INFO 510 Bayesian Modelling and Inference

Lecture 10 – Hierarchical models

Dr. Kunal Arekar
College of Information
University of Arizona, Tucson

October 6, 2025

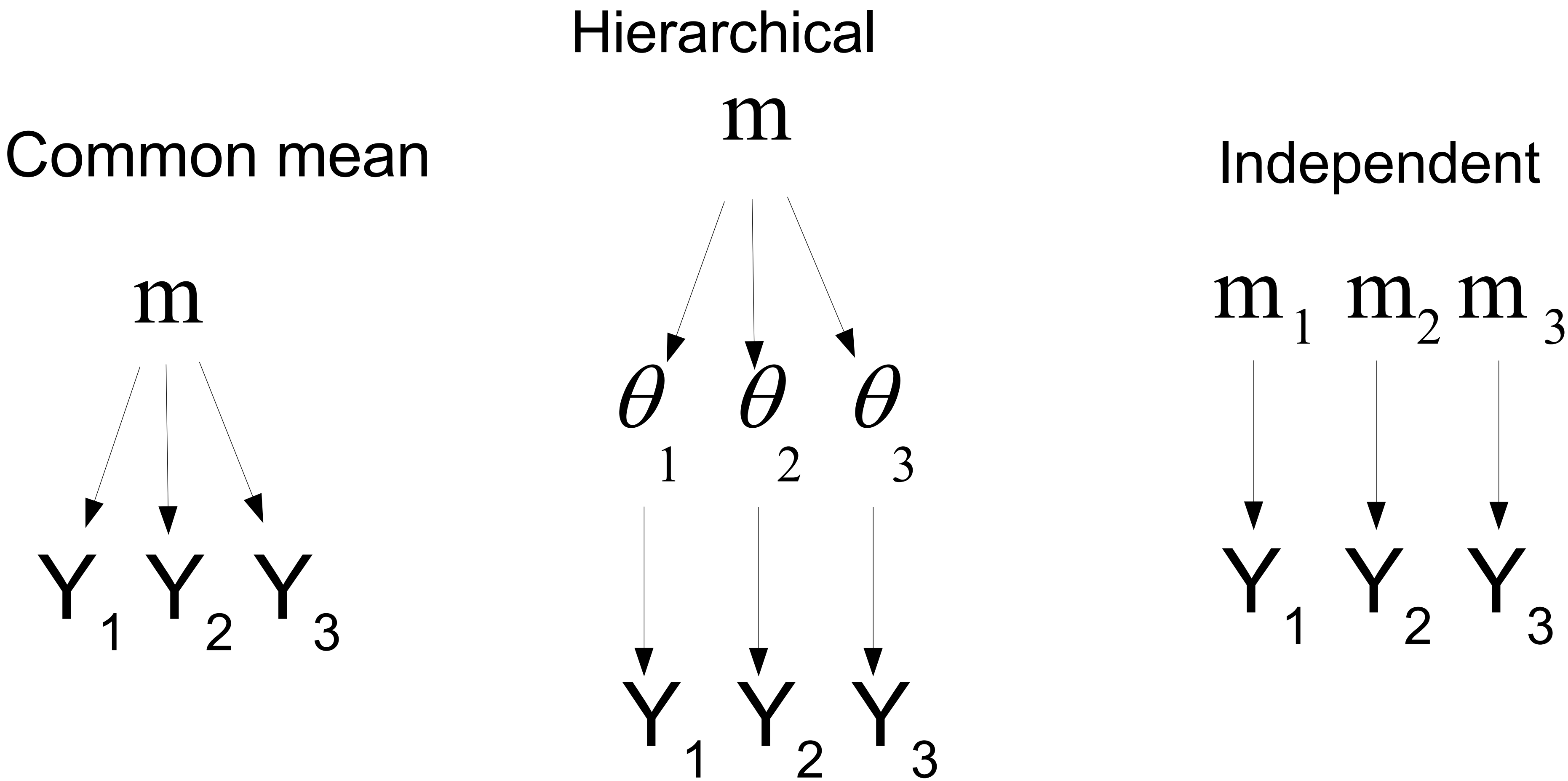
INFO 510 Bayesian Modelling and Inference

Lecture 10 – Hierarchical models

Dr. Kunal Arekar
College of Information
University of Arizona, Tucson

October 6, 2025

Hierarchical Models



Hierarchical modeling is a statistical method used to handle data with multiple levels of variability or nested structures

They are sometimes called multilevel models or mixed-effects models

Hierarchical models are powerful because they:

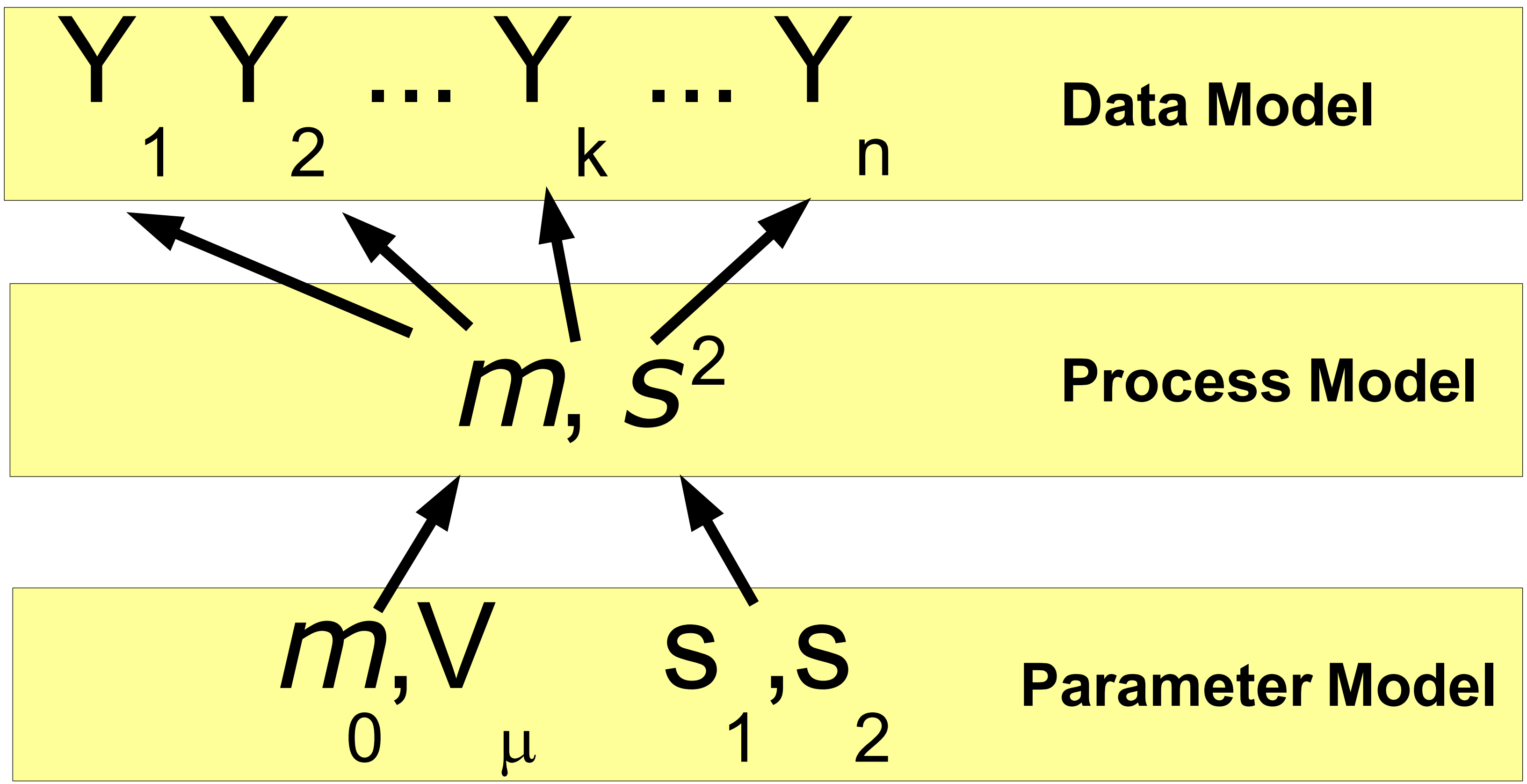
Capture group structure: Allow you to understand the impact of groups (e.g., schools) on individual outcomes (e.g., student scores)

