JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 45, No. 2 AMERICAN WATER RESOURCES ASSOCIATION April 2009

A MULTILEVEL MODEL OF THE IMPACT OF FARM-LEVEL BEST MANAGEMENT PRACTICES ON PHOSPHORUS RUNOFF[1]

K.H. Reckhow, S.S. Qian, and R.D. Harmel[2]

ABSTRACT: Multilevel or hierarchical models have been applied for a number of years in the social sciences but only relatively recently in the environmental sciences. These models can be developed in either a frequentist or Bayesian context and have similarities to other methods such as empirical Bayes analysis and random coefficients regression. In essence, multilevel models take advantage of the hierarchical structure that exists in many multivariate datasets; for example, water quality measurements may be taken from individual lakes, lakes are located in various climatic zones, lakes may be natural or man-made, and so on. The groups, or levels, may effectively yield different responses or behaviors (e.g., nutrient load response in lakes) that often make retaining group membership more effective when developing a predictive model than when working with either all of the data together or working separately with the individuals. Here, we develop a multilevel model of the impact of farm level best management practices (BMPs) on phosphorus runoff. The result of this research is a model with parameters which vary with key practice categories and thus may be used to evaluate the effectiveness of these practices on phosphorus runoff. For example, it was found that the effect of fertilizer application rate on farm-scale phosphorus loss is a function of the application method, the hydrologic soil group, and the land use (crop type). Further, results indicate that the most effective method for controlling fertilizer loss is through soil injection. In summary, the resultant multilevel model can be used to estimate phosphorus loss from farms and hence serve as a useful tool for BMP selection.

(KEY TERMS: nonpoint source pollution; best management practices; nutrients; multilevel models; nutrients; hierarchical models.)

Reckhow, K.H., S.S. Qian, and R.D. Harmel, 2009. A Multilevel Model of the Impact of Farm-Level Best Management Practices on Phosphorus Runoff. *Journal of the American Water Resources Association* (JAWRA) 45(2):369-377. DOI: 10.1111/j.1752-1688.2008.00298.x

INTRODUCTION

In the late 1970s, social scientists expanded the utility of multivariate regression models by developing a "slopes as outcomes" modeling strategy. This approach, which ultimately led to multilevel (hierarchical) models, effectively recognized that in many multivariate regression modeling problems, the regression parameters can be themselves regressed as a function of other predictor variables. This initial perspective has evolved during the past 30 years into a variety of frequentist and Bayesian modeling approaches that retain group identification yet

[1]Paper No. JAWRA-08-0059-P of the *Journal of the American Water Resources Association* (JAWRA). Received April 7, 2008; accepted August 4, 2008. **Discussions are open until October 1, 2009**.

[2]Respectively, Professor (Reckhow) and Associate Professor of the Practice (Qian), Nicholas School of the Environment, Box 90328, Duke University, Durham, North Carolina 27708; and Agricultural Engineer (Harmel), United States Department of Agriculture-Agricultural Research Service, 808 East Blackland Road, Temple, Texas 76502 (E-Mail/Reckhow: reckhow@duke.edu).

"borrow strength" from other, related groups. For example, in the environmental sciences, Reckhow (1996) used empirical Bayes (EB) analysis to develop a model that could be used to predict the index of biotic integrity at specific sites in the Scioto River (Ohio) better than a single global regression model. A key attribute of EB estimators is that they are "shrinkage" estimators; that is, site-specific EB regression parameters shrink toward the grand (global) mean of the parameters from all sites. This process of "borrowing strength" from other sites is an important feature of multilevel models.

To fully appreciate the conceptual basis of the model proposed in this paper, we need to reflect on the underlying logic of multilevel (hierarchical) models. Consider the science of limnology and the task of lake eutrophication modeling. It is universally accepted that limnological science provides a wealth of knowledge concerning the behavior of lakes in response to their environment. Lakes are similar; they have commonalities and respond in known ways; thus we justifiably expect from the science of limnology that we can predict with some confidence how lakes will react to perturbations.

Yet, at the same time, we know that every lake is unique in some ways. Lakes are not identical porcelain bathtubs. They each exist within a unique watershed, climatic zone, geomorphologic setting, and have unique in-lake features. Thus, it is unreasonable to expect that all lakes will respond in exactly the same way to external perturbations or pollutant loadings.

If all lakes were identical, then from a statistical perspective, a single global regression model would serve equally well in predicting the response in any lake to nutrient loading. Correspondingly, if all lakes were strictly unique, then a separate eutrophication response model would be needed for each lake. However, as we know that limnology can scientifically describe the common behavior among lakes and that every lake has unique features, we are wise to seek a middle ground in statistical lake modeling. This is the role played by multilevel, or hierarchical, models.

