Introduction to Applied Bayesian Analysis in Wildlife Ecology

Jeffrey W. Doser

May 11, 2024

Hierarchical Bayesian models: Linear Mixed Models (LMMs)

Terminology

Hierarchical model = multi-level model

Terminology

Any statistical model that includes a random effect

Hierarchical model = multi-level model

Terminology

Any statistical model that includes a random effect

Hierarchical model = multi-level model

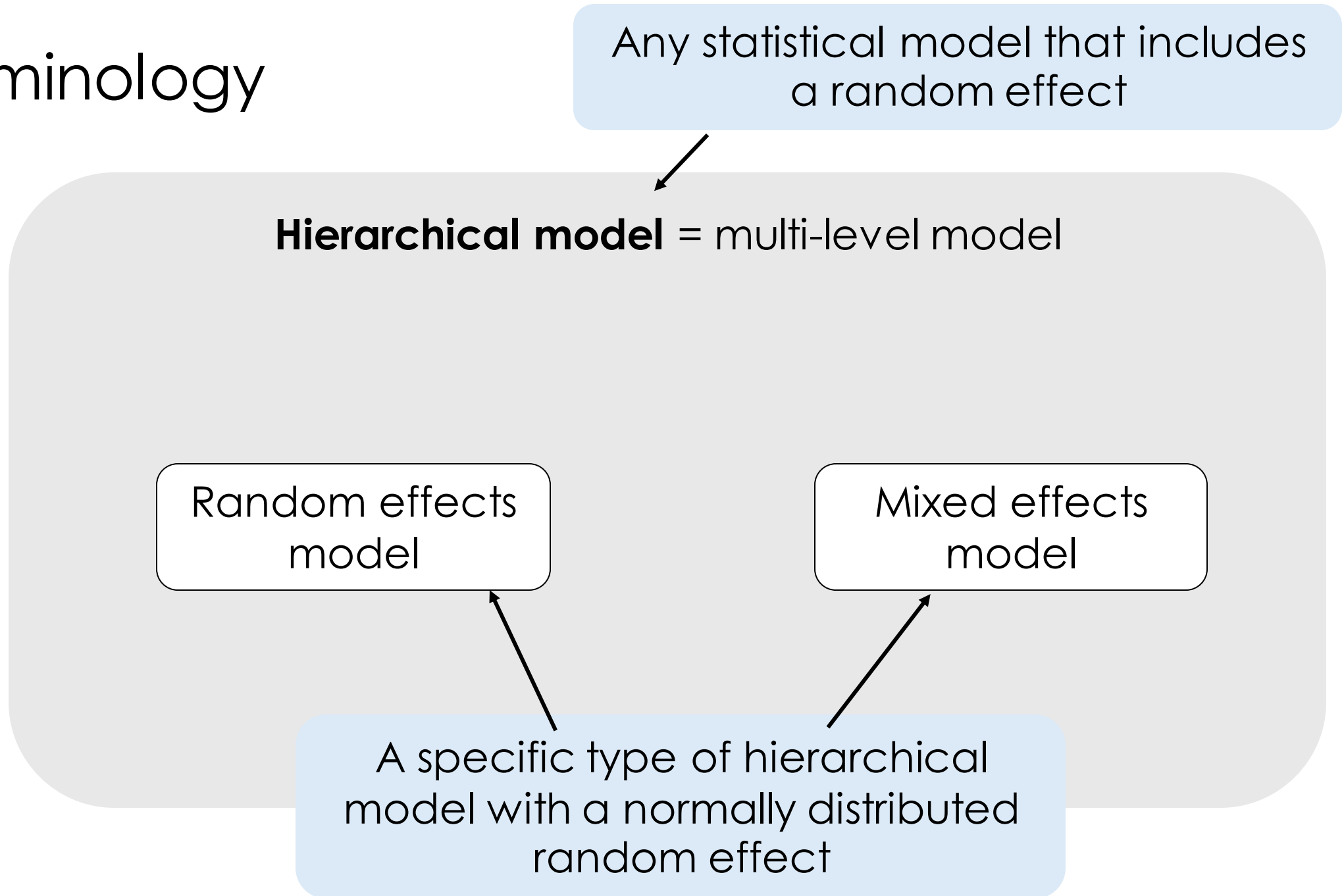Random effects model

Mixed effects model

Terminology

Any statistical model that includes a random effect

Hierarchical model = multi-level model

Random effects model

Mixed effects model

A specific type of hierarchical model with a normally distributed random effect

Some common hierarchical models in wildlife ecology

Occupancy models

Joint species distribution models

Integrated population models

N-mixture models

Hierarchical distance sampling models

Spatial linear mixed models

Time-series models

Some common hierarchical models in wildlife ecology

Occupancy models

Joint species distribution models

Integrated population models

These can all be viewed as extensions of GLMs with different types of random effects!!!

s

Hierarchical distance sampling models

Spatial linear mixed models

Time-series models

Recall the basic linear model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

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

$$i = 1, \ldots, N$$

Recall the basic linear model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$
$$\epsilon_i \sim \text{Normal}(0, \sigma^2)$$
$$i = 1, \dots, N$$