Increase statistical efficiency: They borrow strength from group-level data to help estimate parameters even when individual-level data is sparse

Model complex variability: Enable you to account for different sources of variation at multiple levels of the data

Common Mean

$$\vec{y}_k \sim N(\mu, \sigma^2)$$



Hierarchical Mean, Common Variance

$$Y_k \sim N(\mu_k, \sigma^2)$$

$$\mu_k \sim N(\mu, \tau^2)$$

$$\sigma^2 \sim IG(s_1, s_2)$$

At this point, this model is fitting each data set independently but assume the mean for each has the same prior

Hierarchical Mean, Common Variance

$$Y_k \sim N(\mu_k, \sigma^2)$$

$$\mu_k \sim N(\mu, \tau^2)$$

$$\sigma^2 \sim IG(s_1, s_2)$$

For the hierarchical model, instead assume the prior contains unknown model parameters

Hierarchical Mean, Common Variance

$$Y_k \sim N(\mu_k, \sigma^2)$$

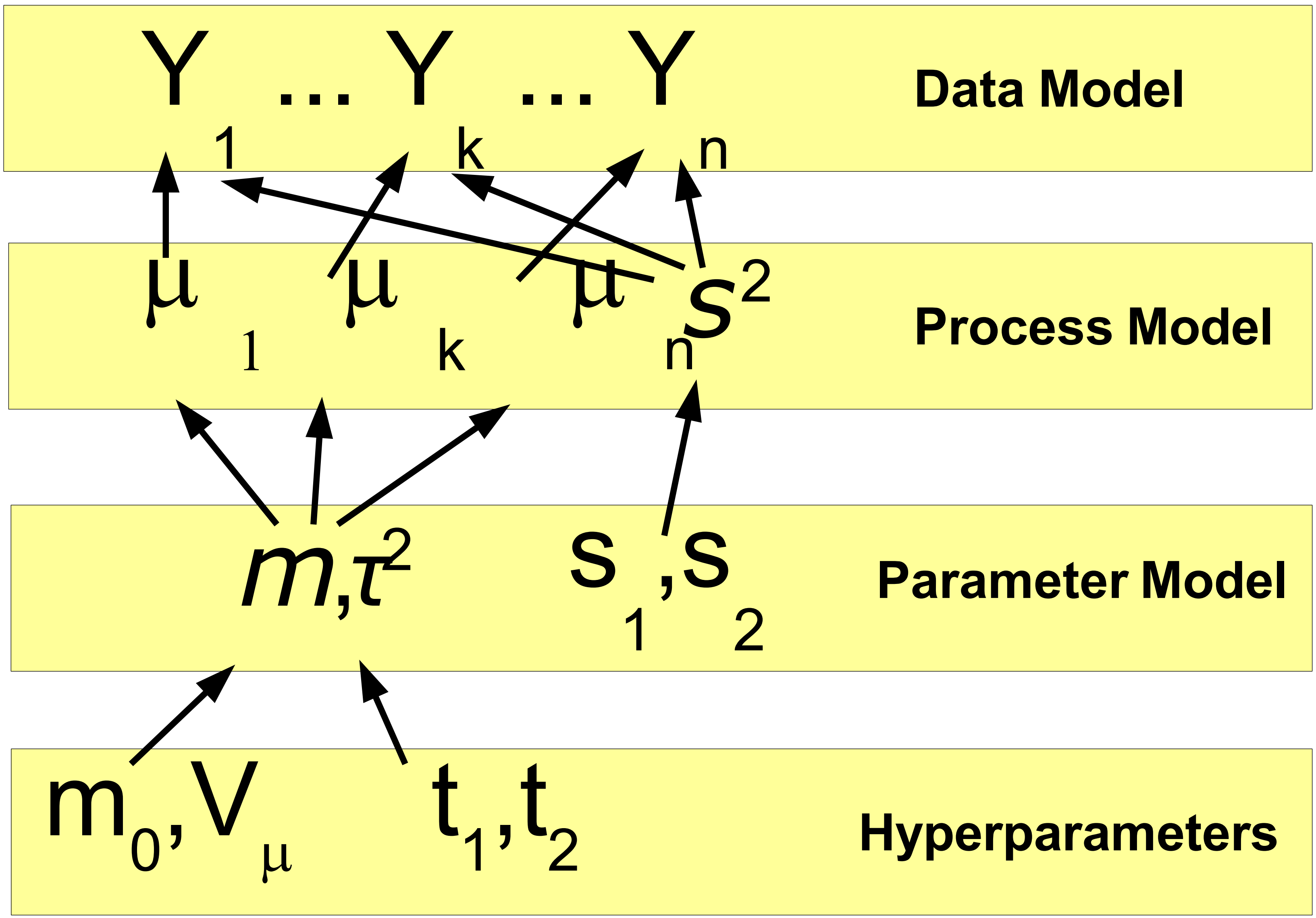
$$\mu_k \sim N(\mu, \tau^2)$$

$$\sigma^2 \sim IG(s_1, s_2)$$

$$\left.\begin{array}{l} \mu \sim N(\mu_0, V_\mu) \\ \tau^2 \sim IG(t_1, t_2) \end{array}\right\}$$

Then need to specify *hyperpriors* on our prior

Hierarchical Mean



Hyperpriors

if you're estimating a parameter θ, you assign it a prior probability distribution $P(\theta)$ based on your current knowledge

$$P(\theta|\text{data}) \propto P(\text{data}|\theta) \times P(\theta)$$

your belief about θ before seeing the data.

