Examining the impacts of increased corn production on groundwater quality using a coupled modeling system

Valerie Garcia\textsuperscript{a,}*, Ellen Cooter\textsuperscript{a}, James Crooks\textsuperscript{b}, Brian Hinckley\textsuperscript{c,1}, Mark Murphy\textsuperscript{d}, and Xiangnan Xing\textsuperscript{e,2}

\textsuperscript{a}Environmental Protection Agency, (MD-E243-02), 109 TW Alexander Drive, RTP, NC 27711, United States

\textsuperscript{b}National Jewish Health, 1400 Jackson St., Denver, CO 80206, United States

\textsuperscript{c}Oak Ridge Institute for Science and Education, 1299 Bethel Valley Rd., Oak Ridge, TN 37830, United States

\textsuperscript{d}Innovate! Inc., Alexandria, VA 22314, United States

\textsuperscript{e}Oak Ridge Institute for Science and Education, 1299 Bethel Valley Rd., Oak Ridge, TN 37830, United States

\textsuperscript{1}Currently with East Carolina University, East 5th St., Greenville, NC 27858, United States.

\textsuperscript{2}Currently with West Virginia University, Morgantown, WV 26506, United States.

Abstract

This study demonstrates the value of a coupled chemical transport modeling system for investigating groundwater nitrate contamination responses associated with nitrogen (N) fertilizer application and increased corn production. The coupled Community Multiscale Air Quality Bidirectional and Environmental Policy Integrated Climate modeling system incorporates agricultural management practices and N exchange processes between the soil and atmosphere to estimate levels of N that may volatilize into the atmosphere, re-deposit, and seep or flow into surface and groundwater. Simulated values from this modeling system were used in a land-use regression model to examine associations between groundwater nitrate-N measurements and a suite of factors related to N fertilizer and groundwater nitrate contamination. Multi-variable modeling analysis revealed that the N-fertilizer rate (versus total) applied to irrigated (versus rainfed) grain corn (versus other crops) was the strongest N-related predictor variable of groundwater nitrate-N concentrations. Application of this multi-variable model considered groundwater nitrate-N concentration responses under two corn production scenarios. Findings suggest that increased corn production between 2002 and 2022 could result in 56\% to 79\% increase in areas vulnerable to groundwater nitrate-N concentrations \( \geq 5 \text{ mg/L} \). These above-threshold areas occur on soils with a hydraulic conductivity 13\% higher than the rest of the domain. Additionally, the average number of animal feeding operations (AFOs) for these areas...
was nearly 5 times higher, and the mean N-fertilizer rate was 4 times higher. Finally, we found that areas prone to high groundwater nitrate-N concentrations attributable to the expansion scenario did not occur in new grid cells of irrigated grain-corn croplands, but were clustered around areas of existing corn crops. This application demonstrates the value of the coupled modeling system in developing spatially refined multi-variable models to provide information for geographic locations lacking complete observational data; and in projecting possible groundwater nitrate-N concentration outcomes under alternative future crop production scenarios.

**GRAPHICAL ABSTRACT**

Keywords

Agricultural impacts; Biosphere modeling; Regression modeling; Groundwater quality

1. **Introduction**

Nitrogen (N) is critical to life on earth, but excess N can be transported to waterways in surface and subsurface runoff, leached into groundwater or emitted to the air and deposited back to underlying surfaces. Exposure to this excess N can result in ecological and human health impacts such as fish kills, human disease and birth defects (Johnson et al., 2010; Ward et al., 2005). Agricultural activities are a significant source of N released into the environment by humans. Corn production, in particular, requires large amounts of N fertilization to achieve the highest yields (Ribaudo et al., 2011; Sobota et al., 2013). Demand for corn grown in the United States (US) is expected to rise because corn is an important commodity for sustaining world populations, and more recently, because of its use as a biofuel. This increase in demand for corn, and subsequent increase in N fertilization, may raise the risk of human exposures to nitrate from contaminated drinking water wells. Exposure through the consumption of contaminated drinking water has been associated with some cancers and birth defects (Ward et al., 2005; Brender et al., 2013).

The rising demand for corn in the US, however, is also expected to be met with technology advancements, cropland redistributions, and a vegetation-enhancing atmosphere higher in carbon dioxide levels (Lark et al., 2015; The Hightower Report, 2015). The ability to anticipate the opposing impacts of evolving political (e.g., agricultural subsidies, biofuel production mandates), economic (e.g., demand for food and livestock feed) and
environmental (e.g., weather, soil health) conditions is essential to understanding the intended and unintended consequences of current and future demand for corn. Integrated or coupled systems-level modeling has the potential to provide the spatially complete, detailed information needed to develop exposure models that meet these emerging demands, while providing improved (refined) identification of locations that would benefit from more rigorous, resource-intensive analyses. In addition, such models are needed to examine the impact of environmental decisions through future scenario analyses.

