Ecology, 88(10), 2007, pp. 2489–2495

ECOLOGICAL APPLICATIONS OF MULTILEVEL ANALYSIS OF VARIANCE

Song S. Qian[1,3] and Zehao Shen[2]

[1]*Nicholas School of the Environment and Earth Sciences, Duke University, Durham, North Carolina 27708 USA*
[2]*Department of Ecology, College of Environment Sciences, Peking University, Beijing, China*

Abstract. A Bayesian representation of the analysis of variance by A. Gelman is introduced with ecological examples. These examples demonstrate typical situations encountered in ecological studies. Compared to conventional methods, the multilevel approach is more flexible in model formulation, easier to set up, and easier to present. Because the emphasis is on estimation, multilevel models are more informative than the results from a significance test. The improved capacity is largely due to the changed computation methods. In our examples, we show that (1) the multilevel model is able to discern a treatment effect that is smaller than the conventional approach can detect, (2) the graphical presentation associated with the multilevel method is more informative, and (3) the multilevel model can incorporate all sources of uncertainty to accurately describe the true relationship between the outcome and potential predictors.

Key words: ANCOVA; conventional ANOVA; Bayesian statistics; hierarchical model; multilevel ANOVA; variance components.

Introduction

Analysis of variance (ANOVA) is widely used in scientific research for testing multiple, often complicated, hypotheses. As originally presented in Fisher's seminal work (Fisher 1925), ANOVA can be viewed as the collection of the calculus of sums of squares and the associated models and significance tests. These tests and models have had a profound impact on ecological studies. ANOVA provides the computational framework for the design and analysis of ecological experiments (Gotelli and Ellison 2004). As a data analysis tool, ANOVA is used in ecology for both confirmative and explorative studies. When used in a confirmative study, the randomized experimental design ensures that the resulting difference between treatments can be attributed to the cause we are interested in testing. When used in explorative studies, the ANOVA framework reflects a basic scientific belief that correlation *implies* a causal relationship (Shipley 2000). The simple steps of ANOVA computation, along with the associated significance test (the *F* test), can be implemented quickly, leading to seemingly straightforward interpretation of the results.

However, interpretation of ANOVA results can be problematic. Difficulties arise when the normality and independence assumptions of the response data are not met, when the experimental design is nested or unbalanced, or when there are missing values. More importantly, ANOVA results are difficult to explain in ecological terms because the significance test is often not very informative (Anderson et al. 2000). On the one hand, when an experiment is proposed, we almost always have reasons to believe that a treatment effect exists. Therefore, we often want to know the strength of the effect a treatment has on the outcome rather than whether the treatment has an effect or not on the outcome. By using a significance test where the inference is based on the assumption of no treatment effect, we emphasize the type I error rate (erroneously rejecting the null hypothesis of no treatment effect) often at the expense of statistical power, especially when multiple comparisons are used. On the other hand, a nonexistent treatment effect can be shown to be statistically significant if one tries often enough (Ioannidis 2005).

From a practical perspective, we find the concept of variance components (Searle et al. 1992) especially useful. The partitioning of total response variable variance into components representing treatment and other factors provides a means for quantifying the effects of the factors of interest. This partitioning is more informative than the results of a significance test, although all the ambiguity and difficulty of ANOVA still exist. The multilevel (or hierarchical) ANOVA of Gelman (2005) can be used to alleviate the ambiguity

Manuscript received 11 December 2006; revised 30 March 2007; accepted 3 April 2007; final version received 30 April 2007. Corresponding editor: A. M. Ellison.
[3] E-mail: song@duke.edu

and difficulty of ANOVA. The multilevel ANOVA can be summarized as the estimation of the variance components and treatment effects using a hierarchical regression. The results are often presented graphically. This approach is intuitively appealing and its implementation is straightforward even when the experimental design is nested and the response variable is not normally distributed. The computation of multilevel ANOVA is Bayesian and inference about treatment effects are made using Bayesian posterior distributions of the parameters of interest. Gelman and Tuerlinckx (2000) suggested that the hierarchical Bayesian approach for ANOVA includes the classical ANOVA as a special case. We introduce Gelman's multilevel ANOVA using three examples. Statistical background is presented in Gelman (2005), Gelman and Hill (2007), and Gelman and Tuerlinckx (2000), and briefly discussed in Appendices A–D.

