Determining Appropriate Size of the Training Data Sets for Neuro-fuzzy Models to Predict Ground Water Vulnerability in Northwest Arkansas

Barnali Dixon[1] & H. D. Scott[2]

[1]University of South Florida

[2] University of Arkansas

Introduction

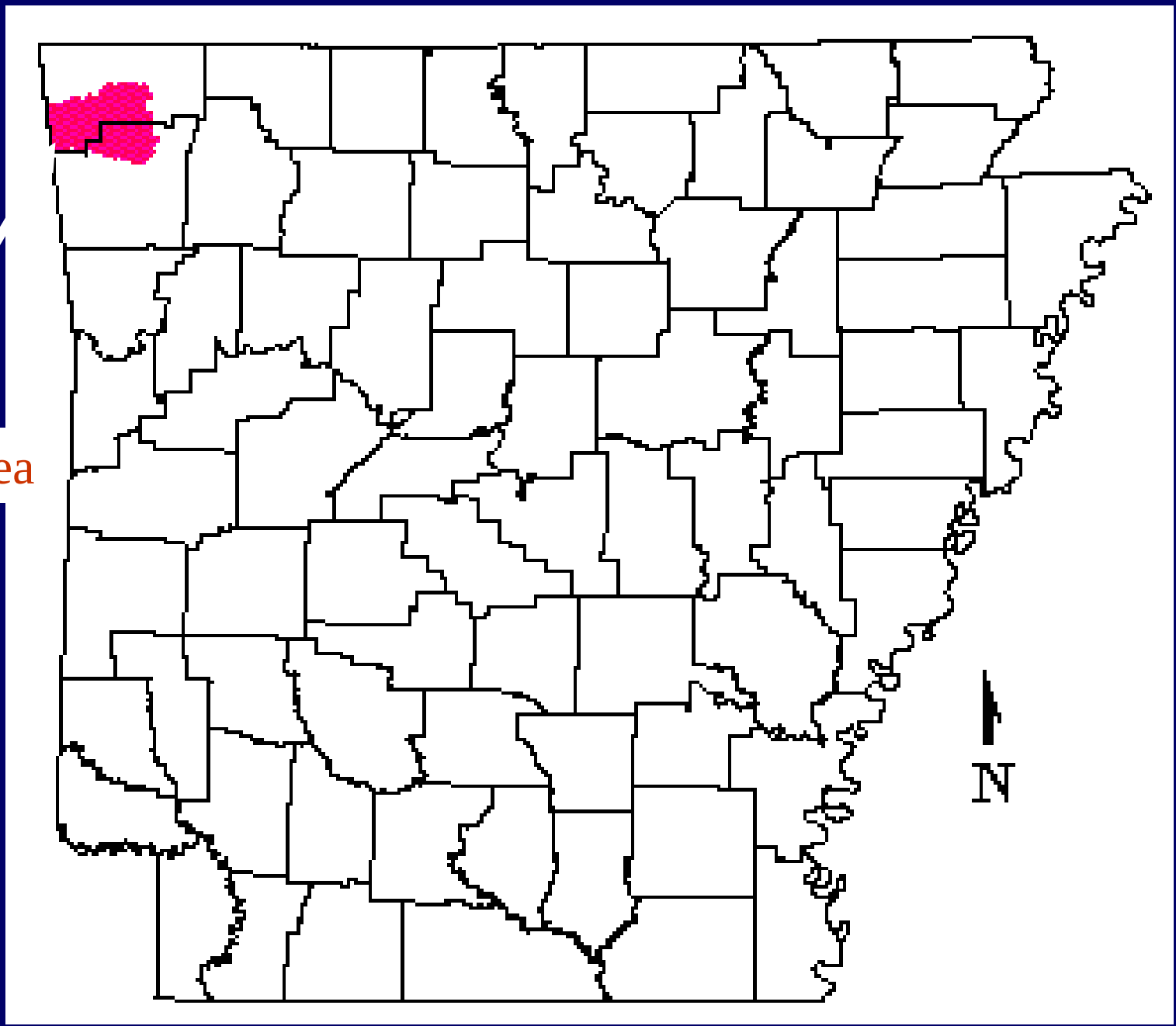
- Delineation of vulnerable areas and selective applications of animal wastes/fertilizer (AW/F) in those areas can minimize contamination of GW.
- However, assessment of GW vulnerability or delineation of the monitoring zones is not easy since uncertainty is inherent in all methods of assessing GW vulnerability

Sources of Uncertainties

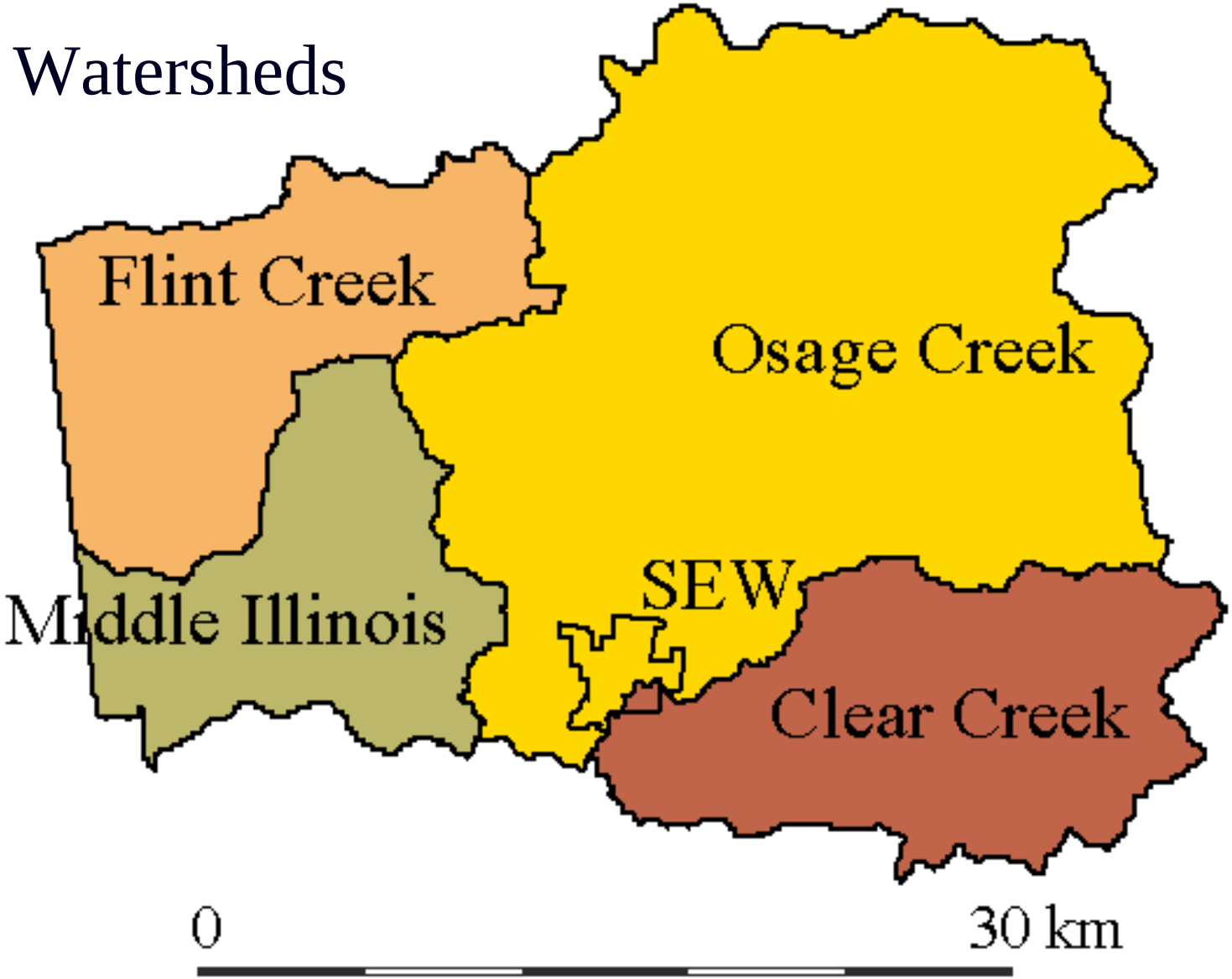
- Errors in obtaining data
- The natural spatial and temporal variability of the hydrogeologic parameters in the field
- The numerical approximation and computerization

Specific Objectives

- To develop Neuro-fuzzy models with the inherent capabilities to deal with uncertainty and to integrate soil hydrologic parameters and LULC in a GIS
- To determine the effects of the size of the training data sets on Neuro-fuzzy model predictions


Study Area
N

Watersheds


Flint Creek
Osage Creek
SEW
Middle Illinois
Clear Creek
0
30 km

Characteristics of the Models

Models	Training data	Application data	Scale
Model1_savoy	SEW	SEW	Field to Field
Mode2_savoy	SEW	Watershed	Field to Watershed
Model3_savoy	Watershed	SEW	Watershed to Field
Model4_savoy	Watershed	Watershed	Watershed to Watershed

Primary Data Layers Used

- Soils
- Landuse and landcover (LULC)
- Location of springs/wells
- Water quality

