

# Logistic Regression Used To Relate Ground Water Quality To Man-made And Natural Causes:

## Better than Multiple Regression

*Jim Tesoriero, Michael G. Rupert,  
Lonna M. Frans, and  
Alex K. (Sandy) Williamson (Presenter)*  
National Water-Quality Assessment Program  
U.S. Geological Survey

# Vulnerability Factors:

- Relate water-quality data to **natural** and **human** factors
- **Natural Hydrogeologic**: ease with which contaminant can reach aquifer or sampling point – called susceptibility or potential vulnerability
- **Human Contaminant Source**: effects of land use activities – called vulnerability

# Vulnerability Methods

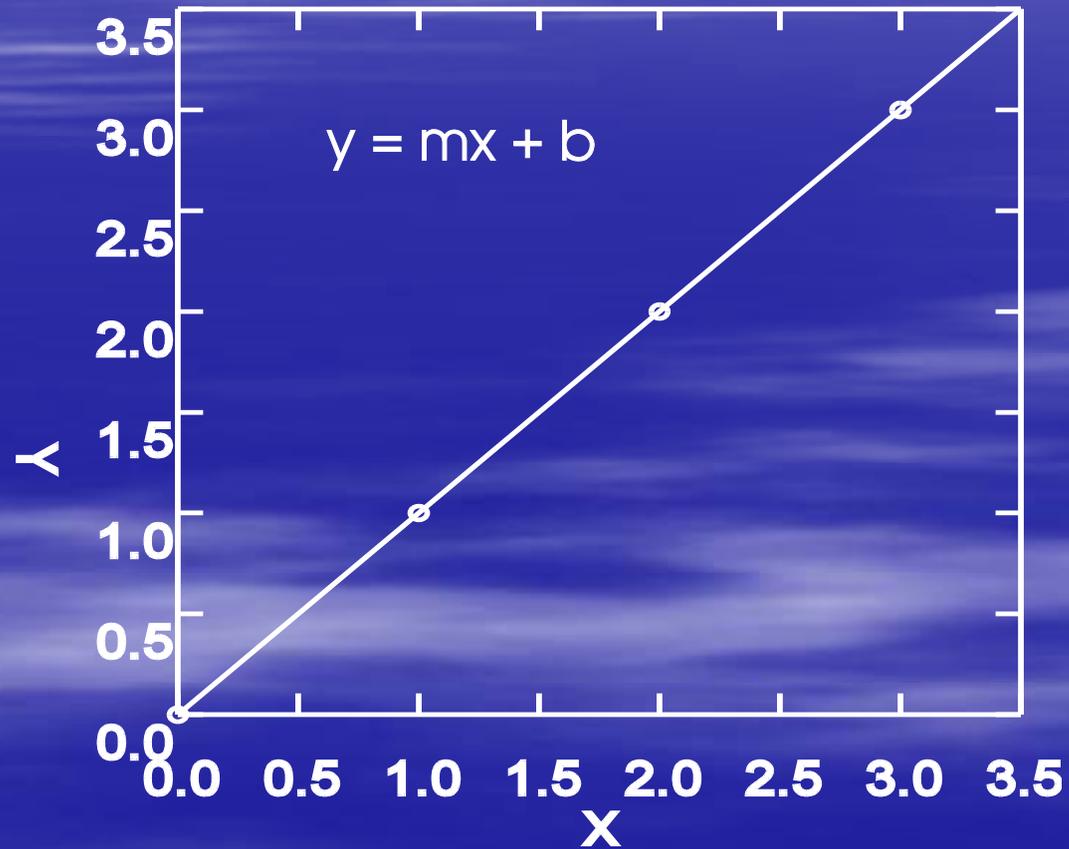
- Process-based Simulation Models
- Overlay and Index Methods, i.e..  
DRASTIC-type
- Statistical Methods – Regression, Logistic –  
using Existing water-quality data

# Logistic Regression

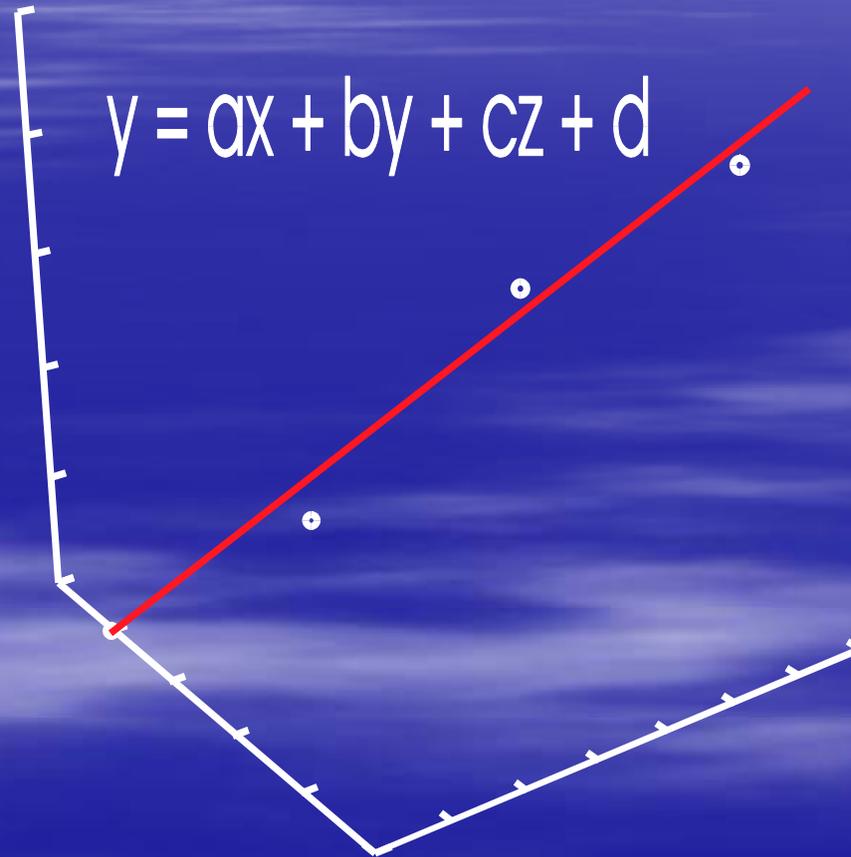
- Answers a simpler Question than linear regression does.
- Therefore it can answer the simpler question with a higher degree of confidence.
- Probability that a constituent concentration is above a specified level is predicted.
- Variables included in the model and their relative weights are chosen statistically.

Let me explain with a few graphs....

