

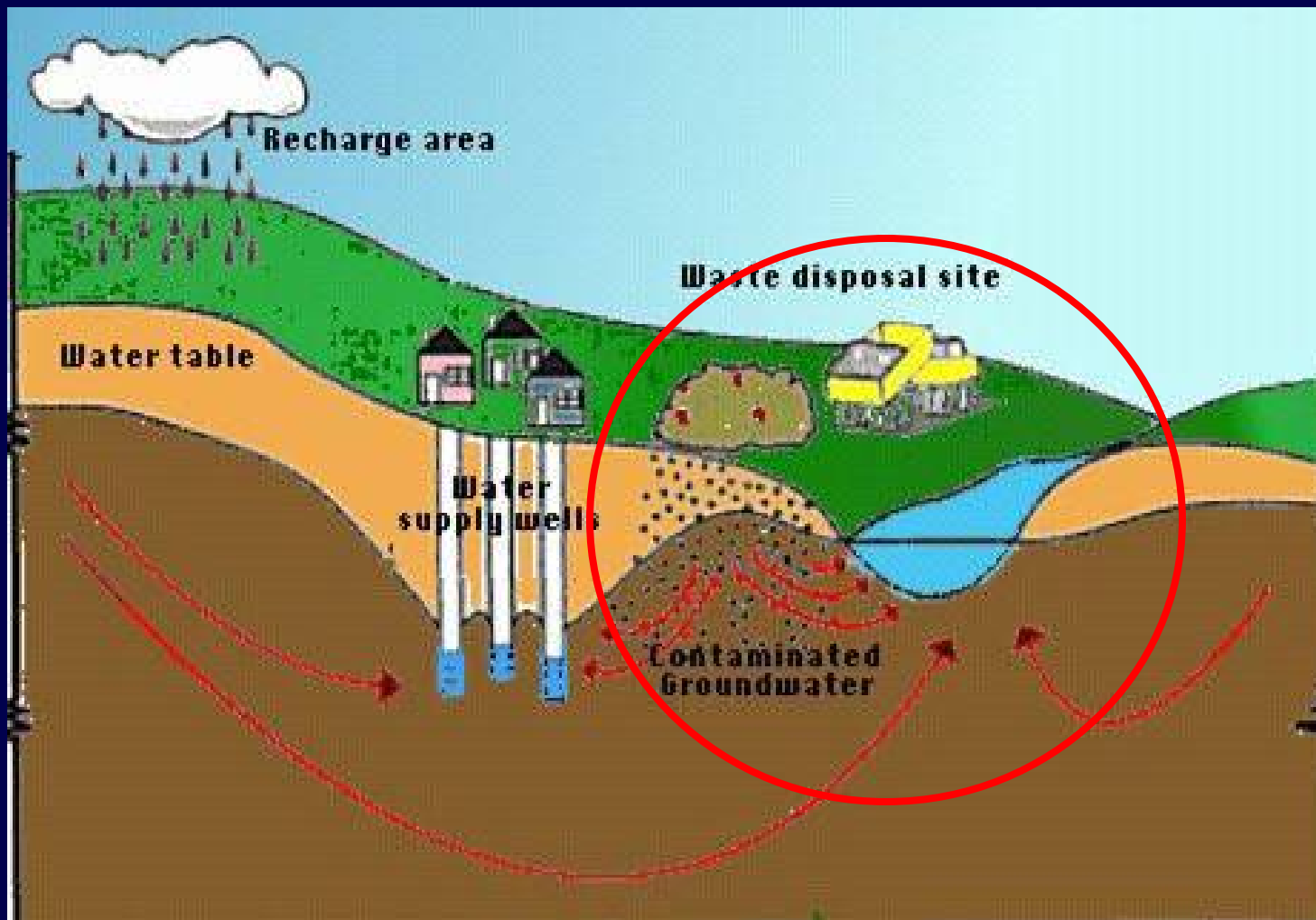
Integrated GIS and Machine Learning Algorithms Applied to Ground Water Contamination Mapping: A Comparative Study

**Barnali Dixon and Nivedita
Candade**

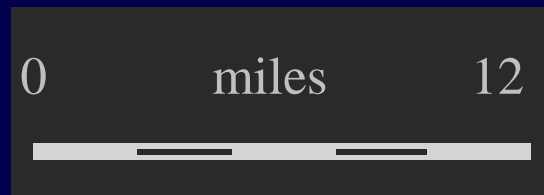
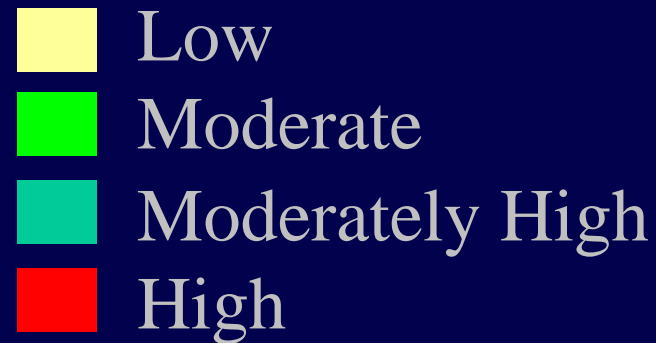
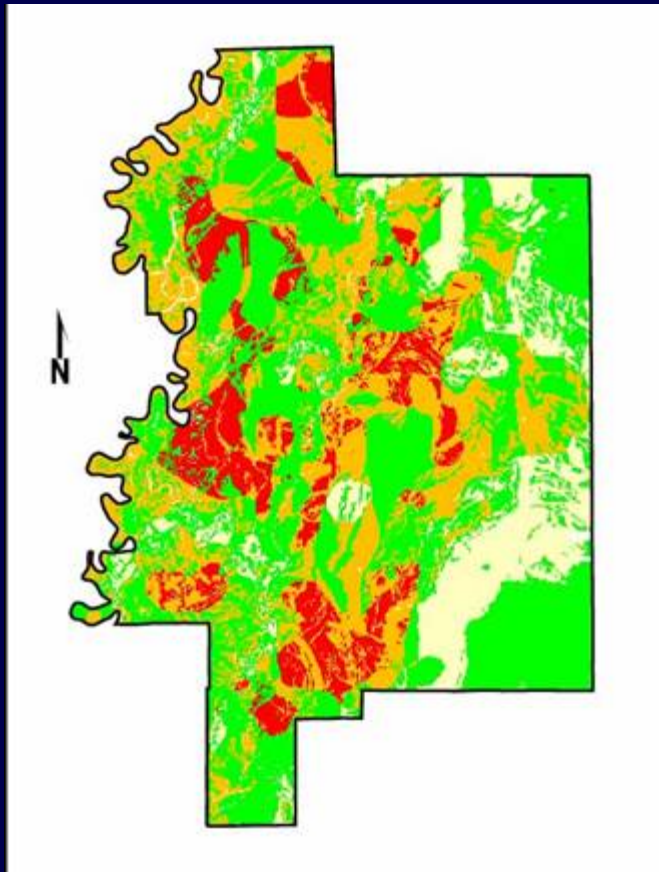
University of South Florida

**Funded by USDA-CSREES: 2001-
35102-10830**

NPS!! The Problem!!



Solution: Contamination Potential Mapping Ex: DRASTIC

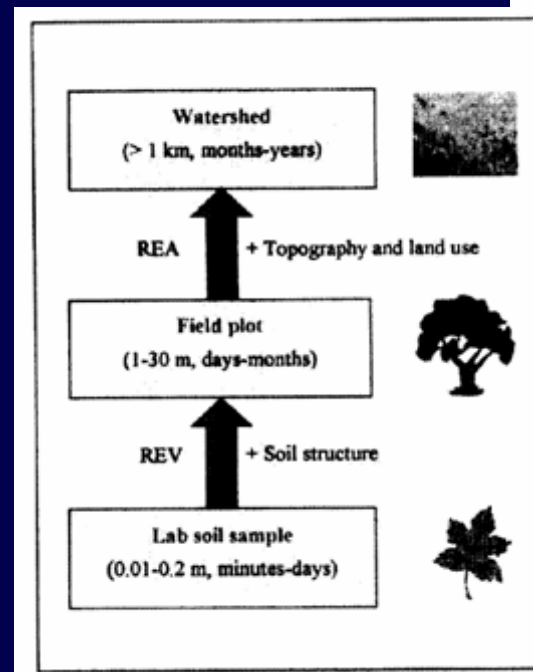
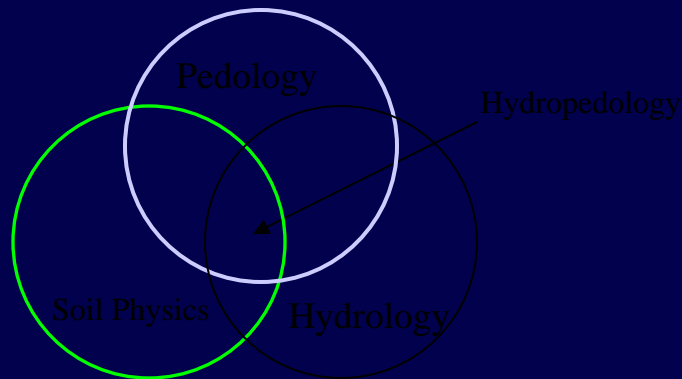
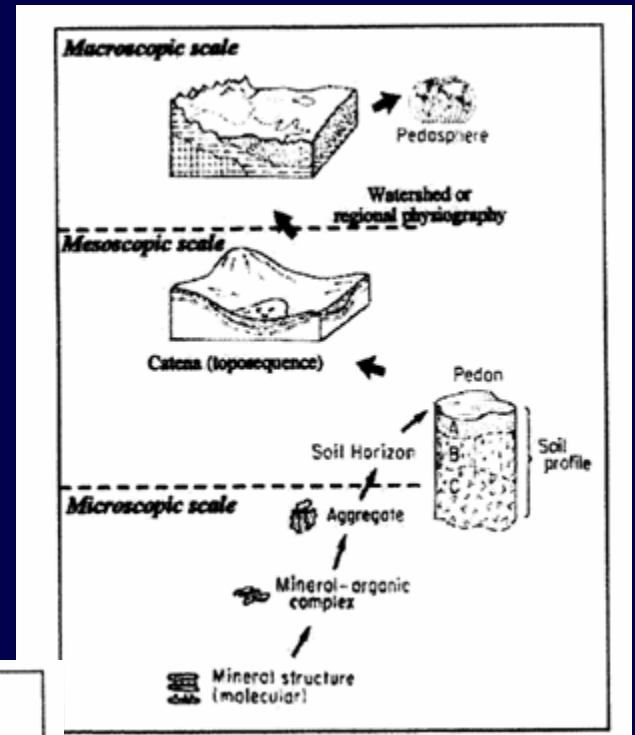


