Assessing Ground Water Vulnerability Through Statistical Methods

Tom Nolan
btnolan@usgs.gov

NWQMC

May 7-11, 2006

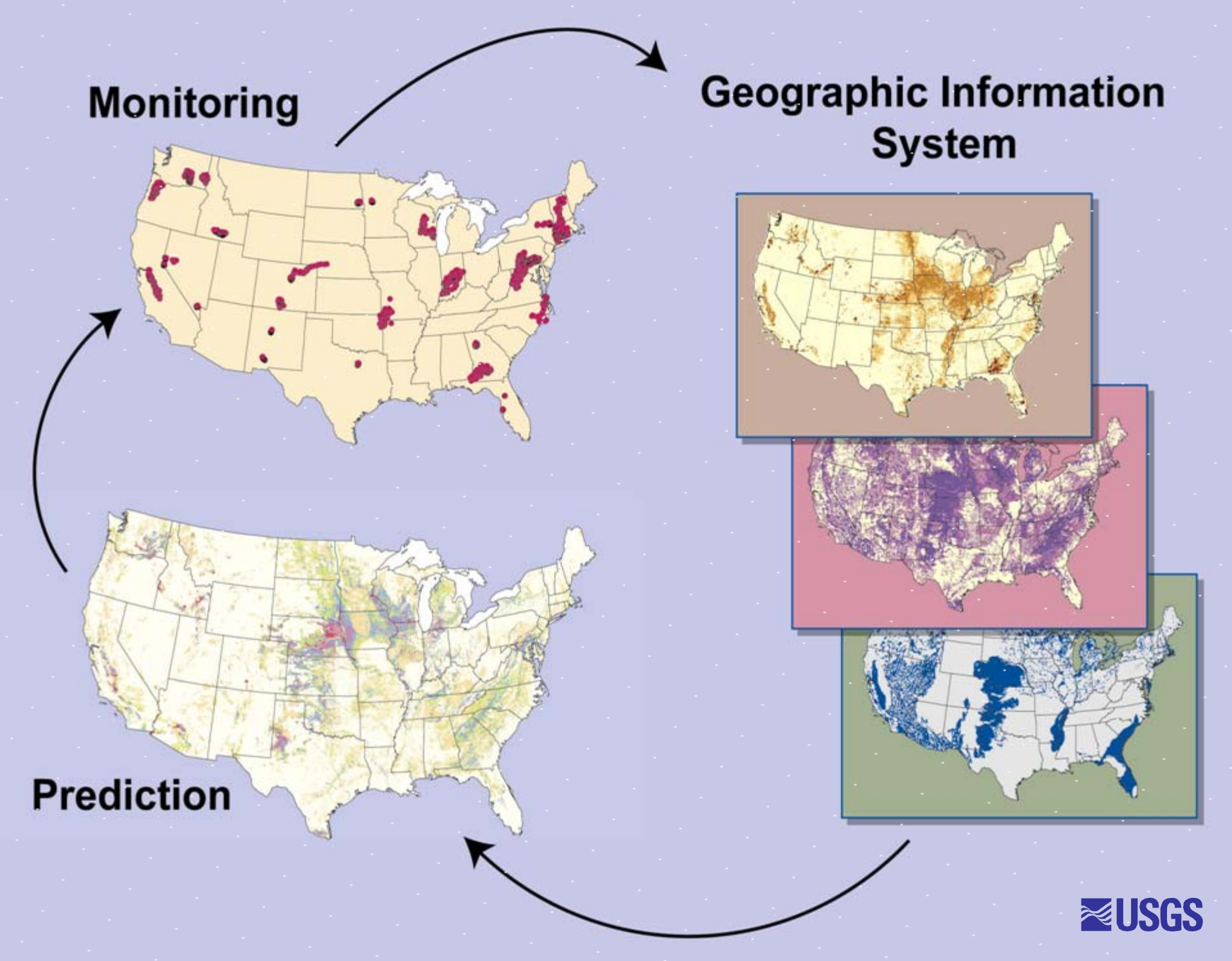
San Jose, CA

USGS

Statistical methods are a good option at large spatial scales.

Example approaches…

- Regression (National)
- Factor analysis (Regional)
- Hybrid deterministic-geostatistical (Subregional)


Monitoring
Geographic Information System
Prediction
USGS

Logistic Regression

- Good way to handle censored data (e.g., > 50%)
- Places results in probabilistic context
- Built-in uncertainty analysis (statistical methods in general)

USGS

Logistic Regression

$$\log\left[\frac{p}{1-p}\right] = b_o + \boldsymbol{bx}$$

p = probability of "event" (NO_3 > 4 mg/L)
b_o = constant
bx = vector of slope coefficients and explanatory variables

USGS

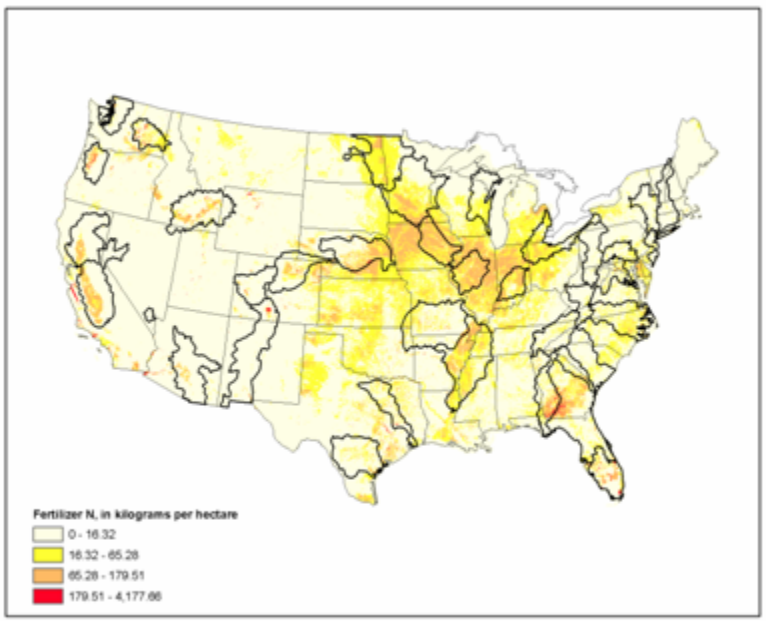
National Example for Nitrate in Shallow GW

- More than 1,200 wells sampled during 1992-1995
- Represent mostly shallow, recently recharged ground water
- Nitrate data plus ancillary (N load, soils, geology, etc.)

