Bayesian methods for ecological and environmental modelling

Trainers:

Lindsay Banin, David Cameron, Pete Henrys & Peter Levy

UK Centre for Ecology & Hydrology

Hierarchical modelling

Part 3: A flexible approach

Lindsay Banin

UK Centre for Ecology & Hydrology

What we will cover in Session 6a

- Deepening our understanding of Bayesian hierarchical models
- Applying to different types of uncertainty
- Visualising and specifying our models
- Directed Acyclic Graphs (DAGs)
- Implementing our model using MCMC (JAGS and R – rjags)
- Practical part 2

- We are not aiming for COMPLETE understanding – this is a starting point that will give you enough to confidently make sense of other resources

Motivation for using hierarchical Bayes…

Why hierarchical models?

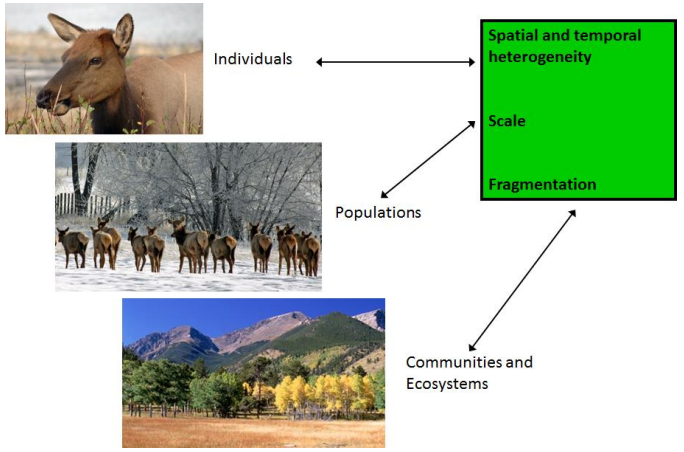
- Allow us to decompose complex, high dimensional problems into parts that can be thought about and analysed individually

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

- Construct models from simple interactions
- Building a network, focussing on local connections amongst elements
- Factor complex relationships into simple pieces
- Models can be constructed and solved in terms of stages
- ‘How does this component work, conditioned on those elements that directly affect it?’ (Clark 2005)

Is my model *hierarchical?*

- 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



Motivation for using hierarchical Bayes…

Clark (2005) Ecol. Letts.

Structure in space, time, among individuals/groups

Integrating data sources (different data/parameters)

Network of components!

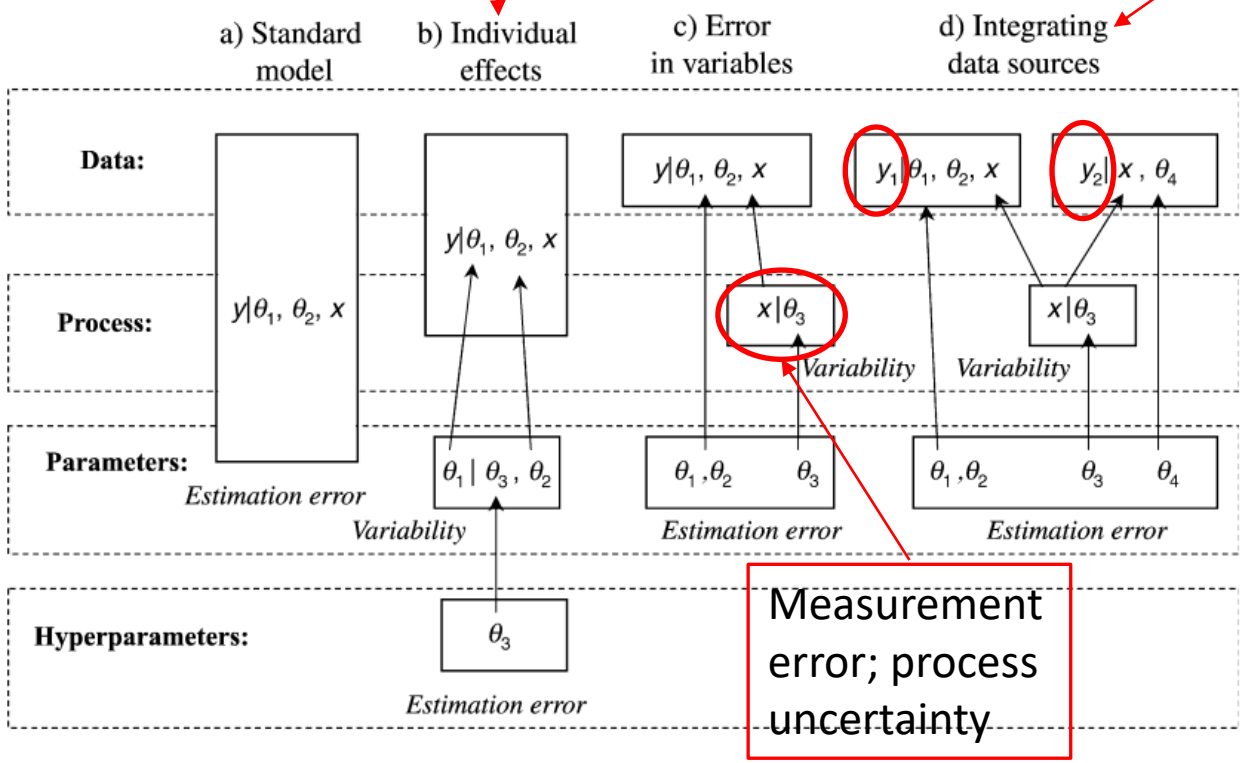


Figure 1 Four examples of how the Bayesian framework admits complexity (see text). Models can be viewed as networks of components, some of which are known and many unknown. The stages shown here, Data, Process, Parameters, and Hyperparameters, represent an overarching structure that admits complex networks. A model might include structure in space, time, or among individuals or groups (b), hidden processes (c,d), and multiple sources of information that bear on the same process (d). Acknowledging variability in a 'parameter' θ_1 (b) is accomplished by conditioning on additional parameters (θ_3). Now θ_1 occupies a middle stage and is truly variable, not just uncertain; θ_3 is asymptotic. Acknowledging variability in a predictor variable x (c) is accomplished in the same fashion.

Bayes law reminder

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

$$[\theta|y] = \frac{[y|\theta][\theta]}{[y]} \longrightarrow [\theta|y] \propto [y|\theta][\theta]$$

Posterior Likelihood Prior

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

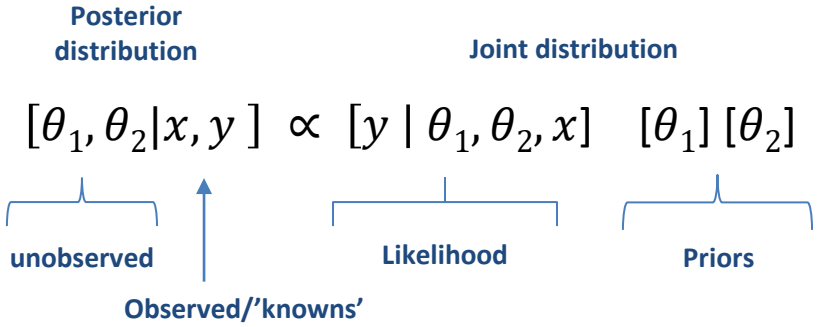
θ are unobserved quantities of interest (e.g., model parameters)

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

In other words, we factor $[y, \theta]$ into **ecologically sensible components** that can be estimated using MCMC as univariate (single-variable) distributions

Writing out a *simple* Bayesian model

Equivalent to Fig. 1a in Clark (2005).



