Environmental Science & Technology
Article
pubs.acs.org/est

Verifiable Metamodels for Nitrate Losses to Drains and Groundwater in the Corn Belt, USA

Bernard T. Nolan,*[,†] Robert W. Malone,[‡] Jo Ann Gronberg,[§] Kelly R. Thorp,[||] and Liwang Ma[⊥]

[†]U.S. Geological Survey, 413 National Center, Reston, Virginia 20192, United States
[‡]U.S. Department of Agriculture, 2110 University Boulevard, Ames, Iowa 50011, United States
[§]U.S. Geological Survey, 345 Middlefield Road, Menlo Park, California 94025-3561, United States
[||]U.S. Department of Agriculture, 21881 North Cardon Lane, Maricopa, Arizona 85138, United States
[⊥]U.S. Department of Agriculture, 2150 Centre Avenue, Fort Collins, Colorado 80526, United States

Supporting Information

ABSTRACT: Nitrate leaching in the unsaturated zone poses a risk to groundwater, whereas nitrate in tile drainage is conveyed directly to streams. We developed metamodels (MMs) consisting of artificial neural networks to simplify and upscale mechanistic fate and transport models for prediction of nitrate losses by drains and leaching in the Corn Belt, USA. The two final MMs predicted nitrate concentration and flux, respectively, in the shallow subsurface. Because each MM considered both tile drainage and leaching, they represent an integrated approach to vulnerability assessment. The MMs used readily available data comprising farm fertilizer nitrogen (N), weather data, and soil properties as inputs; therefore, they were well suited for regional extrapolation. The MMs effectively related the outputs of the underlying mechanistic model (Root Zone Water Quality Model) to the inputs (R^2 = 0.986 for the nitrate concentration MM). Predicted nitrate concentration was compared with measured nitrate in 38 samples of recently recharged groundwater, yielding a Pearson's r of 0.466 (p = 0.003). Predicted nitrate generally was higher than that measured in groundwater, possibly as a result of the time-lag for modern recharge to reach well screens, denitrification in groundwater, or interception of recharge by tile drains. In a qualitative comparison, predicted nitrate concentration also compared favorably with results from a previous regression model that predicted total N in streams.

INTRODUCTION

Nitrate is pervasive in groundwater and has been the focus of numerous water-quality studies at various spatial scales.[1−5] Water-quality models are useful tools that enable prediction of nitrate contamination in unsampled areas for the purpose of assessing aquifer vulnerability. Statistical models typically are used at large spatial scales.[6−11] Such models are data-driven and have comparatively few parameters, but their capability to simulate processes is limited. In contrast, mechanistic models are physically based, simulate controlling processes, and can have many parameters but typically are applied at the smaller scales. The Root Zone Water Quality Model (RZWQM2) is a field-scale systems model that simulates nitrogen (N) cycling processes, the fate and transport of agricultural chemicals, and crop growth.[12] The model provides a detailed accounting of N losses, additions, and transformations in the unsaturated zone and simulates N transport to artificial drains and groundwater. The model has been used in a variety of N studies, primarily in the Midwestern USA.[13−18] Thorough accounting by RZWQM2 of key processes can yield more accurate predictions, but application at large spatial scales is difficult because of the numerous parameters. In this research, we upscaled RZWQM2 using a metamodel (MM) for application to the Corn Belt region, USA. Metamodels are simplified representations of mechanistic models and exploit relations between model inputs and outputs (Figure 1). Although less complex, they retain some of the flexibility and process capability of more physically based simulation models.[19] The main advantage of the approach is reduced data requirements, which enable application at large spatial scales through geographic information systems (GIS). MMs may also be an effective

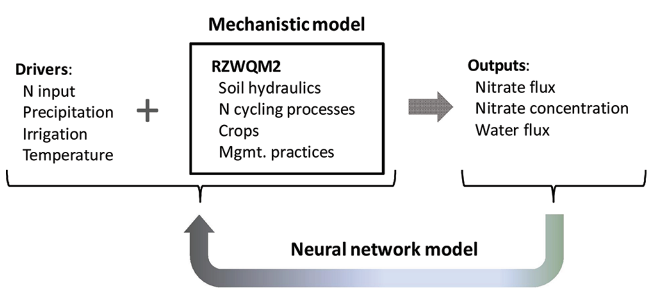

Figure 1. Metamodel concept based on Root Zone Water Quality Model (RZWQM2) inputs and outputs.

Received: August 17, 2011
Revised: November 23, 2011
Accepted: November 30, 2011
Published: November 30, 2011

means of incorporating important processes, such as plant uptake of N, into regional water-quality models. Previous regression models have focused on factors amenable to mapping at large spatial scales, such as N sources, soils, and aquifer type,[20] but not plant processes.

Prior researchers have used various statistical approaches for metamodels. Stenemo et al.[19] developed an artificial neural network (ANN) to relate MACRO model outputs to inputs for prediction of pesticides leaching in Sweden. ANNs are pattern-recognition tools that consist of simple processing elements (neurons) connected to a network by a set of weights.[9] ANNs have been widely used to model groundwater quality,[9,10,21−25] and the reader may consult the literature for detailed discussion of the method.[26] MACRO is a dual porosity mechanistic model that simulates the influence of preferential flow on water and solutes in soil.[27] Stenemo et al. trained ANNs to predict MACRO-simulated 80th percentile pesticide concentration at 1 m for an experimental grid in a 30-ha agricultural field. Overall R^2 for the second of two MMs tested was 0.98, although R^2 varied from 0.22 to 0.98 depending on the class of the target variable (<0.01, 0.01−1, and >1 $\mu g\ L^{-1}$ pesticides concentration). Other MMs have used a MACRO model emulator to predict pesticides risk in England and Wales,[28] multiple linear regression of EuroPEARL output for European scale prediction of pesticides concentration,[29] and nonlinear regression of LEACHM output to predict pesticides leaching in a watershed in eastern Maryland, USA.[30]