Implicit multilevel model structure is surprisingly common in environmental and water quality studies (see Qian and Shen, 2007), yet it is rarely exploited in statistical modeling in these fields. A few exceptions to note are Reckhow (1996) and Wagner *et al.* (2006, 2007). A multilevel model is likely to be preferred over a standard multiple regression model in cross-sectional analyses when there are logical hierarchical groupings (e.g., samples within lakes, lakes within ecoregions, …) or categorical variables (natural *vs*. man-made lakes, run-of-river lakes, …). This advantage arises from retaining the group or categorical membership in construction of the model.

In fact, hierarchical groupings in datasets should be addressed using multilevel modeling for two important reasons – one practical and the other theoretical. From a practical standpoint, Reckhow (1996) and others have shown that by "borrowing strength" across groups, a multilevel model should outperform a standard regression model, particularly when sample sizes vary substantially among groups. From a theoretical standpoint, group data typically have correlated errors for the group variables, and this results in a violation of the least squares assumptions (resulting in a misleading reduction in parameter standard errors). Thus, least squares regression is not actually appropriate for application to a multilevel dataset that includes both individual and group level data.

OBJECTIVE

Given the potential of multilevel modeling to improve models involving data with hierarchical structure, the objective of this research is to develop a multilevel model of field scale phosphorus loading that statistically characterizes the impact of best management practices (BMPs). This model, if successful, could be used to predict how management practices affect phosphorus loads. To make this possible, Harmel *et al.* (2006) compiled a database from multiple studies of measured annual phosphorus (P) and nitrogen (N) loads, soil characteristics, fertilizer application rates and practices, plus other BMPs; see Table 1 for a complete listing of variables from the Measured Annual Nutrient loads from Agricultural Environments database. This database was intended in part to update previous work on nutrient export coefficients by Reckhow and others (Reckhow *et al.*, 1980; Beaulac and Reckhow, 1982).

Phosphorus (P), or more appropriately ortho-phosphorus as it commonly exists in the environment, has recently become a highly scrutinized nonpoint source pollutant. This focus has occurred because P is the limiting nutrient in many freshwater aquatic ecosystems and thus directly affects eutrophication (Carpenter *et al.*, 1998; Sharpley, 2000). Although numerous urban and natural P sources exist, the focus of the present research is runoff from agricultural land uses (LUs), which receive P from the application of commercial (inorganic) fertilizers and animal manures. The present research was designed to enhance contemporary efforts to quantify and model agricultural management practice impacts on water quality and P runoff (e.g., Sharpley *et al.*, 2001; Harmel *et al.*, 2004; Gitau *et al.*, 2005)

TABLE 1. Variables in the MANAGE Database.

Watershed ID – Name of the watershed.
Location (City, State) – City and state/province of the study (occasionally only a county or region was specified).
State – US state (or Canadian province) included to aid state-specific queries.
Location (Lat, Long) – Latitude and longitude of the study.
Date – Beginning and end of period with annual nutrient load data (not necessarily the entire study duration).
Watershed Years (ws yr) – Product of the number of monitored watersheds and the number of years with annual nutrient load data.
Land Use – Identification of crop or vegetation type(s) and crop rotation.
Tillage – Description of the tillage management divided into four options: no-till, conservation, conventional, or pasture.
Conservation Practice: Five options: waterway, terrace, filter strip, riparian buffer, or contour farming.
Dominant Soil Type – Soil textural class and soil series.
Hydrologic Soil Group – NRCS hydrologic soil group (HSG) classification (A, B, C, or D).
Soil Test P (ppm) – Maximum and minimum soil test P values for records with multiple watersheds or multiple years.
Soil Test P Extractant – Extractant used to determine soil test P.
Land Slope (%) – Maximum and minimum land surface slopes for records with multiple watersheds.
Watershed Size (ha) – Maximum and minimum watershed sizes for records with multiple watersheds.
Type of Fertilizer Applied – Macro-nutrient composition (N-P-K).
Fertilizer Application Method – Fertilizer application method divided into four options: surface, injected, incorporated, or other.

Annual maximum, minimum, and average values are provided for the following categories when specified:
N Applied (kg/ha-yr) – The total annual amount of N applied to watershed(s) from all fertilizer sources.
P Applied (kg/ha-yr) – The total annual amount of P applied to watershed(s) from all fertilizer sources.
Precipitation (mm/yr), **Runoff** (mm/yr)**, and Soil Loss** (kg/ha-yr).
Dissolved N (kg/ha-yr) – The total amount of dissolved N lost from the watershed(s).
Particulate N (kg/ha-yr) – The total amount of N lost from the watershed(s) in a particulate form.
Total N (kg/ha-yr) – Total N load was specified in a number of the publications. If the total N load was not specified, it was determined as the sum of dissolved and particulate N loads, when both were specified.
Dissolved P (kg/ha-yr) – The total amount of dissolved P lost from the watershed(s).
Particulate P (kg/ha-yr) – The total amount of P lost from the watershed(s) in a particulate form (associated with sediment).
Total P (kg/ha-yr) – Total P load was specified in a number of the publications. If the total P load was not specified, it was determined as the sum of dissolved and particulate P loads, when both were specified.
Specific form or laboratory analysis technique used to determine dissolved, particulate, and total N or P composition in runoff.
Total, Surface, Base Flow Indication – Indication of the flow transport mechanisms addressed.

including the USDA Conservation Effects Assessment Project (Mausbach and Dedrick, 2004).