In this study, we used a coupled modeling system to simulate the impacts of various corn production scenarios. We used consistent inputs (e.g., emissions, meteorology, land use) to drive the component models within the coupled modeling system, and maintained mass-balanced equations throughout the integrated simulations, resulting in a rich source of information on the fate of N originating from crop fertilization. We used this extensive dataset in a statistical model to describe N-loadings to groundwater as a function of a variety of environmental variables. These variables included measurements of nitrate-N in drinking water wells from the US Geological Survey (USGS), and data from the coupled model at those well locations, to estimate nitrate-N in ground-water at all other locations that met our model criteria but lacked well water observations. With this approach, we estimated N loadings and related groundwater nitrate-N concentrations for 2002 (before the introduction of new US biofuel policies) and two future-year simulations (2022\textsubscript{BASE}, increasing population demand and increasing CO\textsubscript{2} concentrations; 2022\textsubscript{CROP}, and hypothetical biofuel production targets in addition to 2022\textsubscript{BASE} production increases).

While several other generalizable studies (e.g., DeSimone et al., 2009; Nolan et al., 2002; Greene et al., 2005; Nolan and Hitt, 2006) have revealed associations between N-loadings from agricultural management practices (e.g., proximity of animal feeding operations, total amount of inorganic N fertilizer) and groundwater nitrate contamination, these studies were limited by their reliance on historical county level fertilizer sales data, which in turn, are tied to social policies, economic constraints, and spatially incomplete management and weather conditions. Our approach provides more physically and spatially detailed N loading information (e.g., type of fertilizer placed on each crop type, use of drainage tiles, tillage, or irrigation). This allowed us to provide more geographically targeted outcomes, which are of use in fine-tuning future data collection for specific areas, and improving the characterization of wellhead rotation and auger recharge areas. More detailed, process-based characterization of fertilizer applications also allowed us to examine the impact, both collectively and by individual driver, of corn production scenarios on groundwater nitrate-N contamination under environmental and socio-economic conditions that transcend the historical conditions. Accordingly, the objectives of this study are to: (1) refine our understanding of N loadings and interactions related to crop fertilization and groundwater nitrate contamination; (2) predict changes in groundwater nitrate contamination for a base- and two future-year agricultural corn production scenarios; and (3) examine these changes to better understand the impacts of potential corn production expansion on groundwater quality.
2. **Approach**

The US Environmental Protection Agency’s (EPA’s) CMAQ version 5.1 model with bidirectional ammonia exchange (bidiCMAQ) was coupled with a modified version of the US Department of Agriculture’s (USDA’s) Environmental Policy Integrated Climate (EPIC; Williams et al., 2012) agroecosystem model as described in Cooter et al. (2012) and Bash et al. (2013). BidiCMAQ employs a 3-dimensional Eulerian modeling approach to address regional air quality issues such as tropospheric ozone, fine particles, acid deposition and visibility degradation (Appel et al., 2011). EPIC is a field-scale, semi-empirical model that produces daily estimates of fertilizer applied to all crop types grown in the US. More information about how these models were coupled and the related model evaluation can be found in Cooter et al. (2012) and Bash et al. (2013). While the coupled modeling system does not provide estimates of groundwater nitrate-N directly, it does simulate the movement of reactive N through the soil layers. We used the 2002 annually summed or averaged N-loading variables produced by the bidiCMAQ-EPIC modeling system (e.g., N deposition, N fertilizer applied, N soil concentrations, and agriculture management practices such as tilling and irrigation) in a land-use regression model, using nitrate-N measurements taken by USGS in 2002 from drinking water wells located predominantly in agricultural areas throughout the US (Fig. 1a) as the response variable. We then applied the coefficients calculated from the land-use regression approach to predict groundwater nitrate-N concentrations for 2002 throughout the remaining U.S. domain and for two crop production scenarios in 2022. Two policy drivers are included in our scenario analysis; the adoption of hypothetical corn-based biofuel (ethanol) production volumes and the implementation of prescribed Clean Air Act (CAA) emission reductions (http://www.epa.gov/criteria-air-pollutants/naaqs-table/). Our scenarios analysis begins in 2002 to simulate conditions prior to the active implementation of both these policies, and ends in 2022 when ethanol production volume and CAA emission reduction goals were expected to be met representing a common (20-year) US environmental policy projection timeline.

2.1. **Groundwater well measurements**

Nitrate measurements (milligrams (mg) of nitrate-measured-as-nitrogen per liter; nitrate-N) collected from domestic wells in 2002 (Fig. 1a) as part of the USGS NAWQA Program were downloaded from the NAWQA Data Warehouse website (U.S. Geological Survey, 2001). The NAWQA sampling scheme was based on hydrologic systems rather than a single, national-scale assessment of domestic well water quality. Thus, the targeted sampling areas were defined by the extent of major hydrogeologic settings and are not uniformly distributed across the US (see Land-use regression model section for addressing clustered measurements). More information about the well measurements is available from DeSimone et al. (2009). A total of 878 measurements were downloaded. Two or more well locations within a buffer zone (see Processing of data section for selecting buffer size) were averaged, reducing the sample size to 806. Records were removed if they were missing data for the well location (e.g., missing shallow aquifer type or nitrate-N measurement), or contained concentrations below the instrument sensitivity threshold level of 0.05 mg/L, reducing the sample size to 618 (see Land-use regression model section for addressing censored measurements).
2.2. Simulations from the coupled bidiCMAQ-EPIC modeling system