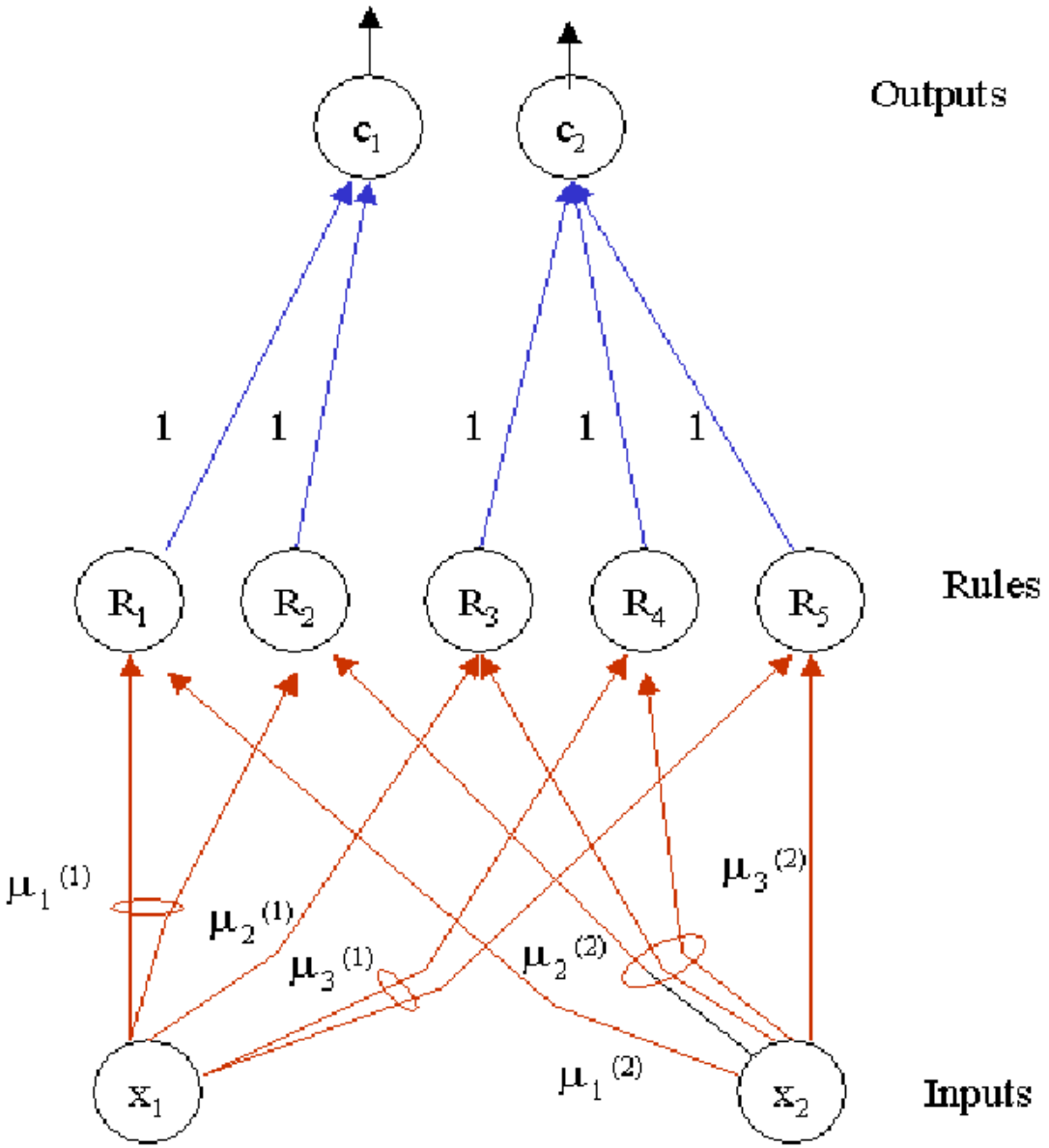
Secondary Data Layers Used

- Soil hydrologic group*
- Soil structure (pedality points)*
- Depth of the soil profile* (excluding Cr and R)
- Slopes
- Elevation

* model inputs

Description of the Primary Data Layers

Data	Scale/resolution	Comments
Soil	1:24,000	NRCS
Location of springs/wells	Field	GPS
Water quality data*	Field	ADEQ and AWRC
LULC	30 m/ 2ha	CAST


Outputs
c1
c2
1
1
1
1
1
Rules
R1
R2
R3
R4
R5
μ1 (1)
μ2 (1)
μ3 (1)
μ2 (2)
μ3 (2)
μ1 (2)
Inputs
x1
x2

Why Neuro-fuzzy?

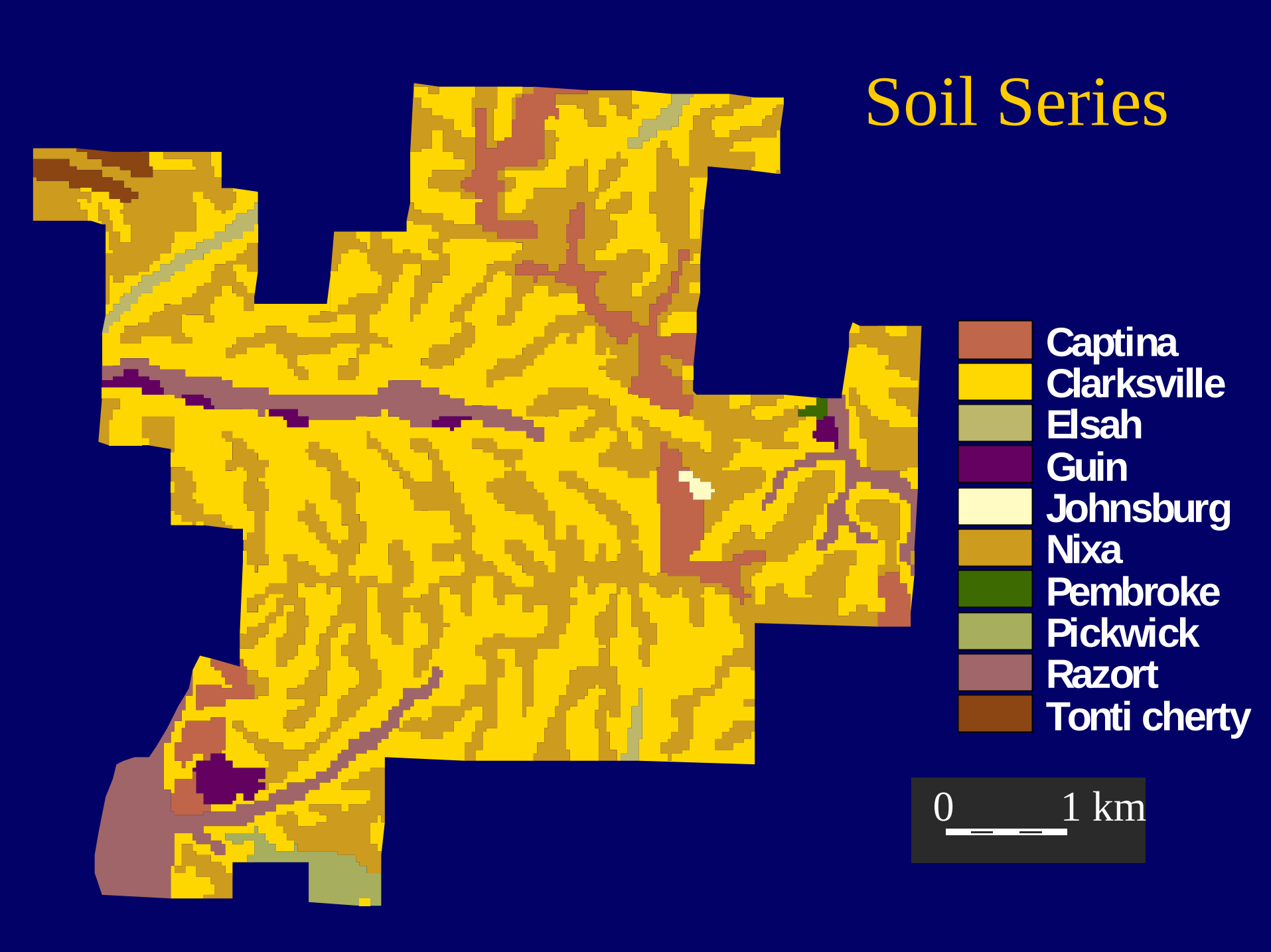
- Schultz and Wieland (1997) suggested that NN could parsimoniously represent non-linear systems and seem to be robust and flexible under data driven situations and allow deeper professional insight into the model.
- Fuzzy logic provides an opportunity to incorporate experts' opinion and robust under uncertainty.