MODEL DEVELOPMENT STRATEGIES

As noted in the introduction, the multilevel, or hierarchical, model strikes a balance between treating data from different studies as completely independent (thereby analyzing them separately, or unpooled) and treating data from different studies as replicates (thereby analyzing the combined data without considering the sources, or completely pooled). This balance is often difficult because of the differences in studies, in terms of sample size, variability, and other factors affecting the outcome.

The result of multilevel analysis is partial pooling, which means two things. First, for the model parameters, the overall (global) average is a weighted mean of study-specific averages. Second, the estimated study-specific parameters are shrunk toward the global mean, and the degree of shrinkage depends on the study-specific uncertainty. From a mathematical standpoint, partial pooling can be seen as an extension of the commonly used analysis of covariance (ANCOVA). To illustrate the concept, we use a simple ANCOVA example of phosphorus loading from an agriculture field as a function of fertilizer application with a dummy variable representing different BMP practices. To clarify the main concept of partial pooling, we assume a common slope. That is, we use a model of the form $y_i = \alpha_{j[i]} + \beta x_i + \epsilon$, where y is the log P loading from field i, x is the fertilizer application rate, β is the common slope, α is the intercept, and the subscript $j[i]$ represents that the i^{th} field receives the j^{th} BMP. With no pooling, the model is fitted by using a dummy variable for BMP and the regression model coefficients are estimated separately for each BMP group. If β is known, the no pooling estimate of α_j is $\bar{y}_j - \beta\bar{x}_j$. If the complete pooling intercept is μ_α, the partial pooling estimate of α_j is (Equation 1):

$$\hat{\alpha}_j \approx \frac{n_j/\sigma_y^2}{n_j/\sigma_y^2 + 1/\sigma_\alpha^2}(\bar{y}_j - \beta\bar{x}_j) + \frac{1/\sigma_\alpha^2}{n_j/\sigma_y^2 + 1/\sigma_\alpha^2}\mu_\alpha \quad (1)$$

which is a weighted average of the complete pooling and no pooling estimate of the intercept. The partial pooling equation is based on the assumption that the group-specific intercepts are from a common distribution (Equation 2):

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \qquad (2)$$

The resulting partial pooling estimate of the group intercept is a weighted average of the no pooling estimate ($\bar{y}_j - \beta\bar{x}_j$) and the complete pooling estimate (μ_α). The weight is determined by group sample size n_j, between group variance (σ_α^2), and within group variance (σ_y^2). If a particular group sample size n_j is large, or the between group variance is very large, the partial pooling estimate is close to the no pooling estimate. In fact when $\sigma_\alpha^2 \to \infty$ or $n_j \to \infty$, the partial pooling estimate converges to the no pooling estimate. Conversely, if the group sample size n_j is small or the within group variance σ_y^2 is large, the partial pooling estimate is closer to the complete pooling estimate. For a group with no data ($n_j = 0$), the partial pooling estimate is the same as the complete pooling estimate. When both the intercept and the slope are allowed to vary by group, the same general conclusion holds (Gelman and Hill, 2007).

Consider the problem addressed in this paper – the prediction of phosphorus loading from a range of agricultural sites. The database includes measurements of phosphorus loading, plus a number of candidate predictor variables, from 55 studies (Harmel *et al.*, 2008). If no pooling of the data from the 55 studies is undertaken, we will have a set of 55 predictive equations of the site-specific phosphorus loading, each with a standard error. However, if complete pooling is used, we pool the data from all 55 studies, yielding a single equation with standard error.

With no pooling, each site will have unique parameter estimates for the predictor variables as the development of the model for each site is conducted separately. With complete pooling, we assume that all studies share the same global parameter estimates and standard errors. With partial pooling, the varying level of shrinkage of each parameter is determined by the ratio of within and between group variances. The estimated site-specific regression parameters are a compromise between the two extremes, completely pooled and unpooled. The hierarchically estimated site parameters are approximately the weighted average of the global parameter and the unpooled site-specific parameter. In effect, the weighted average reflects the amount of information in the parameter from an individual site regression and the information in the global regression based on all sites. A regression parameter from a site with a small sample size has less information, and the hierarchical estimate is pulled closer to the overall average. A regression parameter from a site with a large sample size has more information and the hierarchical estimate is pulled toward the site average.

The models presented in this paper were fitted using the maximum likelihood estimator (MLE; Laird and Ware, 1982) for multilevel models implemented in the R function Linear Mixed-Effects Models (lmer) (Pinheiro and Bates, 2000). The MLE approach is convenient and competing models can be evaluated quickly. However, the MLE has limited flexibility and often can exhibit difficulties in estimating variance parameters when the number of levels is small. In addition, the MLE algorithm cannot handle missing values properly. Thus, while a Bayesian model might be considered to address these difficulties, the conventional frequentist approached employed here facilitates the exploration of a large number of candidate models.