- Much of our focus has been on estimating β_0 and β_1
- These are often referred to as **Fixed Effects**

Random effects

- Factors whose levels are sampled from a larger population
    o Sample 30 species from a community
    o 50 forest plots from all forest plots in Michigan
    o 15 individuals within a population

Random effects

- Factors whose levels are sampled from a larger population
  - o Sample 30 species from a community
  - o 50 forest plots from all forest plots in Michigan
  - o 15 individuals within a population
- Interest lies more in variation among levels rather than the specific effects of each level
  - o Variation among species in a community
  - o Variation among forest plots in Michigan
  - o Variation among individuals in a population

Fixed vs. Random effects

Fixed vs. Random effects

Fixed Effects

Factors whose levels are experimentally determined or whose interest lies in the specific effects of each level

Fixed vs. Random effects

Fixed Effects

Factors whose levels are experimentally determined or whose interest lies in the specific effects of each level

Random Effects

Factors whose levels are sampled from a larger population, or whose interest lies in the variation among them rather than specific effects of each level

Fixed vs. Random effects

Fixed Effects

Factors whose levels are experimentally determined or whose interest lies in the specific effects of each level

Random Effects

Factors whose levels are sampled from a larger population, or whose interest lies in the variation among them rather than specific effects of each level

These definitions are debated! One could say all effects in a Bayesian model are random!!

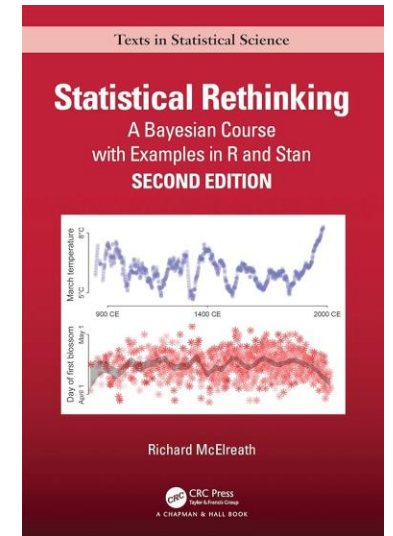# A better way of thinking about random effects

- A hierarchical model assumes the dataset being analyzed consists of a hierarchy of different populations/clusters/groups whose differences relate to that hierarchy
- Measurements come in clusters or groups

A better way of thinking about random effects

- A hierarchical model assumes the dataset being analyzed consists of a hierarchy of different populations/clusters/groups whose differences relate to that hierarchy
- Measurements come in clusters or groups

The term "hierarchical model" tells us more about the data structure than "mixed model" or "random effects model"

A better way of thinking about random effects and hierarchical models

A better way of thinking about random effects and hierarchical models
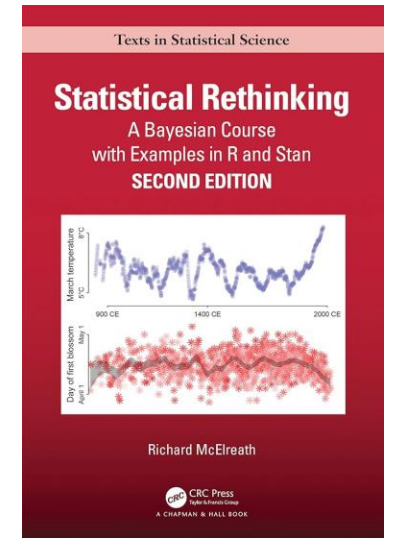
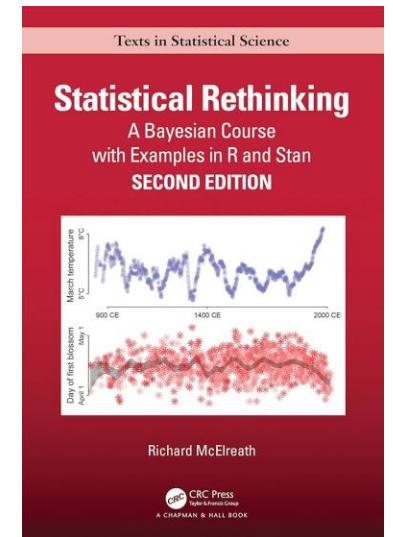- Fixed-effects models have amnesia

Adapted from:
https://github.com/oliviergimenez/Bayesian_Work

A better way of thinking about random effects and hierarchical models

- Fixed-effects models have amnesia
- Every new cluster (individual, species) is a new world
- No information passed among clusters when they are fixed effects

https://www.youtube.com/@rmcelreath/playlists

A better way of thinking about random effects and hierarchical models

- Fixed-effects models have amnesia
- Every new cluster (individual, species) is a new world
- No information passed among clusters when they are fixed effects
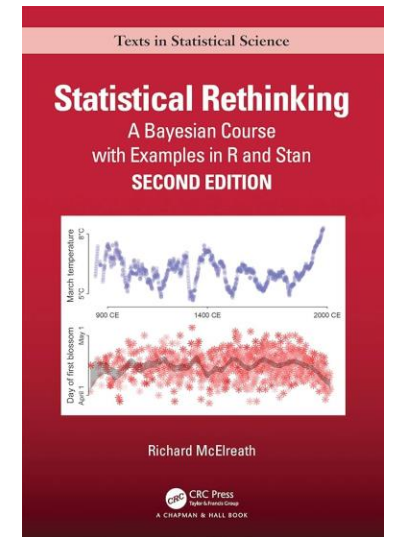- Hierarchical models remember and pool information. They have memory.

