A multilevel model of field-scale nitrogen export from agricultural areas

Phillip Jones
Water and Air Resources
Nicholas School of the Environment
Duke University

Faculty Advisors

Song Qian
Ken Reckhow

Environmental Sciences and Policy

INTRODUCTION

Eutrophication is the process by which aquatic ecosystems become enriched with nitrogen (N) and phosphorus (P). Although eutrophication can result from geologic aging, rapid nutrient enrichment of aquatic ecosystems is strongly associated with human activity (Wetzel 2001); this anthropogenic process is sometimes labeled "cultural eutrophication" (Carpenter et al. 1998). Excessive nutrient loading from human activities causes a wide range of undesirable changes in aquatic ecosystems; these changes include anoxia, fish kills, toxic algal blooms, simplification of biotic communities, increased vulnerability to invasive species, and ecological instability (Carpenter et al. 1998; Smith et al. 1999; Wetzel 2001). Surface waters, including wetlands, streams, and lakes, as well as estuarine and marine environment are all vulnerable to the negative impacts of eutrophication (Carpenter et al. 1998; Craft et al. 2007; Smith et al. 1999).

The impacts of eutrophication extend far beyond aquatic ecology. Nutrient levels represent a significant cause of 303(d) surface water impairment listings, with excessive nutrients listed as the fifth most common cause of stream mile impairment and the third most common cause of lake, pond, and reservoir impairment (US EPA 2004). Designated uses of surface waters impacted by eutrophication include: recreational activities, public water supply, shellfish harvesting, and the protection of aquatic life. The economic impacts of eutrophication can also be substantial, with higher costs associated with additional water treatment, mitigation efforts, and the loss of fisheries and wildlife habitat (Carpenter 1997).

Nitrogen is generally considered to be of secondary importance to phosphorus in cases of freshwater eutrophication. On a landscape level, more organically available N is exported to surface waters relative to P, especially when the ratio of watershed size to water body surface area is large (Wetzel 2001); P is therefore considered to be the limiting nutrient. This paradigm has been subject to recent criticism, however, and for surface waters which experience heavy P-loading, additional inputs of N can cause rapid deterioration in environmental conditions, especially from the explosive growth of nitrogen-fixing blue-green algae (Lewis 2008; Paerl 2009; Wetzel 2001). In addition, the multiple pathways of N in the environment make control of nitrogen input difficult. Nitrate-N is carried in runoff and leached to groundwater; ammonia-

N readily adsorbs to soil colloids; nitrogen can volatilize to the atmosphere as NH3 (Carpenter et al. 1998); and dry and wet atmospheric deposition of N can be substantial (Galloway 1995; Howarth 1996; Boyer et al. 2002). Nitrogen is therefore an important control target in efforts to reduce the eutrophication of freshwater environments.

Human activities have had a profound impact on global N availability (Carpenter 1998; Galloway 1995). Anthropogenic activities have more than doubled the input of N in the terrestrial N cycle (Smith et al 1999). The majority of this increase is due to the Haber-Bosch process, by which atmospheric nitrogen is converted to ammonia. Jenkinson estimates that humans fix approximately 160 million tons of N per year; of this total, 98 million tons are derived from the Haber-Bosch process, with 85% of the output used for agriculture (Jenkinson 2001). As these inputs have entered the global N cycle, excess N has become available to many surface water ecosystems. For instance, estimates of the N flux increase for riverine ecosystems range from a 2-20 fold increase when compared to pre-settlement levels (Howarth 1996). Sources of anthropogenic N-loading can be broadly classed as point and non-point. Point sources can be defined as those that discharge pollutants from a single identifiable point or multiple identifiable points; in the United States, this definition includes runoff from industrial sites, mines, CAFOS, and construction sites, as well as discharge from combined sewer overflows and urban storm sewer outfalls (Novotny 2003). Non-point, or diffuse pollution, is characterized by difficulties in monitoring inputs at the point of origin; non-point pollution often accumulates on a landscape scale and is delivered to surface waters intermittently during meteorological events (Novotny 2003). Common sources of non-point pollution include urban and agricultural runoff, fertilizer and manure leaching from farms, lawns, and animal operations, and atmospheric deposition of anthropogenic pollutants.

Attempts to regulate anthropogenic nitrogen inputs have been most successful for point sources (Carpenter et al. 1998). This is due in large part to the effectiveness of control measures and the ease of source identification. In the US, point sources are regulated under the National Pollutant Discharge Elimination System (NPDES) permit program in the Clean Water Act. Efforts to control non-point sources of N have been less successful. Indeed, non-point pollution is now widely recognized as the dominant source of excess nutrients in aquatic ecosystems, with some studies estimating that the total elimination of point sources would be insufficient for most nutrient impaired water bodies to meet water quality criteria (Carpenter 1998; Smith 1999).

Agriculture represents one of the largest sources of non-point source nitrogen pollution in the environment (Boyer et al. 2002; Chambers et al. 2006; Makarewicz et al. 2008). Agriculture is listed as the number one pollution source for impaired streams and the number three source for impaired lakes and reservoirs in the 2004 USEPA's National Water Quality Inventory report (USEPA 2004). Nitrogen application for agriculture includes both organic and inorganic N species. Common organic forms are delivered in manure, urine, composted waste, mulch, and crop stubble, while inorganic forms include anhydrous ammonia-derived urea and ammonium nitrate. The application of nitrogenous fertilizer beyond what is needed for crop production is of special concern (McMahon and Woodside, 1997). The USDA Economic Research Service calculates that the US consumed 13,194,000 metric tons of nitrogenous fertilizers in 2007 (USDA). The fate of applied nitrogen fertilizer is complex and dependent on multiple factors; soil texture, crop uptake rates, application method, tillage practice, hydrologic soil group, slope and groundwater characteristics, meteorological events, and the presence of ditches or tile drains can all be influential. However, research into transport mechanisms indicates that a substantial portion of nitrogen fertilizer is readily lost to the environment in most agricultural fields. Carpenter found that < 18% of N input as fertilizer is removed from farms as produce (Carpenter et al. 1998); N that is not removed as either agricultural product or as part of a compostable waste stream is vulnerable to export. A range of transport mechanisms have been implicated in N export. For instance, estimates of volatilization rates for urea-based fertilizers can reach 15% (Boyer et al. 2002). Additional pathways by which N is lost from agricultural fields include leaching and runoff for dissolved forms and soil erosion for particulate forms (Wetzel 2001).

Field-scale attempts to limit the export of N from agricultural fields involve the implementation of Best Management Practices (BMPs). BMPs seek to limit nutrient loss due to erosion, leaching, runoff, and other pathways. Examples of nutrient based BMPs include soil nutrient testing to determine appropriate fertilizer application rates, the use of buffer areas and small grain cover crops, soil erosion control practices, and the coordination of appropriate fertilizer application rates with irrigation and plant growth (USEPA 2003). Fertilizer management strategies such as soil nutrient testing and coordinated application are designed to limit the amount of nitrogen that is applied to fields, while other BMPs seek to prevent unused fertilizer-N from leaving the agricultural area. The effectiveness of BMPs in limiting nutrient loss is a matter of some debate. Fertilizer management programs have demonstrated that a 20-50 lbs/acre reduction in N fertilizer application can be achieved for many fields with no loss in crop yield (USEPA 2003). BMPs designed to limit N-export via volatilization, soil erosion, and runoff, have been less well studied. Meek et al. found that no-till practices reduced nitrate leaching over other tillage methods (Meek et al. 1995). Makarewicz documented that watershed-level reductions (>30%) of N-containing chemical species can be achieved with the implementation of structural and cultural BMPs (Makarewicz et al. 2008). However, other studies have published conflicting results. For instance, Baker found that conservation tillage practices increase soil infiltration and hence the rate of nitrate leaching (Baker 1993). The complexity of transport mechanisms and possible interactions among factors is one likely explanation for contradictory findings. Given the impacts of excessive nutrient loading on surface water quality, there is a thus a strong need to resolve the debate over BMP effectiveness and characterize how field-scale factors affect nitrogen export to surface waters.

OBJECTIVE

The ability to reliably predict field-scale N export is of great practical concern for water quality and watershed management. Statistical analysis of the problem, however, is complicated by the large number of predictor variables that influence nutrient export from agricultural fields. One promising approach to the characterization of farm-level N export is multilevel or hierarchical statistical modeling. Multilevel modeling is a generalized form of regression that is well suited for analyzing data with a hierarchical structure. A hierarchical data structure is one in which individual sites or studies can also be classed as groups that share certain commonalities. To illustrate this point, imagine a river basin that is divided into several distinct subwatersheds. Water quality samples are taken at the outflow points of each subwatershed. The measurement outcomes at each site occur at an individual level, and predictor variables are likely to include information at both individual (flow in ft^3/s at the outflow point) and group (dominant soil texture) levels. Three general classical regression approaches can be applied to this situation. In the first case, data from all sites can be combined and a single regression fit to characterize all of the subwatersheds. This approach is known as "complete pooling." The statistical assumption here is that all sites share the same regression coefficients.   In the second case, each subwatershed can be included in a single classical regression using group indicators; however, no group-level predictors are allowed due to collinearity and there is no model for group-level coefficients (Gelman and Hill 2008). This approach is labeled "partial pooling." Each subwatershed in a partial-pooling model will have a group indicator (i.e., dummy variable), and the regression model coefficients for each dummy variable will be estimated separately (Reckhow et al. 2009). An additional assumption in the no-pooling approach is that the distribution of residuals is identical for all groups. In the third case, a separate regression model can be fit for each subwatershed. The statistical assumption here is that none of the regression models for individual subwatersheds share any coefficients and that the distribution of residuals across groups is not identical.