MULTILEVEL MODEL FITTING AND SELECTION

For this study, three key continuous predictor variables emerged based on our initial exploratory analysis: water runoff, soil loss, and phosphorus (fertilizer) application rate. Our initial analysis also revealed that, for this multilevel regression model, the regression parameters may vary depending on the following categorical factors: LU (crop type), hydrologic soil group (HSG), fertilizer application method, tillage method, conservation practice, and study site. In our exploratory stage, we used lmer to fit various combinations of continuous and categorical predictors and evaluated the model fit. The model evaluation process is mostly subjective, as we cannot use conventional methods such as the Akaike information criterion or the Bayesian information criterion because missing data for individual predictor variables results in different sample sizes for different combinations of variables. If we eliminated all cases with missing values, we would be left with an extremely small dataset.

In selecting a model, we wanted to simplify the final model by using the minimum combination of the three continuous predictor variables and the six potential categorical variables. A full model allows the intercept and three slopes (for water runoff, soil loss, and fertilizer rate) to vary according to all six categorical variables. In simplifying this full model, we needed to determine whether the six potential categorical variables could be dropped from one or more slopes or from the intercept. Our first task in the exploratory study using lmer was to determine whether or not different study sites resulted in different model coefficients. Our initial hypothesis was that there would be a site effect on the intercept term, which represents the mean phosphorus delivery rate; this term was expected to vary from site to site reflecting the differences in soil, climate, and other factors among study sites. Our exploratory analysis confirmed this hypothesis. However, the site effect on the three slopes is statistically insignificant (at the 5% level). When examining the LU effects, we noticed that the LU effects and site effects cannot be effectively estimated together. Upon further examination, we discovered that at most sites, experiments were on one LU and on one HSG. As a result, the effects of study site and these two categorical variables are confounded. To address this and simplify the model, we removed study site as a categorical predictor; thus, the study site effects are represented in the LU and the HSG effects.

After fitting over 20 models, we found the following model provides the best fit and physical interpretation for predicting phosphorus loading from agricultural fields:

$$\log(P_{Li}) = \beta_{0,LU[i],AppM[i],Till[i],HSG[i]} + \beta_{1,LU[i],AppM[i],HSG[i]} \log(x_{1i}) + \beta_{2,Till[i]} \log(x_{2i}) + \beta_3 \log(x_{3i}) + \epsilon, \quad (3)$$

where P_{Li} is annual total phosphorus runoff (kg/ha-yr) from the study site (i), x_1 is the annual amount of fertilizer applied (kg/ha-yr), x_2 is the annual water runoff (mm/yr) from the field, x_3 is the annual soil loss (kg/ha-yr), the β's are the regression intercept/slopes, and ϵ is the regression error term. This basic model structure involving these three continuous variables seems plausible as fertilizer rate determines to a great extent the source (the amount of P available for runoff), runoff provides the transport mechanism and thus gives an indication of the dissolved P portion, and soil loss gives an indication of the particulate P loss. The categorical variables [LU, fertilizer application method (AppM), tillage method (Till), and HSG] determine various β-parameters in the multilevel model. Thus for example:

$$\beta_{0,LU[i],AppM[i],Till[i],HSG[i]} = \mu_{\beta 0} + \gamma_{\beta 0,LU} + \gamma_{\beta 0,AppM} + \gamma_{\beta 0,Till} + \gamma_{\beta 0,HSG} + \epsilon, \quad (4)$$

where $\mu_{\beta 0}$ is the overall mean effect intercept, γ_{LU} is the LU effect, γ_{AppM} is the fertilizer application method, γ_{Till} is the tillage method, and γ_{HSG} is the hydrologic soil group.

TABLE 2. Estimated Overall Mean Effects.

	β_0	β_1	β_2	β_3
Estimate	0.144	0.232	0.373	0.549
Standard Error	0.422	0.191	0.095	0.052

TABLE 3. Estimated Categorical Variable Effects.

Land Use (Crop Type)	$\gamma_{\beta 0,\ LU}$	$\gamma_{\beta 1,\ LU}$
Alfalfa	0.645197	−0.125893
Corn	−0.157255	−0.153294
Cotton	−0.056698	−0.101043
Fallow	0.226142	−0.150826
Oats/Wheat	0.195751	0.205736
Pasture/Range	−0.225113	0.039476
Peanuts	0.111403	−0.010232
Rotation	−0.212704	0.318791
Sorghum	−0.526721	−0.022715
Hydrologic Soil Group	$\gamma_{\beta 0,\ HSG}$	$\gamma_{\beta 1,\ HSG}$
B	−0.564179	−0.092124
B and C	0.438210	0.157649
B and D	−0.017252	0.049782
C	0.436505	0.193347
C/D	−0.090889	−0.046704
D	−0.202394	−0.261949
Tillage Methods	$\gamma_{\beta 0,\ Till}$	$\gamma_{\beta 2,\ Till}$
Conservation	−0.344965	−0.1527216
Conventional	0.071645	0.0317183
No-Till	0.260853	0.1154838
Pasture	0.012467	0.0055195
Fertilizer Application Methods	$\gamma_{\beta 0,\ AppM}$	$\gamma_{\beta 1,\ AppM}$
Incorporated	0.46801	0.22160
Injected	−0.59687	−0.28262
Surface Applied	0.39698	0.18797
Other	−0.26811	−0.12695