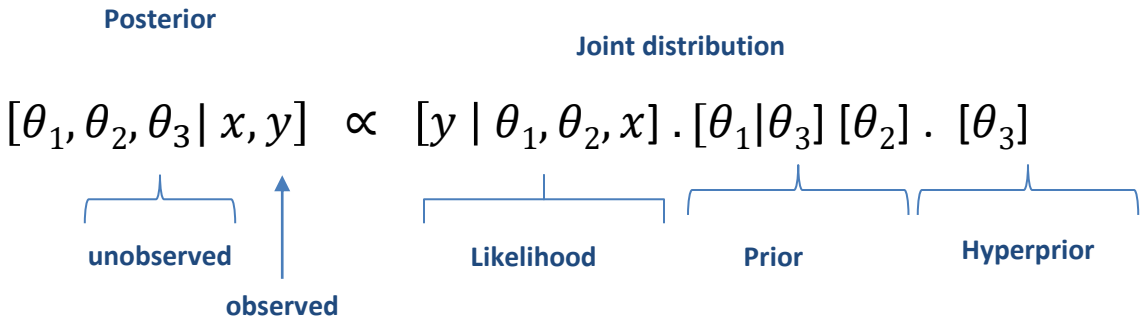
- This model is not hierarchical because there is no conditioning beyond the dependence of the data, y, on the unobserved quantities, θ_1, θ_2

- X is a predictor variable and is assumed to be known, so does not have a prior distribution

- The parameters on the right hand side of the conditioning symbol in the likelihood are *found in a prior – the priors allow for uncertainty*

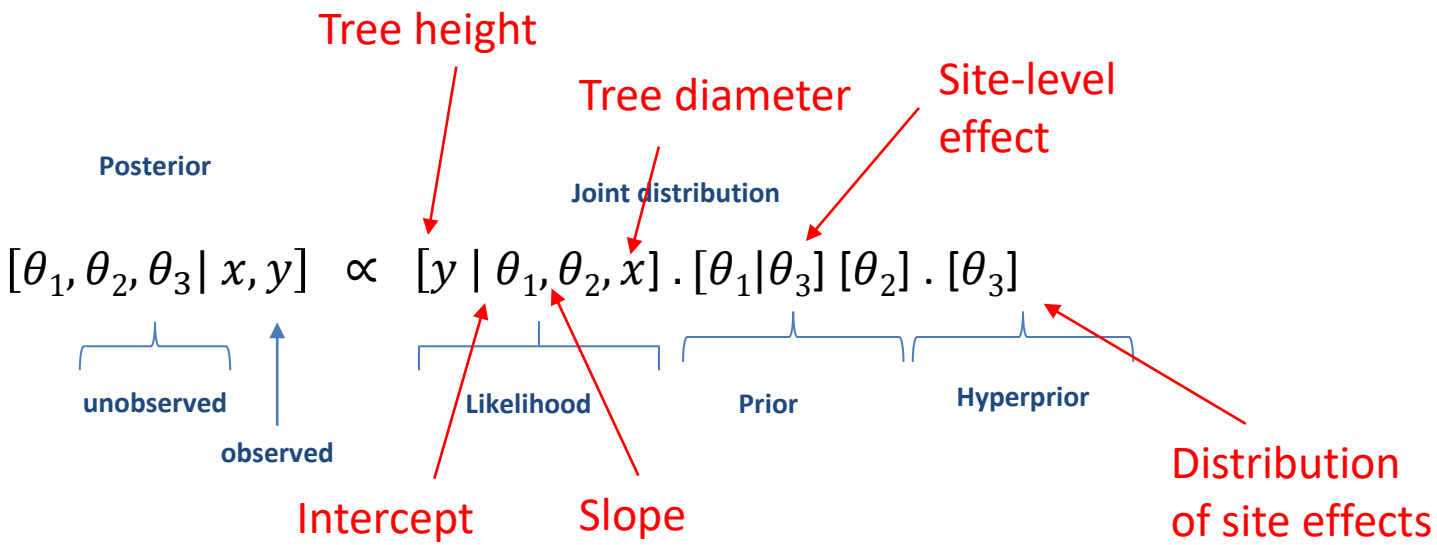
Writing out a *hierarchical* Bayesian model (1)

Equivalent to Fig. 1b in Clark (2005).



- Now θ_1 is treated like a random variable or population with its own parameters
- No prior on θ_1 because it is conditional on θ_3
- **A Bayesian model is hierarchical whenever we use probability rules for factoring to express the joint distribution as a product of conditional distributions**

Writing out a *hierarchical* Bayesian model (1)

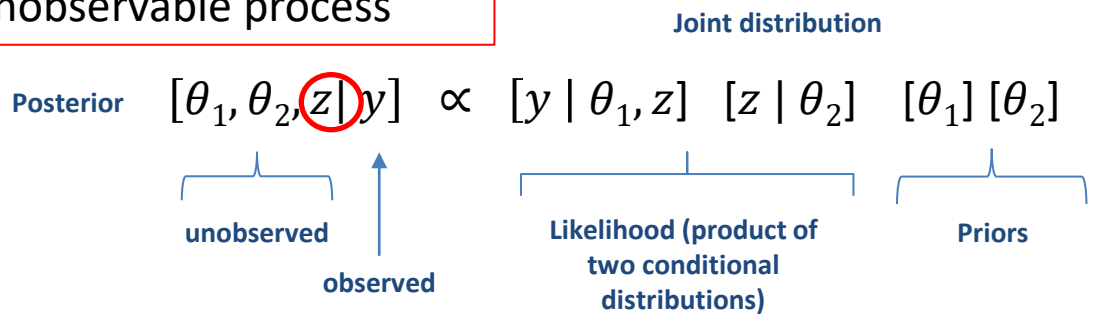


- **A Bayesian model is hierarchical whenever we use probability rules for factoring to express the joint distribution as a product of conditional distributions**

Writing out a *hierarchical* Bayesian model (2)

Equivalent to Fig. 1c in Clark (2005).

Sampled with error;
latent state;
unobservable process



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

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

- Process; latent states…

Defining models

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

True State we are trying to model

Parameters of interest

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

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

Data observations

Model

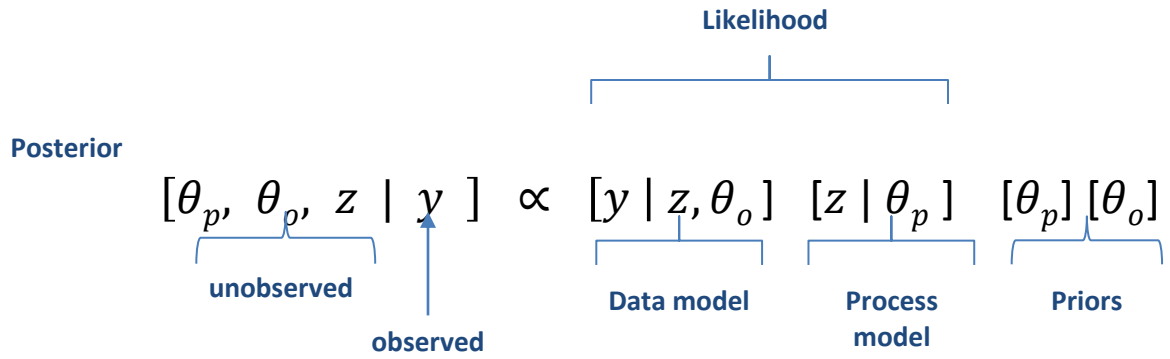
We learn about the process (our model) using our data