Previous MMs focused on pesticides leaching and most did not attempt to evaluate model predictions against measured data. In the current study, we developed two MMs to predict nitrate concentrations and fluxes, respectively, in the shallow subsurface (≤10 m). Whereas previous vulnerability models in the region considered groundwater and surface water as separate resources,[6,11,31] the MMs used an integrated approach. Each MM considered both leaching and drainage through aggregation of predictions from RZWQM2 models previously calibrated to data from Maple Creek, Nebraska (NE), Walnut Creek, Iowa (IA), and Morgan Creek, Maryland (MD)[32,33] (see the Supporting Information, SI, for site descriptions). The NE and MD sites were dominated by leaching and the IA site was tile drained. We evaluated the nitrate concentration MM by comparing predictions with measured nitrate in groundwater samples from the U.S. Geological Survey's National Water Quality Assessment (NAWQA) Program and qualitatively through comparison with previous regression results for total N in streams.[31]

Study objectives were (1) to develop neural-network MMs to predict nitrate concentrations and fluxes based on RZWQM2 predictions; (2) to evaluate the nitrate concentration MM by comparison with measured groundwater nitrate data and previous regression results; and (3) to extrapolate both MMs to assess the vulnerability of streams and groundwater in the Corn Belt.

■ METHODS

The MMs comprised ANNs that related RZWQM2 outputs (ANN response variables) to RZWQM2 inputs and other quantities estimated by the model (ANN explanatory variables) (Figure 1). We used ANNs because linear regression has several limiting assumptions, including linear relation of the response variable to the explanatory variables, and the independence and normality of model residuals.[34] The modeled system in the current study is highly nonlinear, and the observations were not independent because they were all generated by RZWQM2. In contrast, ANN is less restrictive because it does not rely on hypothesis testing.

Explanatory variables for MMs consisted of driving variables, selected RZWQM2 parameters and outputs, and measured soil properties (see Table SI-1). Soil hydraulic parameters, N inputs, irrigation, and climatic variables were systematically varied by the SENSAN utility in PEST parameter estimation software.[35] Other explanatory variables such as measured soil properties were constant at each field site. Soil hydraulic parameters consisted of water content at field capacity (WFC), saturated hydraulic conductivity (Ks), and bulk density (BD) of the predominant soil layer at each site. Predominant soil layers were identified based on sensitivities of the soil hydraulic parameters as determined during previous analyses conducted with PEST.[33,36] Weather and soil hydraulic parameters were varied by less than a single standard deviation of their ranges so as not to cause numerical instability of RZWQM2. RZWQM2 outputs included plant and N cycling quantities (potential evapotranspiration, plant N uptake, N mineralization, and N fixation) obtained from each SENSAN model run. These outputs, referred to here as "plant variables", represented processes not commonly found in statistical models applied at large spatial scales. Reliable estimates of such variables are difficult to obtain at regional scales, and RZWQM2 provides a potential alternative means for their estimation. We generated a total of 3,643 RZWQM2 runs for the three sites, resulting in wide ranges of explanatory variables meant to encompass regional conditions.

Metamodel response variables comprised RZWQM2-predicted nitrate-N fluxes and concentrations in tile drains and lateral flow in the case of IA or in soil water in the unsaturated-zone simulation profile at NE and MD. Unsaturated-zone fluxes are referred to in this study as "leaching". We emphasized nitrate because RZWQM2-predicted N losses by volatilization and denitrification were negligible at all three field sites,[32,33] and because nitrate was the dominant N species in unsaturated-zone samples from NE and MD.[2] Nitrate flux was expressed on a per-year basis (kg N ha^{-1} yr^{-1}) and nitrate concentration (mg L^{-1}) was calculated as (nitrate flux/water flux) × 10. Nitrate concentrations and fluxes predicted by RZWQM2 are referred to here as "observations" in the context of MM training.

We used the "newff" function in R's Amore package[37] to develop multilayer feed-forward ANNs for the two response variables (nitrate concentration and flux) as described in the SI. Metamodel development proceeded in two phases. First, we evaluated up to four MMs for each response variable, designated the "full model", "base model", "no plant variables" (NPV), and the "simple" model (Table 1). Each response variable for a particular type of MM used the same explanatory variables. The objective was to deconstruct the MMs by eliminating selected variables at each step (full, base, and so on), to see at what point the model performed poorly. The full model segregated fertilizer, irrigation, precipitation, and temperature data by crop (corn or soybean). The base model did not distinguish between crop type and aggregated or averaged (in the case of temperature) these amounts over the entire simulation. The NPV model further excluded RZWQM-simulated plant variables, and the simple model further excluded soil hydraulic parameters. In this phase, 70% of the cases were used for training and the remainder for model evaluation. Second, we tested the base and NPV MMs using the

Table 1. Descriptions of Neural Network Metamodels[a]

explanatory variable	full	base	no plant variables	simple
water content at field capacity[b], cm^3 cm^{-3}	X	X	X	
saturated hydraulic conductivity[b], cm h^{-1}	X	X	X	
bulk density[b], g cm^{-3}	X	X	X	
total fertilizer, kg N ha^{-1} yr^{-1}		X	X	X
corn only	X			
soybean only	X			
total irrigation, cm yr^{-1}		X	X	X
corn only	X			
soybean only	X			
average temperature, °C		X	X	X
corn only	X			
soybean only	X			
total precipitation, cm yr^{-1}		X	X	X
corn only	X			
soybean only	X			
potential evapotranspiration, cm yr^{-1}	X	X		
plant N uptake, kg N ha^{-1} yr^{-1}	X	X		
mineralization, kg N ha^{-1} yr^{-1}	X	X		
N fixation, kg N ha^{-1} yr^{-1}	X	X		
ratio of sand to silt, layer 1	X	X	X	X
organic matter content, average, percent	X	X	X	X

[a]Each response variable (nitrate concentration, nitrate flux) used all indicated variables for a particular type of metamodel. [b]Of predominant soil layer.

combined data to identify final MMs for spatial extrapolation. Metamodel fit criteria included the mean square error (MSE) and the coefficient of determination (R^2) for observations and predictions.

