A Review of “Synthesizing ecological experiments and observational data with hierarchial Bayes” (Clark, LaDeau)

Emily Austin Michael Saum

Department of Ecology and Evolutionary Biology
The University of Tennessee, Knoxville

March 15, 2010

Outline

Goal

The main objective of this chapter is to illustrate how “Hierarchial Bayes” methodology can be utilized to address modeling issues which arise in many ecological field experiments.

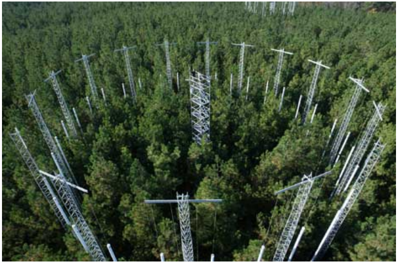
Common problems which exist in many ecological field experiments include:

- Low replication rates.
- Identification of key variables and inability to measure.
- Limited control of key variables.
- Uncertainty in data measurements.
- Model may need to change after experiment begins.

FACE - Free Air CO_2 Experiment

"Simple" field experiment to determine the effects of elevated CO_2 levels on cone production.

- Loblolly pine trees in monoculture.
- CO_2 treatments established in 1996.
- Ambient CO_2 (365 ppm).
- Elevated CO_2 (565 ppm).
- 1998 is the first "treatment year".
- Seed rain data used to predict fecundity.



FACE - Free Air CO_2 Experiment

- 3 ambient and three elevated “rings”.
- Each ring 30 m in diameter.
- Each ring contains 75-170 saplings that were 14 years old.
- Tree diameters range from 5-25 cm in 1996.

FACE Summary

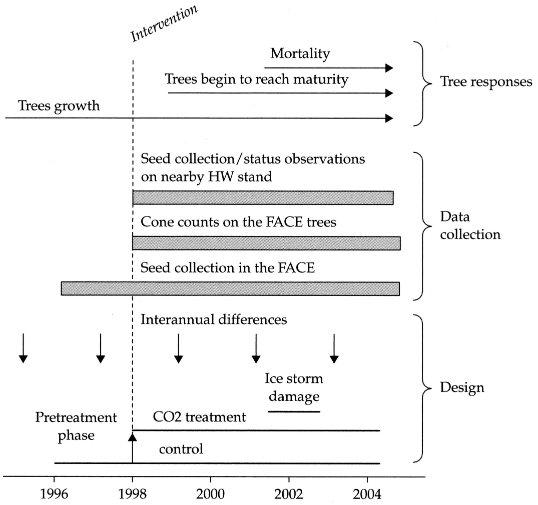


Figure 3.1. Summary of the variables, data, and design for the experimental example. The intervention is indicated in 1998 to reflect the first year of reproduction that developed under the elevated CO_2 treatment. Cones have a 2-year development time.

A Standard Model

Expectations were that a model based on seed production and dispersal would work.

$$s_{jt} \sim \text{Pois}(A_j g_j(\boldsymbol{y}_t)), \quad j = 1, \dots, m, t = t_j, \dots, T_j$$

where

- s_{jt} is the number of seeds recovered from seed trap j at time t
- A_j is the area of seed trap j
- $g_j(\boldsymbol{y}_t)$ is the density of seed arriving at location j, producing seeds $\boldsymbol{y}_t = \{y_{it}, i = 1, \dots, n\}$.
- The model $g_j(\boldsymbol{y}_t)$ can include effects of tree sizes and dispersal distances, CO_2 levels, etc.

Counting Cones

- Scientists realized they could climb a tower at the center of each ring and count cones on individual trees using binoculars.
- *What Fun!*
- Changed to a model on the individual tree scale, with the idea that seed production could be related allometrically to tree diameter:

$$y_{it} = a_0 d_{it}^{a_1}.$$

Linear Model

Redefining $Y_{it} = \log(y_{it})$, $D_{it} = \log(d_{it})$, $\alpha_0 = \log(a_0)$, $\alpha_1 = a_1$ gives

$$Y_{it} = \alpha_0 + \alpha_1 D_{it} + \alpha_2 C_{it} + \varepsilon_{it}$$

or

$$Y_{it} = \boldsymbol{x}_{it}\boldsymbol{a} + \varepsilon_{it}, \quad i = 1, \ldots, n, t = t_i, \ldots, T_i$$

where $\boldsymbol{x}_{it} = [1, D_{it}, C_{it}]$ is the vector of fixed effects and $\boldsymbol{a} = [\alpha_0, \alpha_1, \alpha_2]^{\mathrm{T}}$ is the coefficient vector. Note that C_{it} are scaled log CO_2 concentrations and ε_{it} are Gaussian error terms, i.e.,

$$\varepsilon_{it} \sim \mathcal{N}(0, \sigma^2).$$

Linear Model, contd.

