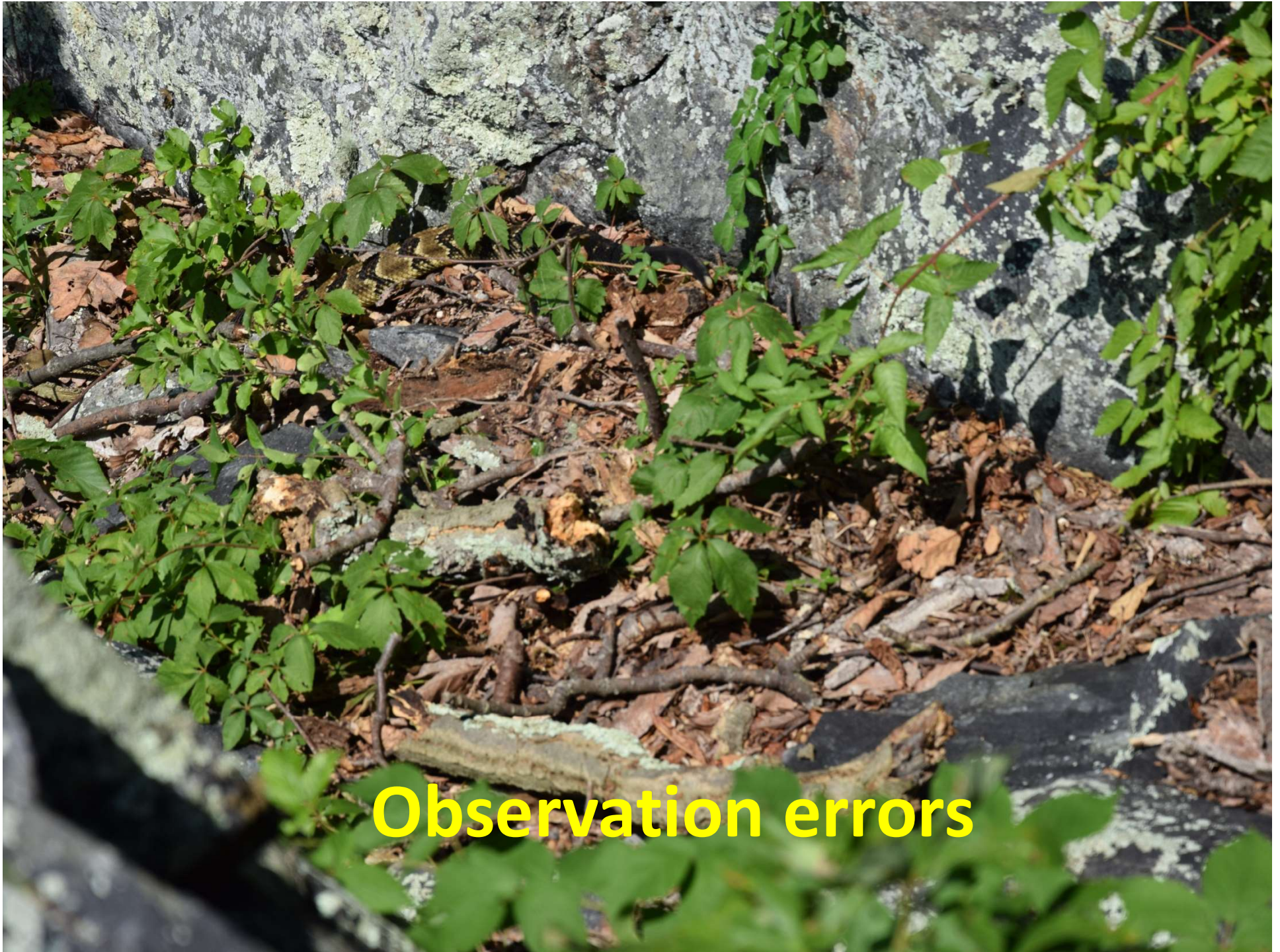
$$E(y_{i.}) = \alpha_i$$

(2) do **NOT** cancel out

$$E(y_{i.}) = z_i * p_{ij} \neq z_i$$

=> Should care for measurement errors





Modeling of measurement errors in HMs

- What do you see ?



Importance of linear models and random effects

=> Your major interaction with a HM/IPM will be to specify linear models and making some factors random