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

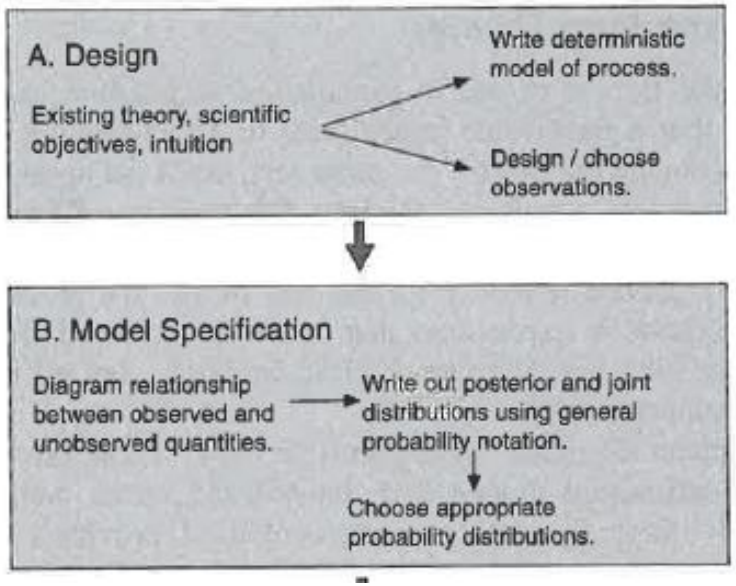
- We relate that state to observable data, y, using a model with a vector of parameters, θ_o

- The behaviour of the true state, or the process, is predicted by a model with parameters, θ_p




© Lindsay Banin

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

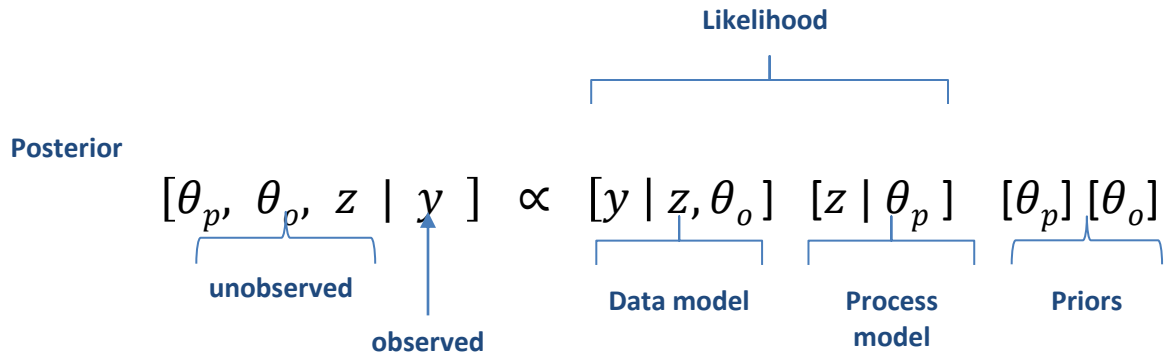
- Diagram it
- Write out posterior and joint distributions
- Choose probability distributions

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

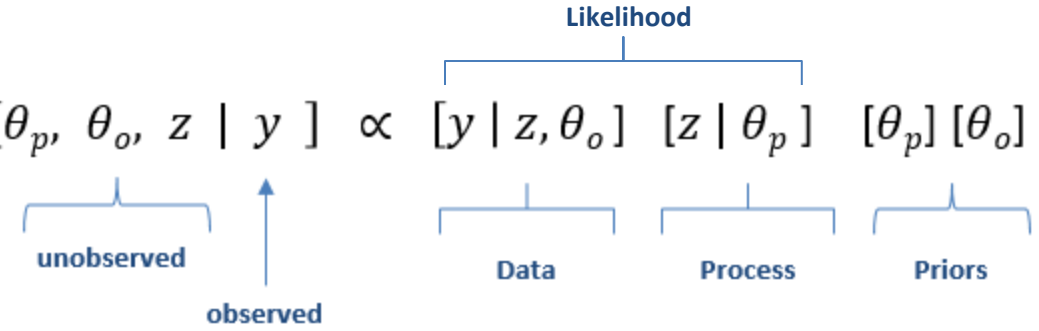
- There is a true ecological state of interest, z, that is not directly observable

- 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



Let's break this down….



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

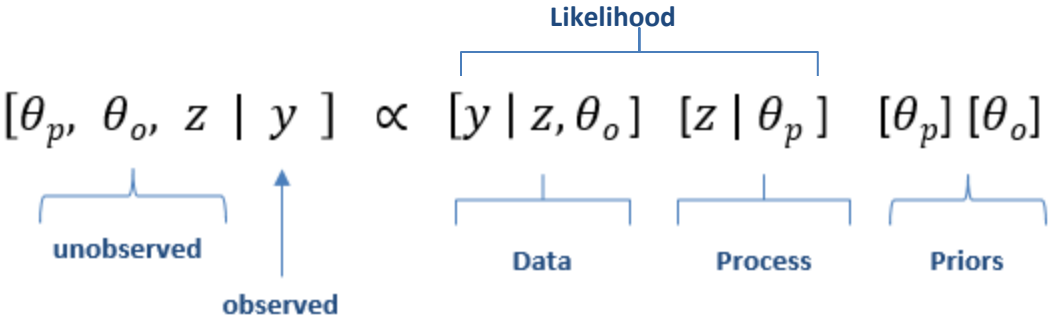
- *z* is the quantity we would observe if we could perfectly observe the instance of the true state, without any bias injected by our observation process

- 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

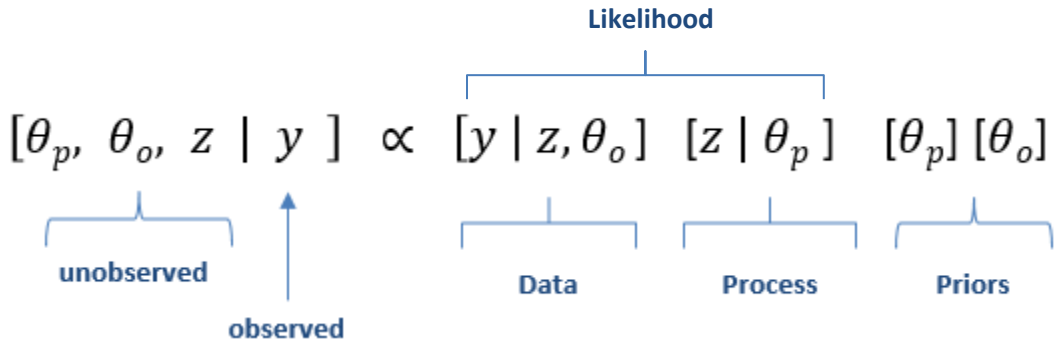
- We think about the true state of a system, *z* (e.g. the size of a population, the number of offspring per individual)

