Bayesian Hierarchical Models – an introduction

Kate Searle

Motivation for using hierarchical Bayes…

Why Bayes?

- All scientific models are abstractions

- Because models are abstractions and reduce the dimensions of a problem, we must deal with the elements we choose to leave out

- so assessing uncertainty is fundamental to science:
  - "process uncertainty"
  - "observation uncertainty"

Why hierarchical models?

- Allow us to decompose complex, high dimensional problems into parts that can be thought about and analysed individually

- Broad and flexible approach, allowing us to tackle virtually any ecological problem

Ecological systems are fundamentally hierarchical in nature…

Ecological variation – levels of biological organisation

Individuals

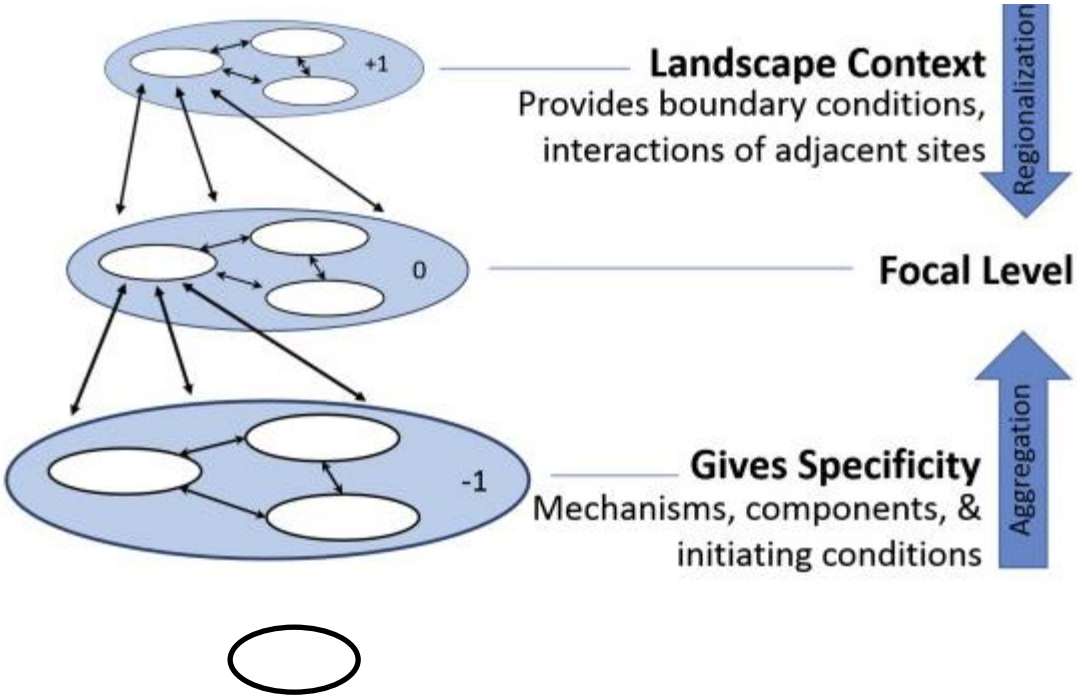
Spatial and temporal heterogeneity

Scale

Populations

Communities and Ecosystems

Ecological variation – spatial scales



Landscape

Habitat type

Plots

Individual plants

Figure 1. Scales of organization showing feedback and feedforwards from the immediate higher and lower scale in a hierarchical system. Adapted from Urban et al.[8] and O'Neill et al.[10]

Adapted from Salley et al. 2016. Rangelands 38(6):313—317 doi: 10.1016/j.rala.2016.10.003

Ecological research commonly has to deal with issues such as:

- Variation among individuals (e.g., location or genotype)

- Ecological processes operating at more than one spatial scale (plot → habitat type) or level of ecological organisation (individuals → populations → communities)

- The need to accommodate uncertainty arising from modelling a process ***as well as*** uncertainty derived from imperfect observations

- Dealing with changes in state that cannot be observed directly (age transitions of individuals that are hard to observe)

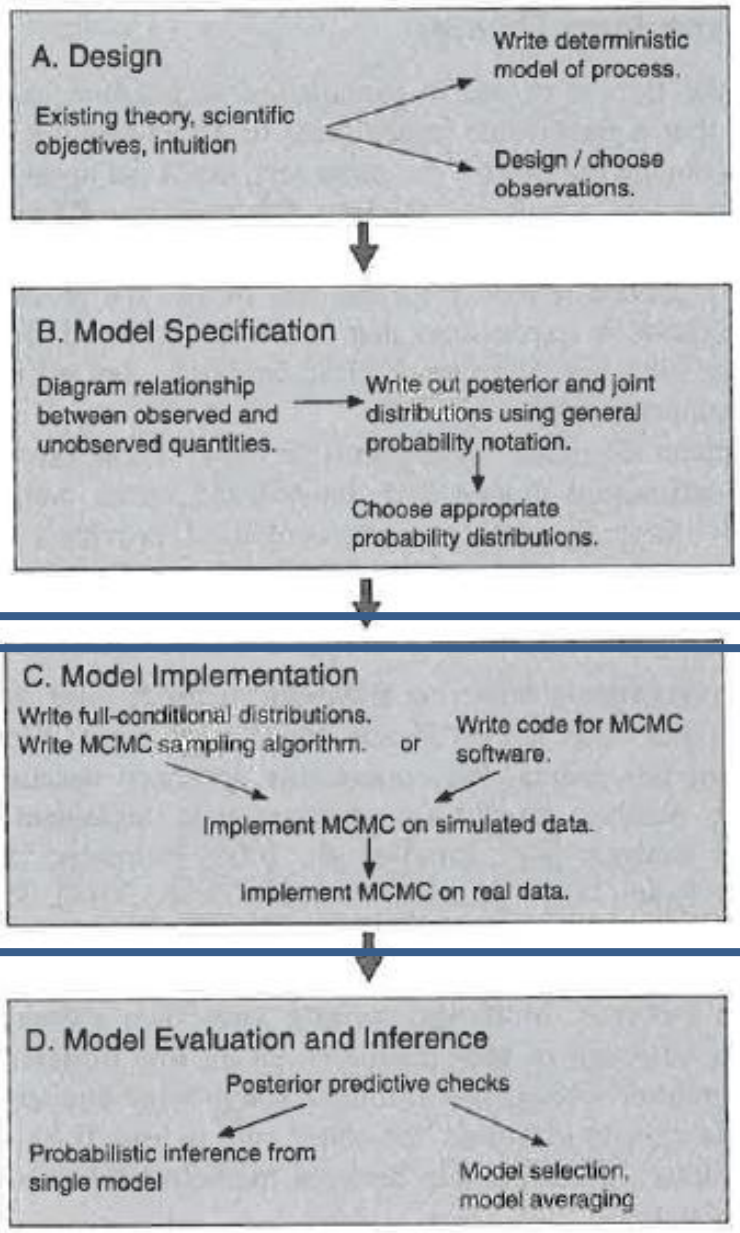
Hierarchical models provide a natural framework for addressing all of these issues

Ecological systems are fundamentally hierarchical in nature…

…so it follows that our models should be hierarchical, also

A framework for thinking about and fitting Bayesian hierarchical models in ecology…

Modelling in Ecology using Bayes



Define question, think about ecological process and observations

Define the model to represent the ecological process and observations

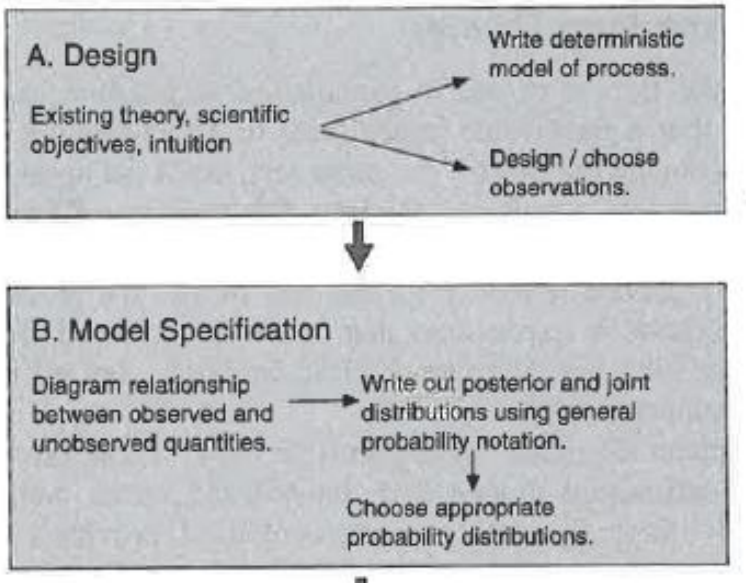
- Diagram it
- Write out posterior and joint distributions
- Choose probability distributions

Implement the model using your expression for the posterior and joint distributions

Perform model checking

Hobbs & Hooten (2015)

Part 1: defining our model and its relationship to our data

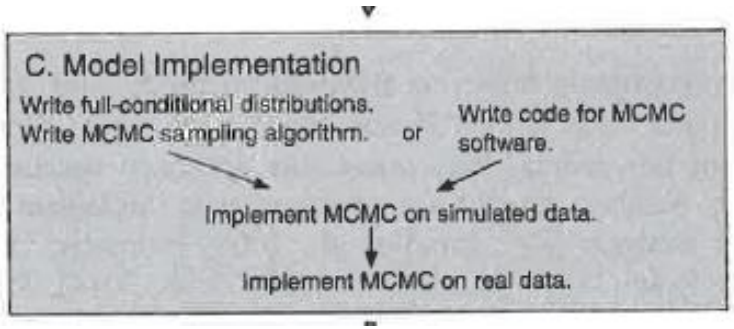


Define question, think about ecological process and observations

Define the model to represent the ecological process and observations

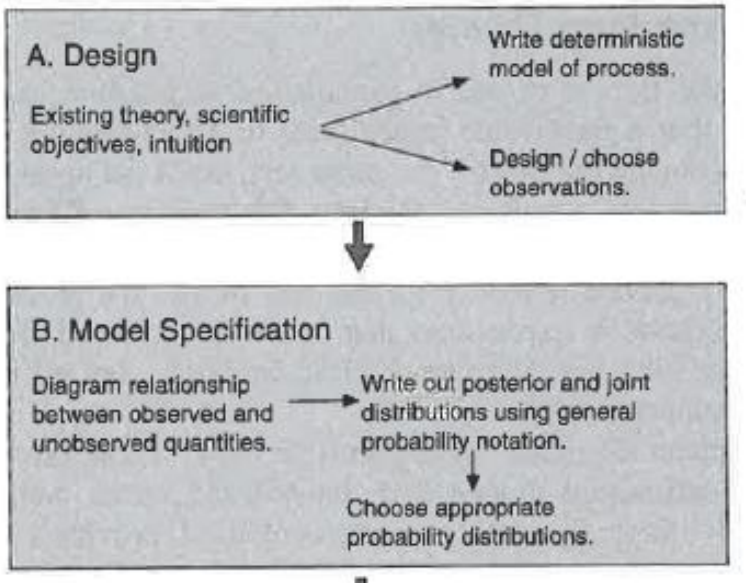
- Diagram it
- Write out posterior and joint distributions
- Choose probability distributions

Part 2: implementing our model using MCMC (R and JAGS)



Implement the model using your expression for the posterior and joint distributions

Part 1: defining our model and its relationship to our data



Define question, think about ecological process and observations

Define the model to represent the ecological process and observations

- Diagram it
- Write out posterior and joint distributions
- Choose probability distributions

Start with the fundamentals…

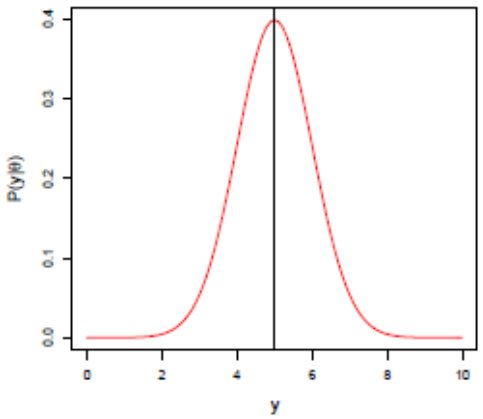
Concept to be taught	Why do you need to understand this concept?
Conditional probability	It is the foundation for Bayes' Theorem and all inferences we will make.
The law of total probability	Basis for the denominator of Bayes' Theorem $[y]$
Factoring joint distributions	This is the procedure we will use to build models.
Independence	Allows us to simplify fully factored joint distributions.
Marginal distributions	Bayesian inference is based on marginal distributions of unobserved quantities.
Statistical distributions	Our toolbox for representing uncertainty and for linking observed quantities to unobserved ones.
Moments	Basis for inference from MCMC
Moment matching	Allows us to embed the predictions of models into any statistical distribution.