https://www.youtube.com/@rmcelreath/playlists

Adapted from:
https://github.com/oliviergimenez/Bayesian_Work

A better way of thinking about random effects and hierarchical models

- Fixed-effects models have amnesia
- Every new cluster (individual, species) is a new world
- No information passed among clusters when they are fixed effects
- Hierarchical models remember and pool information. They have memory.
- Properties of clusters come from a population

Adapted from:
https://github.com/oliviergimenez/Bayesian_Work

A better way of thinking about random effects and hierarchical models

- Fixed-effects models have amnesia
- Every new cluster (individual, species) is a new world
- No information passed among clusters when they are fixed effects
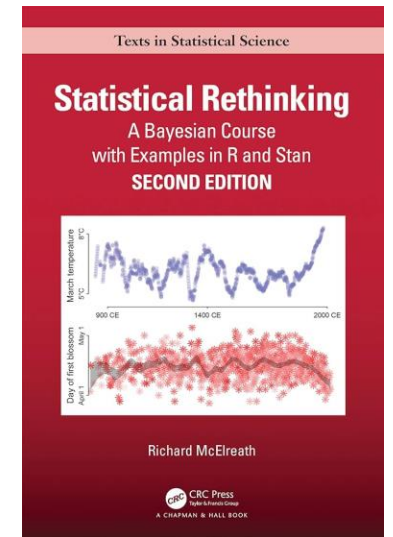- Hierarchical models remember and pool information. They have memory.
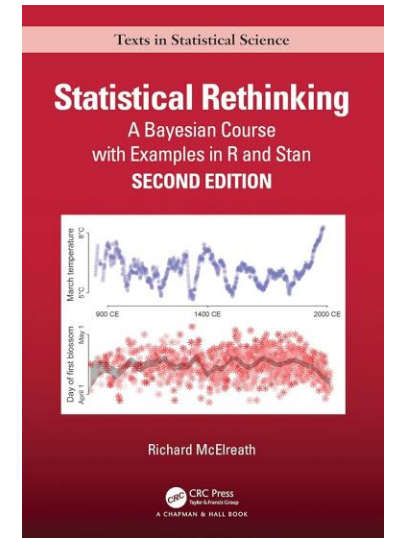- Properties of clusters come from a population
- If other clusters improve your guess about a new cluster, you want to use pooling.

https://www.youtube.com/@rmcelreath/playlists

Adapted from:
https://github.com/oliviergimenez/Bayesian_Work

Example: wing length in little owls

- We measured wing length in 20 individuals in 5 different populations of little owls (total of 100 owls)
- Is there a difference in wing length among populations?

Example: wing length in little owls

- We measured wing length in 20 individuals in 5 different populations of little owls (total of 100 owls)
- Is there a difference in wing length among populations?

$$y_{i,j} = \beta_j + \epsilon_{i,j}$$
$$\epsilon_{i,j} \sim \text{Normal}(0, \sigma^2)$$

- y_{ij} = observed wing length of owl i in population j
- β_j = mean wing length of an owl in population j (mean parameterization)
- ϵ_{ij} = random wing deviation of owl i in population j from its population mean (residual)

Scenario 1: Background

- You are a wildlife manager in charge of these 5 specific populations
- You want to know the mean wing length for each population, and the differences between them. You only care about these 5 populations.
- You specifically chose these 5 populations. Not a random sample of populations from some broader area.
- Here it makes sense to fit a fixed effects model

$$y_{i,j} \sim \mathrm{Normal}(\beta_j, \sigma^2)$$
$$\beta_j \sim \mathrm{Normal}(0, 1000)$$
$$\sigma \sim \mathrm{Student\text{-}t}(\mathrm{df}, 0, \mathrm{scale})$$

Scenario 1: Fixed effects model

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$

$$\beta_j \sim \text{Normal}(0, 1000)$$

$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Scenario 1: Fixed effects model

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$
$$\beta_j \sim \text{Normal}(0, 1000)$$
$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Each β_j is given a separate prior. No information is passed from one to the other (amnesia).

Scenario 1: Fixed effects model

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$
$$\beta_j \sim \text{Normal}(0, 1000)$$
$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Each β_j is given a separate prior. No information is passed from one to the other (amnesia).

Nothing new here. We've done this before.

Example: little
owl wing length
(fixed effects)

5a-lmm-brms.R

Scenario 2: Background

- You randomly sample 20 individuals from 5 random populations across the owl's range.
- You don't necessarily care about these 5 populations, but rather you want to know if there is substantial variability in wing length across populations.
- These 5 populations are representative of populations across the little owl's range.
- Here it makes sense to fit a random effects model (your first hierarchical Bayesian model!)