Hyperparameters: while θ is a parameter we are trying to estimate, the hyperparameters (μ and σ^2) govern your initial guess about where θ might lie.

Hyperpriors: Hyperpriors are priors on the hyperparameters aka assign a probability distribution to hyperparameters

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

first, you specify a prior for your parameter θ

μ and σ^2 are hyperparameter

$\mu \sim \text{Normal}(\mu_0, \tau^2)$

$\sigma^2 \sim \text{InverseGamma}(\alpha, \beta)$

Instead of assuming fixed values for μ and σ^2, you assign them hyperpriors

Here $\mu_0, \tau^2, \alpha, \beta$ are the hyperparameters of the hyperpriors.

The overall structure now looks like

$$P(\theta|\mu, \sigma^2) \cdot P(\mu|\mu_0, \tau^2) \cdot P(\sigma^2|\alpha, \beta)$$

Why Use Hyperpriors?

Hyperpriors are useful because they let you capture a richer structure of uncertainty

In many cases, it's hard to confidently set fixed values for hyperparameters like μ and σ^2. By assigning hyperpriors, you acknowledge this uncertainty and allow the data to inform not only the parameter of interest (θ) but also the underlying assumptions about the hyperparameters

Hierarchical Models

- Model variability in the parameters of a model
- Partition variability more explicitly into multiple terms
- Borrow strength across data sets
- Hierarchical with respect to parameters