Methods

We illustrate the multilevel ANOVA approach using a one-way ANOVA setting. For a one-way ANOVA problem, we have a treatment with several levels, and the statistical model is

$$y_{ij} = \beta_0 + \beta_i + \varepsilon_{ij} \tag{1}$$

where β_0 is the overall mean, β_i is the treatment effect for level i and $\Sigma\beta_i = 0$, and j represents individual observations in treatment i. This is a multilevel problem because we are interested in parameters at two levels: the data level and the treatment level. At the data level, individual response variable values are governed by the group level parameters, and the group level parameters are further governed by a distribution with hyperparameters. The term "multilevel" is used to describe the data structure (data points are clustered in multiple treatment levels) and the hierarchical model structure. We avoid terms "fixed" and "random" effects to avoid confusion as described in Gelman and Hill (2007: sections 1.1 and 11.4) and Gelman (2005: section 6). The term "multilevel" encompasses both fixed and random effects. The total variance in y_{ij} is partitioned into between-group variance [var(β_i)] and within-group variance [var(ε_{ij})]. Instead of using the sums-of-squares calculation, we use a hierarchical formulation and model the coefficients β_i as a sample from a normal distribution with mean 0 and variance σ_β^2:

$$\beta_i \sim \mathcal{N}(0, \sigma_\beta^2). \tag{2}$$

The model error term ε_{ij} is also modeled as from a normal distribution:

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

Or, equivalently

$$y_{ij} \sim \mathcal{N}(\beta_0 + \beta_i, \sigma^2). \tag{3}$$

The variance component for the treatment can be naturally estimated by σ_β, or the standard deviation of $\hat{\beta}_i$ (s_β, the finite population standard deviation). The computation expressed here is standard for random effect coefficients under the classical random effect models (Clayton 1996). We can view fixed effects as special cases of random effects ($\sigma_\beta = \infty$) in a Bayesian context. Therefore, this computation is not unique to the Bayesian framework.

Model coefficients (β's and σ_β, or s_β) can be estimated using the maximum likelihood estimator, and the likelihood function of this setting is a product of two normal distribution density functions defined by Eqs. 1 and 2. Analytical solutions are available, but it is easy to implement the computation using Markov chain Monte Carlo simulation (MCMC [Gilks et al. 1996, Qian et al. 2003]).

When there is more than one factor affecting the outcome, we can easily extend the approach by using the same hierarchical representation of the additional factors. Furthermore, as suggested by Eq. 3, this approach is not limited by the normality assumption. That is, the normal distribution in Eq. 3 can be replaced with any distribution from the exponential family, similar to the generalization from linear models to the generalized linear models (GLM [McCullagh and Nelder 1989]).

Data sets and models

We illustrate the multilevel ANOVA using three data sets. The intertidal seaweed grazers example, a textbook example of ANOVA, is intended to make a direct comparison between the multilevel ANOVA and the classical ANOVA. The Liverpool moths example is used to illustrate the application of multilevel model under a logistic regression setting, where the response variable follows a binomial distribution. The seedling recruitment data set illustrates the use of this approach for count data (Poisson regression) that may be spatially correlated. Results from applying the classical ANOVA or linear modeling are presented in Appendices A–D.

1. *Intertidal seaweed grazers*.—This example was used in the text by Ramsey and Schafer (2002: Case Study 13.1, p. 375), describing a randomized experiment designed to study the influence of three ocean grazers—small fish, large fish, and limpets—on the regeneration rate of seaweed in the intertidal zone of the Oregon coast. The experiments were carried out in eight locations to cover a wide range of tidal conditions and six treatments were used to determine the effect of different grazers (control, no grazer allowed; only limpets allowed; only small fish allowed; large fish excluded; limpets excluded; and all allowed). The response variable is the seaweed recovery of the experimental plot, measured as the percentage of the plot covered by regenerated seaweed. The standard approach illustrated in Ramsey and Schafer (2002) is a two-way ANOVA (plus the interaction effect) on the

logit-transformed percent regeneration rates. The logit of percent regeneration rate is the logarithm of the regeneration ratio (percent regenerated/percent not regenerated).

