Data Mining Analysis of Nitrate Occurrence in Ground Water & Overview of Stormwater Best Management Practices for Nitrate Control

Andy O'Reilly, USGS
Marty Wanielista, UCF
Ni-Bin Chang, UCF

USGS science for a changing world

May 3, 2007

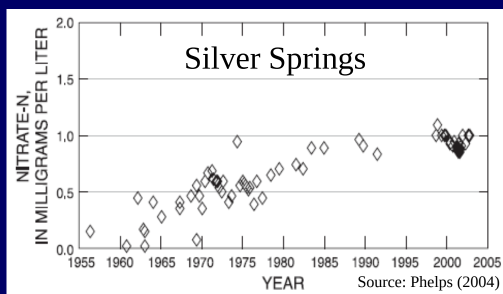
UCF

PARTNERS

- Marion County
- Florida Department of Environmental Protection
- Southwest Florida Water Management District
- St Johns River Water Management District

PROBLEM

- About 45% of the lakes and streams listed as impaired by the FDEP are the result of elevated nutrient levels.
- Nitrate concentrations have increased in many Floridan aquifer springs since the 1950s.
- These trends indicate that human activities at land surface likely are impacting ground-water quality, which may in turn impact surface-water quality.



OBJECTIVES

- Identify hydrogeological factors and biogeochemical processes that are important in controlling nitrate occurrence in central and northeast Florida.
- Develop novel technologies for nutrient removal using in-situ permeable reactive media units in stormwater retention ponds.

DATA MINING

"We are drowning in information and starving for knowledge."

- Rutherford D. Roger

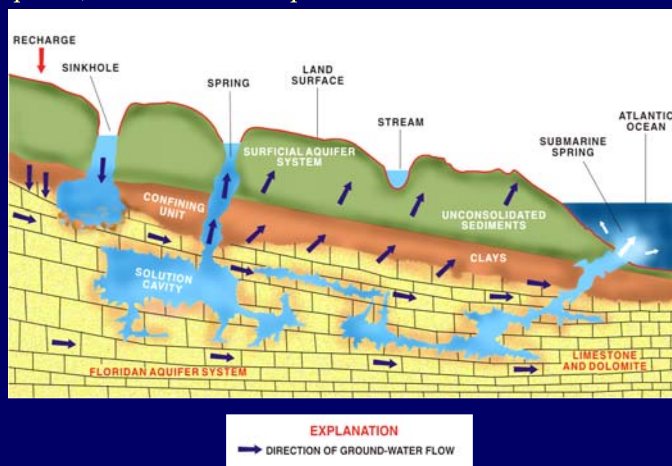
- Apply data mining techniques to sift through large volumes of data to identify associations among the data that may provide insight into the factors and processes that are important in controlling nitrate occurrence.

APPROACH

- Compile a database of nitrate concentrations and related water-quality, land-use, and geologic parameters for wells sampled in central and northeast Florida from 1990 to 2006.
- Develop several models (tree, regression, and neural network) to predict the presence of nitrate possibly attributable to human influences.
- For these purposes, nitrate is considered to be present when its concentration is equal to or greater than 0.2 mg/L and absent (at background concentrations) below this level.
- Madison and Brunett (1985) identified 0.2 mg/L as the level below which nitrate concentrations probably represent natural background conditions, based on a statistical analysis of nearly 124,000 wells throughout the United States.
- Select the best model for further interpretation.

BACKGROUND
Florida Hydrogeology

- Three principle aquifers: Surficial aquifer, Intermediate aquifer, and Floridan aquifer



BACKGROUND
Factors Controlling Nitrate Occurrence

Transport of nitrogen from land surface into the aquifer:

- Source of nitrogen (e.g. fertilizer and animal or human waste)
- Hydrogeologic properties of the subsurface (e.g. clay versus sand and depth of the aquifer or well).