Practically speaking, the assumptions behind the complete and partial pooling models are likely to be false. For instance, in the present example it seems reasonable to conclude that the subwatersheds are likely to share some similarities since they occur within the same river basin. At the same time, each subwatershed is likely to contain unique features which differentiate it from the other sites. The sites are not exactly alike, nor completely different. When data from all sites are combined as in complete pooling, the resulting model ignores any variation in the average response variable between subwatersheds; group-level information, or variation, is ignored (Gelman and Hill 2008). Partial pooling models, by contrast, are likely to overstate the variance between different subwatersheds; the models can be seen as over-fitting the data within each subwatershed (Gelman and Hill 2008). Finally, fitting separate regressions for each subwatershed is only possible if sample size is sufficiently large and data collection is complete; for large numbers of group, this approach can be cumbersome.

Multilevel regression circumvents these issues and provides a reasonable framework for utilizing both individual and group-level predictor variables (Gelman and Hill 2008). In essence, the approach allows exploration of factors at the group level that may explain variation in model coefficients (Qian 2010).

A key feature of the multilevel approach is that it retains study-specific group identity while "borrowing strength" from related groups (Reckhow et al. 2009). This concept of "borrowing strength" is expressed in multilevel regression through the use of partial-pooling. Partial pooling represents a balance between the complete pooling and no pooling options described above; the method estimates the degree to which group membership changes the global or overall mean regression coefficients. Thought of in this manner, the overall mean regression coefficient estimates, or fixed-effects, are analogous to the case of complete pooling, where all sites are assumed to share the same coefficient estimates and standard error (Reckhow et al. 2009). The influence of group membership, or random effects, is similar to the no pooling scenario, where coefficients vary by group membership. The balance between the overall or global average and group-specific estimates is achieved via a weighted average that considers the amount of information contained in each group-level variable. This weighted average is based on sample size, within-group variance, and between group variance (Reckhow et al. 2009). For example, a group that has a large sample size, small within-group variance, and large between-group variance, contains more information than a group with the opposite characteristics. The estimate for a group-level predictor variable that contains less information will be closer to the overall mean coefficient than the estimate for a group-level variable that contains more information; in other words, all group-level estimates are shrunk against the global average, and the degree of movement towards the global mean depends on the amount of information contained in the group as defined by sample size, within-group variance, and between-group variance. This key feature of multilevel regression is known as a "shrinkage" effect (Reckhow et al. 2009). The shrinkage estimator also allows us to obtain a more reasonable estimate of model coefficients for sites with small sample sizes; in general, the partial-pooling estimated group means will be less variable than those obtained from no-pooling estimates (Qian 2009).

For many environmental and ecological problems, both individual and group level predictors are likely to be important. However, large spatial or temporal variation between individual and group-level predictors can be problematic. In statistics, the spatial/temporal variation between individual and group-level predictors for a given response variable is termed cross-scale interactions. Cross-scale interactions occur when outcomes measured at one scale are affected by predictors at that scale and also by predictors working at different spatial/temporal scales (Qian 2010). For instance, nitrogen export from a field may depend on both individual predictors such as runoff or soil loss and group-level predictors such as hydrologic soil group that operate on a larger spatial scale and, more importantly, do not vary on a field-scale level. If a predictor does not vary within an individual site, that predictor does not provide useful information regarding the response variable. Multilevel regression accounts for large spatial/temporal variation by preserving information contained in group membership. In the case of nitrogen export from individual agricultural fields, two collections of farms in different geographic regions may have the same hydrologic soil group, dominant soil texture, and tillage method, but different crop type, fertilizer application rate, fertilizer application method, and soil loss rate. Since soil characteristics and tillage method do not vary across any of the sites, these categorical variables will not be useful predictors for nitrogen export. However, differences in other larger spatial scale variables such as crop type and fertilizer application method can be used to divide the farms into distinct groups. When many different fields are included in the analysis, multiple combinations of group membership become possible; the impact of variables that operate over large spatial scales can then be investigated. The ‘random effects’ component of the multilevel regression output can therefore be interpreted as group interactions; group membership interacts with ‘main level’ predictors to change coefficient estimates. As mentioned earlier, the influence of these group-level variables is estimated with partial pooling – for instance, the within-group variance of a certain hydrologic soil group, the number of samples, and how much this hydrologic soil group varies from other hydrologic soil groups. Multilevel regression thus provides powerful tools to investigate group membership even when these predictors vary over large spatial scales.

The main objective of this paper, then, is to characterize the problem of farm-level nitrogen loading through the application of a multilevel statistical model. A successful model should be useful in i) identifying the effectiveness of various agricultural BMPs on nitrogen export and ii) predicting farm-level nitrogen runoff. The model will therefore address the problem of farm-level nitrogen loading by serving as a tool to assess the impacts of proposed management changes in agricultural areas. For instance, specific group-level information such as conservation practice, soil texture, tillage method, etc., can be used to predict field-scale nitrogen export. In sensitive watersheds, this model could be used to guide BMP implementation in an effort to reduce nitrogen export for a given tillage method, hydrologic soil group, etc.

A secondary objective is to contribute to the development of multilevel modeling applications in environmental science. As Reckhow et al. note, the use of hierarchical data structures in environmental modeling is relatively common; however, few researchers explicitly use multilevel techniques in their statistical models. Studies involving cross-scale interactions are also common. In the case of nutrient export modeling, multilevel techniques allow empirical models to advance beyond export coefficients (Reckhow et al. 2009). A wider application of the technique seems justified.

METHODS

Mathematical/Statistical characterization of multilevel regression

A multilevel model can be written in at least five different ways (Gelman and Hill 2008). One of the more straightforward conceptual presentations involves thinking of multilevel regression as a linear regression in which coefficients vary across groups. Group membership can influence intercept, slope, or both intercept and slope depending on how the model is specified. In the case of a linear regression with a single predictor variable in which both the intercept and slope vary by group, the response variable distribution can be represented by (1), or by a classical linear regression equation (2); both intercept and slope in this example are indexed by the *j*th group.

$$y_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]}X_i,\ \sigma_y^2\right), \text{ for } i = 1, \ldots, n \qquad (1)$$

$$y_i = \alpha_{j[i]} + \beta_{j[i]}X_i + \varepsilon_i \tag{2}$$

The multilevel aspect of this model comes from the assignment of a multivariate normal distribution to the parameters within each group (3) (Gelman and Hill 2008). Note that in this example there are no additional predictors that operate within a given group (for instance, if the intercept estimate within a group were allowed to vary by a predictor variable γ, the μ_α estimate shown in (3) would be a function of an additional regression: $\gamma_0 + \gamma_1\mu_j$). Note also the between-group correlation parameter ρ; a multivariate normal distribution is used because regression coefficients are usually correlated (Qian 2009).

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right), \text{ for } j = 1, \ldots, J \tag{3}$$

In addition to the normality assumption, multilevel regression also assumes exchangeability. Individual units within a type are exchangeable as long as there is no information other than observed data to distinguish one unit from any other unit within a type (Qian 2009). For example, if a researcher knows that fields (units) planted in corn (a type of land use) always export more nitrogen than fields planted in soybean, then fields within land use type are not exchangeable; in other words, the researcher knows that α_j and β_j will be different for fields planted in corn versus fields planted in soybean. A key point with exchangeability is whether the researcher knows how to represent information that might distinguish units within a type in a statistical model. Using the previous example, a researcher may know that corn has a relatively low ability to take up N from fertilizer application relative to other crops. Fields in corn may then be expected to export more N via soil loss than other crops and the exchangeability assumption appears to be violated. However, soybeans fix nitrogen in the soil, so perhaps soil erosion from soybean fields contains as much or even more particulate N as erosion from corn fields. Unless information on type has a clear implication for the response variable estimate for a unit within that type, then the exchangeability assumption is likely to be met.

The use of a common a priori distribution, or hyperdistribution, for multilevel regression coefficients leads directly to the shrinkage effect observed in multilevel models. As mentioned above, the shrinkage effect acts to pull group estimates closer to the overall mean. This shrinkage is the result of partial-pooling. Partial-pooling can be thought of as a mathematical means of reconciling information contained in complete-pooling and no-pooling estimates (Qian 2009). For a simple model in which only the intercept varies by group membership (4), and where the group-specific intercept estimates come from a common distribution and are exchangeable (5), the partial-pooling estimate of α_j is given by equation (6):

$$y_i = \alpha_{j[i]} + \beta X_i + \varepsilon_i \tag{4}$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \tag{5}$$

$$estimate\ of\ \alpha_j \approx \frac{\frac{n_j}{\sigma_y^2}}{\frac{n_j}{\sigma_y^2}+\frac{1}{\sigma_\alpha^2}}\left(\bar{y}_j - \beta\bar{x}_j\right) + \frac{\frac{1}{\sigma_\alpha^2}}{\frac{n_j}{\sigma_y^2}+\frac{1}{\sigma_\alpha^2}}\mu_\alpha \tag{6}$$

Mathematically, the partial-pooling estimate is a weighted average of the no-pooling estimate $\left(\bar{y}_j - \beta\bar{x}_j\right)$ and the complete-pooling estimate (μ_α); the weights are determined by group sample size (n_j), between group variance (σ_α^2), and within group variance (σ_y^2). As Qian notes, the multilevel estimate of the overall or global mean is closer to the group mean when the group sample size is large, or the between group variance is large, or the within group variance is small. If all three conditions hold, we would trust the group mean as a reliable estimate of a group-level, or random effect (Qian 2009). In other words, the amount of information contained in the group is high. The balance between complete-pooling and no-pooling can be seen by the fact that Equation 5 contains both approaches as special cases. For instance, Equation 5 collapses to the no-pooling estimate when either between-group variance (σ_α^2) or group sample size (n_j) is set to infinity (∞). When a group contains no data ($n_j = 0$), Equation 5 is the same as the complete-pooling estimate. As Gelman and Hill note, the same approach holds when both intercept and slope are allowed to vary by group (Gelman and Hill 2007).