The bidiCMAQ-EPIC modeling system documentation and model code are downloadable at http://www.cmascenter.org as CMAQ and FEST-C. The BidiCMAQ air quality component emphasizes the characterization of surface/atmosphere exchange (flux, i.e., emission and deposition) of N species using an hourly output interval. The EPIC soil biochemical component estimates the inorganic and organic N and phosphorus fertilizer applied to major commercial crops grown throughout the US. The coupled bidiCMAQ-EPIC modeling system relies on EPIC estimates of fertilization timing, rate, and fertilizer type, while bidiCMAQ simulates the soil ammonium pool by conserving the ammonium mass due to fertilization, evasion, deposition, and nitrification processes. Both EPIC and bidiCMAQ were driven with the same meteorology (Weather Research and Forecast Model; Skamarock et al., 2008) and the same land use data (National Land Cover Data; Homer et al., 2007). The coupled modeling system used a 12 km × 12 km grid cell coverage across the Continental US. Grid cells with 40 acres or more of any crop type were included in the analysis.

In addition to daily fertilizer application information, bidiCMAQ requires emissions inventory inputs for non-cropland agriculture sources. Each scenario inventory was developed from a combination of the National Emissions Inventory (NEI), and the USEPA Motor Vehicle Emission Simulator (MOVES; USEPA, 2010a). NEI version 2 (USEPA, 2013) was used for simulating the 2002 scenario. The 2022BASE projection was simulated using renewable fuel volumes published by the US Energy Information Administration (EIA) in its Annual Energy Outlook (AEO) for 2007 (EIA, 2007), including a biofuel projection of 0.25 billion gallons of cellulose-based ethanol production and 12.29 billion gallons of corn starch-based ethanol. AEO 2007 represents ethanol volumes projected for 2022 in the absence of the renewable fuel requirements of the Energy Independence and Security Act (EISA) of 2007. This inventory (called AEO 2022) was used by the USEPA as the “base” 2022 case in the generation of the Renewable Fuel Standards Program (RFS2) Regulatory Impact Analysis (USEPA, 2010b). Here, however, the 2022 emissions projection has been modified to reflect our simulation domain, 2002 weather, and an updated emissions temporal profile for animal feeding operations.

The 2022CROP scenario represents a hypothetical future projected by an integrated agriculture and energy markets model (Elobeid et al., 2013) and begins by assuming a solution of 10 billion gallons of cellulosic-based fuel, and 18 billion gallons corn starch-based fuel (USEPA, 2010b). In order to translate this fuel production into emission changes (from AEO 2022), modifications were made to non-point input files for biodiesel, cellulosic, and corn ethanol plants to reflect changes in levels of oxidized N, sulfur dioxide, fine and course particulate matter as well as volatile organic carbons from ethanol transfers. The 2022CROP scenario could be considered a more highly-intensive scenario, in that in addition to high levels of corn ethanol, the cellulosic ethanol was based heavily on corn-stover removal. It should be emphasized that the 2022CROP scenario is not intended to be predictive. Any outcomes modeled using this inventory are purely hypothetical and are illustrative of one possible production trajectory.
Both 2022 scenarios include yield trends projected by the coupled markets model, emission reductions and trends in ambient CO$_2$ concentrations. Advancements in agricultural management technology and methods leading to increased yields were based on extrapolation of historical USDA reports. Future CO$_2$ concentrations were estimated using an annual growth rate of ~2.0 ppm/yr (1960-current, http://www.esrl.noaa.gov/gmd/ccgg/trends). Emissions reductions were included as prescribed by the US CAA (http://www.epa.gov/criteria-air-pollutants/naaqs-table/).

2.3. Ancillary data

Ancillary data were assigned to a buffer surrounding each drinking water well location for regression modeling, and to a 12 km rectangular grid structure (144 km$^2$) coinciding with the bidiCMAQ-EPIC modeling system for predicting each scenario. Land cover data from the National Land Cover Database 2001 (Homer et al., 2007) were classified using a modified Anderson Level 1 classification scheme (Supplementary Table 1). The areal percent contribution of land cover classes within each 1500 m well buffer and cell from the 12 km national grid were calculated using a zonal histogram function in ArcGIS. USGS principal aquifer data for the conterminous US were downloaded from the USGS National Map Small-Scale Collection (U.S. Geological Survey, 2000) and aggregated to a four-category classification scheme. Aquifer data were initially processed in the same manner as NLCD, but were later aggregated to create one categorical variable with two factor levels in order to improve model performance (Supplementary Table 2). Hydraulic conductivity ($K_{sat}$) and depth to water table in the upper 2 m of the soil profile were extracted from the USDA’s Soil Survey Geographic (SSURGO) database (2015). Both variables were averaged to the 1500 m well buffers and the 12 km national grid using a zonal statistics function. Population density data at block, block group and tract geometries from US Census Bureau (2004) were attributed directly to well locations using a spatial join process and estimated at the 12 km national grid using a spatially weighted average approach. Animal feeding operations (AFOs) were extracted from the Dun and Bradstreet Standard Industrial Classification (2013) using SIC codes 0211, 0213, 0214, 0241, 0251, 0252, 0253, 0254, 0272. These data were attributed to both the wells and the national grid based on a count of facilities within a 1000 m buffer or 12 km grid cell, respectively (see below for explanation of buffers used).