- 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



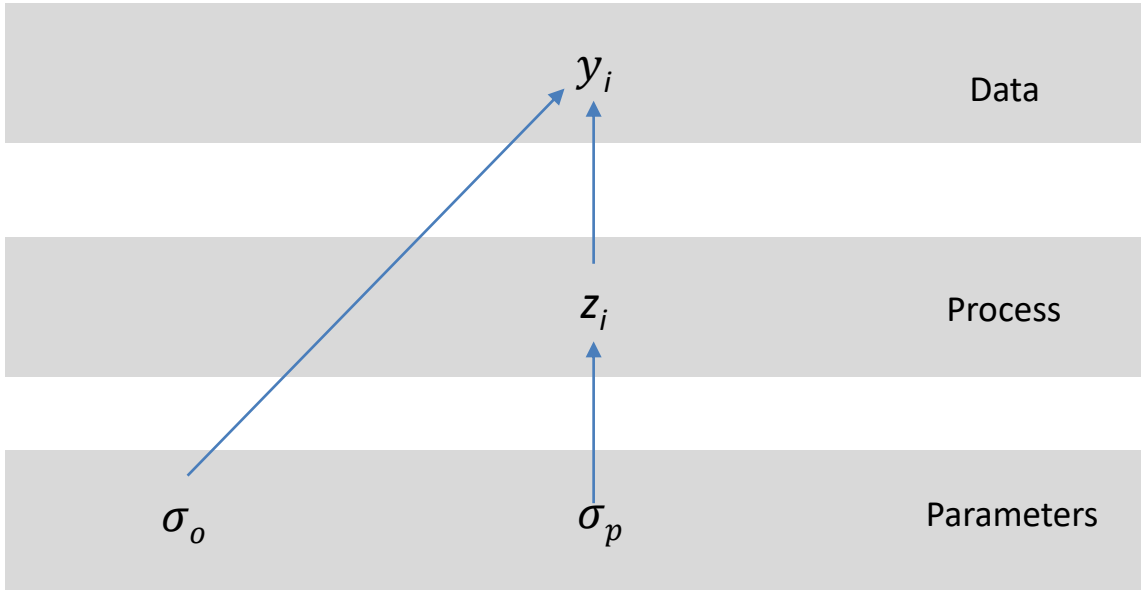
data model + process model + priors =

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)**


© Lindsay Banin

How can DAGs help us specify our models?



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

Likelihood

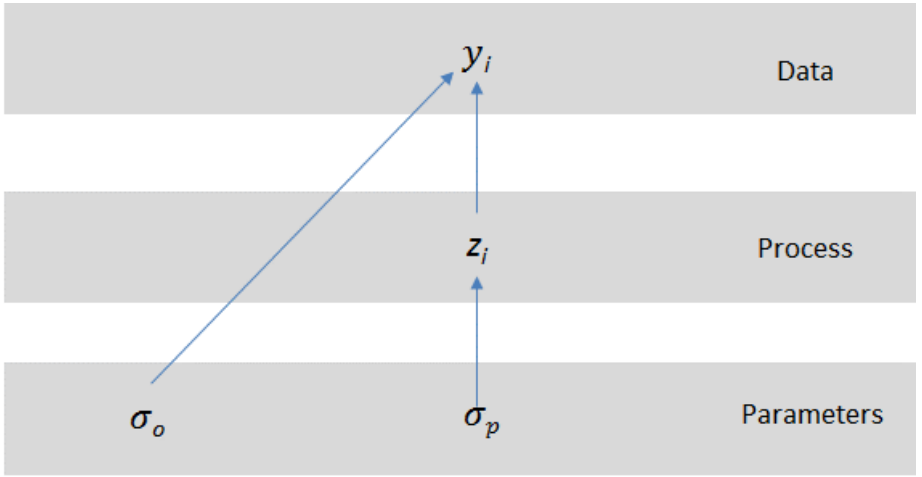
unobserved

observed

Data

Process

Priors



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

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

Example: Modelling light limitation of plant growth

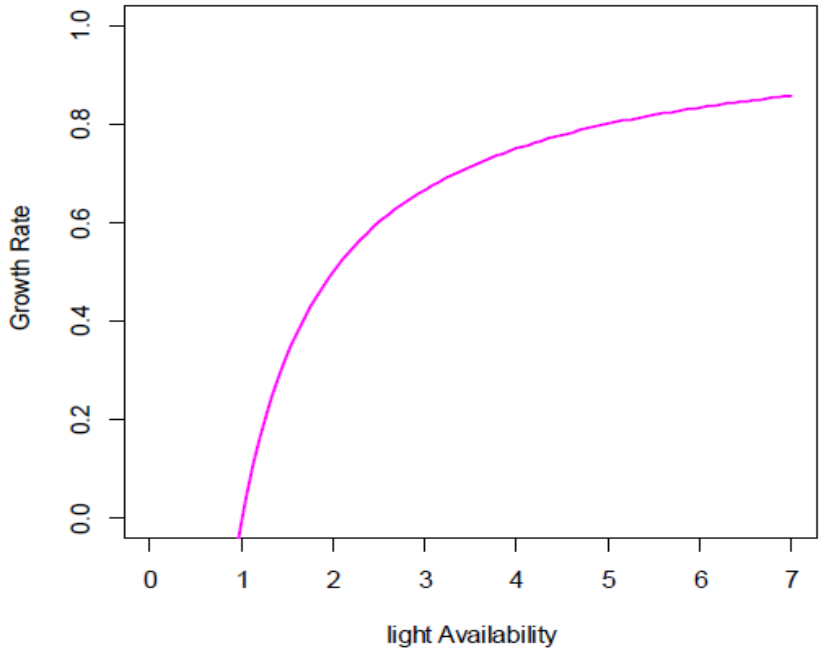
The relationship between plant growth rate and light tends to be non-linear, approaching an asymptote under high light conditions (no matter how much extra light you give it, it can't grow faster!)

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*) = this curve is our **process model,** and it has three parameters to describe it.

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)



Example: Modelling light limitation of plant growth

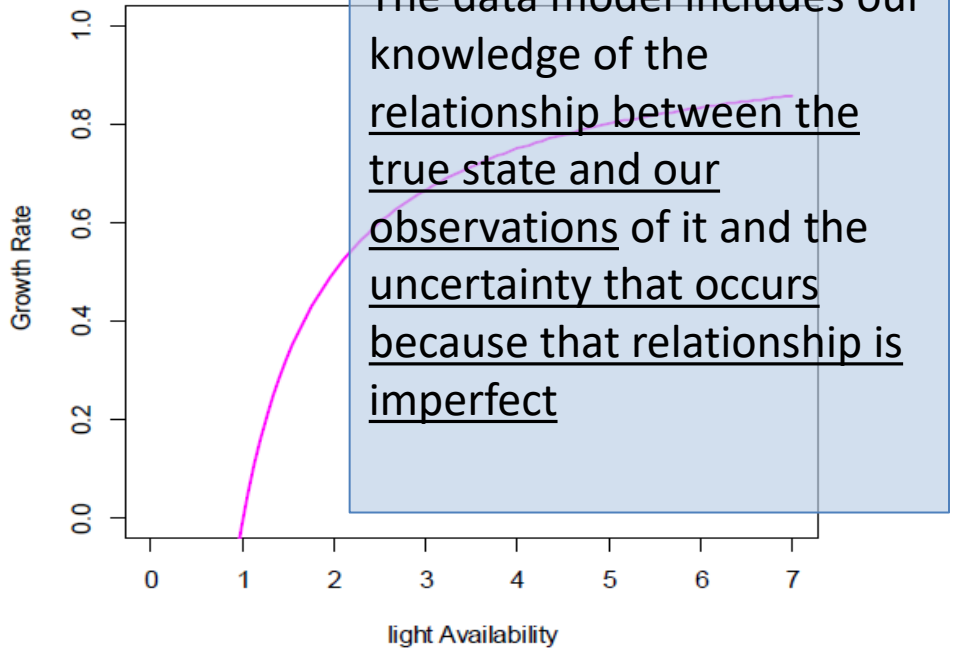
The relationship between plant growth rate and light tends to be non-linear, approaching an asymptote under high light conditions (no matter how much extra light you give it, it can't grow faster!)

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*) = this curve is our **process model,** and it has three parameters to describe it.