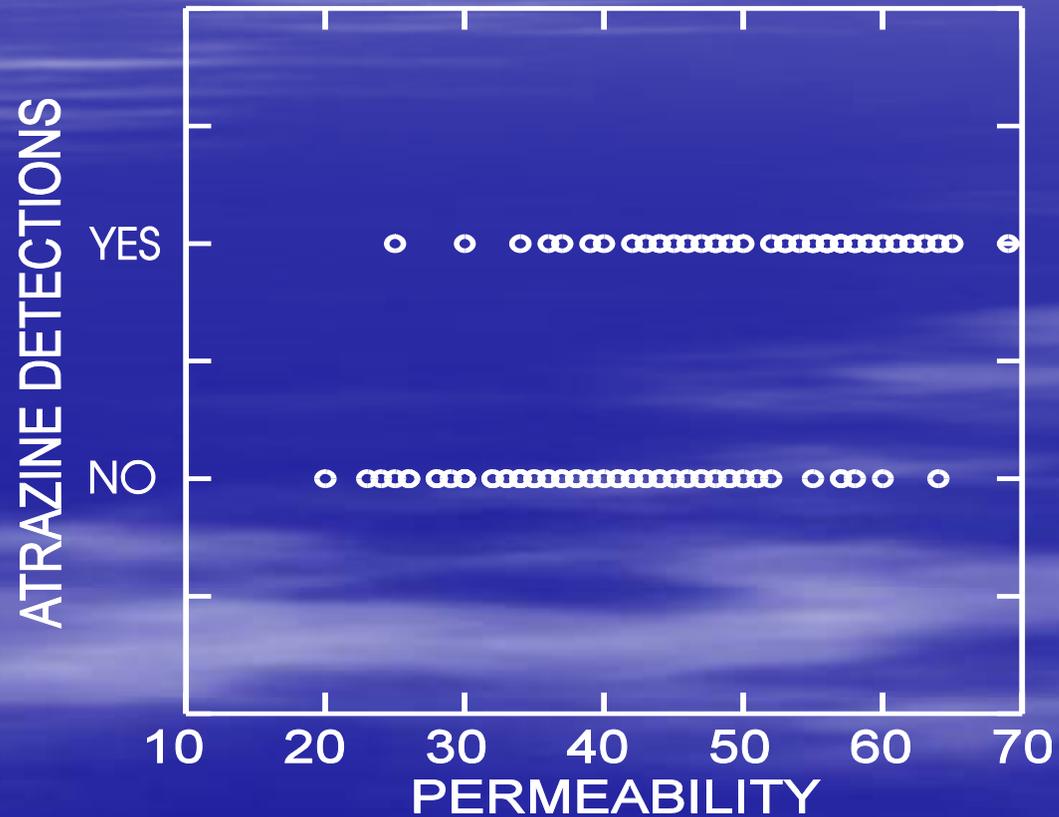
# Linear Regression



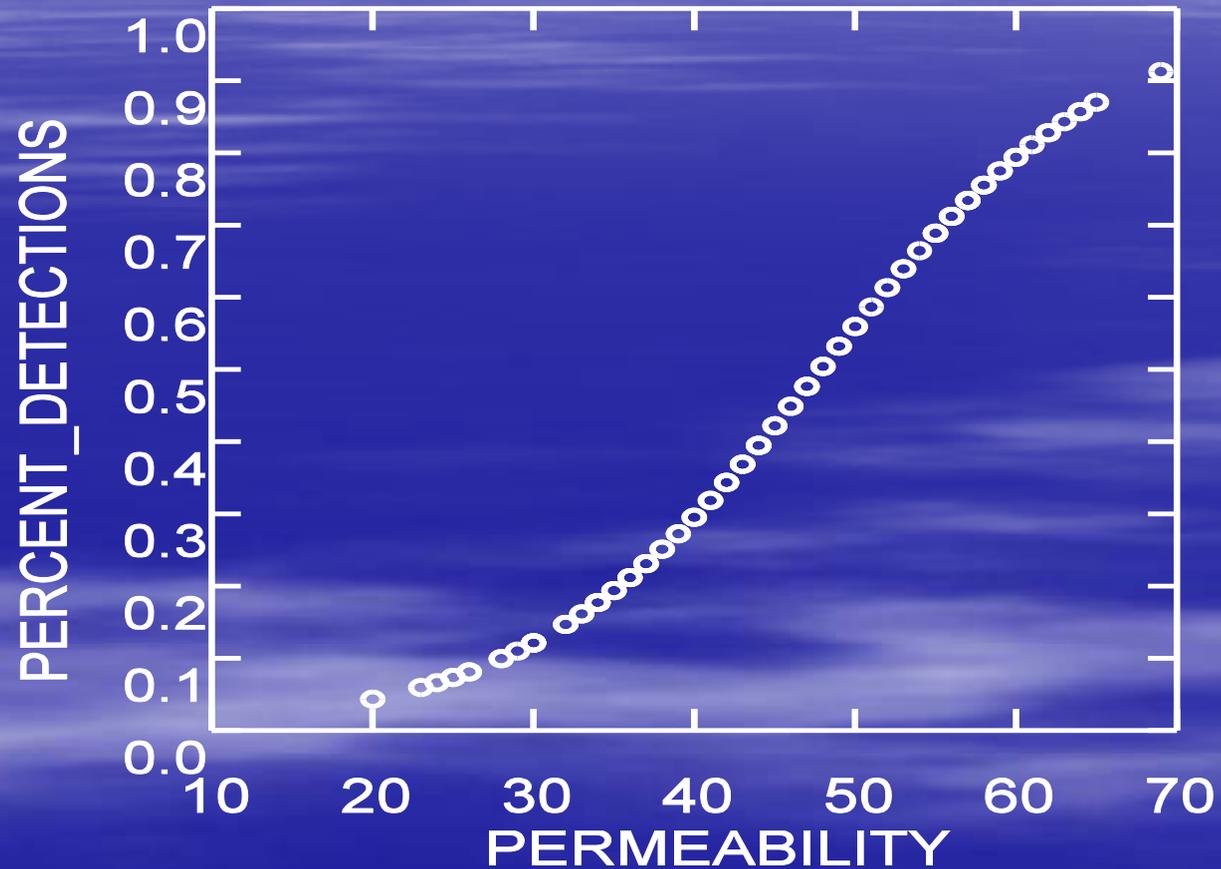
# Multiple Regression