From Tom Hobbs, Colorado State University

Start with the fundamentals …

Random variables

- A random variable is a quantity whose values are subject to chance
- The values it may take are governed by a probability distribution
- Bayesian inference treats all unobserved quantities as random variables
- Our goal in Bayesian modelling is to understand those distributions (*draw inference about unobserved quantities*)



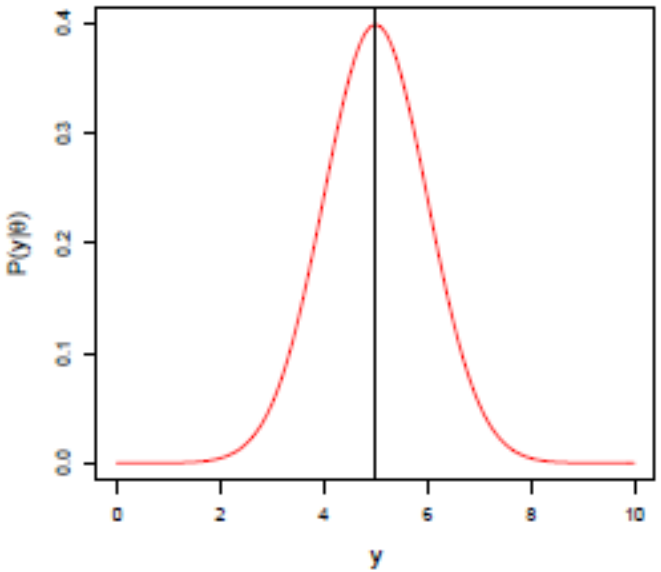
Start with the fundamentals …

Probability distributions

- The values of random variables are governed by a probability distribution

Probability model:

$y_i \sim f(\mu_i, \sigma)$, μ_i and σ^2 are parameters of the distribution $f()$



From Tom Hobbs, Colorado State University

Start with the fundamentals …

A toolbox of *f()*'s for ecological data (and later, parameters, latent states)

- Discrete
  - Poisson
  - binomial
  - negative binomial
  - Bernoulli
  - multinomial
- Continuous
  - normal
  - multivariate normal
  - lognormal
  - uniform
  - beta
  - gamma
  - Dirichlet

From Tom Hobbs, Colorado State University

Start with the fundamentals …

Our line of inference for defining models and data within the Bayesian framework…

We need a way to talk about models and data:

Models $[z|\theta]$ an ecological process of interest

State we are trying to model

Parameters of interest

$$[z|\theta] = z = a + b(x)$$

Data $[\mathrm{y}|z, \theta]$ some observations that help us model and understand the process

Data observations

Model

We learn about the process (our model) using our data

Start with the fundamentals…Bayes law

Our line of inference for defining models and data within the Bayesian framework

In Bayesian statistics, we use Bayes law to learn about our process, using the model and the data

$$[\theta|y\,] = \frac{[\mathrm{y}|\theta][\theta]}{[y]}$$

y are our observed data, which become fixed after we have observed them

θ are unobserved quantities of interest (e.g., model parameters)

We factor joint conditional probabilities to define and estimate our model…

In other words, we factor $[\mathrm{y}, \theta]$ into **ecologically sensible components** that can be estimated using MCMC as univariate distributions

Allows us to decompose complex, high dimensional problems into parts that can be thought about and analysed individually

Why is factoring joint conditional distributions so important?

Factoring joint distributions allows us to decompose complex, high dimensional problems into parts that can be thought about and analysed individually

$$p(z_1, z_2) = p(z_1 \mid z_2)\, p(z_2) = p(z_2 \mid z_1)\, p(z_1)$$

The sequence of conditioning is arbitrary, when we build models we choose a sequence that makes sense

Bayes law:

$$\overbrace{[\theta|y]}^{\text{Posterior}} = \frac{[y,\theta]}{[y]} = \frac{\overbrace{[y|\theta]}^{\text{likelihood}}\ \overbrace{[\theta]}^{\text{prior}}}{\underbrace{\int_\theta [y|\theta]\,[\theta]\,d\theta}_{\text{marginal}}}$$

Useful, ecological models will be more complex…

$$\underbrace{[\theta_1, \theta_2, \theta_3, \ldots, \theta_n, \mathbf{z}_1, \mathbf{z}_2 .. \mathbf{z}_n | \mathbf{y}_1, \mathbf{y}_2]}_{\text{Multiple parameters, latent states, multiple datasets}} \propto \underbrace{[\theta_1, \theta_2, \theta_3, \ldots, \theta_n, \mathbf{z}_1, \mathbf{z}_2 .. \mathbf{z}_n, \mathbf{y}_1, \mathbf{y}_2]}_{\text{Factor into conditional distributions}}$$

And we use MCMC to find the marginal posterior distributions of all the unobserved quantities

Fundamentally, we need to be able to write out our ecological problem as follows…

$$\overbrace{[\theta|y]}^{\text{Posterior}} = \frac{[y,\theta]}{[y]} = \frac{\overbrace{[y|\theta]}^{\text{likelihood}}\ \overbrace{[\theta]}^{\text{prior}}}{\underbrace{\int_\theta [y|\theta]\,[\theta]\,d\theta}_{\text{marginal}}}$$

First – a simple Bayesian model:

Joint distribution

$$\text{Posterior} \quad [\theta_1, \theta_2, z|\, y\,] \propto [y \mid \theta_1, \theta_2, z] \quad [\theta_1]\,[\theta_2]\,[z]$$

unobserved | observed | Likelihood | Priors

- This model is not hierarchical because there is no conditioning beyond the dependence of the data, y, on the unobserved quantities, θ_1, θ_2, z

- Every quantity found on the RHS of the conditioning symbol in the likelihood is found in a prior

Let's all just take a moment to stare at this, and make sure we get it…

Fundamentally, we need to be able to write out our ecological problem as follows…

$$\underbrace{[\theta|y]}_{\text{Posterior}} = \frac{[y,\theta]}{[y]} = \frac{\overbrace{[y|\theta]}^{\text{likelihood}}\ \overbrace{[\theta]}^{\text{prior}}}{\underbrace{\int_\theta [y|\theta]\,[\theta]\,d\theta}_{\text{marginal}}}$$

Now - a hierarchical Bayesian model:

Joint distribution

Posterior $[\theta_1, \theta_2, z|\, y] \propto [y \mid \theta_1, z]\ [z \mid \theta_2]\ [\theta_1]\,[\theta_2]$

unobserved — observed

Likelihood (product of two conditional distributions)

Priors

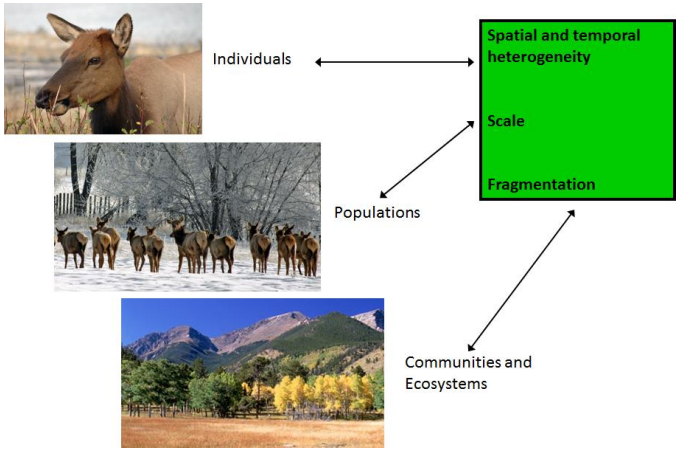
- **A Bayesian model is hierarchical whenever we use probability rules for factoring to express the joint distribution as a product of conditional distributions**

- Note there is no prior for z because it is conditional upon a quantity, θ_2, for which there is a prior distribution

Cue more staring….

Summary: how does this help us with ecological models?

- We use our knowledge of:
  - ecological systems (the context)
  - the ecological process
  - how we observe the process
  - the assumptions we make to simplify it (represent it as a model) and the parts we have left out



hierarchical Bayesian model:

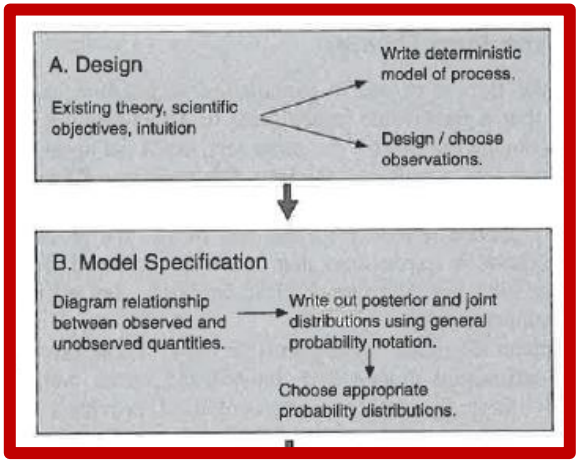
Joint distribution $[\theta_1, \theta_2, z | y] \propto [y | \theta_1, z] \; [z | \theta_2] \; [\theta_1] \; [\theta_2]$

unobserved

observed

Likelihood (product of two conditional distributions)

Priors



The mechanics…

How do we go from our conceptual understanding of an ecological process and our data…

…to a fully functioning model we can fit?

Start with the fundamentals …

Concept to be taught	Why do you need to understand this concept?
Conditional probability	It is the foundation for Bayes' Theorem and all inferences we will make.
The law of total probability	Basis for the denominator of Bayes' Theorem $[y]$
Factoring joint distributions	This is the procedure we will use to build models.
Independence	Allows us to simplify fully factored joint distributions.
Marginal distributions	Bayesian inference is based on marginal distributions of unobserved quantities.
Statistical distributions	Our toolbox for representing uncertainty and for linking observed quantities to unobserved ones.
Moments	Basis for inference from MCMC
Moment matching	Allows us to embed the predictions of models into any statistical distribution.

From Tom Hobbs, Colorado State University

We need to be able to define our model in terms of a set of factored joint distributions

- **Conceptual model →**

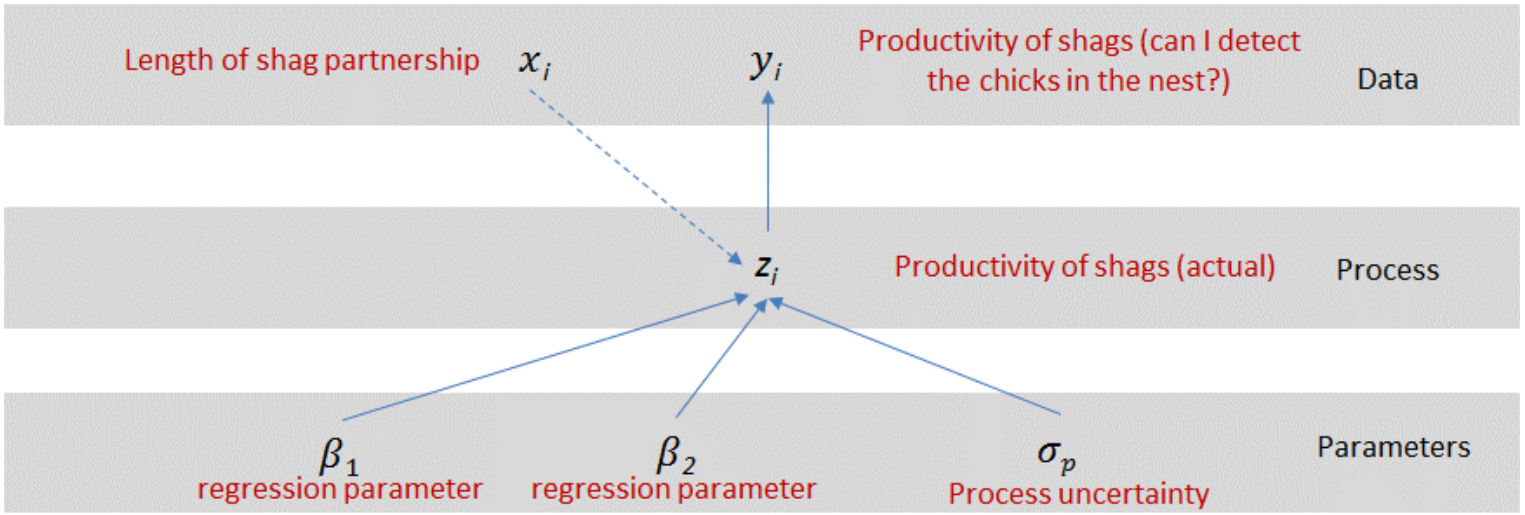
- **Factored joint distribution for the posterior**

Graphical modelling, or Directed Acyclic Graphs (DAGs)

DAG and factoring practice:

Diagramming joint and conditional probabilities

An ecological model – coming up later….

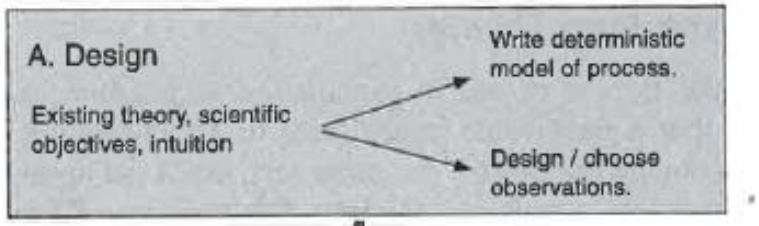


….but first, something simpler!