Necessary steps

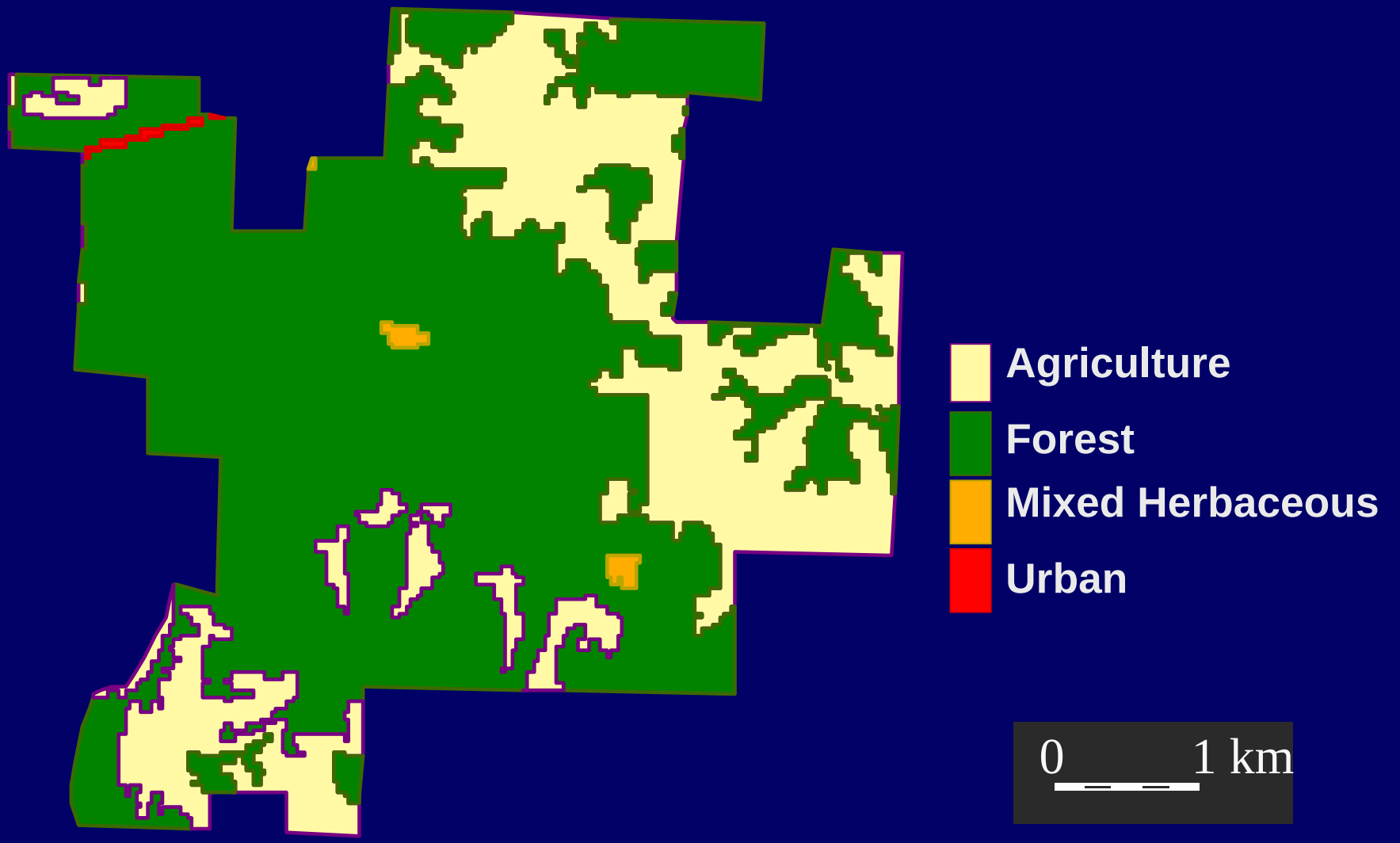
- Training data
- Testing data

Assessment of Models

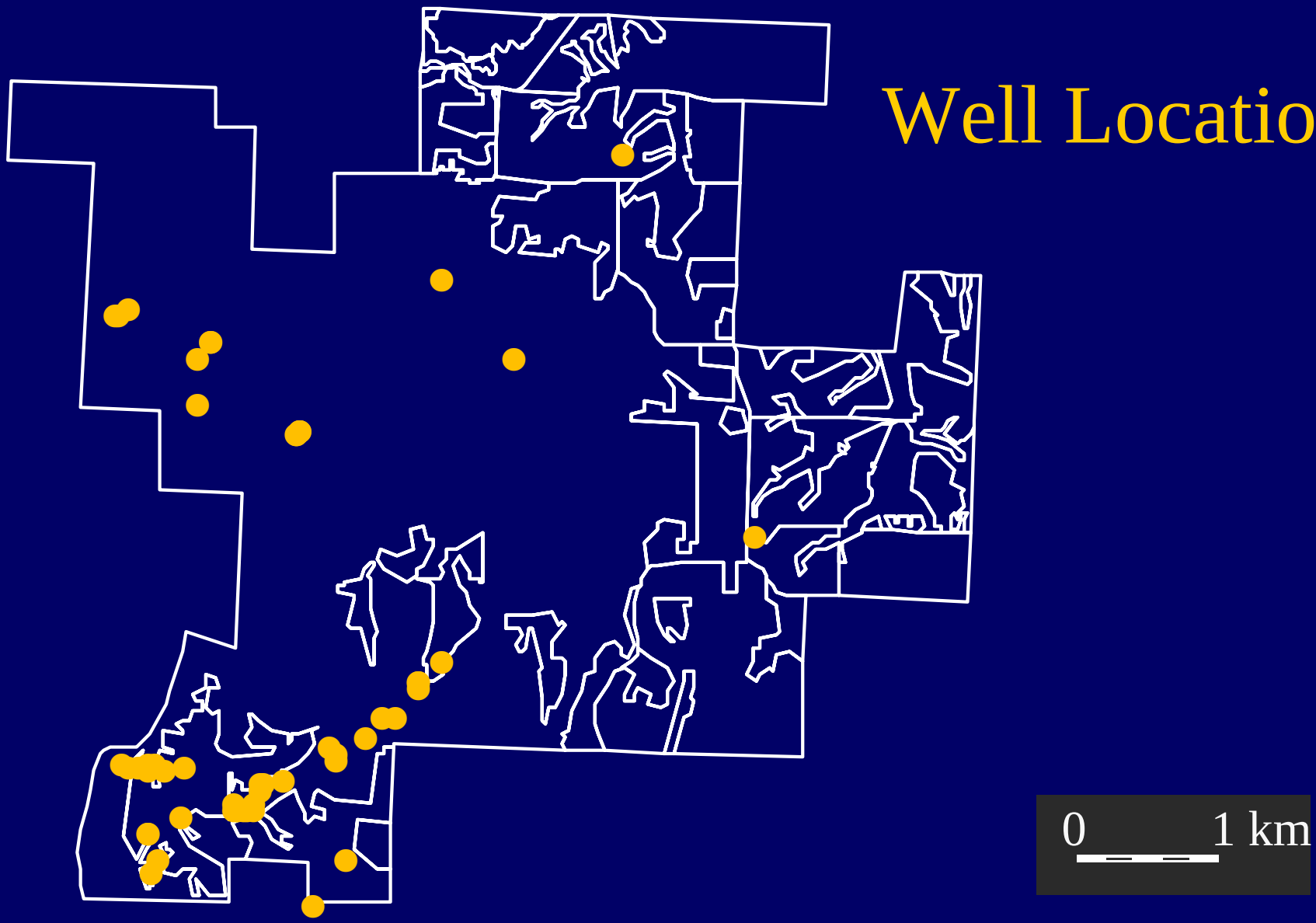
- Comparison of models and Field data
  - Coincidence analyses