Scenario 2: Hierarchical Bayesian model

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$
$$\beta_j \sim \text{Normal}(\mu, \tau^2)$$
$$\mu \sim \text{Normal}(0, 1000)$$
$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$
$$\tau \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Scenario 2: Hierarchical Bayesian model

All the β_j effects now come from a shared distribution with parameters that we estimate.

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$
$$\beta_j \sim \text{Normal}(\mu, \tau^2)$$
$$\mu \sim \text{Normal}(0, 1000)$$
$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$
$$\tau \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Scenario 2: Hierarchical Bayesian model

All the β_j effects now come from a shared distribution with parameters that we estimate.

Average wing-length across populations

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$

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

$$\mu \sim \text{Normal}(0, 1000)$$

$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

$$\tau \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Scenario 2: Hierarchical Bayesian model

All the β_j effects now come from a shared distribution with parameters that we estimate.

Average wing-length across populations

Standard deviation of wing-length across populations

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$

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

$$\mu \sim \text{Normal}(0, 1000)$$

$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

$$\tau \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Scenario 2: Hierarchical Bayesian model

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$

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

$$\mu \sim \text{Normal}(0, 1000)$$

$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

$$\tau \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Hyperparameters: parameters of a distribution for a parameter.

Scenario 2: Hierarchical Bayesian model

$$y_{i,j} \sim \text{Normal}(\beta_j, \sigma^2)$$
$$\beta_j \sim \text{Normal}(\boxed{\mu, \tau^2})$$
$$\mu \sim \text{Normal}(0, 1000)$$
$$\sigma \sim \text{Student-t}(\text{df}, 0, \text{scale})$$
$$\tau \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

Hyperparameters: parameters of a distribution for a parameter.

An alternative definition of a hierarchical model is any model with hyperparameters

Example: little
owl wing length
(random effects)

5a-lmm-brms.R

Fixed vs. Random: which to choose?

Fixed vs. Random: which to choose?

Fixed

- You have particular interest in the measured factor levels
- Minimal interest in variance among levels
- Minimal interest in generalizing results to factor levels that you did not study

Fixed vs. Random: which to choose?

Fixed

- You have particular interest in the measured factor levels
- Minimal interest in variance among levels
- Minimal interest in generalizing results to factor levels that you did not study

Random

- You don't have particular interest in the measured factor levels, or you could not sample all levels
- Interested in the variation among levels (but still may be interested in the effect for observed levels)
- You want to generalize results to a larger population (levels are a sample from some population)

Reasons to use random effects/HBMs

- Extrapolation to a wider population of inference

Reasons to use random effects/HBMs

- Extrapolation to a wider population of inference
- Improved accounting for system uncertainty

Reasons to use random effects/HBMs

- Extrapolation to a wider population of inference
- Improved accounting for system uncertainty
- Account for non-independence among data points (pseudoreplication)

Reasons to use random effects/HBMs

- Extrapolation to a wider population of inference
- Improved accounting for system uncertainty
- Account for non-independence among data points (pseudoreplication)
- Efficiency of estimation -> "shrinkage"
  - Parameters are no longer independent and will be pulled towards the mean
  - Also called "borrowing strength"

Reasons you might not want to use random effects/HBMs

- Models take longer to run (run time increases as number of levels increases)

Reasons you might not want to use random effects/HBMs

- Models take longer to run (run time increases as number of levels increases)
- You have exhaustively sampled all possible factor levels

Reasons you might not want to use random effects/HBMs

- Models take longer to run (run time increases as number of levels increases)
- You have exhaustively sampled all possible factor levels
- You need an "adequately large" number of factor levels (populations in our little owl example)
  - There is no clear minimum number that you need, it will depend on the model and data characteristics
  - Bolker recommends at least 5-6 levels

Linear mixed models (another type of HBM)

Linear mixed models (another type of HBM)

- A "mixed" model is one that has both fixed and random effects

Linear mixed models (another type of HBM)

- A "mixed" model is one that has both fixed and random effects
- I would guess that at least 33% of papers in wildlife ecology journals use some sort of mixed model
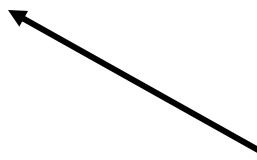
Linear mixed models (another type of HBM)

- A "mixed" model is one that has both fixed and random effects
- I would guess that at least 33% of papers in wildlife ecology journals use some sort of mixed model
- Particularly useful for:
  - Complex experimental designs (split plot, RCBD)
  - Spatial autocorrelation
  - Temporal autocorrelation
  - Missing data

Linear mixed models (another type of HBM)

- A "mixed" model is one that has both fixed and random effects
- I would guess that at least 33% of papers in wildlife ecology journals use some sort of mixed model
- Particularly useful for:
  - Complex experimental designs (split plot, RCBD)
  - Spatial autocorrelation
  - Temporal autocorrelation
  - Missing data

We can fit all of these in `brms` (although some are better than others)

Hierarchical modeling is where Bayesian analysis thrives

- Bayesian approaches are much more flexible in dealing with random effects compared to frequentist approaches

Hierarchical modeling is where Bayesian analysis thrives

- Bayesian approaches are much more flexible in dealing with random effects compared to frequentist approaches
- **Remember: the benefits of going Bayesian increase as the complexity of the model you want to fit increases**

Hierarchical Bayesian models: Generalized Linear Mixed Models (GLMMs)

Generalized Linear Mixed Models (GLMMs)

- Extension of LMMs to different distributions
- Four ingredients to a GLMM

GLMM Recipe

Statistical Distribution

Describes random variation in the response y. This is the stochastic (random) part of the model.