USGS

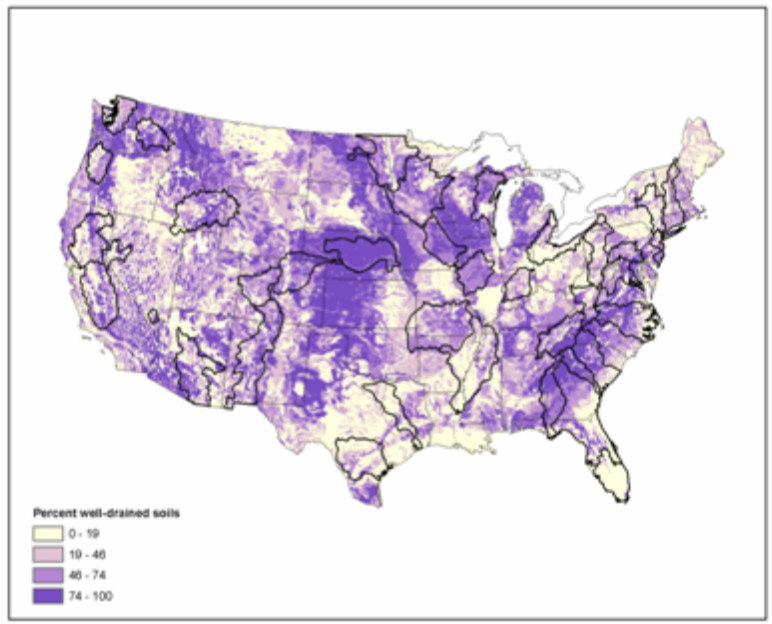
Test spatial data in the model…

Load



+

Susceptibility

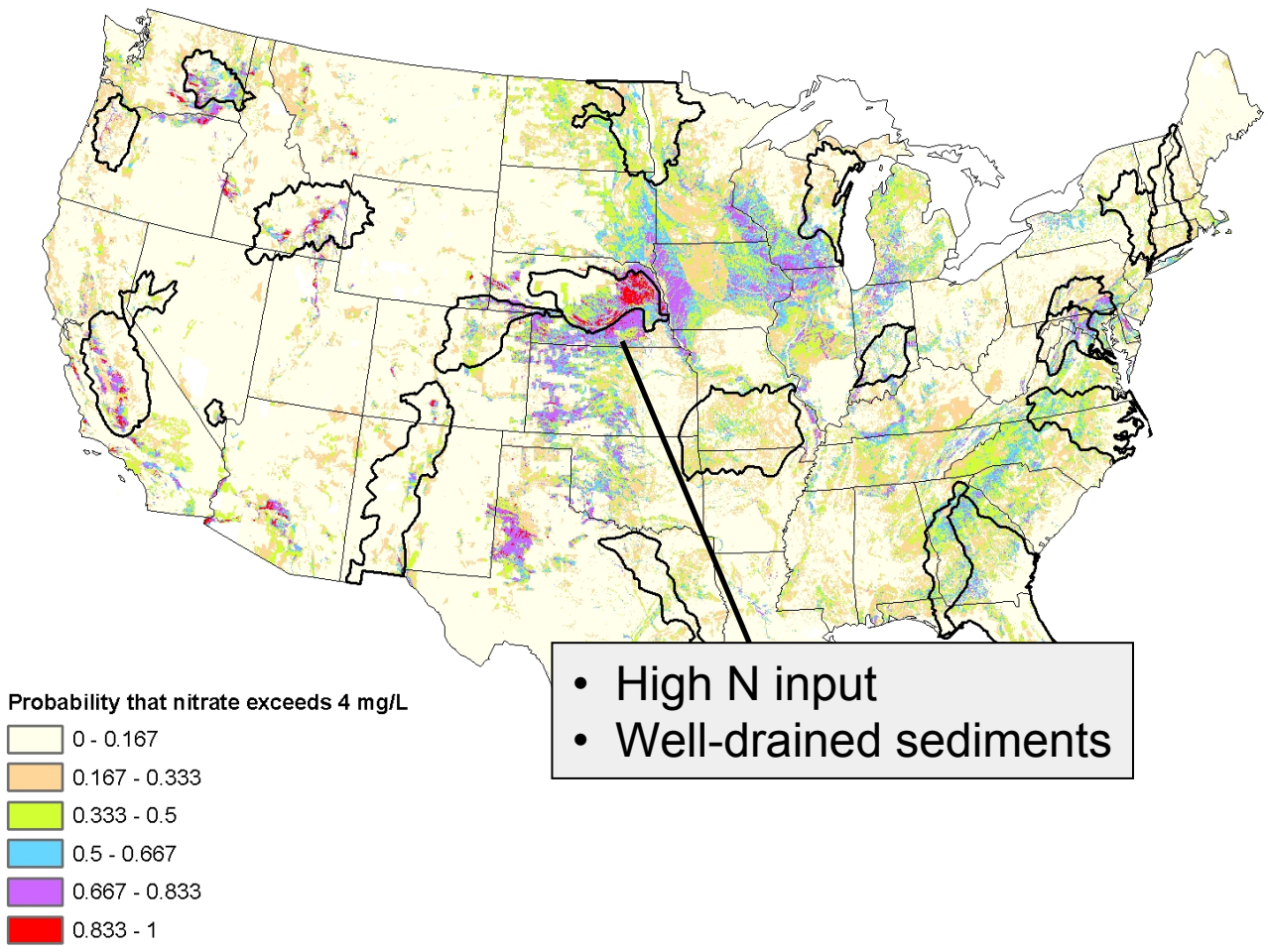


Final Model

Variable	Standard. coefficient	Wald p-value
N Load		
1992 Farm-nonfarm fertilizer N, kg/ha	0.101	<0.001
Cropland-pasture (NLCD), %	0.272	<0.001
ln(population density), ln(people/km^2)	0.223	<0.001
Aquifer susceptibility		
Well-drained soils, %	0.190	<0.001
Depth to season. high water table, m	0.159	<0.001
Presence or absence of unconsol. sand and grav. aquif.	0.091	0.002