Michealis-Menton equation

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

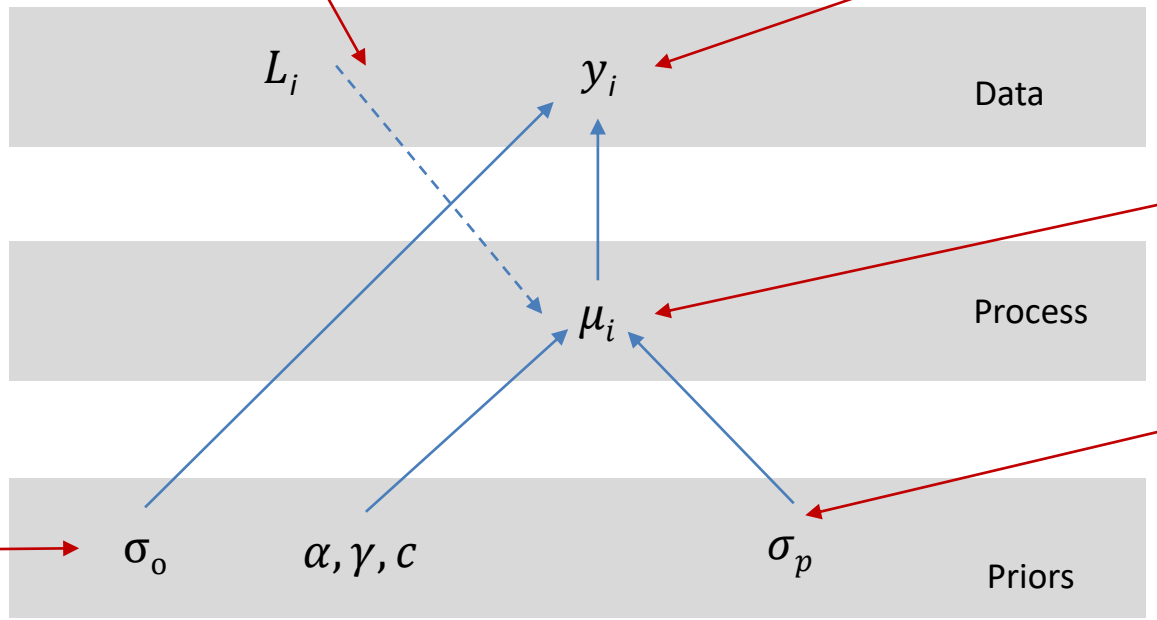
α = max. growth at infinite light
γ = rate at which curve tails off
c = light level at which growth is zero (x intercept)



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

Example: Bayesian model

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



Observations made with error

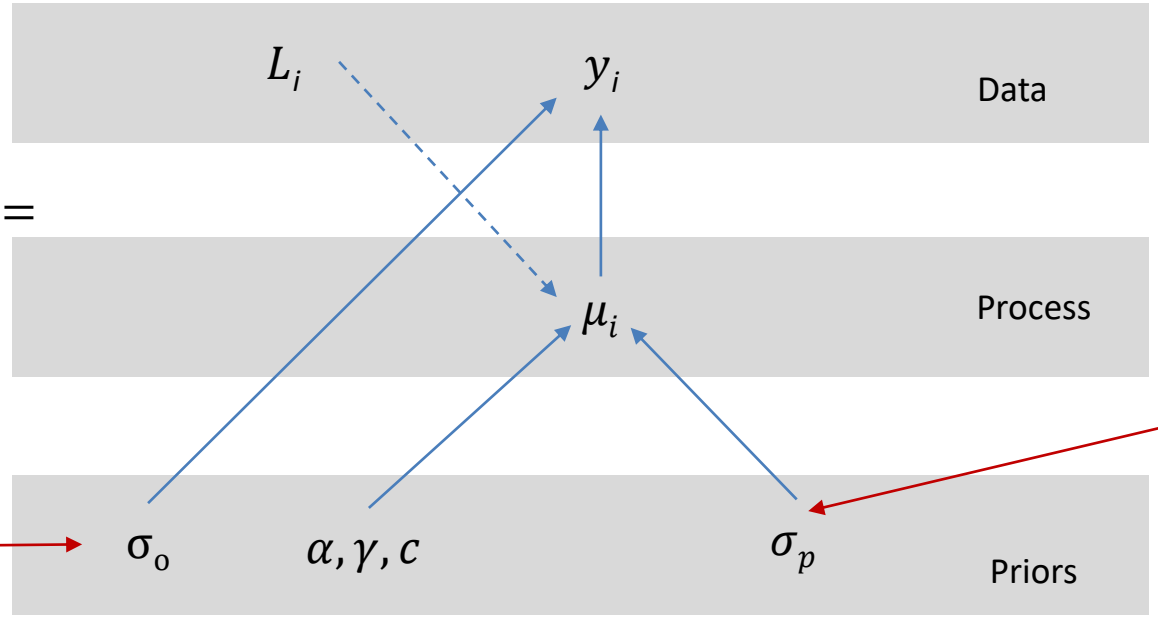
Unobserved true state

Process uncertainty

Observation uncertainty

Process model = $g(\alpha, \gamma, c, Li) = \mu_i =$

$$\frac{\alpha(L_i - c)}{(\alpha/\gamma) + (L_i - c)}$$



Observation uncertainty

Process uncertainty

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

unobserved

observed

Likelihood

Priors

Process model = $g(\alpha, \gamma, c, Li) = \mu_i =$

$$\frac{\alpha(L_i - c)}{(\alpha/\gamma) + (L_i - c)}$$

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

unobserved | observed | Likelihood | Priors

- We choose a normal distribution for *y* growth rate (can be + or -)
- We choose a normal distribution for μ 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 (asymptote) and γ (rate) because they are positive random variables
- We choose a uniform for *c (intercept)* 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)

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

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

unobserved

observed

Likelihood

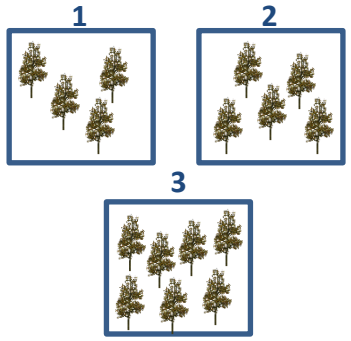
Priors

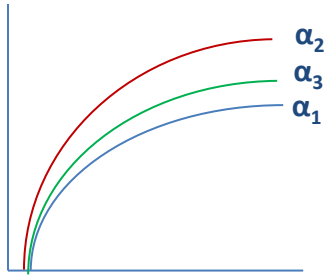
```
# JAGS model

# y is the observed growth rate
# x is the measurement of light, L

# a is alpha, the max. growth rate at infinite light
# c is light level at which growth is zero
# b is rate at which curve tails off (gamma)
```

```
##############################################################################
model{
  ### likelihood
  # Data model
  for (i in 1:n)
  {
    y[i] ~ dnorm(mu[i], tau.o) #Note JAGS uses precision, not variance
  }

  # process model
  for (i in 1:n)
  {
    mu[i] ~ dnorm(mu2[i], tau.p) #Note JAGS uses precision, not variance

    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: convert precisions to standard deviations
  tau.o <- 1/(sigma.o * sigma.o)
  tau.p <- 1/(sigma.p * sigma.p)

} # end of model
##############################################################################
```

