Modeling the Impact of CAFOs on Nitrate Contamination in the Middle Trinity Aquifer of Central Texas

Kartik Venkataraman[1], Jesse Crawford[2], Keith Emmert[3]

[1] Asst. Professor, Department of Engineering and Computer Science, Box T-0390, Tarleton State University, Stephenville, TX 76402. venkataraman@tarleton.edu
[2] Asst. Professor, Department of Mathematics, Tarleton State University, Stephenville, TX 76402. jcrawford@tarleton.edu
[3] Assoc. Professor, Department of Mathematics, Tarleton State University, Stephenville, TX 76402. emmert@tarleton.edu

Abstract

Elevated levels of nitrate, well in excess of drinking water standards, occur in the Middle Trinity aquifer underlying Central Texas. Agricultural businesses such as Concentrated Animal Feed Operations (CAFOs) are common in this region; improper management of waste, particularly manure, at these sites has resulted in the impairment of water resources by nutrients such as nitrogen. While several studies have investigated the impact of such chemicals released from CAFOs on the Bosque River, there have been limited attempts to characterize the vulnerability of the Middle Trinity aquifer, a major source of drinking water to the largely rural community.

In this study, we have evaluated the risk of groundwater nitrate contamination via logistic regression techniques coupled with a GIS framework. Model inputs included hydrogeologic, geographic, land-use and climate data compiled from a variety of local and state sources. A unique aspect of this study is the development of a feature that captures the potential migration pathways from CAFOs by considering the gradient of the water table, proximity of CAFOs to drinking-water wells, and permitted waste application-rates at CAFOs. A logistic regression model for predicting nitrate contamination based on this feature demonstrated a statistically significant relationship ($p < 10^{-11}$), and the area under the ROC curve was 0.77. Future models will incorporate additional hydrogeologic variables such as depth to water table, soil properties, and annual rainfall. The results of this study highlight the need for promoting best management practices (BMPs) to minimize the impact of nutrients from CAFOs on natural ecosystems and human health.

Introduction

Groundwater is a vital natural resource that supplies over half of the water needs in the state of Texas (Texas Water Development Board [TWDB], 2012). While most of the groundwater is used for irrigation, a vast majority of rural areas and small municipalities in the state rely on groundwater for domestic consumption needs as well. Protection of the aquifers that serve these communities is hence critical in the interest of public health. Groundwater contamination by nitrate has been widely acknowledged as one of the most common environmental issues in the United States, particularly in agricultural watersheds (Lockhart et al., 2013). Elevated levels of nitrate in water supplies can cause deleterious health effects such as methemoglobinemia (blue-baby disease) in infants, and spontaneous abortions and non-Hodgkins lymphoma in adults (Nolan, 2001); consequently, the United States Environmental Protection Agency has set a Maximum Contaminant Level (MCL) of 10 mg/L (as nitrogen) to protect public health. Nitrate is a component of the global nitrogen cycle and occurs naturally in the environment as a result of various microbiological and redox processes. While the background concentrations of nitrate in groundwater exhibit great geographic variability and may even be difficult to quantify in some regions, concentrations of nitrate in waters unaffected by human activities have been shown to be well below the MCL (Burkartaus and Stoner, 2008). It is generally accepted that concentrations of nitrate in excess of 3 mg/L (as nitrogen) may be indicative of human inputs, particularly agricultural. Aquifers impacted by agricultural activities such as Concentrated Animal Feed Operations (CAFOs) may face a higher risk of nitrate contamination due to (a) increased nitrogen loading resulting from application of nitrogen-based fertilizers and manure and (b) increased leaching due to irrigation. Leaching of nitrate from poorly-managed dairy farms, ponds and CAFOs has resulted in significant nitrate contamination in several states in the United States such as California, Iowa and Texas (Harter et al., 2002; Burkartaus and Stoner, 2008; Rekha et al., 2011). With livestock production becoming increasingly spatially concentrated, manure loading rates at these sites often reach or exceed the natural absorptive capacity of the land. In many farms, land application of manure is practiced simply to reduce disposal costs rather than to serve as a nutrient for crops. CAFOs are also often located away from major population centers, where water supplies may be decentralized in nature and at a higher risk due to lack of adequate infrastructure or resources needed for monitoring and treatment.