We performed sensitivity analyses with the base and NPV nitrate concentration MMs as described in the SI. The objective of the sensitivity analysis was to determine the influence of variables on MM output and to identify potential surrogates for plant variables to facilitate spatial extrapolation.

We compared predictions from the final nitrate concentration MM to measured nitrate in 38 wells sampled by the NAWQA Program during 1993−2003. The wells were identified through classification and regression tree (CART) analysis and represented recently recharged groundwater in intensively cropped areas (see the SI). Limitations to the verification approach include the fact that the MM represented the shallow subsurface, but the nitrate samples were from the underlying aquifer. The RZWQM2 simulations did not extend to the water table at NE and MD. However, in lieu of lysimeters which were not available at these sites, the shallow groundwater data are the best available for the MM evaluation. Because the MMs represent drainage as well as leaching, spatial patterns of predicted nitrate were qualitatively compared with patterns of total N in streams from a previous regression model.[31]

Predictions by the final MMs were mapped in a GIS to show the vulnerability of streams and groundwater in the study area. The Corn Belt was delineated by identifying areas with >50% corn, >50% soybean, or >20% each of corn and soybean based on GIS classification[38] of 1997 Census of Agriculture crop data.[39] We emphasized 1990s land use data because they are current with the 1993−2003 period of groundwater sampling. Separate GIS layers were developed at 1 km^2 resolution for each explanatory variable in the NPV model in Table 1, resulting in over 900,000 grid cells for which predictions of nitrate concentration and flux were made (see the SI).

RESULTS AND DISCUSSION

Metamodel Development and Testing. In phase 1, we deconstructed the nitrate concentration MM by eliminating selected variables at each modeling step to see at what point the model performed poorly. The main objective was to see if the MM could perform adequately without plant data, which are difficult to obtain regionally. The base MM for nitrate concentration (R^2 = 0.984, MSE = 0.00197) slightly outperformed the full MM (R^2 = 0.983, MSE = 0.00209) (Table 2). Thus, model fit was not improved by segregating input data by crop type. The NPV model fit the data only slightly less well (R^2 = 0.966, 0.00405) than the base model, suggesting that the MM can perform adequately without specifying plant variables. Variables such as plant N uptake could possibly be estimated from Census of Agriculture crop yield data.[40] For example, prior researchers showed that annual RZWQM-estimated corn yield was correlated with total N uptake (R^2 = 0.60).[41] However, Census of Agriculture data are compiled at the county level, which would introduce spatial error because crops typically vary within counties. Although plant variables were not explicitly included in the NPV MMs, they are reflected in the

Table 2. Metamodel Fit Criteria for Both Phases of Testing

model	training MSE	R^2 training (n = 2,429)	R^2 verification (n = 1,083)
Phase 1 − Training and Verification[a]			
nitrate concentration, mg L^{-1} as N			
full model	0.00209	0.983	0.969
base model	0.00197	0.984	0.978
no plant variables	0.00405	0.966	0.958
simple model	0.0483	0.616	0.605
		R^2 training (n = 2,429)	verification (n = 1,203)
nitrate flux, kg N ha^{-1} yr^{-1}			
base model	0.00282	0.983	0.936
no plant variables	0.00764	0.914	0.905
Phase 2 − All Observations[b] (n = 3,055)			
nitrate concentration, mg L^{-1} as N			
base model	0.000755	0.996	NA
no plant variables	0.00248	0.986	NA
nitrate flux, kg N ha^{-1} yr^{-1}			
base model	0.00269	0.986	NA
no plant variables	0.0189	0.911	NA

[a]High observations (i.e., RZWQM2 predictions) of nitrate concentration (>130 mg L^{-1}) and/or flux (>300 kg N ha^{-1} yr^{-1}), which occurred at NE, were excluded from the metamodels. The nitrate value was more than twice the maximum concentration in the shallow lysimeter at the field site.[33] [b]Low or high RZWQM2 predictions of nitrate flux (≤10 or >100 kg N ha^{-1} yr^{-1}) and low estimates of N fixation (≤70 kg N ha^{-1} yr^{-1}) were excluded in keeping with conditions typical of the three field sites (NE, IA, MD). RZWQM2-predicted fluxes were 27−59 kg N ha^{-1} yr^{-1} and fixation rates were 79−131 kg N ha^{-1} yr^{-1} at these sites.[32,33]

RZWQM2 predictions on which these MMs are based; therefore, the NPV MMs indirectly incorporate the effects of plant uptake and other N cycling processes.

The fit of the simple nitrate concentration MM was comparatively poor (R^2 = 0.616, MSE = 0.0483), indicating the importance of soil hydraulic properties (and by extension, water fluxes) to the modeling (Table 2). In RZWQM2, WFC and BD are used in a subroutine that scales built-in water-retention functions, θ(h), representing different soil types.[12] Ks and θ(h) are used to estimate the unsaturated hydraulic conductivity function, K(h), which is part of Richards' equation for water flow in unsaturated soils.