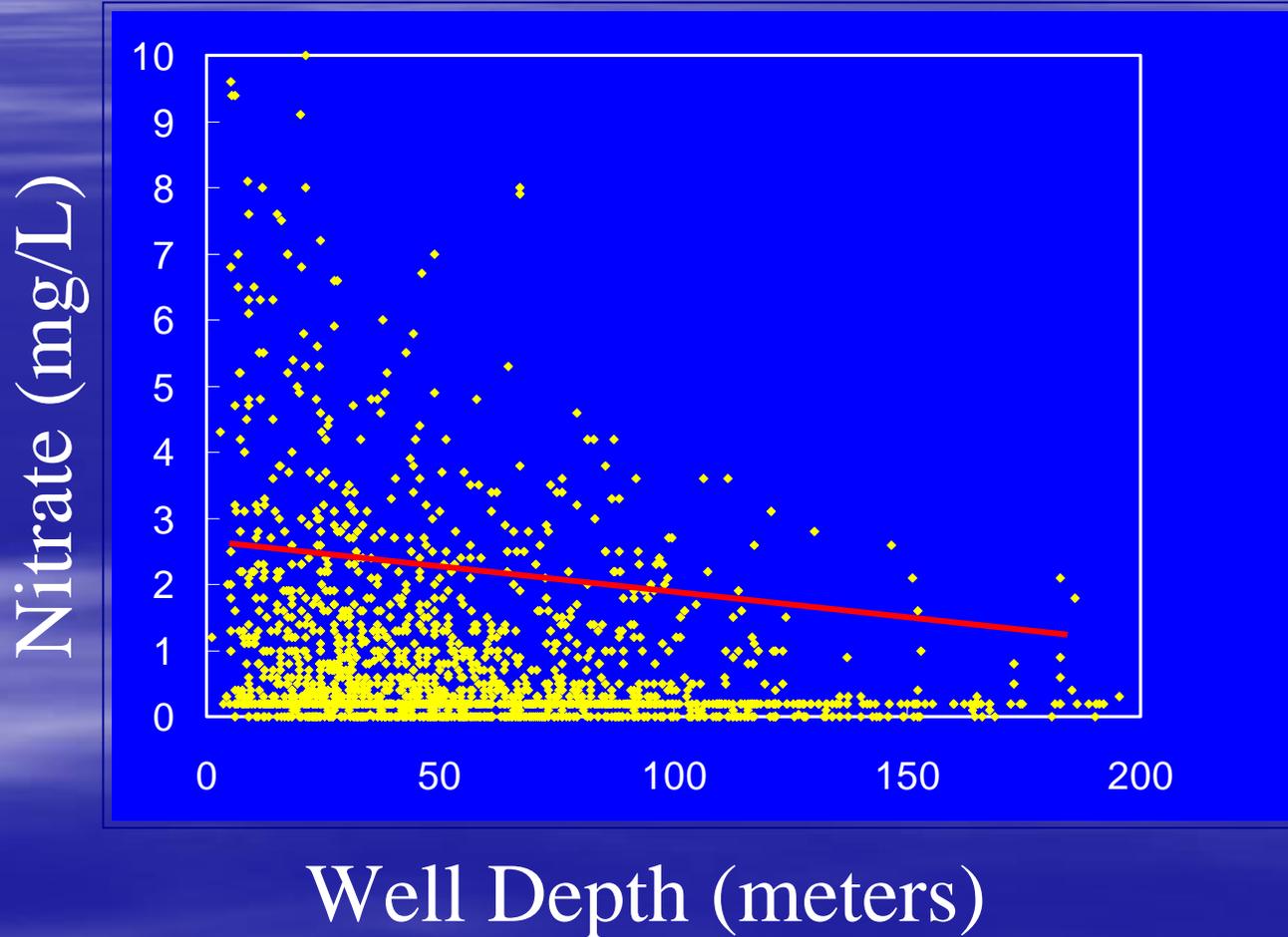
# Logistic Regression



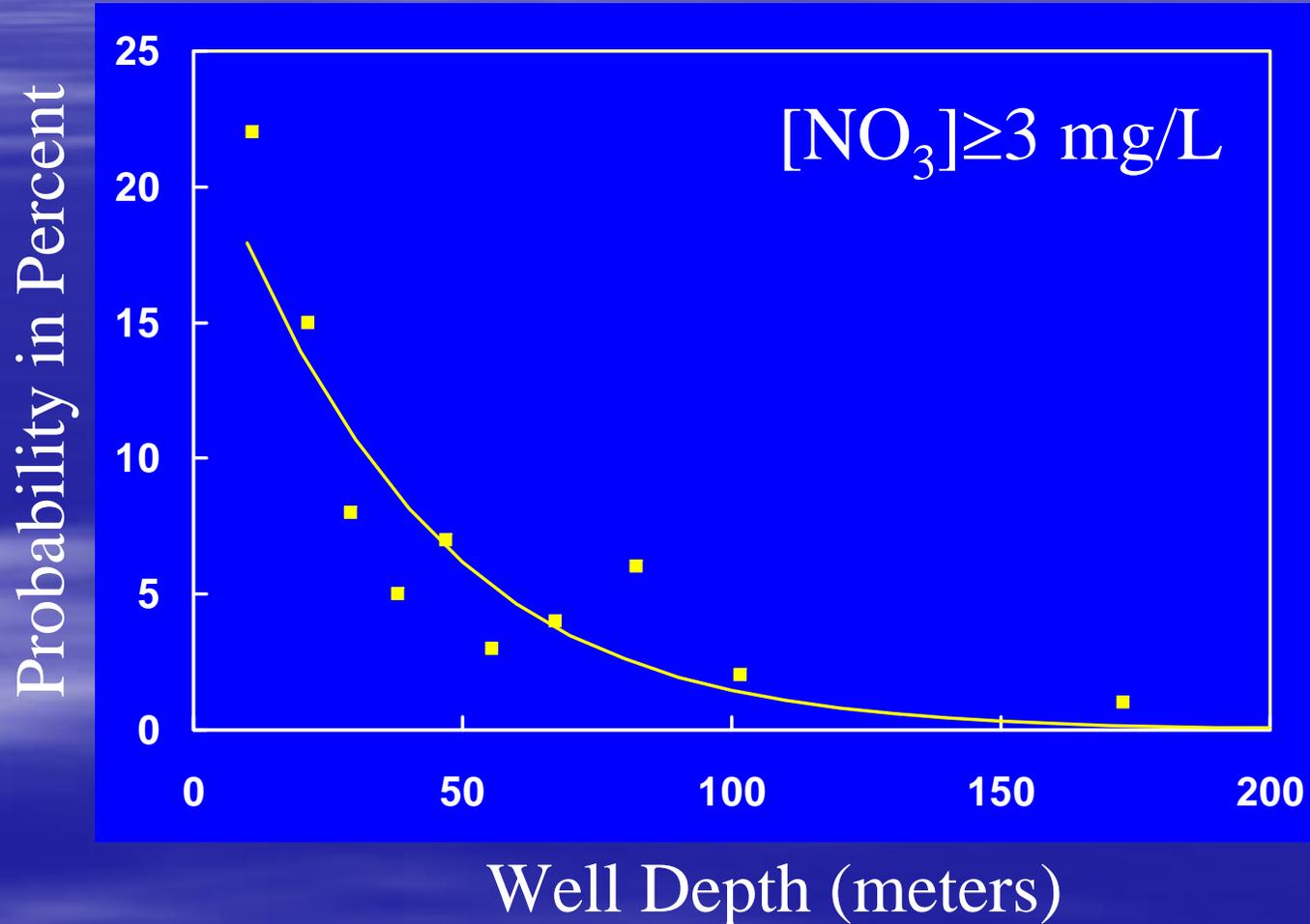
# Logistic Regression