To understand the application of Equation (4), let us calculate the multilevel value of β_0 (overall mean effect value of +0.144) using the fitted parameters in Tables 2 and 3 for a field planted with corn (−0.157255), using an injection method for the fertilizer application (−0.59687), using conventional tillage (+0.071645), and with a HSG of B (−0.564179). Thus, substituting these values from Tables 2 and 3, we get, for these field conditions:

$$\beta_{0,LU[corn],AppM[injection],Till[conventional],HSG[B]} = 0.144 - 0.157255 - 0.59687 + 0.071645 - 0.564179 = -1.103$$

Based on the same notation, the slope for $\log(x_1)$ (P applied) is written as:

$$\beta_{1,\mathrm{LU[i]},\mathrm{AppM[i]},\mathrm{HSG[i]}} = \mu_{\beta 1} + \gamma_{\beta 1,\mathrm{LU}} + \gamma_{\beta 1,\mathrm{AppM}} + \gamma_{\beta 1,\mathrm{HSG}} + \epsilon \tag{5}$$

and the slope for $\log(x_2)$ (water runoff) is:

$$\beta_{2,\mathrm{Till[i]}} = \mu_{\beta 2} + \gamma_{\beta 1,\mathrm{Till}} + \epsilon \tag{6}$$

For the third regression parameter, we found that the slope of soil loss does not vary. Thus in summary, the effect of β_0 (the intercept of mean annual P loss) varies based on LU, fertilizer application method, tillage, and HSG at each site; the effect of β_1 (the slope of fertilizer applied) varies based on LU, fertilizer application method, and HSG at each site; the effect of β_2 (the slope of water runoff) varies based on tillage; and the effect of β_3 (the slope of soil erosion) is constant. The overall model is summarized graphically in Figure 1, and a comparison of observations with fitted values is shown in Figure 2.

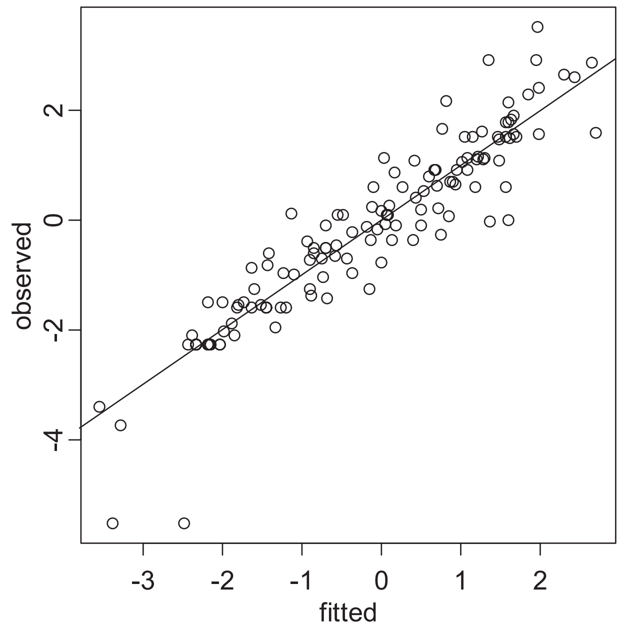

FIGURE 2. Model Estimated Natural Log P Loadings Are Plotted Against the Observed Log P Loadings. The reference line has an intercept of 0 and slope of 1. For a perfect model, all data points should fall on the line.

DISCUSSION

Figures 3-6 illustrate the multilevel categorical effects on the overall regression model that is presented in equation form above; Tables 2 and 3 present the actual parameter values. The LU effects are reflected in both the intercept and slope of P application (Figure 3). As the continuous predictors are all