In general, multilevel regression is preferred for studies which seek to estimate regression parameters for observational datasets containing a large number of groups with unbalanced sample sizes (Qian 2009). In such cases, multilevel regression offers predictable benefits over classical multiple linear regression. Multilevel models are often superior to classical regression in predictive accuracy; multilevel models may also hold advantages in data reduction and causal inference (Gelman 2006). For data obtained from randomized controlled samples, multilevel regression is unlikely to produce results that differ greatly from classical techniques (Qian 2009). However, since classical approaches are included as special cases in the multilevel computation, not much is lost by the application of multilevel regression in these instances other than the additional time that may be required to formulate and interpret the model.

Data Description

Data on farm-level nitrogen loading were obtained from the updated Measured Annual Nutrient loads from Agricultural Environments (MANAGE September 2009) database. MANAGE is a compilation of data from >55 peer-reviewed studies of field-scale N and/or P export (Harmel et al. 2008). The database contains information on field-level variables as well as larger scale site characteristics such as watershed size, land slope, and soil type (Table 1). The compilation is limited to field-scale studies of single land-use areas of cultivated agriculture or pasture/range/hay with contributing area > 0.009 hectares; modeled/simulated nutrient loading estimates are not included. In total, MANAGE contains over 1600 watershed-years of data from a wide spatial distribution of agricultural fields. However, it is important to note that the database contains limited or no information on farms within certain regions of the US, including the Northeast and West.

BMP information for nitrogen use is captured in the database through five variables: tillage, conservation practice, fertilizer application method, land use, and the amount of N fertilizer applied. Tillage practice is divided into four options: conventional, conservation, no-till, and pasture. Conservation tillage represents methods designed to leave crop residue on the field after harvest (Harmel et al. 2006); no-till is the least intrusive of commonly used tillage practices. The conservation practice variable lists any nutrient management strategies used in the study watershed. Options include waterway, terrace, filter strip, riparian buffer, and contour farming. Fertilizer application categories include surface application, incorporation, and injection.

Software

The models described in this paper were fit using a restricted (or residual) maximum likelihood estimator (REML) for linear mixed-effects models implemented in the R package 'lme4,' authored by Douglas Bates and Martin Maechler. In general, REML estimates of variance components are regarded as less biased than full maximum likelihood (ML) estimates. However, the maximum likelihood algorithm is not efficient in estimating variances when the number of groups is small (Qian 2009). The algorithm also does not handle missing values properly, which can be problematic for some datasets (Reckhow et al. 2009).

Model Fitting

Log transformation of the field-scale nitrogen export data resulted in an approximately normal response variable. All predictor variables except one were also log-transformed to stabilize residual variance. The continuous predictor variables were then centered. As Gelman notes, centering continuous predictors may help remove correlations between intercepts and slopes in multilevel regressions (Gelman and Hill 2007). Centering variables also aids in interpreting model coefficients. Conditional plots were used to check for potential interactions among possible predictor variables.

The task of predictor variable selection in multilevel modeling is largely subjective. Scatterplot matrixes were examined for clear relationships between predictors and the response. Two data mining techniques, Classification and Regression Tree (CART) and Multivariate Adaptive Regressive Splines (MARS), were also employed to identify likely continuous predictors as well as potential categorical group variables. Data mining resulted in the common identification of two continuous predictor variables and two categorical variables. Additional variable selection was guided by the minimized profile deviance of the model (negative twice the log-likelihood), group-level variance reduction, and physical interpretation.

RESULTS

Model Selection

Exploratory data analysis led to the identification of three important continuous predictor variables: soil loss, water runoff, and N fertilizer application rate. However, multilevel regressions fit to total nitrogen as the response variable exhibited significant variations in coefficient estimation. These models exhibited adequate fit to the data, but often lacked a valid physical interpretation. In addition, BMP information for total nitrogen produced little change in the overall model coefficients. For these reasons, the original dataset was divided into particulate and dissolved forms. Where recorded, particulate N is listed in the MANAGE database as Total Kjeldahl Nitrogen (TKN), organic N, or TKN – NH4-N. Dissolved N is recorded as NO3 + NH4-N, NO3-N, NO2-N + NO3-N, NO2-N + NO3-N + NH4-N, or NH4-N. However, the exact N-species of particulate and dissolved forms is undocumented for many observation points.

Particulate Nitrogen

The final selected model for particulate nitrogen is shown below:

$$Log(PN_i) = \alpha_{SoilT[i],FertM[i],CPS[i]} + \beta_{1,CPS[i]} \log(X_{1i}) + \beta_2 \log(X_{2i}) + \epsilon_i \quad (7)$$

where PN_i is the annual particulate nitrogen load (kg/ha-yr) from field site (i), X_1 is the annual soil loss (kg/ha-yr), X_2 is the annual amount of water runoff (mm/yr), α is the intercept, the β's represent the slopes, and ϵ_i is the regression error term. Since particulate N is sediment attached, the inclusion of soil loss as a predictor is reasonable. The significance of runoff is likely due to the role of water in sediment transport. Soil loss is associated with higher rates of particulate N export than water runoff (Table 2); all three model coefficients are significant at a 95% confidence interval. The overall fit of the model is reasonable, and the distribution of residuals is approximately normal, symmetric, and centered at a value of zero (Figure 1).

The overall mean effects for particulate N are influenced by several group-level categorical variables in this model. Intercept values vary by soil texture (SoilT), fertilizer application method (FertM), and conservation practice (CPS). The slope of annual soil loss also varies by conservation practice. Conceptually, the model can be envisioned as a flowchart in which the coefficient estimates for the intercept and the slope of soil loss vary by the specific group memberships of an individual study site (Figure 2). Estimates for these "random effects" are provided as part of the model summary in R (Table 3). These estimates can be combined with the overall mean estimates to calculate predictions for a specific field. For example, to calculate the intercept for a loam field with no conservation practices and injected fertilizer, we take the overall intercept estimate and add/subtract the random effects estimates for each group-level variable: α + (γα, CPS [0]) + (γα, SoilT [loam]) + (γα, FertApp [Injected]), or 1.01425 + 0.255312 + 1.147798 + (–0.037030) = 2.38033. Since the continuous predictor variables in this model have been log-transformed and centered, this number estimates $Log(PN_i)$ for our case study when annual soil loss and annual water runoff are at their respective geometric means.

Dissolved Nitrogen

The final selected model for dissolved N involves individual and group level predictors that were not included in the model for particulate N. These differences are likely due to the distinct fate and transport characteristics of particulate versus dissolved forms of N. The model for dissolved N can be represented as:

$$Log(DN_i) = \alpha_{Till[i],SoilT[i],CPS[i]} + \beta_{1,SoilT[i]} \log(X_{1i}) + \beta_{2,CPS[i]} \log(X_{2i}) + \epsilon_i \quad (8)$$

where DN_i is the annual dissolved nitrogen load from field site (i), X_1 is the annual amount of water runoff (mm/yr), X_2 is the nitrogen fertilizer application rate (kg/ha-yr),), α is the intercept, the β's represent the slopes, and ϵ_i is the regression error term. Water runoff is the likely transport mechanism for dissolved nitrogen export from agricultural fields; the high coefficient estimate for this predictor variable indicates that runoff is strongly associated with the level of dissolved N export (Table 3). The rate of nitrogen fertilizer application is a likely source of dissolved N in runoff, especially for surface applications of fertilizer. Both main level predictor variables are statistically significant at a 95% confidence interval (Table 3). The fitted versus observed plot for dissolved N indicates that there is a greater amount of variation associated with dissolved N that is not explained by information contained in the MANAGE database (Figure 3); model residuals are approximately normal, symmetric, and centered at zero (Figure 3).

As in the previous model, the overall mean effects for dissolved N are influenced by several group-level categorical variables. The intercept value varies by soil texture (SoilT), tillage practice (Till), and conservation practice (CPS). The slope of annual water runoff varies by soil texture (SoilT), and the slope of fertilizer application rate varies by conservation practice (CPS). This model formulation highlights the flexibility of multilevel regression; intercept, slope and intercept, or slope alone can be allowed to vary by any given factor variable (Figure 4). Coefficient estimates for an individual field site can be calculated by using a table of random-effects estimates (Table 5). For example, to calculate the runoff slope estimate for a field with a clay soil texture, we take the overall mean effect estimate for water runoff and add/subtract the random effect estimate for soil texture: β_1 + ($\gamma\beta1$, SoilT [Clay]), or 1.01722 + 0.576821 = 1.594041. Since both the response and predictor variables in this model have been log-transformed, the value of 1.59 predicts that a one unit increase in annual runoff will result in a 159% increase in dissolved N export for our case study field.