Fundamental Research Issues

- Data need
- Sensitivity of the models (spatial and location)
- Scalability

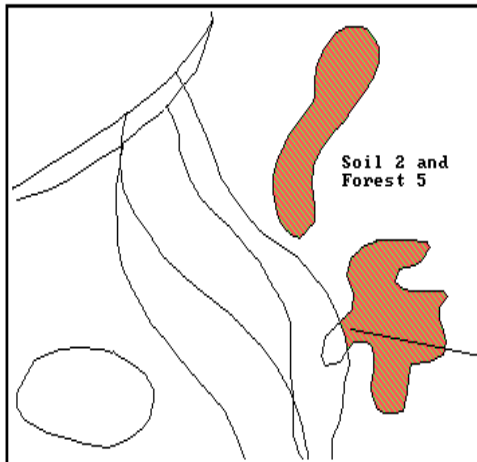
Challenges

- Format Compatibility
- Scale Issues
- Representative Mapping Unit

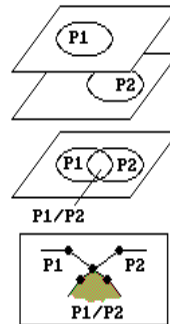


Map Accuracy (Error Propagation)

The topological overlay assumes that the features on both maps are precisely positioned hence the results are precisely positioned.



TOPOLOGICAL OVERLAY

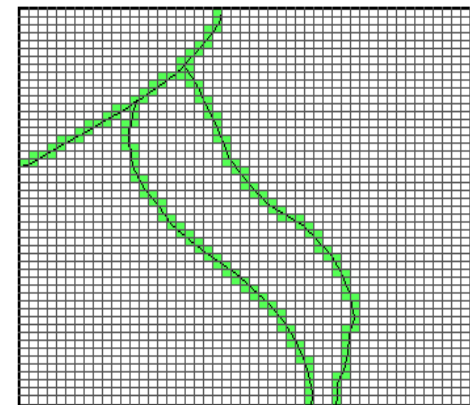


COINCIDENCE SEARCH

100% certain anywhere in the derived polygon?
...be honest, there is a lot of uncertainty

Map Accuracy (Error Propagation)

Uncertainty in boundary positioning can be treated as a gradient of certainty (fuzzy line) creating a zone of transition.



Using a Raster format, each cell containing the implied boundary is identified. These locations are the most uncertain as they form the immediate transitions among map features.

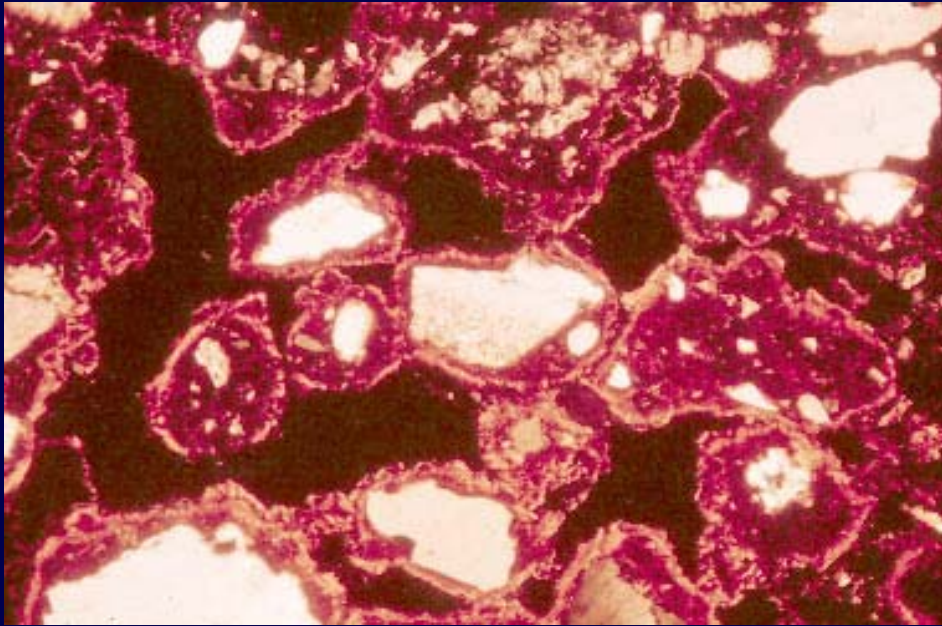
SOIL
Soil



Challenges

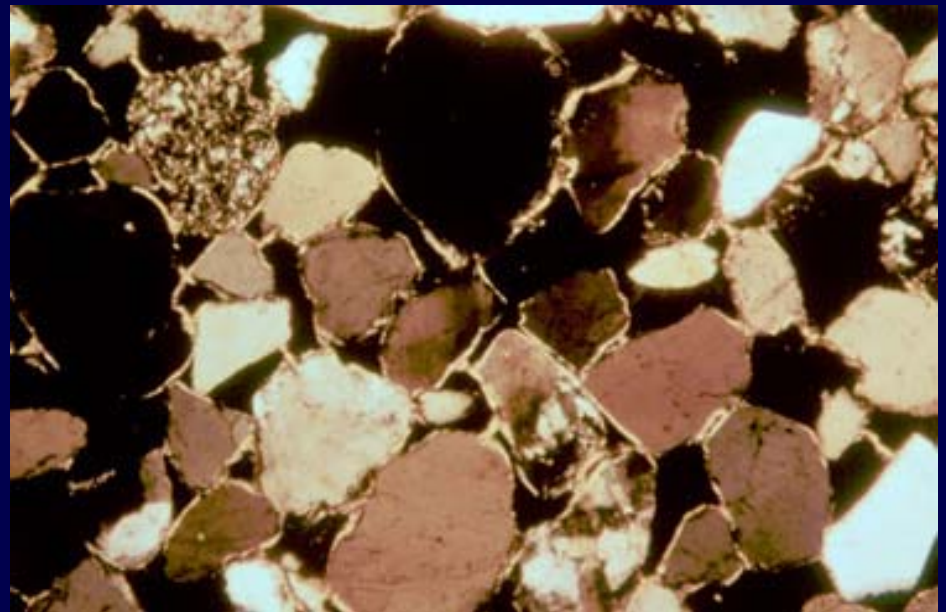
Challenges

**How are we going to
organize the data??**



**Photomicrograph of a thin
section of soil structure**

How do we incorporate
this layer in the model??