Using the multilevel notation, this two-way ANOVA model can be expressed as

$$Y_{ijk} = \beta_0 + \beta_{1i} + \beta_{2j} + \beta_{3ij} + \varepsilon_{ijk} \tag{4}$$

where Y is the logit of regeneration rate, β_{1i} is the treatment effect ($i = 1, \ldots, 6$ and $\Sigma\beta_{1i} = 0$), β_{2j} is the block effect ($j = 1, \ldots, 8$ and $\Sigma\beta_{2j} = 0$), and β_{3ij} is the interaction effect ($\Sigma\beta_{3ij} = 0$). The residual term ε_{ijk} is assumed to have a normal distribution with mean 0 and a constant variance, where $k = 1, 2$ is the index of individual observations within each block–treatment cell. The total variance in Y is partitioned into four components: treatment, block, interaction effects, and residual.

2. *Liverpool moths.*—Bishop (1972) reported a randomized experimental study on natural selection. The experiment was designed to answer the question whether tree trunks blackened by air pollution near Liverpool, England were the cause of the increase of a dark morph of a local moth. The moths in question were nocturnal, resting during the day on tree trunks. In Liverpool, a high percentage of the moths were of a dark morph, whereas a higher percentage of the typical (pepper-and-salt) morph were observed in the Welsh countryside, where tree trunks were lighter. Bishop selected seven locations progressively farther away from Liverpool. At each location, Bishop chose eight trees at random. Equal numbers of dead light and dark moths were glued to the trunks in lifelike positions. After 24 hours, a count was taken of the numbers of each morph that had been removed, presumably by predators. The original study was published before the time of GLM, but the data set has since been used in several regression textbooks as an example of logistic regression (e.g., Ramsey and Schafer 2002). We chose to model the moth data using a binomial distribution and the typical logistic regression model:

$$\begin{aligned} y_{ij} &\sim \mathrm{Bin}(p_{ij}, n_{ij}) \\ \mathrm{logit}(p_{ij}) &= \beta_0 + \beta_{1i} + \beta_{2i} \times \mathrm{dist}_j \end{aligned} \tag{5}$$

where y_{ij} is the number of moths removed for morph i ($i = 1$ [dark] or 2 [light]), at the jth distance (distance from Liverpool), n_{ij} is the total number of moths placed on each tree, p_{ij}, the parameter of interest, is the probability a moth being removed, β_{1i} is the morph effect, and β_{2i} is the slope of distance, representing the interaction between morph and distance. Bishop (1972) and Ramsey and Schafer (2002) used a categorical predictor "site" instead of distance from Liverpool to account for an apparent outlier at distance of 30.2 km. We choose to use distance as a continuous predictor to illustrate the interaction effect between a categorical predictor and a continuous predictor.

3. *Seedling recruitment.*—Shen (2002) reported an observational study on factors affecting the species composition and diversity of a mixed evergreen–deciduous forest community in southwest China. The original study had observations at multiple spatial scales. We use the seedling recruitment data collected along a transect of 128 consecutive 5 × 5 m plots to study factors affecting seedling recruitment. A total of 49 species of seedlings were observed in the field and were classified into five types according to their status in the community (Shen et al. 2000): Pioneer, early dominant, early companion, later dominant (including evergreen species), and tolerant. Within each plot, the number of seedlings with a height below 1 m were recorded, along with several physical and biological variables, including canopy (gap, %), the position of each plot measured as the relative position along a hillside between valley (position = 1) and ridge (position = 5), and soil total organic carbon (TOC, %).