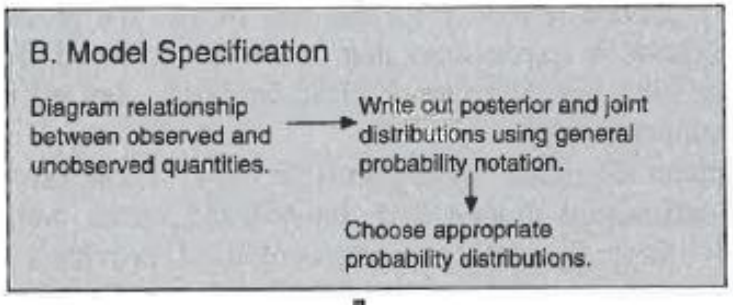
Bayesian networks or directed acyclic graphs (DAGs)

- We use these to draw and then write out factored expressions for joint distributions

- The expression for the joint distribution is then implemented within a statistical package (e.g., BUGS, JAGS, STAN) to fit the model and estimate the parameters of interest



Define question, think about ecological process and observations

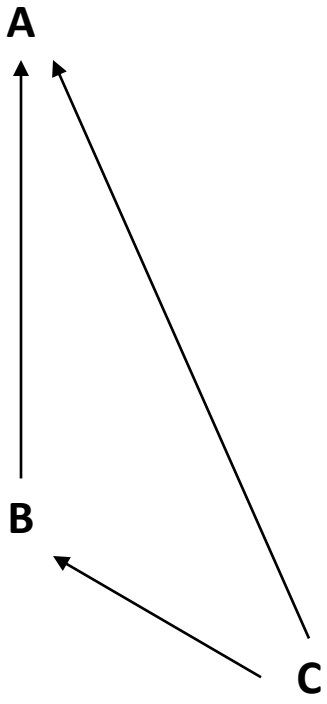


Define the model to represent the ecological process and observations

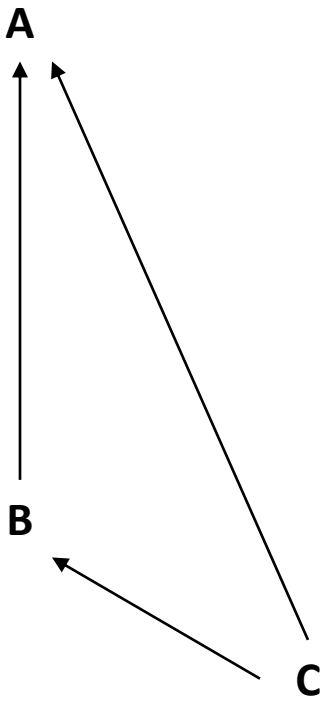
- Diagram it
- Write out posterior and joint distributions
- Choose probability distributions

Rules for factoring joint distribution using Bayesian networks

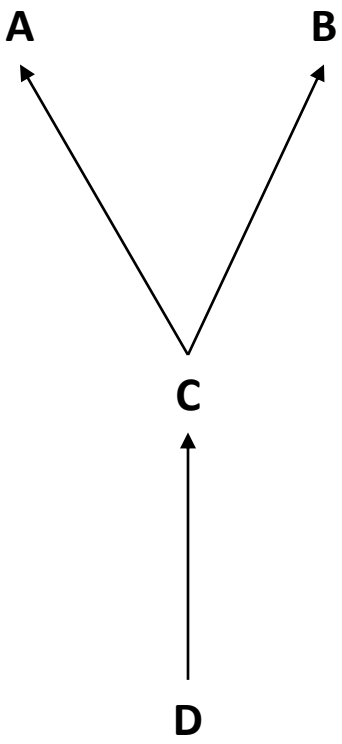
- All nodes at head of arrows must be on left hand side of conditioning symbol "|"
- All nodes at tails of arrows must be on right hand side of conditioning symbol.
- Any node at the tail of an arrow without an arrow leading into it must be expressed unconditionally.



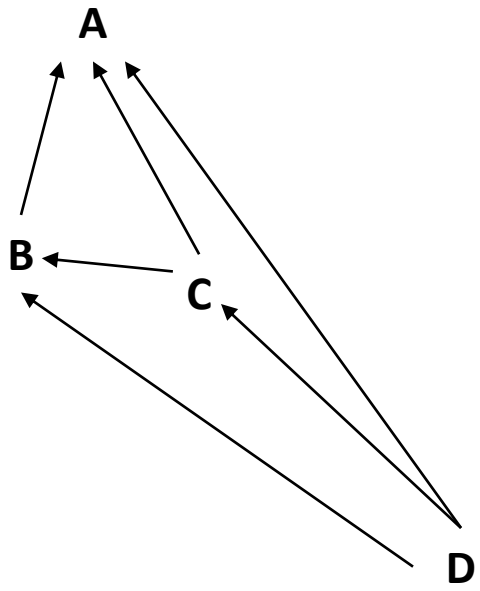
$$\Pr(A,B) = \Pr(A \mid B)\ \Pr(B)$$



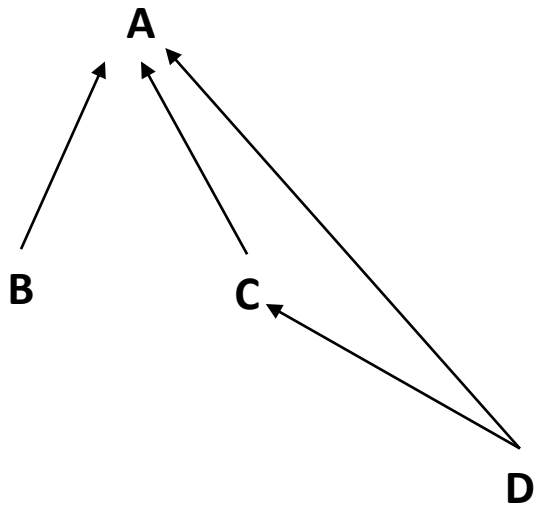
$$\Pr(A,B,C) = \Pr(A \mid B, C) \times \Pr(B \mid C) \times \Pr(C)$$



$$\Pr(A,B,C,D) = \Pr(A \mid C) \times \Pr(B \mid C) \times \Pr(C \mid D) \times \Pr(D)$$



$$\Pr(A,B,C,D) = \Pr(A \mid B, C, D) \times \Pr(B \mid C, D) \times \Pr(C \mid D) \times \Pr(D)$$

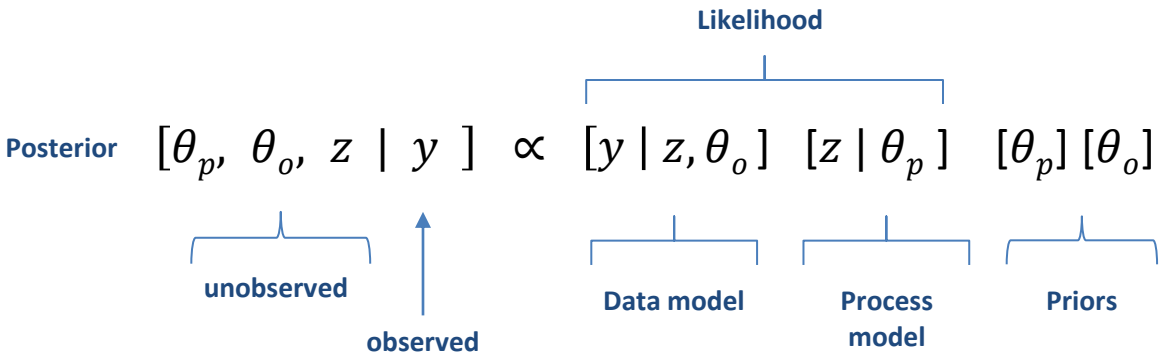


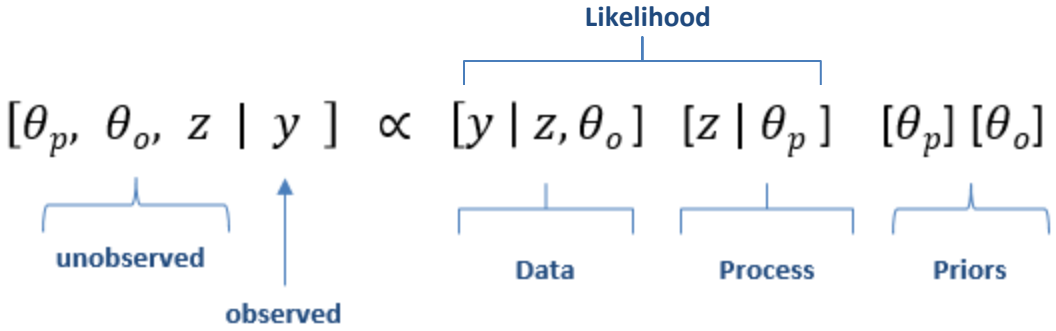
$$\Pr(A,B,C,D) = \Pr(A \mid B, C, D) \times \Pr(C \mid D) \times \Pr(B) \Pr(D)$$

And now for something more ecological…

- We most often **factor the joint distribution** in a way that allows us to deal with a broad range of ecological questions:

  - There is a true ecological state of interest, z, that is not directly observable
  - We relate that state to observable data, y, using a model with a vector of parameters, θ_o
  - The behaviour of the true state, or the process, is predicted by a model with parameters, θ_p





Data model (observation model)

- When we count animals, some are overlooked…the mismatch between what we observe and the true state requires a model of the observations

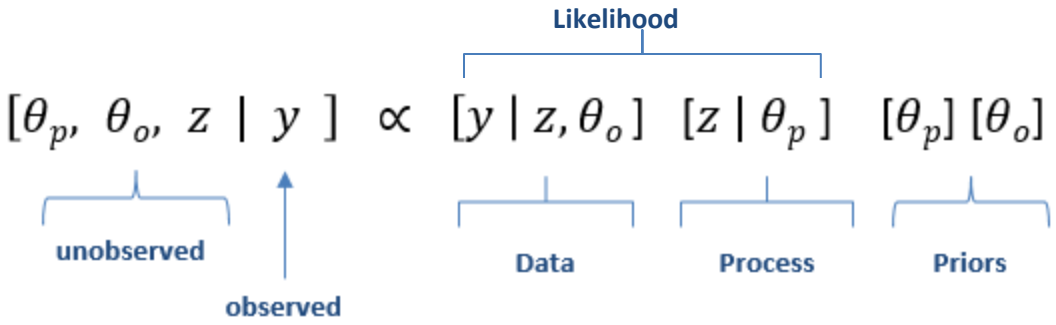
- *z* is the quantity we would observe if we could perfectly observe the instance of the true state, without any bias injected by our observation process

- The data model includes our knowledge of the relationship between the true state and our observations of it and the uncertainty that occurs because that relationship is imperfect

- We estimate θ_o to represent our observation uncertainty or sampling error

Parameter model (priors)

- what we know about the parameters when we began our investigation, that is, our prior knowledge



Process models are a mathematical statement depicting a process and a way to account for uncertainty about the process

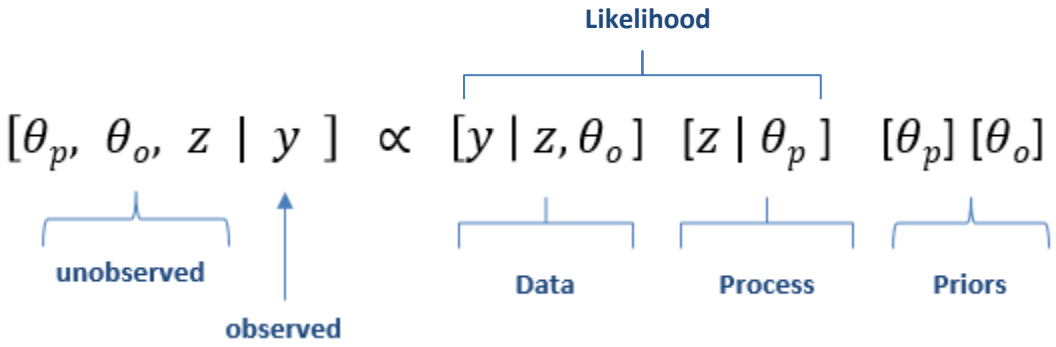
- We think about the true state of a system, *z* (the size of a population, the number of offspring per individual)

- We write an equation, a deterministic model that represents how we think the state of interest behaves, and the quantities that influence it

- We recognise there are missing parts to our model that may shape the behaviour of the true state, and we estimate these using a parameter, σ_p, the process variance

- if we know the functional form of the deterministic model, the values of its parameters, and the process variance, we can specify the probability of the true state…in other words, we can make predictions about the probability of various values of the true state

- We evaluate the predictions of the process model against data to refine and fit our model