Random Effects

- Common special case of Hierarchical models

$$Y_k \sim N(\mu_k, \sigma^2) \qquad Y_k \sim N(\mu_g + \alpha_k, \sigma^2)$$

$$\mu_k \sim N(\mu, \tau^2) \qquad \alpha_k \sim N(0, \tau^2)$$

$$\sigma^2 \sim IG(s_1, s_2) \qquad \sigma^2 \sim IG(s_1, s_2)$$

$$\mu \sim N(\mu_0, V_\mu) \qquad \mu_g \sim N(\mu_0, V_\mu)$$

$$\tau^2 \sim IG(t_1, t_2) \qquad \tau^2 \sim IG(t_1, t_2)$$

Random Effects

- Common special case of Hierarchical models

$$Y_k \sim N(\mu_g + \alpha_k, \sigma^2)$$

$$\alpha_k \sim N(0, \tau^2)$$

$$\sigma^2 \sim IG(s_1, s_2)$$

$$\mu_g \sim N(\mu_0, V_\mu)$$

$$\tau^2 \sim IG(t_1, t_2)$$

Random Effects

$$Y_k \sim N(\mu_g + \alpha_k, \sigma^2)$$

$$\alpha_k \sim N(0, \tau^2)$$

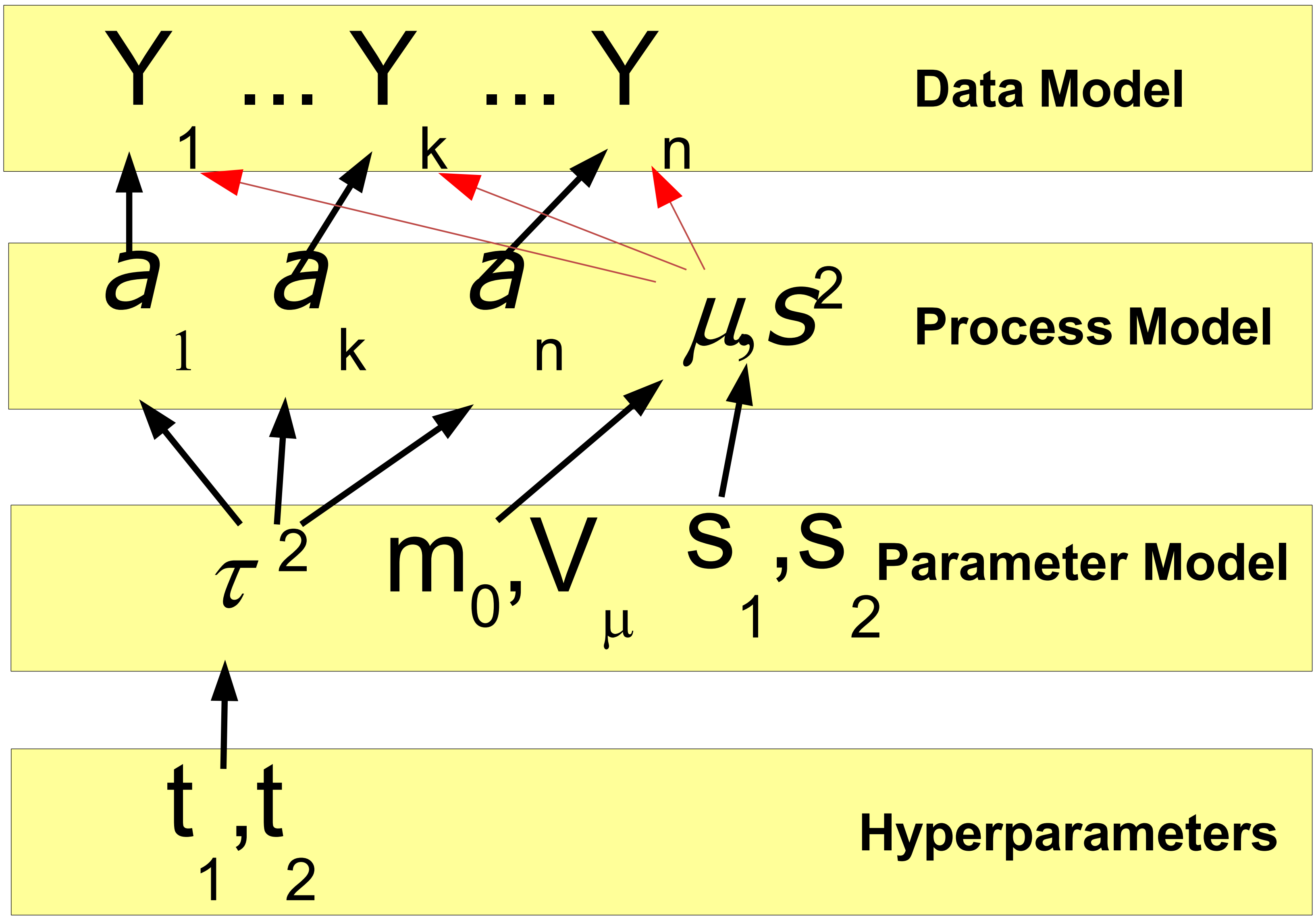
$$\sigma^2 \sim IG(s_1, s_2)$$

$$\mu_g \sim N(\mu_0, V_\mu)$$

$$\tau^2 \sim IG(t_1, t_2)$$

- Random effects always have mean 0
- Random effects variance attributes a portion of uncertainty to a specific source
- Can be used to try an account for a lack of independence

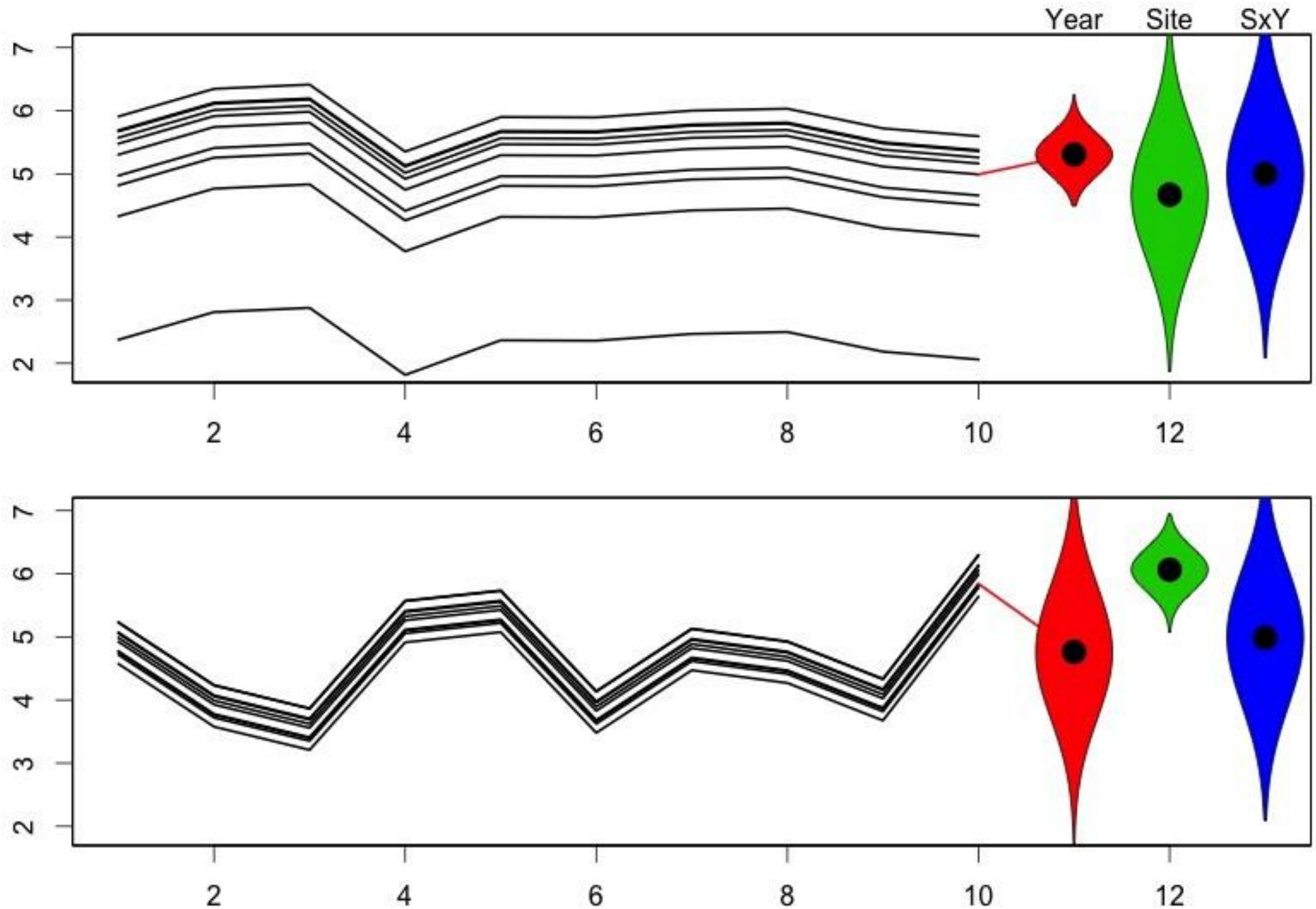
Random Effects Mean



What things can be random effects?

- Traditionally, random effects apply to aspects of the study that would not be the same if replicated
  - e.g. Plot, Block, Year, individual, etc.
  - Often used to account for a lack of independence
- Treatments and covariates of interest are usually treated as **fixed effects**
- Typically there is some degree of replication otherwise the random effect is not identifiably different from the residual “noise” term $e \sim N(0, \sigma^2)$

Why bother?
Impacts on inference...