The observed number of seedlings (the response variable) from different plots are likely correlated, and the correlation is likely due to the spatial layout of the transect. A natural strategy to deal with this problem is to introduce a spatial random effects term ϵ using the intrinsic conditional autoregressive model (CAR [Besag et al. 1991, Qian et al. 2005]):

$$\begin{aligned} y_{ijkl} &\sim \mathrm{Pois}(\mu_{ijk}) \\ \log(\mu_{ijkl}) &= \beta_0 + \beta_{1j} + \beta_{2k} + \beta_{3j} \times \mathrm{logit}(\mathrm{gap}_i) \\ &\quad + \beta_{4j} \times \mathrm{TOC} + \epsilon_i + \varepsilon_i \end{aligned} \tag{6}$$

where i is the index of plot, j is tree type index ($j = 1, \ldots, 5$), k is the index of position, and l is the index of observations within a plot. The spatial random effect term ϵ_i has a CAR prior. It is used to model the spatially structured variation. The error term ε_i is used to account for unstructured over-dispersion. The sum of ϵ_i and ε_i is termed as the convolution prior (Besag et al. 1991). Only two-way interactions between tree type and the two continuous predictors were considered.

Results

The results from classical ANOVA are presented in Appendix A. All multilevel results are presented graphically, showing the estimated posterior mean (the circle), and the 50% (thick line) and 95% (thin line) posterior credible intervals. The ANOVA display shows the estimated posterior distributions of variance components (in standard deviation), and the effects plots are based on estimated posterior distributions of effects.

Seaweed grazers

The multilevel model results are qualitatively similar to the conventional ANOVA results (Fig. 1). In addition, the estimated main effects are similar to results from a conventional ANOVA (Figs. 2 and 3). The emphasis on estimation is clearly displayed in these

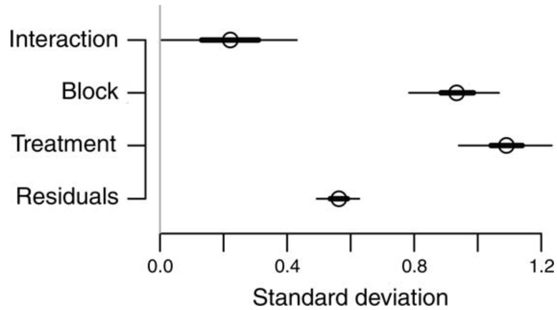

FIG. 1. Seaweed example, with ANOVA display of the estimated standard deviation of the estimated variance components showing a general pattern similar to that of the conventional ANOVA. Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

plots. From the main effects plots, we know that the maximum difference between the treatment and control is about 3 in the logit scale, meaning the mean regeneration ratio of the control sites is about 20 times (e^3) larger than that of the sites where all grazers were allowed. Traditional ANOVA does not emphasize the interaction effect beyond whether or not it is statistically significant. See Appendix A for a comparison of the multilevel interaction plot and the commonly used interaction plot in ANOVA.

Moths

Using the multilevel model, we can use the traditional analysis of covariance (ANCOVA) to calculate the variance components of the main morph effect, main distance effect and the interaction (Figs. 4 and 5). For this particular example, we see a strong interaction effect (clearly expressed by the difference in the distance slope

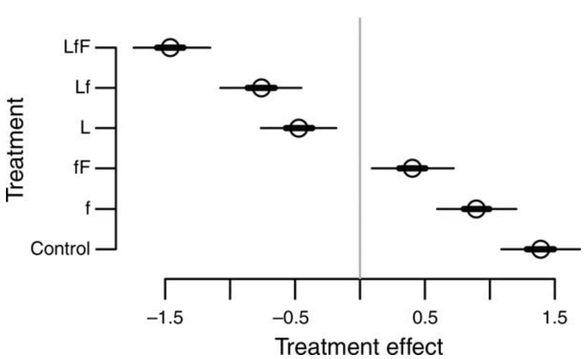

FIG. 2. Estimated treatment main effect of the seaweed grazer example shows that the regeneration rate decreases as grazing pressure increases. The six treatments are: C, control, no grazers allowed; L, only limpets allowed; f, only small fish allowed; Lf, large fish excluded; fF, limpets excluded; and LfF, all grazers allowed. The largest difference between treatments is about 3 (in logit scale); the regeneration ratio in the control is about 20 times (e^3) larger than the regeneration ratio in the LfF treatment (all grazers allowed). Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