2.4. Processing of data

A literature search and sensitivity analysis was conducted to determine the appropriate radius (buffer size) surrounding each well location for assigning modeled variables and ancillary data for regression modeling. For example, Hay and Battaglin (1990) found that a buffer size of 800 to 1200 m produced a maximum correlation between land cover and nitrate-N concentrations. Tesoriero and Voss (1997) found that a 3200 m radius produced the best logistic-regression model relating land cover to nitrate-N concentration. Finally, Greene et al. (2005) found that a 1500 m radius around each well location produced the best model fit after testing 8 radii that ranged from 500 to 4000 m in 500-m steps. In this study, a buffer size of 1500 m was selected for all variables except for animal feeding operations and a buffer size of 1000 m was selected for animal feeding operations based on a sensitivity analysis of 3 buffer sizes (1500 m, 1200 m and 1000 m) applied to each spatial variable. Aggregated variables (weighted average or total) were assigned to a buffer based on percent...
(e.g., percent of land cover, aquifer type, crop type that lie within buffer zone) or totals (e.g.,
total revenue or number of animal feeding operations that lie within a buffer zone).

2.5. Land-use regression model

A generalized least squares (gls) statistical approach available in the nlme package (Pinheiro
et al., 2016) for the R Statistical Computing software environment was the primary
regression model used in this study. Because the groundwater nitrate-N measurements
(response variable) were not randomly selected and showed strong evidence of spatial
clustering, we applied a gls model with a rational quadratics spatial covariate function to
account for spatial autocorrelation (measurements related based on proximity). Selection of
this spatial autocorrelation function over other functions (e.g., spherical, exponential, cubic)
was based on examination of the semivariogram for the measurements used in the study and
model selection criteria. Other regression models, including ordinary least squares,
generalized linear, ESRI’s ArcGIS geographically weighted regression function and
censored regression models were also used to examine the stability of the results and to
check for the effect of data characteristics. For example, the Companion to Applied
Regression library Tobit model was used to examine the impact of removing censored data
removal of measurements below the instrument detection threshold of 0.05 mg/L).
Regardless of this censorship and the type of regression modeling employed, regression
model results were similar, with the same statistically significant variables as the primary gls
model used in the study.

Explanatory variables used in the study represented plausible biological pathways and
included fertilizer placed on all crop types, irrigation, tiling drainage, tilling, shallow aquifer
type, soil characteristics, depth-to-water-table, slope, population density and land cover (e.g.,
percent forested, developed and cultivated). The Variance Inflation Factor (VIF) was
calculated for all model variants to assess collinearity among the explanatory variables.
Variables were removed or appropriately combined to address collinearity. For example,
inorganic N fertilizer was combined with organic N fertilizer to address collinearity between
the two variables. Model selection was based on conducting forwards, backwards and
combined forwards and backwards variable selection and observing significance (p-value ≤
0.001), residuals, Aikake’s Information Criterion (AIC), Shwarz’s BIC, and iterative cross-
validation (k = 10; Alfons, 2012). The final model used to predict the response of the log-
transformed variable Y (NO$_3$-N concentration in groundwater) at site s was of the form

\[
\ln(Y_s) = \beta_0 + \beta_1 AqTy_s + \beta_2 PctClt_s + \beta_3 CAFO_s + \beta_4 NFerts_s + \beta_5 Ksat_s + \epsilon_s
\]

Here $\beta_0$ is the intercept and the $\beta_i$’s ($i = 1, \ldots, 5$) are the regression coefficients. AqTy is the
aquifer type (a 2-level factor), PctClt is the percent of cultivated land area within the grid
cell, AFO is the number of animal feeding operations, and NFert is the N fertilizer rate (kg-
N/ha) applied to irrigated grain corn. The error term ($\epsilon_s$) represents unexplained variance at
each site.
3. **Results and discussions**

The results and discussions are presented in the context of each study objective below:

### 3.1. Refine our understanding of N loadings and interactions related to crop fertilization and groundwater nitrate contamination

Groundwater measurements used in the regression analysis were predominantly taken in agricultural areas throughout the US. Accordingly, the results of this study are limited to agricultural areas in the US, which is consistent with our interest in the impact of increased corn production. Within this national cropland extent, the coupled modeling system allowed us to investigate associations between groundwater nitrate-N contamination and N fertilizer application in more detail than is typically feasible. Factors we considered included the type of crop (e.g., corn, wheat, soy, cotton) and sub-crop type (e.g., grain corn, silage corn), agriculture management practice (e.g. tile drainage, irrigation, fertilizer type/rate) and soil characteristics (e.g., total soil carbon, porosity, N uptake by the plant, vertical leaching, \( K_{sat} \)). For example, we first examined total N fertilizer applied to all crops and did not find a significant association with groundwater nitrate-N measurements. The subset of total organic N fertilizer placed on all corn, which comprises rainfed and irrigated grain corn (head of stalk harvested for consumption) and silage corn (whole plant harvested for feedstock) was more strongly associated. Eventually, we found the strongest N fertilizer explanatory variable to be the rate (kg-N/ha versus total N-tons applied per 144 km² grid cell) of inorganic and organic N fertilizer, applied to irrigated grain corn. Fig. 1b is a graphical representation of the difference in the domain extents for all crops, and rainfed and irrigated grain-corn crops, to demonstrate the subsetting process.