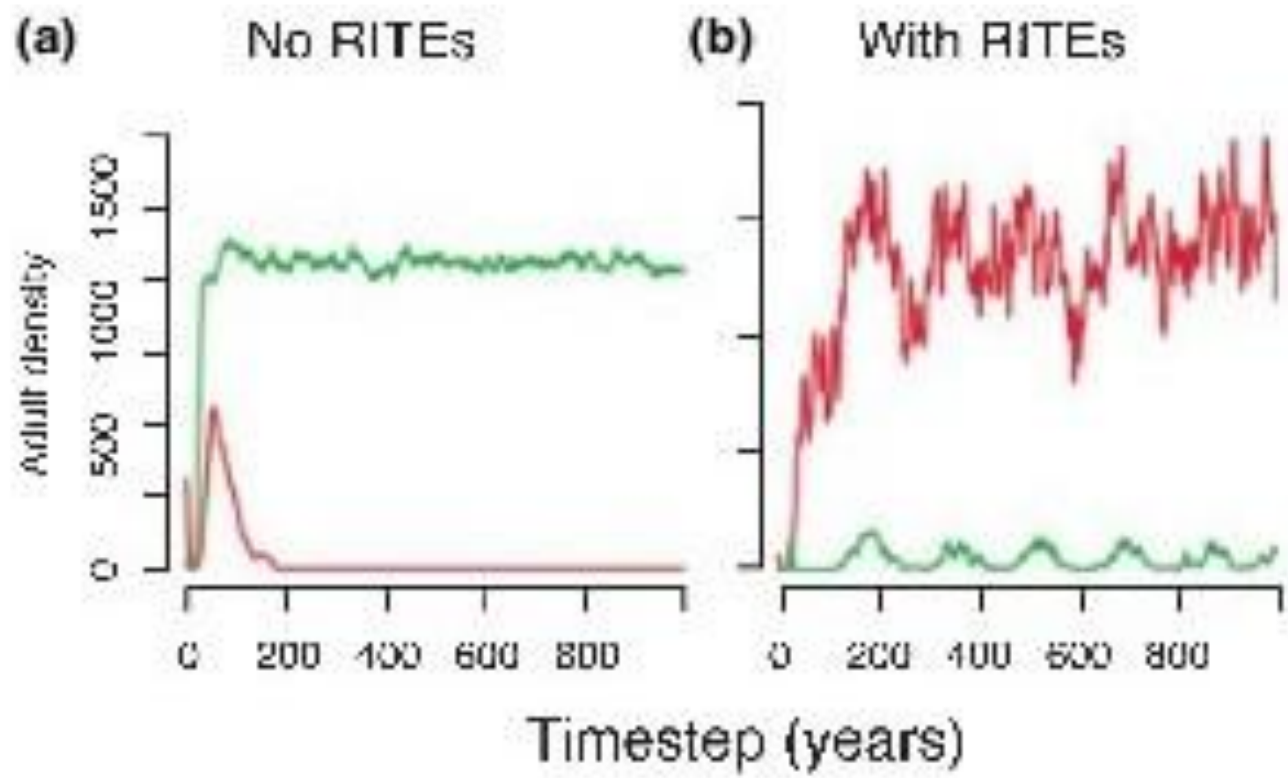
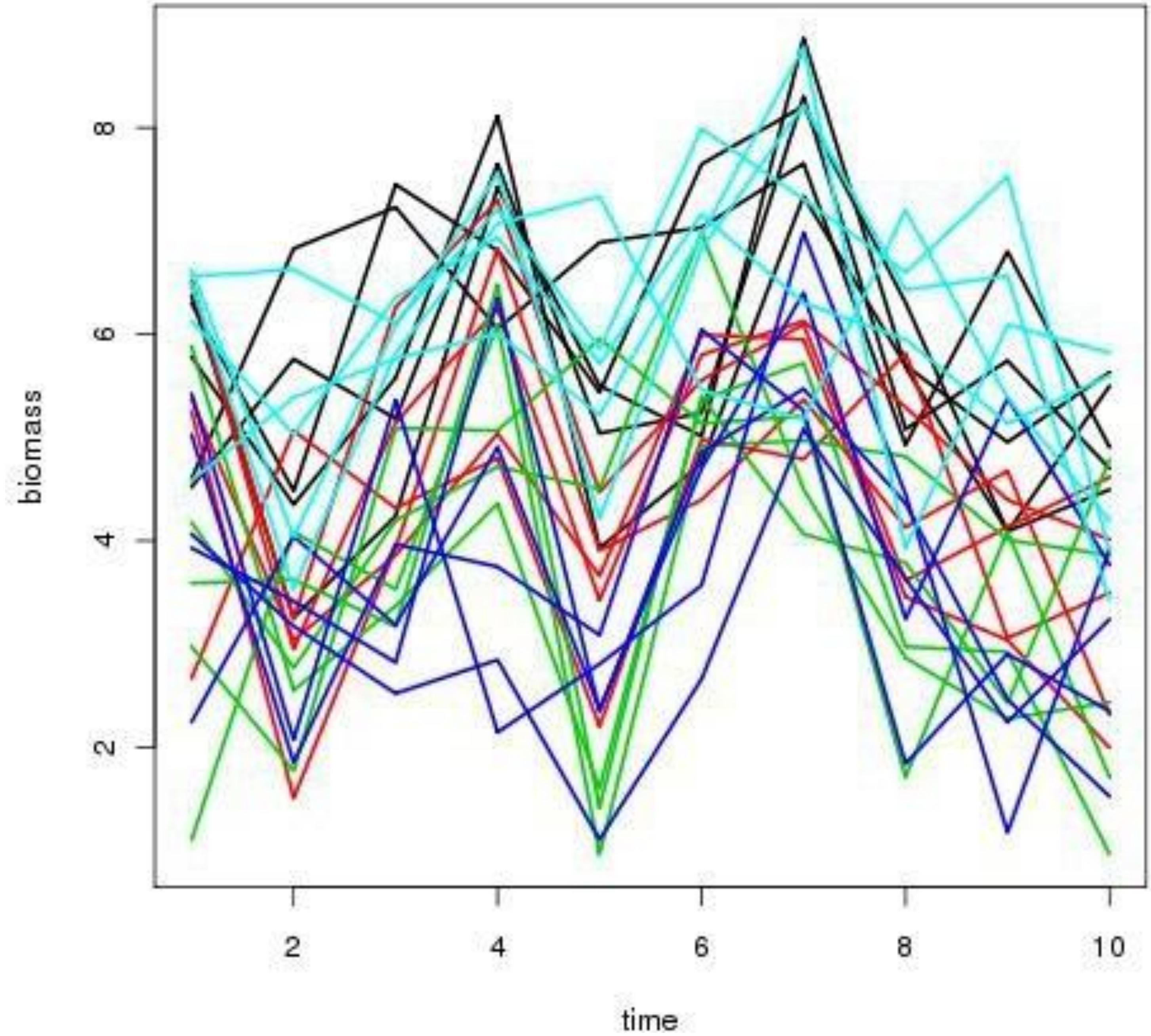


Figure 3 The impact of random individual effects (RITEs) on coexistence of two competing species. Two spatiotemporal and individual-based simulations were run using recruitment processes that are parameterized with data, summarized in Fig. 1. Panel (a) is the traditional approach having deterministic species differences and stochasticity in time, but no within-population heterogeneity, reflecting that fact the green species is the deterministic winner (Fig. 1a). Population heterogeneity in (b) means that green is not the deterministic winner, but rather both species win with some probability.