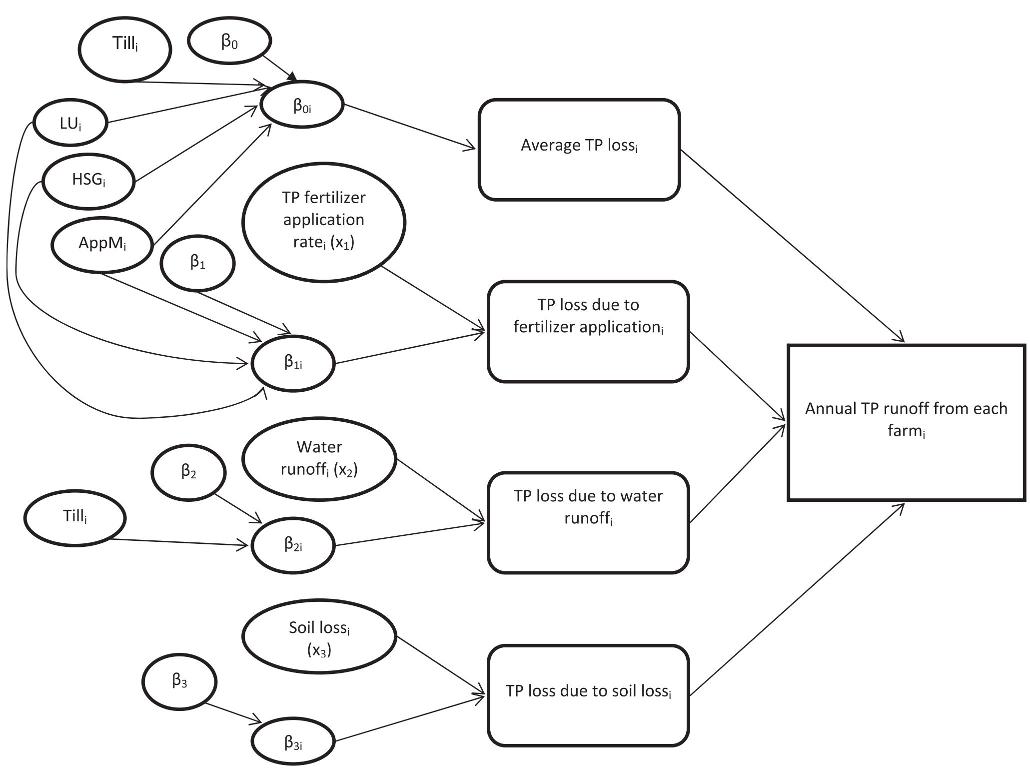

FIGURE 1. A Graphical Representation of the Multilevel Phosphorus Runoff Model.

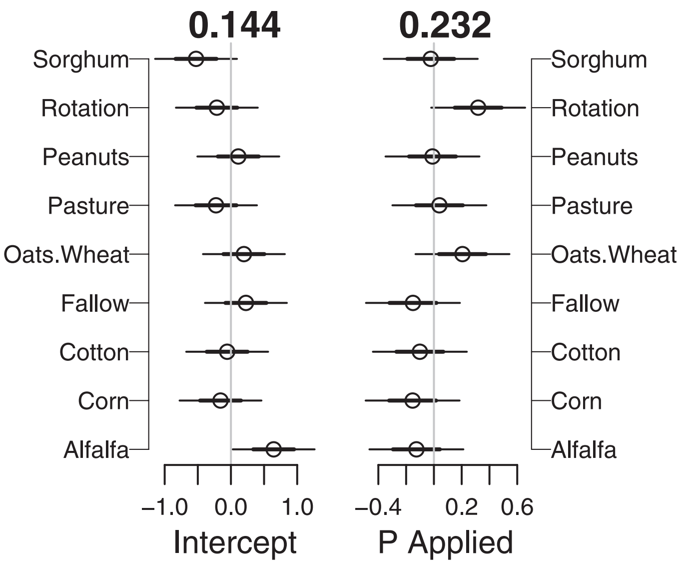

FIGURE 3. The Land Use Type Effects on the Intercept and on the Slope of P Applied. The estimated mean effects are shown as open circles. The thick lines are the estimated mean ± 1 SE, and the long thin lines are the estimated mean ± 2 SE. The numbers on top of the plot are the average intercept and slope. Thus, the intercept for land used for growing alfalfa would have an intercept about 0.6 units above the overall average of 0.144, and a slope for P application about 0.2 units less than the overall average of 0.232. Because the response variable is the log of P loading, the difference reported here is a multiplicative factor. For example, 0.6 units above the average is $e^{0.6}$ (or 1.8) times the overall average (or 80% above the average).

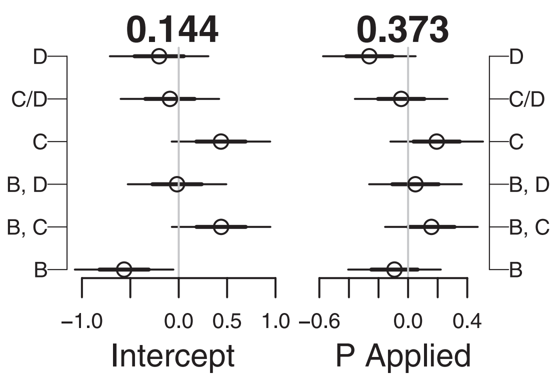

FIGURE 4. The Hydrologic Soil Group Effects on the Intercept and on the Slope of P Application.