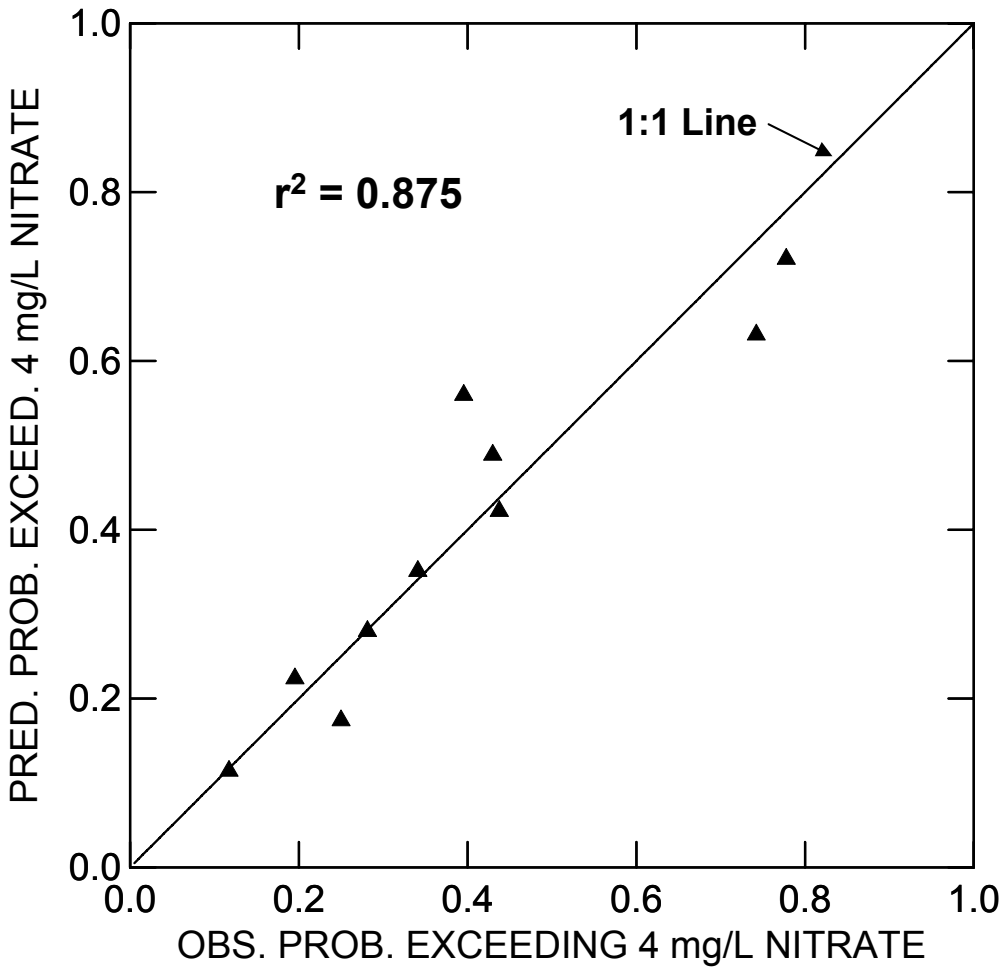
USGS

Nitrate Probability Map



- High N input
- Well-drained sediments

Observed vs. Predicted Probabilities for Calibrated Model



Diagnostics

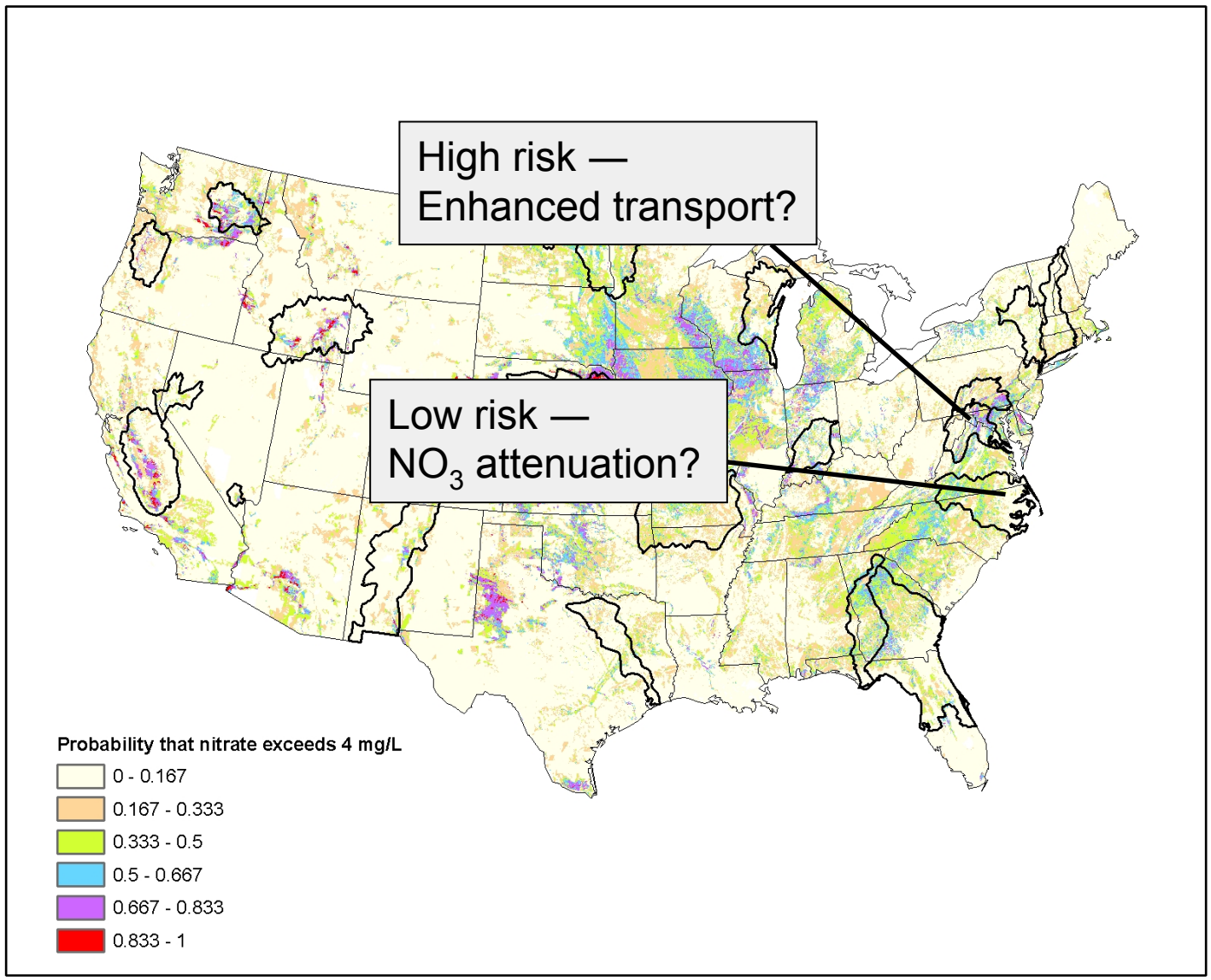
For each observation…

- Pearson residual (RESCHI)
- Leverage (H)
- Change in Pearson residual (DIFCHISQ)
- Standardized change in regression coefficients (DBETA)

Menard (2002), Applied Logistic Regression Analysis, Sage Publication 07-106

USGS

Factor Analysis for Aquifer Processes (Regional)


High risk —
Enhanced transport?
Low risk —
NO_3 attenuation?
Probability that nitrate exceeds 4 mg/L
0 - 0.167
0.167 - 0.333
0.333 - 0.5
0.5 - 0.667
0.667 - 0.833
0.833 - 1

Principal Components Analysis

- Reduces complex data set to a few "interpretable" components.
- Subjective—must know your data set.
- Terminology:

Components: linear combinations of original variables

Loadings: correlations between component and original variables

Scores: show component influence on individual samples

Eigenvalue: variance explained by component

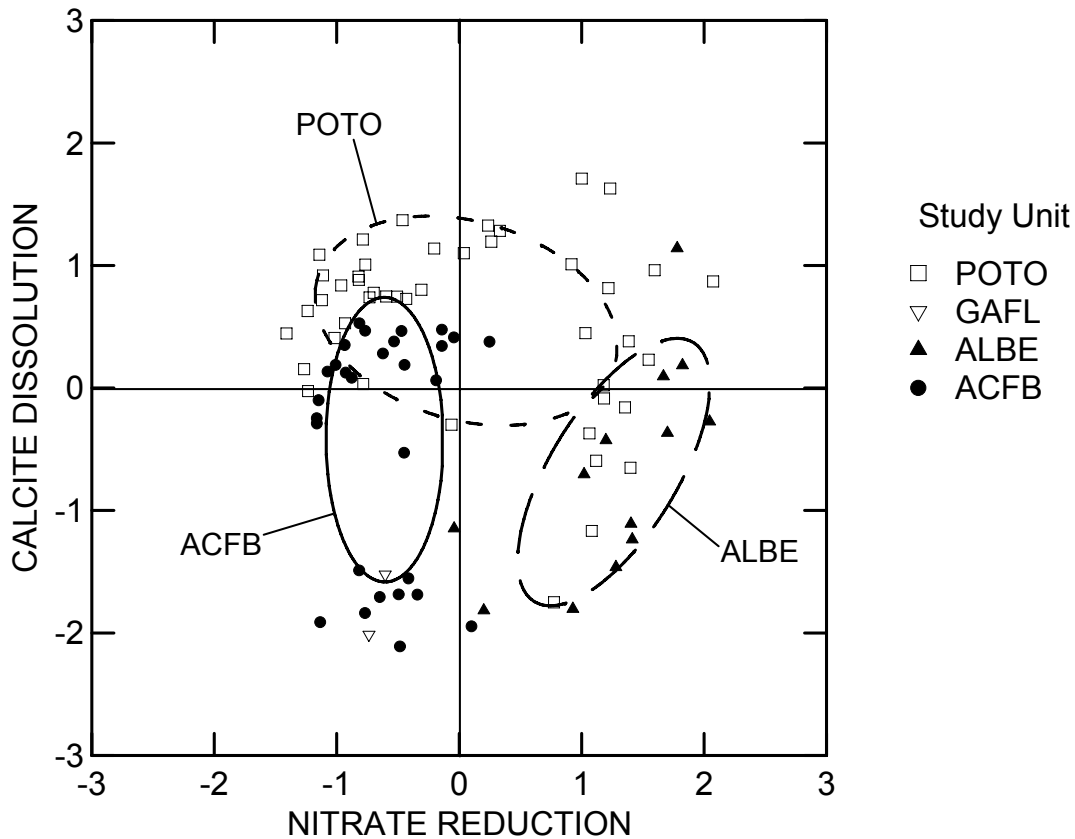
USGS

PCA Loadings

Water-quality variable	Nitrate reduction	Calcite dissolution	Phosphate dissolution
Dissolved oxygen	**-0.89**	-0.06	-0.00
Manganese	**0.81**	-0.34	-0.05
Nitrite-plus-nitrate as N	**-0.79**	0.13	-0.14
Ammonium as N	**0.77**	0.14	0.06
Iron	**0.76**	-0.40	0.12
Bromide	0.62	0.00	0.06
Ammonia and org. nitrogen as N	0.60	0.26	0.02
Sodium	0.59	0.13	0.17
Silica as SiO_2	0.54	-0.24	0.33
Calcium	-0.06	**0.85**	-0.12
Akalinity as $CaCO_3$	-0.08	**0.84**	0.03
pH	-0.06	**0.79**	0.13