Thus, our final model included \( K_{sat} \) (Table 1; Fig. 2a), percent cultivated land cover (Table 1), a categorical variable for primary shallow aquifer (Table 1; Fig. 2b), count of animal feeding operations (Table 1; Fig. 2c), and N fertilizer placed on irrigated grain corn (Table 1; Fig. 3a). Ultimately, we subset our data for applying the land-use regression model to include only those areas with irrigated grain-corn croplands, because this produced the strongest regression model and met the needs of the study to examine the impact of corn production. Thus, 498 measurements were used in the final regression model. Fig. 4 shows the residuals (observed—fitted) of the final land-use regression model used in this study. Note that our regression approach tends to under-predict estimated groundwater nitrate-N concentrations.

Our results are consistent with past studies, but offers a more refined N fertilizer loading variable than revealed in these studies. Similar to our study, Greene et al. (2005), Nolan et al. (2002), and Nolan and Hitt (2006) used inorganic N fertilizer rates in a multi-variable regression model. These rates were derived from county-level fertilizer sales data available from the Association of American Plant and Food Control. Fertilizer sales were allocated to either “farmed” or “non-farmed” land cover classifications. Although valuable, these data contain relatively large spatial uncertainty (Cooter et al., 2012). Thus, one county-level average rate was applied to all crop types as opposed to our study, which used more spatially resolved simulated fertilizer application rates reflecting typical regional management practices, crop type and environmental conditions (e.g., time of year, precipitation, precipitation,
temperature). In our study, including this refined N-fertilizer loading variable in the land-use regression model resulted in a stronger prediction, indicating that using this refined variable will result in more targeted scenario comparisons and informed decision making.

The coefficients from this land-use regression model were applied using the 2002 coupled model variables to estimate groundwater nitrate-N concentrations for all irrigated grain-corn croplands (Figs. 2a–c and 3a). Of particular note, use of the variables from the coupled model suggests an interesting hierarchy of environmental conditions associated with high groundwater nitrate-N concentrations. For purposes of this study, “high” groundwater nitrate-N concentrations were defined as ≥5 mg/L because this level has been associated with several health effects, including some birth defects and cancers (Brender et al., 2013; Ward et al., 2005). Table 2 shows the minimum, mean and maximum values calculated across the domain for $K_{sat}$, AFOs, and N fertilizer rate. Percent unconsolidated aquifers are also shown. These metrics were calculated for all areas (144 km$^2$ grid cells) with irrigated grain-corn croplands, as well as for areas predicted to have high groundwater nitrate-N concentrations.

In examining these metrics, we found that soils with at least a $K_{sat}$ value of 3.6 μm/s underlay areas with high groundwater nitrate-N concentrations (Table 2; Fig. 2a). This minimum $K_{sat}$ value was 13% higher for high groundwater nitrate-N areas as compared to the rest of the domain. While the mean $K_{sat}$ values were also significantly higher for high groundwater nitrate-N areas, it is interesting to note that the maximum values were not significantly different. Our findings suggest that clayey soils (defined as $K_{sat}$ < 1.41 μm/s) are protective of groundwater nitrate-N contamination (none of the high groundwater concentrations were in clayey soils), as opposed to loamy or sandy soils (defined as $K_{sat}$ > 1.41 μm/s; categories defined by Soil Survey Staff, 2014). Once the minimum $K_{sat}$ value was met, high groundwater nitrate-N concentrations tended to cluster on unconsolidated aquifers (Table 2; Fig. 2b). Table 2 shows that the average percent of unconsolidated aquifers underlying areas with high groundwater nitrate-N predictions is statistically greater than the average percent of unconsolidated aquifers underlying the entire domain (63% vs. 46%). Table 2 and Fig. 2c support that many of the high groundwater nitrate-N areas that do not lie on unconsolidated shallow aquifers are located near a relatively high average number of AFOs, with a 5-fold increase between the mean values for high groundwater nitrate areas and the rest of the domain (8.6/cell vs. 1.8/cell). Finally, Table 2 suggests a 4-fold increase in the minimum N fertilizer rate between areas with high groundwater nitrate-N concentrations and the rest of the domain (85.85 kg-N/ha vs. 21.05 kg-N/ha).