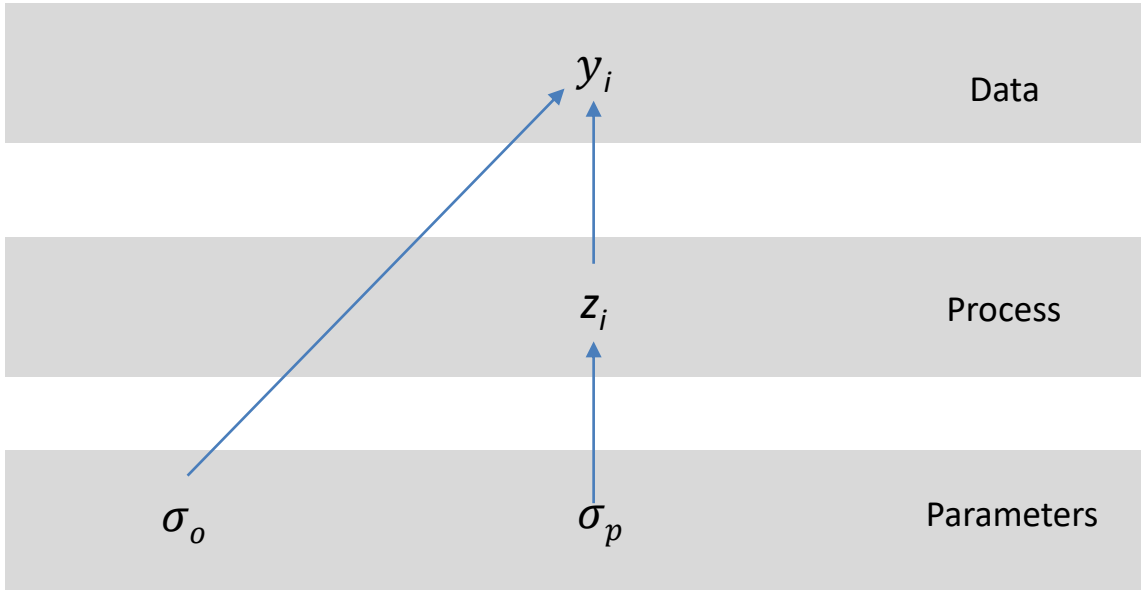


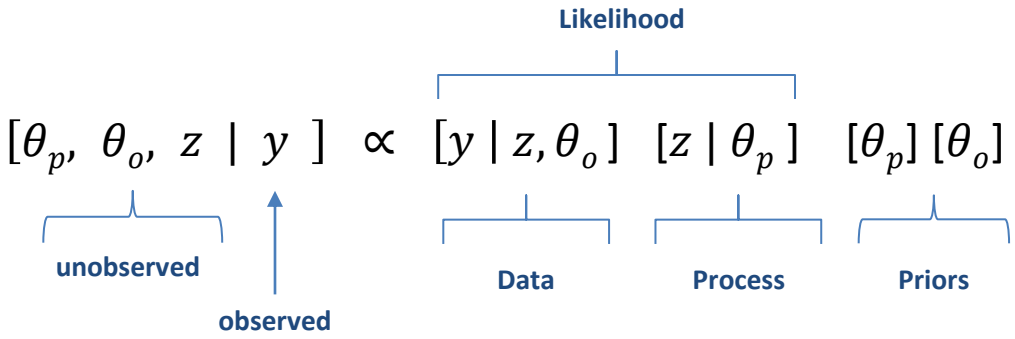
data model + process model + priors =

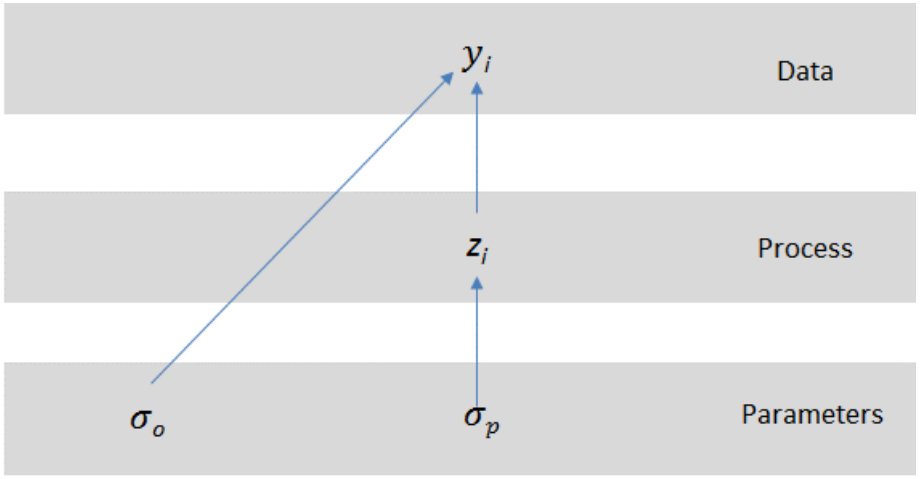
full mathematical expression for:

- **our ecological process (process model), …**
- **linked to data (data model), …**
- **in a way that includes all sources of uncertainty (observation uncertainty and process uncertainty), …**
- **and allows us to include prior understanding (priors)**

So, what does that look like? How can we use DAGs to help us with our model specification?







$$[\theta_p,\ \theta_o,\ z \mid y\ ] \propto [y \mid z, \theta_o]\ [z \mid \theta_p\ ]\ [\theta_p]\ [\theta_o]$$

unobserved — observed — Data — Process — Priors

- Nodes (random variables) at the heads of arrows appear on the LHS of the conditioning |
- Nodes at the tails of arrows appear on the RHS of the conditioning |
- Nodes at the tails of arrows with no arrow leading to it are expressed as priors

- Nodes are random variables
- Solid arrows are stochastic relationships among random variables
- Tails of arrows specify parameters defining the distribution of the random variable at the head of the arrow

Let's make this a bit less abstract….

We can use DAGs to write out our full factored joint distribution for the posterior

Now let's try with an ecological example –

- ***We need to define probability distributions for our random variables***

- And we have data that is not normal

Let's make this a bit less abstract….variation in processes caused by variation among individuals

- We want to model how the productivity (number of chicks) of shags is affected by the length of time that a pair of shags has been bonded together

- We have some data on individual shags (productivity and length of pairing)

- **Step 1**, draw the DAG…

- **Step 2**, write out the joint distribution

- **Step 3**, define the probability distributions we need for each random variable

- **Step 4**, add in the process model for productivity

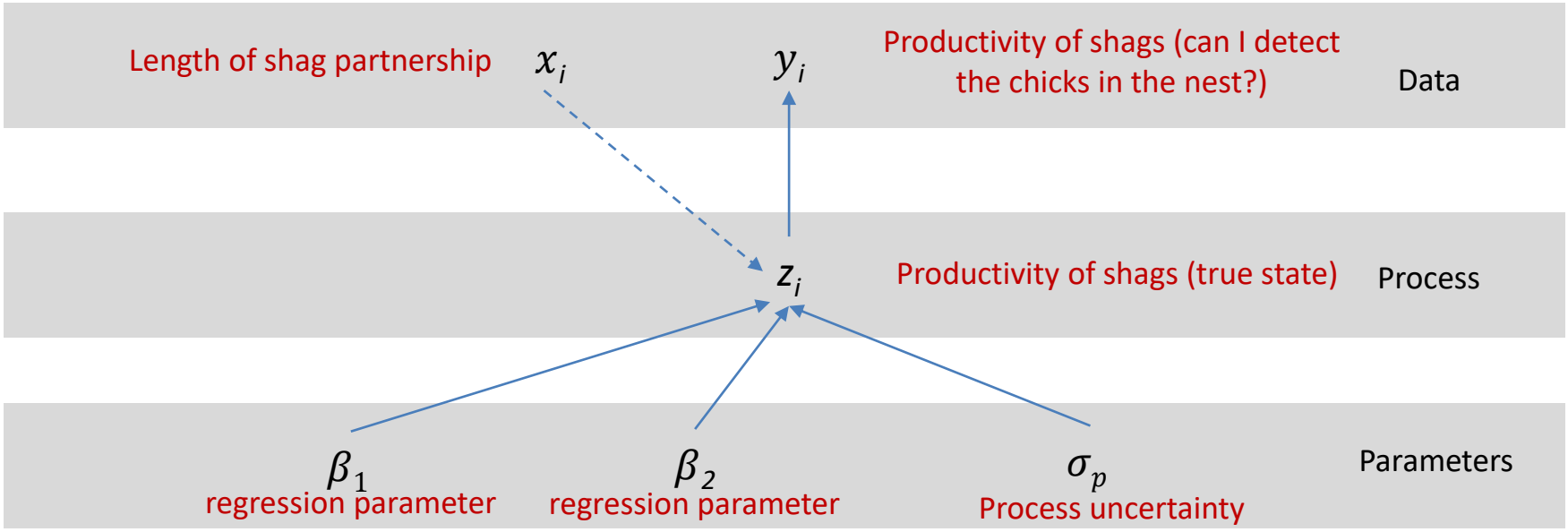
Let's make this a bit less abstract….variation in processes caused by variation among individuals

Process model = $z_i = g(\beta_1, \beta_2 xi) = \beta_1 + \beta_2 x_i$

= modelling the process for how productivity is affected by length of partnership

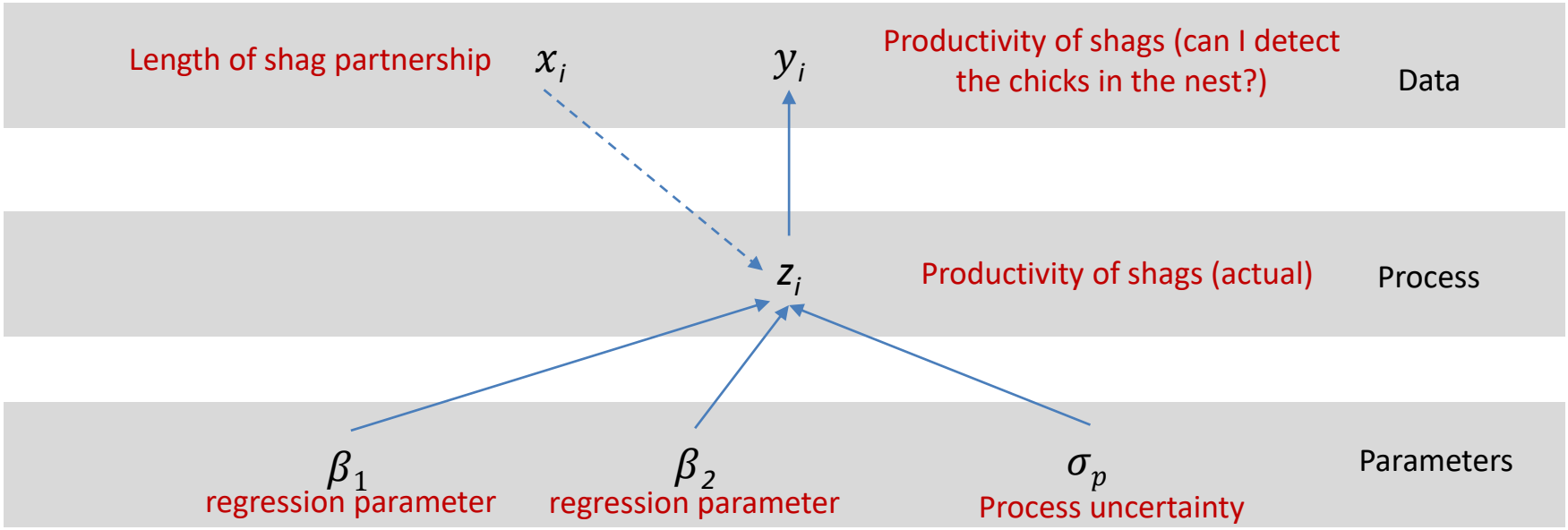
- β_1 is the average productivity of a shag with a partnership of length 0
- β_2 is the parameter that controls how length of partnership affects productivity
- σ_p is the process variance or process uncertainty

- **Step 1**, draw the DAG…

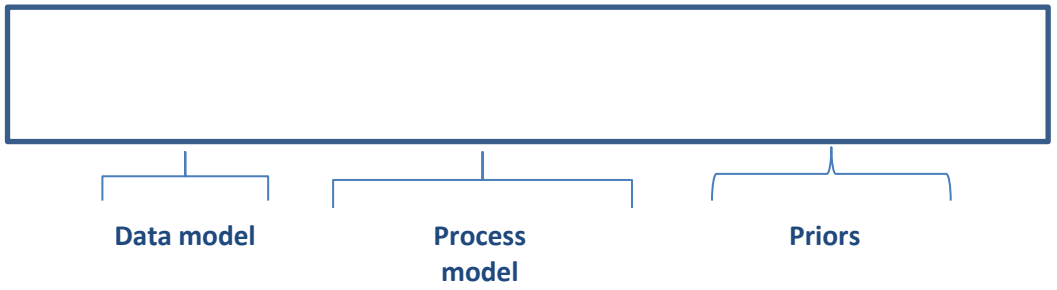


Let's make this a bit less abstract….variation in processes caused by variation among individuals

Now use the DAG to factor out the joint distribution for the posterior….



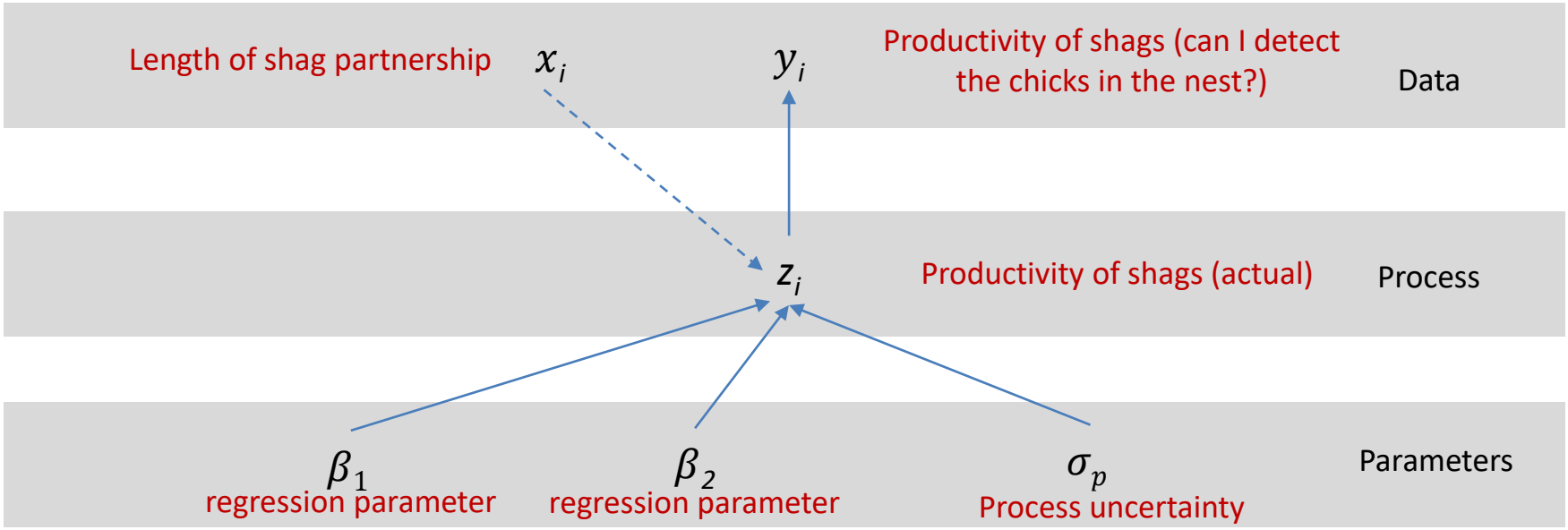
$$[\sigma_p,\ \beta_1, \beta_2,\ z \mid y\ ] \propto$$



Step 2, write out the joint distribution

Let's make this a bit less abstract….variation in processes caused by variation among individuals

Now use the DAG to factor out the joint distribution for the posterior….