We evaluated predictions of nitrate flux by the base and NPV MMs to check the performance of the latter. The NPV MM had a moderate decrease in performance (R^2 = 0.914, MSE = 0.00764) compared with the base version (R^2 = 0.983, MSE = 0.00282), suggesting that plant variables are more critical to nitrate flux than nitrate concentration (Table 2). This is reasonable, because nitrate fluxes at all three sites were previously shown to be dominated by plant N uptake and fixation, according to RZWQM2-derived mass balances. At NE and MD, estimated contributions of N by fixation (131 and 99 kg N ha^{-1} yr^{-1}, respectively) exceeded those by N fertilizer (73 and 37 kg N ha^{-1} yr^{-1}, respectively).[33] The time-averaged fertilizer amounts represent both corn and soybean years; corn received 84 to 160 kg N ha^{-1} yr^{-1} in our modeling scenarios, and soybean received 0 to 15 kg N ha^{-1} yr^{-1}. Estimates of plant N uptake at NE and MD (237 and 176 kg N ha^{-1} yr^{-1}, respectively) were 3−4 times those lost by leaching (56 and 59 kg N ha^{-1} yr^{-1}, respectively). At IA, N fixation estimates (79−80 kg N ha^{-1} yr^{-1}) slightly exceeded time-averaged fertilizer N (75 kg N ha^{-1} yr^{-1}) for all three treatments, which comprised fall or spring fertilizer applications or side-dress in conjunction with the late season nitrate test (LSNT).[32] Estimated plant uptake of N at IA (212−224 kg N ha^{-1} yr^{-1}) was up to 11 times greater than N loss to tile drains (20−27 kg N ha^{-1} yr^{-1}).

The performance of all MMs was about the same for training and verification data sets (Table 2). For example, the nitrate concentration NPV MM yielded R^2 values of 0.966 and 0.958 for training and verification, respectively. Because these differences were consistently small, in phase 2 we tested the MMs using the combined data set (n = 3,055).

Phase 2 objectives were to confirm the influence of plant variables when all of the data were considered, and to identify final MMs for spatial extrapolation. As before, the nitrate concentration MM saw only a slight decrease in performance when plant variables were excluded; R^2 = 0.996 and 0.986 with and without plant variables (Table 2), and mean square error increased by a factor of 3.3. The nitrate flux MM saw moderate degradation of R^2 when plant variables were excluded (0.986 with plant variables and 0.911 without), and MSE was 7 times greater. The model fit criteria were corroborated by scatterplots of observed and predicted values. The nitrate concentration MM without plant variables had only slightly more scatter than the base version (Figure 2A-B). Although the nitrate flux MM without plant variables had more scatter (Figure 2C-D), it fit the data reasonably well (R^2 > 0.90). Therefore we considered both the nitrate concentration and flux NPV MMs as final models suitable for spatial extrapolation.

We constructed partial dependence plots to ascertain the direction of influence of explanatory variables, including plant variables, on nitrate flux (see Figure SI-1). N source terms (fertilizer, mineralization, and fixation) had a positive effect on nitrate flux, while plant N uptake had a negative effect. The effect of N sources was consistent with prior regression studies encompassing the region.[6,11] Among soil hydraulic variables, Ks had a positive effect, which is consistent with N transport tied to unsaturated zone water fluxes. WFC had a negative effect, which is reasonable because high values of this variable (>0.25) are associated with silts and clays. Water inputs such as precipitation and irrigation had a positive effect, which is reasonable. In a previous study, precipitation was positively correlated with groundwater recharge,[42] which is directly related to unsaturated-zone nitrate flux; and other researchers showed that significantly higher N leaching occurred with less efficient irrigation practices consisting of multiple irrigations.[43]

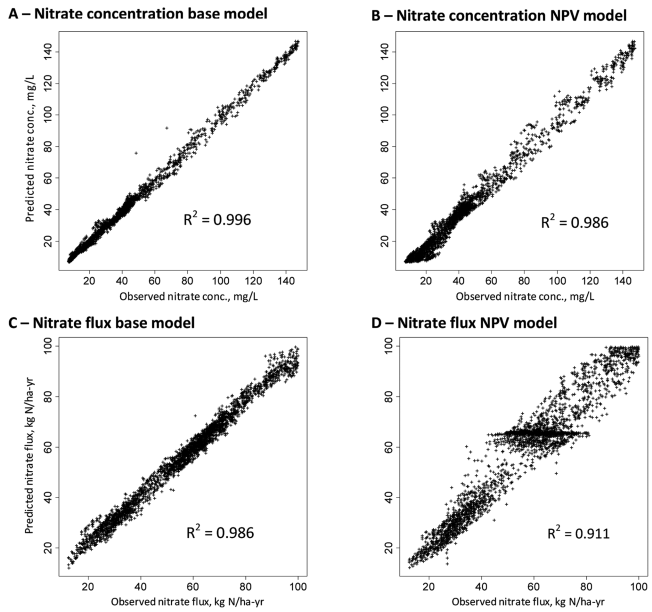

Figure 2. Comparison of observed and predicted values for base and no plant variable (NPV) metamodels of nitrate concentration (A-B) and nitrate flux (C-D) in RZWQM2 simulation profiles. Observed values are RZWQM2 predictions at Nebraska, Iowa, and Maryland.

Sensitivity Analysis. Sensitivity analyses were conducted for both the base and NPV versions of the nitrate concentration MM to assess the influence of plant variables and their potential surrogates. The four most sensitive variables in the base MM in order of decreasing sensitivity were N mineralization > plant N uptake > sand-silt ratio > N fixation, according to Sobol's main effect index (see Figure SI-3A). Mineralization and fixation are important sources of N and along with plant uptake directly affect nitrate accumulation in the shallow subsurface. The four most sensitive variables in the NPV MM were organic matter content > precipitation > water content > sand-silt ratio (see Figure SI-3B). Normalizing the main effect indices by the maximum value resulted in sensitivity values of 0.034 and 1.00, respectively, for organic matter content in the base and NPV MMs. The normalized sensitivity of this variable increased by a factor of 29.7 in the NPV MM. Organic matter evidently functioned as a surrogate for key aspects of the N cycle, such as organic N pools and mineralization (organic matter is the source of mineralized N). This provided evidence that the more parsimonious NPV models (i.e., the final MMs) can effectively substitute for the base versions in spatial extrapolation.