Letting $S_i = T_i - t_i$, $\boldsymbol{Y}_i = [Y_{t_i} \cdots Y_{T_i}]^{\mathsf{T}}$, and

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & D_{it_i} & C_{it_i} \\ \vdots & \vdots & \vdots \\ 1 & D_{iT_i} & C_{iT_i} \end{bmatrix}$$

one obtains the liklihood for the data given by

$$\prod_{i=1}^{n} \mathcal{N}_{S_i}\left(\boldsymbol{Y}_i | \boldsymbol{X}_i \boldsymbol{a}, \sigma^2 \boldsymbol{I}_{S_i}\right)$$

and the joint posterior for parameters conditioned on the data is

$$p(\boldsymbol{a}, \sigma^2 | \boldsymbol{X}, \boldsymbol{Y}, \ldots) \propto \prod_{i=1}^{n} \mathcal{N}_{S_i}\left(\boldsymbol{Y}_i | \boldsymbol{X}_i \boldsymbol{a}, \sigma^2 \boldsymbol{I}_{S_i}\right) \mathcal{N}_3\left(\boldsymbol{a} | \boldsymbol{a}_\alpha, \boldsymbol{V}_\alpha\right) \mathrm{IG}\left(\sigma^2 | a_\alpha, b_\alpha\right).$$

Available FACE Data

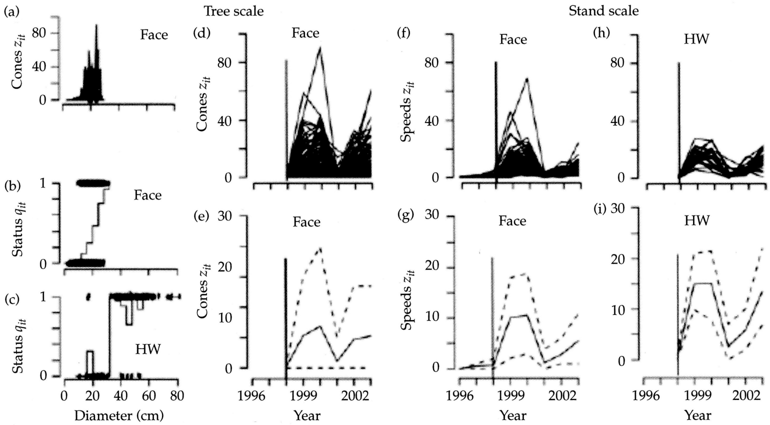


Figure 3.2. Summary of data available for estimating fecundity in Duke Forest. Tree scale data include counts of cones (FACE trees in (a) and by year in (d, e)) and statuses (plotted by diameter in (b, c)). The jittered status observations are summarized by histograms in (b) and (c). Stand scale data are seeds collected in seed traps, beginning in 1996 for FACE trees (f, g) and in 1998 for HW trees (h, i). The vertical lines on time plots indicate first season of reproductive effects, in 1998. Other lines in (e), (g), and (i) are 10th, 50th, and 90th percentiles.

Complications

The above model posterior could be sampled with a Gibbs sampler and is described in Box 3.1, page 46 in the text.

Model Assumptions

- log fecundity is a linear function of log diameter and CO_2 levels,
- all individuals respond independently and identically,
- residual variance σ^2 is everywhere and always the same,

Model Design Changes

Would like to also incorporate known information into the model, including

- Reproductive maturity was reached only after treatment began.
- Ice storm damage.
- Threshold of tree maturity.
- Nearby loblolly pines in Hardwood (HW) stands have something to offer:
  - Cone production synchronized to these trees.
  - Trees are much larger, providing data on diameters not present at FACE.

Hierarchial Bayes

Bayesian Model

Given liklihood $\Pr(X|\theta)$ and prior $\Pr(\theta)$, one can determine the posterior $\Pr(\theta|X)$ as

$$\Pr(\theta|X) \propto \Pr(X|\theta)\Pr(\theta).$$

Hierarchial Bayes (HB) Model

An extension of the basic Bayesian Model,

$$\Pr(\theta, \psi|X) \propto \Pr(X|\theta)\Pr(\theta|\psi)\Pr(\psi),$$

and one can continue this process

$$\Pr(\theta, \psi, \phi|X) \propto \Pr(X|\theta)\Pr(\theta|\psi)\Pr(\psi|\phi)\Pr(\phi).$$

HB, contd.