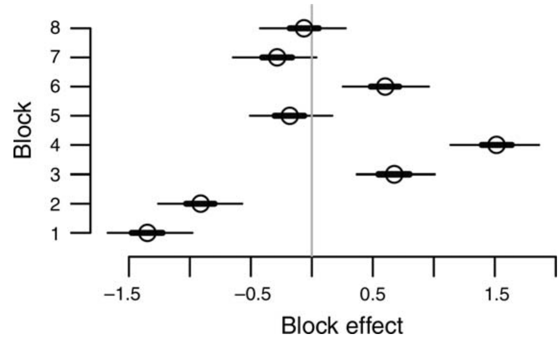

FIG. 3. Estimated block main effect of the seaweed grazer example showing that the block effect has approximately the same magnitude as the treatment effect (3 in logit scale). Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

in Fig. 5) and an unambiguous morph main effect (Fig. 4b). The morph main effect is obvious because it is evaluated at a distance of 27.2 (the average distance).

Seedling recruitment

The largest variance component is the unstructured over-dispersion term, while the structured spatial random effects (CAR) contributes a smaller than expected variance (Fig. 6). The inclusion of spatially correlated predictors may have accounted for some of the spatial autocorrelation in the response variable data. Tree type index explains the most variation in recruitment (Figs. 6 and 7a), which is expected since tree species were classified to reflect their different ecological strategies and roles in community dynamics. Similar to the GLM results, we found that the effect of gap is uncertain. However, the type × gap interaction effect, as shown in terms of type-specific gap slope (Fig. 8b), indicates that type 4 (late dominant) trees are likely to respond negatively to increased gap, while the other types likely respond positively. The interaction effect between type and TOC (Fig. 8a) is unambiguous. Because soil carbon concentration is likely to be similar in neighboring plots, including a spatial autocorrelation term reduces uncertainty on type-specific TOC slopes. Type 4 and 5 trees include all evergreen species and the shade-tolerant deciduous species which tend to be restricted to relatively steep and higher hillside positions, corresponding to a lower soil TOC value, while the deciduous dominant and pioneer species normally experienced rapid recruitment and fast growth in richer habitat. This pattern has also been reported in similar contexts (Tang and Ohsawa 2002). Compared to the type-specific TOC slopes from the multilevel model (Fig. 8a), the GLM fit (see Appendices A–D) is quite different. Because spatial autocorrelation is accounted for, the multilevel model results are more reliable. The position main effect (Fig. 7b) shows a clear pattern indicating increased recruitment from valley to ridge.

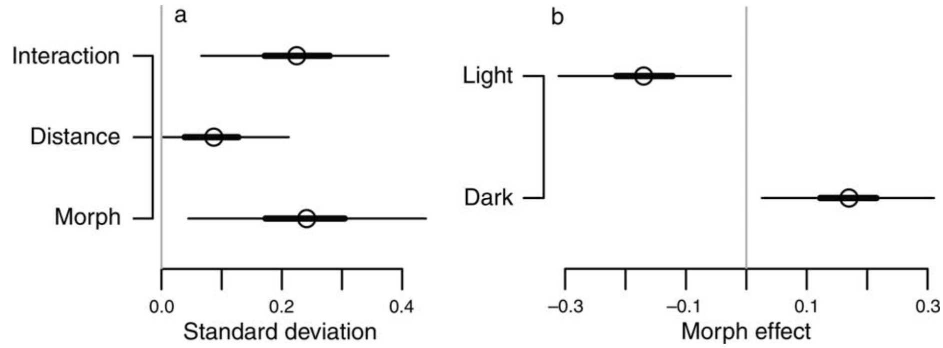

FIG. 4. Liverpool moth example: (a) the ANOVA table indicating strong morph main effect and the morph × distance interaction effect; (b) the estimated morph color main effects. Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

The large uncertainty associated with the position main effect can be attributed to the qualitative nature of this variable. That is, topographic relief has multiple scales while the plot size is fixed. Assigning position to a plot can be ambiguous depending on the length of a hillside. Plots with the same position level could be at quite different absolute positions on hillsides of different sizes. As a result, an emphasis on estimation is more informative than the hypothesis-testing approach which will almost surely lead to a nonsignificant result.