Metamodel Evaluation. We evaluated the final nitrate concentration MM with measured nitrate data from a subset of NAWQA wells identified by CART (see the SI). An additional, qualitative evaluation involved comparison with previous regression results and is discussed below. The 38 wells represented recently recharged groundwater beneath intensively cropped areas of the Corn Belt. Median depth of the wells was 5.6 m, and median open interval width was 1.5 m. Predicted nitrate in the shallow subsurface was significantly and positively correlated with measured groundwater nitrate (Pearson's $r = 0.466$, $p = 0.003$) (Figure 3). Although a fair

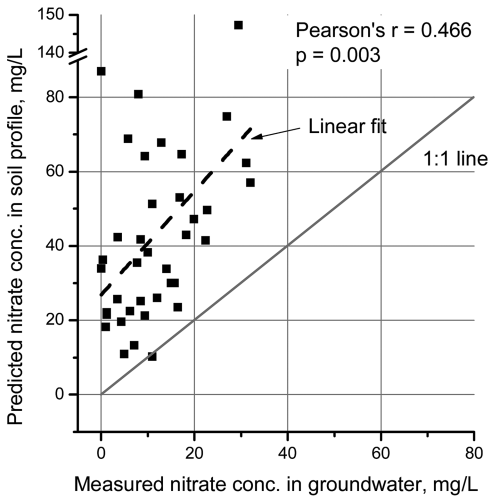

Figure 3. Comparison of metamodel (version without plant variables) predictions of shallow subsurface nitrate concentration with measured nitrate in groundwater wells.

amount of scatter is evident, we viewed the evaluation as favorable for several reasons. Most of the predictions are above the 1:1 line in Figure 3, and nitrate concentrations in the shallow subsurface, which includes the unsaturated zone, would be expected to exceed those in the aquifer. The groundwater samples reflect a mixture of ages at the well screen, with older groundwater more likely to have low nitrate concentration. It is also possible that some of the 38 wells had travel times and redox conditions favorable for denitrification, which prior researchers observed in portions of a sand and gravel aquifer in eastern NE.[44] Despite the positive bias between predicted and measured nitrate, the slope of a linear fit line (1.39) is not significantly different from the 1:1 line ($p = 0.380$ as determined by a t-ratio test).

Spatial Extrapolation. Regional maps of shallow subsurface nitrate concentration and flux were prepared from MM predictions for inputs compiled in GIS. Predicted nitrate concentration was greater in the northern part of the region and corresponded to the spatial distribution of sensitive variables in that MM (Figure 4A). For example, organic matter content was the most sensitive variable (Figure SI-3B), and STATSGO values of this parameter were higher in the northern part of the Corn Belt (data not shown). High values of organic matter occur in the Des Moines Lobe in north-central IA, a prominent high nitrate area in Figure 4A. The Des Moine Lobe is an area of recent glaciation (12,000 yrs) consisting of poorly sorted glacial till. Precipitation was the second most sensitive variable in the nitrate concentration MM and a GIS map of this variable shows a distinct north−south gradient, with higher values in the south (data not shown). Higher precipitation is associated with higher predicted water fluxes, and the predicted nitrate concentrations are inversely proportional to water flux.

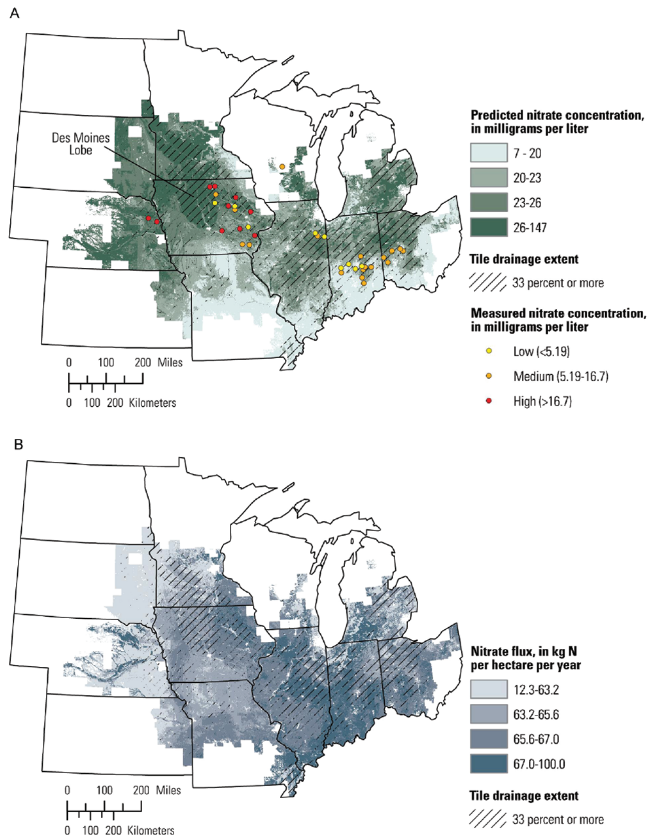

Figure 4. (A) Shallow subsurface nitrate concentration (mg L^{-1}) and (B) nitrate flux (kg N ha^{-1} yr^{-1}) predicted by metamodels without plant variables. The predictions represent leaching, tile drainage, and lateral flow in the soil profile. Subsurface nitrate in ≥33% tile-drained areas is more likely to impact streams, and nitrate in the remaining areas is more likely to leach to groundwater.

Because the MM integrates the effects of leaching and drainage, predictions of nitrate concentration in the shallow subsurface differ from groundwater nitrate risk predicted by previous statistical models in the region. A logistic regression (LR) model developed for the glacial aquifer system in the northern USA predicted low probability of nitrate in groundwaters of north-central IA and also for large areas of Illinois (IL), Indiana (IN), Michigan (MI), and Ohio (OH).[11] These areas are extensively tile drained, which restricts nitrate transport to groundwater but results in N transport to streams.[1] Additionally, the median depth of private wells used in the study was 23 m, which is well below the depth of tile drains. In contrast, the MM predicted nitrate concentration in drainage as well as by leaching in the shallow subsurface; therefore, areas such as north-central IA are shown as high risk in Figure 4A. An analysis of farm drainage in the USA showed that subsurface drains are extensive in northern IA, IL, IN, MI, Minnesota (MN), and OH.[45] This was further corroborated by a map of farmed areas in tile drains developed by researchers at the U.S. Department of Agriculture[46] (see Figure SI-5).