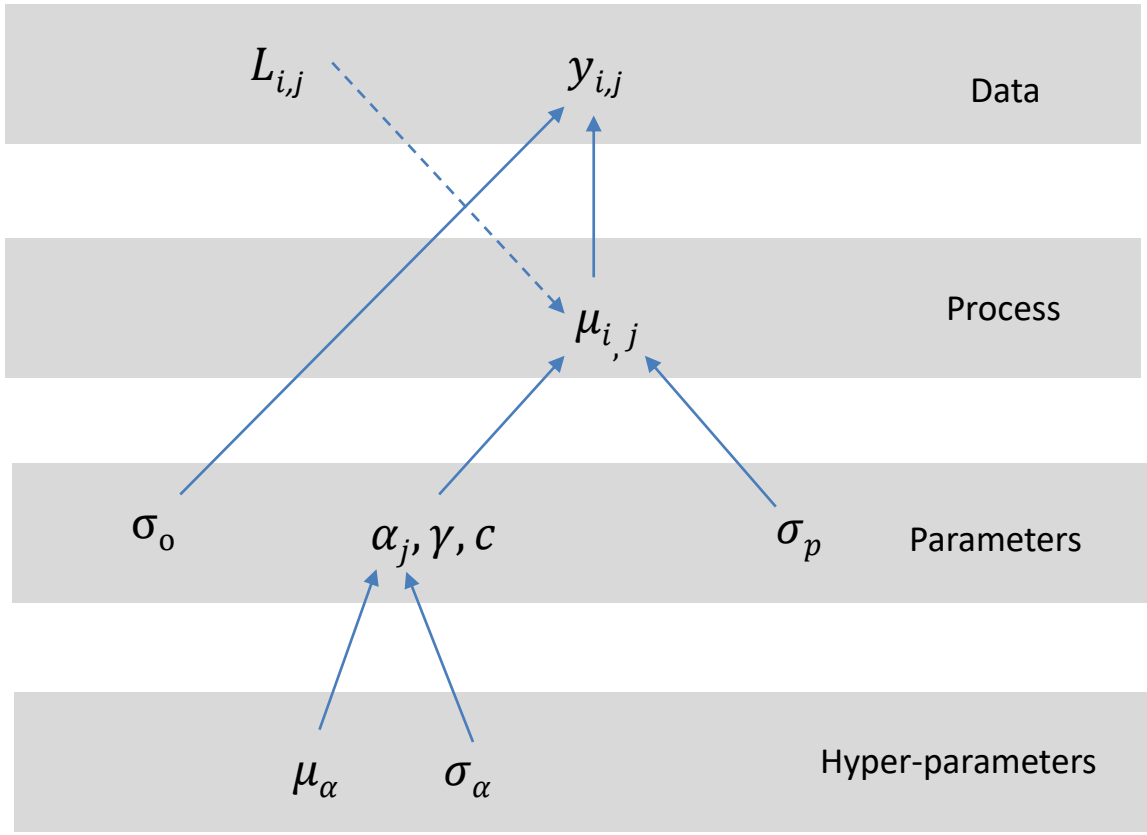
Practical 6a. Fit hierarchical Bayesian model in R and JAGS

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





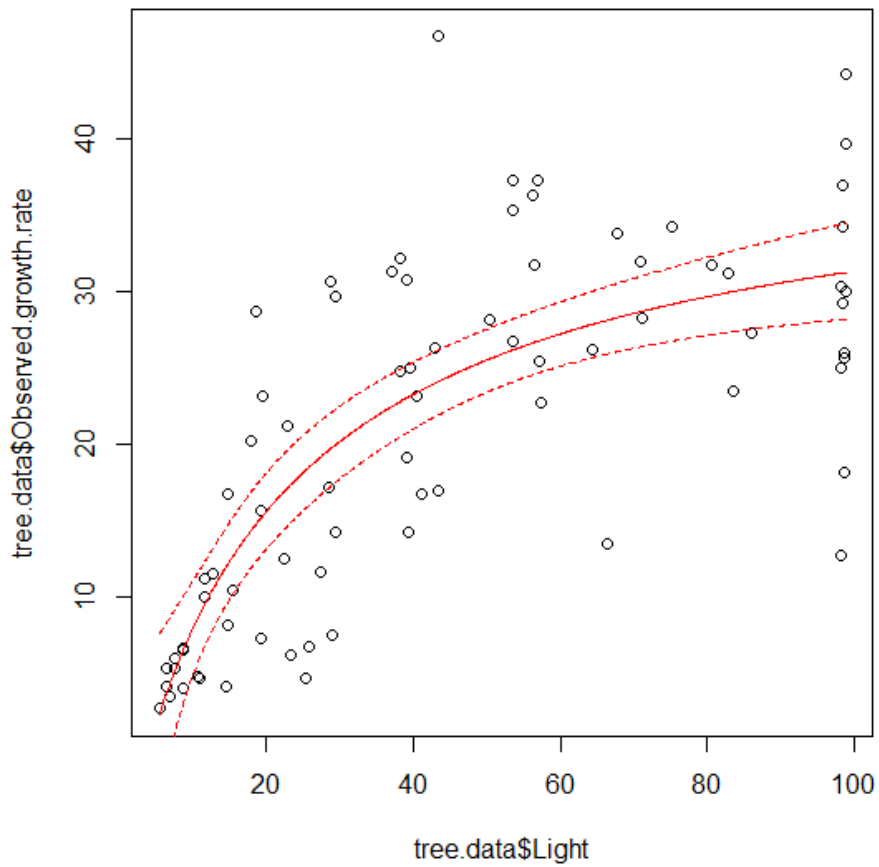
Multiple sites, *j*, and we expect there to be differences in the maximum growth rate per site, α_j, for instance due to soil water availability

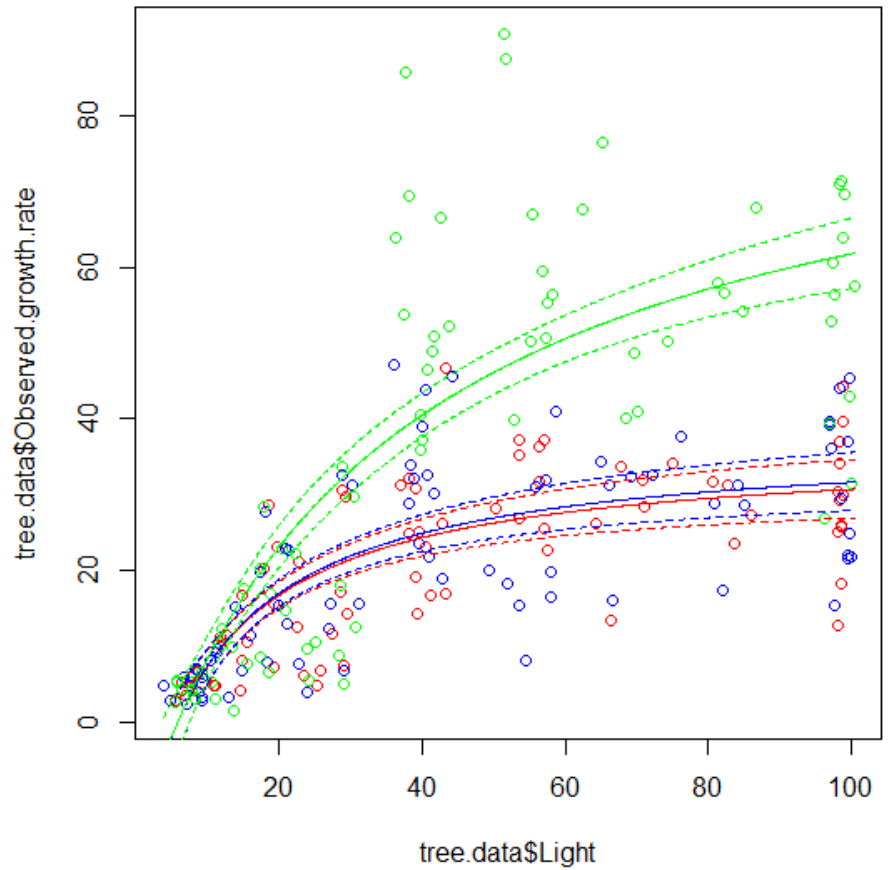


$$[\alpha_j, \mu_{i,j}, \gamma, c, \sigma_p, \sigma_o, \mu_\alpha, \sigma_\alpha | y_{i,j}] \propto \prod_{i=1}^{n} \prod_{j=1}^{3} [y_{i,j} | \mu_{i,j}, \sigma_o] \times \prod_{i=1}^{n} \prod_{j=1}^{3} [\mu_{i,j} | g(c, \gamma, \alpha_j), \sigma_p] \times \prod_{j=1}^{3} [\alpha_j | \mu_\alpha, \sigma_\alpha]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