In the state of Texas, the location of CAFOs and the amount of manure, sludge or wastewater they dispose into natural water bodies are regulated by the Texas Commission on Environmental Quality (TCEQ). CAFO operators are also required to routinely monitor all discharges into surfacewater bodies and establish and periodically update a nutrient management plan (NMP). However, a monitoring plan for groundwater is mandated only when a playa is used as a retention control structure (RCS) or if required by the TCEQ. The effect of nutrient management at CAFOs on water quality in the Bosque River watershed, particularly with regards to phosphate, has been well-documented (e.g., McFarland and Hauck, 1997; McFarland and Adams, 2007). McFarland and Hauck (1997) suggest that the largest contributing source of phosphate to the Bosque River was dairy waste application fields while row-crop fields contributed significant nitrogen loads. However, with the exception of a few vulnerability assessment studies such as Fritch et al. (1999), little information exists about the impact of land use on the quality of groundwater underlying the Bosque River watershed.

Considering that several potential sources of nitrate exist in this region, it is imperative to evaluate the impact and significance of CAFOs on groundwater nitrate occurrence to better manage water disposal at these sites. However, monitoring groundwater quality on regional scale, particularly when non-point sources (NPS) such as CAFOs are involved can often be extremely difficult and expensive. It has been shown that statistical hydrologic models integrated within a Geographic Information Systems (GIS) framework are a suitable alternative as they can produce scientifically-reliable information on risk-based management for natural resource planners (Twarakavi and Kaluarachchi, 2005; Venkataraman and Uddameri, 2012). In this study, one such technique, Logistic Regression (LR), has been used to test the associations between groundwater nitrate occurrence and CAFOs using several hydrogeologic, climatologic and land-use variables. A novel aspect of this study is the development of a parameter that captures the effect of migration of nutrients from a CAFO by incorporating the hydraulic gradient, the proximity of CAFOs to drinking-water wells and the permitted discharge rates at these CAFOs. As such, this study represents the first step towards a more thorough and comprehensive evaluation of groundwater quality in this area.

Study Area Characteristics

The study area comprises 17 counties in a rural part of north-central Texas as shown in Figure 1. This area is underlain by the Middle Trinity Aquifer, a leaky semi-confined system comprised of the Twin Mountains, Glen Rose and Paluxy Formations of the Cretaceous Comanche Series (Muller and McCoy, 1987; Mace et al., 2000). The western boundary of the study area represents the outcrop (recharge area) of the Trinity; the aquifer is unconfined throughout the study area and is thus vulnerable to contamination from sources originating from the land surface. Groundwater generally flows towards the southeast where the aquifer dips underground and becomes confined. Agriculture is the predominant land use (~ 66% of the area), as shown in Figure 1. Currently, 86 CAFOs are permitted to operate in the area, mostly clustered around the aquifer outcrop in Comanche and Erath counties (Figure 1). Information about permitted waste application rates at these CAFOs was obtained from the Texas Commission on Environmental Quality ([TCEQ], 2014). At these sites, solid manure is generally applied at the land surface while liquid manure is usually dispersed with a variety of irrigation systems. Water quality data for wells located in the study area were obtained from the TWDB Groundwater Database (TWDB, 2014). In all, 344 wells with the most comprehensive set of water quality and hydrologic information and minimal sampling bias were selected. Many of these wells are sampled frequently by the TWDB and hence the mean value of a parameter across all sampling events was chosen for the study. Nitrate concentration in approximately 8% of these wells (27 out of 344) exceeds the MCL (Figure 1), while 23% (78 out of 344) have nitrate concentrations in excess of 3 mg/L, often used as an indicator of anthropogenic inputs. In addition to nitrate levels, the depth to the water table (DWT) was compiled for all wells and regionalized using GIS-based kriging to develop the local hydraulic gradient. To characterize the general water chemistry, pH, Total Dissolved Solids (TDS) and major ion concentrations were also compiled for all wells. Trilinear diagram analysis indicated that the groundwater samples are chemically dominated by bicarbonate; shallow wells (DWT < 10 m) are characterized by $Ca\text{-}HCO_3$ type waters, suggesting these wells may have been recently recharged.