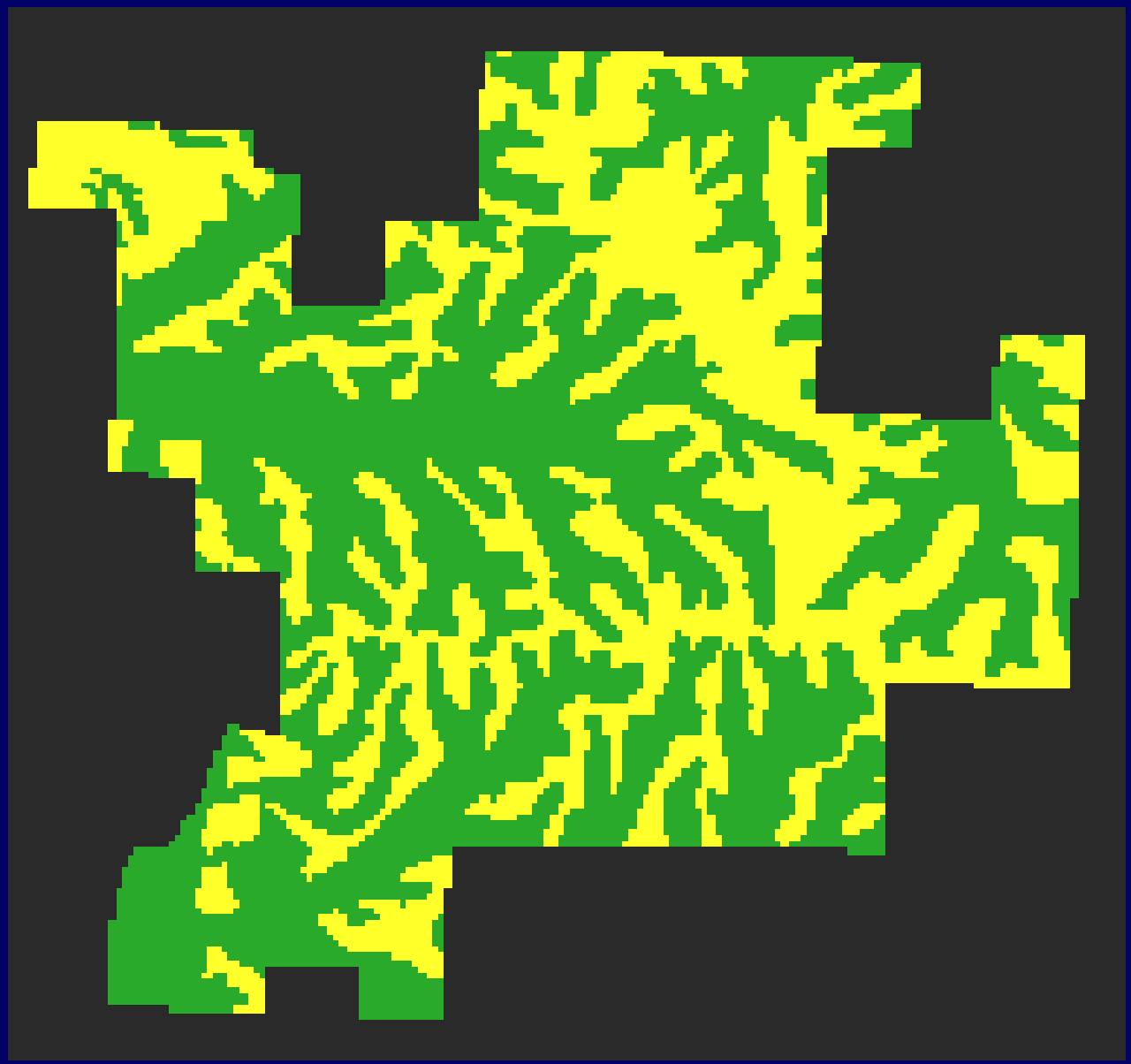

Soil Series
Captina
Clarksville
Elsah
Guin
Johnsburg
Nixa
Pembroke
Pickwick
Razort
Tonti cherty
0
1 km

Landuse



Well Locations





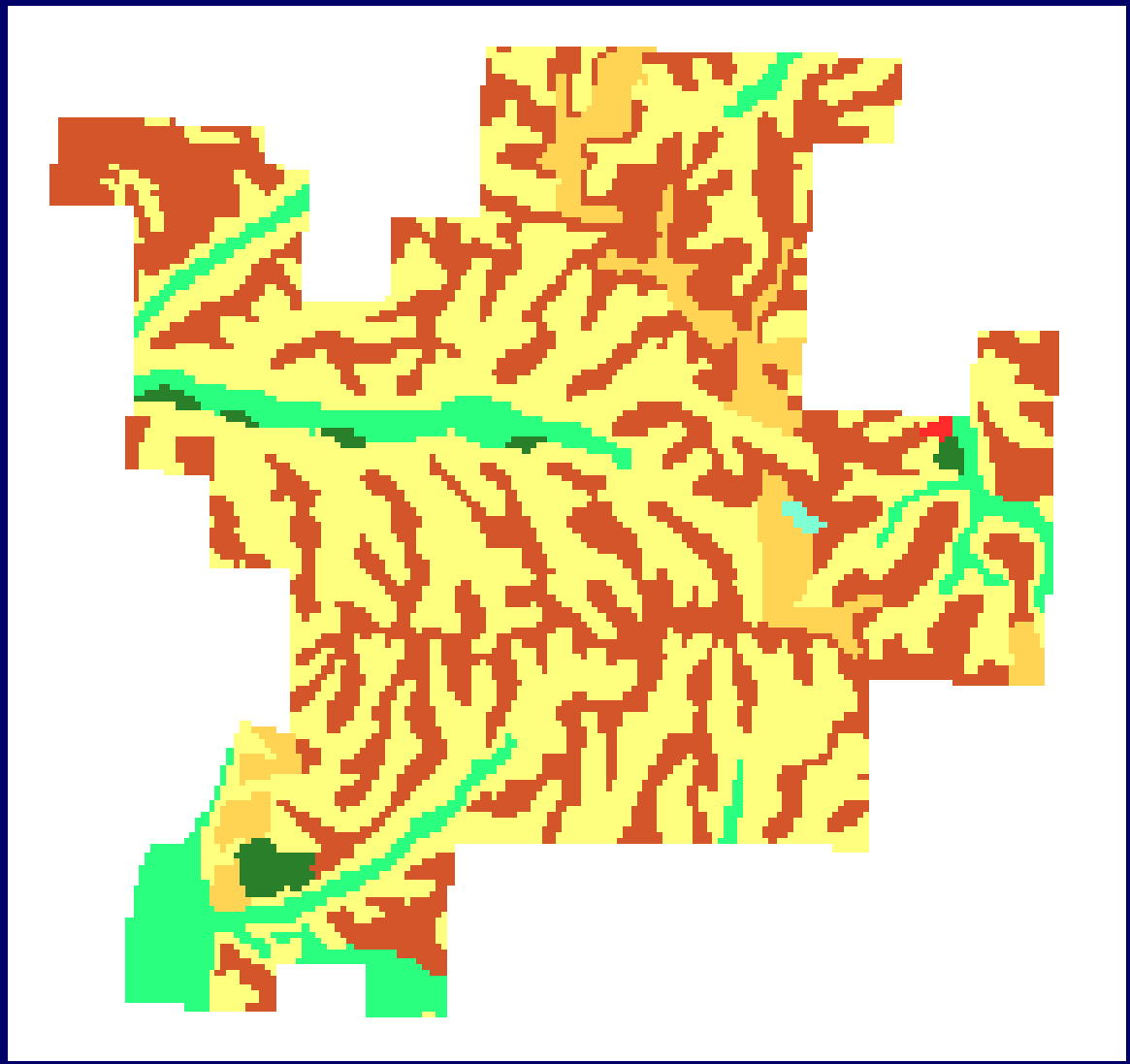
Hydrologic Units

C

B

0 1 km

Soil Depth

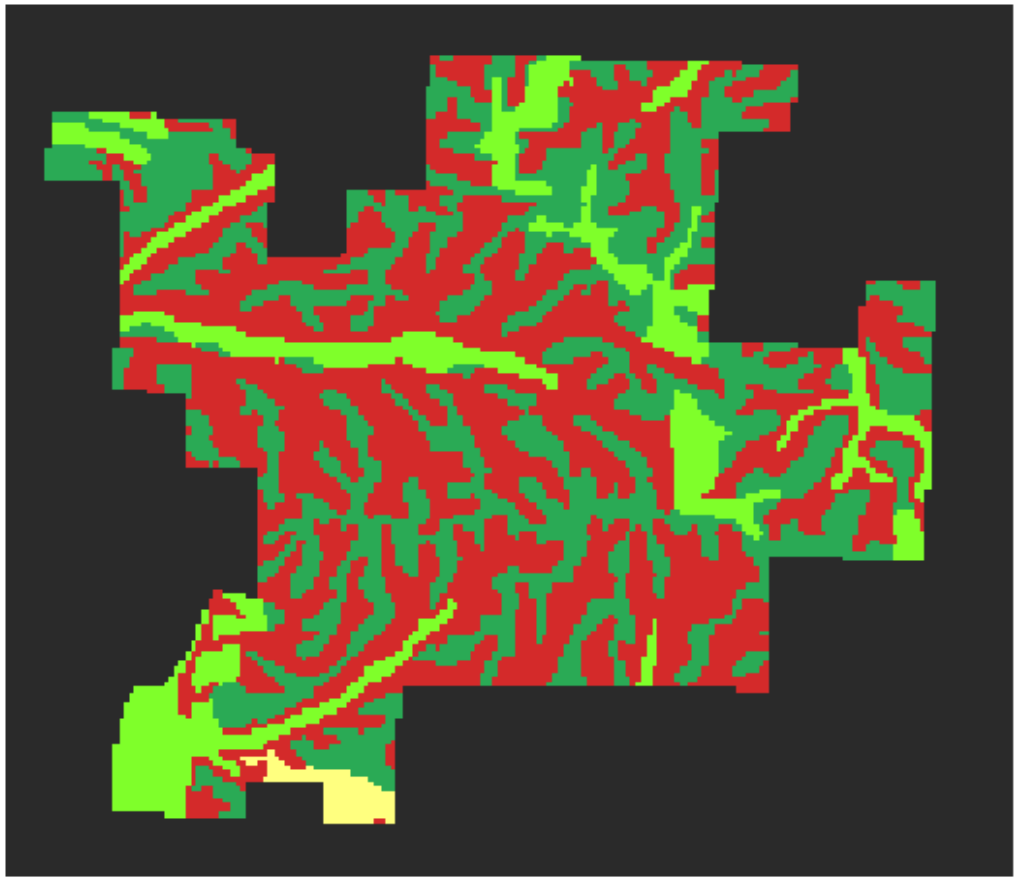


96
80
90
72
60
66
108

0 1 km

Depth (inches) : Shallow = 9 – 30, Moderately shallow = 31 – 50, Moderately deep = 51 – 69, Deep = 70 – 85 and Very Deep = > 85