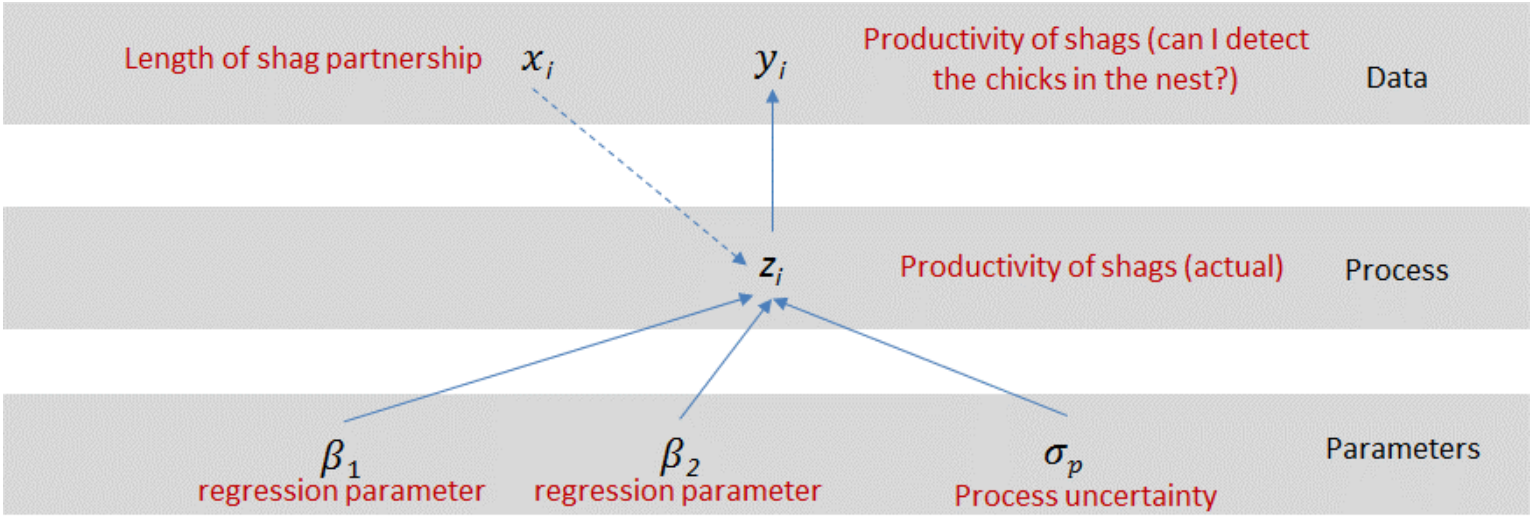


$$[\sigma_p,\ \beta_1, \beta_2,\ z \mid y\ ] \propto [y \mid z\ ]\ \ [z \mid \sigma_p,\ \beta_1,\ \beta_2]\ \ [\beta_1]\,[\beta_2]\,[\sigma_p]$$

Data model | Process model | Priors

Step 2, write out the joint distribution

Step 3, define the probability distributions we need for each random variable



$$[\sigma_p,\ \beta_1, \beta_2,\ z \mid y\ ] \propto [y \mid z\,]\quad [z \mid \sigma_p,\ \beta_1,\ \beta_2]\quad [\beta_1]\,[\beta_2]\,[\sigma_p]$$

$$[\sigma_p, \theta_o, \beta_1, \beta_2, z \mid y\ ] \propto \prod_{i=1}^{n} \text{Poisson}(y_i \mid z_i)\ \ \text{gamma}(z_i \mid \beta_1,\ \beta_2, \sigma_p)$$

x normal(β_1) normal(β_2)

x inverse gamma(σ_p)

Step 3, define the probability distributions we need for each random variable

Why did we choose those probability distributions?

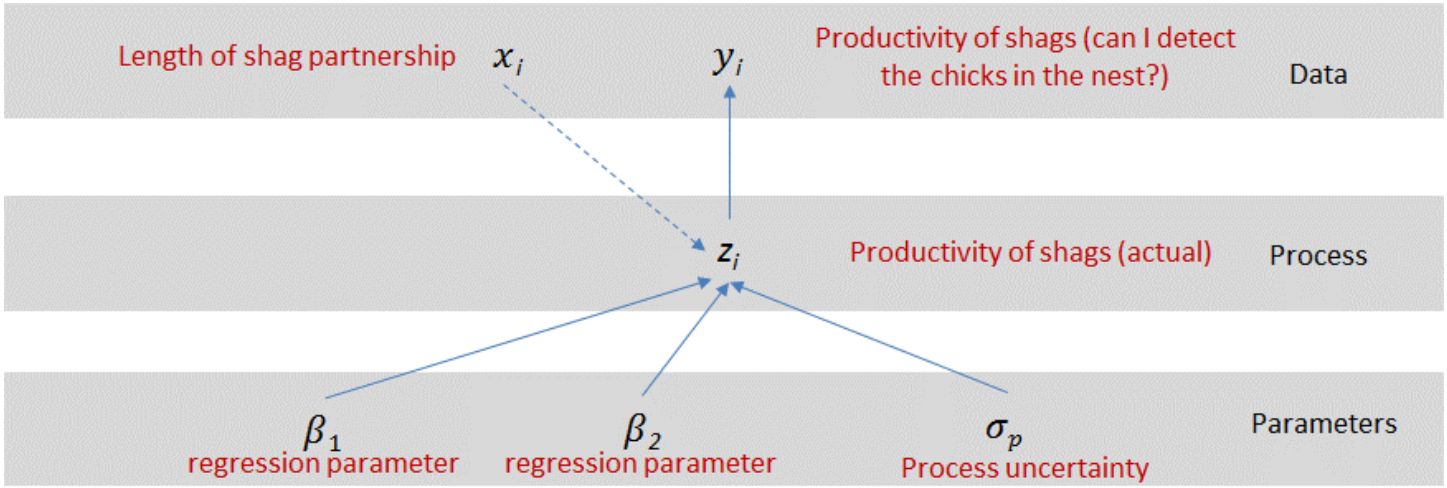
$$[\sigma_p, \theta_o, \beta_1, \beta_2, z \mid y] \propto \prod_{i=1}^{n} \text{Poisson}(y_i \mid z_i) \ \text{gamma}(z_i \mid \beta_1, \beta_2, \sigma_p)$$

$$\text{x} \quad \text{normal}(\beta_1)\ \text{normal}(\beta_2)$$

$$\text{x} \quad \text{inverse gamma}(\sigma_p)$$

- We choose a Poisson distribution for productivity (count data)
- We choose a gamma distribution for z because it is a conjugate for the Poisson, and because it is continuous and non-negative, and has **two parameters**
- We use an inverse gamma for σ_p because it is a variance
- We choose normal distributions for the β's because they are continuous random variables that can take on any value
    - We make them minimally informative priors by centring on zero and assigning a variance that is large relative to their values

Step 4, add in the process model for productivity explicitly…



Process model = $z_i = g(\beta_1, \beta_2\, xi) = \beta_1 + \beta_2 x_i + \varepsilon$

= modelling the process for how productivity is affected by length of partnership

- β_1 is the average productivity of a shag with a partnership of length 0
- β_2 is the parameter that controls how length of partnership affects productivity
- σ_p is the process variance or process uncertainty

Gamma has two parameters

Deterministic process

Process uncertainty

$$[\sigma_p, \beta_1, \beta_2, z \mid y] \propto \prod_{i=1}^{n} \text{Poisson}(y_i \mid z_i)\ \text{gamma}(z_i \mid g(\beta_1, \beta_2\, xi), \sigma_p)$$

$$\text{x}\quad \text{normal}(\beta_1 \mid 0,100)\ \text{normal}(\beta_2 \mid 0,100)$$

$$\text{x}\quad \text{inverse gamma}(\sigma_p \mid 0.001, 0.001)$$

Step 4, add in the process model for productivity explicitly…

gamma($z_i \mid g(\beta_1, \beta_2\, xi), \sigma_p$)

$$[\sigma_p, \beta_1, \beta_2, z \mid y\ ] \propto \prod_{i=1}^{n} \text{Poisson}(y_i \mid z_i)\ \text{gamma}\left(z_i \mid \frac{g(\beta_1,\beta_2\, xi)^2}{\sigma_p}, \frac{g(\beta_1,\beta_2\, xi)}{\sigma_p}\right)$$

x normal(β_1 |0,100) normal(β_2|0,100)
x inverse gamma(σ_p |0.001,0.001)

What's going on here?

Moment matching…

Process model = $z_i = g(\beta_1, \beta_2\, xi) = \beta_1 + \beta_2 x_i$

- Our process model predicts the mean (or sometimes the median) of a distribution, and we want to estimate process uncertainty
- But, the parameters for the gamma distribution are not the mean and the variance – they are the *shape* and the *rate*
- We need to be able to calculate these parameters from the mean and the variance…

$$\alpha = \frac{\mu^2}{\sigma^2}$$

$$\beta = \frac{\mu}{\sigma^2}$$

The mean of the gamma distribution is

$$\mu = \frac{\alpha}{\beta}$$

and the variance is

$$\sigma^2 = \frac{\alpha}{\beta^2}.$$

where α is the shape and β is the rate.

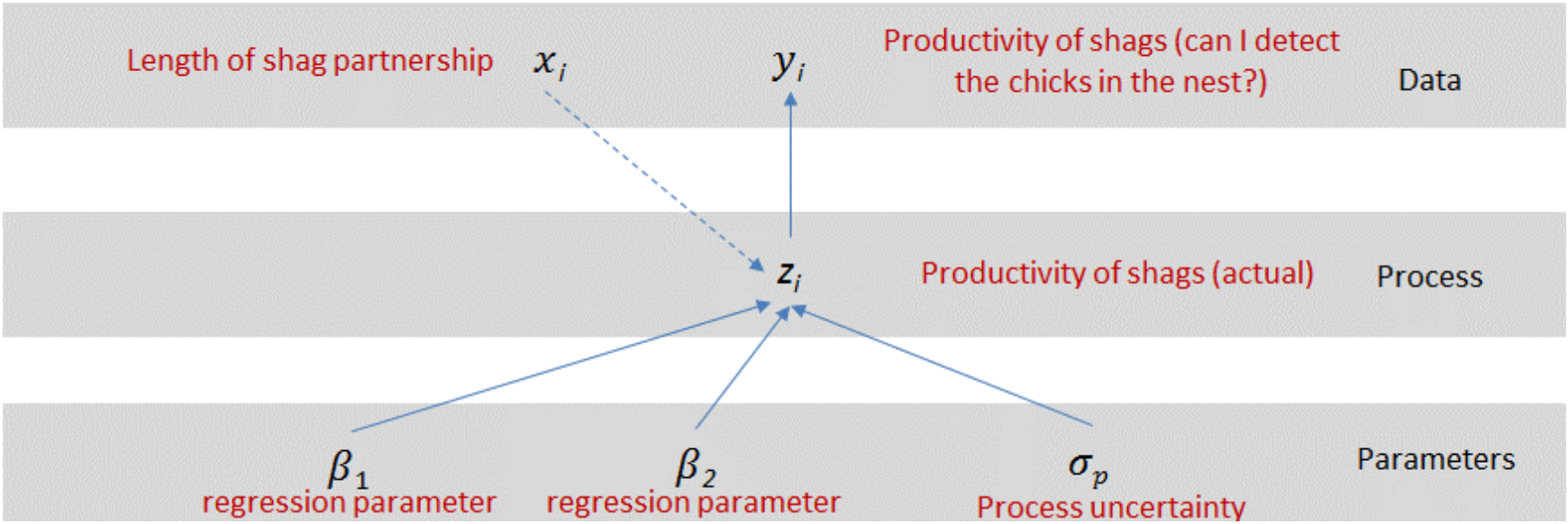
Step 5 (I didn't tell you about step 5) – some additional observations on this model:

What about sampling variance/observation error?

- There is an explicit parameter for the process variance, σ_p, but there does not appear to be a parameter controlling variance in the observations, y_i

- Does this mean we are assuming no observation variance? No…

    - Because the Poisson distribution assumes the variance is the same as the mean, so observation variance is implicit in the likelihood



$$[\sigma_p, \beta_1, \beta_2, z \mid y\ ] \propto \prod_{i=1}^{n} \text{Poisson}(y_i \mid z_i)\ \ \text{gamma}(z_i \mid \frac{g(\beta_1,\beta_2\, xi)^2}{\sigma_p}, \frac{g(\beta_1,\beta_2\, xi)}{\sigma_p})$$

x normal(β_1 |0,100) normal(β_2|0,100)

x inverse gamma(σ_p |0.001,0.001)

Why doesn't the dataset, *x* (length of partnership), appear in the posterior distribution?