Process model

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


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


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


© Lindsay Banin

What we covered in this session

- We have linked Bayesian theory more explicitly to our modelling process
- We have shown how diagramming can help to grapple with model structure, especially as they grow in complexity (variables, latent variables, grouping factors….)
- We have demonstrated another flexible tool (AKA JAGS) for explicitly writing, running and evaluating our own models

Any Questions?

This concludes

Hierarchical modelling Part 3

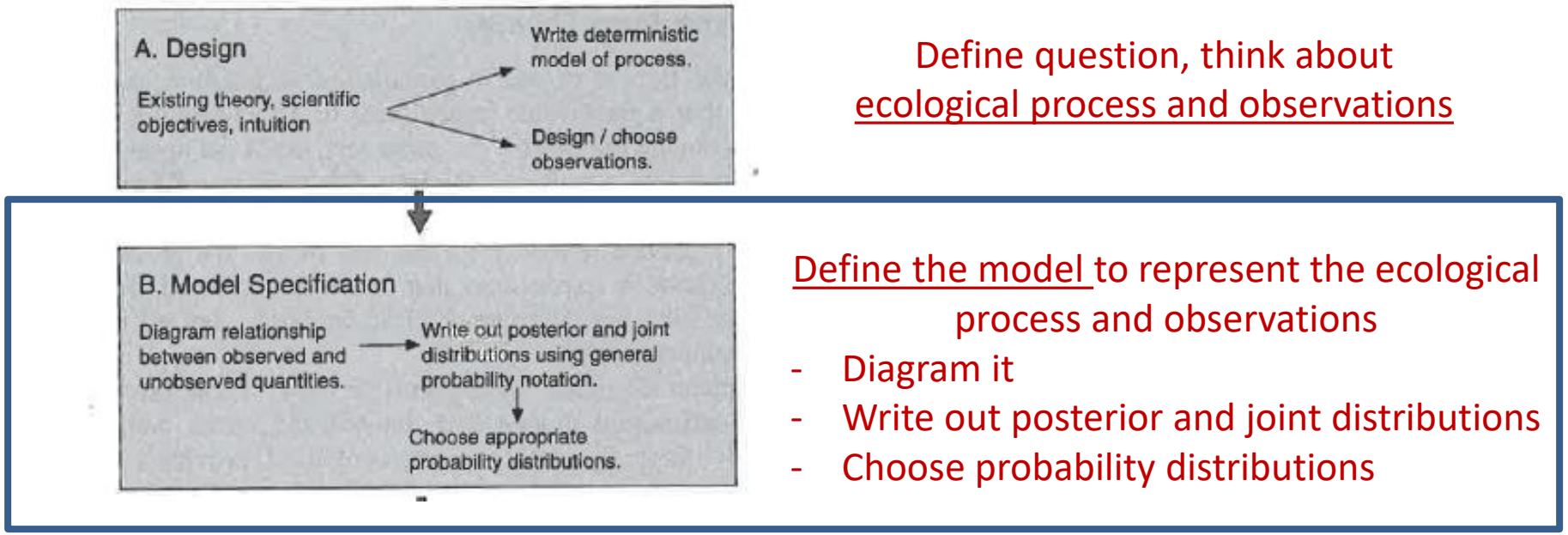
UK Centre for Ecology & Hydrology

Directed Acyclic Graphs (DAGs)
(Graphical modelling, Bayesian Networks)

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 software/package (e.g., BUGS, JAGS, STAN) to fit the model and estimate the parameters of interest



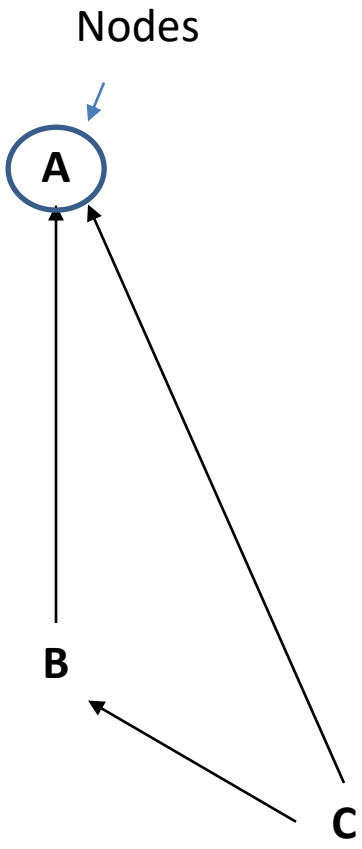
Define question, think about ecological process and observations

Define the model to represent the ecological process and observations

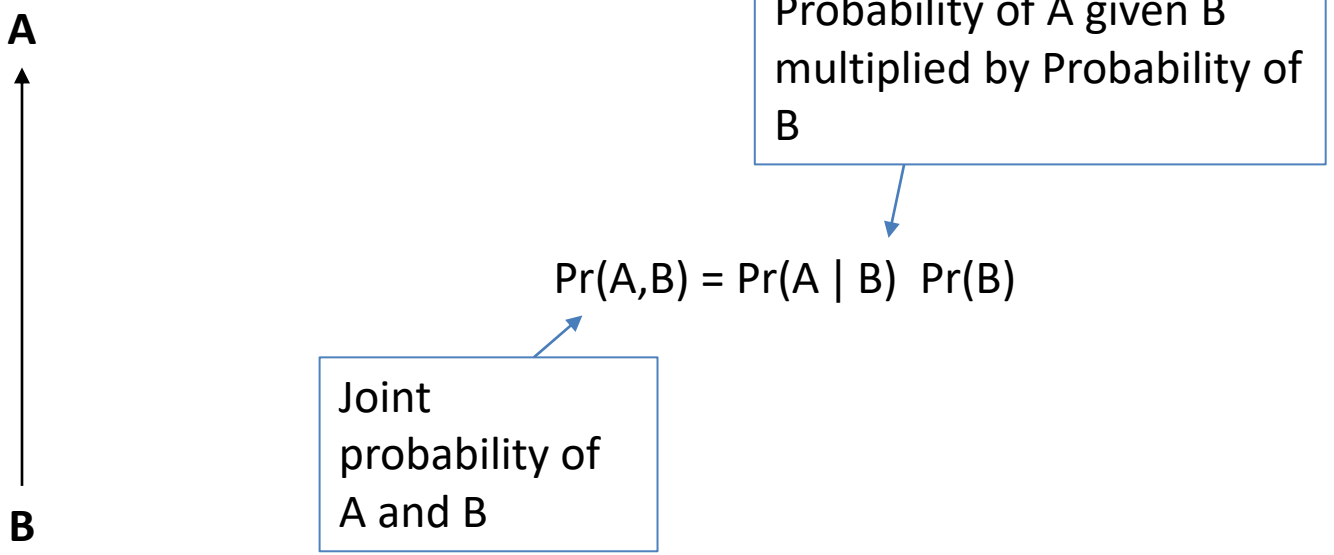
- Diagram it
- Write out posterior and joint distributions
- Choose probability distributions