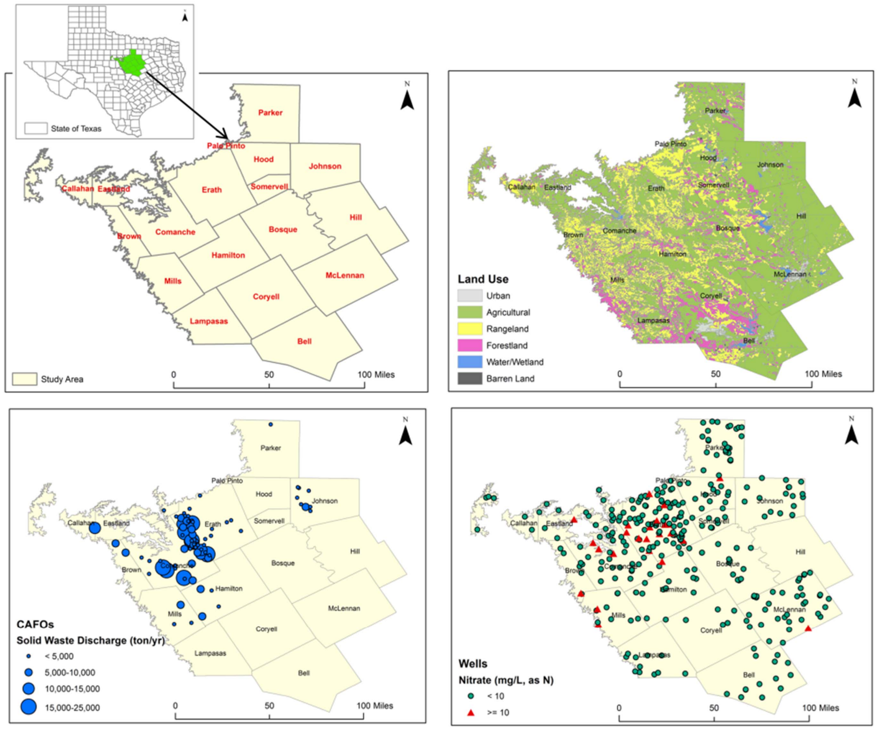

Figure 1. Study Area Location (top left), Land Use/Land Cover (top right), Permitted Discharge at CAFOs (bottom left) and Nitrate Levels at Water Wells (bottom right)

Development of the CAFO Migration Score (CMS)

The relationship between nitrate occurrence at a well and its distance to the nearest CAFO and the DWT are shown in Figure 2. It is evident that the median nitrate levels in wells located within 2 miles (3.2 km) of a CAFO are higher than those located further away; likewise, median nitrate levels in wells shallower than 33 ft (10 m) are higher than those in deeper wells. The statistical significance of the difference in these respective samples was confirmed using the Wilcoxon rank sum test ($p < 0.001$ in both instances). The association between CAFOs and nitrate contamination in water wells can be explored further by modeling potential flow paths in the aquifer from each CAFO using the hydraulic gradient. Each well is then assigned a CAFO migration score (CMS), which is a weighted sum incorporating the well's distance to each CAFO's flow path, the lengths of those flow paths, and the permitted waste application rates at the CAFOs. In the following discussion, we have assumed that nitrate is conserved during transport from the land surface to the water table and its migration thereafter is entirely advective in nature. Although natural processes may alter nitrate levels during subsurface transport, this scenario likely presents the greatest risk to a consumer at a water well downstream from a potential source.

For $i = 1, \ldots, 344$, let $w_i = (w_{i1}, w_{i2})$ denote the ith well's location, where w_{i1} and w_{i2} are its longitude and latitude in decimal degrees. The location of the jth CAFO is similarly represented by $c_j = (c_{j1}, c_{j2})$, for $j = 1, \ldots, 86$. The geodesic distance between two locations w and c is then approximated using the haversine formula,

$$d(w,c) = 2R\sin^{-1}\left(\sqrt{\sin^2\left(\frac{\pi(c_2-w_2)}{360}\right) + \cos\left(\frac{\pi w_2}{180}\right)\cos\left(\frac{\pi c_2}{180}\right)\sin^2\left(\frac{\pi(c_1-w_1)}{360}\right)}\right),$$

where R is the average radius of the Earth in km.