- The *x* are not treated as random variables in this formulation – we assume the *x* data are perfectly observed and fixed, whilst the *y* data are random variables

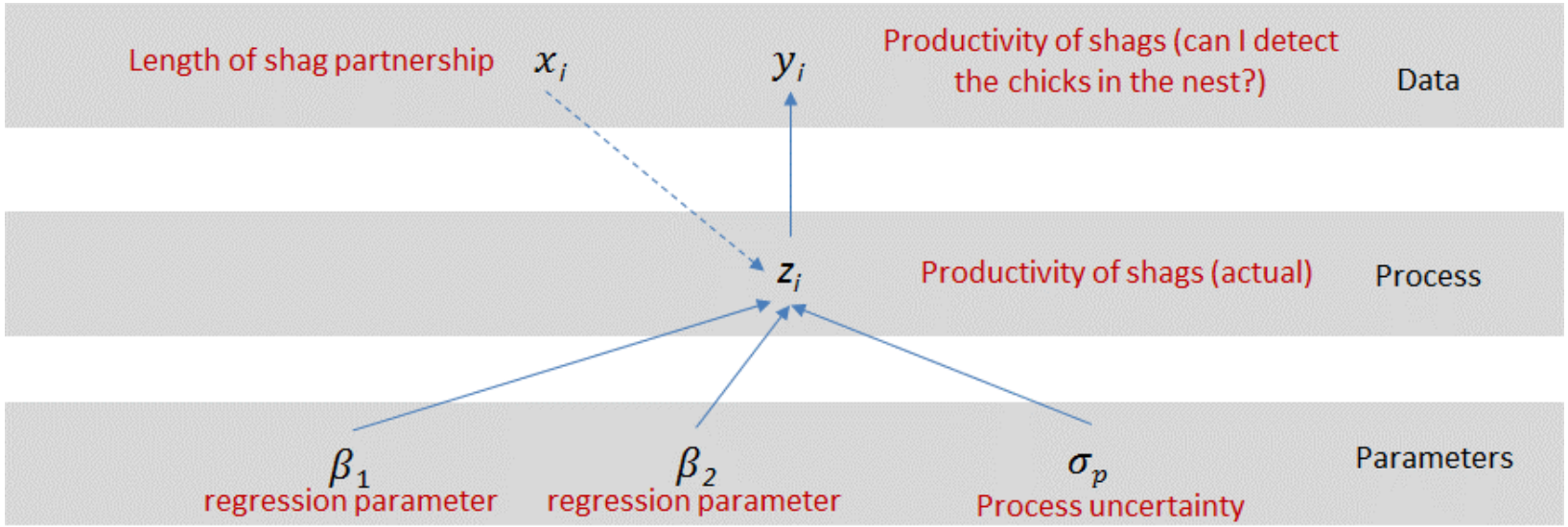
- This means the *x* are known, fixed quantities, treated no differently than a constant – they are not random variables and therefore should not appear in the expression for the posterior distribution, which by definition, is composed of random variables

- The predictor variables do appear as arguments to the deterministic function for shag productivity:

Process model = $g(\beta_1, \beta_2\, xi)$

Why do we need to do this again (…my head hurts)?

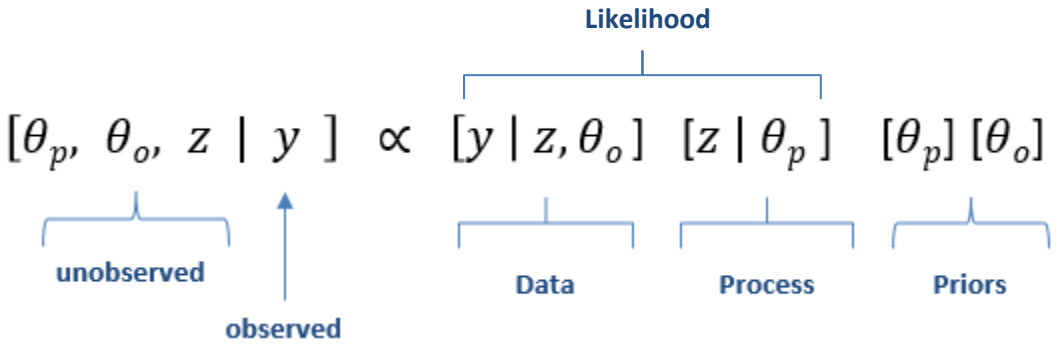
- clear, transparent science (What have we left out? What are our uncertainties?)
- We need to be able to express the posterior distribution as a set of joint factored likelihoods in order to fit our models in **any software package**
- E.g., JAGS – the 'model statement' in JAGS is the joint distribution for the posterior of your model…



$$[\sigma_p, \beta_1, \beta_2, z \mid y] \propto \prod_{i=1}^{n} \text{Poisson}(y_i \mid z_i) \ \text{gamma}(z_i \mid \frac{g(\beta_1,\beta_2 xi)^2}{\sigma_p}, \frac{g(\beta_1,\beta_2 xi)}{\sigma_p})$$

x normal(β_1 |0,100) normal(β_2|0,100)

x inverse gamma(σ_p |0.001,0.001)



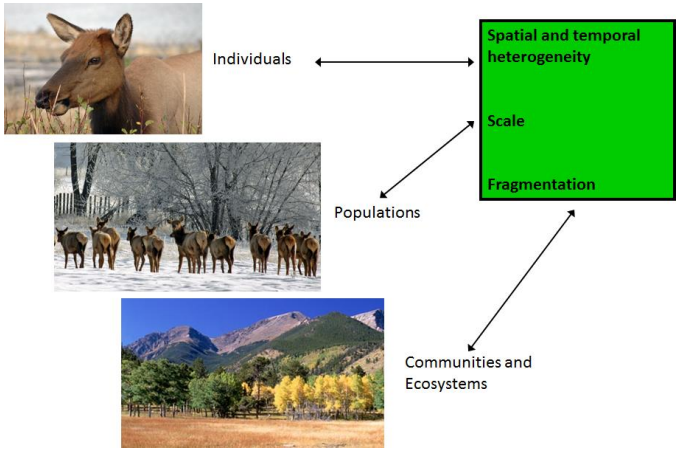
data model + process model + priors =

full mathematical expression for:

- **our ecological process (process model), …**
- **linked to data (data model), …**
- **in a way that includes all sources of uncertainty (observation uncertainty and process uncertainty), …**
- **and allows us to include prior understanding (priors)**

Summary: how does this help us with ecological models?

- We use our knowledge of:
  - ecological systems (the context)
  - the ecological process
  - how we observe the process
  - the assumptions we make to simplify it (represent it as a model) and the parts we have left out



hierarchical Bayesian model:

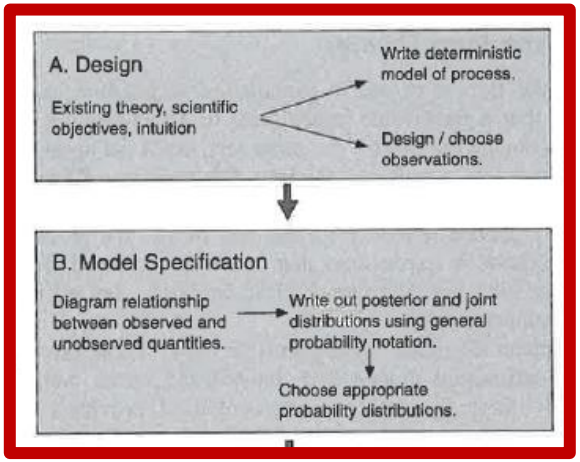
Joint distribution $[\theta_1, \theta_2, z | y] \propto [y | \theta_1, z] \; [z | \theta_2] \; [\theta_1] \; [\theta_2]$

unobserved

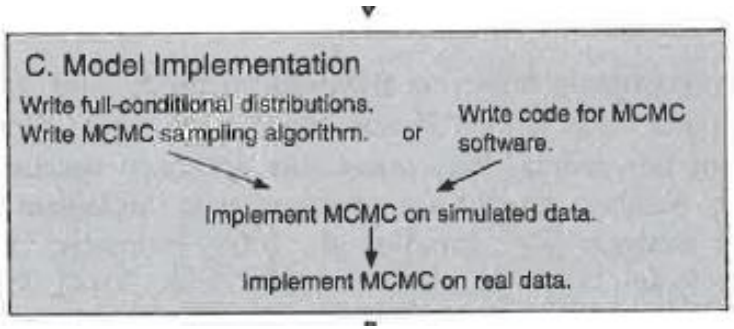
observed

Likelihood (product of two conditional distributions)

Priors



Part 2: implementing our model using MCMC (R and JAGS)



Implement the model using your expression for the posterior and joint distributions

$$[\sigma_p, \beta_1, \beta_2, z \mid y] \propto \prod_{i=1}^{n} \text{Poisson}(y_i \mid z_i) \ \text{gamma}(z_i \mid \frac{g(\beta_1, \beta_2 xi)^2}{\sigma_p}, \frac{g(\beta_1, \beta_2 xi)}{\sigma_p})$$

x normal(β_1 |0,100) normal(β_2|0,100)

x inverse gamma(σ_p |0.001,0.001)

PRACTICAL

Modelling light limitation of tree growth

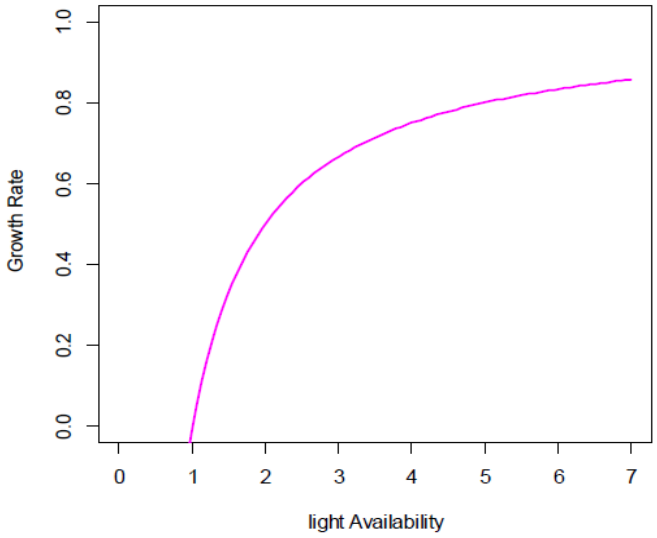
The relationship between tree growth rate and light tends to be non-linear, approaching an asymptote under high light conditions

Here, we model this simple curve using a Bayesian approach, where our response (*y*) is observed growth rate and our only predictor variable is light (*L*) = **process model**

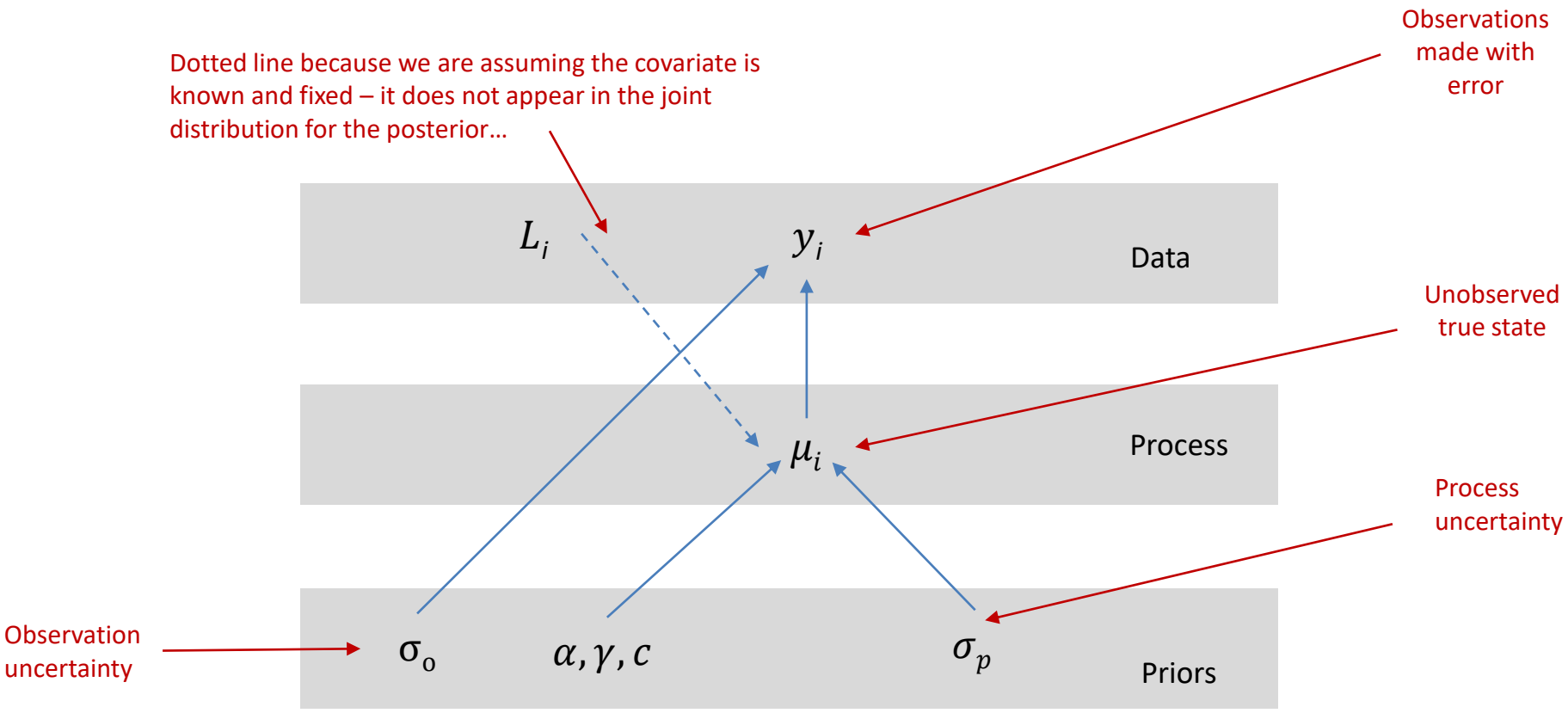
Michealis-Menton equation

Process model = $g(\alpha, \gamma, c, Li) = \dfrac{\alpha(L_i - c)}{(\alpha/\gamma) + (L_i - c)}$

α = max. growth at infinite light
γ = rate at which curve tails off
c = light level at which growth is zero (x intercept)