GLMM Recipe

Statistical Distribution

Describes random variation in the response y. This is the stochastic (random) part of the model.

Linear Predictor

Relates the response variable to the predictor variables (parameters are linear)

GLMM Recipe

Statistical Distribution

Describes random variation in the response y. This is the stochastic (random) part of the model.

Linear Predictor

Relates the response variable to the predictor variables (parameters are linear)

Link Function

Applied to the expectation of the response variable. Connects the mean to its linear predictor.

GLMM Recipe

Statistical Distribution

Describes random variation in the response y. This is the stochastic (random) part of the model.

Linear Predictor

Relates the response variable to the predictor variables (parameters are linear)

Link Function

Applied to the expectation of the response variable. Connects the mean to its linear predictor.
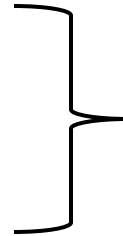
Specifying fixed vs. random effects

GLMMs are a big reason why I'm a Bayesian

GLMMs are a big reason why I'm a Bayesian

- Fitting GLMMs in a frequentist paradigm is tricky. Lots of algorithms (amusement parks)
  - Penalized quasilikelihood
  - Laplace approximation
  - Gauss-Hermite quadrature

Different pros and cons. Best method depends on objectives.

GLMMs are a big reason why I'm a Bayesian

- Fitting GLMMs in a frequentist paradigm is tricky. Lots of algorithms (amusement parks)
  - Penalized quasilikelihood
  - Laplace approximation
  - Gauss-Hermite quadrature
- **Bayesian GLMMs are more flexible, easier to fit, and supported more by statistical theory**

Different pros and cons. Best method depends on objectives.

Trends in Ecology & Evolution — Supports open access

This journal | Journals | Publish | News & events | About

REVIEW | VOLUME 24, ISSUE 3, P127-135, MARCH 2009 | Download Full Issue

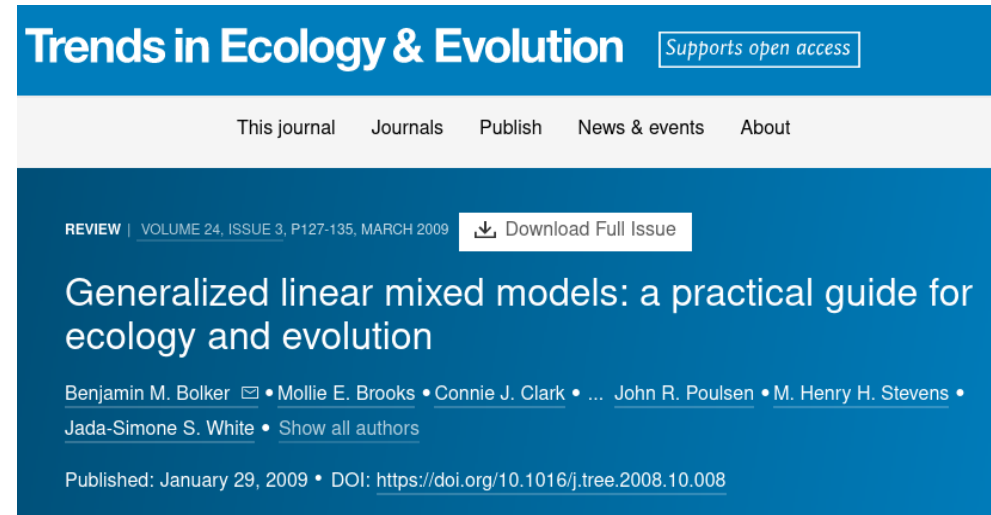Generalized linear mixed models: a practical guide for ecology and evolution

Benjamin M. Bolker • Mollie E. Brooks • Connie J. Clark • ... John R. Poulsen • M. Henry H. Stevens • Jada-Simone S. White • Show all authors

Published: January 29, 2009 • DOI: https://doi.org/10.1016/j.tree.2008.10.008

Exercise: Amphibian species richness

- We want to model amphibian species richness across some area of interest
- Sample richness at 15 wetlands in each of 10 protected parks (150 measurements)