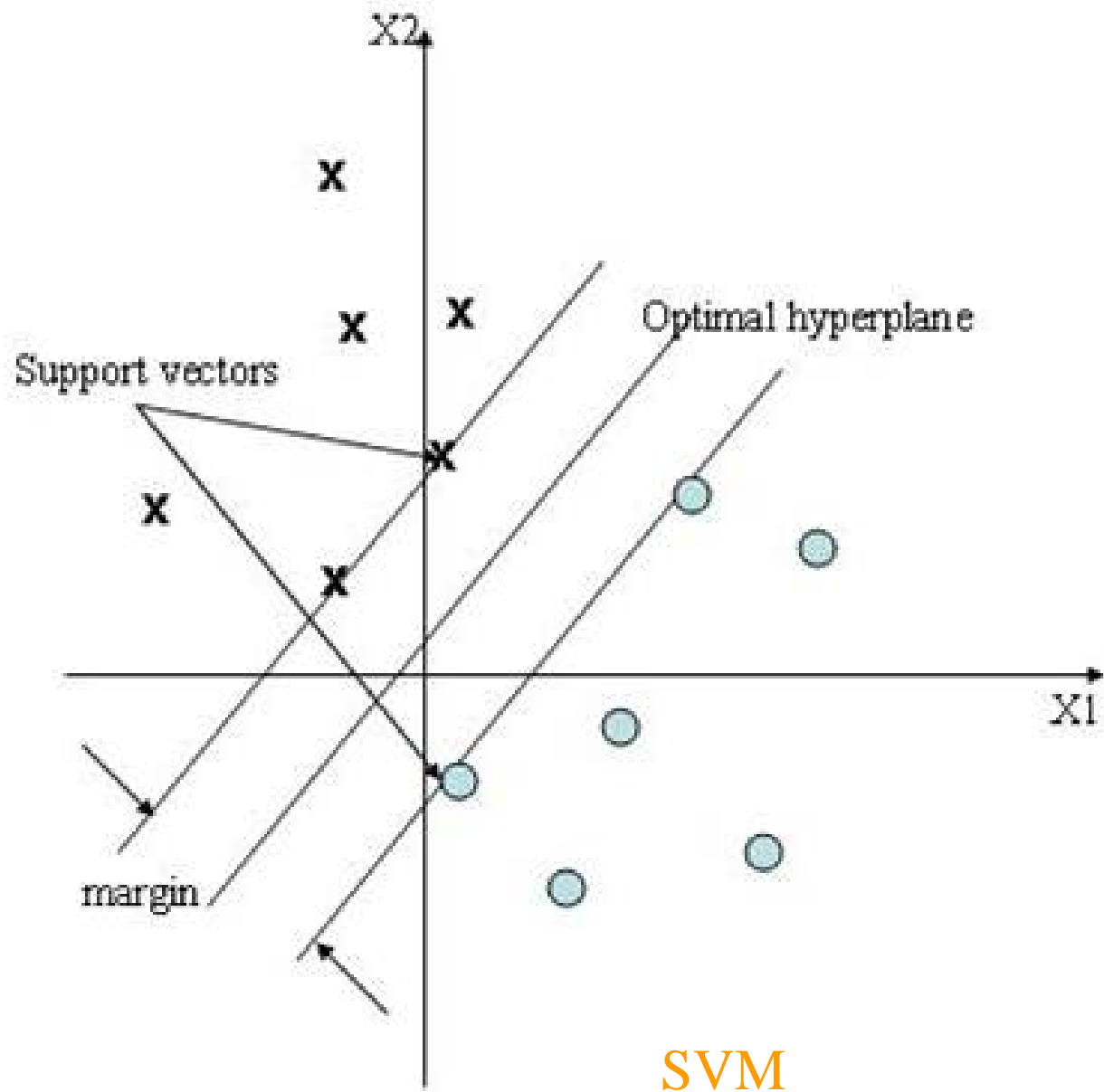


Model Characteristics

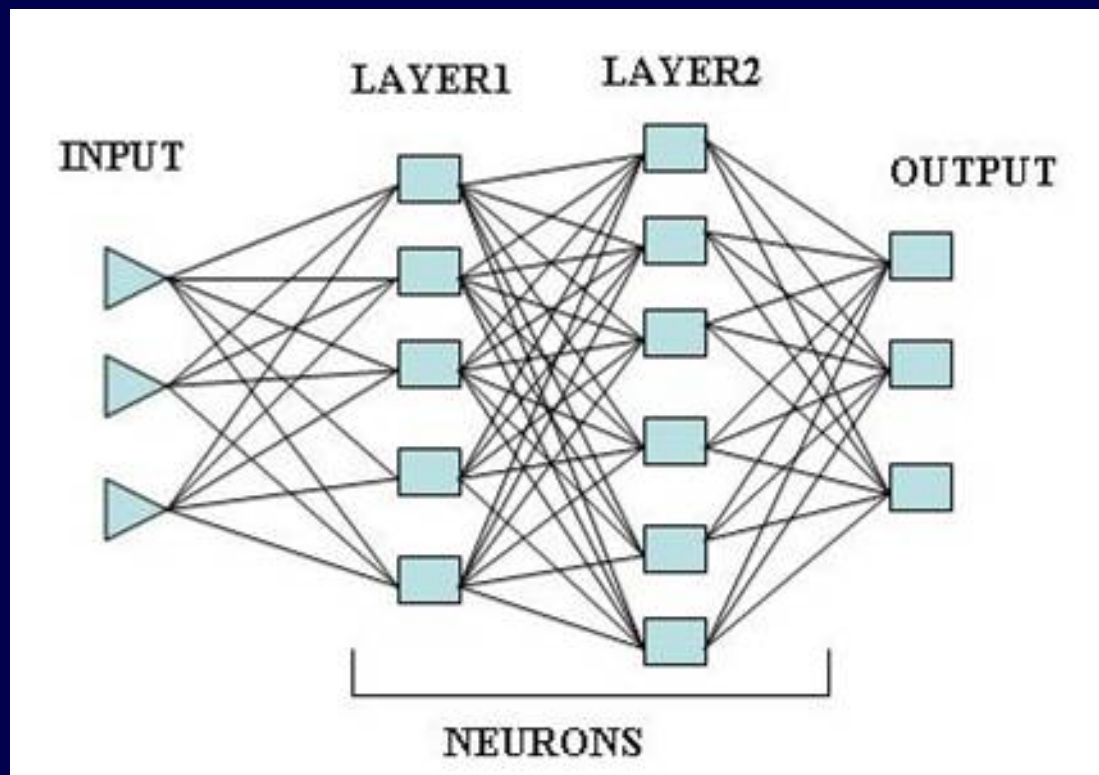
- Flexibility
- Uncertainty
- Robust

Machine Learning Algorithms

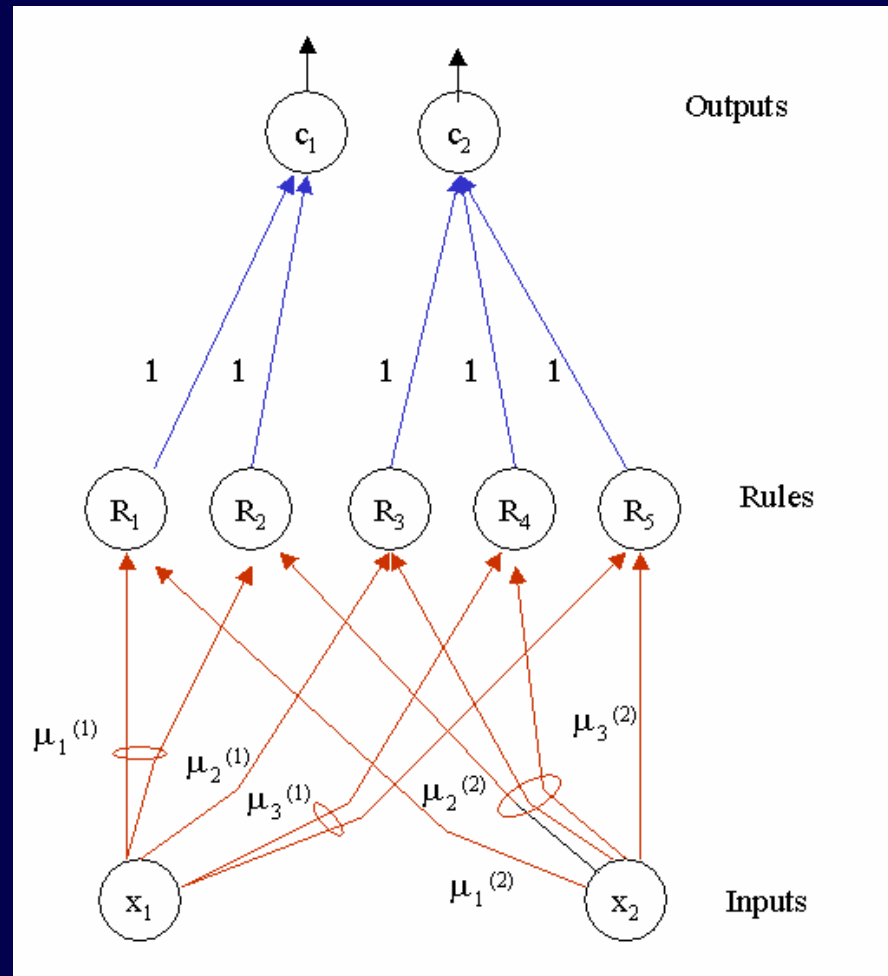
- Neural Networks (NN)
- Support Vector Machine (SVMs)
- Neuro-Fuzzy



Neural Networks



Neuro-Fuzzy



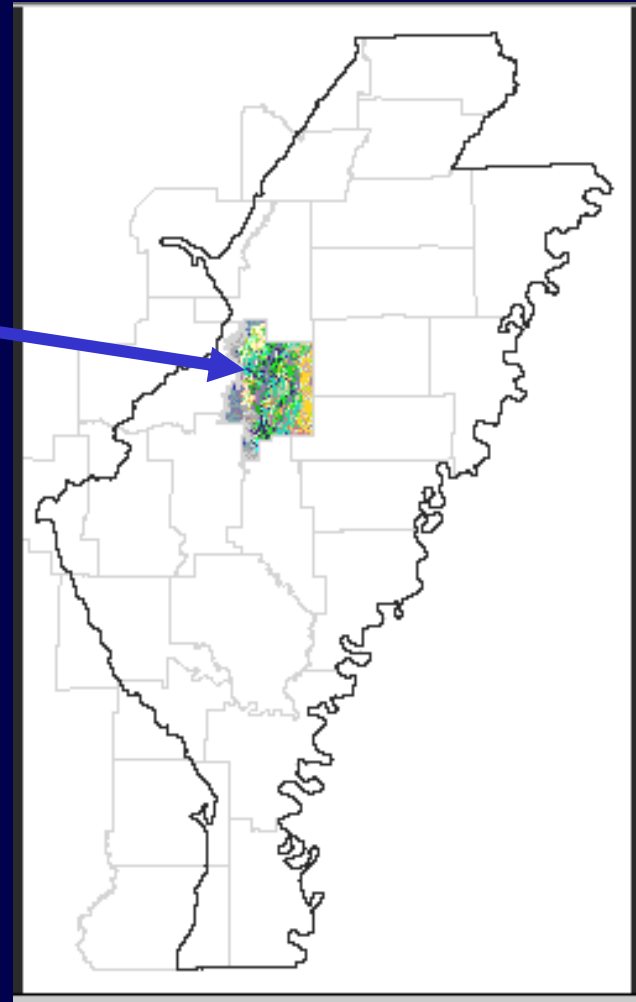
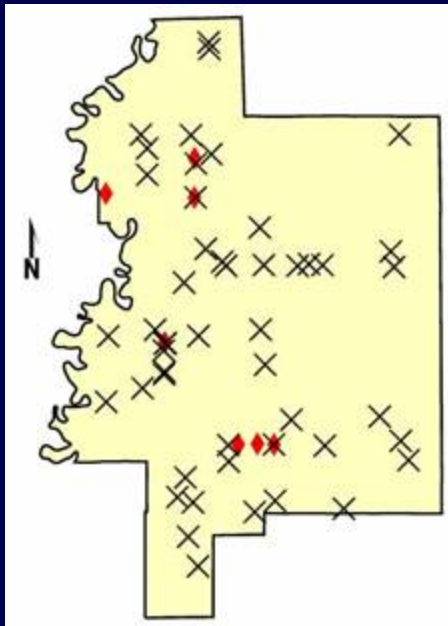
Software

- Neural Network: PREDICT (Neuralware, 2001, version 2.4)
- Support Vector Machine: LIBSVM (Chang and Lin, 2003, version 2.5)
- Neuro-fuzzy: Nefclass-J (Nauck and Kruse, 1999 version 1.0)

Models and Variables

- Model-1: D, R, S, I
- Model-2: Pedality, drainage, hydrologic group, BD
- Model-3: pH, OM, BD