Spatially, these factors are evident when examining areas vulnerable to high groundwater nitrate-N concentrations (Figs. 2a–c and 3b). For example, areas of predicted high groundwater nitrate-N concentrations in the California Central Valley are underlain by unconsolidated aquifers, and have soils with relatively high $K_{sat}$ values, a high percentage of cultivated lands and are near a relatively high density of AFOs. Two of these factors, high $K_{sat}$ values and high AFO counts, are associated with high groundwater nitrate-N concentrations in the upper Midwest. In the central Midwest, however, areas predicted to have high ground-water nitrate-N are not near high AFO counts, but are on unconsolidated aquifers, and are in areas with a high percentage of cultivated lands and high N fertilizer
rates (lower-central Midwest); or in areas with very high $K_{\text{sat}}$ values (upper-central Midwest). Conversely, areas of high N fertilization rates overlain on unconsolidated aquifers in the Southeast do not result in estimates of high groundwater nitrate-N concentrations because the warm, humid climate promotes vegetation growth in this area. The vegetation, in turn, contributes to relatively high concentrations of organic carbon in the soil, improving the nutrient holding capacity of the soil.

Compared to previous studies, our study provides unique information on N loadings associated with high groundwater nitrate-N areas (rate of fertilizer applied to irrigated grain corn at a 144 km$^2$ grid resolution, versus county inorganic fertilizer sales data). In addition, these findings suggest a hierarchical structure in the cropland characteristics associated with high groundwater nitrate-N concentrations; high groundwater nitrate-N concentrations (1) occur in areas of loamy or sandy soils (according to our model, $K_{\text{sat}} > 3.6 \mu m/s$), and not in clayey soils, (2) overlie unconsolidated aquifers or are near a high-density of AFOs, or (3) have irrigated grain corn, fertilized at a rate > 85 kg-N/ha (as calculated by our model). Understanding these characteristics is important to agricultural and environmental managers in prioritizing efforts to reduce groundwater nitrate-N contamination.

3.2. Predict changes in groundwater nitrate contamination for a base and future corn production scenario; and examine these changes to better understand the impacts of increased corn production on groundwater quality

In addition to predicting groundwater nitrate-N concentrations in 2002, the land-use regression model was applied to the 2022$_{\text{BASE}}$ and 2022$_{\text{CROP}}$ scenarios to predict respective groundwater nitrate-N concentrations. The N fertilizer rates (kg-N/ha) simulated by the bidiCMAQ-EPIC model were different for the three scenarios, but all other explanatory variables (aquifer type, percent cultivated land cover, count of AFOs and $K_{\text{sat}}$; Table 2) remained the same. We examined differences among the scenarios and among the high groundwater nitrate-N concentrations ($\geq 5$ mg/L). In investigating changes between the 2022$_{\text{BASE}}$ and 2022$_{\text{CORN}}$ scenarios (Supplementary Table 3), we calculated the difference in the averaged N-fertilizer rates for each grid cell, as well as the difference in the fertilizer rate for just the increased corn cropland within each grid cell (2022$_{\text{CORN-ONLY}}$). We then used these “biofuel-only” fertilizer rates to predict groundwater nitrate-N across our 2022 domain, but the modeling scenario remained the same for both the 2022$_{\text{CORN}}$ and 2022$_{\text{CORN-ONLY}}$ calculations. This approach was used to isolate the characteristics associated with the corn production expansion scenario from the 2022$_{\text{BASE}}$ changes.

As expected, Table 2a–b shows little difference in the overall averaged data between the scenarios for $K_{\text{sat}}$, AFOs or unconsolidated aquifers. There is evidence, however, of corn production expansion in response to increased biofuel demand, with a 27% increase in the number of grid cells projected to contain irrigated grain-corn croplands between the 2002 and 2022$_{\text{CORN}}$ domains (Table 2a; Fig. 5a–b; Supplementary Table 3). Overall, the expansion of corn crops was simulated to occur in 77% of the total number of grid cells in the domain (14,633 out of 19,078; Table 2c). This hypothetical increase in biofuel demand resulted in expansion of corn croplands onto less productive lands such as those set aside for protection (e.g., buffers along streams and rivers) and non-renewal of Conservation Reserve...
Program (CRP) contracts (Lark et al., 2015). Under our 2022\textsubscript{CROP} simulation, the growth in area of the irrigated corn domain due to the demand for biofuels often occurs in these more marginal areas, particularly in the upper and central Midwest (Figs. 1b and 5).