Soil Structure



Low
Moderately high
High
Very high

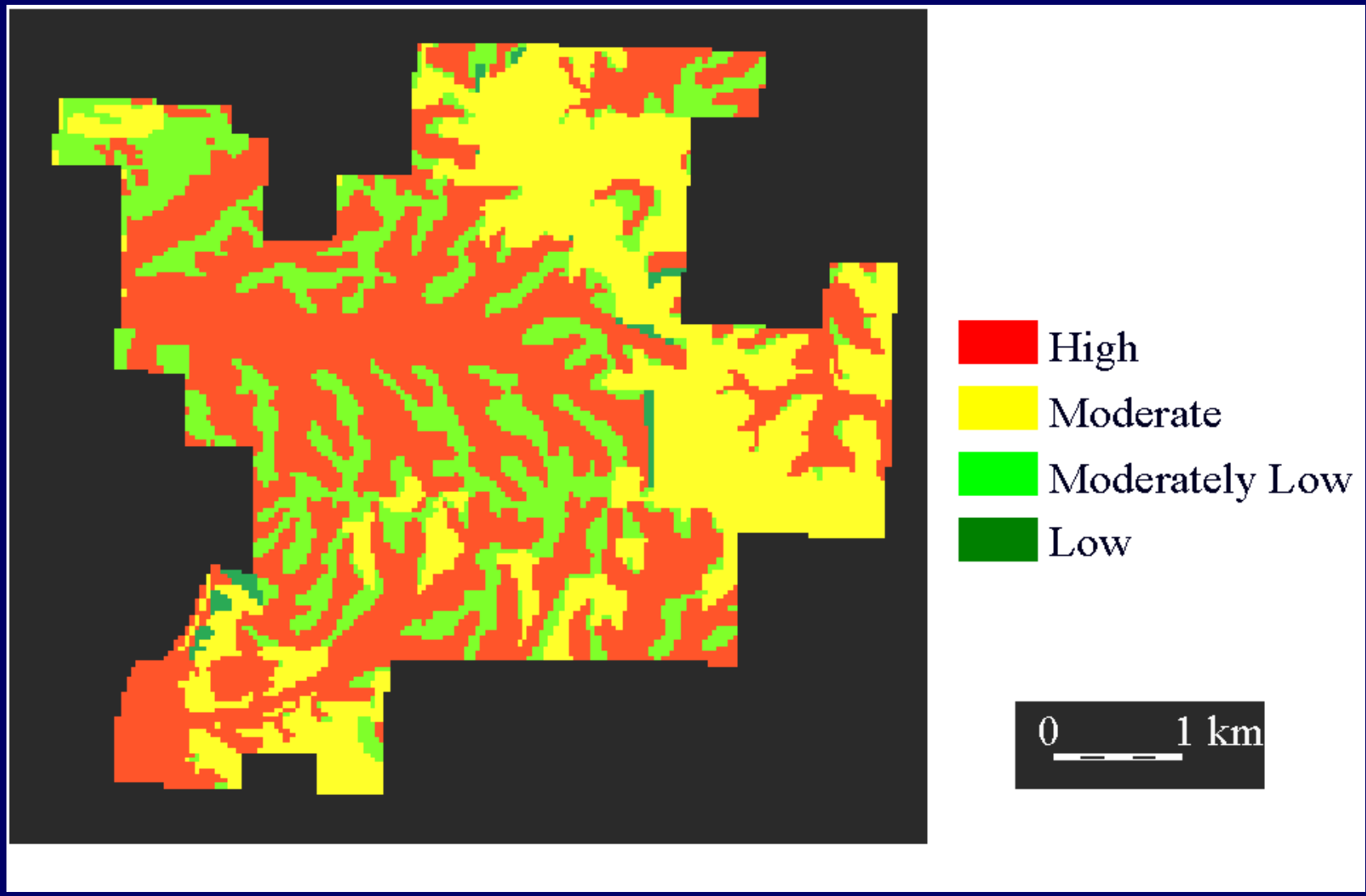
0 1 km

Low = 14 – 17, Moderate = 20 – 30, Moderately high= 31 – 40, High = 40 – 50 and very high> 51

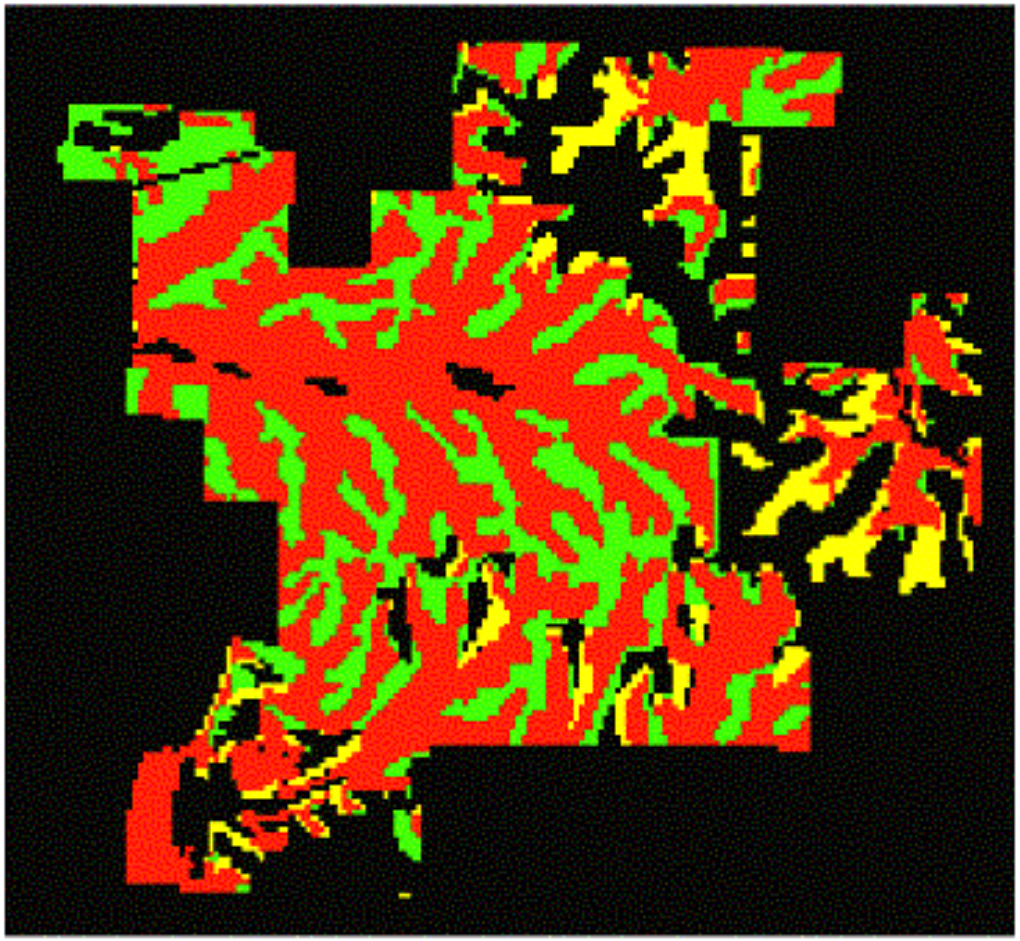
* points = ped grade + ped size + ped shape