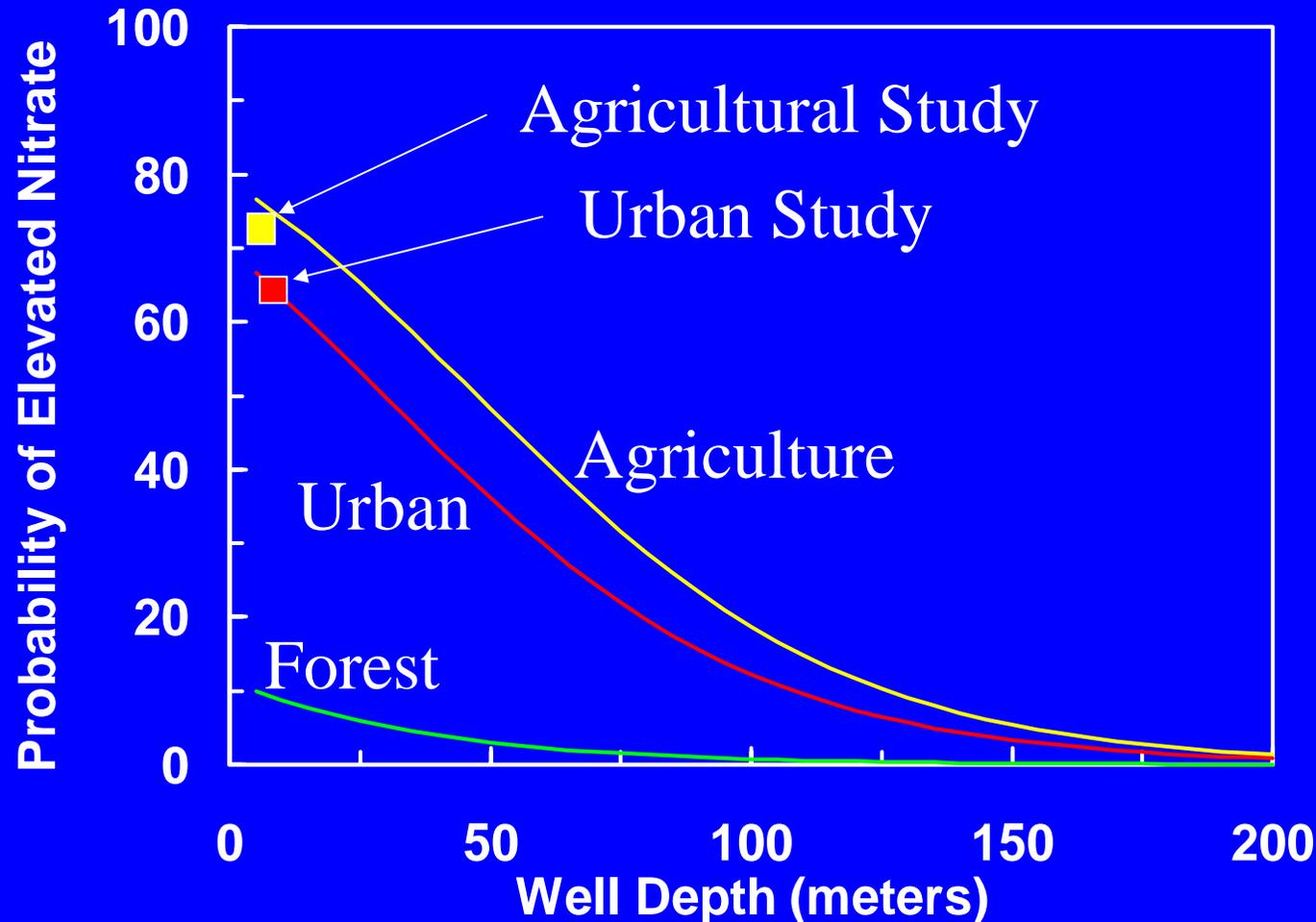
# Predictions Of Concentrations By Linear Regression May Not Be Possible At Acceptable Certainty...