DISCUSSION

Particulate Nitrogen

The impact of random-effects on the overall model coefficients can be easily evaluated using the ‘plot.lmer.ranef’ function written by Professor Song Qian for the R package ‘lme4’ (Qian). This function produces graphs which allow the user to evaluate a) whether the impact of a particular level within a categorical variable is significantly different from zero and b) whether any two levels within a categorical variable are significantly different from one another. As an example, consider the impact of soil texture on the intercept of the particulate N model (Figure 5). The x-axis label lists the overall model coefficient that has been allowed to vary; the numeric value for the coefficient appears at the top of the graph. Because each continuous predictor variable has been log-transformed and centered, the intercept value represents the amount of particulate N loading when annual soil loss and annual water runoff are at their respective geometric means. The impact of each level within the categorical variable is represented by a circle; the range of each point, when read from the x-axis, corresponds to the estimated group-level effect (Table 3). Values to the right of zero represent an increase in the model coefficient, while negative values to the left of zero represent a decrease in the model coefficient; in the case of the intercept, these positions can be interpreted as above or below average, respectively. Each circle is bisected by a line; the thicker portion of the line represents one standard error; the thinner portion of the line represents two standard errors. Given the well-known 68-95-99.7 rule for normal distributions, two standard errors represent 95 percent of the total distribution; if this spread does not bisect 0 on the x-axis, the effect can be seen as being statistically significant at a 95% confidence interval. Similarly, if the 2σ lines for two levels within a categorical variable do not overlap, then these two levels can be interpreted as being statistically different from one another at a 95% confidence interval. With this explanation in mind, the impact of soil texture on average particulate N loading is clear: loam soils are associated with above average particulate N export, while clay soils are associated with below average particulate N export (Figure 5). There are several possible explanations for this result. Loam soils are preferred for fertilizer-based agriculture; we would expect agricultural soils to contain higher levels of particulate N due to the application of ammonia in fertilizer, which readily adsorbs to soil colloids, as well as increased organic matter from crop residue and even the possible effects of N-fixing crops. In addition, loam soils in agricultural areas may be more prone to erosion due to tillage practices that leave the soil surface exposed to wind and rain. Finally, the physical properties of the soils themselves may also play a role.

The impact of fertilizer application method on particulate N shows a clear trend. However, due to large standard errors, the results are not considered to be statistically significant. The impact of application method is likely due both to N-specific biogeochemical cycling and the physical process of the application method itself. For instance, incorporated application methods, which mix fertilizer with the upper layer of the soil, show a trend towards higher particulate N loading. This is not surprising, since incorporation loosens the soil surface and leaves sediment-attached ammonia vulnerable to erosion. By contrast, injection methods generally deposit N at a depth of 6-10cm below the soil surface; this application method results in sediment-attached ammonia that is less vulnerable to surficial erosion. Finally, due to the relatively fast conversion of ammonia-N to nitrate-N in oxidized environments, surface applications of nitrogenous fertilizers may be more associated with dissolved forms of N and water runoff than sediment-attached export.

The conservation practice variable impacts both the overall intercept and the slope of annual soil loss in the particulate N model. In the MANAGE database, conservation practice is comprised of three columns to account for agricultural fields that use up to three different practices. The variable itself is divided into five levels: waterway, terrace, filter strip, riparian buffer, and contour farming. In order to analyze the impact of more than one conservation practice, this categorical variable was converted to a numeric value (0 for no conservation practice, 1 if a conservation practice was present) and summed. This transformation produced a conservation practice score for each field site that ranged from 0 to 3. The resulting sample sizes, however, were very low, especially for values of 2 and 3. As a result, conservation practice was then converted to a binary variable with 0 denoting the absence of any conservation practice and 1 indicating the use of one or more conservation practices in the field. The model results indicate that fields with no conservation practice in place are associated with higher average particulate N loads than fields that have one or more such practices in place (Figure 6; Table 3). This result is statistically significant; the x-axis indicates that the magnitude of the difference is +/-0.255312, or 25%, respectively. The impact of conservation practice on the slope of annual soil loss shows the opposite trend. Fields with one or more conservation practices in place export slightly more particulate N per unit increase of soil loss than fields with no conservation practices (Figure 6; Table 3). One possible explanation for this trend is that fields with conservation practices use incorporated application methods more often than fields with no conservation practices (Table 6); this application method is associated with higher N particulate loading (Figure 5). Fields with no conservation practices are also more likely to use surface application methods, which may result in faster oxidation of sediment attached ammonia to nitrate (Table 6). In addition, some combination of conservation practices may function to keep more N in the soil and/or minimize soil loss. However, when these soils are eroded, they may contribute more to particulate N loading than fields in which nitrogen is primarily lost as nitrate.

Dissolved Nitrogen

Soil texture exhibits a clear influence on the average export of dissolved nitrogen (Figure 7). Soils with a loam component are associated with higher dissolved nitrogen loads; this may be explained by the fact that loam soils undergo higher rates of agriculture-based nitrogen fertilization. Clay soils, which are unlikely to experience much fertilizer application, and sandy soil textures, which drain quickly and produce little runoff, are associated with average to below average dissolved nitrogen loading (Figure 7). Any impact of tillage method on average dissolved N export is obscured by high standard errors (Figure 7).

The complexity of processes influencing field-scale N export is shown by the impact of conservation practice on dissolved N loading. In contrast to the model for particulate N, agricultural fields using one or more conservation practices are associated with slightly higher dissolved N loading when annual runoff and the rate of N fertilizer application are at their geometric means (Figure 8). The magnitude of the impact is smaller than the reduction in particulate N loading seen in the previous model (Table 3; Table 5). The use of one or more conservation practices is also associated with higher dissolved N loading as the rate of fertilizer application increases (Figure 8); the magnitude of this impact, while statistically significant, is extremely small (Table 5). The impact of conservation practice is therefore dependent on the form of N that is exported from the field.

Biogeochemical Complexity

The N-form specific effect of conservation practice on N loading highlights the biogeochemical complexity of nitrogen in the environment. This form specific result, however, is not surprising when one considers the properties of common N-species. Ammonium-N is a reduced form of nitrogen that is readily adsorbed to soil particles; this species is therefore highly associated with sediment-attached N loading. Over time, ammonium-N is oxidized by bacteria to form nitrate, which is highly soluble. Nitrate-N is highly associated with dissolved N loading. The use of separate multilevel regressions for these two forms therefore has a clear justification in the chemical and physical properties of nitrogen.

At the same time, the relative proportion of particulate and dissolved forms in agricultural fields is not static. For instance, the microbially-mediated rate of conversion from ammonium to nitrate is dependent on temperature, soil carbon levels, and many additional factors. N-fixation can occur with common crops in the family *Fabaceae*, such as soybean, clover, alfalfa, and peanuts. Organic N in the form of manure or urea includes both particulate and soluble forms of nitrogen; organic-N can undergo both mineralization and hydrolysis to form ammonium. This ammonium, if not taken up by vegetation, can result in high rates of ammonia volatilization. The potential for cycling between various forms of N is therefore large. Characterization of nitrogen loading is also complicated by temporal factors that are not captured in the MANAGE database. For example, immediately after fertilization, much of the exportable nitrogen will be in a reduced, sediment-attached form. Over time, this nitrogen will be oxidized to nitrate. The relative impact of dissolved and sediment-attached nitrogen loading will thus depend in part on the length of time between when samples are taken and when nitrogenous fertilizer was last applied to the field. These biogeochemical complexities help explain why regression models for total nitrogen often produce inconclusive results; management practices effective for one form of nitrogen loss may not be effective for other forms. In other words, the sum BMP effect may cancel out. The behavior of N in the environment may also explain why the rate of fertilizer application was not included in the regression model for particulate N export; the relatively rapid conversion of ammonia to nitrate may obscure any relationship between fertilizer application and particulate N loading (and instead highlight other factors such as soil texture). These spatial and temporal complexities highlight the need for a statistical tool like multilevel regression. In addition to providing a more reasonable estimate of group-level effects via partial pooling, multilevel regression is also well suited for handling many of the cross-scale interactions mentioned above (climate, soil carbon, etc.). Indeed, the results for both particulate and dissolved nitrogen show that cross-scale interactions are important in characterizing N loading. In both models, individual level predictors are significantly influenced by variables that occur at larger spatial scales; these variables are also unlikely to vary across the spatial extent of a single farm. The inclusion of additional predictors associated with larger spatial and temporal scales is therefore likely to increase the predictive power of nitrogen loading models.

Management Applications and General Recommendations

Despite the conclusion that BMP impact is likely to vary by nitrogen species, several general management recommendations can be reached. Agricultural management should focus on identifying whether soil erosion or water runoff is more important for a given field. This information, together with data on fertilizer schedules and local climate data, will help identify whether dissolved or sediment-attached nitrogen loading is of more concern. Of course, it is possible that both forms of loading may be significant for a given field, especially for fields located within sensitive watersheds. A further recommendation is based on the fact that annual water runoff is a significant predictor for both dissolved and particulate nitrogen loading. Management practices that address field-scale water runoff, such as contour farming, should be implemented where appropriate since runoff is a transport mechanism for both particulate and dissolved nitrogen. A more general recommendation is soil-N testing. Soil nitrogen testing provides an estimate of how much plant-available nitrogen is available in the soil; these tests, which are often performed at low cost by state agencies and universities, enable the farmer to moderate N application to meet guidelines based on soil-series-specific yield expectations (US EPA 2003).

The models developed in this analysis also have several applications to management decisions. The models can be used to predict field-scale export of both particulate and dissolved nitrogen using individual and group-level predictors. These load estimates could then be used as part of an assessment process such as a TMDL study or as the front-end of various watershed-level water quality models. The models could also help decision makers evaluate the likely impact of best management practice implementation in specific watersheds. A key benefit of multilevel regression in this regard is that models focused on specific management practices can be easily formed and evaluated.

The most pressing further research need is for randomized, controlled experiments with sample sizes sufficient to identify the impacts of individual conservation practices on specific nitrogen forms. Development of temporal data that could be used to characterize the rate of nitrogen cycling between reduced and oxidized forms in agricultural areas would also be useful.