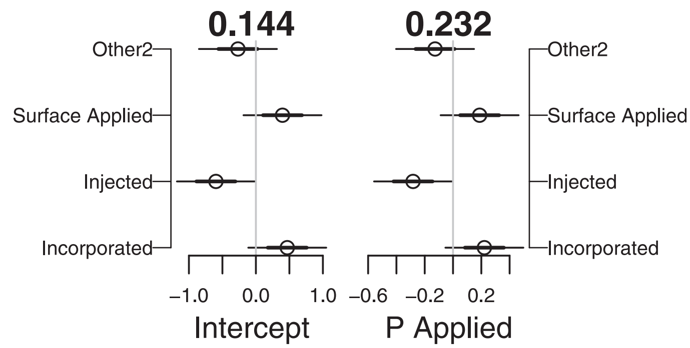

FIGURE 5. The Fertilizer Application Method Effects on the Intercept and on the Slope of P Application.

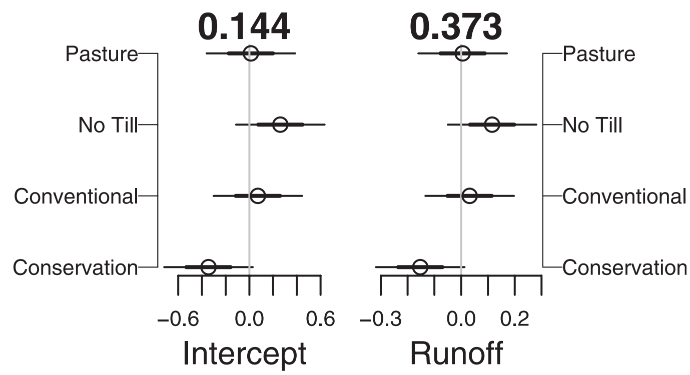

FIGURE 6. The Tillage Method Effects on the Intercept and on the Slope of Runoff.

TABLE 4. Geometric Mean for the Continuous Predictor Variables.

Variable	**Geometric Mean**
Amount of Fertilizer Applied (x_1)	13.812
Water Runoff (x_2)	57.439
Soil Loss (x_3)	679.776

centered at their respective geometric mean (presented in Table 4), the overall average intercept in the model is the estimated log P loading when all continuous predictors are at their geometric mean. The LU effect on the intercept shows the change in log total P loading for different LU, or crop types, when the three continuous variables are at their respective geometric mean. The LU type also changes the slope for fertilizer application. Fallow land and land planted with corn, cotton, and alfalfa have the smallest slopes, and all of these slopes are not significantly different from zero; hence increased fertilizer usage is not directly associated with increased P loss for these LU types. In contrast, the land for various crop rotations has the highest slope (indicating a positive association between fertilizer use and P loss for this LU). HSG affects the model intercept and the fertilizer usage slope. Land with HSG of B has the smallest intercept (low contribution of HSG B to average P loading) followed by group D. HSG of C and B/C are associated with a high intercept (high contribution of these HSGs to P loading) (Figure 4). HSG D has the smallest slope for P application (not different from zero) indicating insignificant direct association between P loading and fertilizer application. Groups B/C and C have the highest slopes, suggesting a positive association between P loading and the amount of P applied in these hydrological soil groups. Fertilizer application methods affect the intercept and the slope of annual P applied (Figure 5). When injection was used, both the intercept and the slope are the smallest, suggesting that the most effective method for controlling fertilizer loss is injection. Tillage methods affect the intercept and the slope of runoff (Figure 6). The tillage effects are somewhat counterintuitive, in that fields using the no-till method are associated with the highest intercept (average P loading) and a large runoff effect. This is attributed to lack of fertilizer incorporation at no-till sites. It should, however, be noted that few no-till fields are included in the database.

Although the estimated model coefficients (Tables 2 and 3) have varying levels of uncertainty (as indicted by the standard error), we retained all three continuous predictors and four categorical variables for two reasons. The first reason is that these predictors are physically meaningful in predicting the P loading from a field. The second reason has practical aspects – about half of the observations were removed from the analysis because of missing values in order to fit the mixed effect regression model displayed in Figures 1-6. If more predictors were included in the final model, we would have a much smaller sample size. The small sample size problem is also the reason why no validation was performed; it is expected that validation will be possible in the future as more data are obtained.

CONCLUSIONS

The flexibility of the multilevel model allows us to advance empirical modeling of agricultural practices on nutrient loading beyond export coefficients (Reckhow *et al.*, 1980) and beyond bivariate relationships such as those presented in Harmel *et al.* (2006). Such models can provide an attractive alternative to daily time step, long-term simulation models, such as erosion/productivity impact calculator (Williams and Sharpley, 1989) in situations where detailed input data are not available.

Using a multilevel modeling approach, we developed a predictive model for P loading from agricultural land using cross-sectional data. Our model suggests that P loss from an agricultural field is a function of the amount of fertilizer applied, annual water runoff, and annual soil loss. Further, the model coefficients vary due to different LUs, HSGs, fertilizer application methods, and tillage methods. These factors all contribute to the slope and intercept factors developed in this multilevel modeling approach.