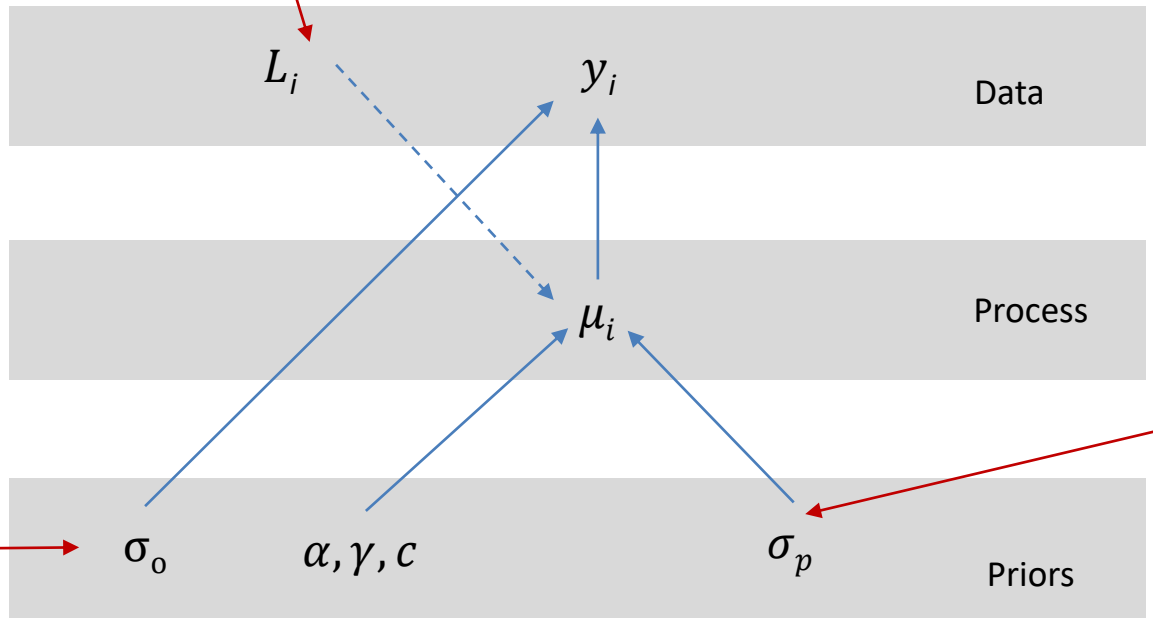
Hierarchical Bayesian model…



Hierarchical Bayesian model…

Process model = $g(\alpha, \gamma, c, Li) = \mu_i = \dfrac{\alpha(L_i - c)}{(\alpha/\gamma) + (L_i - c)}$

Dotted line because we are assuming the covariate is known and fixed – it does not appear in the joint distribution for the posterior…



$$[\alpha, \gamma, c, \mu_i, \sigma_p, \sigma_o | \ y_i \ ] \propto \prod_{i=1}^{n} [y_i \mid \mu_i, \sigma_o] \times \prod_{i=1}^{n} [\mu_i \mid g(\alpha, \gamma, c), \sigma_p] \times [\alpha]\,[\gamma]\,[c]\,[\sigma_p]\,[\sigma_o]$$

unobserved

observed

Likelihood

Priors

$$[\alpha, \gamma, c, \mu_i, \sigma_p, \sigma_o | \ y_i \ ] \propto [y_i \mid \mu_i, \sigma_o] \times [\mu_i \mid g(\alpha, \gamma, c), \sigma_p] \times [\alpha]\,[\gamma]\,[c]\,[\sigma_p]\,[\sigma_o]$$

unobserved — observed — Likelihood — Priors

$$\text{Process model} = g(\alpha, \gamma, c, Li) = \mu_i = \frac{\alpha(L_i - c)}{(\alpha/\gamma) + (L_i - c)}$$

- We choose a normal distribution for *y* growth rate (can be + or -)
- We choose a normal distribution for z because it is a conjugate for the normal, and because it can be + or -
- We use a normal for σ_o because we have prior knowledge about the mean and variance of our observation error
- We use a uniform for σ_p because we know the process variance is positive and bounded within a sensible range
- We choose gamma distributions for the α and γ because they are positive random variables
- We choose a uniform for *c* because we know it is bounded on the x-axis
    - We make them minimally informative priors by centring on zero and assigning a variance that is large relative to their values (normal) or placing most of the density mass at zero (gamma)

2a. Fit hierarchical Bayesian model in R and JAGS…

Open '*Kate_Tree light example.R*'

$$[\alpha, \gamma, c, \mu_i, \sigma_p, \sigma_o | \ y_i \ ] \propto \prod_{i=1}^{n} [y_i \mid \mu_i, \sigma_o] \times \prod_{i=1}^{n} [\mu_i \mid g(\alpha, \gamma, c), \sigma_p] \times [\alpha]\,[\gamma]\,[c]\,[\sigma_p]\,[\sigma_o]$$

unobserved

observed

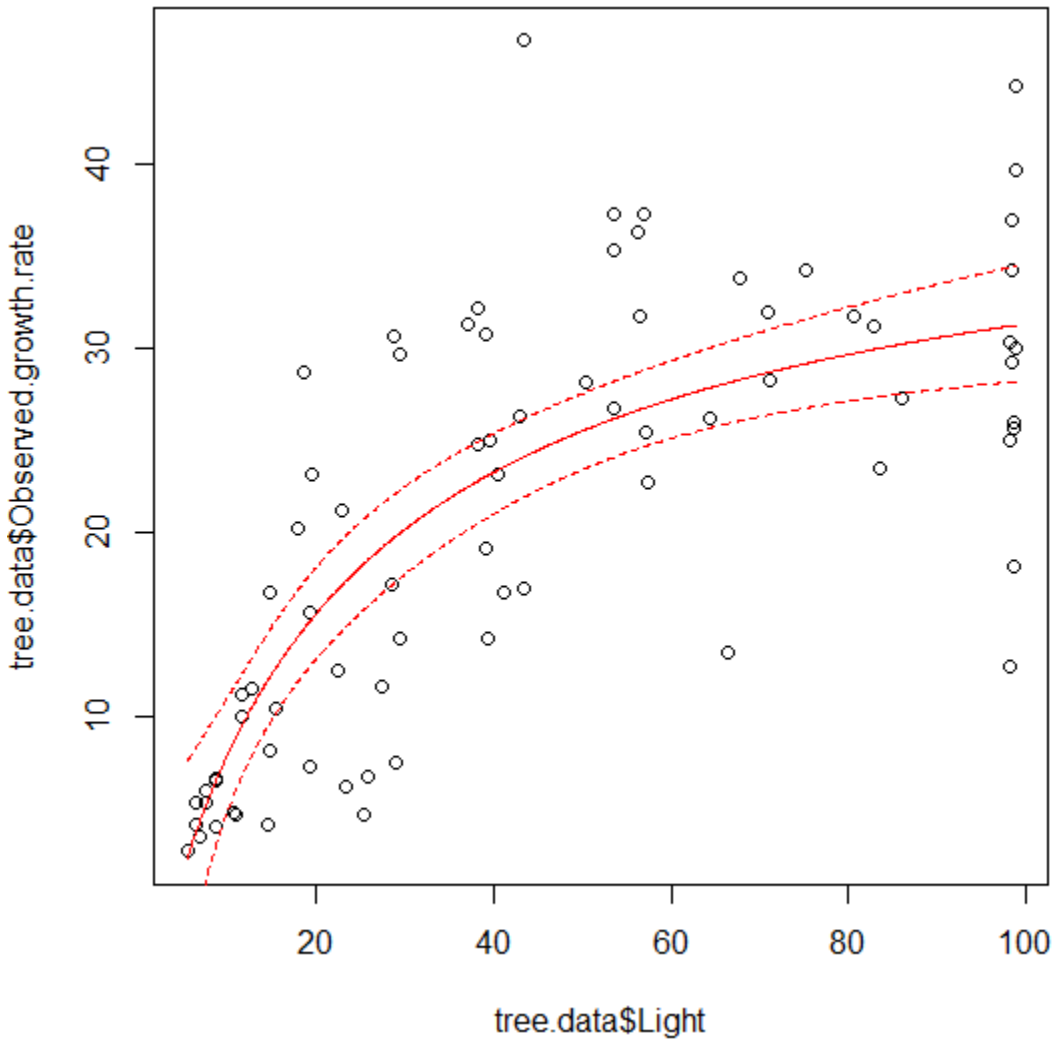
Likelihood

Priors

```
model{
  ### likelihood
  # Data model
  for (i in 1:n)
  {
    y[i] ~ dnorm(mu[i],tau.o)
  }

  # process model
  for (i in 1:n)
  {
    mu[i] ~ dnorm(mu2[i],tau.p)

    mu2[i] <- a * (x[i]-c) / ((a/b)+(x[i]-c))

  }

  # priors
  a ~ dgamma(0.01,0.01)
  c ~ dunif(-10,10)
  b ~ dgamma(0.01,0.01)

  sigma.o ~ dnorm(5, 1/(0.5*0.5)) ## assume prior knowledge of observation error (5 with SD of 0.5)
  sigma.p ~ dunif(0, 50)

  # derived quantities
  tau.o <- 1/(sigma.o * sigma.o)
  tau.p <- 1/(sigma.p * sigma.p)

} # end of model
```

Likelihood

Priors

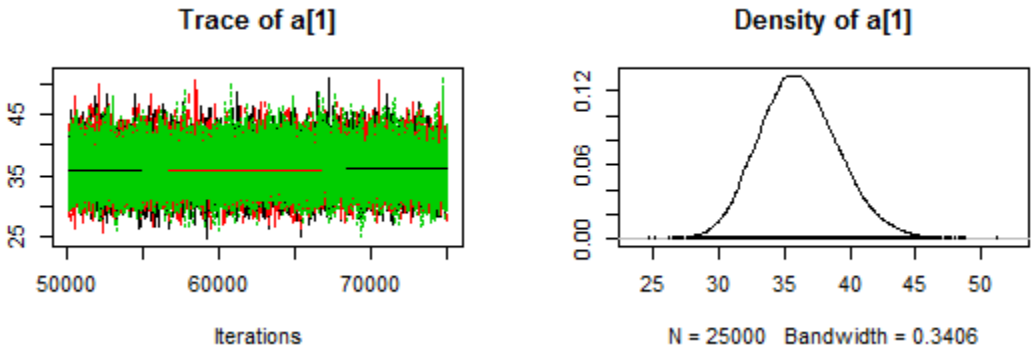

tree.data$Observed.growth.rate
10
20
30
40
20
40
60
80
100
tree.data$Light

Brief notes on model convergence…more later in the course

	mean	sd	2.5%	50%	97.5%	overlap0	f	Rhat	n.eff
a[1]	36.133264	3.049424	30.555743	35.998572	42.548168	FALSE	1.000000	1.000044	54897
a[2]	37.438024	3.083481	31.728760	37.322796	43.812127	FALSE	1.000000	1.000102	17228
a[3]	88.782200	7.820278	75.539225	88.072088	106.097899	FALSE	1.000000	1.000155	26351
b	2.217662	0.273113	1.718108	2.204837	2.785714	FALSE	1.000000	1.000347	10859
c	5.782560	1.014120	3.494090	5.887440	7.472791	FALSE	0.999853	1.000152	46979
sigma.o	5.015905	0.501903	4.029126	5.015210	5.999768	FALSE	1.000000	1.000357	12215
sigma.p	9.975478	0.652395	8.724783	9.965442	11.290310	FALSE	1.000000	1.000244	8331
sigma.a	32.726562	9.398469	15.756380	32.594954	48.925852	FALSE	1.000000	1.000020	56604
mu.a	53.594328	18.240695	16.061351	53.670157	89.897252	FALSE	1.000000	1.000011	75000
deviance	1398.395824	51.646216	1291.712879	1400.127933	1494.590955	FALSE	1.000000	1.000301	16346

```
Successful convergence based on Rhat values (all < 1.1).
Rhat is the potential scale reduction factor (at convergence, Rhat=1).
For each parameter, n.eff is a crude measure of effective sample size.

overlap0 checks if 0 falls in the parameter's 95% credible interval.
f is the proportion of the posterior with the same sign as the mean;
i.e., our confidence that the parameter is positive or negative.

DIC info: (pD = var(deviance)/2)
pD = 1333.5 and DIC = 2731.934
DIC is an estimate of expected predictive error (lower is better).
```

Trace plots provide an important tool for assessing mixing of a chain. **Density plot**s are smoothed histograms of the samples, that is they show the function that we are trying to explore – the posterior density of our unobserved parameter…



Burn-in

It is standard practice to discard the initial iterations of iterative simulation as they are too strongly influenced by starting values and do not provide good information about the target distribution

n.eff

the effective sample size, that is an estimate for the number of *independent* samples (taking into account autocorrelations) generated by the MCMC run

Convergence

Our usual approach is, for each parameter or quantity of interest,

- compute the variance of the simulations from each chain
- average these within-chain variances = average within-chain variance
- Then compare this to the variances of all the chains mixed together = mixture variance

We take the mixture variance divided by the average within-chain variance, compute the square root of this ratio, and call it **R.hat** or the "potential scale reduction factor" (Gelman and Rubin, 1992, following ideas of Fosdick, 1959).

At convergence, the chains will have mixed, so that the distribution of the simulations between and within chains will be identical, and the ratio **R.hat** should equal 1. If **R.hat** is greater than 1, this implies that the chains have not fully mixed and that further simulation might increase the precision of inferences.