- Eventually the process must terminate with priors which do not depend on any unknown parameters.
- Allows creation of a complex model by constructing a joint distribution which is the product of *simpler* conditional relationships.
- Given **Data** linked to **Processes**, not directly observed, one is able to construct a mathematical model which allows one to estimate **Parameters**.
- $$[data, process, parameters] = [data|process, parameters] \times [process|parameters] \times [parameters].$$

The Full Model

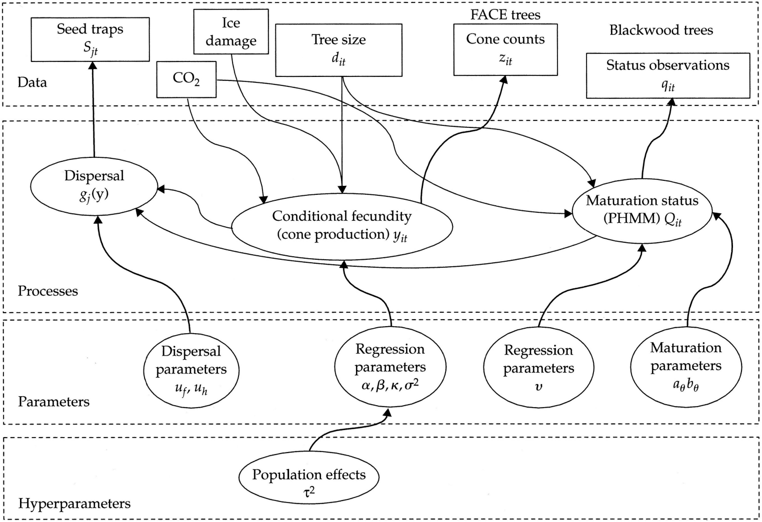


Figure 3.4. Graph of the full model, showing observables in the upper **Data** stage, state variables in the **Process** stage, **Parameters**, and **Hyperparameters**. Heavy arrows have stochastic components, thin arrows are deterministic. For clarity, prior parameters are not shown.

Extending the Linear Model

- Allow for differences among individuals and non-independence by introducing random effects $\beta_i \sim \mathcal{N}(0, \tau^2)$ in a *random intercept model*.

$$\begin{aligned} Y_{it} &= \boldsymbol{x}_{it}\boldsymbol{a} + \beta_i + \varepsilon_{it} \\ &= (\alpha_0 + \beta_i) + \alpha_1 D_{it} + \alpha_2 C_{it} + \varepsilon_{it}. \end{aligned}$$

- Allow for the fact that assignment of ambient and elevated treatments may not have resulted in samples of trees with identical fecundity potential by introducing an indicator variable (α_1) into the design matrix. Letting t^* indicate the year treatment begins, expected responses are

$$\begin{aligned} \mathbb{E}[Y_{it}|\beta_i] &= (\alpha_0 + \beta_i) + \alpha_2 D_{it} \quad \text{Ambient} \\ \mathbb{E}[Y_{it}|\beta_i, t \leq t^*] &= (\alpha_0 + \alpha_1 + \beta_i) + \alpha_2 D_{it} \quad \text{Pre-treatment} \\ \mathbb{E}[Y_{it}|\beta_i, t > t^*] &= (\alpha_0 + \alpha_1 + \beta_i) + \alpha_2 D_{it} + \alpha_3 C_{it} \quad \text{Post-treatment.} \end{aligned}$$

Extending the Linear Model, contd.

- Trees synchronized in terms of seed production implies that year effects may play a role. Allow that some part of the variation is shared among individuals, which changes year to year. Letting κ_t represent these fixed year effects gives

$$Y_{it} = \boldsymbol{x}_{it}\boldsymbol{a} + \beta_i + \kappa_t + \varepsilon_{it}$$

- Allow for changes over time which affect only certain individuals, such as ice storm damage, by introducing another indicator variable into the design matrix (α_4) gives

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{a} + \mathbf{1}_{S_i}\beta_i + \boldsymbol{\kappa}_i + \boldsymbol{\epsilon}_i.$$

Extending the Linear Model, contd.