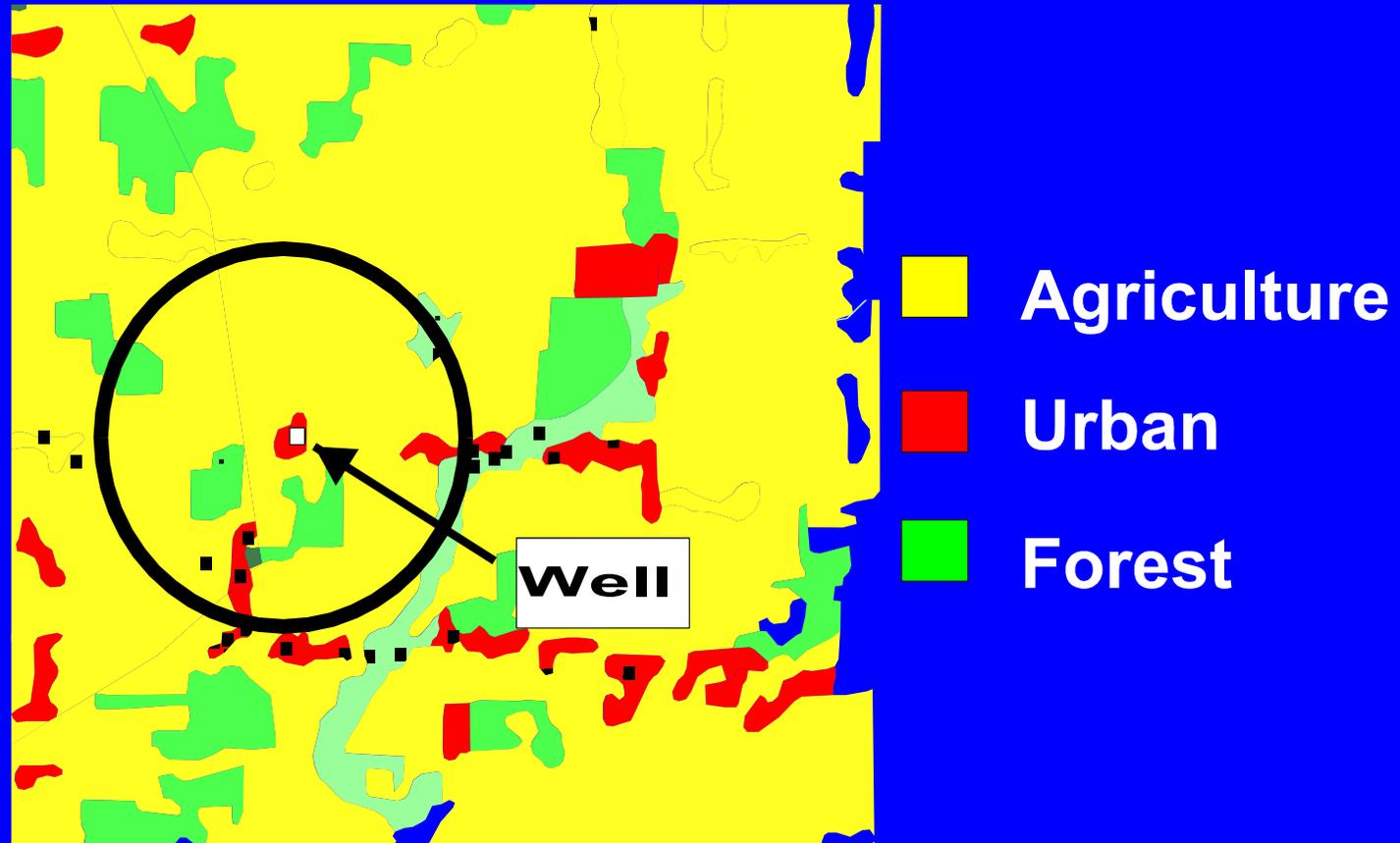
# It May Be Possible to Predict the *Probability* that a Concentration Exceeds a Threshold