Specific conductance	0.12	**0.76**	0.03
Dissolved solids	0.18	**0.74**	0.05
Phosphorous	0.17	0.01	**0.81**
Orthophosphorous as P	0.10	-0.05	**0.79**
Temperature	0.10	-0.17	-0.04
Magnesium	0.05	0.29	0.07
Potassium	0.00	0.05	-0.05
Nitrite as N	-0.16	-0.07	0.41
Chloride	0.07	0.03	-0.03
Dissolved organic carbon, as C	0.38	0.06	0.07
Fluoride	-0.21	0.16	**0.47**
Sulfate as SO_4	0.42	0.37	0.07
Percent variance explained:	23	18	9

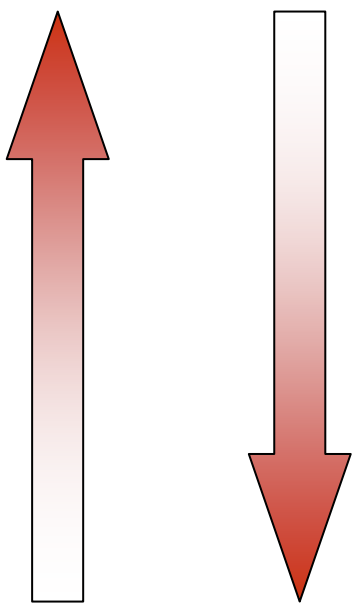
USGS

PCA Scores by Study Unit



Redox Reaction Sequence

kcal **pε**



O_2 consumption (respiration) → CO_2

Denitrification → N_2

Manganese-oxide reduction → Mn^{2+}

Nitrate reduction (if any remaining) → NH_4^+

Iron-oxide reduction → Fe^{2+}

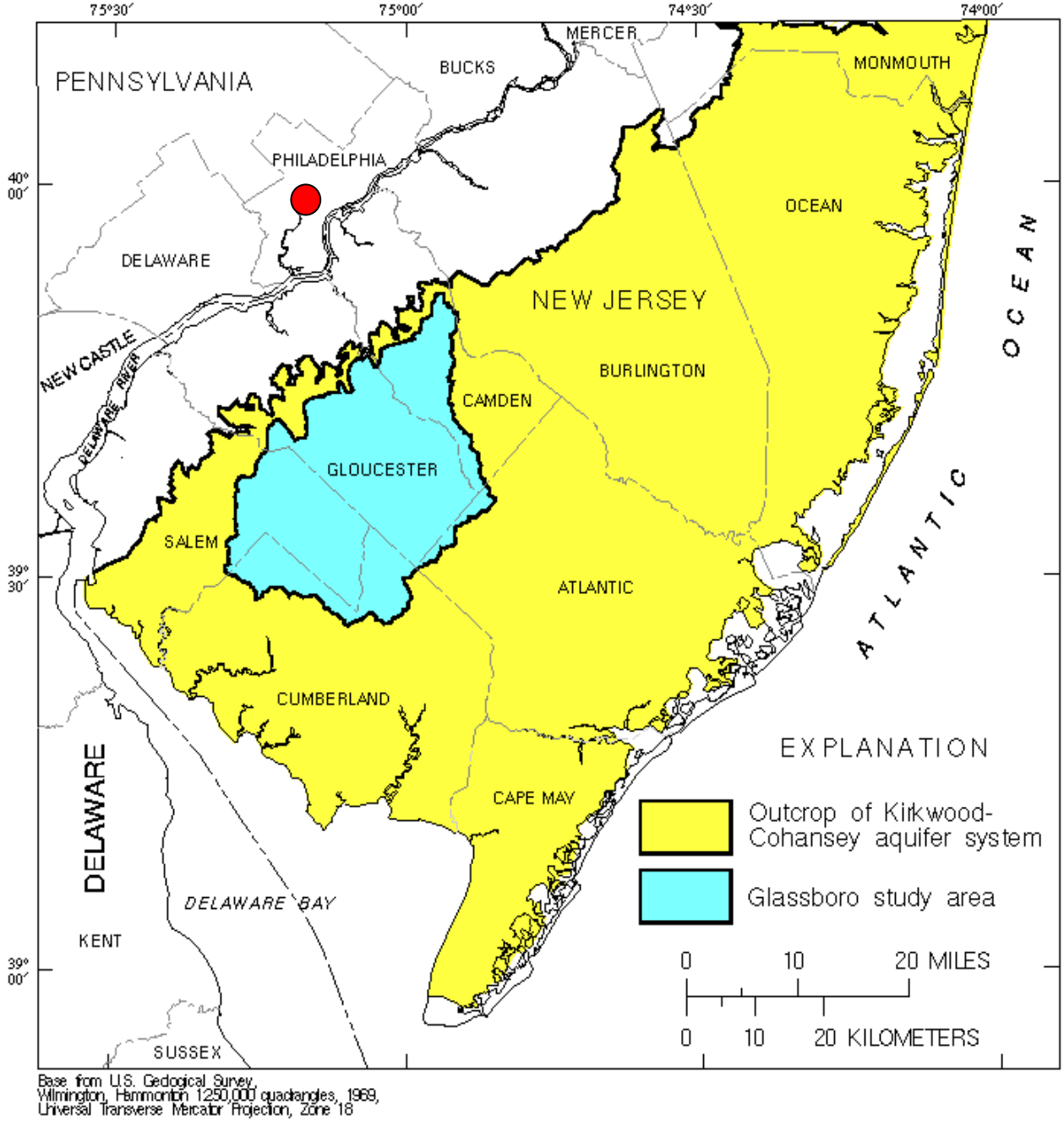
(Other reducing reactions…)