Posterior

$$p(\boldsymbol{a}, \sigma^2 | \boldsymbol{X}, \boldsymbol{Y}, \dots) \propto \prod_{i=1}^{n} \mathcal{N}_{S_i}\left(\boldsymbol{Y}_i | \boldsymbol{X}_i \boldsymbol{a}, \sigma^2 \boldsymbol{I}_{S_i}\right) \mathcal{N}_5\left(\boldsymbol{a} | \boldsymbol{a}_\alpha, \boldsymbol{V}_\alpha\right)$$
$$\prod_{t=1}^{T} \mathcal{N}\left(\kappa_t | a_{\kappa_t}, v_\kappa\right) \prod_{i=1}^{n} \mathcal{N}\left(\beta_i | 0, \tau^2\right)$$
$$\mathrm{IG}\left(\sigma^2 | a_\alpha, b_\alpha\right) \mathrm{IG}\left(\tau^2 | a_\tau, b_\tau\right).$$

Scale of Observations

- Tree level and trap level data obtained at different times.
- Tree status is partially known (mature trees produce cones).
- Condition fecundity model on tree status, regression only applies to mature trees.
- Tree state $Q_{it} = 0$ for immature trees, $Q_{it} = 1$ for mature trees.
- Tree changes status according to gamma cumulative distribution with probability

$$\theta_{it} = \text{CGamma}(d_{it}; a_\theta, b_\theta), \quad i = 1, \ldots, n, t = t_i, \ldots, T_i.$$

- Allow that ambient and elevated trees may differ by using $a_\theta^{(amb)}$ and $a_\theta^{(ele)}$.

Data Models

There are three types of observable data:

- **Seed traps**: sampling distribution is Poisson, with two dimensional Student's t dispersion kernel.
- **Cone counts**: sampling distribution also Poisson, but is conditioned on Q_{it}.
- **Tree status**: conditioned on the true but uncertain state Q_{it} with recognition probability v. This recognition probability q_{it} is Bernoulli distributed when $Q_{it} = 1$, zero otherwise.
- Because status is unknown before the first cones are observed, model the conditional probability that an unknown tree is in the mature state is

$$\Pr(Q_{it} = 1 | q_{it} = 0) = \frac{\theta_{it}(1 - v)}{1 - v\theta_{it}}.$$

Parameter Models

- Prior densities and hyperprior (τ^2) for random effects.
- For example, mean $\boldsymbol{a}_\alpha = [2, 0, 0.7, 0.5, -1]^\mathsf{T}$, $\boldsymbol{V}_\alpha = \text{diag}(100, 100, 0.1, 1, 10)$.
- These choices reflect prior information and associated weights one assigns to that prior information.
- The choice of $\alpha_2 = 0.7$ with variance 0.1 reflects the fact that the while the diameter effect is representative of a large number of species over a large number of years, the small prior variance has insignificant weight to have much influence. Thus, this is a relatively *noninformative prior*.
- Other prior information is described fully in Section 3.5.3 (pp. 52–53).
- Gibbs sampler for full model is described fully in Box 3.2 (pp. 54–55).

Full Model Posterior

$$\prod_{i=1}^{n_f}\prod_{t=t_i}^{T_i}\text{Pois}(z_{it}|\gamma y_{it}) \qquad \text{(cones)}$$

$$\prod_{j=1}^{m}\prod_{t=t_j}^{T_j}\text{Pois}(s_{jt}|A_j g_j(\boldsymbol{y}_t)) \qquad \text{(traps)}$$

$$\text{Bin}(q_f|Q_f, v_f)\text{Bin}(q_h|Q_h, v_h) \qquad \text{(status observations)}$$

$$\prod_{i=1}^{n_h}\prod_{t=t_i}^{T_i}\text{Bernoulli}(Q_{it}|\theta_{it}) \qquad \text{(hidden Markov)}$$

$$\prod_{i=1}^{n}\mathcal{N}_{S_i}\left(\boldsymbol{Y}_i|\boldsymbol{X}_i\boldsymbol{a}+\mathbf{1}_{S_i}\beta_i+\boldsymbol{\kappa}_i, \sigma^2\boldsymbol{I}_{S_i}\right) \qquad \text{(Condtional fecundity)}$$

$$\mathcal{N}_5\left(\boldsymbol{a}|\boldsymbol{a}_\alpha, \boldsymbol{V}_\alpha\right)\prod_{t=1}^{T}\mathcal{N}\left(\kappa_t|a_{\kappa_t}, v_\kappa\right)\prod_{i=1}^{n}\mathcal{N}\left(\beta_i|0, \tau^2\right)$$