# Recent Study Results Consistent with Logistic Regression Fit

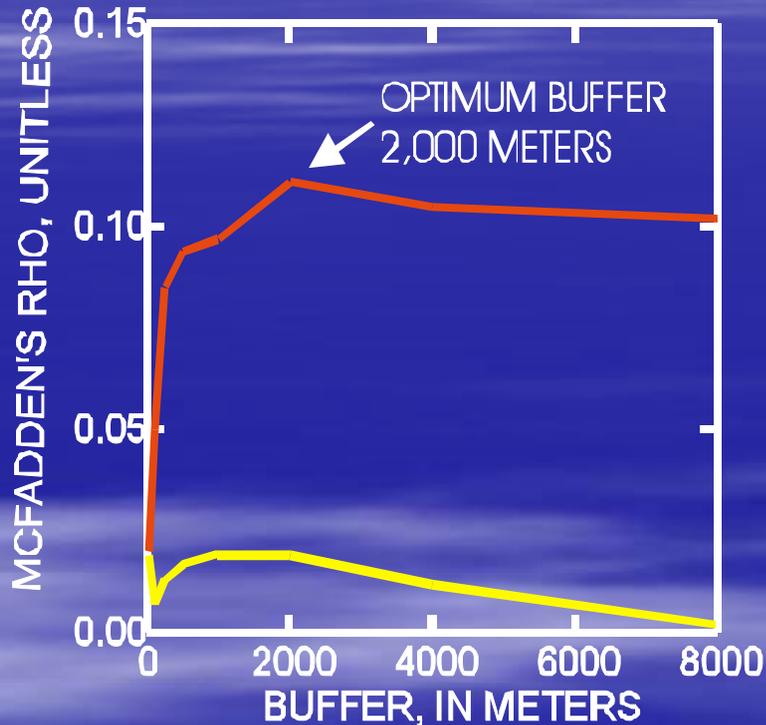


# LandUse:PercentLandUseTypeWithin 3.2kmofWellUsedinModel



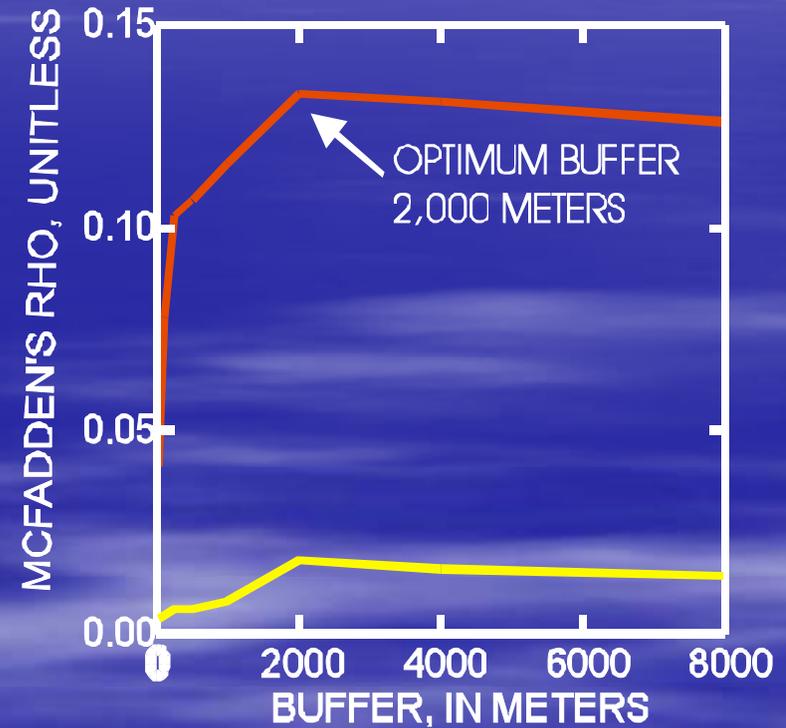
# AGRICULTURAL LANDS

## ATRAZINE



— ROW CROPS  
— SMALL GRAINS

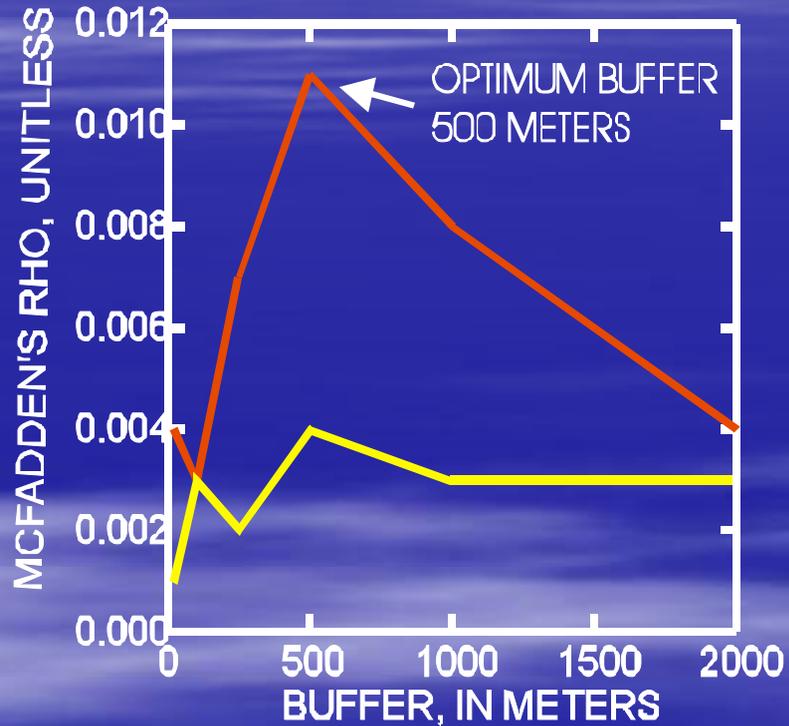
## NITRATE



— ROW CROPS  
— SMALL GRAINS

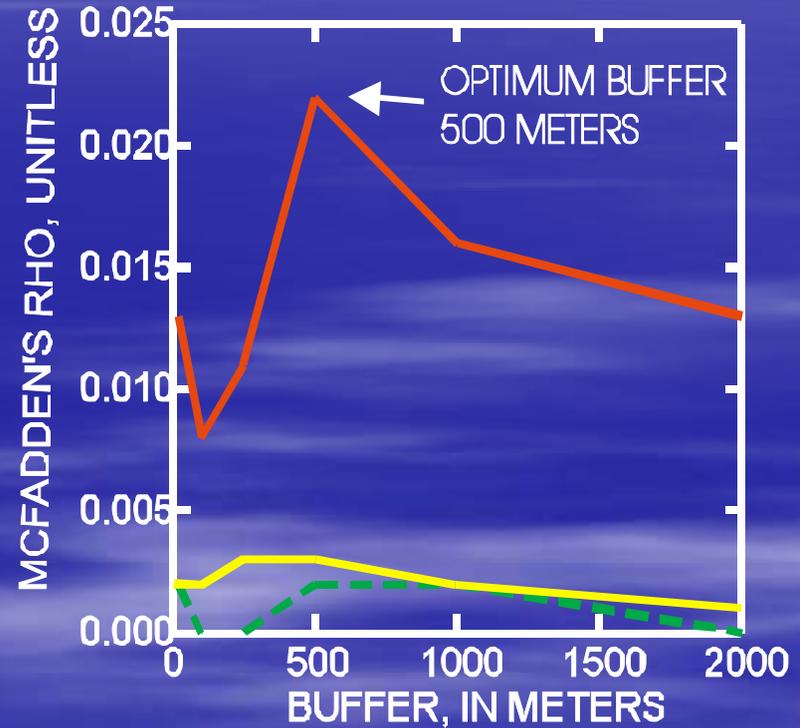
# URBAN LANDS

## NITRATE



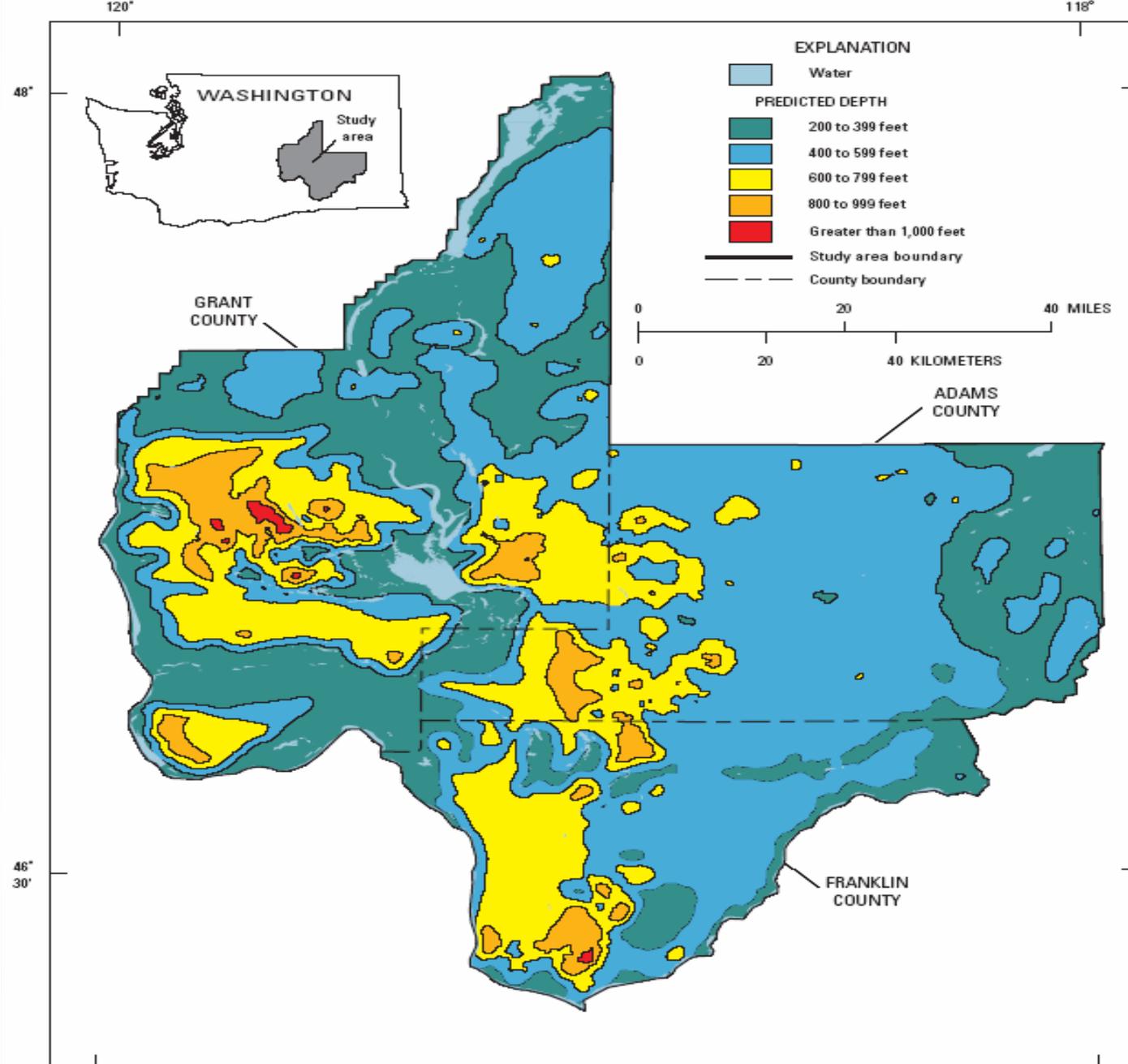
— COMMERCIAL  
— URBAN\_GRASSLAND

## ATRAZINE



— COMMERCIAL  
— URBAN GRASSLANDS  
- - - HIGH DENSITY RESIDENTIAL

Predicted depth to which wells need to be cased to have at least an 80-percent chance of obtaining water with a nitrate concentration less than 3 milligrams per liter.



**Figure 11.** Predicted depth to which wells need to be cased to have at least an 80-percent chance of obtaining water with a nitrate concentration less than 3 milligrams per liter.

# Conclusions

- **Ground-water vulnerability assessed using statistical approach.**
- **Useful tool for prioritizing monitoring and the selection of source areas.**
- **Method can be used to evaluate other contaminants and levels (e.g., MCLs).**

# Advantages of Logistic Regression

- **Calculates actual probability of detections, instead of assigning subjective categories such as “High Vulnerability”**