Hybrid Deterministic-Geostatistical Approach (Subregional)

Art Baehr

Leon Kauffman

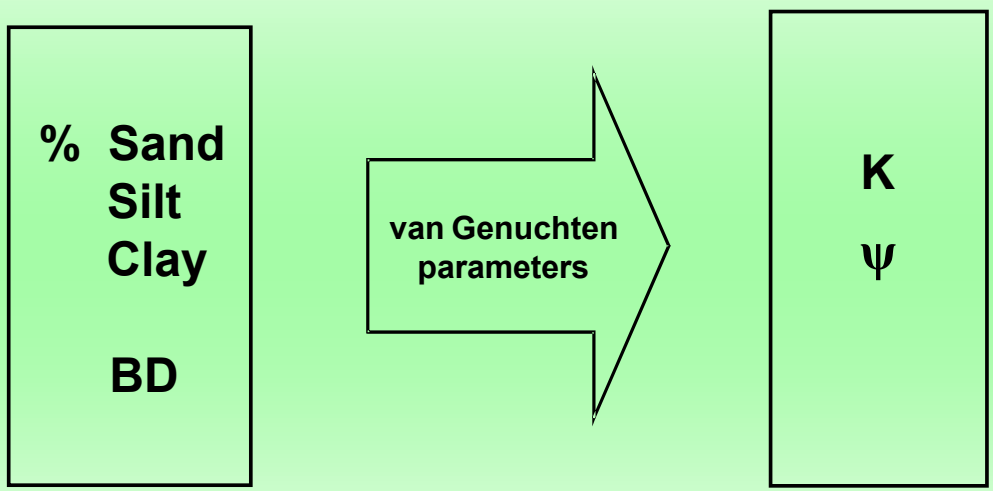
USGS



Base from U.S. Geological Survey, Wilmington, Hammonton 1:250,000 quadrangles, 1969, Universal Transverse Mercator Projection, Zone 18

Location of Glassboro, New Jersey study area

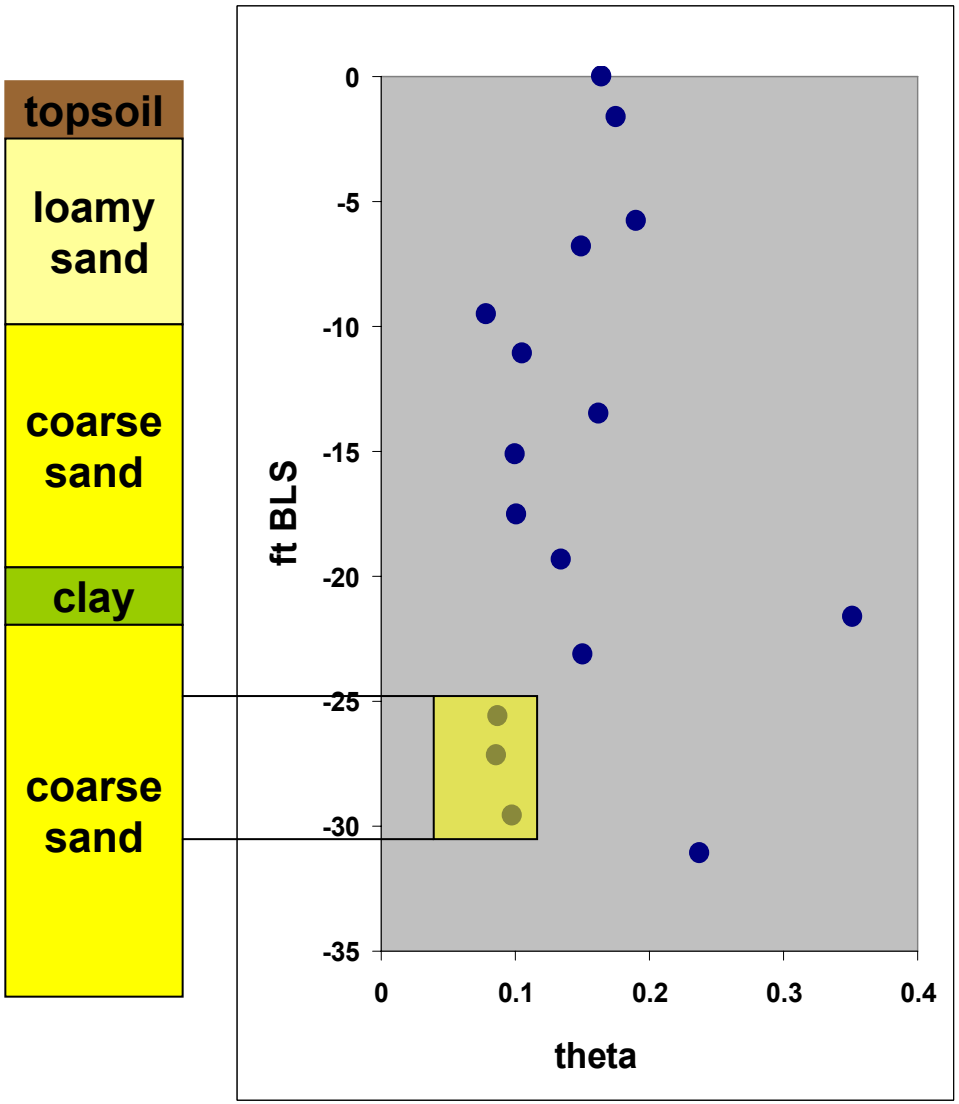
Darcian Approach to Aquifer Susceptibility