TABLES AND FIGURES

Variable Name	Description
Watershed ID	Watershed name.
Location (city, state)	Study location.
Date	Beginning and end period of annual nutrient measurement.
Watershed Years (ws yr)	Number of watersheds * number of years of data.
Land Use	Crop or vegetation type(s).
Tillage	Four options: No-till, conservation, conventional, and pasture.
Conservation Practice	Five options: waterway, terrace, filter strip, riparian buffer, or contour farming.
Dominant Soil Type	Soil textural class.
Hydrologic Soil Group	NRCS hydrologic soil group classification: A, B, C, or D.
Land Slope (%)	Maximum and minimum land surface slopes for studies with multiple watersheds.
Watershed Size (ha)	Maximum and minimum sizes for studies with multiple watersheds.
Fertilizer Formulation	Nutrient composition of fertilizers used on field.
Fertilizer Application Method	Four options: surface, injected, incorporated, or other.
N applied (kg/ha-yr)	Maximum, minimum, and average annual values.
Precipitation (mm/yr)	Maximum, minimum, and average annual values.
Runoff (mm/yr)	Maximum, minimum, and average annual values.
Soil Loss (kg/ha-yr)	Maximum, minimum, and average annual values.
Form Dissolved N	Chemical species measured.
Dissolved N (kg/ha-yr)	Total amount of dissolved N lost from WS; maximum, minimum, and average annual values.
Form Particulate N	Chemical species measured.
Particulate N (kg/ha-yr)	Total amount of particulate N lost from WS; maximum, minimum, and average annual values.
Form Total N	Chemical species used to calculate Total N.
Total N (kg/ha-yr)	Maximum, minimum, and average annual values; if not specified, calculated from particulate and dissolved values.

Table 1: Selected variables in the MANAGE database.

Coefficient	**α**	**β_1**	**β_2**
Variable Name	Intercept	Soil Loss	Water Runoff
Estimate	1.01425	0.63630	0.18430
Standard Error	0.47643	0.08717	0.07364

Table 2: Estimated overall mean effects for particulate N loading model.

Conservation Practice	**γα, CPS**	**γβ1, CPS**
0 (no conservation practices)	0.255312	–0.117926
1 (one or more conservation practices)	–0.255312	0.117926
Soil Texture	**γα, SoilT**	–
Clay	–0.597145	–
Loam	1.147798	–
Sandy Loam	–0.552598	–
Silt Loam	0.001944	–
Fertilizer Application Method	**γα, FertApp**	–
Incorporated	0.156467	–
Injected	–0.037030	–
Surface Applied	–0.119437	–

Table 3: Estimated group-level effects for particulate N loading model.

Coefficient	α	β_1	β_2
Variable Name	Intercept	Runoff	App. Rate
Estimate	0.04467	1.01722	0.25442
Standard Error	0.31698	0.21710	0.09680

Table 4: Estimated overall mean effects for dissolved N loading model.

Soil Texture	γα, SoilT	γβ1, SoilT
Clay	–0.680435	0.576821
Clay Loam	0.478397	–0.405548
Loam	0.337201	–0.285854
Loamy Sand	–0.462006	0.391653
Sandy Loam	–0.129593	0.109859
Silt Loam	0.363475	–0.308126
Silty Clay	0.092961	–0.078805
Conservation Practice	**γα, CPS**	**γβ2, CPS**
0 (no conservation practices)	–0.155780	–0.088702
1 (one or more conservation practices)	0.155780	0.088702
Tillage	**γα, Till**	–
Conservation	0.078750	–
Conventional	–0.058330	–
No Till	0.188632	–
Pasture	–0.209052	–

Table 5: Estimated group-level effects for dissolved N loading model.

CPS	Fertilizer Application Method			
	Incorporated	**Injected**	**Other**	**Surface Applied**
0	37 (33%)	18 (16%)	3 (2%)	53 (47%)
1	16 (45%)	6 (17%)	0	13 (37%)

Table 6: Fertilizer application method as a function of conservation practice; percent of total sample size is listed in parentheses.

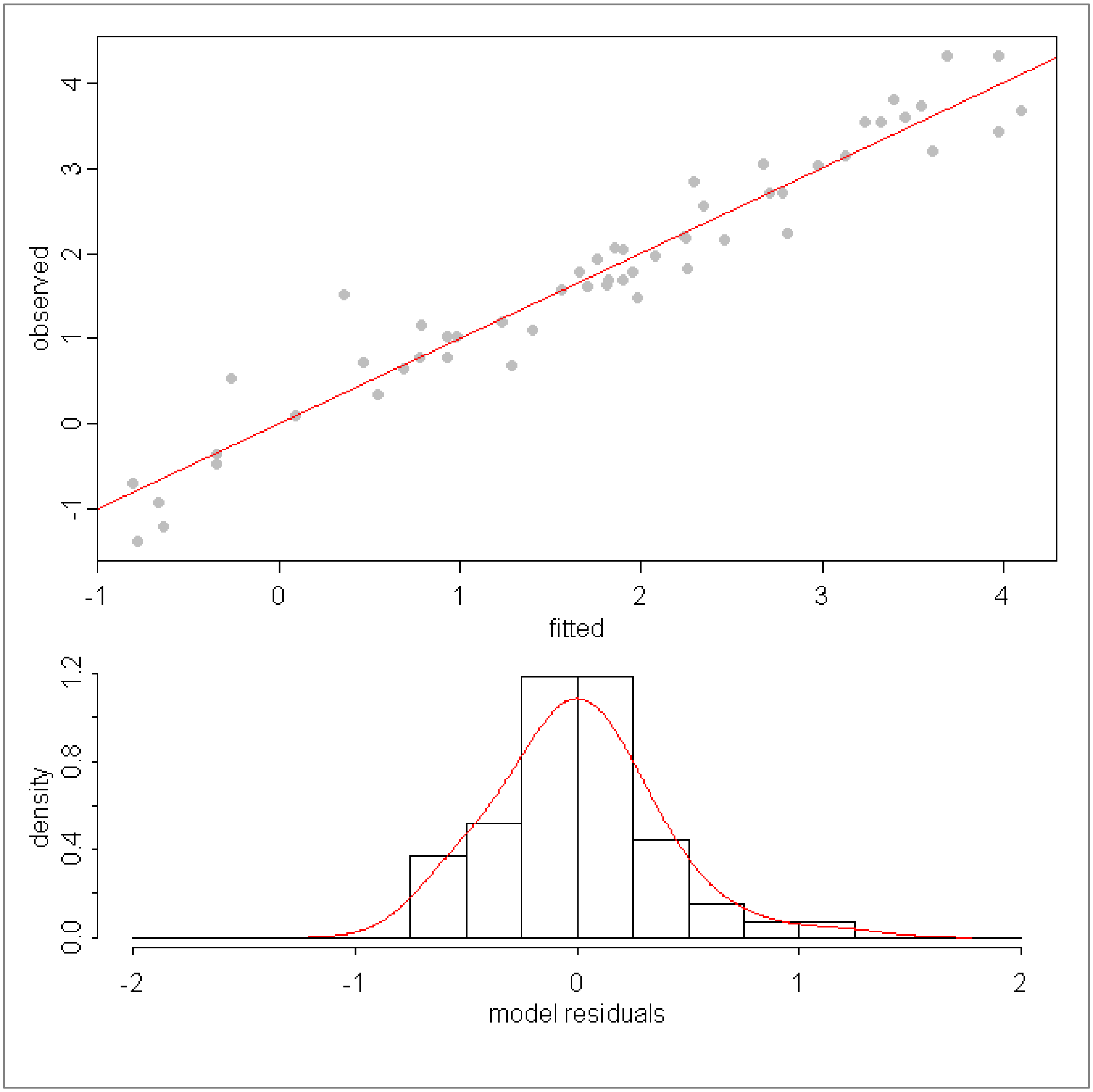

Figure 1: Fitted vs. observed plot and residual distribution density for the particulate N model. In a perfect model, each point in the fitted vs. observed plot would fall along the linear red reference line. The red line in the model residual graph represents an approximation of residual density.

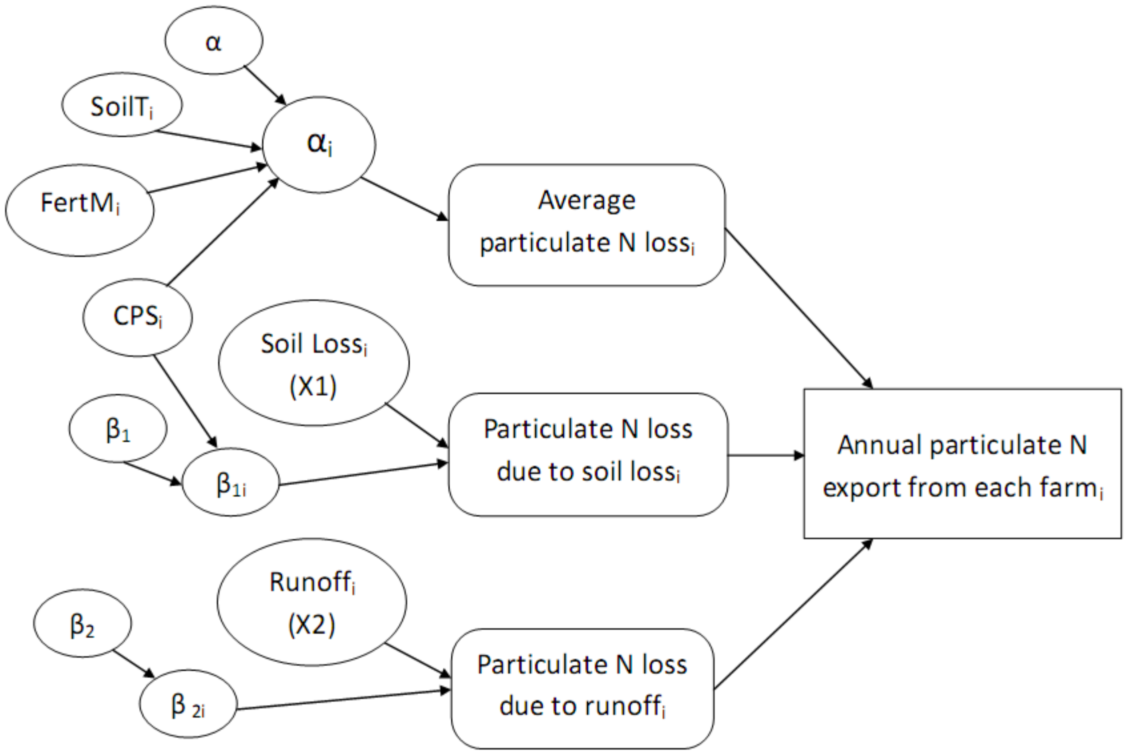

Figure 2: Flowchart of the final particulate N model.