DISCUSSION

Our examples used typical data sets encountered in ecological studies. Although ANOVA is well suited for analyzing the seaweed grazer data, multilevel ANOVA can be more informative and the graphical display is easier to understand and interpret. In many ecological studies, data are collected from observations or from experiments applied to a limited number of plots with unobserved confounding factors. Large natural variability plus small sample size often lead to nonsignificant results from ANOVA or *t* tests, because the significance test is based on the comparison of the variance due to treatment and the residual variance. This situation is very common because of the high cost of collecting ecological data. When using the multilevel ANOVA, we estimate the treatment effect directly. The estimated treatment effect posterior distribution is not directly associated with the residual variance. Consequently, we are more likely to discern a treatment effect.

The seaweed regeneration example compared the multilevel ANOVA to the conventional ANOVA. The comparison illustrates the multilevel ANOVA's emphasis on estimation. This emphasis yields more informative results presented in terms of the estimated effects and the associated uncertainty. As we discuss in Appendices A–D, conventional hypothesis testing on treatment effects can be performed using the 95% posterior distributions of effects. As a result, our emphasis on estimation does not lead to the loss of information in terms of comparisons of treatment effects. More importantly, the hierarchical computational framework allows the ANOVA concept to be applied to non-normal response variables, as illustrated in the Liverpool moth and seedling-recruitment examples.

The philosophical basis of the traditional ANOVA is Popper's falsification theory (Popper 1959). Although not fully compatible with methods practiced by most scientists, Popper's falsification philosophy had an immense impact on Fisher. Because statistical theories are not strictly falsifiable, Fisher devised his methodology based on a quasi-falsificationist view. Fisher held that a statistical hypothesis should be rejected by any experimental evidence which, on the assumption of that hypothesis, is relatively unlikely, relative that is to other possible outcomes of the experiment. Such tests, known as significance tests, or null hypothesis tests are controversial (see for example, Anderson et al. 2000 and Quinn and Keough 2002).

Although Fisher's principles of randomized experimental design provide a mechanism for discerning the true causal effect of interest from confounding correlations, in practice ANOVA and associated significance

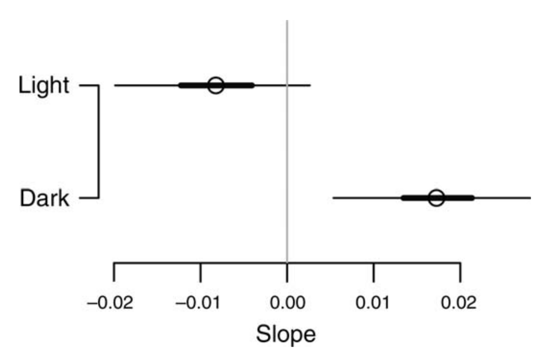

FIG. 5. The estimated distance slope is positive for dark moths, indicating increased risk of removal for dark moths away from Liverpool. The distance slope for light moths is most likely negative, indicating increased risk of removal closer to Liverpool. Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

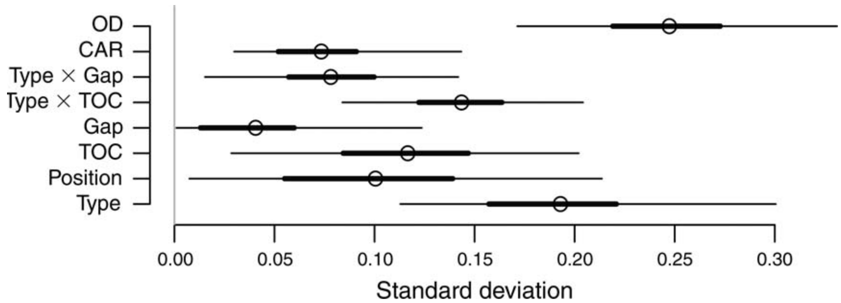

FIG. 6. Seedling example, with ANOVA display of the estimated standard deviations of the estimated variance components showing that the unstructured overdispersion (OD) is the main contributor of the total variance, followed by tree type, type × TOC interaction, TOC, position, gap, and spatial autocorrelation (CAR; conditional autoregressive model). Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