$$\text{IG}\left(\sigma^2|a_\alpha, b_\alpha\right)\text{IG}\left(\tau^2|a_\tau, b_\tau\right)$$

$$\text{Gam}\left(a_\theta^{(amb)}|a_{\theta_1}, a_{\theta_2}\right)\text{Gam}\left(a_\theta^{(ele)}|a_{\theta_1}, a_{\theta_2}\right)$$

$$\text{Gam}\left(b_\theta|b_{\theta_1}, b_{\theta_2}\right)\text{Gam}\left(u_f|5, 1)\right)\text{Gam}\left(u_h|100, 1)\right)$$

$$\text{Beta}\left(v_f|a_{v_f}, b_{v_f}\right)\text{Beta}\left(v_h|a_{v_h}, b_{v_h}\right).$$

Running MCMC

- The full model requires up to 200K iterations for convergence, using previous runs to provide starting values.
- Parameter estimation then performed on samples of 5000 iterations, post-convergence.
- Posterior estimates are contained in Table 3.1 (p. 57).
- Convergence assessment included running multiple chains, starting from overdispersed initial conditions.
- One diagnostic useful in assessing convergence is the Gelman-Rubin shrink factor, which calculates statistics involving *within chain* and *between chain* variances.
- One usually will require other convergence diagnostics to be sure that convergence has been obtained.

Gibbs Sampler Output

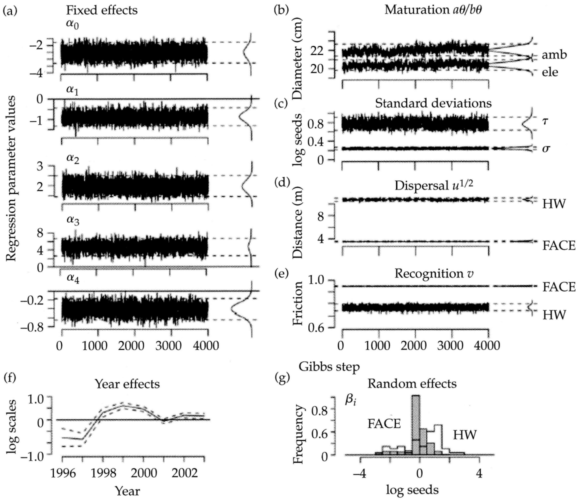


Figure 3.5. (a–e) Example of Gibbs sequence following 200,000 iterations to achieve convergence. Smooth posterior density estimates are shown to the right of each sequence. (f) 95% credible intervals for year effects. (g) Distribution of random effects across all trees, with FACE trees shown as a shaded histogram.

Posterior Predictions

- Cone count predictions are very good when compared to observed FACE cone count data.
- Elevated CO_2 levels increased fecundity of trees that are mature.
- Elevated CO_2 levels increased the rate at which smaller trees became mature (Fig. 3.6(c), p. 57).

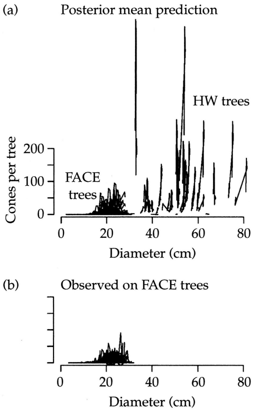


Figure 3.7. Posterior predicted seed production for individual trees and years (a) compared with cone counts for FACE trees (b).

Why Hierarchial Bayes

- HB allows one to embed standard models within a structure which allows for relatively easy modification based on known prior information and model design changes.
- Analyzing data obtained from simple or complex field experiments, once the model is set up correctly, can be done with powerful computational tools such as MCMC methods. Assesing convergence of these methods though is somewhat of an art.
- Working with local processes, identifying links to other data, processes, and parameters, one can then combine into a robust and complex HB model.
- Don't be intimidated by the complexity!

References

- Carlin, B. and Louis, T. (2009).
  Bayesian Methods for Data Analysis.
  Texts in Statistical Science. Chapman and Hall/CRC, New York, third edition.

- Clark, J. and LaDeau, S. (2006).
  Synthesizing ecological experiments and observational data with hierarchial Bayes.
  In Clark, J. and Gelfand, A., editors, *Hierarchial Modelling for the Environmental Sciences - Statistical Methods and Applications*, chapter 3, pages 41–58. Oxford University Press, New York.

- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004).
  Bayesian Data Analysis.
  Texts in Statistical Science. Chapman and Hall/CRC, New York, second edition.