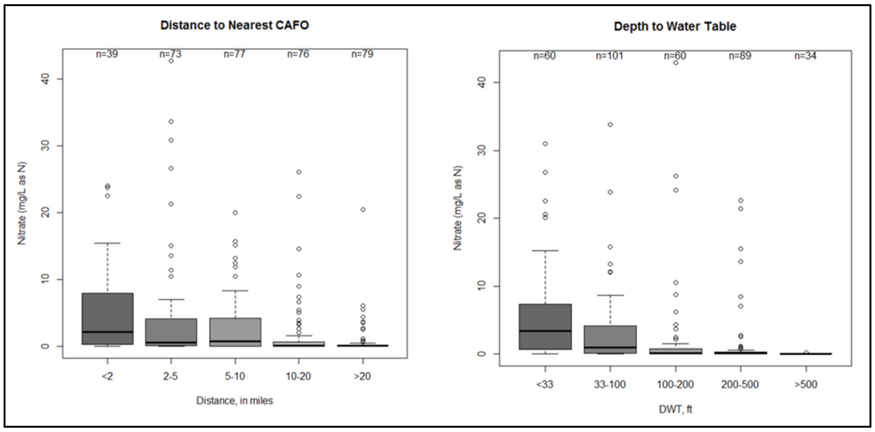

Figure 2. Impact of Proximity to a CAFO (left) and Depth to Water Table (right) on Nitrate Occurrence.

Construction of the CAFO flow paths requires the direction of the hydraulic gradient $\nabla h(c)$ at various locations c in the study area. More specifically, let $\theta(c)$ be the angle in radians between $\nabla h(c)$ and a vector pointing North, where $\theta(c)$ increases in the clockwise direction (see Figure 3). Gradient angles were obtained using GIS for a grid (a raster model of the study area) of locations with 348 rows, 466 columns; the Southwest corner of the grid has the coordinates $(-100.15°, 30.43°)$ and the horizontal and vertical spacing between grid points (λ) are equal to 0.00872°. The dimensions of this grid are $338\text{ km} \times 390\text{ km}$, the average distance between horizontally adjacent points is $\Delta x = 0.84$ km, and the average distance between vertically adjacent points is $\Delta y = 0.97$ km .We are now prepared to describe the construction of the CAFO flow paths. The idea is to start at a CAFO c_j and to follow the hydraulic gradient from that location, resulting in a flow path consisting of grid points $c_j^{(k)}, k = 0, \ldots, N_j$. The flow path is extended until it enters a cycle or leaves the study area. The gradient angle $\theta(c_j^{(k)})$ determines the next grid point $c_j^{(k+1)}$, e.g., if $\theta(c_j^{(k)}) = \frac{\pi}{2}$, then $c_j^{(k+1)}$ is the grid point East of $c_j^{(k)}$, and if $\theta(c_j^{(k)}) = \phi := \tan^{-1}\left(\frac{\Delta x}{\Delta y}\right)$, then $c_j^{(k+1)}$ is the grid point Northeast of $c_j^{(k)}$. Of course, it is unlikely that the gradient will point directly from one grid point to another, so it is necessary to discretize the gradient angle θ. More precisely, let $c_j^{(0)}$ be the grid point closest to c_j, and recursively define $c_j^{(k+1)} = c_j^{(k)} + \Delta c_j^{(k)}$, where $\Delta c_j^{(k)}$ is determined by $\theta(c_j^{(k)})$ according to Table 1.

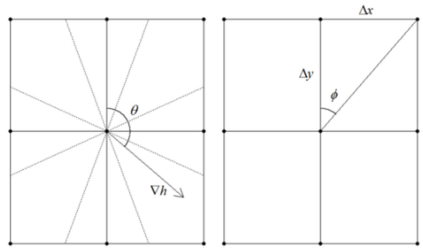

Figure 3. Hydraulic gradient and other important features of the GIS raster model.

The interval boundaries in Table 1 are simply midpoints between angles that lead directly to neighboring grid points, and they are illustrated by dotted lines in Figure 3. For example, $\theta_1 = \phi$ and $\theta_2 = \frac{\pi}{2}$ correspond to Northeastern and Eastern transitions to neighboring grid points, respectively, and their midpoint $\frac{\theta_1 + \theta_2}{2} = \frac{\pi}{4} + \frac{\phi}{2}$ is the boundary between the Northeastern and Eastern transitions given in Table 1.

Table 1. Gradient Angle Discretization.