Despite expansion onto less productive lands, however, the minimum N fertilizer rate for areas with high groundwater nitrate-N concentrations (≥ 5 mg/L) is projected to increase from 75\% (2022\textsubscript{BASE}) to 79\% (2022\textsubscript{CORN}) as compared to areas of high nitrate-N concentrations in 2002 (Table 2b). In examining the biofuels-only rate calculation, the increase is as high as 84\% (2022\textsubscript{CORN-ONLY} Table 2c). The number of grid cells predicted to have high groundwater nitrate-N concentrations was also projected to rise between 56\% (2022\textsubscript{BASE}) and 79\% (2022\textsubscript{CORN}), and as high as 91\% for the biofuels-only calculations (2022\textsubscript{CORN-ONLY}; Table 2b–c; Fig. 6). In comparing the characteristics of areas with high groundwater nitrate-N, the biofuels-only areas had higher N-fertilizer rates, but a lower percent of unconsolidated aquifers and lower maximum AFO counts (Table 2c), i.e., expansion into areas with lower historical risk. In addition, we found that high groundwater nitrate-N areas attributable to the increased corn demand scenario did not occur in new grid cells of expanded irrigated grain-corn croplands because of their relatively lower fertilizer rates and lower percent unconsolidated aquifers (Fig. 6b). Instead, they cluster around existing hotspots where existing corn crops were expanded to meet increased demand. These findings indicate that areas already prone to high groundwater nitrate-N concentrations could expand in 2022 as compared to 2002, and that this increase could be greater with demand for biofuels (in our simulations, 18\%–24\% of the increase was due to biofuel demand; Supplementary Table 3). These increases are projected regardless of expansion onto less productive lands and other factors simulated for 2022 that are anticipated to decrease N fertilizer demand. For example, increasing atmospheric CO\textsubscript{2} projected in the future scenarios allows corn to grow more efficiently, increasing yield and reducing the need for N fertilizer. This factor, along with agricultural management technology advancements, offset the demand for increased corn production and associated N fertilizer from growth in population and higher biofuel targets. Thus, the response of the environment to changing N loadings is non-linear and complex, and the coupled model is needed to simulate these opposing effects to holistically assess the impact of protective actions. In addition, models are needed to predict future conditions where observations are unavailable.

4. Conclusion

In this study, we demonstrated the value of using the coupled bidiCMAQ-EPIC modeling system to examine environmental variables associated with groundwater nitrate-N contamination, as well as to predict future impacts of increased corn production on groundwater nitrate-N concentrations. In our land-use regression analysis, we used the model output for the N-loading explanatory variable, and found that the rate of fertilizer (versus total fertilizer) applied to irrigated (versus rainfed) grain corn (versus silage corn or other crops) was the strongest N-loading predictor of groundwater nitrate-N concentrations for areas (144 km\textsuperscript{2}) across the US with > 40 acres of croplands. Previous studies have also shown N fertilizer is a statistically significant predictor of groundwater nitrate-N concentrations but these studies used total county-level fertilizer sales as a surrogate for N fertilizer loadings. The use of the coupled model variables allowed us to examine a full suite
of factors directly and indirectly related to N fertilizer and its infiltration into groundwater, including agriculture management practices (e.g., tilling, tiling drainage, irrigation, fertilizer type and rate) and soil conditions (e.g., total organic carbon, vertical and horizontal flow of excess N, N yield rates). Our findings also point to a hierarchical structure associated with areas of high groundwater nitrate-N concentrations (≥ 5 mg/L). Areas prone to high groundwater nitrate-N concentrations occurred in loamy and sandy soils (as opposed to clayey soils) with a $K_{sat}$ value > 3.6 μm/s, in combination with overlying unconsolidated aquifers, occurring near a relatively high number of AFOs, or having a fertilizer rate N 86 kg-N/ha (as calculated by our models). Areas with high groundwater nitrate-N had minimum $K_{sat}$ values that were 13 times higher than the entire irrigated grain-corn domain. Similarly, the mean number of AFOs for high groundwater nitrate-N areas was 5 times higher and the minimum N fertilizer rate was 4 times higher as compared to the entire domain.

In addition, we applied the coupled model and the land-use regression approach to predict groundwater nitrate-N concentrations from two future corn production scenarios (2022$_{\text{BASE}}$ scenario simulating technology advancements and atmospheric carbon dioxide increases; and 2022$_{\text{CROP}}$ scenario simulating these changes as well as additional corn production increases to support biofuel production). These simulations revealed an estimated 27% increase in irrigated grain-corn crop-land areas (144 km$^2$ grid cells) resulting from the 2022 biofuel scenario. This expansion occurs in areas of less productive soils, particularly in the upper Midwest and Texas. In addition, our simulations suggest that corn production between 2002 and 2022 could result in a 56% (2022$_{\text{BASE}}$) to 79% (2022$_{\text{CROP}}$) increase in the number of 144 km$^2$ grid cells projected as having high groundwater nitrate-N concentrations. The characteristics of these same areas indicate that the minimum fertilizer rate is an important factor (over percent aquifer and density of AFOs), in projecting high groundwater nitrate-N concentrations resulting from the increased corn demand scenario. In summary, the coupled bidiCMAQ-EPIC modeling system provides additional, more spatially resolved information regarding N fertilizer loadings leading to nitrate contamination of groundwater, and facilitated the estimation of possible changing levels of groundwater nitrate contamination under alternative future corn production scenarios.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgements**

We wish to thank Jesse Bash for his contribution to the modeled data used in the study and Brandon Hayes who conducted data analysis and worked on the regression modeling. The views expressed in this article are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**References**


Brender JD, Weyer PJ, Romitti PA, Mohanty BP, Shinde MU, Vuong AM, et al., 2013 Prenatal nitrate-N intake from drinking water and selected birth defects in offspring of participants in the National Birth Defects Prevention study. Environ. Health Perspect. 121, 1083–1089.23771435


Dun and Bradstreet Standard Industrial Classification, 2013.