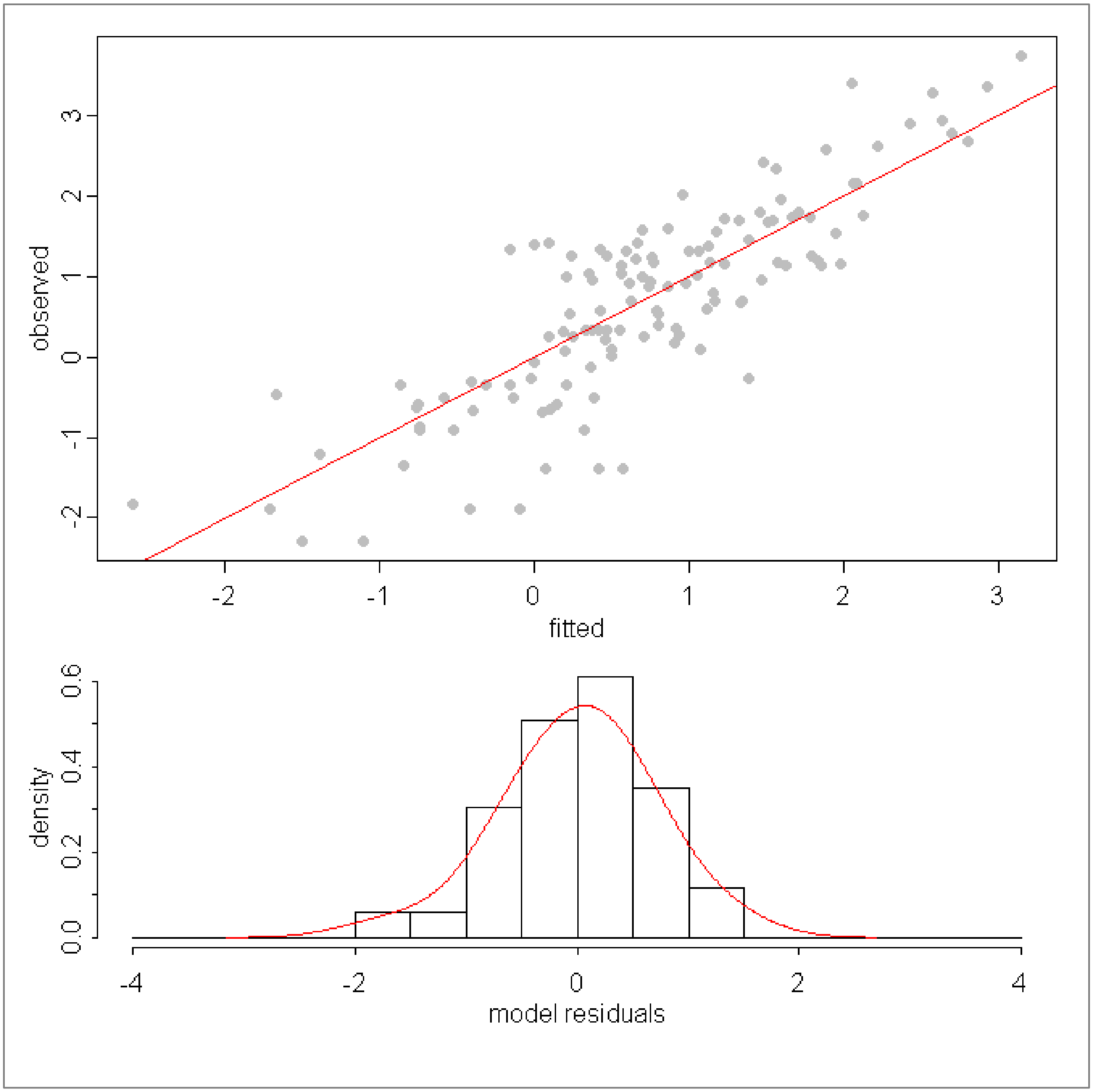

Figure 3: Fitted vs. observed plot and residual distribution density for the dissolved N model. In a perfect model, each point in the fitted vs. observed plot would fall along the linear red reference line. The red line in the model residual graph represents an approximation of residual density.

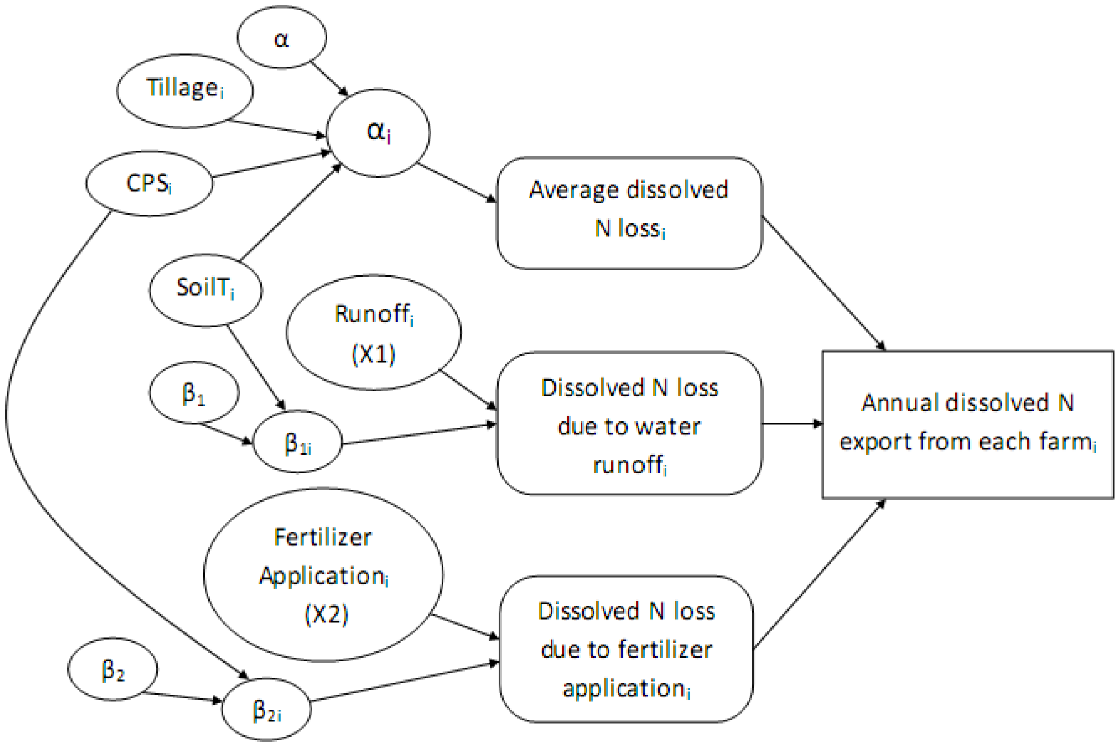

Figure 4: Flowchart of the final dissolved N model.