Exercise: Amphibian species richness
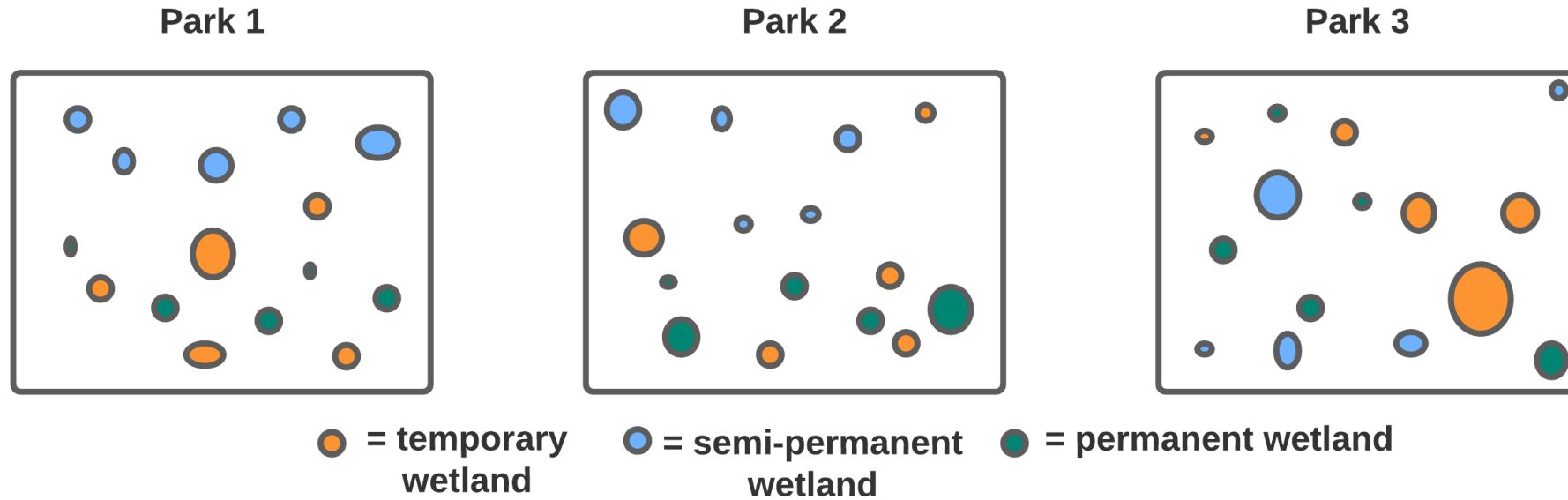
- We want to model amphibian species richness across some area of interest
- Sample richness at 15 wetlands in each of 10 protected parks (150 measurements)
- Two variables we hypothesize will influence richness:
  - Area of the wetland: continuous
  - Hydroperiod (length of time a wetland retains water): categorical with three levels (permanent, semi-permanent, and temporary)
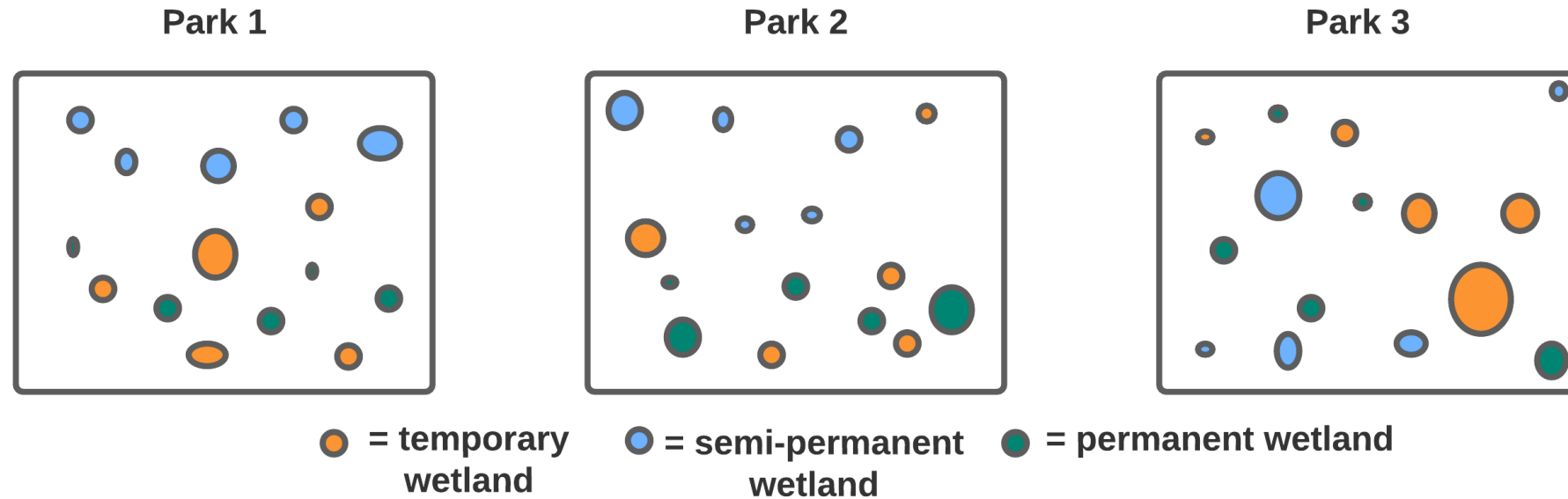
Exercise: Amphibian species richness