## HIGHLIGHTS

- Corn ethanol demands can increase corn production in 2022.
- Nitrogen (N) from fertilizer can leach into groundwater causing health impacts.
- Coupled models were used to regress on measurements and project increased corn production impacts.
- The rate of N fertilizer placed on *irrigated grain* corn was the strongest N-loading predictor.
- Our scenario resulted in a 56%–79% increase in areas with high groundwater nitrate.
Fig. 1.
Location of groundwater nitrate well locations overlain on total N fertilizer placed on all corn crops in 2002 (panel a); extent of rainfed versus irrigated grain-corn domains for 2002 and 2022_{BASE} (same extent), and 2022_{CORN} scenarios (panel b).
Fig. 2.
High groundwater nitrate measurements and predictions overlain on soil hydraulic conductivity (panel a), aquifer type (panel b), and high density animal feeding operations (panel c).
Fig. 3.
N fertilizer rate predicted by coupled modeling system for 2002 (panel a), and for the 2022\textsubscript{CORN} crop expansion scenario (panel b).
Fig. 4.
Residuals (observed-fitted) resulting from generalized least squares model prediction.
**Fig. 5.**
N fertilizer rate predicted by coupled modeling system for the 2022$_{CORN}$ expansion scenario (panel a), and the change between scenarios (2022$_{CORN}$—2002; panel b).
Fig. 6.
Predictions of groundwater nitrate-N overlain with above-threshold grid cells for the 2022\textsubscript{CORN} (panel a). Change in Above-threshold groundwater nitrate-N added by each scenario. Magenta squares are cells above-threshold in 2002, visible yellow squares are grid cells added by the 2022\textsubscript{BASE} scenario, and black cells are added by the 2022\textsubscript{CORN} scenario (panel b).
Table 1

Metrics from regression of groundwater nitrate-N concentrations on independent variables used in generalized least squares model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.581</td>
<td>0.231</td>
<td>0.012</td>
</tr>
<tr>
<td>Aquifer type</td>
<td>-0.034</td>
<td>0.144</td>
<td>0.810</td>
</tr>
<tr>
<td>Percent cultivated</td>
<td>0.012</td>
<td>0.003</td>
<td>≤0.001</td>
</tr>
<tr>
<td>AFO count</td>
<td>0.024</td>
<td>0.006</td>
<td>≤0.001</td>
</tr>
<tr>
<td>N fertilizer irrig. grain corn</td>
<td>0.003</td>
<td>0.001</td>
<td>≤0.001</td>
</tr>
<tr>
<td>$K_{sat}$</td>
<td>0.014</td>
<td>0.003</td>
<td>≤0.001</td>
</tr>
</tbody>
</table>

*a* Categorical variable; unconsolidated (default) and ‘all other’.
Table 2

Summary statistics of independent variables.

<table>
<thead>
<tr>
<th></th>
<th>( K_{sat} ) (Mm/sec)</th>
<th>Animal feeding operations (count)</th>
<th>Acquifer type (percent)</th>
<th>N fertilizer rate (kg/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Mean</td>
<td>Max</td>
<td>Min</td>
</tr>
<tr>
<td>a. All irrigated grain corn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002 domain (15,035 cells)</td>
<td>0.27</td>
<td>19.53</td>
<td>106.10</td>
<td>0.00</td>
</tr>
<tr>
<td>2022BASE domain (15,035 cells)</td>
<td>0.27</td>
<td>19.54</td>
<td>106.10</td>
<td>0.00</td>
</tr>
<tr>
<td>2022CORN domain (19,078 cells)</td>
<td>0.27</td>
<td>18.70</td>
<td>106.10</td>
<td>0.00</td>
</tr>
<tr>
<td>b. Groundwater NO3 &gt; 5 mg/L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002 above threshold (250 cells)</td>
<td>3.57</td>
<td>49.51</td>
<td>97.14</td>
<td>0.00</td>
</tr>
<tr>
<td>2022BASE above threshold (389 cells)</td>
<td>3.57</td>
<td>46.48</td>
<td>97.14</td>
<td>0.00</td>
</tr>
<tr>
<td>2022CORN above threshold (445 cells)</td>
<td>3.57</td>
<td>45.70</td>
<td>97.14</td>
<td>0.00</td>
</tr>
<tr>
<td>c. Areas of increase corn production only(^d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022CORN-ONLY (14,633 cells)</td>
<td>0.27</td>
<td>18.55</td>
<td>97.14</td>
<td>0.00</td>
</tr>
<tr>
<td>2022CORN-ONLY above threshold (86 cells)</td>
<td>7.45</td>
<td>41.93</td>
<td>90.42</td>
<td>0.00</td>
</tr>
</tbody>
</table>

\(^d\) 2022CORN-ONLY represents only those grid cells that had an increase in the N fertilizer rate because of the biofuel demand scenario.