We believe that this model can serve as either a standalone model to assess the impact of proposed practices or as an empirical augmentation to watershed-scale models, such as SPAtially Referenced Regressions On Watershed attributes (SPARROW) (Smith *et al.*, 1997) or Soil and Water Assessment Tool (SWAT) (Santhi *et al.*, 2005). When applied in conjunction with a watershed model, the model presented in this paper could be applied for agricultural fields incorporating (or proposed to incorporate) one or more of the modeled practices to provide an empirically based estimate of the impact of BMPs and other modeled practices. For example for SPARROW with an "agricultural land" category, the load computed using our model could be substituted for the SPARROW agricultural parameter to assess the watershed impact of the practices being modeled. This has obvious value for estimating the load allocation component of a Total Maximum Daily Load (TMDL). Future multilevel modeling work should involve a similar model for nitrogen, and perhaps multilevel modeling within a Bayesian framework.

ACKNOWLEDGMENTS

We thank Melissa Kenney, Conrad Lamon, and Craig Stow for helpful reviews. Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency through grant R830883 to Kenneth H. Reckhow, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

LITERATURE CITED

Beaulac, M.N. and K.H. Reckhow, 1982. An Examination of Land Use – Nutrient Export Relationships. Water Resources Bulletin 18(6):1013-1024.

Carpenter, S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, and V.H. Smith, 1998. Nonpoint Pollution of Surface Waters With Phosphorus and Nitrogen. Ecological Applications 8:559-568.

Gelman, A. and J. Hill, 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York.

Gitau, M.W., W.J. Gburek, and A.R. Jarrett, 2005. A Tool for Estimating Best Management Practice Effectiveness for Phosphorus Pollution Control. Journal of Soil and Water Conservation 60(1):1-10.

Harmel, D., S. Potter, P. Casebolt, K. Reckhow, C. Green, and R. Haney, 2006. Compilation of Measured Nutrient Load Data for Agricultural Land Uses in the United States. Journal of the American Water Resources Association 42(5):1163-1178.

Harmel, R.D., S.S. Qian, K.H. Reckhow, and P. Casebolt, 2008. The MANAGE Database: Nutrient Load and Site Characteristic Updates and Runoff Concentration Data. Journal of Environmental Quality 37:2403-2406.

Harmel, R.D., H.A. Torbert, B.E. Haggard, R. Haney, and M. Dozier, 2004. Water Quality Impacts of Converting to a Poultry Litter Fertilization Strategy. Journal of Environmental Quality 33:2229-2242.

Laird, N.M. and J.H. Ware, 1982. Random-Effects Models for Longitudinal Data. Biometrics 38:963-974.

Mausbach, M.J. and A.R. Dedrick, 2004. The Length We Go: Measuring Environmental Benefits of Conservation Practices. Journal of Soil and Water Conservation 59(5):96-103.

Pinheiro, J.C. and D.M. Bates, 2000. Mixed-Effects Models in S and S-PLUS. Springer, New York.

Qian, S.S. and Z. Shen, 2007. Ecological Applications of Multilevel Analysis of Variance. Ecology 88:2489-2495.

Reckhow, K.H., 1996. Improved Estimation of Ecological Effects Using an Empirical Bayes Method. Journal of the American Water Resources Association 32(5):929-935.

Reckhow, K.H., M.N. Beaulac, and J.T. Simpson, 1980. Modeling Phosphorus Loading and Lake Response Under Uncertainty: A Manual and Compilation of Export Coefficients. Office of Water Regulations and Standards, U.S. Environmental Protection Agency. EPA400/5-80-011, 212 pp.

Santhi, C., R. Srinivasan, J.G. Arnold, and J.R. Williams, 2005. A Modeling Approach to Evaluate the Impacts of Water Quality Management Plans Implemented in a Watershed in Texas. Environmental Modelling & Software 21:1141-1157.

Sharpley, A.N. (Editor), 2000. Agriculture and Phosphorus Management: The Chesapeake Bay. CRC Press, Boca Raton, Florida.

Sharpley, A.N., R.W. McDowell, J.L. Weld, and P.J.A. Kleinman, 2001. Assessing Site Vulnerability to Phosphorus Loss in an Agricultural Watershed. Journal of Environmental Quality 30:2026-2036.

Smith, R.A., G.E. Schwarz, and R.B. Alexander, 1997. Regional Interpretation of Water-Quality Monitoring Data. Water Resources Research 33:2781-2798.

Wagner, T., M.T. Bremigan, and K.S. Cheruvelil, 2007. A Multilevel Modeling Approach to Assessing Regional and Local Landscape Features for Lake Classification and Assessment of Fish Growth. Environmental Monitoring and Assessment 130:437-454.

Wagner, T., D.B. Hayes, and M.T. Bremigan, 2006. Accounting for Multilevel Data Structures in Fisheries Data Using Mixed Models. Fisheries 31:180-187.

Williams, J.R. and A.N. Sharpley (Editors), 1989. EPIC – Erosion/Productivity Impact Calculator: 1. Model Documentation, USDA Technical Bulletin No. 1768.