- We want to model amphibian species richness across some area of interest
- Sample richness at 15 wetlands in each of 10 protected parks (150 measurements)
- Two variables we hypothesize will influence richness:
  - Area of the wetland: continuous
  - Hydroperiod (length of time a wetland retains water): categorical with three levels (permanent, semi-permanent, and temporary)

What distribution(s) could work for this data set?
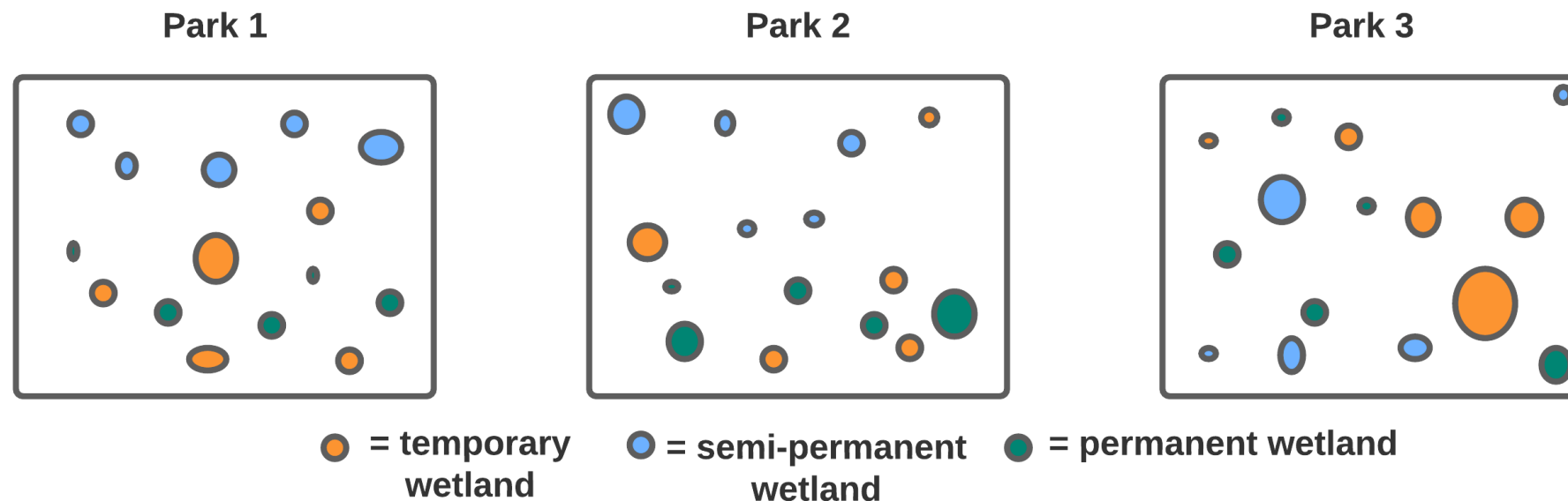
Exercise: Amphibian species richness

Park 1 **Park 2** **Park 3**

● = temporary wetland ● = semi-permanent wetland ● = permanent wetland

Exercise: Amphibian species richness

Park 1   **Park 2**   **Park 3**
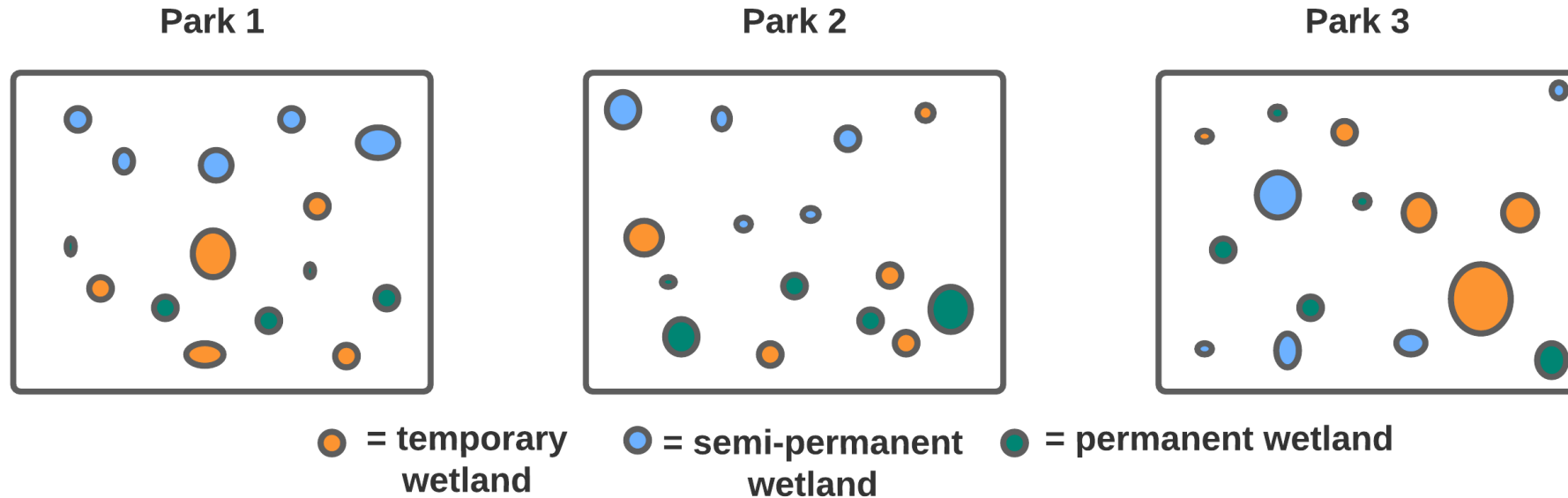
● = temporary wetland   ● = semi-permanent wetland   ● = permanent wetland