- **Weighting and interaction of variables such as geology and soils is automatically evaluated by logistic regression**
- **All possible combinations of independent variables such as land use and soils can be numerically evaluated, and the best model selected.**
- **Can be used for pesticide data (few detections, concentrations near laboratory reporting limits)**

# Advantages of Statistical Method for Vulnerability Assessment

- Probability that a threshold is exceeded can be predicted both spatially and temporally
- Variables in model and their relative importance is determined statistically
- Data collection needs are minimal if existing water quality and ancillary data can be used

# References

- <http://water.usgs.gov/pubs/wri/wri02-4269/pdf/WRIR02-4269.pdf>
- <http://wa.water.usgs.gov/pubs/misc/ps.gw.vol35.no.6.html>
- <http://www.ecy.wa.gov/events/hg/abstracts2000.pdf>  
page 10 & 77
- <http://webserver.cr.usgs.gov/midconherb/html/texas.html>
- <http://wa.water.usgs.gov/pubs/wri/wri004110>

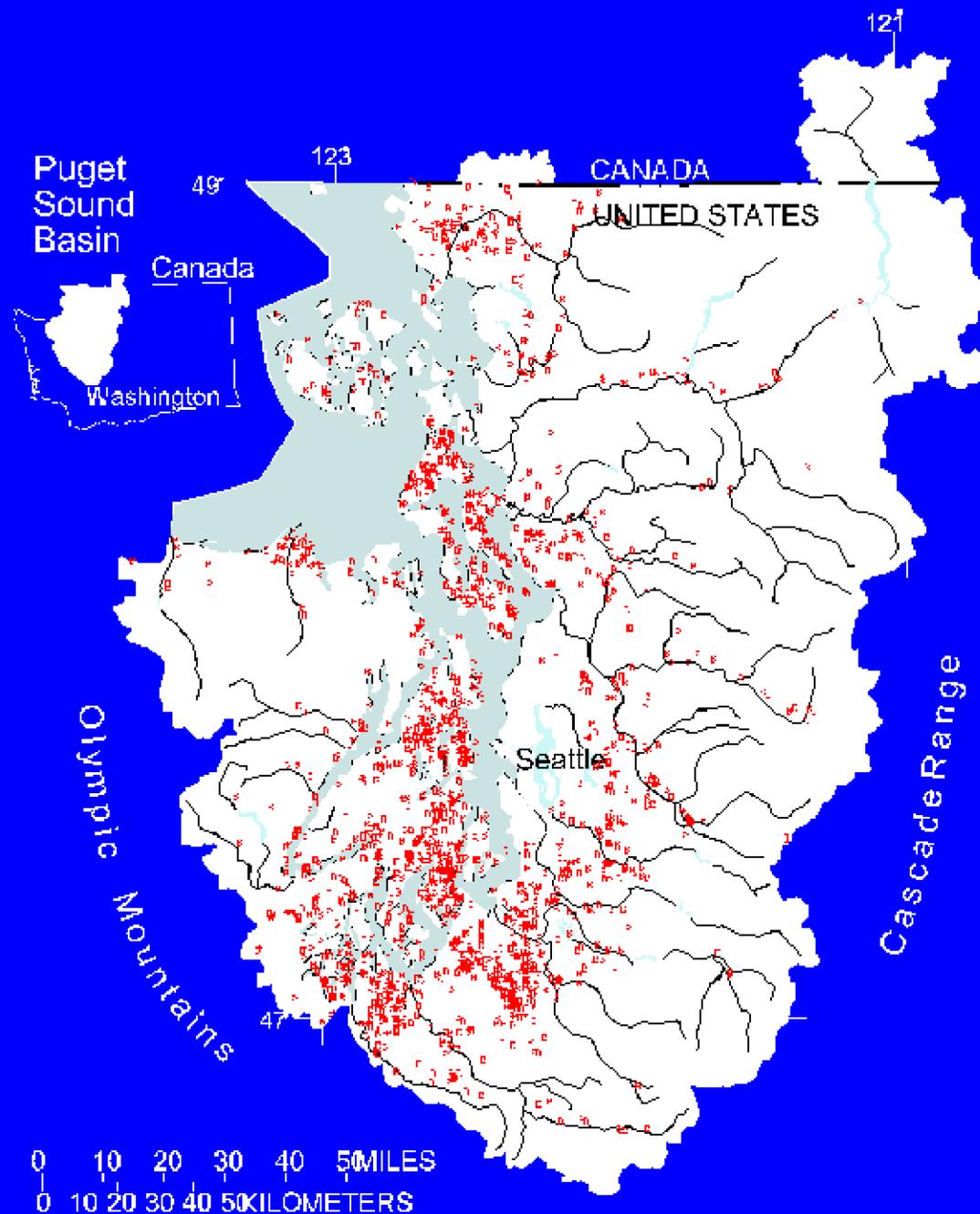
# The End

- *Sandy Williamson, presenter*
- *U.S. Geological Survey, Water Resources,*
- *<http://water.usgs.gov/nawqa/data>*
- *1201 Pacific Avenue, Suite 600, Tacoma, WA  
98402*
- *(253) 428-3600 ext. 2683; cell (253) 376-8273*
- *E-mail [akwill@usgs.gov](mailto:akwill@usgs.gov)*