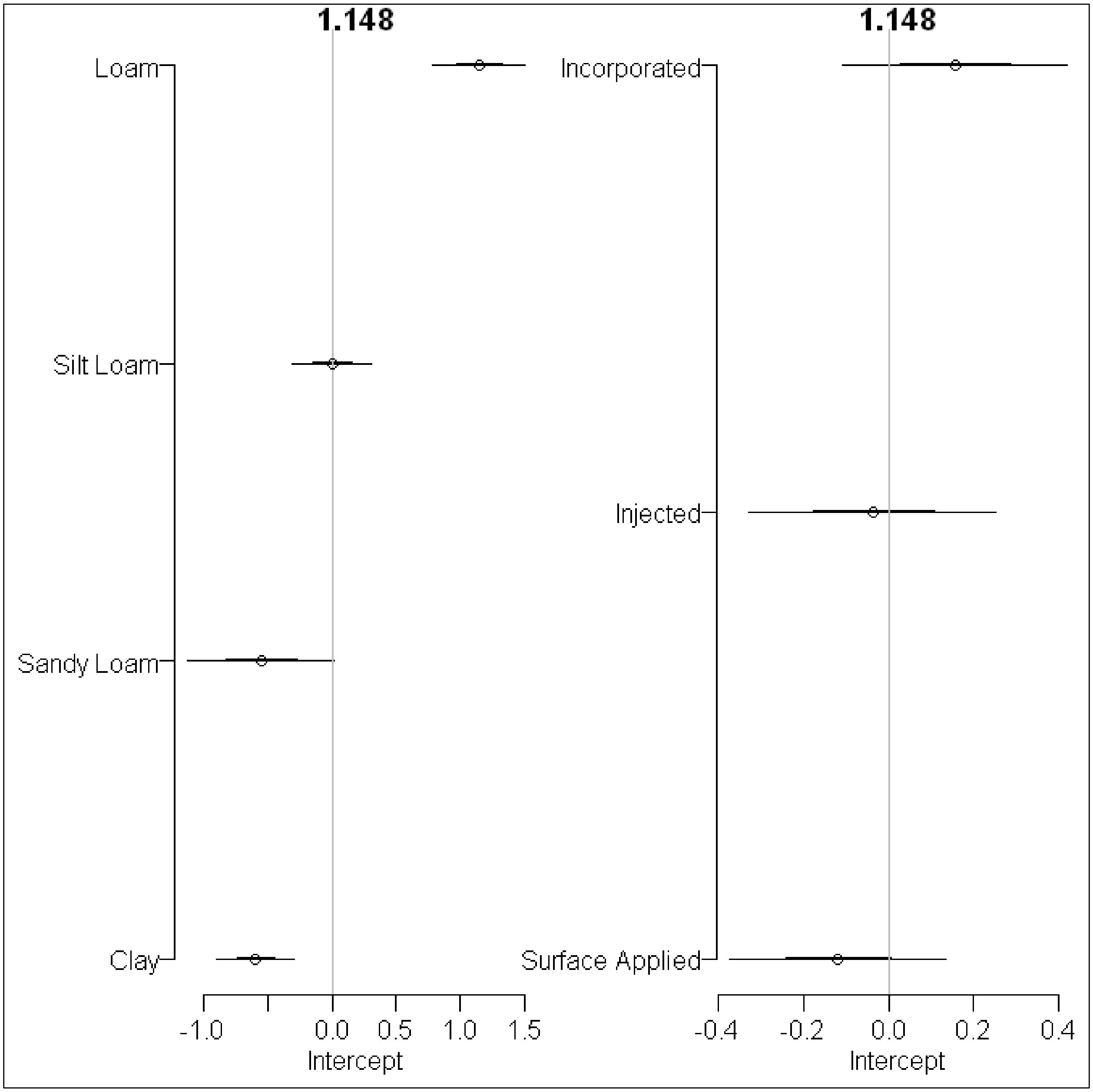

Figure 5: Effect of soil texture (left) and fertilizer application method (right) on the intercept of the particulate N model.

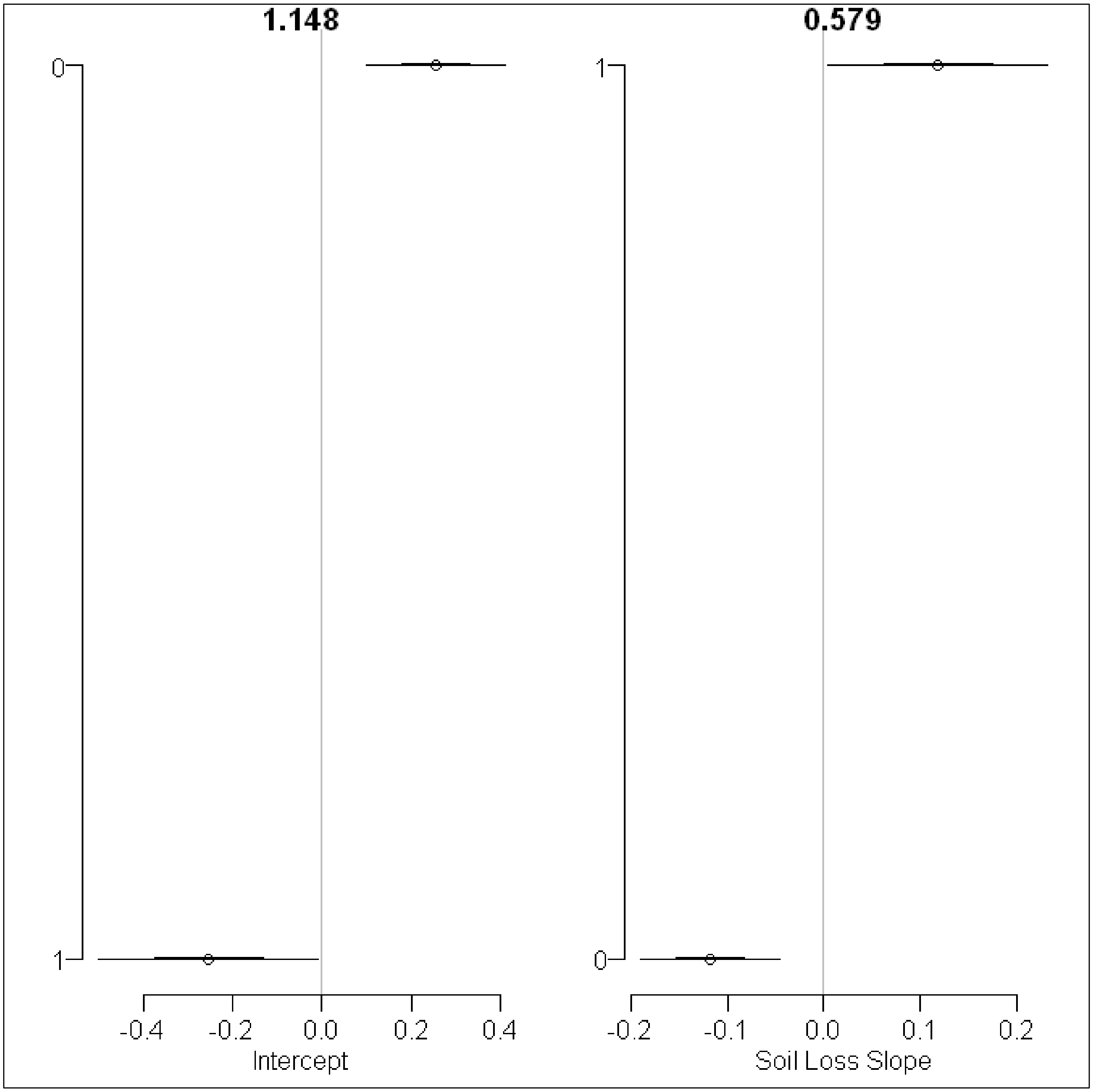

Figure 6: Effect of conservation practice on the intercept (left) and the slope of annual soil loss (right) of the particulate N model.

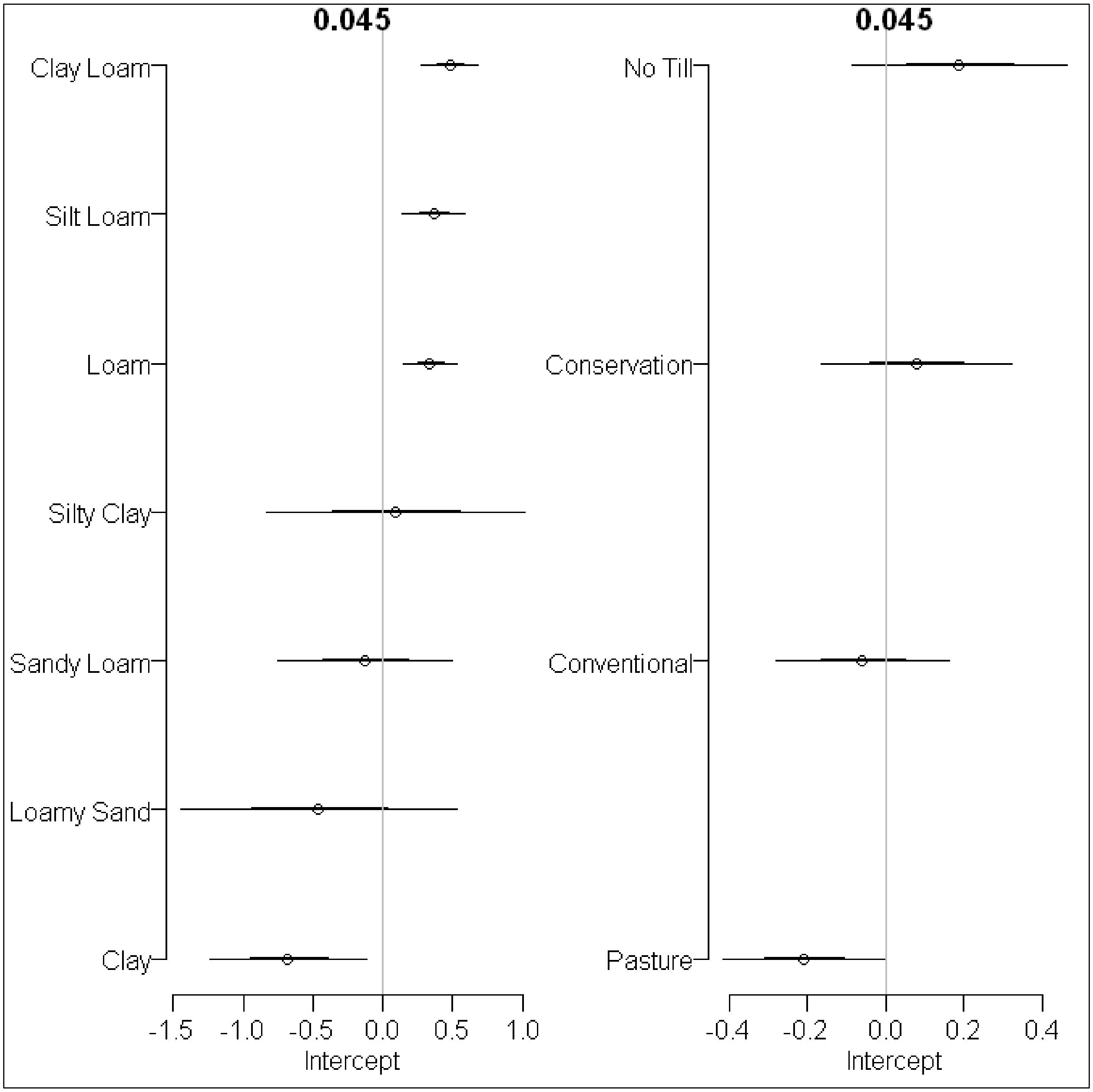

Figure 7: Effect of soil texture (left) and tillage practice (right) on the intercept of the dissolved N model.