Location of Woodruff County, AR

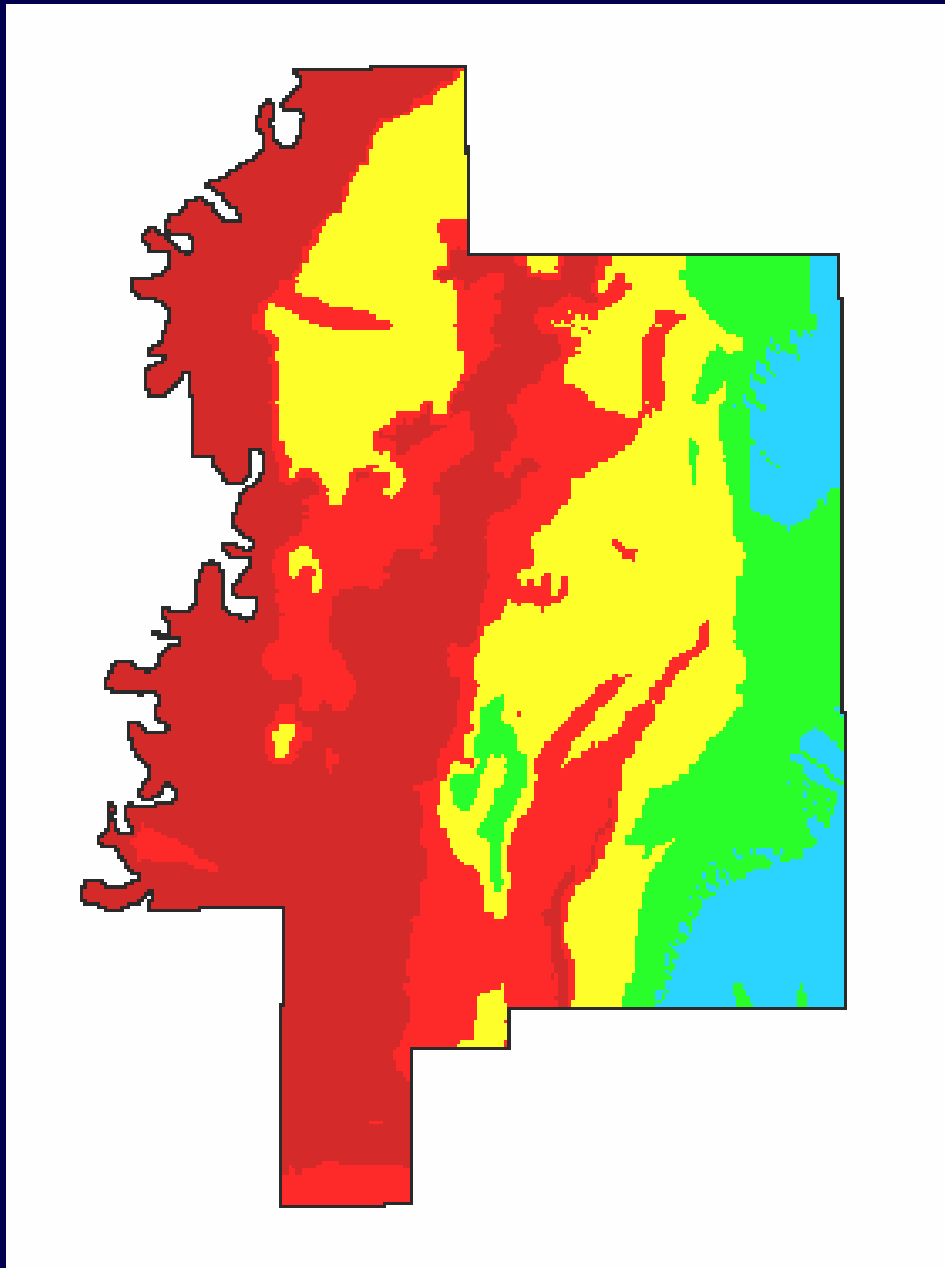


Spatial Data Layers

- Well Location/Contamination
- Soils
- Geology
- Thickness of the Confining Unit (claycap)
- Recharge of Ground Water
- Depth to GW

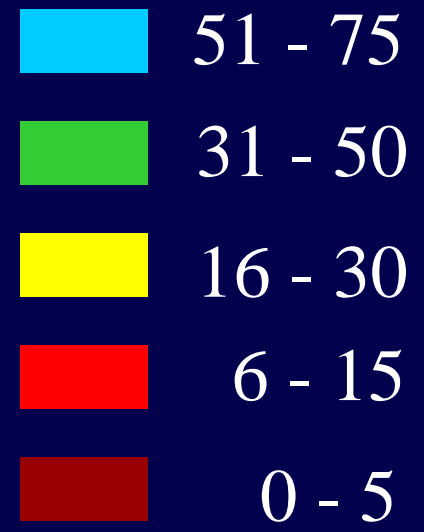
Sources of Spatial Data

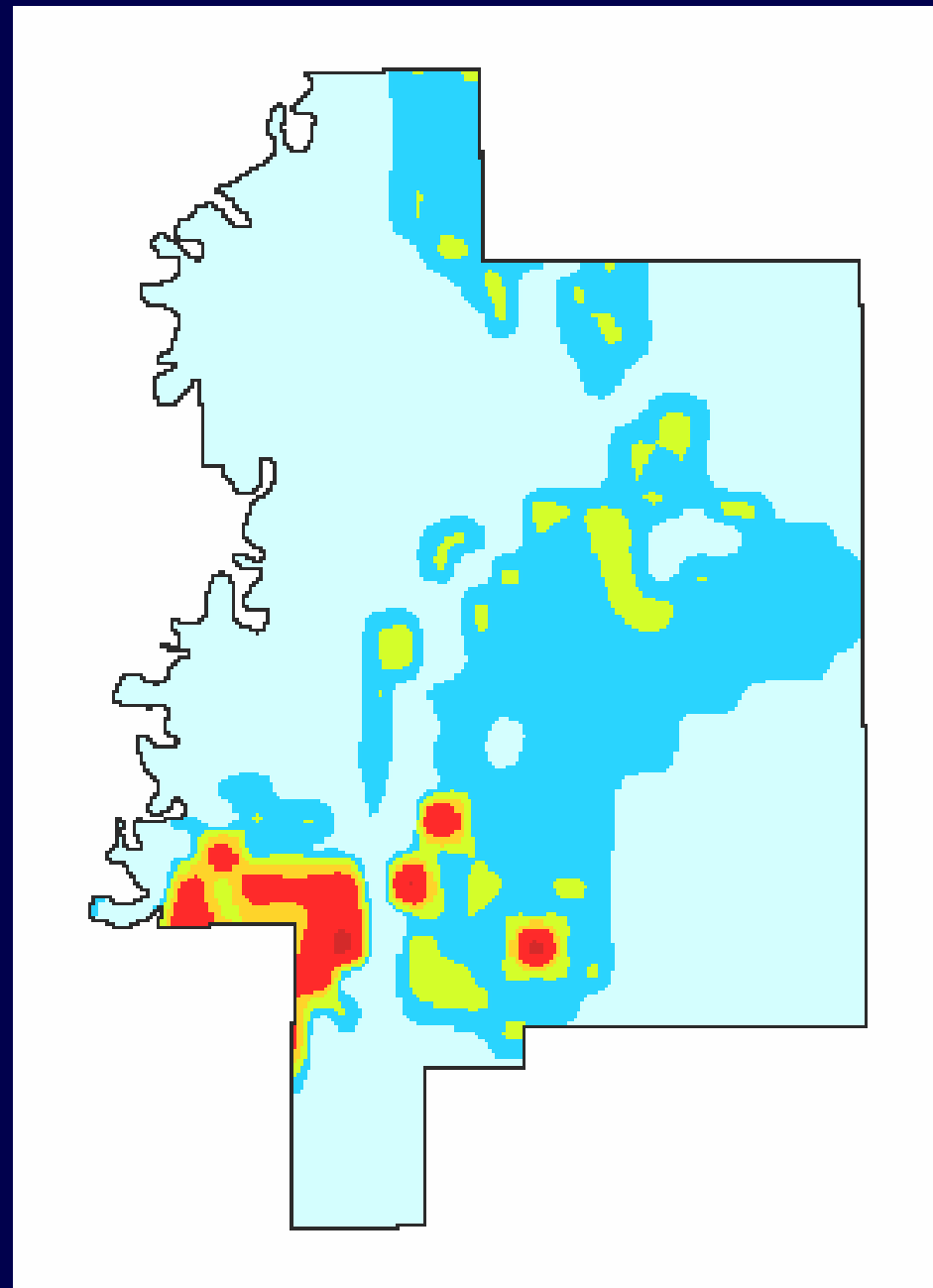
- Well Location/Contamination DEQ
- Soils NRCS
- Geology LGS
- Claycap USGS
- Recharge USGS
- Depth to GW USGS