Interval containing $\theta(c_j^{(k)})$	$\Delta c_j^{(k)}$	Interval containing $\theta(c_j^{(k)})$	$\Delta c_j^{(k)}$
$\left[0, \frac{\phi}{2}\right] \cup \left(2\pi - \frac{\phi}{2}, 2\pi\right]$	$(0, \lambda)$	$\left(\pi - \frac{\phi}{2}, \pi + \frac{\phi}{2}\right]$	$(0, -\lambda)$
$\left(\frac{\phi}{2}, \frac{\pi}{4} + \frac{\phi}{2}\right]$	(λ, λ)	$\left(\pi + \frac{\phi}{2}, \frac{5\pi}{4} + \frac{\phi}{2}\right]$	$(-\lambda, -\lambda)$
$\left(\frac{\pi}{4} + \frac{\phi}{2}, \frac{3\pi}{4} - \frac{\phi}{2}\right]$	$(\lambda, 0)$	$\left(\frac{5\pi}{4} + \frac{\phi}{2}, \frac{7\pi}{4} - \frac{\phi}{2}\right]$	$(-\lambda, 0)$
$\left(\frac{3\pi}{4} - \frac{\phi}{2}, \pi - \frac{\phi}{2}\right]$	$(\lambda, -\lambda)$	$\left(\frac{7\pi}{4} - \frac{\phi}{2}, 2\pi - \frac{\phi}{2}\right]$	$(-\lambda, \lambda)$

Figure 4 shows the flow path for a CAFO c_j, a nearby well w_i, and the point $c_j^{(k_{ij})}$ on the flow path that is nearest to the well. The CAFO migration score (CMS) defined below depends on the distance d_{ij} between w_i and $c_j^{(k_{ij})}$ and the arc length L_{ij} of the flow path from the CAFO c_j to $c_j^{(k_{ij})}$, that is,

$$k_{ij} = \arg\min_k d(w_i, c_j^{(k)})$$

$$d_{ij} = d(w_i, c_j^{(k_{ij})}) = \min_k d(w_i, c_j^{(k)})$$

$$L_{ij} = \sum_{k=1}^{k_{ij}} d(c_j^{(k-1)}, c_j^{(k)}) \ .$$

Denoting the permitted waste application rate at CAFO c_j by A_j, the CMS for the well w_i is defined by

$$\text{CMS}_i = \sum_{j=1}^{86} \frac{A_j K(d_{ij})}{1 + L_{ij}^{\gamma}},$$

where $\gamma = 2$, and K is the Epanechnikov kernel with bandwidth $b = 100 \text{ km}$,

$$K(d) = \begin{cases} \frac{3}{4}\left[1 - \left(\frac{d}{b}\right)^2\right], \text{ for } d \leq b \\ 0, \text{ otherwise.} \end{cases}$$

A CAFO will only contribute to the CMS of wells within the bandwidth of the corresponding flow path, and the contribution will decrease as the distance d_{ij} of the well from the flow path increases. A CAFO's contribution to a well's CMS value is also inversely related to the corresponding flow path arc length, L_{ij}. The Epanechnikov kernel is an example of a kernel function that is used in classification algorithms, such as weighted *k*-nearest neighbors (Hechenbichler and Schliep, 2004). The logistic regression model discussed in the following section is based on these choices for the kernel function K and the constants b and γ, but all of these parameters will be tuned to optimize the performance of future models.

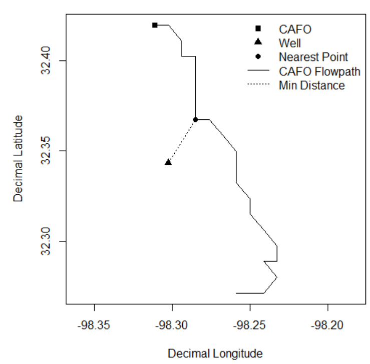

Figure 4. Sample CAFO Flow Path.

Logistic Regression Model for Nitrate Occurrence

A statistical model for predicting nitrate occurrence in water wells based on CAFO migration scores will now be presented. As mentioned in the introduction, the Maximum Contaminant Level for nitrate is 10 mg/L as nitrogen, but concentrations exceeding 3 mg/L may be indicative of human inputs. We define $Y_i = 1$ if the *i*th well has a concentration of nitrate (as nitrogen) exceeding 3 mg/L, and $Y_i = 0$ otherwise. The dependent variable Y is dichotomous, and the independent variable, CMS, is quantitative. A classification algorithm that is often effective in this setting is logistic regression (Hosmer et al., 2013). The logistic regression model postulates that the probability that $Y_i = 1$ is the logistic function of the linear predictor $g_i = \beta_0 + \beta_1 \mathrm{CMS}_i$, that is, $\Pr[Y_i = 1] = \frac{1}{1+e^{-g_i}}$. The model summary is provided in Table 2, and the estimate for β_1 is positive and statistically significant ($p < 10^{-11}$). A decile plot for the model is shown in Figure 5, where the wells have been divided into ten deciles based on CMS, represented by the ten points in the plot. The x-coordinate of each point is the average value of CMS for that decile, the y-coordinate is the observed proportion of wells in that decile with nitrate concentrations exceeding 3 mg/L, and the regression curve shows the predicted probabilities of nitrate concentration exceeding 3 mg/L for each value of CMS. Before drawing any inferences from the logistic regression model, we turn our attention to assessing its performance in terms of discrimination and goodness of fit.