tests are applied in both exploratory and confirmatory studies. In a confirmatory study, significance tests associated with ANOVA are used as the "seal of approval," while in an exploratory study ANOVA is often used to infer potential factors that may affect the outcome. While the null hypothesis of a significance test is of little interest, the variance component concept of ANOVA provides a convenient structure that allows scientists to form a causal model and develop hypotheses. The new computational method of multilevel ANOVA allows the classical ANOVA concept be applied to more complicated situations and can be accepted by both Bayesian and frequentist practitioners.

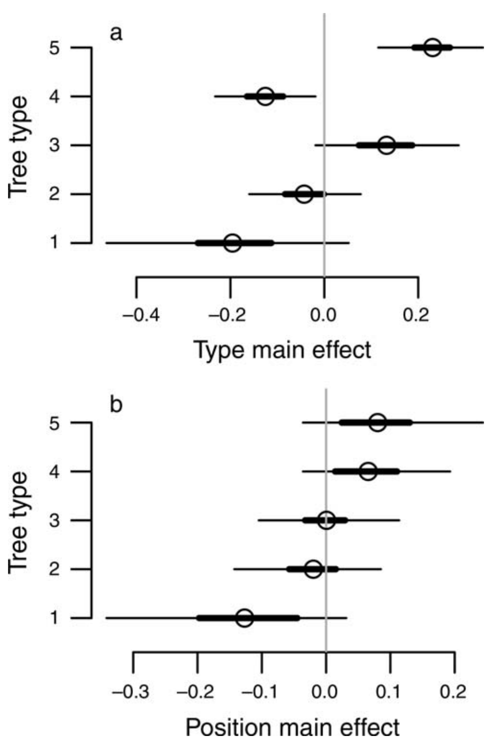

FIG. 7. Seedling example. (a) The tree type main effect shows that pioneer (1) and late-dominant (4) types tend to have fewer seedlings and tolerant (5) types tend to have much higher recruitment, while early-dominant (2) and early-companion (3) types are close to average. (b) The position main effect shows that recruitment increases when moving from valley to ridge. Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

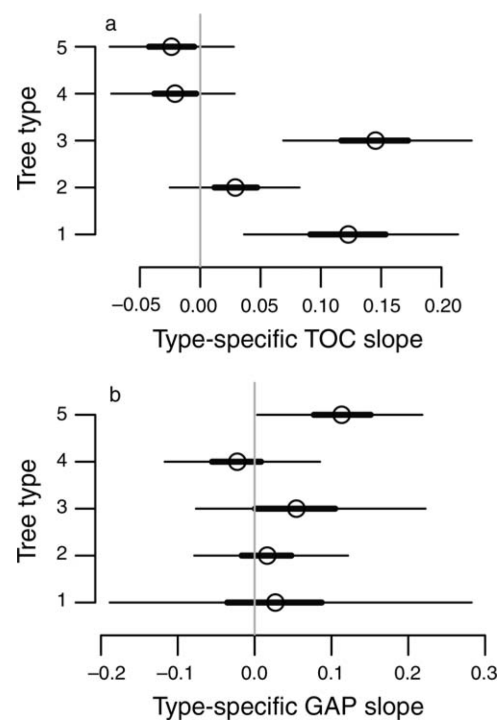

FIG. 8. Seedling example. (a) The type × TOC interaction shows generally positive slopes for pioneer (1), early-dominant (2), and early-companion (3) species and generally negative slopes for late-dominant (4) and tolerant (5) species. (b) The type × gap interaction shows that only tolerant species respond positively to gaps. Circles are estimated posterior means, short thick lines are the 50% posterior credible intervals, and the long thin lines are the 95% posterior credible intervals.