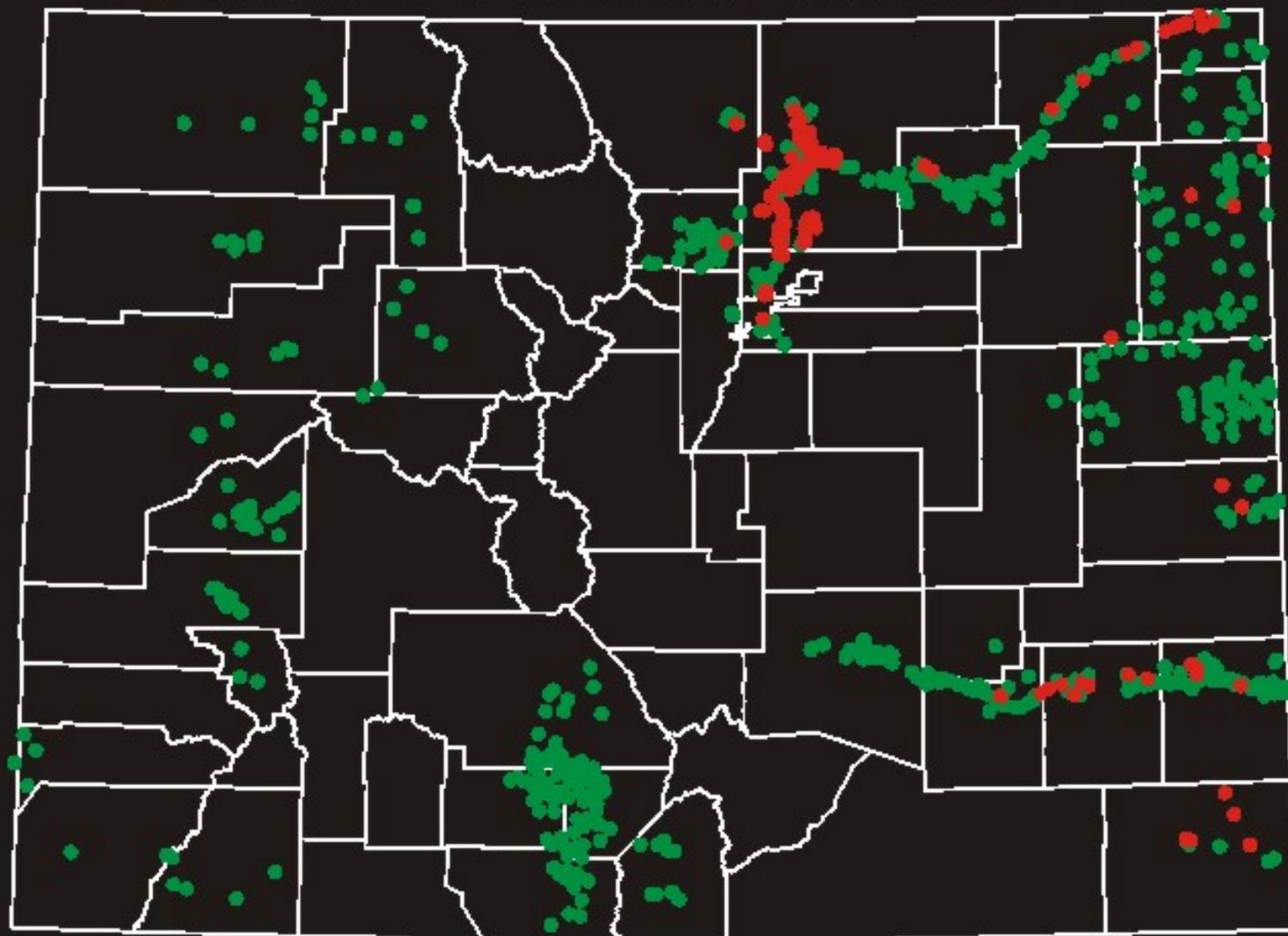
# GIS DATA

- Chemical Use
- Elevation
- Geology
- Hydrogeomorphic regions
- Land cover
- Precipitation
- Soils

# Well Locations



# ATRAZINE DETECTIONS IN COLORADO'S GROUND WATER DATA COLLECTED BY THE STATE OF COLORADO, 1992-99



● No atrazine detection ● Atrazine detection

# OPTIMUM NITRATE CONCENTRATION

- Three nitrate concentrations were evaluated to determine which concentration generated the most effective logistic regression model: 2 mg/L, 5mg/L, and 10mg/L.
- The nitrate data were transformed to detect/no detect nomenclature at each of these three concentrations, and logistic regression models were developed.

- Ground-water vulnerability typically has typically been assessed using qualitative methods expressed as relative measures of risk, like DRASTIC. The logistic regression approach has the advantage of having both model variables and coefficient values determined on the basis of existing water\_ quality information. Unlike DRASTIC-type methods, the logistic regression approach it does not depend on the somewhat arbitrary assignment of variables and weighting factors based on qualitative criteria. Logistic regression is a great way to rigorously relate man-made and natural factors to ground\_ water quality. It usually produces more statistical confidence than regular multiple regression because logistical regression tries to answer a simpler yes or /no question (contaminant exceeds threshold) rather than the multiple regression question relating a wide range of concentration values to contributing factors. Logistic regression can be used to assess ground-water aquifer susceptibility (relative ease with which contaminants will can reach aquifer) and ground-water vulnerability (relative ease with which contaminants will reach aquifer for a given set of land-use practices). In 3three USGS studies in Washington and Colorado, the variables that best explain the occurrence of high nitrate or pesticides included the (1) well and/or casing depth, (2) the percentage of urban and agricultural land or the amount of fertilizer applied within a radius of 3.2 kilometers of the well, (3) surficial geology, and the (4) mean soil hydrologic group, which is a measure of soil \_infiltration rate. Maps can be made of the predicted depth to which wells would need to be cased in order to have an X-percentage probability of drawing water with high nitrate or pesticides or the predicted probability of high nitrate or pesticides for wells cased to median casing depth.