Table 2. Logistic Regression Model Summary.

Parameter	Estimate	Std. Error	z value	Pr(>|z|)
β_0	-2.858E+00	3.074E-01	-9.296E+00	1.460E-20
β_1	1.650E-05	2.407E-06	6.854E+00	7.187E-12

Using the regression model, a new well w_i can be classified as follows based on its predicted probability of nitrate occurrence, $\hat{p}_i$, and a chosen probability threshold, p_0

$$\hat{Y}_i = \begin{cases} 1, \text{ if } \hat{p}_i \geq p_0 \\ 0, \text{ if } \hat{p}_i < p_0. \end{cases}$$

The choice of the probability threshold will always involve a compromise between maximizing the true positive rate and minimizing the false positive rate. Performance of the regression model can be assessed with the receiver operating characteristic (ROC) curve, which shows the tradeoff between the true positive rate (sensitivity) and the false positive rate $(1-\text{specificity})$ for different values of the probability threshold (Figure 5).

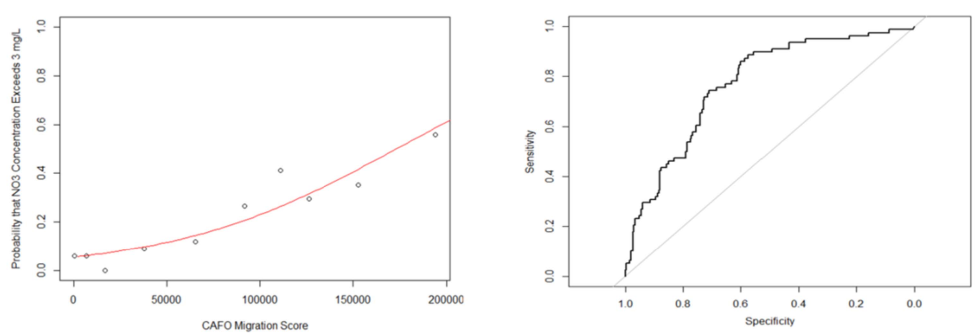

Figure 5. Decile Plot (left) and ROC Curve (right) for the Logistic Regression Model.

The area under the curve (AUC) for this model is 0.769, which is the proportion of the time $\hat{p}_{i_1} > \hat{p}_{i_2}$ when $Y_{i_1} = 1$ and $Y_{i_2} = 0$. In other words, the AUC measures the model's capability to discriminate between wells with nitrate contaminations that are above or below the 3 mg/L threshold. Values of AUC between 0.7 and 0.8 are considered to provide acceptable discrimination. In comparison, a perfect model would have $\mathrm{AUC} = 1$, and a useless model obtained from random guessing would have $\mathrm{AUC} = 0.5$. Another method for evaluating the model is the Hosmer-Lemeshow goodness of fit test, which tests the null hypothesis that the logistic regression model adequately fits the data. The *p*-value obtained from this test was 0.43, so the null hypothesis of an adequate logistic regression model was not rejected. Returning to the model summary in Table 2, the positive estimate for β_1 provides statistically significant evidence that higher values of CMS are associated with increased probabilities of nitrate concentrations exceeding 3 mg/L, as confirmed by the generally increasing trend in the decile plot.

Summary and Conclusions

The assessment of the effect of agricultural waste management practices on water quality in the Middle Trinity Aquifer is a critical natural resource and public health issue. Most drinking-water supplies in this region are served through individual wells. Extensive agriculture, particularly livestock operations, is practiced in and around the aquifer outcrop area. However, few programs currently exist that routinely monitor private groundwater systems for agricultural nitrogen contamination. In order to better understand the relationship between waste application at CAFOs and nitrate occurrence in groundwater, a novel parameter, the CAFO migration score was developed and its relationship with groundwater nitrate levels was evaluated using logistic regression. Overall, these preliminary results of the study show that higher CAFO migration scores are associated with higher probabilities of nitrate occurrence in water wells, suggesting that nutrient management practices at these sites likely represent significant nutrient inputs into the environment. There are several options for exploring the relationship in more depth. Additional hydrogeologic variables should be included in the model, and the parameters K, γ, and b used to define CMS should be tuned to optimize model performance. Other classification techniques, such as decision trees, support vector machines, and neural networks should also be explored in this context. The integration of other fate and transport processes, both in the unsaturated and saturated zones will provide water resource managers with a Decision Support System for characterizing the spatial distribution and behavior of nitrate and aid in the risk-based management of rural water supplies.