Start Simple

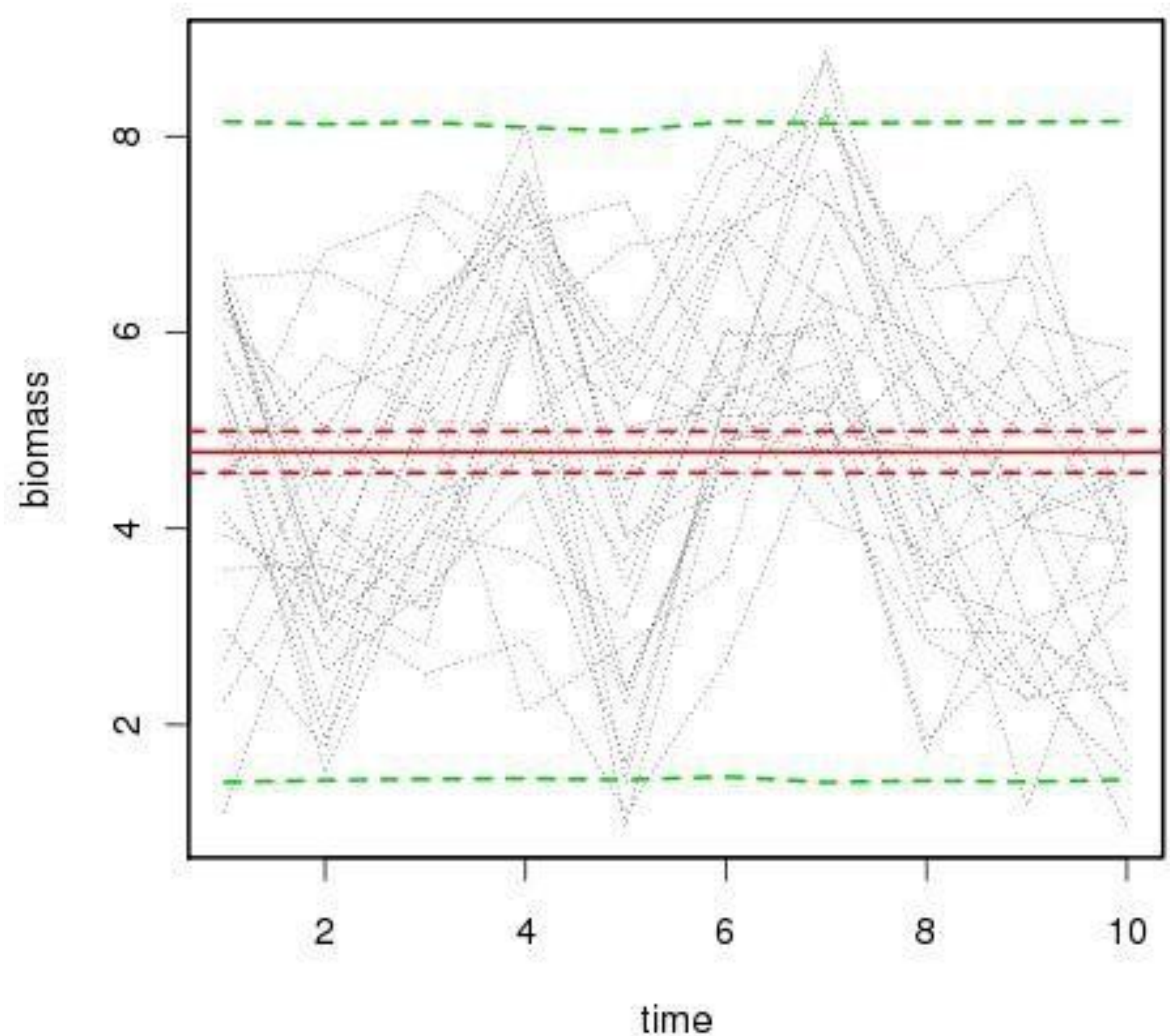
Progressively
Add Complexity

Example: Biomass by Block and Time



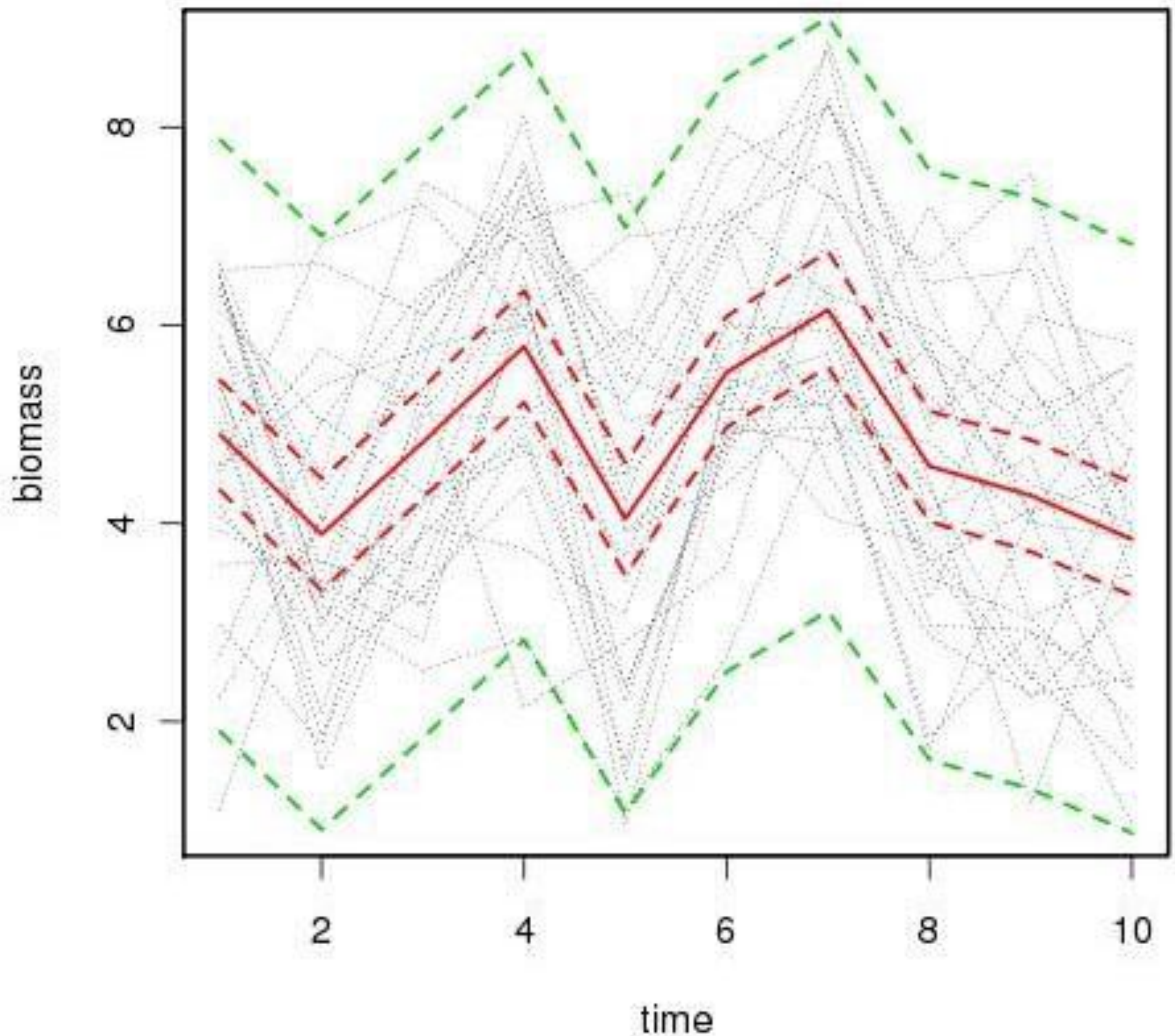
Model 1: Global Mean

```
model{
  mu ~ dnorm(0,0.001)                  ## priors
  sigma ~ dgamma(0.001,0.001)

  for(t in 1:nt){            ## time
    for(b in 1:nb){          ## block
      for(i in 1:nrep){      ## individual
        x[t,b,i] ~ dnorm(mu,sigma)
      }
    }
  }
}
```