D

ft

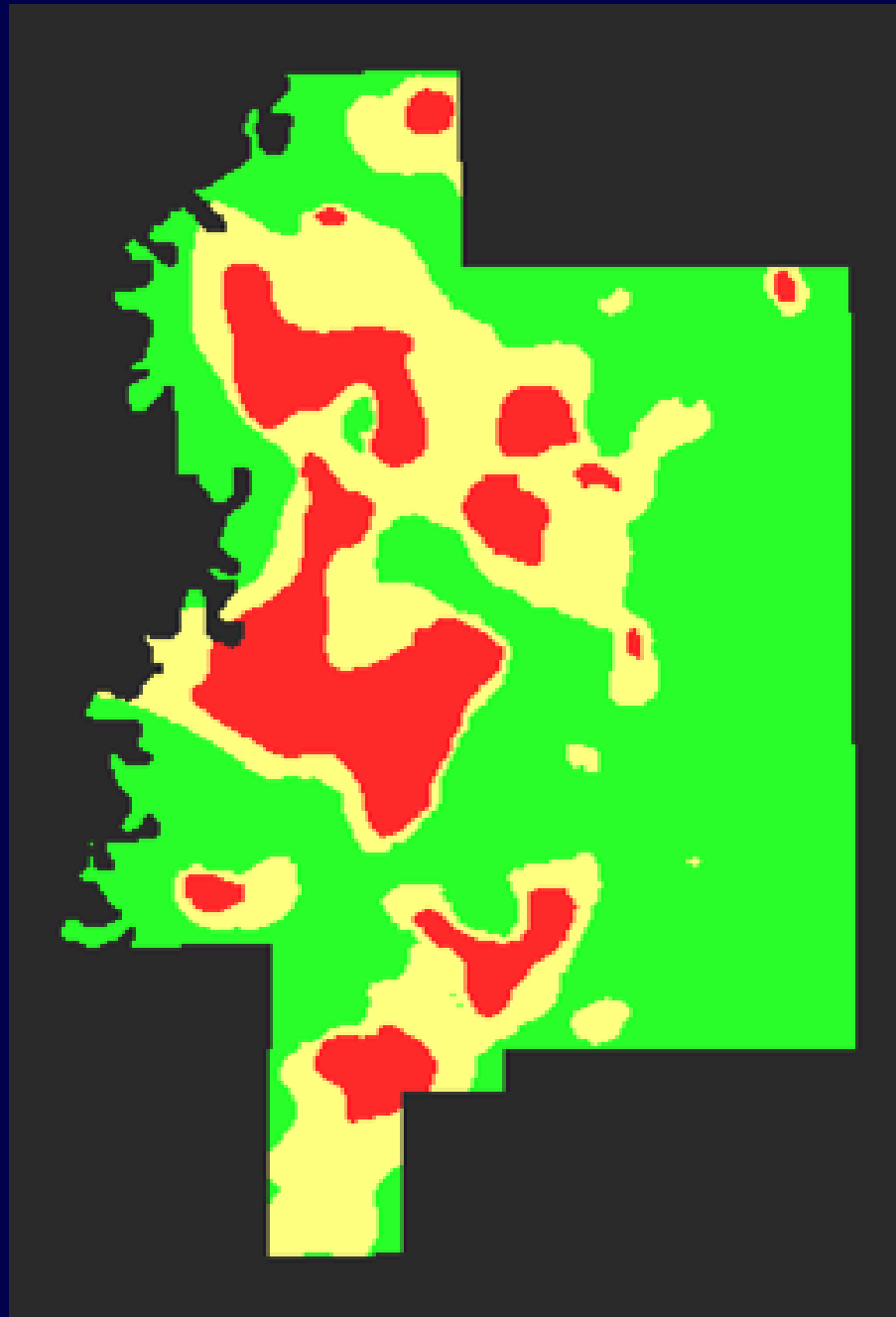




R

in/yr





I

ft



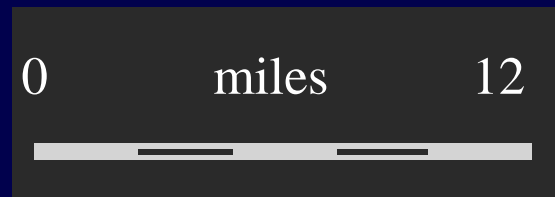
> 10

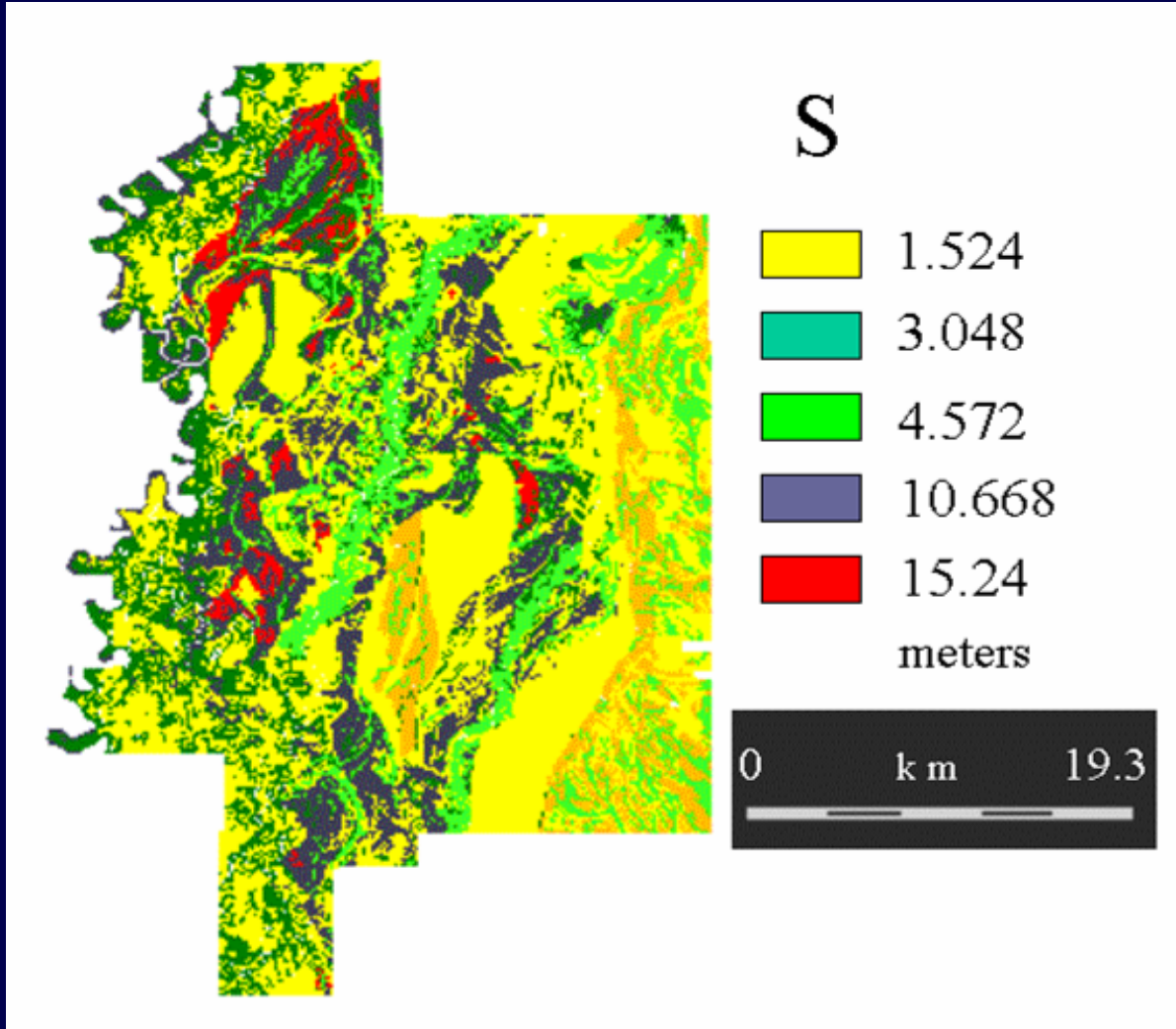


1 - 10

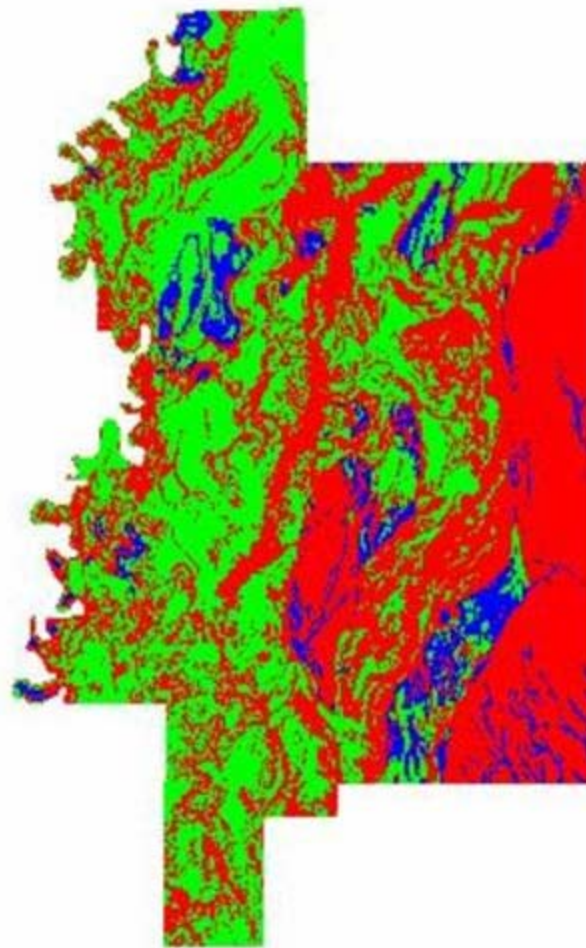


0





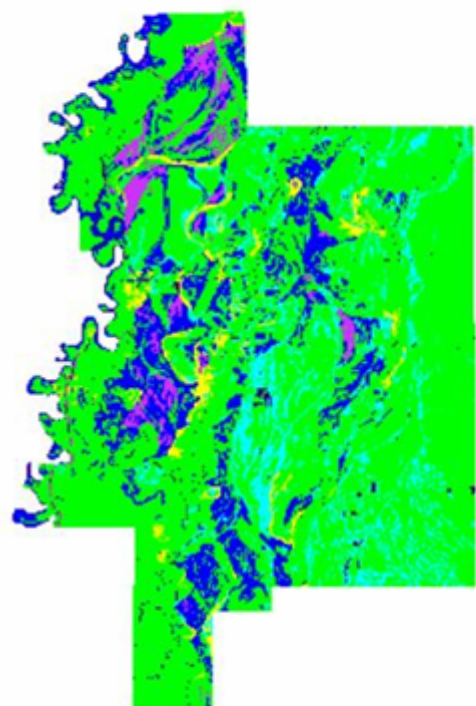
Pedality



- Low (14-18)
- Mod (19-35)
- High (>35)