Acknowledgments

S. Qian's work is supported by the U.S. EPA's STAR grant (# RD83244701). Z. Shen's work is supported by the Chinese National Natural Science Foundation (No. 30000024) and a China Scholarship Council grant for visiting study. Andrew Gelman generously shared the computer programs used in his 2005 paper. Collaboration with John Terborgh allowed us to explore the method. Many students at Nicholas School of the Environment and Earth Sciences of Duke University provided valuable data sets that facilitated our learning of the method. Comments and suggestions from two reviewers (Gerry Quinn and Mike Meredith), Kevin Craig, Craig Stow, Ben Poulter, and Ariana Sutton-Grier are greatly appreciated.

Literature Cited

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64:912–923.

Besag, J., J. York, and A. Mollie. 1991. Bayesian image restoration, with two applications in spatial statistics (with discussion). Annals of the Institute of Statistical Mathematics 43:1–59.

Bishop, J. A. 1972. An experimental study of the cline of industrial melanism in *Biston betularia* [Lepidoptera] between urban Liverpool and rural North Wales. Journal of Animal Ecology 41:209–243.

Clayton, D. G. 1996. Generalized linear mixed models. Pages 275–301 *in* W. R. Gilks, S. Richardson, and D. J. Spielelhalter, editors. Markov chain Monte Carlo in practice. Chapman and Hall, London, UK.

Fisher, R. A. 1925. Statistical methods for research workers. First edition. Oliver and Boyd, Edinburgh, UK.

Gelman, A. 2005. Analysis of variance: why it is more important than ever (with discussions). Annals of Statistics 33(1):1–53.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, New York, USA.

Gelman, A., and F. Tuerlinckx. 2000. Type S error rates for classical and Bayesian single and multiple comparison procedures. Computational Statistics 15:373–390.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, editors. 1996. Markov chain Monte Carlo in practice. Chapman and Hall, New York, New York, USA.

Gotelli, N. J., and A. M. Ellison. 2004. A primer of ecological statistics. Sinauer Associates, Sunderland, Massachusetts, USA.

Ioannidis, J. P. A. 2005. Why most published research findings are false. PLoS Medicine 2(8):e124 [doi: 10.1371/journal.pmed.0020124].

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Chapman and Hall, London, UK.

Popper, K. P. 1959. The logic of scientific discovery. Hutchinson Education, London, UK.

Qian, S. S., K. H. Reckhow, J. Zhai, and G. McMahon. 2005. Nonlinear regression modeling of nutrient loads in streams: a Bayesian approach. Water Resources Research 41(7): W07012, 2005.

Qian, S. S., C. A. Stow, and M. Borsuk. 2003. On Monte Carlo methods for Bayesian inference. Ecological Modelling 159: 269–277.

Quinn, G. P., and M. J. Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press, Cambridge, UK.

Ramsey, F. L., and D. W. Schafer. 2002. The statistical sleuth, a course in methods of data analysis. Second edition. Duxbury, Pacific Grove, California, USA.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. Wiley, New York, New York, USA.

Shen, Z. 2002. A multiscale study on the vegetation–environment relationship of a mountain forest transect. [In Chinese with English abstract.] Acta Ecologica Sinica 22:461–470.

Shen, Z., X. Zhang, and Y. Jin. 2000. An analysis of the topographical pattern of the chief woody species at Dalaoling Mountain in the Three Gorges. [In Chinese with English abstract.] Acta Phytoecologica Sinica 24:581–589.

Shipley, B. 2000. Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference. Cambridge University Press, Cambridge, UK.

Tang, C. Q., and M. Ohsawa. 2002. Coexistence mechanisms of evergreen, deciduous, and coniferous trees in a mid-montane mixed forest on Mt. Emei, Sichuan, China. Plant Ecology 161:215–230.

APPENDIX A

A comparison of the conventional ANOVA and the multilevel, or Bayesian, ANOVA (*Ecological Archives* E088-150-A1).

APPENDIX B

Using Bayesian ANOVA results for hypothesis testing (*Ecological Archives* E088-150-A2).

APPENDIX C

Results of the three examples from using the conventional linear modeling approach (*Ecological Archives* E088-150-A3).

APPENDIX D

A tutorial of the computer programs (WinBUGS and R) used in the paper (*Ecological Archives* E088-150-A4).

SUPPLEMENT

Data sets used in the paper (*Ecological Archives* E088-150-S1).