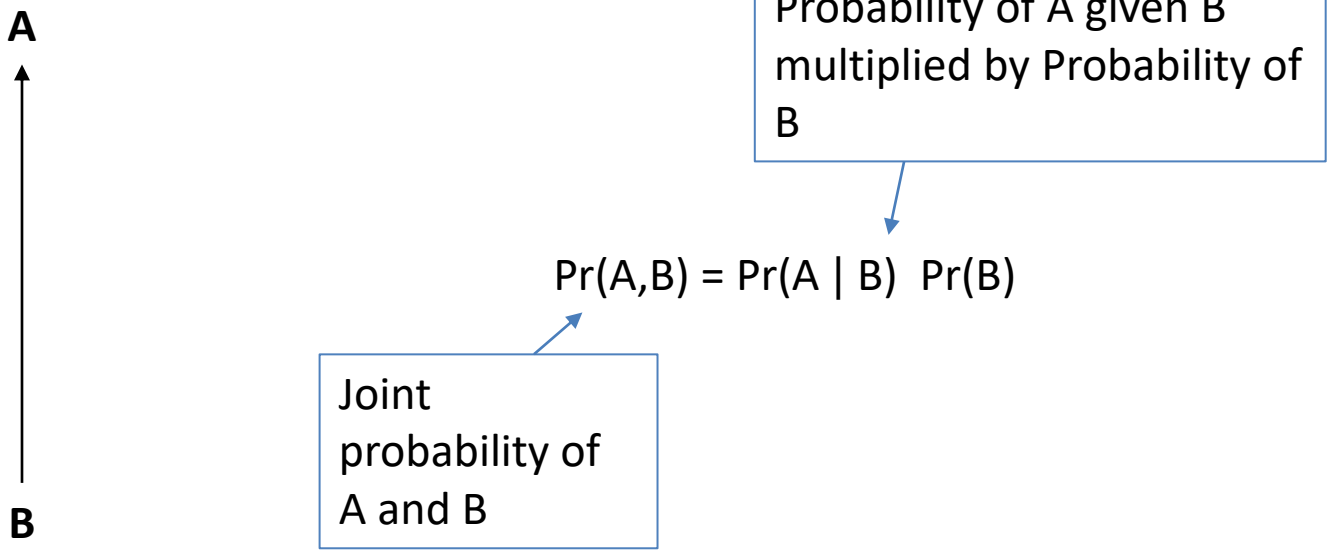
Directed Acyclic Graphs (DAGs)

- Describe a complex system in a simple way
- Make statements about conditional dependence and independence
- Provide a basis for computation
- Nodes are random variables
- Two nodes linked by an arrow are dependent (direction of arrow shows direction of dependence) – parent and child nodes
- Nodes not connected, with no common ancestors, are marginally independent

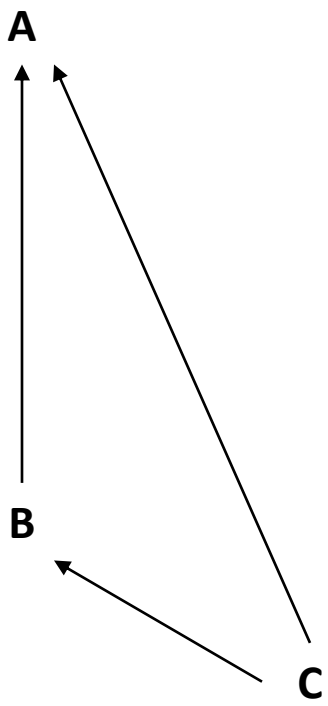


- Nodes (random variables) at the heads of arrows appear on the LHS of the conditioning |
- Nodes at the tails of arrows appear on the RHS of the conditioning |
- Nodes at the tails of arrows with no arrow leading to it are expressed as priors
- Solid arrows are stochastic relationships among random variables
- Tails of arrows specify parameters defining the distribution of the random variable at the head of the arrow


A
B
Probability of A given B multiplied by Probability of B
Pr(A,B) = Pr(A | B) Pr(B)
Joint probability of A and B

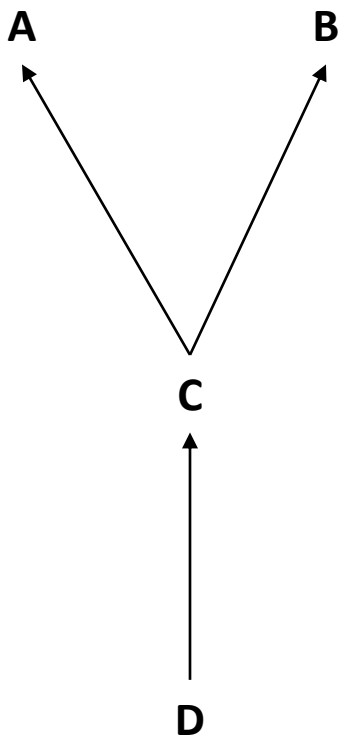


- All random variables (letters; joined by solid arrows) in the diagram are in the joint probability distribution
- If it is at the head of the arrow, there is dependency – conditional on the quantity at the end of the arrow
- Anything on the end of an arrow with nothing else feeding in and therefore expressed unconditionally - prior

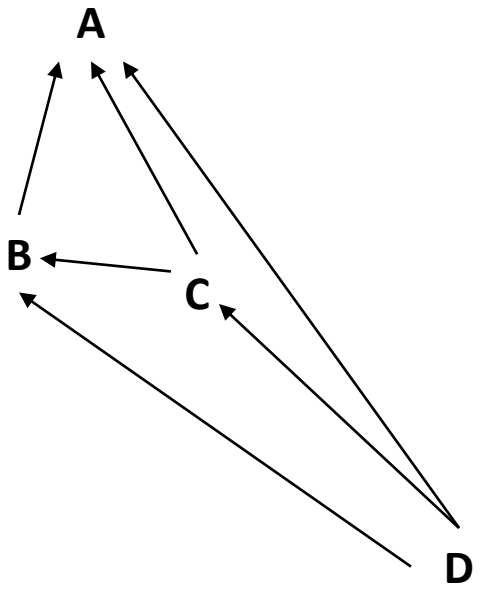


Here, A is conditional on B and C; B is conditional on C

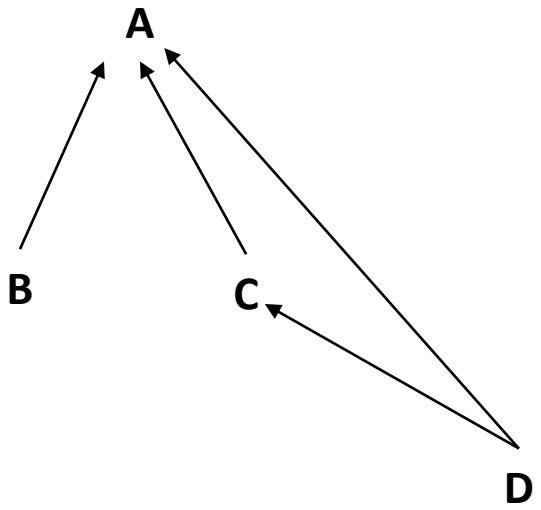
$$Pr(A,B,C) = Pr(A \mid B, C) \times Pr(B \mid C) \times Pr(C)$$



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



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



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

Linking DAGs to our modelling process