Drainage

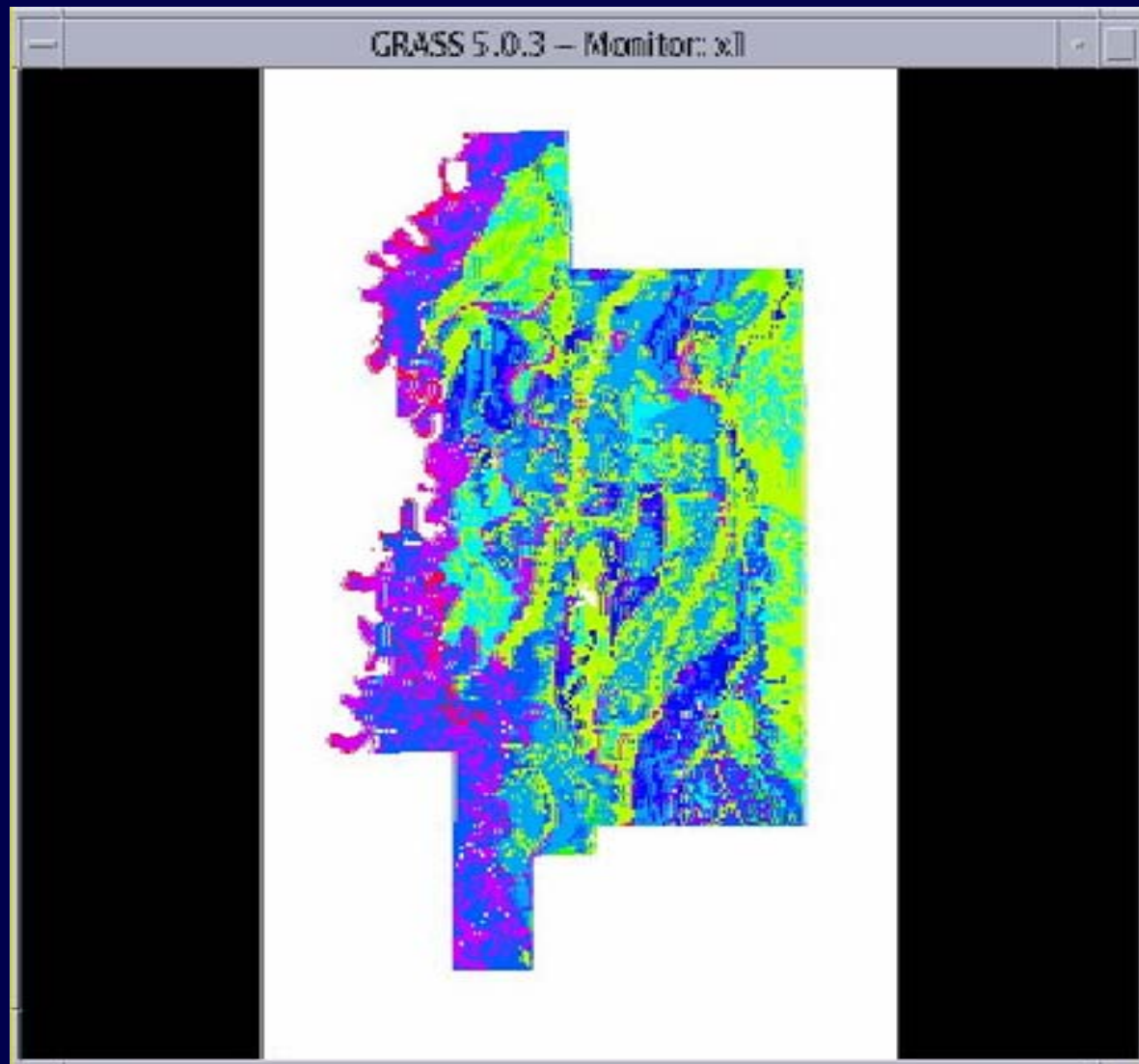


-  Very poor
-  Somewhat poor
-  Poor
-  Moderately well
-  Well
-  Somewhat excessive
-  Levee
-  Water

0 k m 19.3

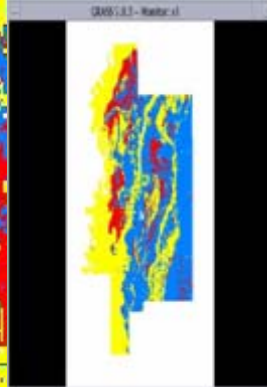
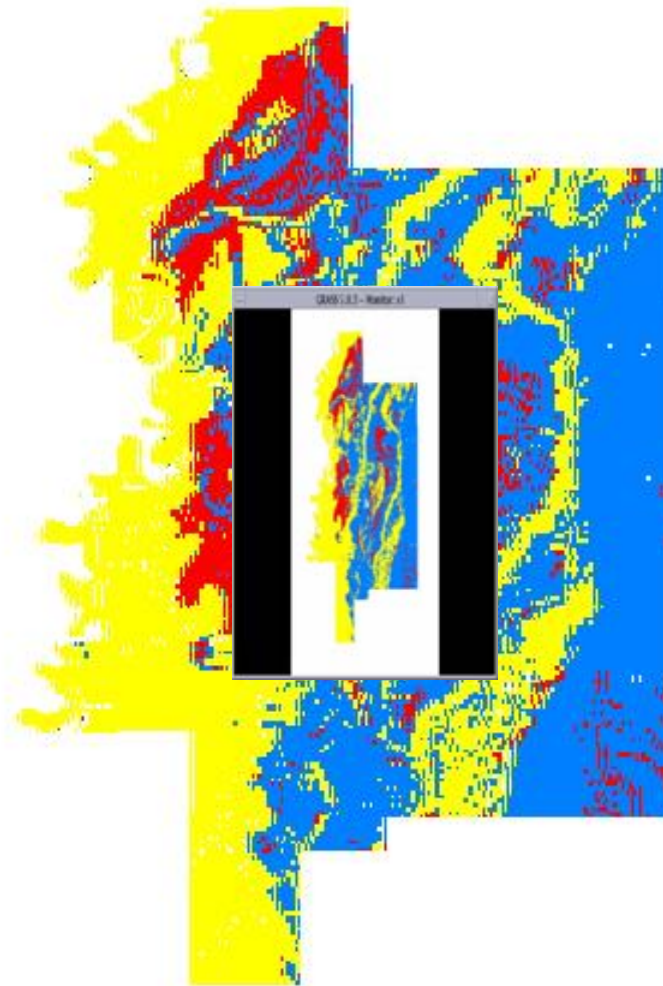


pH





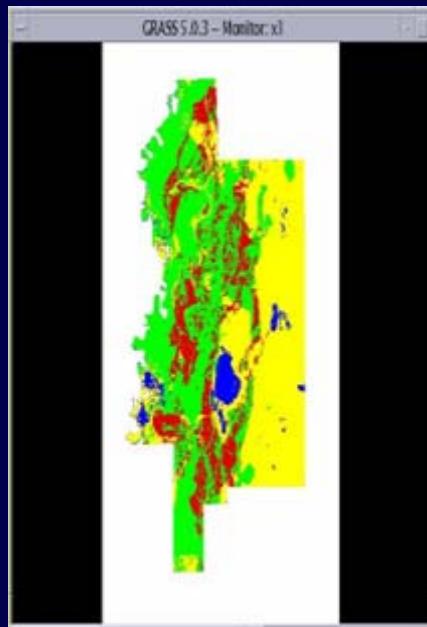
4 0112



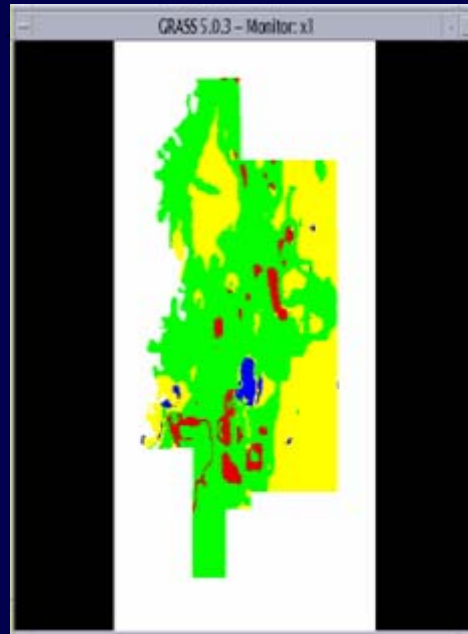
Results

DRSI

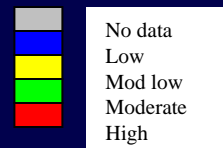
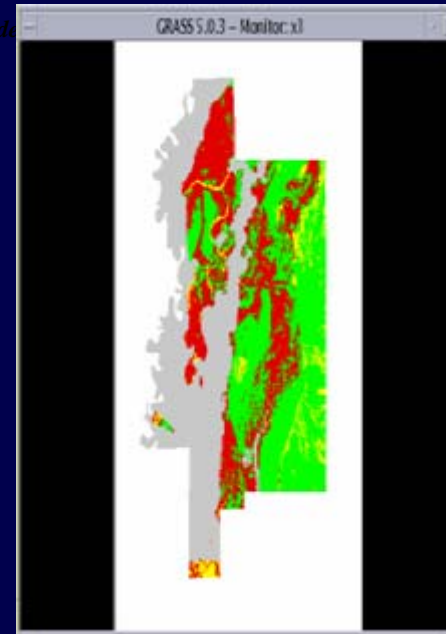
NN



SVM



NF

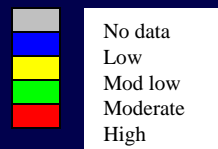
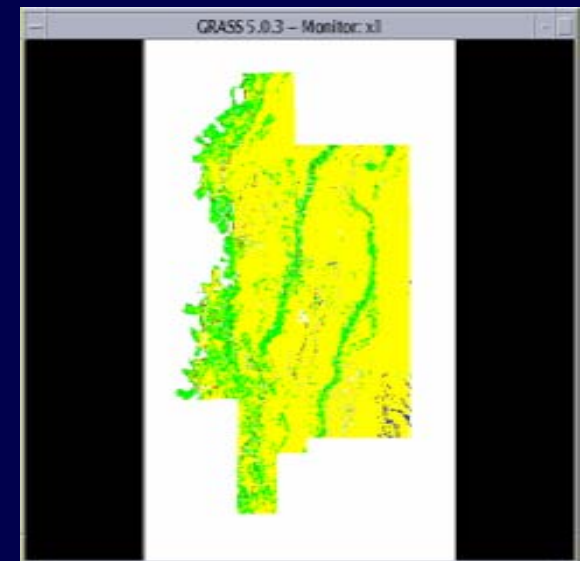
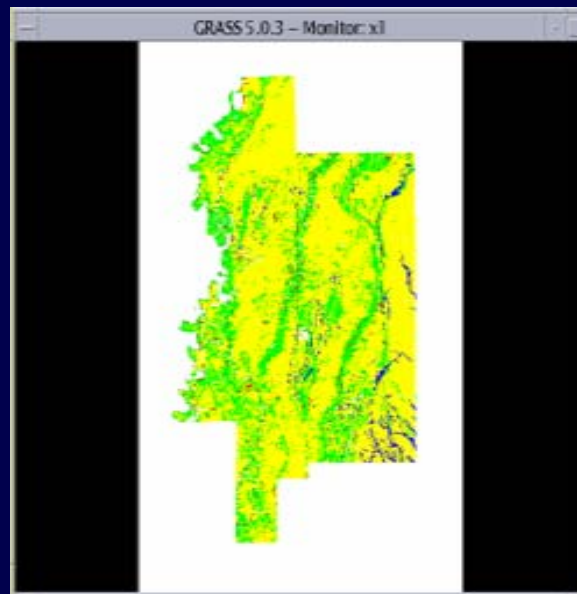
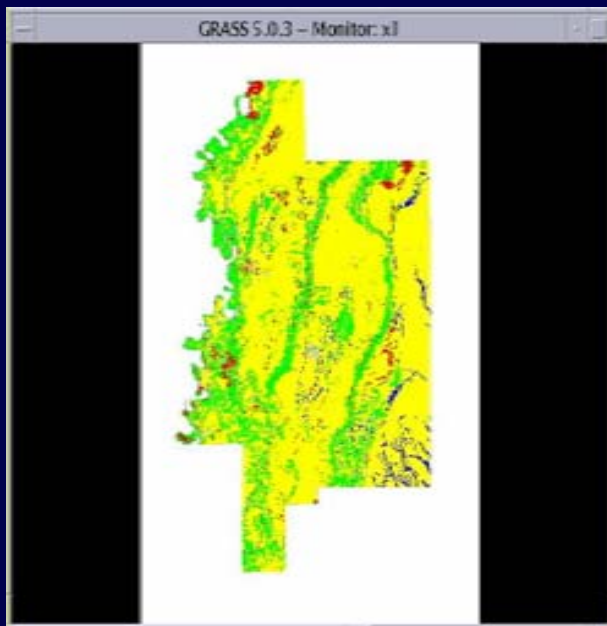