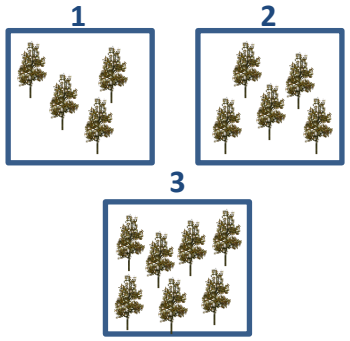
In practice we typically go until **R.hat** is less than 1.1 for all parameters and quantities of interest

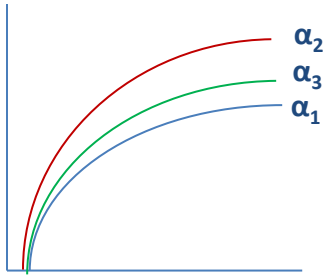
http://www.mcmchandbook.net/HandbookChapter6.pdf

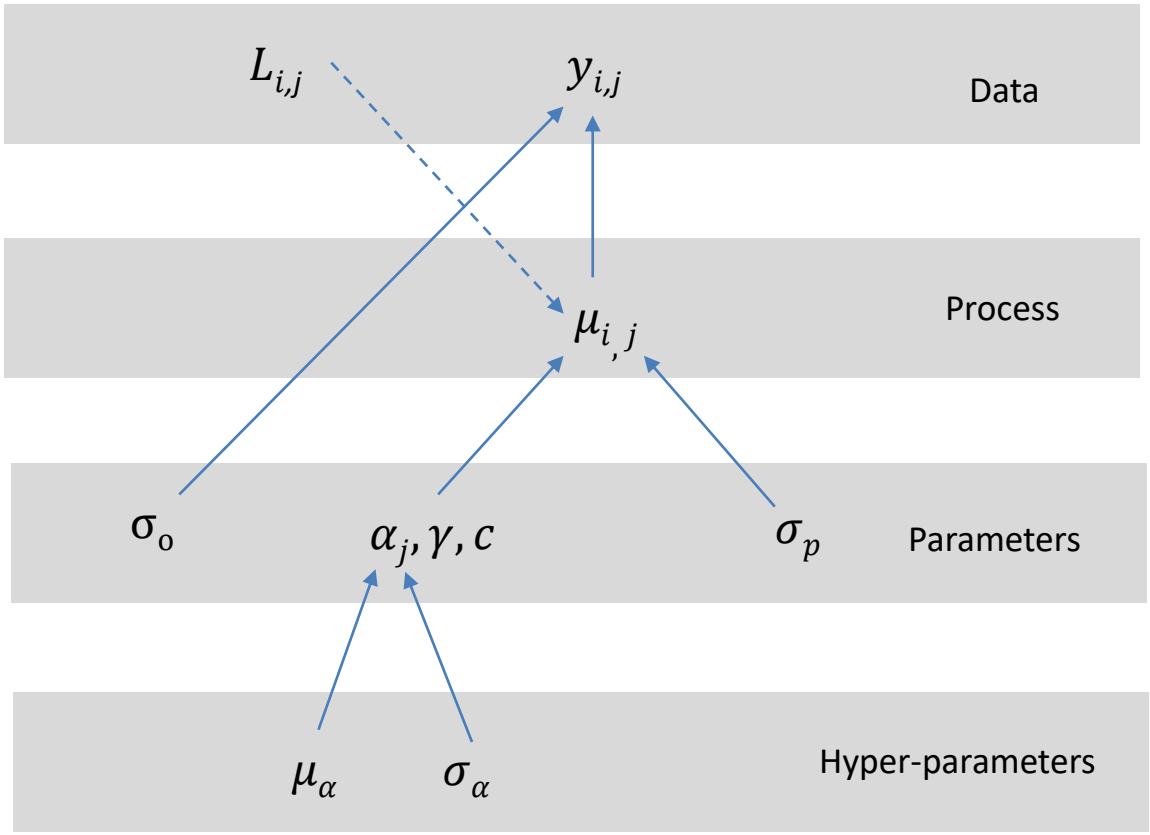
2b. Fit hierarchical Bayesian model with multiple sites in R and JAGS…

Open '*Kate_Tree light example multi-level.R*'

Hierarchical Bayesian model…now with multiple sites, *j*





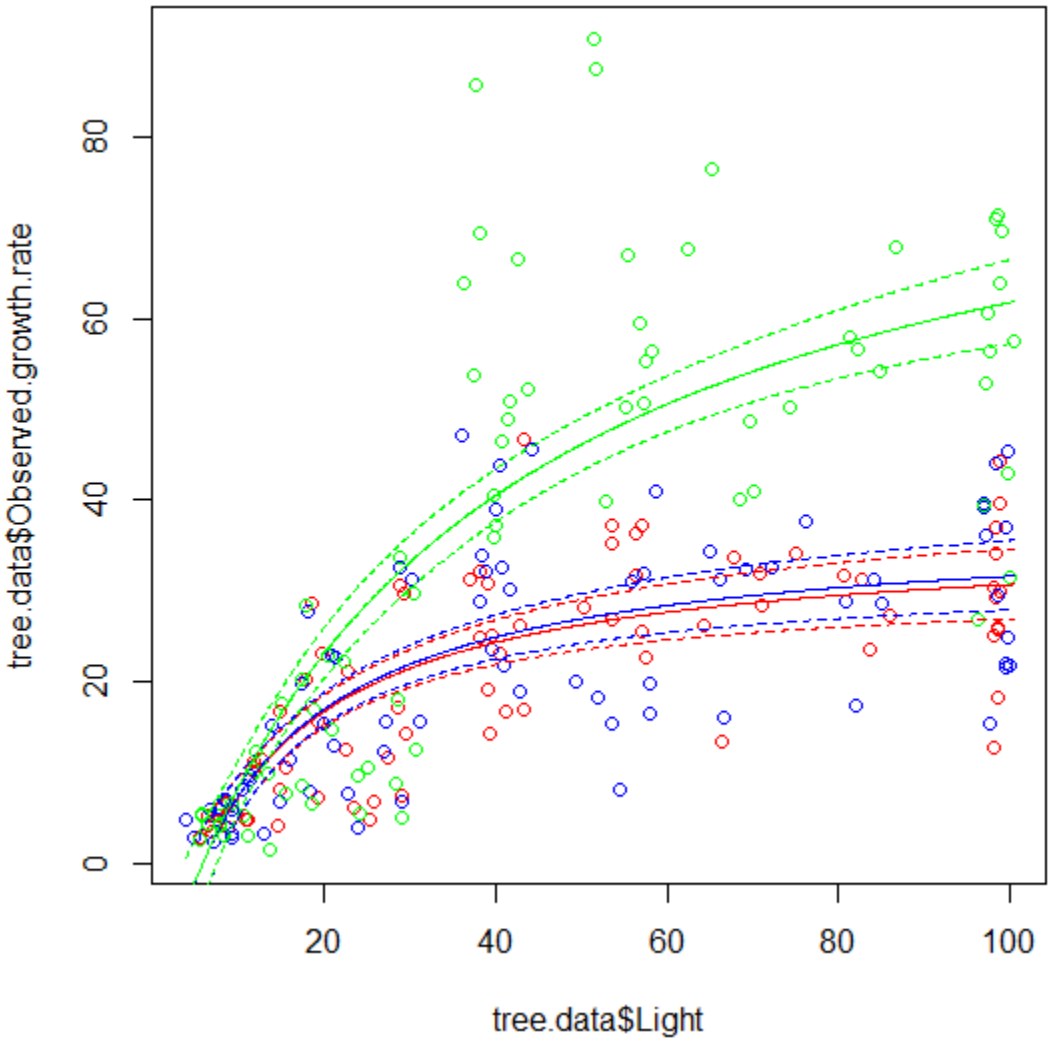


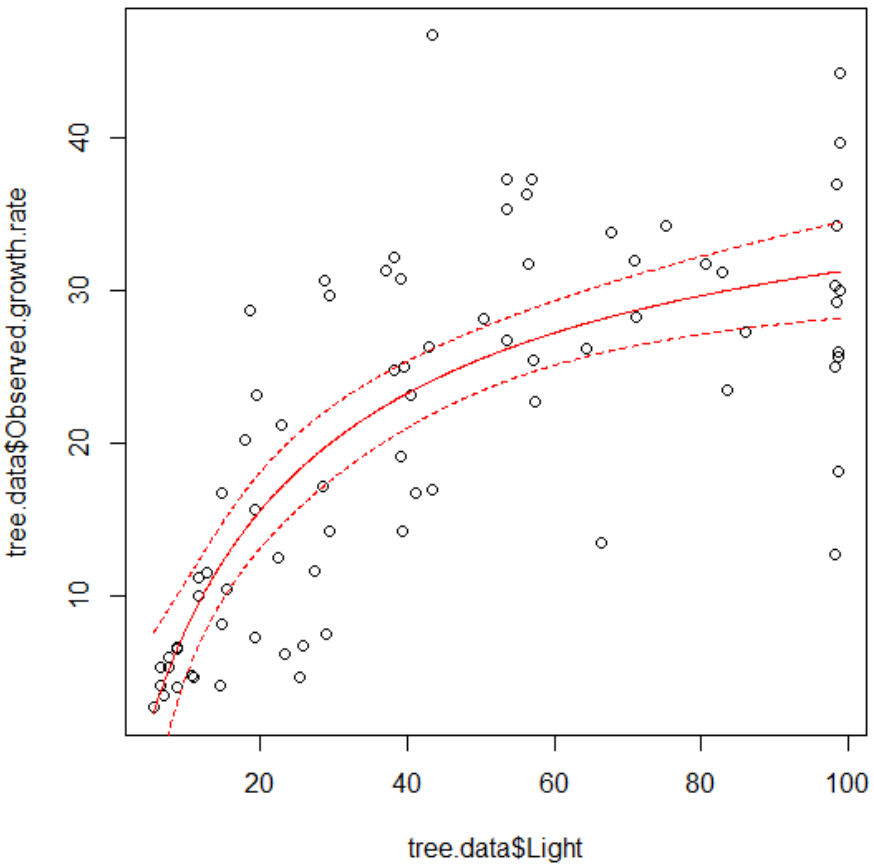
Multiple sites, *j*, and we expect there to be differences in the maximum growth rate per site, α_j, for instance due to soil water availability

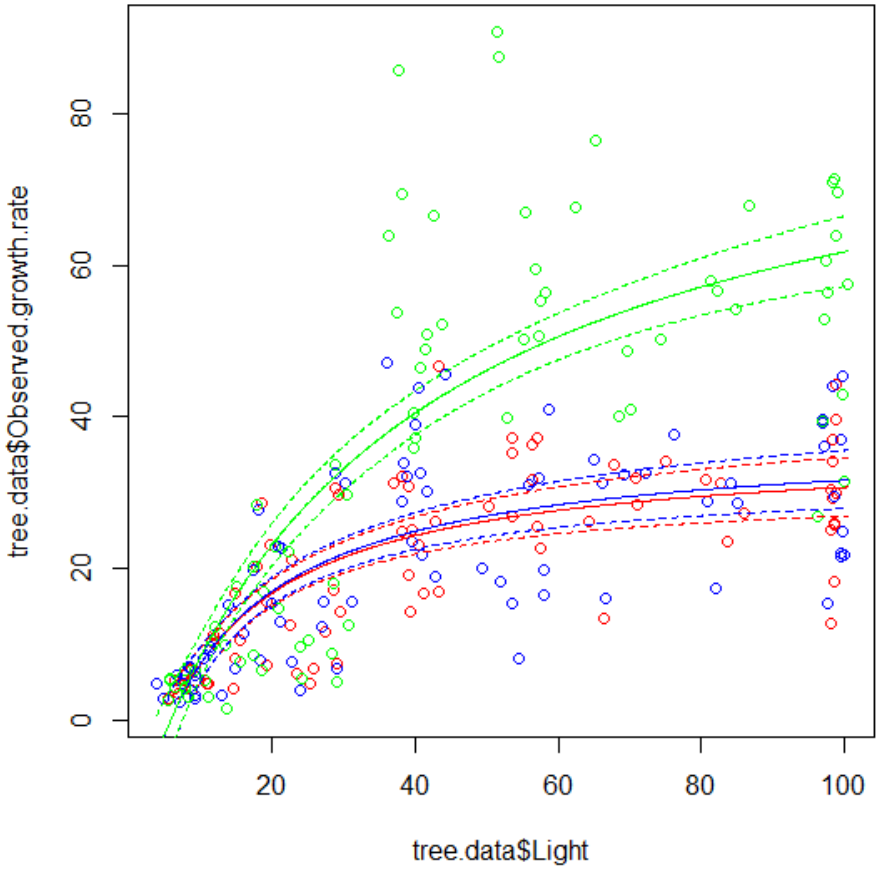
$$[\alpha_j, \mu_{i,j}, \gamma, c, \sigma_p, \sigma_o, \mu_\alpha, \sigma_\alpha | y_{i,j}] \propto \prod_{i=1}^{n} \prod_{j=1}^{3} [y_{i,j} | \mu_{i,j}, \sigma_o] \times \prod_{i=1}^{n} \prod_{j=1}^{3} [\mu_{i,j} | g(c, \gamma, \alpha_j), \sigma_p] \times \prod_{j=1}^{3} [\alpha_j | \mu_\alpha, \sigma_\alpha] \times [\mu_\alpha][\sigma_\alpha][\gamma][c][\sigma_p][\sigma_o]$$

Process model

$$= g(\alpha, \gamma, c, Li_{,j}) = \mu_{i,j} = \frac{\alpha(L_i - c)}{(\alpha/\gamma) + (L_i - c)}$$


tree.data$Observed.growth.rate
0
20
40
60
80
20
40
60
80
100
tree.data$Light


tree.data$Observed.growth.rate
10
20
30
40
20
40
60
80
100
tree.data$Light


tree.data$Observed.growth.rate
0
20
40
60
80
20
40
60
80
100
tree.data$Light


THE END
(phew)