BACKGROUND
Factors Controlling Nitrate Occurrence

Biogeochemical transformation of nitrogen:

- Two important processes that result in the transformation of nitrate are nitrification and denitrification
- *Nitrification*: oxidation of NH_4^+, end product NO_3^-
  - facilitated by bacteria
  - requires the presence of oxygen (aerobic)
- *Denitrification*: reduction of NO_3^-, end product N_2
  - facilitated by bacteria
  - requires the absence of oxygen (anaerobic)
  - requires an electron donor, commonly C, S, Fe, Mn

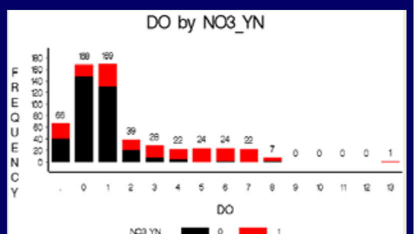
DATABASE

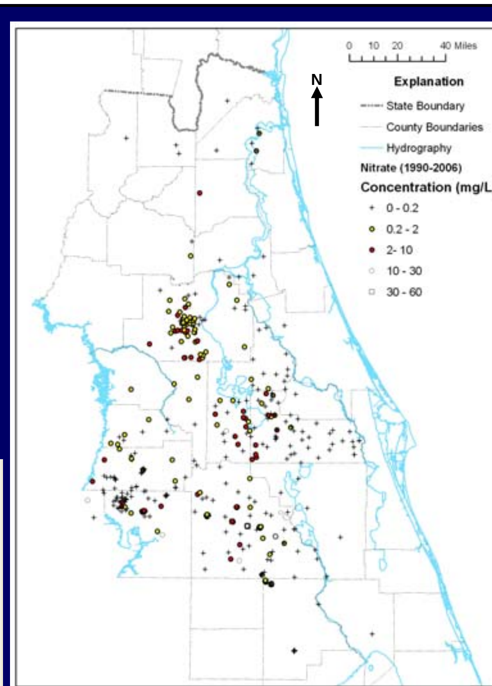
- 570 wells: 49% SAS, 46% FAS, 5% IAS/ICU
- Target variable, binary (NO3_YN):

  1 → elevated nitrate ($\geq$ 0.2 mg/L), 38.4%
  0 → background nitrate (< 0.2 mg/L), 61.6%
- 25 input variables: Major ions, Temp, SC, pH, DO, ANC, electron donors (TOC/DOC, SO_4, Fe, Mn), land use, hydrogeological factors
- Water quality parameters compiled from the USGS National Water Information System for 1990 – 2006.
- Land use and hydrogeological parameters derived from GIS analysis using well location and relevant maps.
- Missing data varies: 0% to 78% , averages 15%

EXPLORATORY DATA ANALYSIS

- Geographic analysis
- Variable histograms
- Descriptive statistics





DATA PREPARATION & PARTITIONING

- For regression and neural network models, data were consolidated, imputed, and transformed
  - $\geq$ 5 categories → consolidated via tree model
  - Missing data imputed via tree surrogate model
  - Data transformed to maximize normality
- For tree model, *no* consolidation, imputation, or transformation
- For all models, data partitioned into training dataset (70%, 399 wells) and validation dataset (30%, 171 wells).

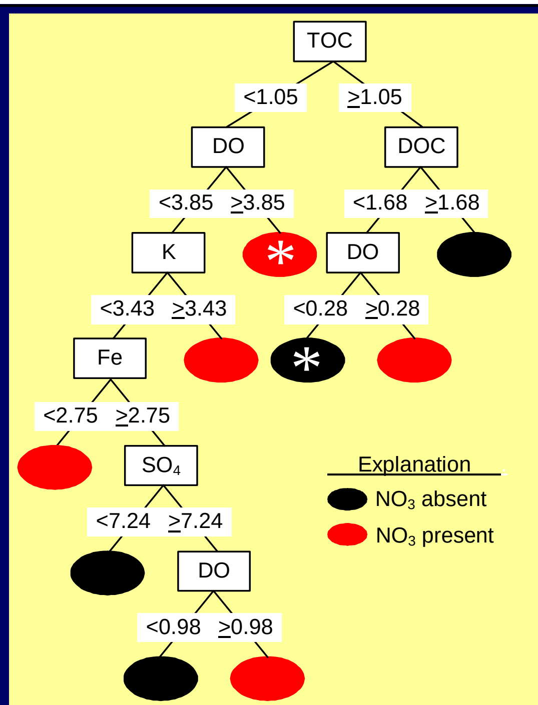
MODEL COMPARISON

Model	Role	False Negative	True Negative	False Positive	True Positive	Misclassification Rate (%)
Tree	Train	20	222	24	133	
	Validate	11	95	10	55	12.28
Linear Logistic Regression	Train	38	231	15	115	
	Validate	15	99	6	51	12.28
Neural Network	Train	29	234	12	124	
	Validate	14	97	8	52	12.87
AutoNeural Network	Train	30	234	12	123	
	Validate	18	100	5	48	13.45
Nonlinear Logistic Regres.	Train	34	234	12	119	
	Validate	18	96	9	48	15.79