Acknowledegements

This research was made possible through support provided by the Tarleton State University Organized Research Grant No. 150325-51704. The views expressed herein are those of the authors and do not necessarily reflect on views of Tarleton State University.

References

Burkartaus, D.M.R., and Stoner, J.D. (2008). "Nitrogen in Groundwater Associated with Agricultural Systems." In Nitrogen in the Environment: Sources, Problems

and Management; ed. J.L. Hatfield and R.F. Follett. Publications from USDA-ARS/UNL Faculty, Paper 259.
Harter, T., Davis, H., Matthews, M.C., and Meyer, R.D. (2002). "Shallow Groundwater Quality on Dairy Farms with Irrigated Forage Crops." J. Contam. Hydrol., 55, 287-315.
Hechenbichler K. and Schliep K.P. (2004). "Weighted *k*-Nearest-Neighbor Techniques and Ordinal Classification". Discussion Paper 399, SFB 386, Ludwig-Maximilians University Munich http://www.stat.uni-muenchen.de/sfb386/papers/dsp/paper399.ps (Dec 20, 2014).
Hosmer, D., Lemeshow, S., and Sturdivant, R. (2013). *Applied Logistic Regression, 3rd ed.* John Wiley and Sons, Inc., Hoboken, NJ.
Lockhart, K.M., King, A.M., and Harter, T. (2013). "Identifying Sources of Groundwater Nitrate Contamination in a Large Alluvial Groundwater Basin with Highly Diversified Intensive Agricultural Production." J. Contam. Hydrol., 151, 140-154.
Mace, R.E., Chowdhury, A.H., Anaya, R., and Way, S.C. (2000). "A Numerical Groundwater Flow Model of the Upper and Middle Trinity Aquifer." Hill Country Area: Texas Water Development Board Open File Report 00-02.
McFarland, A., and Adams, T. (2007). "Semiannual Water Quality Report for the North Bosque River Watershed." Texas Institute for Applied Environmental Research Technical Report 0701.
McFarland, A. and Hauck, L. (1997). "Livestock and the Environment: A National Pilot Project Report on Stream Water Quality in the Upper North Bosque River Watershed." Texas Institute for Applied Environmental Research Progress Report 97-03.
Muller, D.A., and McCoy, J.W. (1987). Groundwater Conditions of the Trinity Group Aquifer in Western Hays County. Texas Water Development Board.
Nolan, B.T. (2001). "Relating Nitrogen Sources and Aquifer Susceptibility to Nitrate in Shallow Ground Waters of the United States." J. Groundwater, 39(2), 290-299.
Rekha, P., Kanwar, R.S., Nayak, A.K., Hoang, C.K., and Pederson, C.H. (2011). "Nitrate Leaching to Shallow Groundwater Systems from Agricultural Fields with Different Management Practices." J. Environ. Monit., 13, 2550-2558.
Texas Commission on Environmental Quality (2014). Water Quality General Permits Search. http://www2.tceq.texas.gov/wq_dpa/index.cfm#permit_types (Mar 06 2014).
Texas Water Development Board (2014). TWDB Groundwater Database Reports. http://www.twdb.state.tx.us/groundwater/data/gwdbrpt.asp (Mar 06 2014).
Texas Water Development Board (2014). *Water for Texas – 2012 State Water Plan*. Texas Water Development Board, Austin, TX.
Twarakavi, N.K.C., and Kaluarachchi, J.K. (2005). "Aquifer Vulnerability Assessment to Heavy Metals using Ordinal Logistic Regression." Groundwater, 43, 200-214.
Venkataraman, K., and Uddameri, V. (2012). "Modeling Simultaneous Exceedance of Drinking-water Standards of Arsenic and Nitrate in the Southern Ogallala Aquifer using Multinomial Logistic Regression." J. Hydrol., 458-459, 16-27.