Results

Vulnerability Results: Model1_Savoy



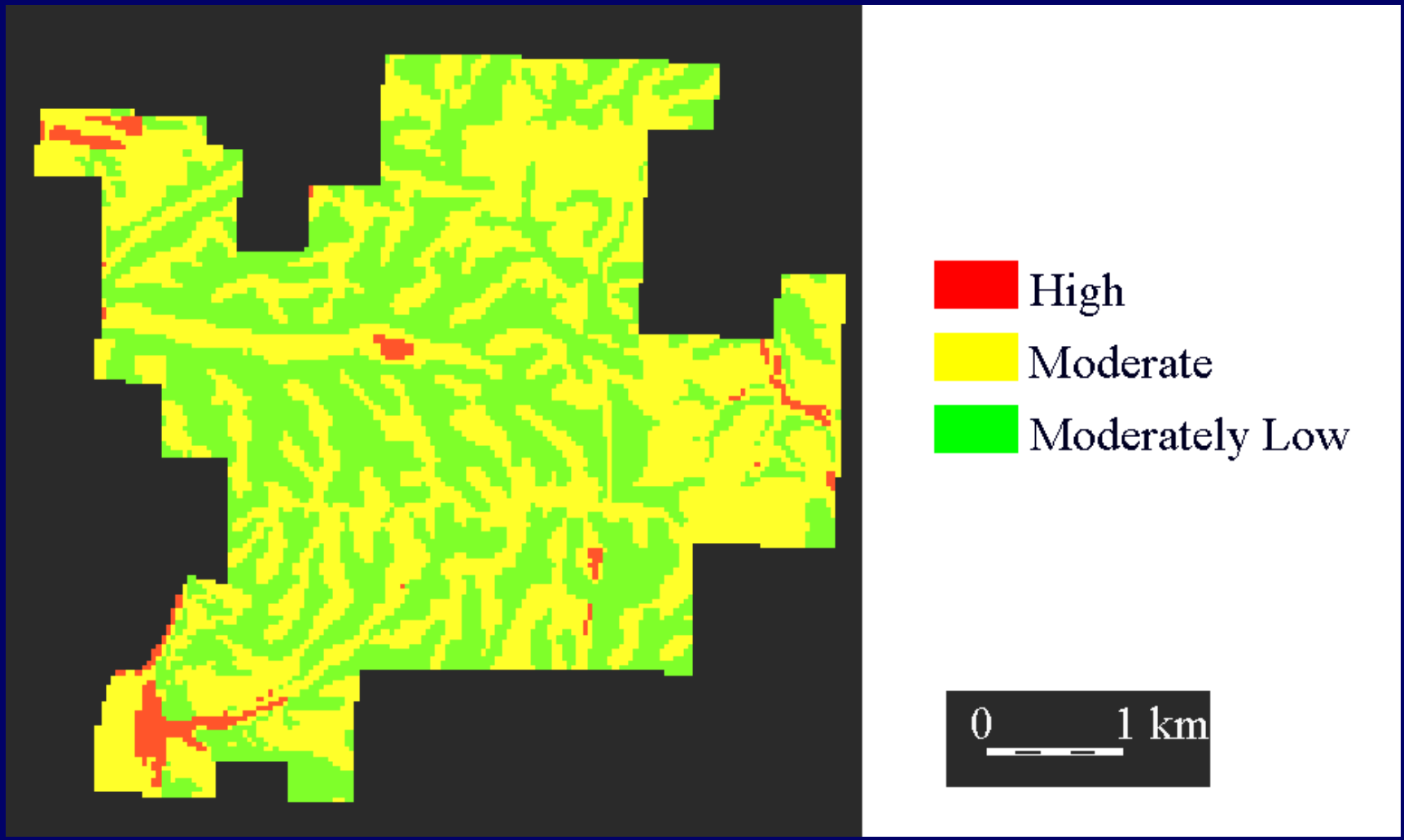
Vulnerability Results: Model2_Savoy



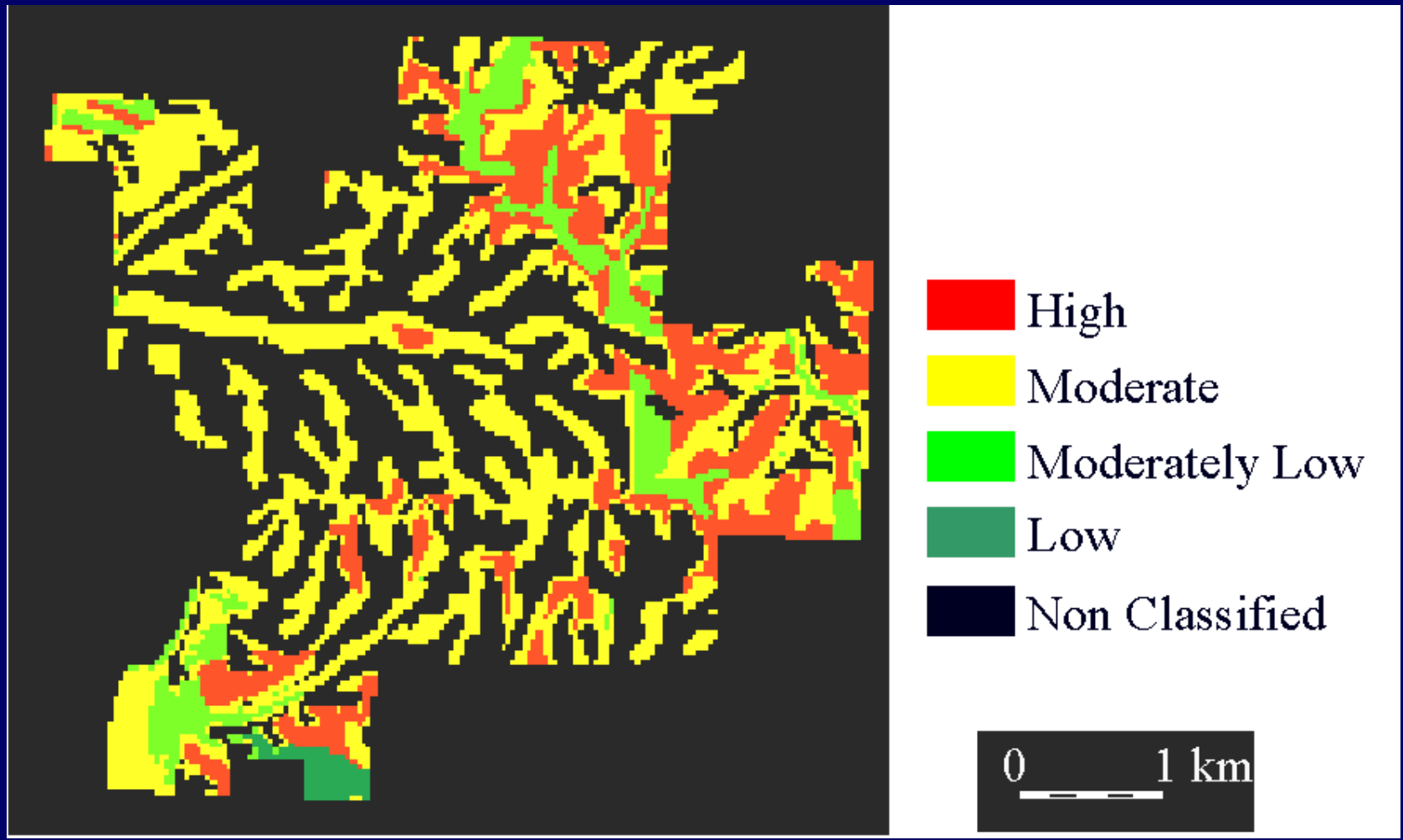
High
Moderate
Moderately Low
Non Classified