Another difference from previous aquifer vulnerability models[6,11] is that in the current study, significant amounts of nitrate in drainage and leachate are attributable to unsaturated zone organic matter, which is a surrogate for mineralization. RZWQM2-estimated mineralization rates at NE and MD (88 and 93 kg N ha^{-1} yr^{-1}, respectively) exceeded time-averaged fertilizer application rates (73 and 37 kg N ha^{-1} yr^{-1}, respectively).[33] Similarly, mineralization estimates at IA (113–116 kg N ha^{-1} yr^{-1}) exceeded fertilizer inputs (75 kg N ha^{-1} yr^{-1}). The effect of organic matter is consistent with previous research that showed an average increase in organic N concentration of 8.8% (320 kg N ha^{-1}) following soybean growth on various soils in IA.[47] The organic N concentrations represented the potential amount of N mineralized during the growing season.

The MM predicted high nitrate concentration in southeastern NE (Figure 4A), which is not extensively tile drained according to Figure SI-5. These areas have coarse textured soils that commonly are irrigated and receive high N input, factors that combine to promote nitrate leaching to groundwater. This is consistent with results of a national groundwater vulnerability model that included variables for fertilizer N, water input, and soil texture.[6]

Spatial patterns of nitrate concentration predicted by the MM generally are consistent with total N concentrations in streams predicted by CONDOR, a multiple linear regression model developed at national scale.[31] Total N comprises nitrate, nitrite, ammonia, and organic N, but nitrate is the primary form of N in streams and groundwater.[1] CONDOR predicted high total N concentration (>5 mg L^{-1}) in the same areas of the Corn Belt shown as high risk in Figure 4A, namely eastern NE, southern MN, and all but the southern portions of IA, IL, IN, and OH. Excluding eastern NE, these areas are extensively tile drained. Total N concentrations predicted by CONDOR were less than nitrate concentrations predicted by the MM, likely because of dilution by precipitation and overland flow in basins with low N input.[31] Precipitation and overland flow both had negative signs in the CONDOR model. The MM indicated high nitrate concentration for some areas of the Corn Belt that were predicted by CONDOR to have comparatively low total N (e.g., northern Missouri (MO) and southern MI). Such disparities suggested areas where leaching, rather than drainage, is the predominant N loss pathway. Both MO and MI are less extensively tile-drained than IA, IL, IN, and OH according to Figure SI-5. Additionally, northern MO has high values of organic matter content according to STATSGO data, which suggested increased N input from mineralization. Southern MI has high organic matter content and high sand content, which can promote nitrate leaching. Overall, CONDOR results corroborated predictions by the nitrate concentration MM and underscored the effect of tile drainage on streams in the Corn Belt.

Following verification, we used the metamodel to estimate nitrate concentration and flux for drained and undrained areas of the Corn Belt. We separated the region into areas likely to be tile drained (grid cells with percent drained areas ≥33%, which is the 75th percentile of drainage extent for all 1 km^2 grid cells in the region; see the SI) and unlikely to be drained (grid cells with percent drained areas <33%), and calculated mean predicted nitrate concentration for each area. Mean predicted nitrate concentration was 26 mg L^{-1} in likely tile drained areas and 24 mg L^{-1} in areas unlikely to have drains. Predicted nitrate concentration in tile drained areas is less than that reported for drains in corn-soybean areas of Indiana (39 mg L^{-1}).[48] However, that study reported higher N fertilization rates on corn and soybean than considered here (168 kg N ha^{-1}).

Predicted nitrate fluxes generally were greater in the eastern portion of the study area (Figure 4B), which reflected regional patterns of precipitation and shallow subsurface water fluxes. The latter generally were higher in the southeastern Corn Belt (data not shown). Predicted annual nitrate mass in kg N km^{-2} was summed for all grid cells in the likely drained and undrained areas. In areas where leaching predominates, the shallow subsurface contained an estimated 4.1×10^6 metric tons of nitrate, and areas likely to be tile-drained contained an estimated 1.4×10^6 metric tons of nitrate. Because the drain locations are uncertain, we repeated the analysis using the median drained area (20%) of Midwestern states substantially within the study area, derived from previous estimates.[45,46] This yielded estimates of 3.5×10^6 metric tons of nitrate in the likely undrained areas and 2.0×10^6 metric tons of nitrate in the likely drained areas. Based on these calculations, an estimated $1.4–2.0 \times 10^6$ metric tons of nitrate impacts streams annually via drains in the Corn Belt, and an estimated $3.5–4.1 \times 10^6$ metric tons leaches beyond the root zone annually.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Site descriptions; explanatory variables; GIS derivation of explanatory variables; partial dependence plots; ANN metamodels; model evaluation; determination of farmed areas in tile drainage. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 703-648-4000. Fax: 703-648-6693. E-mail: btnolan@usgs.gov.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the USGS and USDA field personnel who collected data used in this study. We thank Terri L. Arnold for providing an insightful review of the draft manuscript, and Dan B. Jaynes for reviewing the manuscript and providing critical insight into tile drained lands in the study area. Lastly, David L. Lorenz provided valuable assistance with R coding.

■ REFERENCES

(1) Dubrovsky, N. M.; Burow, K. R.; Clark, G. M.; Gronberg, J. M.; Hamilton, P. A.; Hitt, K. J.; Mueller, D. K.; Munn, M. D.; Nolan, B. T.; Puckett, L. J.; Rupert, M. G.; Short, T. M.; Spahr, N. E.; Sprague, L. A.; Wilber, W. G. The quality of our Nation's waters—Nutrients in the Nation's streams and groundwater, 1992–2004. U.S. Geological Survey Circular 1350, 2010.

(2) Green, C. T.; Fisher, L. H.; Bekins, B. A. Nitrogen fluxes through unsaturated zones in five agricultural settings across the United States. *J. Environ. Qual.* **2008**, *37*, 1073–1085.