- In a basic GLM, we assume each data point is independent from each other

Exercise: Amphibian species richness

Park 1	**Park 2**	**Park 3**

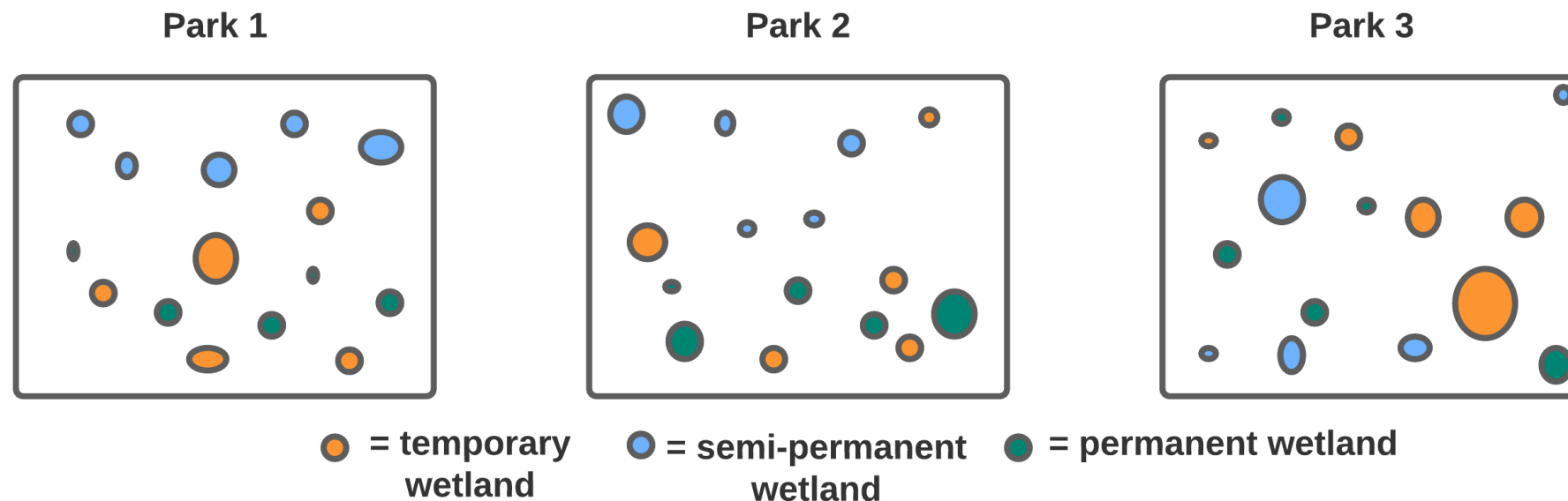● = temporary wetland   ● = semi-permanent wetland   ● = permanent wetland

- In a basic GLM, we assume each data point is independent from each other
- Here, we have sub-samples (wetlands) within each park

Exercise: Amphibian species richness

Park 1

Park 2

Park 3

● = temporary wetland   ● = semi-permanent wetland   ● = permanent wetland

- In a basic GLM, we assume each data point is independent from each other
- Here, we have sub-samples (wetlands) within each park
- Sub-samples are not independent of each other, so we should include a random effect of park

Exercise: Amphibian species richness

Park 1

Park 2

Park 3

● = temporary wetland   ● = semi-permanent wetland   ● = permanent wetland

- In a basic GLM, we assume each data point is independent from each other
- Here, we have sub-samples (wetlands) within each park
- Sub-samples are not independent of each other, so we should include a random effect of park
- Random effect of park tells us that samples within a park are likely more related to each other than samples from two different parks

Exercise: Amphibian species richness

$$\text{richness}_{i,j} \sim \text{Poisson}(\lambda_{i,j})$$

$$\log(\lambda_{i,j}) = \alpha_j + \beta_1 \cdot \text{SEMI}_{i,j} + \beta_2 \cdot \text{TEMP}_{i,j} + \beta_3 \cdot \text{AREA}_{i,j}$$

$$\alpha_j \sim \text{Normal}(\beta_0, \tau^2)$$

$$\beta_0 \sim \text{Normal}(0, 1000)$$

$$\beta_1 \sim \text{Normal}(0, 1000)$$

$$\beta_2 \sim \text{Normal}(0, 1000)$$

$$\beta_3 \sim \text{Normal}(0, 1000)$$

$$\tau \sim \text{Student-t}(\text{df}, 0, \text{scale})$$

i = wetland, j = park

Exercise: Amphibian species richness

5c-glmm-brms.R