Pedality, Drainage, Bulk Density

NN

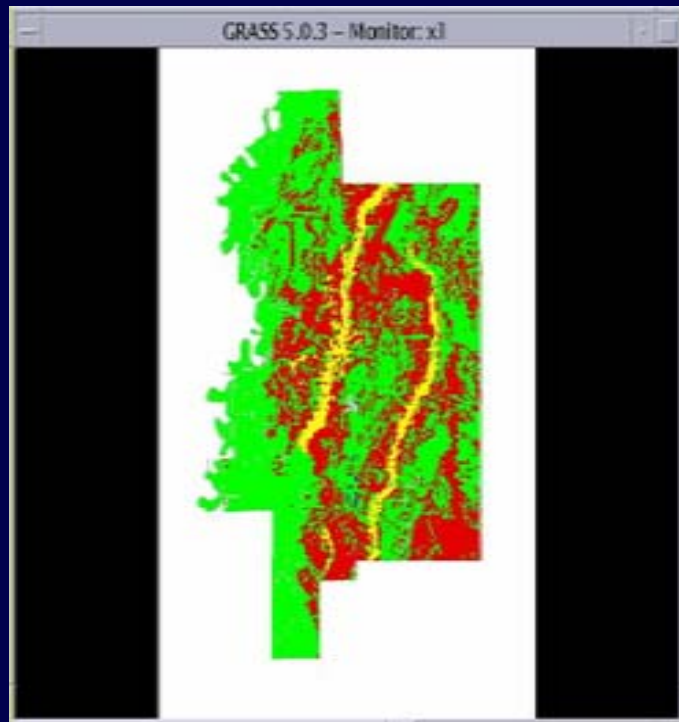
SVM

NF

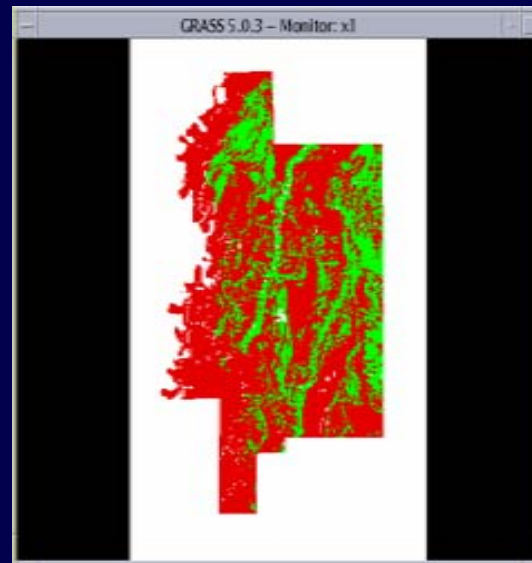


pH, OM, BD

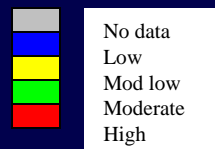
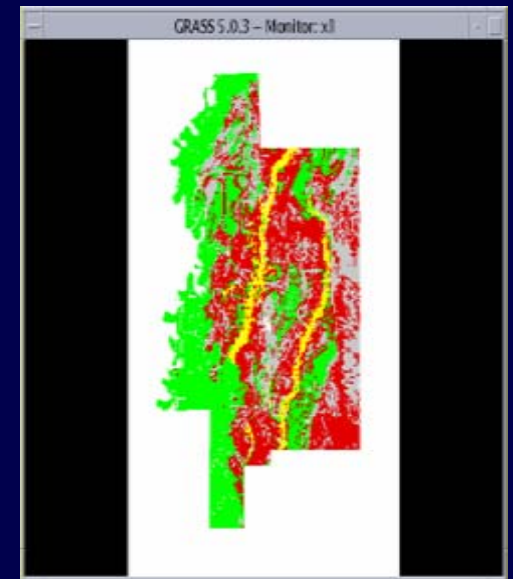
NN



SVM



NF



FROC Curve

	+1	-1
+1	<i>TP</i>	<i>FN</i>
-1	<i>FP</i>	<i>TN</i>

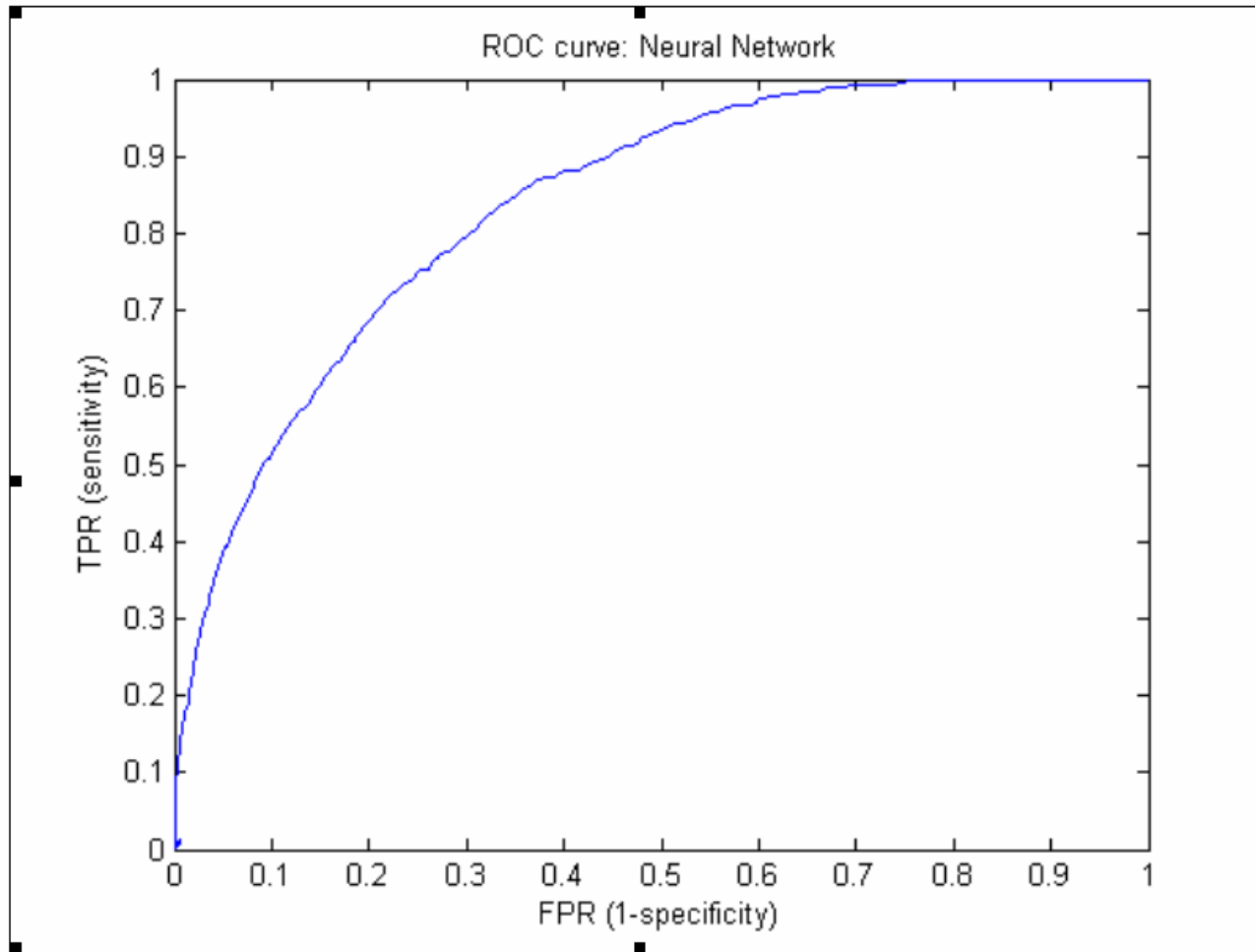
TP = number of correct predictions that an instance is positive

FP = number of incorrect predictions that an instance is positive

TN = number of correct predictions that an instance is negative (or zero)

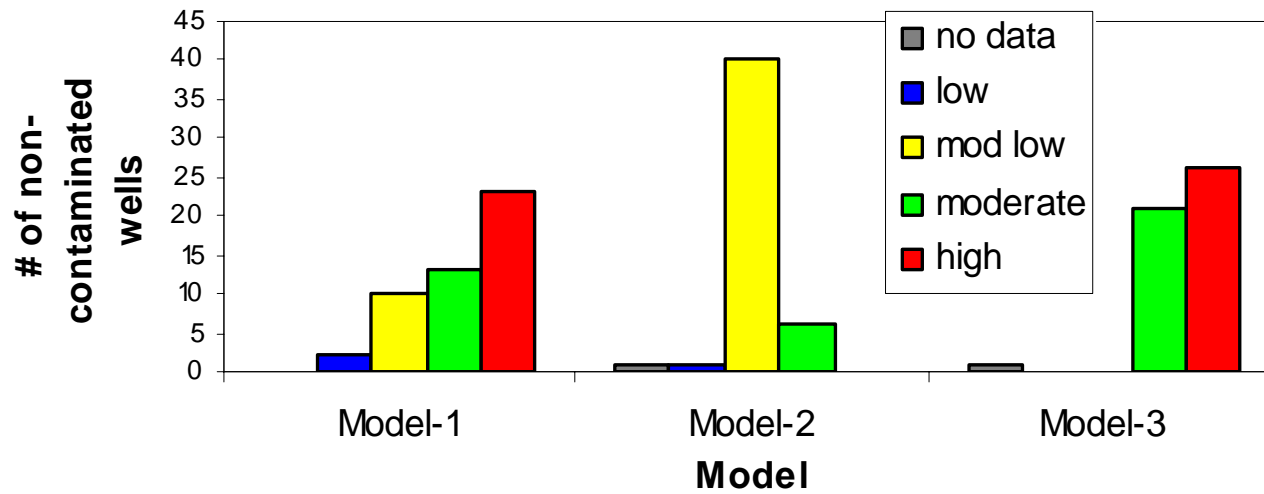
FN = number of incorrect predictions that an instance is negative (or zero)