(3) Puckett, L. J.; Cowdery, T. K.; Lorenz, D. L.; Stoner, J. D. Estimation of nitrate contamination of an agro-ecosystem outwash aquifer using a nitrogen mass-balance budget. *J. Environ. Qual.* **1999**, *28*, 2015–2025.

(4) Burow, K. R.; Nolan, B. T.; Rupert, M. G.; Dubrovsky, N. M. Nitrate in groundwater of the United States, 1991–2003. *Environ. Sci. Technol.* **2010**, *44*, 4988–4997.

(5) Böhlke, J. K.; Denver, J. M. Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic coastal plain, Maryland. *Water Resour. Res.* **1995**, *31*, 2319–2339.

(6) Nolan, B. T.; Hitt, K. J. Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States. *Environ. Sci. Technol.* **2006**, *40*, 7834–7840.

(7) Rupert, M. G. Probability of detecting atrazine/desethyl-atrazine and elevated concentrations of nitrate in ground water in Colorado.

U.S. Geological Survey Water-Resources Investigations Report 02-4269, 2003.
(8) Ayotte, J. D.; Nolan, B. T.; Nuckols, J. R.; Cantor, K. P.; Robinson, G. R. Jr; Baris, D.; Hayes, L.; Karagas, M.; Bress, W.; Silverman, D. T.; Lubin, J. H. Modeling the probability of arsenic in groundwater in New England as a tool for exposure assessment. *Environ. Sci. Technol.* **2006**, *40*, 3578−3585.
(9) Yesilnacar, M. I.; Sahinkaya, E.; Naz, M.; Ozkaya, B. Neural network prediction of nitrate in groundwater of Harran Plain, Turkey. *Environ. Geol.* **2008**, *56*, 19−25.
(10) Gemitzi, A.; Petalas, C.; Pisinaras, V.; Tsihrintzis, V. A. Spatial prediction of nitrate pollution in groundwaters using neural networks and GIS: An application to South Rhodope aquifer (Thrace, Greece). *Hydrol. Processes* **2009**, *23*, 372−383.
(11) Warner, K. L.; Arnold, T. L. Relations that affect the probability and prediction of nitrate concentration in private wells in the glacial aquifer system in the United States. U.S. Geological Survey Scientific Investigations Report 2010-5100, 2010.
(12) Ahuja, L. R.; Rojas, K. W.; Hanson, J. D.; Shaffer, M. J.; Ma, L. *Root Zone Water Quality Model - Modelling Management Effects on Water Quality and Crop Production*; Water Resources Publications LLC: Highlands Ranch, CO, 2000.
(13) Jaynes, D.; Miller, J. Evaluation of the Root Zone Water Quality Model using data from the Iowa MSEA. *Agron. J.* **1999**, *91*, 192−200.
(14) Kumar, A.; Kanwar, R. S.; Singh, P.; Ahuja, L. R. Evaluation of the Root Zone Water Quality Model for predicting water and NO3-N movement in an Iowa soil. *Soil Tillage Res.* **1999**, *50*, 223−236.
(15) Ma, L.; Malone, R. W.; Heilman, P.; Jaynes, D. B.; Ahuja, L. R.; Saseendran, S. A.; Kanwar, R. S.; Ascough, I. J. C. RZWQM simulated effects of crop rotation, tillage, and controlled drainage on crop yield and nitrate-N loss in drain flow. *Geoderma* **2007**, *140*, 260−271.
(16) Ma, L.; Malone, R. W.; Heilman, P.; Karlen, D. L.; Kanwar, R. S.; Cambardella, C. A.; Saseendran, S. A.; Ahuja, L. R. RZWQM simulation of long-term crop production, water and nitrogen balances in Northeast Iowa. *Geoderma* **2007**, *140*, 247−259.
(17) Malone, R. W.; Ma, L.; Heilman, P.; Karlen, D. L.; Kanwar, R. S.; Hatfield, J. L. Simulated N management effects on corn yield and tile-drainage nitrate loss. *Geoderma* **2007**, *140*, 272−283.
(18) Malone, R. W.; Ma, L.; Karlen, D. L.; Meade, T.; Meek, D.; Heilman, P.; Kanwar, R. S.; Hatfield, J. L. Empirical analysis and prediction of nitrate loading and crop yield for corn-soybean rotations. *Geoderma* **2007**, *140*, 223−234.
(19) Stenemo, F.; Lindahl, A. M. L.; Gärdenäs, A.; Jarvis, N. Metamodeling of the pesticide fate model MACRO for groundwater exposure assessments using artificial neural networks. *J. Contam. Hydrol.* **2007**, *93*, 270−283.
(20) Nolan, B. T.; Hitt, K. J.; Ruddy, B. C. Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States. *Environ. Sci. Technol.* **2002**, *36*, 2138−2145.
(21) Chowdhury, M.; Alouani, A.; Hossain, F. Comparison of ordinary kriging and artificial neural network for spatial mapping of arsenic contamination of groundwater. *Stoch. Environ. Res. Risk Assess.* **2010**, *24*, 1−7.
(22) Ray, C.; Klindworth, K. K. Neural networks for agrichemical vulnerability assessment of rural private wells. *J. Hydrol. Eng.* **2000**, *5*, 162−171.
(23) Sahoo, G. B.; Ray, C.; Mehnert, E.; Keefer, D. A. Application of artificial neural networks to assess pesticide contamination in shallow groundwater. *Sci. Total Environ.* **2006**, *367*, 234−251.
(24) Singh, R. M.; Datta, B. Groundwater pollution source identification and simultaneous parameter estimation using pattern matching by artificial neural network. *Environ. Foren.* **2004**, *5*, 143−153.
(25) Wu, Q.; Xu, H.; Pang, W. GIS and ANN coupling model: An innovative approach to evaluate vulnerability of karst water inrush in coalmines of north China. *Environ. Geol.* **2008**, *54*, 937−943.