biomass
8
6
4
2
2
4
6
8
10
time

Model 2: Random Temporal Effect

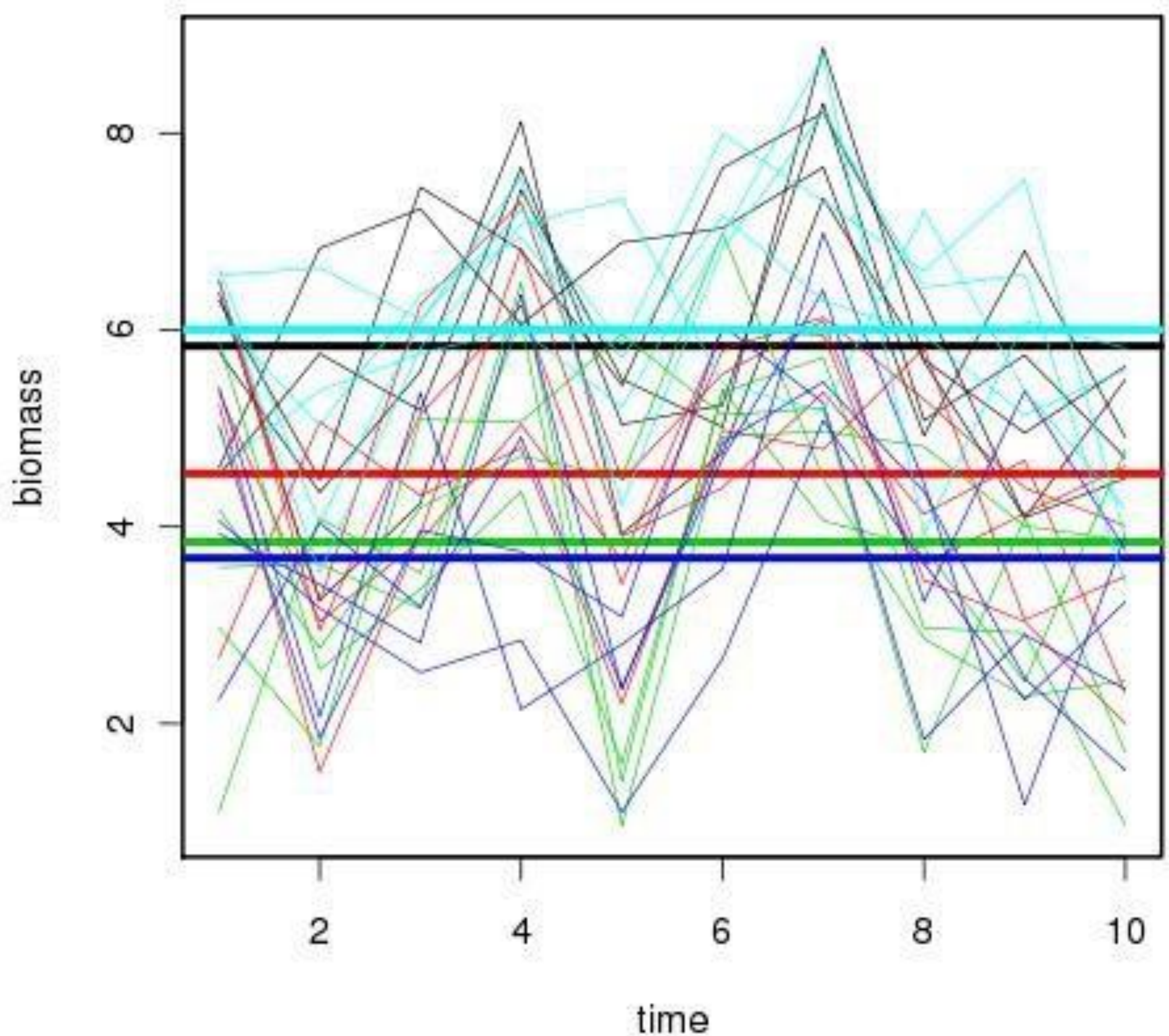
```
model{
  mu ~ dnorm(0,0.001)                        ## priors
  sigma ~ dgamma(0.001,0.001)
  for(t in 1:nt){alpha.t[t] ~ dnorm(0,tau.t)}
  tau.t ~ dgamma(0.001,0.001)        ## hyperprior

  for(t in 1:nt){
    Ex[t] <- mu + alpha.t[t]                  ## process model
    for(b in 1:nb){
      for(i in 1:nrep){
        x[t,b,i] ~ dnorm(Ex[t],sigma)   ## data model
      }
    }
  }
}
```



biomass
2
4
6
8
time
2
4
6
8
10

Model 3: Random Block Effect

```
model{
  mu ~ dnorm(0,0.001)          ## priors
  sigma ~ dgamma(0.001,0.001)
  tau.b ~ dgamma(0.001,0.001)
  for(b in 1:nb){ alpha.b[b] ~ dnorm(0,tau.b)}

  for(b in 1:nb){
    Ex[b] <- mu + alpha.b[b]
    for(t in 1:nt){
      for(i in 1:nrep){
        x[t,b,i] ~ dnorm(Ex[b],sigma)
      }
    }
  }
}
```



biomass
8
6
4
2
2
4
6
8
10
time

Model 4: Random Block & Time

```
model{

  mu ~ dnorm(0,0.001)                ## priors
  sigma ~ dgamma(0.001,0.001)
  tau.b ~ dgamma(0.001,0.001)
  tau.t ~ dgamma(0.001,0.001)
  for(t in 1:nt){alpha.t[t] ~ dnorm(0,tau.t) }
  for(b in 1:nb){alpha.b[b] ~ dnorm(0,tau.b) }

  for(t in 1:nt){
    for(b in 1:nb){
      Ex[t,b] <- mu + alpha.b[b] + alpha.t[t]
      for(i in 1:nrep){
        x[t,b,i] ~ dnorm(Ex[t,b],sigma)
      }
    }
  }
}
```