http://www.ussl.ars.usda.gov/MODELS/rosetta/rosetta.htm

USGS

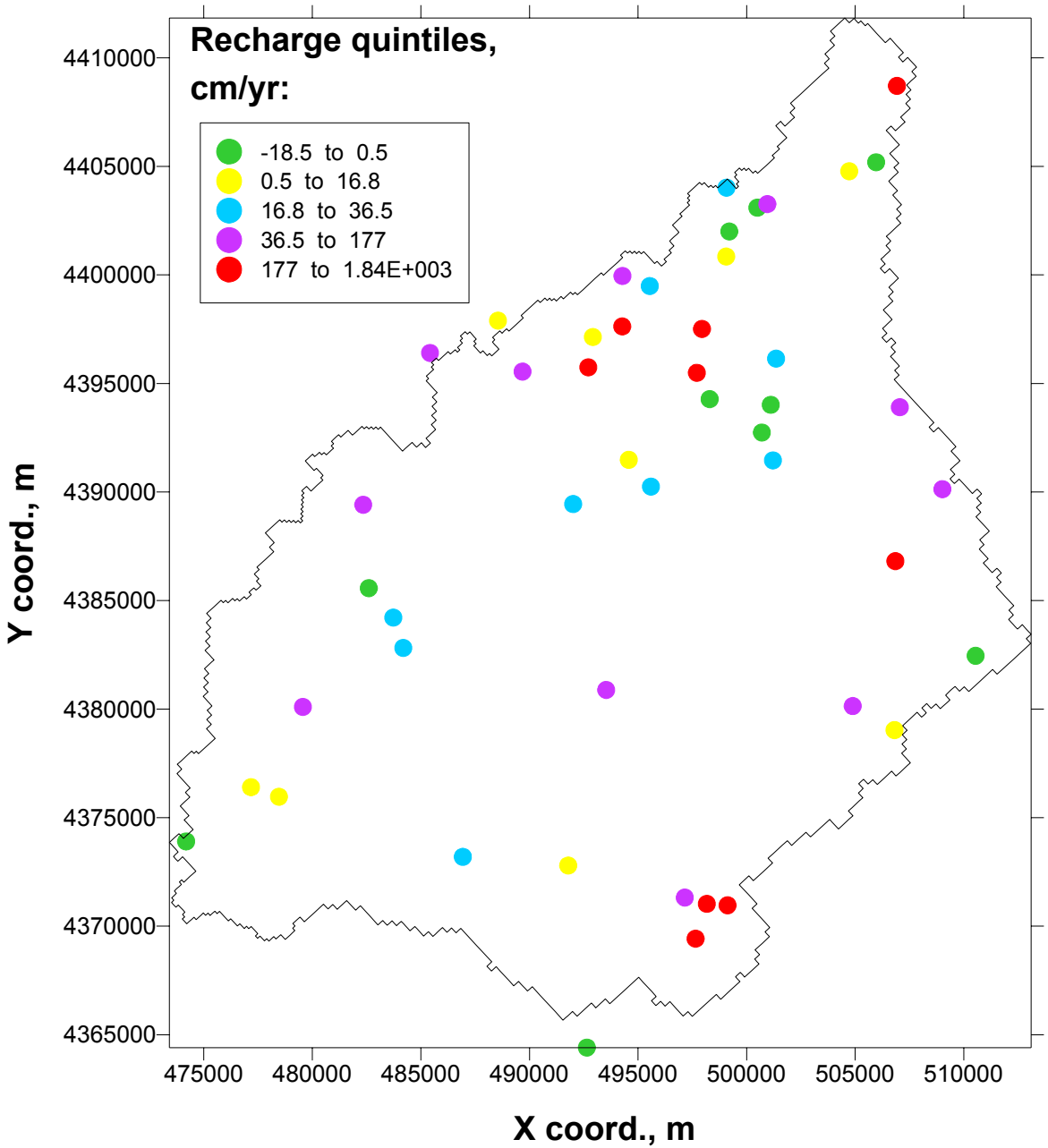
Darcian Method



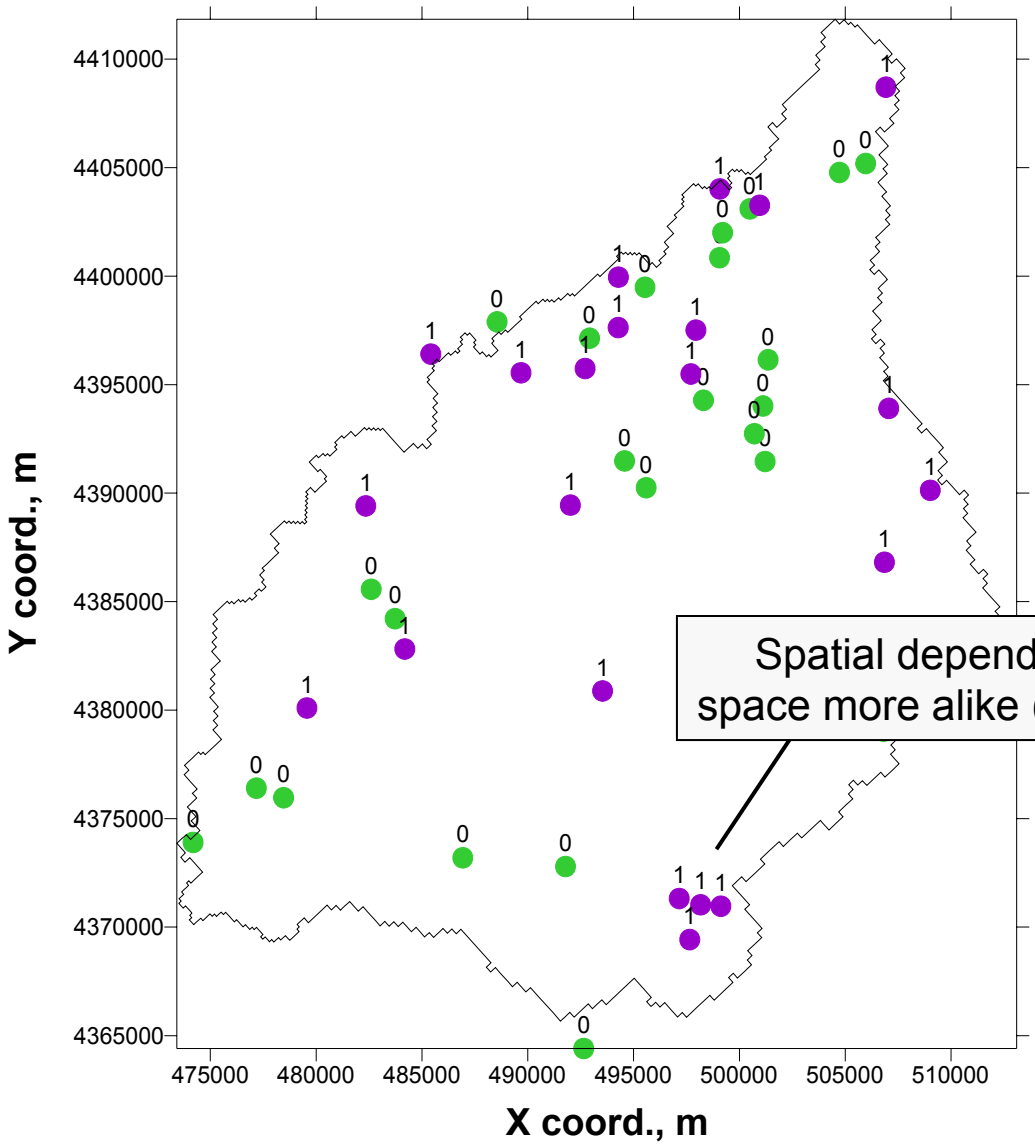
Calculation made in lowermost unit above capillary fringe:

$$q = -K(\theta)\left[\frac{d\psi}{dz} + 1\right]$$

Recharge Estimates

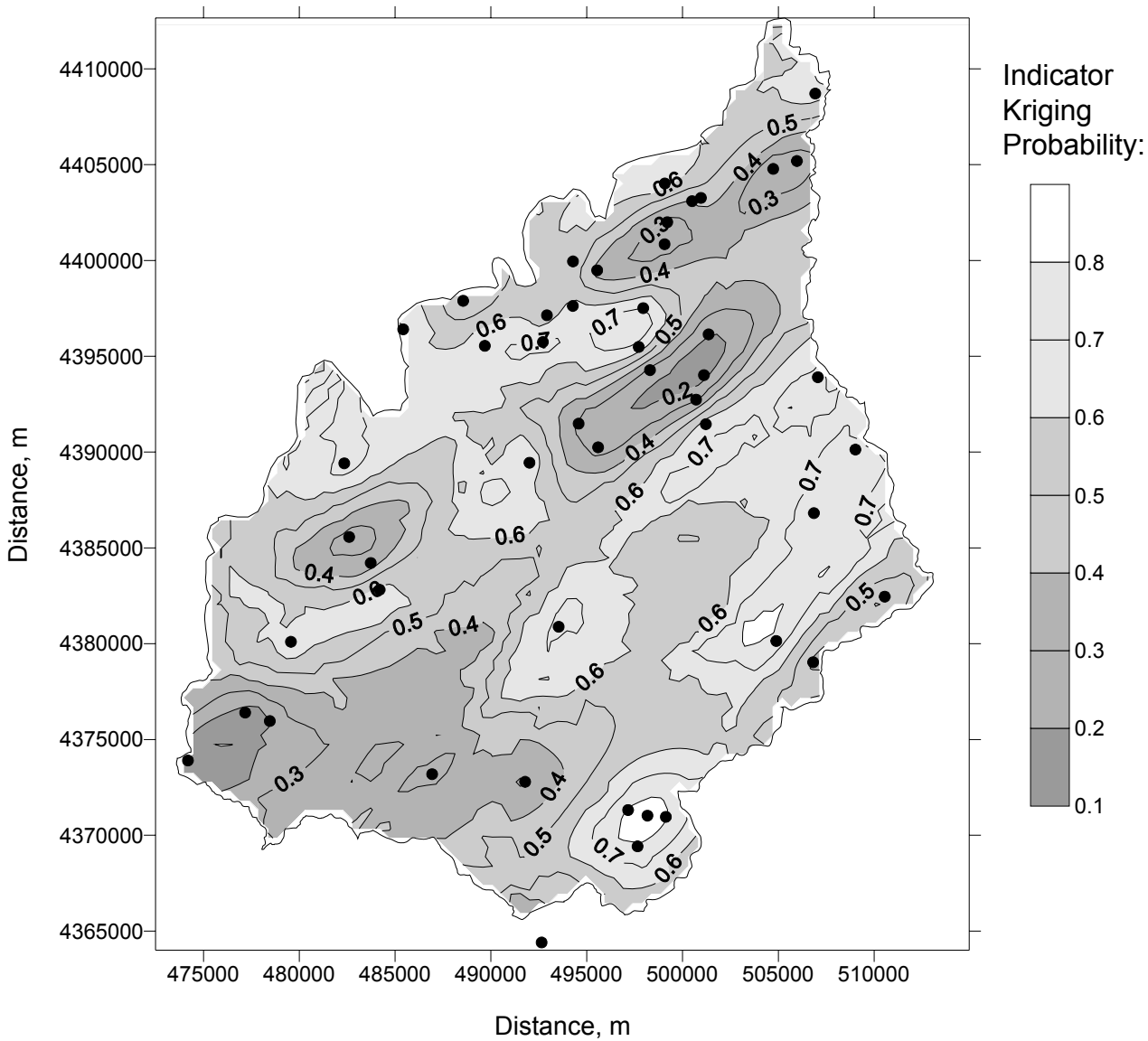


Indicator Transform Based on Median Recharge (29.1 cm/yr)



Spatial dependence—obs. closer in space more alike (variogram).

Probability of Exceeding Median Recharge (29.1 cm/yr)



Conclusions

Statistical methods…

- Effectively use available data at large spatial scales.
- Enable prediction in unsampled areas.
- Reveal bottom-line processes.
- Exploit spatial dependence (geostatistics)

USGS