(26) Abdi, H.; Valentin, D.; Edelman, B. *Neural Networks*; Sage Publications, Inc.: Thousand Oaks, CA, 1999.
(27) Larsbo, M.; Jarvis, N. J. MACRO 5.0. A model of water flow and solute transport in macroporous soil. Technical description. Swedish University of Agricultural Sciences Report Emergo 2003:6, 2003.
(28) Holman, I. P.; Dubus, I. G.; Hollis, J. M.; Brown, C. D. Using a linked soil model emulator and unsaturated zone leaching model to account for preferential flow when assessing the spatially distributed risk of pesticide leaching to groundwater in England and Wales. *Sci. Total Environ.* **2004**, *318*, 73−88.
(29) Tiktak, A.; Boesten, J. J. T. I.; van der Linden, A. M. A.; Vanclooster, M. Mapping ground water vulnerability to pesticide leaching with a process-based metamodel of EuroPEARL. *J. Environ. Qual.* **2006**, *35*, 1213−1226.
(30) Webb, R. M. T.; Wieczorek, M. E.; Nolan, B. T.; Hancock, T. C.; Sandstrom, M. W.; Barbash, J. E.; Bayless, E. R.; Healy, R. W.; Linard, J. Variations in pesticide leaching related to land use, pesticide properties, and unsaturated zone thickness. *J. Environ. Qual.* **2008**, *37*, 1145−1157.
(31) Spahr, N. E.; Mueller, D. K.; Wolock, D. M.; Hitt, K. J.; Gronberg, J. M. Development and application of regression models for estimating nutrient concentrations in streams of the conterminous United States, 1992−2001. U.S. Geological Survey Scientific Investigations Report 2009-5199, 2010.
(32) Malone, R. W.; Jaynes, D. B.; Ma, L.; Nolan, B. T.; Meek, D. W.; Karlen, D. L. Soil-test N recommendations augmented with PEST-optimized RZWQM simulations. *J. Environ. Qual.* **2010**, *39*, 1711−1723.
(33) Nolan, B. T.; Puckett, L. J.; Ma, L.; Green, C. T.; Bayless, E. R.; Malone, R. W. Predicting unsaturated zone nitrogen mass balances in agricultural settings of the United States. *J. Environ. Qual.* **2010**, *39*, 1051−1065.
(34) Helsel, D. R.; Hirsch, R. M. *Statistical Methods in Water Resources*; U.S. Geological Survey Techniques of Water Resources Investigations: Book 4, Chapter A3, 2002.
(35) Doherty, J. *PEST - Model Independent Parameter Estimation*; Watermark Numerical Computing: Brisbane, Australia, 2010.
(36) Nolan, B. T.; Malone, R. W.; Ma, L.; Green, C. T.; Fienen, M. N.; Jaynes, D. B. Inverse Modeling with RZWQM2 to Predict Water Quality. In *Methods of Introducing System Models into Agricultural Research*; Ahuja, L., Ma, L., Eds.; ASA-CSSA-SSSA: Madison, WI, 2011.
(37) R, The R project for statistical computing. 2011. <http://www.r-project.org/> (accessed June 2011).
(38) Gilliom, R. J.; Thelin, G. P. Classification and mapping of agricultural land for national water quality assessment. U.S. Geological Survey Circular 1131, 1997.
(39) USDA 1997 Census of Agriculture, Geographic area series 1A, 1B, 1C, U.S. summary and county level data file. National Agricultural Statistics Service, United States Department of Agriculture [CD-ROM]: Washington, DC, 1999.
(40) USDA, Quick Stats. National Agricultural Statistics Service, *United States Department of Agriculture*, 2011. <http://www.nass.usda.gov/QuickStats/Create_County_All.jsp> (accessed June 2011).
(41) Malone, R. W.; Ma, L. N uptake effects on N loss in tile drainage as estimated by RZWQM. In *New Advances in Understanding and Quantification of Plant N Uptake*; Ma, L., Bruulsema, T., Ahuja, L., Eds.; CRC Press: Boca Raton, FL, 2008.
(42) Nolan, B. T.; Healy, R. W.; Taber, P. E.; Perkins, K.; Hitt, K. J.; Wolock, D. M. Factors influencing ground-water recharge in the eastern United States. *J. Hydrol.* **2006**, *332*, 187−205.
(43) Nakamura, K.; Harter, T.; Hirono, Y.; Horino, H.; Mitsuno, T. Assessment of root zone nitrogen leaching as affected by irrigation and nutrient management practices. *Vadose Zone J.* **2004**, *3*, 1353−1366.
(44) Green, C. T.; Puckett, L. J.; Böhlke, J. K.; Bekins, B. A.; Phillips, S. P.; Kauffman, L. J.; Denver, J. M.; Johnson, H. M. Limited occurrence of denitrification in four shallow aquifers in agricultural areas of the United States. *J. Environ. Qual.* **2008**, *37*, 994−1009.
(45) Pavelis, G. A., Ed. *Economic survey of farm drainage*; U.S. Gov. Print. Office: Washington, DC, 1987.

(46) Jaynes, D. B.; James, D. E. The extent of farm drainage in the United States. U.S. Department of Agriculture, 2007. <http://www.ars.usda.gov/SP2UserFiles/Place/36251500/TheExtentofFarmDrainageintheUnitedStates.pdf> (accessed November 11, 2011).
(47) Martens, D. A.; Jaynes, D. B.; Colvin, T. S.; Kaspar, T. C.; Karlen, D. L. Soil organic nitrogen enrichment following soybean in an Iowa corn−soybean rotation. *Soil Sci. Soc. Am. J.* **2006**, *70*, 382−392.
(48) Brouder, S.; Hofmann, B.; Kladivko, E.; Turco, R.; Bongen, A.; Frankenberger, J. Interpreting nitrate concentration in tile drainage water. Purdue University Agronomy Guide AY-318-W, 2004. <http://www.extension.purdue.edu/extmedia/AY/AY-318-W.pdf> (accessed June 2011).