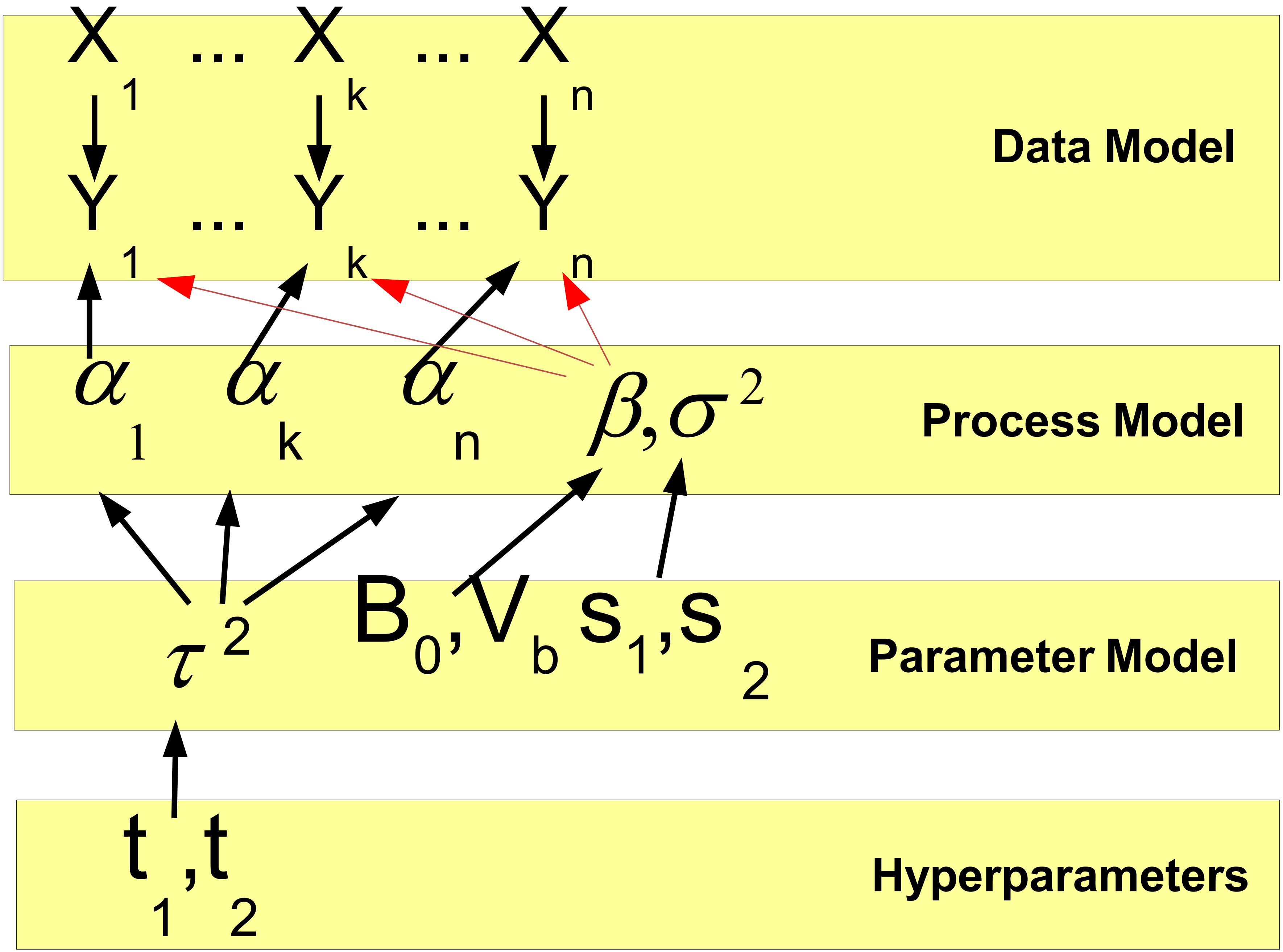
Summary Table

Model	mu	sigma	tau.t	tau.b	DIC
Global Mean	4.78 (0.11)	2.92 (0.27)			977.9
Random Time	4.75 (0.33)	2.23 (0.21)	0.97 (0.64)		919.8
Random Block	4.82 (0.69)	1.92 (0.18)		2.36 (3.62)	878.0
Random B x T	4.85 (0.75)	0.84 (0.08)	1.31 (0.67)	0.80 (0.60)	766.8

Mixed Model

Fixed Effects	Random Effect	Residual Error	
$\mu_{i,k} = X_i \beta + \alpha_k + \epsilon_{i,k}$			**Process model**
$\epsilon_{i,k} \sim N(0, \sigma^2)$			**Data model**
$\alpha_k \sim N(0, \tau^2)$			**Random effect**
$\sigma^2 \sim IG(s_1, s_2)$			**Error variance prior**
$\beta \sim N(B_0, V_\beta)$			**Fixed effects prior**
$\tau^2 \sim IG(t_1, t_2)$			**Random effects variance prior**

Mixed Model



X_1 ... X_k ... X_n
Y_1 ... Y_k ... Y_n
Data Model
α_1 α_k α_n β, σ^2
Process Model
τ^2 B_0, V_b s_1, s_2
Parameter Model
t_1, t_2
Hyperparameters

Explaining unexplained variance

- Random effects attempt to account for the unexplained variance associated with some group (plot, year, etc.) due to all the things that were not measured
- May point to scales that need additional explanation
- Adding covariates may explain some portion of this variance, but there's always something you didn't measure
- Sometimes additional fixed effects not justified (model selection)

Modeling Uncertainty

Overall take home message:

The proper accounting of uncertainty can be JUST AS IMPORTANT to making valid inference from your model as the process model and covariates

Random effects are used to account for the impacts of unmeasured/unmeasurable covariates