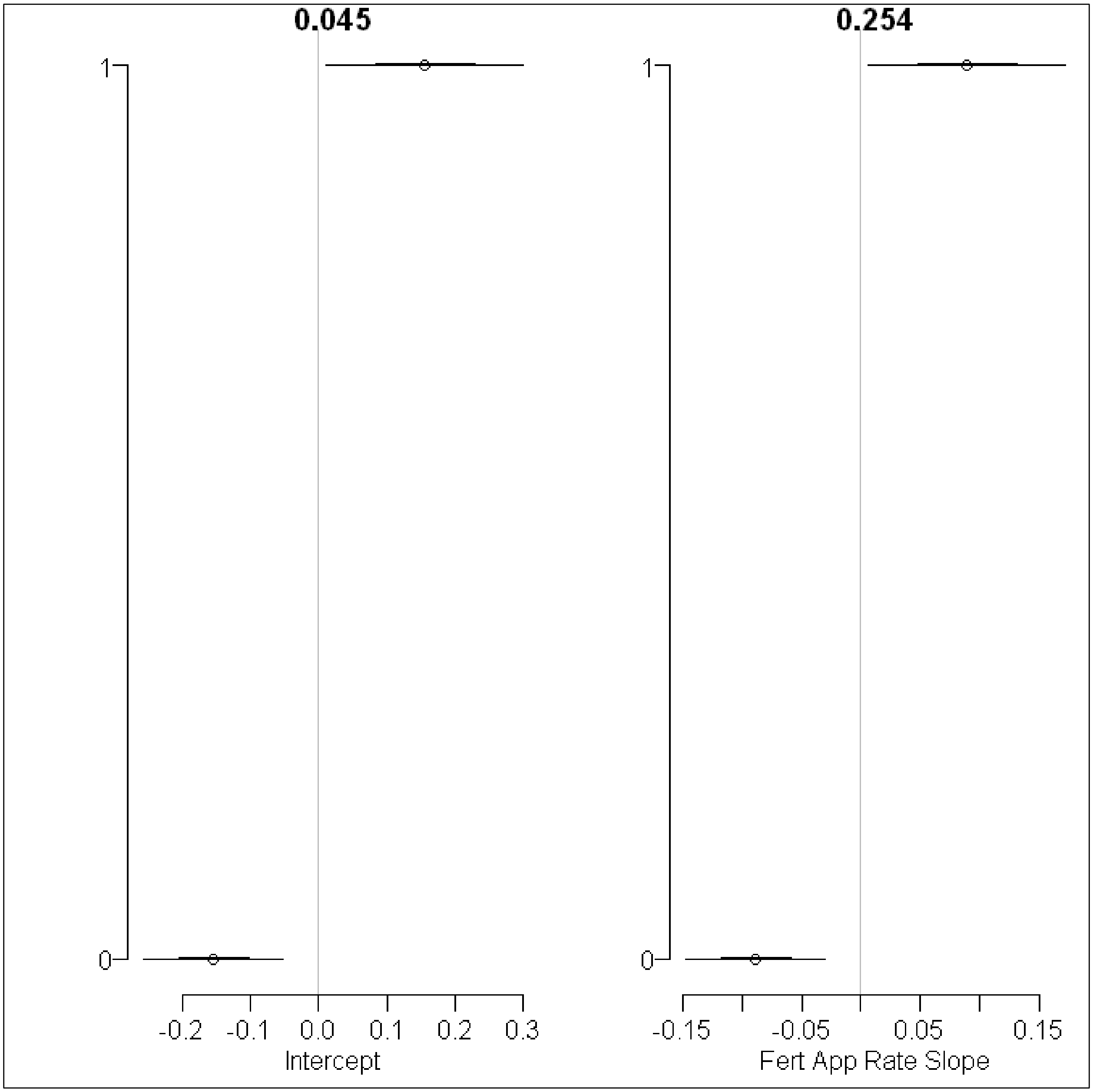

Figure 8: Effect of conservation practice on the intercept (left) and the slope of fertilizer application rate (right) of the dissolved N model.

LITERATURE CITED

Baker, D.B. (1993). The Lake Erie agroecosystem program: water quality assessments. *Agriculture, Ecosystems and Environment*, 46, 197-215.

Boyer, E., Goodale, C., Jaworski, N., and Howarth, R. (2002). Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern USA. *Biogeochemistry* 57/58, 137-169.

Carpenter, S.R., Bolgrien, D., Lathrop, R.C., Stowe, C.A., Reed, T. and Wilson, M.A., (1997). Ecological and economic analysis of lake eutrophication by nonpoint pollution. *Australian Journal of Ecology* 23, pp. 68–79.

Carpenter, S., Caraco, N., Correll, D., Howarth, A., Sharpley, A., Smith, V. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, 8, 559-568.

Cassell, E.A. and J.C. Clausen. (1993). Dynamic Simulation Modeling for Evaluating Water Quality Response to Agricultural BMP Implementation. *Water Science & Technology*, 28(3-5), 635-648.

Chambers, P. A., Meissner, R., Wrona, F., Rupp, H, Guhr, H., Seeger, J., Culp, J., Brua, R. (2006). Changes in nutrient loading in an agricultural watershed and its effects on water quality and stream biota. *Hydrobiologia*, 556, 339-415.

Craft, C., Krull, K., & Graham, S. (2007). Ecological indicators of nutrient enrichment, freshwater wetlands, Midwestern United States. *Ecological Indicators*, 7, 733-750.

Galloway, JN, Schlesinger, WH, Levy, H, et al. (1995). Nitrogen fixation – anthropogenic enhancement – environmental response. *Global Biogeochemical Cycles*, 9(2), 235-252.

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

Gelman, A. (2006). Multilevel (Hierarchical) modeling: what it can and cannot do. *Technometrics*, 48(3), 432-435.

Harmel, D., Potter, S., Casebolt, P., Reckhow, K., Green, C., Haney, R. (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, D., Qian, S., Reckhow, K., Casebolt, P. (2008). The MANAGE database: Nutrient load and site characteristic updates and runoff concentration data. *Journal of Environmental Quality*, 37, 2403-2406.

Howarth RW, Billen G, Swaney D, et al. (1996). Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: Natural and human influences. *Biogeochemistry*, 35(1), 75-139.

Jenkinson, DS. (2001). The impact of humans on the nitrogen cycle, with focus on temperate arable agriculture*. Plant and Soil*, 228(1), 3-15.

Lewis, WM., Wurtsbaugh, W. (2008). Control of lacustrine phytoplankton by nutrients: Erosion of the phosphorus paradigm. *International Review of HydroBiology*, 93(4-5), 446-465.

Makarewicz, J., Lewis, T., Bosch, I., Noll, M., Herendeen, N., Simon, R., Zollweg, J., Vodacek, A. (2008). The impact of agricultural best management practices on downstream systems: Soil loss and nutrient chemistry and flux to Conesus Lake, New York, USA*. Journal of Great Lakes Research*, 35, 23-36.

McMahon, G. and M.D. Woodside. 1997. Nutrient mass balance for the Albemarle-Pamlico drainage basin, North Carolina and Virginia, 1990. *Journal of the American Water Resources Association*, 33(3), 573-589.

Meek, B.D., D.L. Carter, D.T. Westerman, J.L. Wright, R.E. Peckenpaugh. 1995.
Nitrate leaching under furrow irrigation as affected by crop sequence and tillage.
Soil Science Society of America Journal, 59, 204-210.

Novotny, Vladimir. (2003). Water Quality. 2nd edition. Hoboken: John Wiley & Sons.

Paerl, Hans. (2009). Controlling eutrophication along the freshwater-marine continuum: Dual nutrient (N and P) reductions are essential. *Estuaries and Coasts*, 32, 593-601.

Qian, Song. *R function: plot.lmer.ranef* [R script]. Retrieved from http://web.duke.edu/~song/eesrfuns.r

Qian, Song, et al. (2010). On the application of multilevel modeling in environmental and ecological studies. *Ecology*, 91, 355–361.

Qian, Song. (2009). Environmental and Ecological Statistics with R. Baton Rouge: CRC.

Reckhow, K., Qian, S., Harmel, R. (2009). A multilevel model of the impact of farm- level best management practices on phosphorus runoff. *Journal of the American Water Resources Association*, 45(2), 369-377.

Smith, V. H., Tilman, G. D., Nekola, J. C. (1999). Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution*, 100, 179-196.

USDA. Economic Research Service. US Fertilizer Use and Price. http://www.ers.usda.gov/Data/FertilizerUse/

US EPA. (2003). National management measures to control nonpoint source pollution from agriculture (EPA Publication 841-B-03-004). Office of Water (4503T). US Government Printing Office, Washington, DC, USA.

US EPA. (2004). National water quality inventory. 2004 Report to Congress. Office of Water. US Government Printing Office, Washington, DC, USA.

Wetzel, RG. (2001). Limnology. 3rd ed. Orlando: Academic Press.