0 1 km

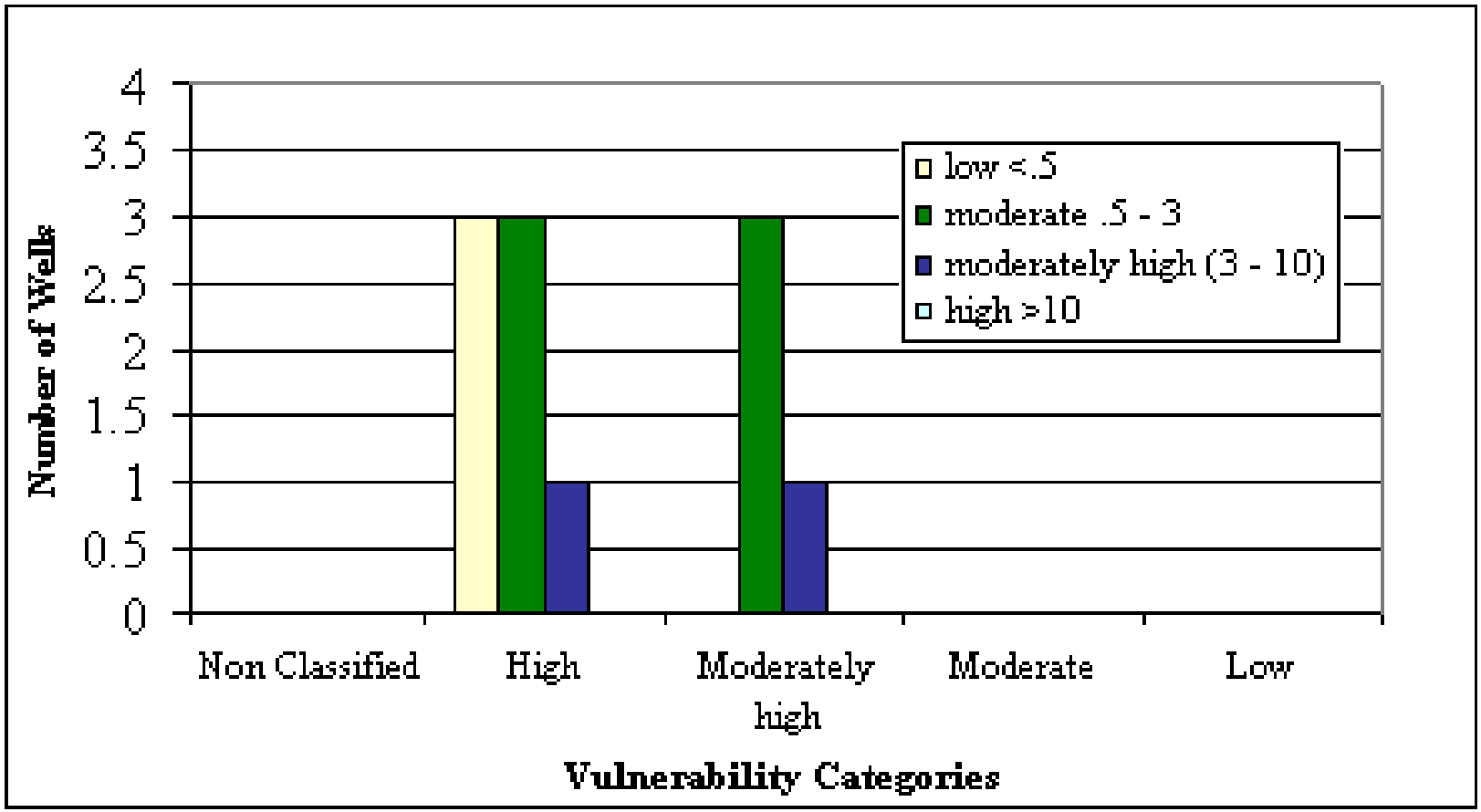
Vulnerability Results: Model3_Savoy



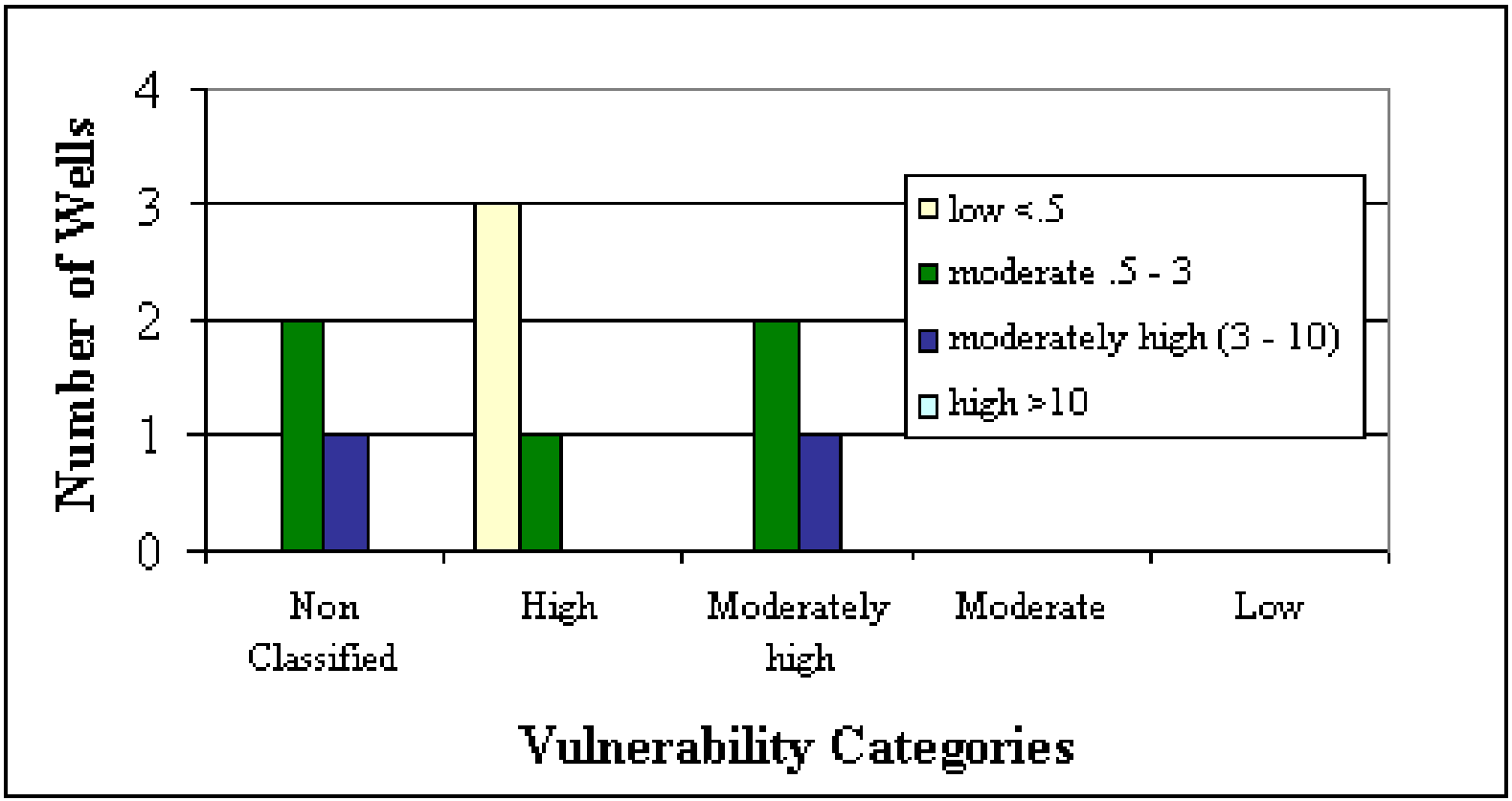
Vulnerability Results: Model4_Savoy



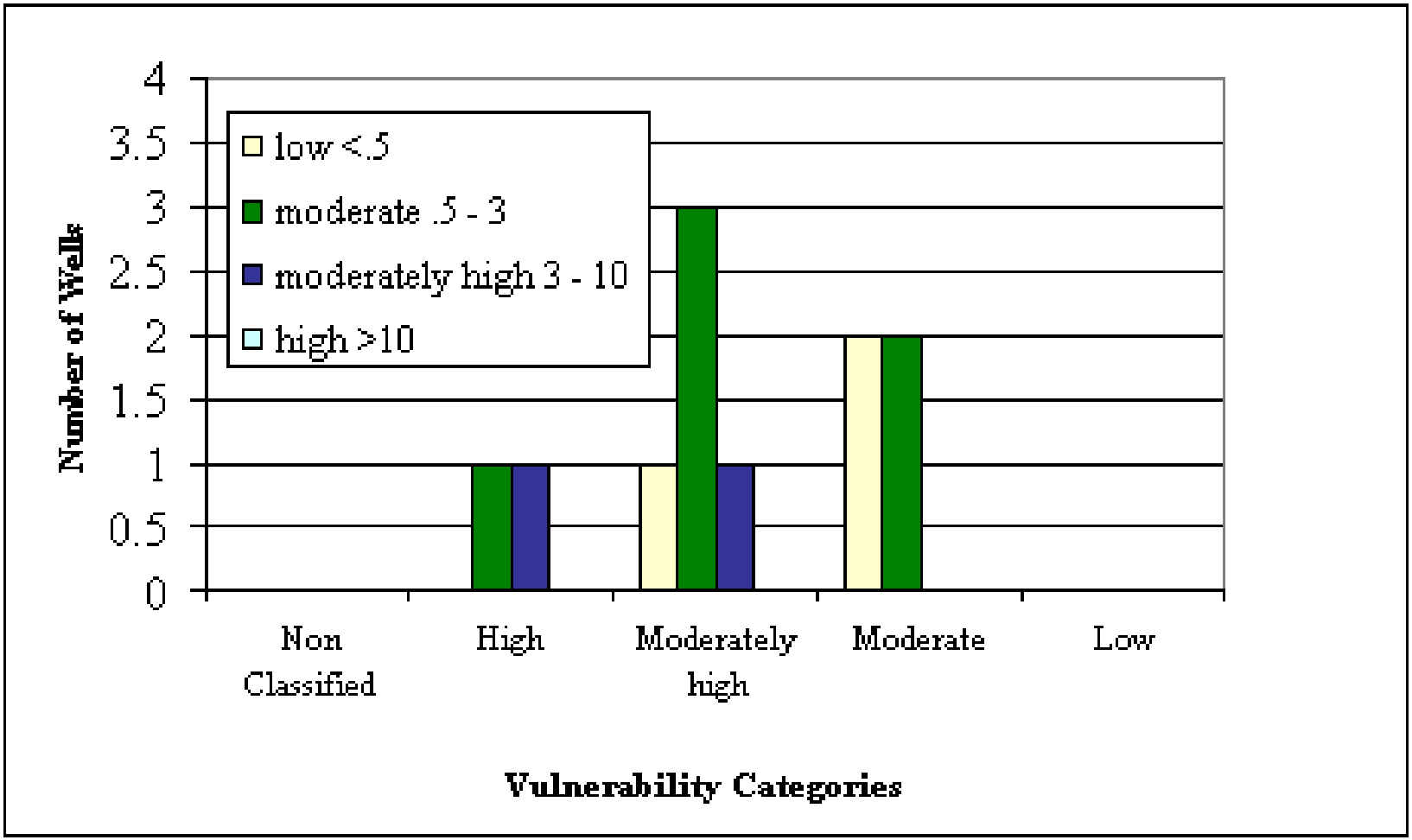
Coincidence Results: Model1_Savoy