1
2

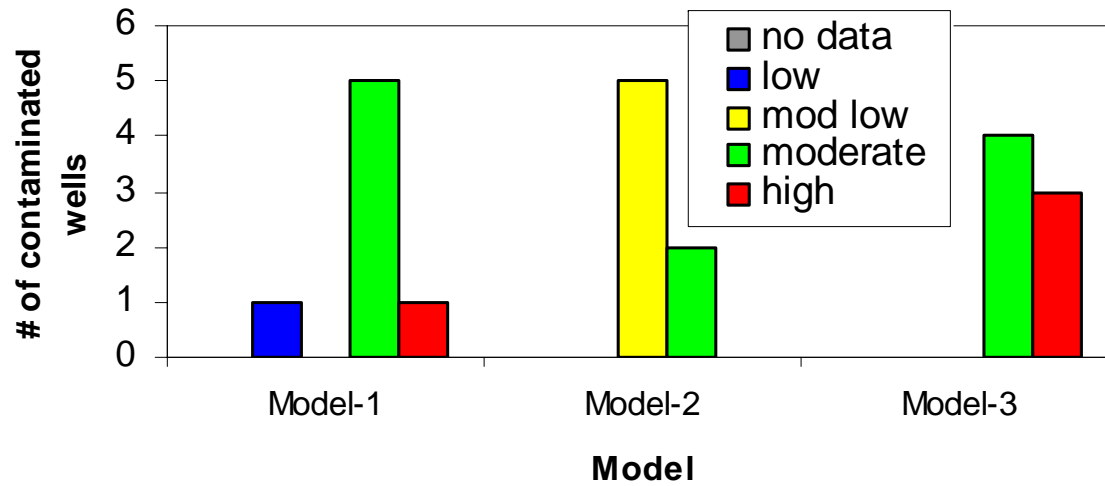


Area under Curve (AUC) = 0.8385

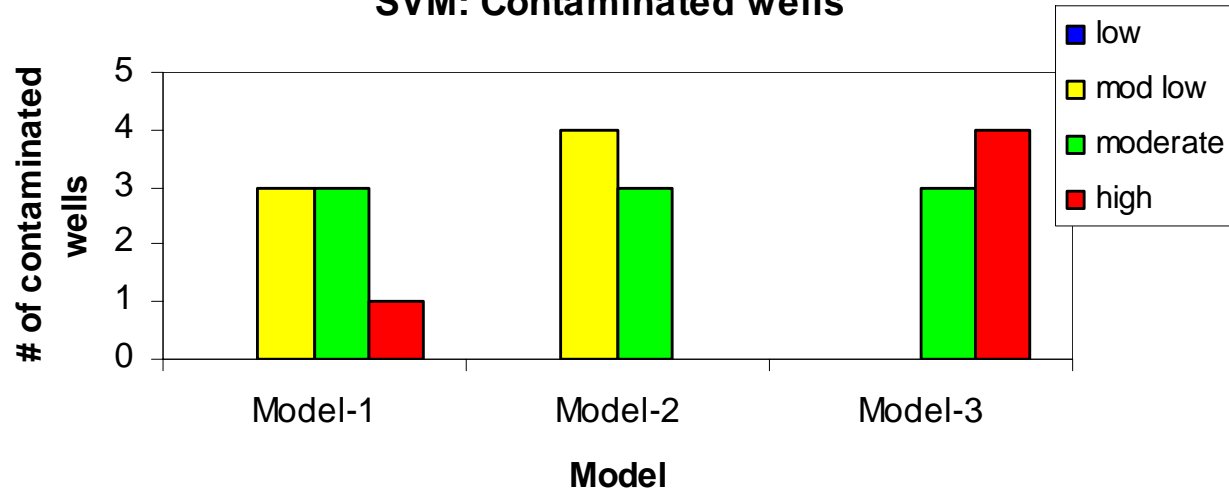
NN: non-contaminated wells



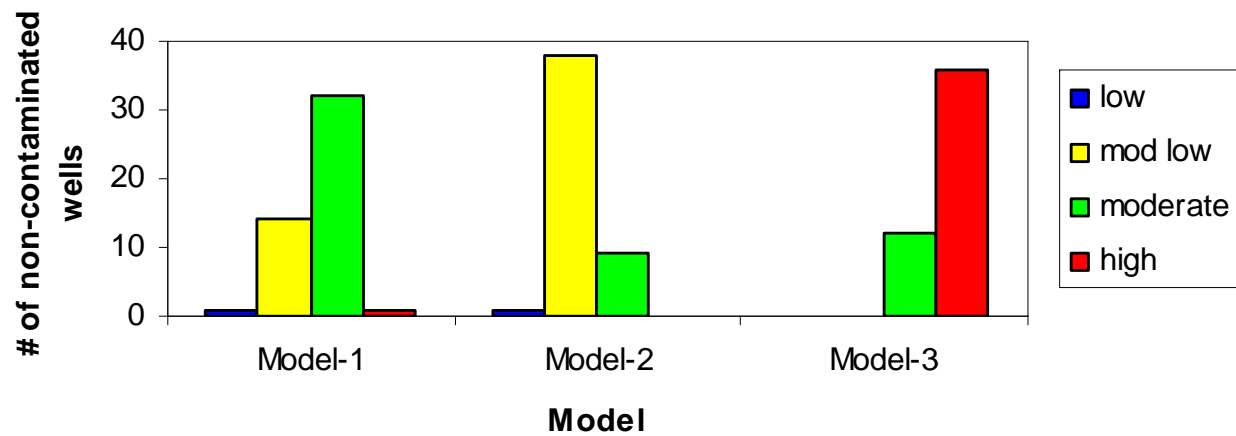
NN: Contaminated wells



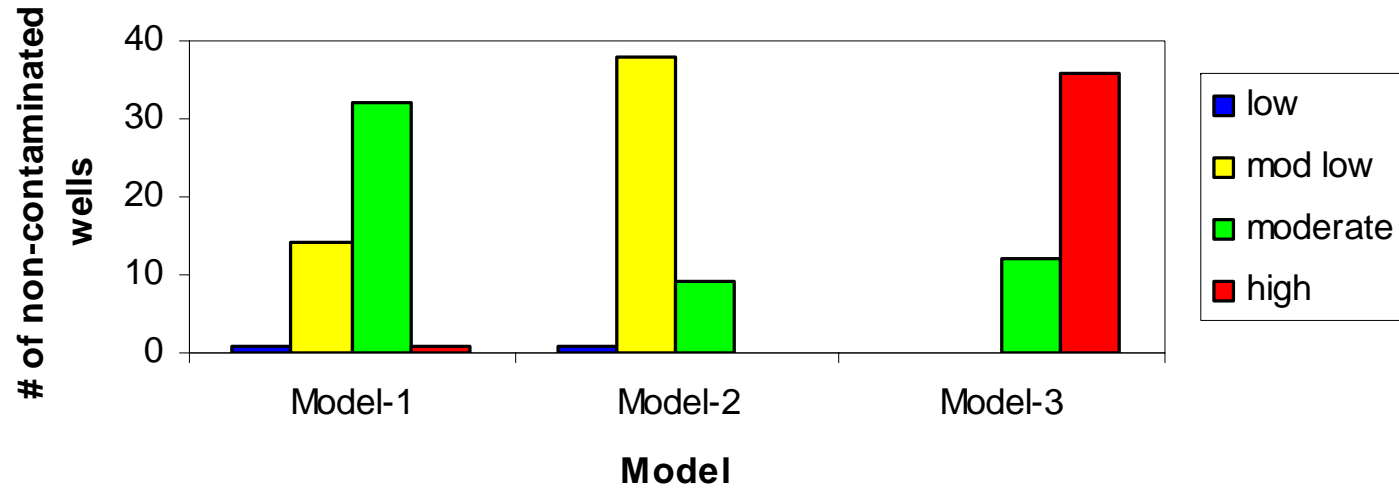
SVM: Contaminated wells



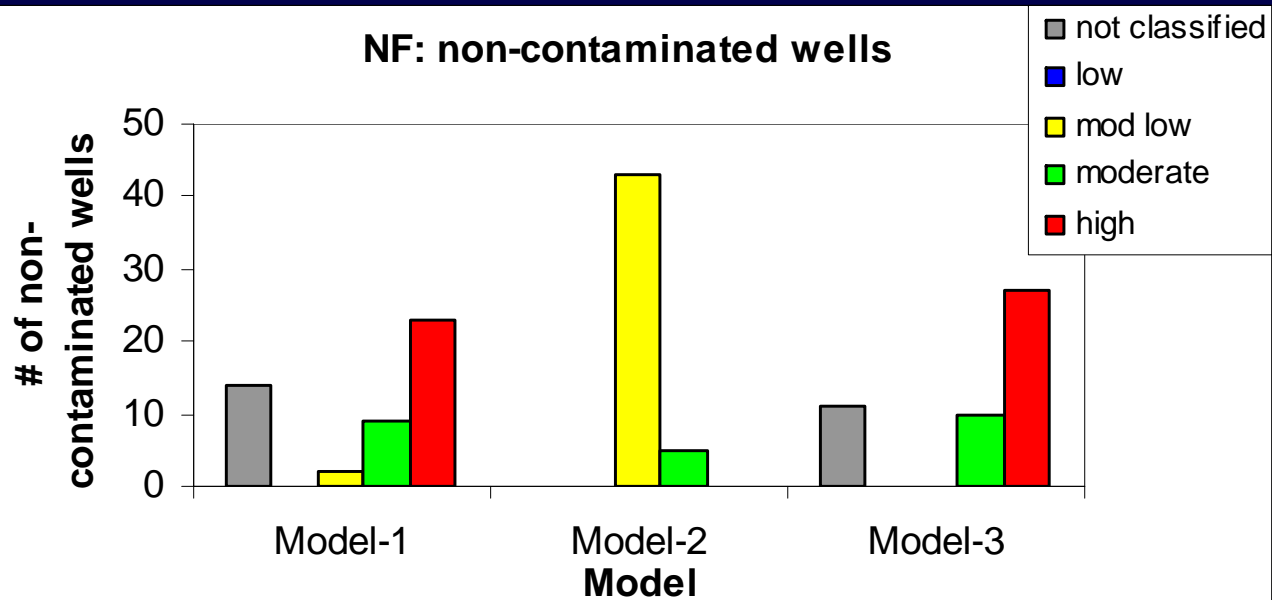
SVM: non-contaminated wells



SVM: non-contaminated wells



NF: non-contaminated wells



Summary

- Which one is better??
 - More sensitive to input variables than models.
 - How can we find out which variables are more meaningful than other (without investing the money)