- Tree model selected because it does not require variable consolidation or imputation. Therefore, its results are not affected by assumptions inherent in these techniques.

FINAL TREE MODEL

- A classification tree recursively partitions the dataset into increasingly homogenous (in terms of the target variable) subsets.
- To yield a more generalized model (not overfit), the final tree was reduced in size, yielding a misclassification rate of 14.62%.
- In general, the more important variables are more frequently used in a split.
- DO is the most important variable, i.e. has the greatest explanatory value for predicting nitrate occurrence



FINAL TREE MODEL
Some Example Splitting Rules

- Aerobic conditions with little carbon availability are *not* conducive for denitrification, thus nitrate present in the ground water is relatively stable.

```
IF      DO >= 3.85
  AND TOC < 1.05
THEN
  N       :      50
  0       :    4.0%
  1       :   96.0%
```

- Anaerobic conditions with high carbon availability *are* conducive for denitrification, thus nitrate entering the ground water may be transformed to different nitrogen species.

```
IF      DO < 0.275
  AND DOC < 1.675
  AND TOC >= 1.05
THEN
  N       :      51
  0       :   82.4%
  1       :   17.6%
```

FINAL TREE MODEL
Some Additional Observations

- Other factors can be identified that suggest the potential for transport of nitrogen from land surface into the aquifer.
- Elevated nitrate concentrations occur when potassium (K) concentrations are high. Potassium concentrations may likely be an indicator of fertilizer application. Potassium, in addition to nitrogen and phosphorus, is an important plant nutrient included in commercial fertilizers.
- Variable AQ_TYPE is used in a surrogate splitting rule indicating elevated nitrate concentrations for unconfined aquifers and background concentrations for confined aquifers. Hydrogeologic conditions, such as the presence of clay confining sediments, may restrict the movement of contaminated ground water into the aquifer.

BMP Strategies for Nitrate Control

- From data mining results, high TOC and low DO imply low nitrates.
- Develop novel technologies for mimicking the natural conditions that are conducive for nutrient attenuation.
- Selection and emplacement of permeable reactive media is a critical component of successful BMP.

Technology Matrix

- Permeable reactive barriers (PRBs) at strategic locations of the watershed
- Bioswale and biofiltration
- LID options for urban areas
- Riparian buffer and vegetation type in river corridors
- Changes in agricultural management
- Microbial nitrogen removal enhancement - Denitrification enzyme activity (DEA)

Optimal Design and Location of BMPs using Reactive Media

- Reactive media selection
- Recipe and module design
- Kriging analysis for identifying nutrient problem areas (regional scale)
  - Spatial and temporal scales
  - Geostatistic methodologies (kriging, co-kriging, factor kriging, stochastic simulation)
  - Geographic Information Systems
- Groundwater modeling (subregional and local scale) for identifying optimal location of BMPs with reactive media

Permeable Reactive Media Selection
~ Results of Literature Review ~

- Florida Peat
- Sandy/Loamy/ Clayey soils
- Sawdust (untreated wood)
- Paper/Newspaper
- Tire Crumb + electron donor
- Limestone + electron donor
- Crushed Oyster + electron donor
- Wood Fiber/Chips/ Compost

CONCLUSIONS AND RECOMMENDATIONS

- In general, the results of this analysis suggest that nitrate occurrence in ground water in central and northeastern Florida is controlled, to some degree, by a combination of
  - biogeochemical transformation processes;
  - nitrogen sources; and
  - hydrogeologic conditions.
- Permeable reactive media units may serve as a cost-effective BMP for the in-situ attenuation of nutrients in the subsurface.

T H E… E N D

A
N
Y
?'s