ESS 575: Multi-Level Regression Lab

Team England

18 October, 2022

Contents

Team England:

- Caroline Blommel
- Carolyn Coyle
- Bryn Crosby
- George Woolsey

cblommel@mail.colostate.edu, carolynm@mail.colostate.edu, brcrosby@rams.colostate.edu, george.woolsey@colostate.edu

Motivation

Each section of this lab has two parts – a model *building* exercise and a model *coding* exercise. The material covered here is important and broadly useful – building multi-levels models is a true workhorse for understanding ecological processes because so many problems contain information at nested spatial scales, levels of organization, or categories. It will be worthwhile to dig in deeply to understand it. The big picture is to demonstrate the flexibility that you gain as a modeler by understanding basic principles of Bayesian analysis. To accomplish that, these exercises will reinforce the following:

1. Diagramming and writing hierarchical models
2. Using data to model parameters
3. JAGS coding
4. Creating index variables, a critically important and useful skill
5. Posterior predictive checks

Introduction

Ecological data are often collected at multiple scales or levels of organization in nested designs. Group is a catchall term for the upper level in many different types of nested hierarchies. Groups could logically be composed of populations, locations, species, treatments, life stages, and individual studies, or really, any sensible category. We have measurements within groups on individual organisms, plots, species, time

periods, and so on. We may also have measurements on the groups themselves, that is, covariates that apply at the upper level of organization or spatial scale or the category that contains the measurements. Multilevel models represent the way that a quantity of interest responds to the combined influence of observations taken at the group level and within the group.

Nitrous oxide N_2O, a greenhouse gas roughly 300 times more potent than carbon dioxide in forcing atmospheric warming, is emitted when synthetic nitrogenous fertilizers are added to soils. Qian and colleagues (2010) conducted a Bayesian meta-analysis of N_2O emissions (g N $\cdot$ ha^{-1} $\cdot$ d^{-1}) from agricultural soils using data from a study conducted by Carey (2007), who reviewed 164 relevant studies. Studies occurred at different locations, forming a group-level hierarchy (we will use only sites that have both nitrogen and carbon data, which reduces the number of sites to 107 in the analysis here). Soil carbon content (g $\cdot$ organic C $\cdot$ g^{-1} soil dry matter) was measured as a group-level covariate and is assumed to be measured without error. Observations of N_2O emission are also assumed to be measured without error and were paired with measurements of fertilizer addition (kg N$\cdot$ ha^{-1} $\cdot$ year^{-1}). The effect of different types of fertilizer was also studied.

You are going to use these data to build increasingly complex models of N_2O emission. The initial models will ignore some important covariates as well as how the data are structured hierarchically into sites. This is ok! When writing for a multi-level model like this one, do it incrementally, starting with a separate model for each site (the no-pool model) or a model that ignores sites entirely (the pooled model). After getting these models to work you can add complexity by drawing the intercept for each model from a distribution, before pursuing further refinements. We **strongly sugggest** this approach because it is always best to do the simple thing first: there is less to go wrong. Also, when things do go wrong it will be clearer as to what is causing the problem.

```
### hey
```