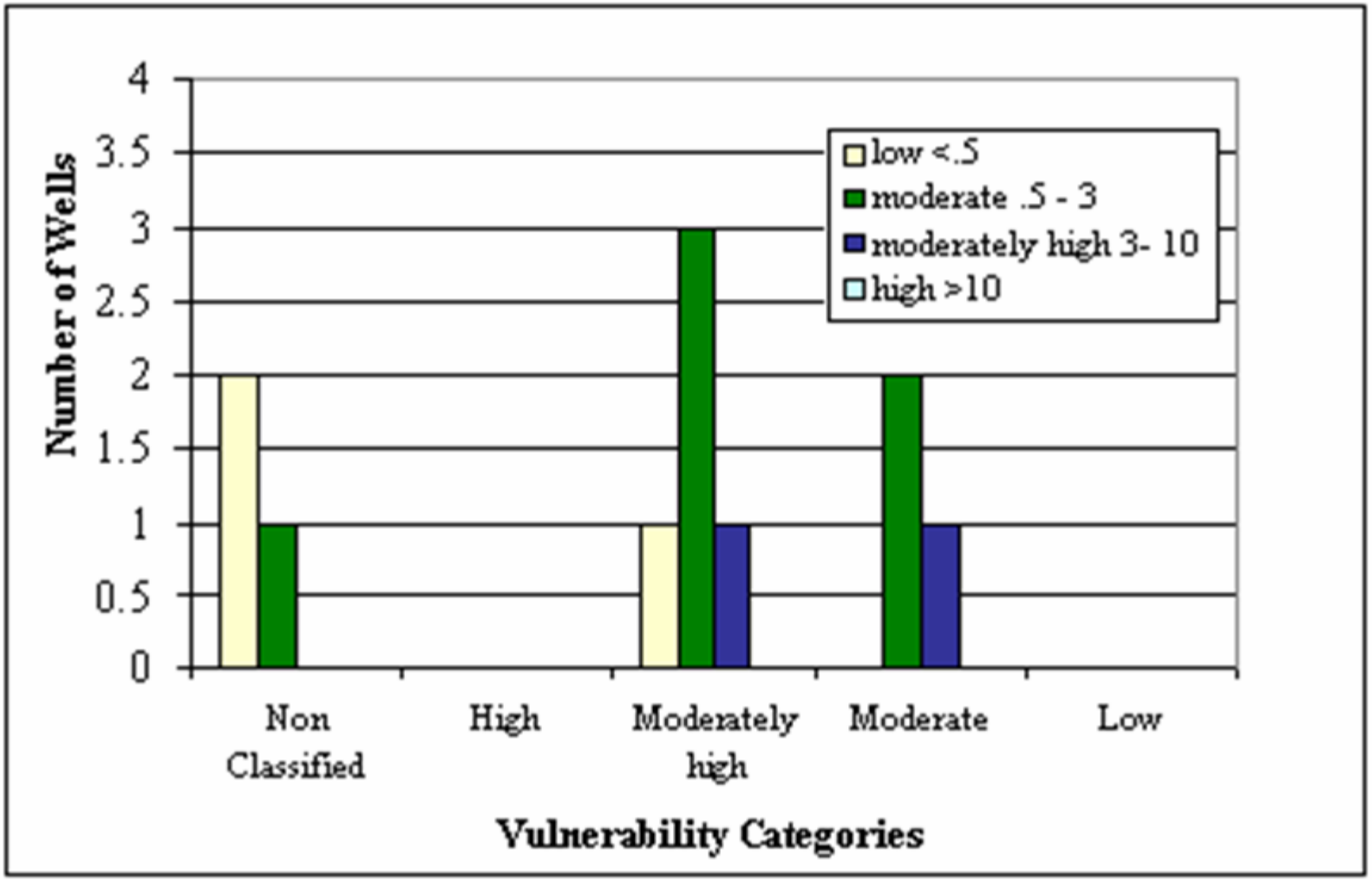
Coincidnece Results: Model2_Savoy



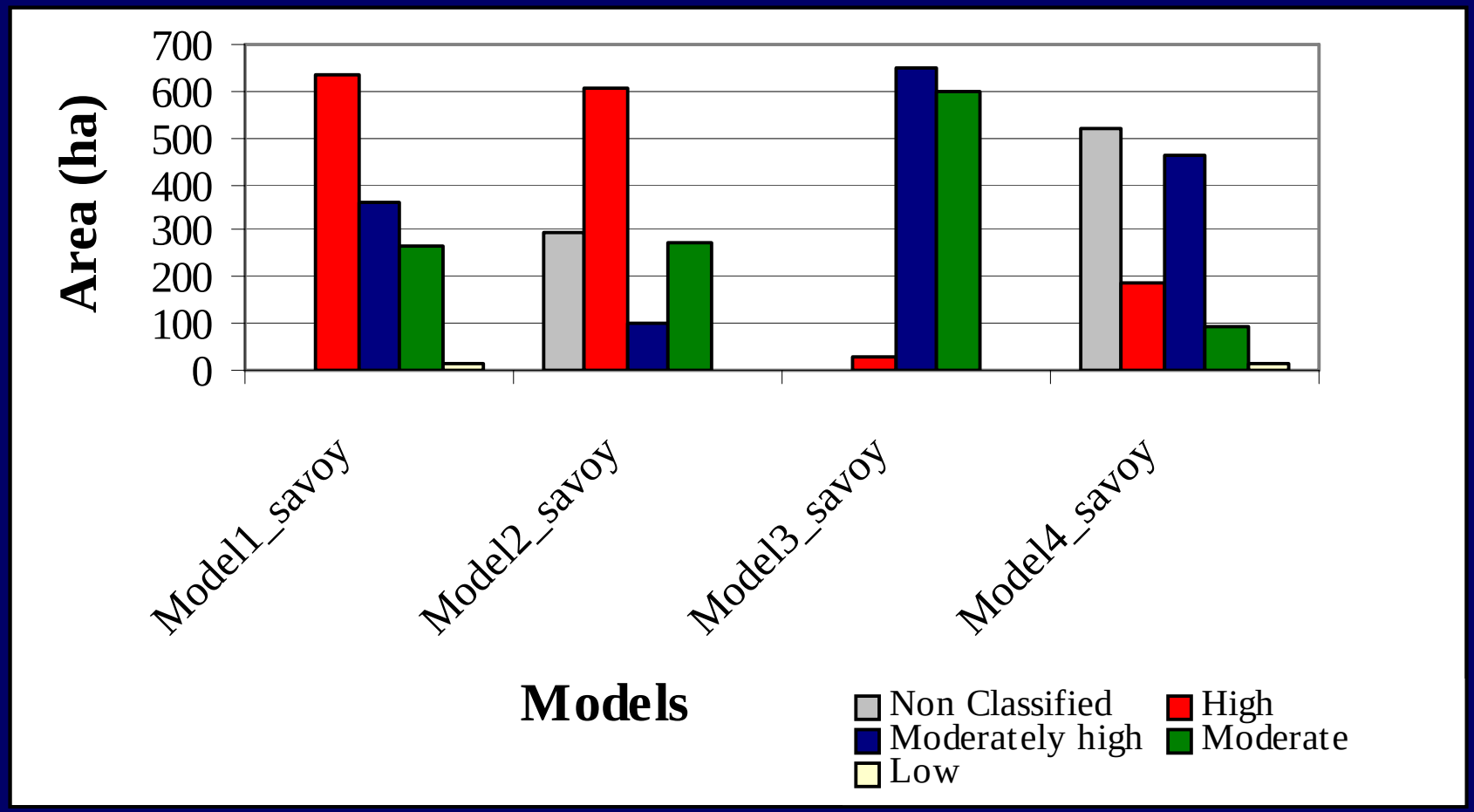
Coincidence Results: Model3_Savoy



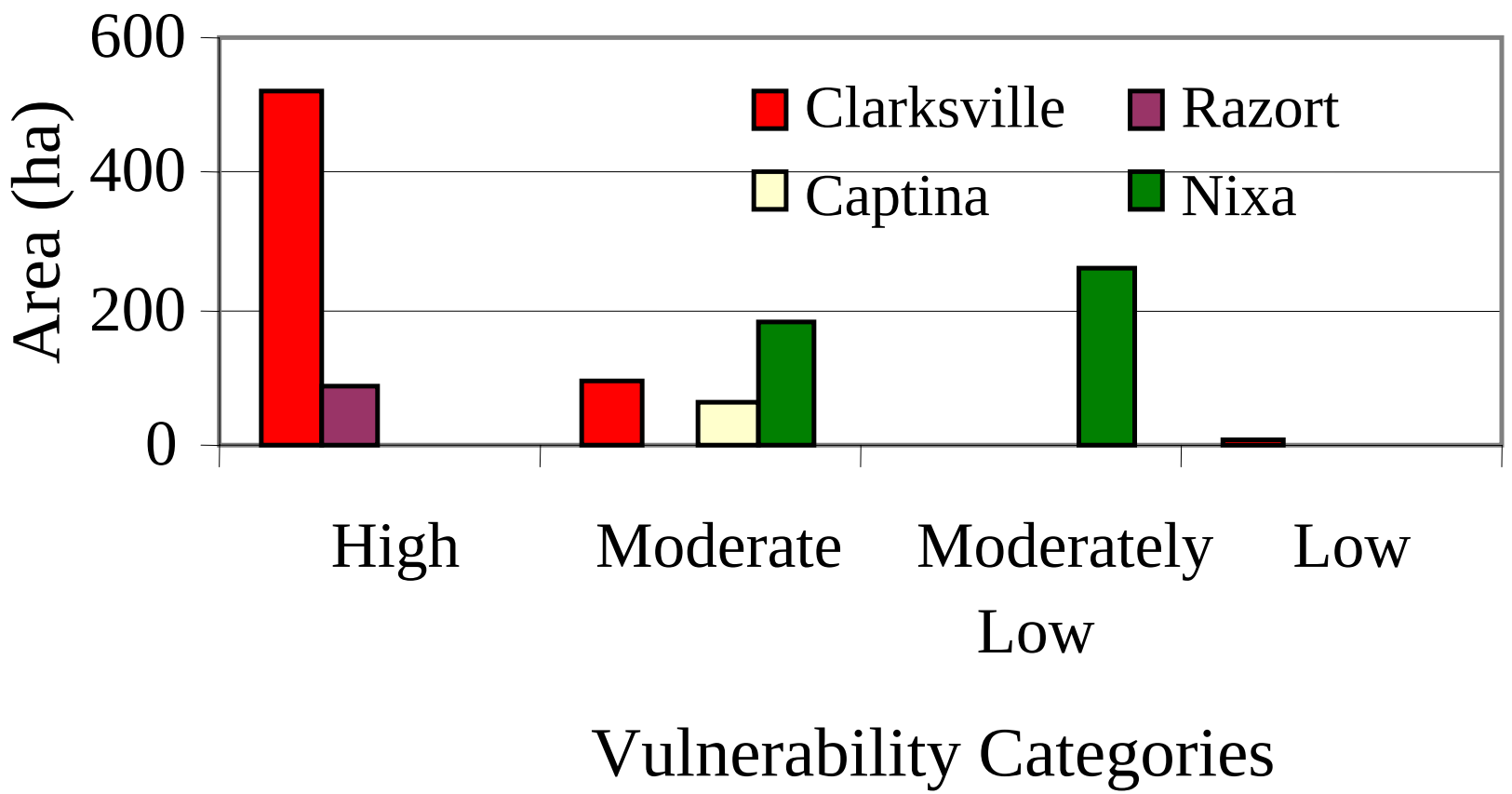
Coincidence Results: Model4_Savoy



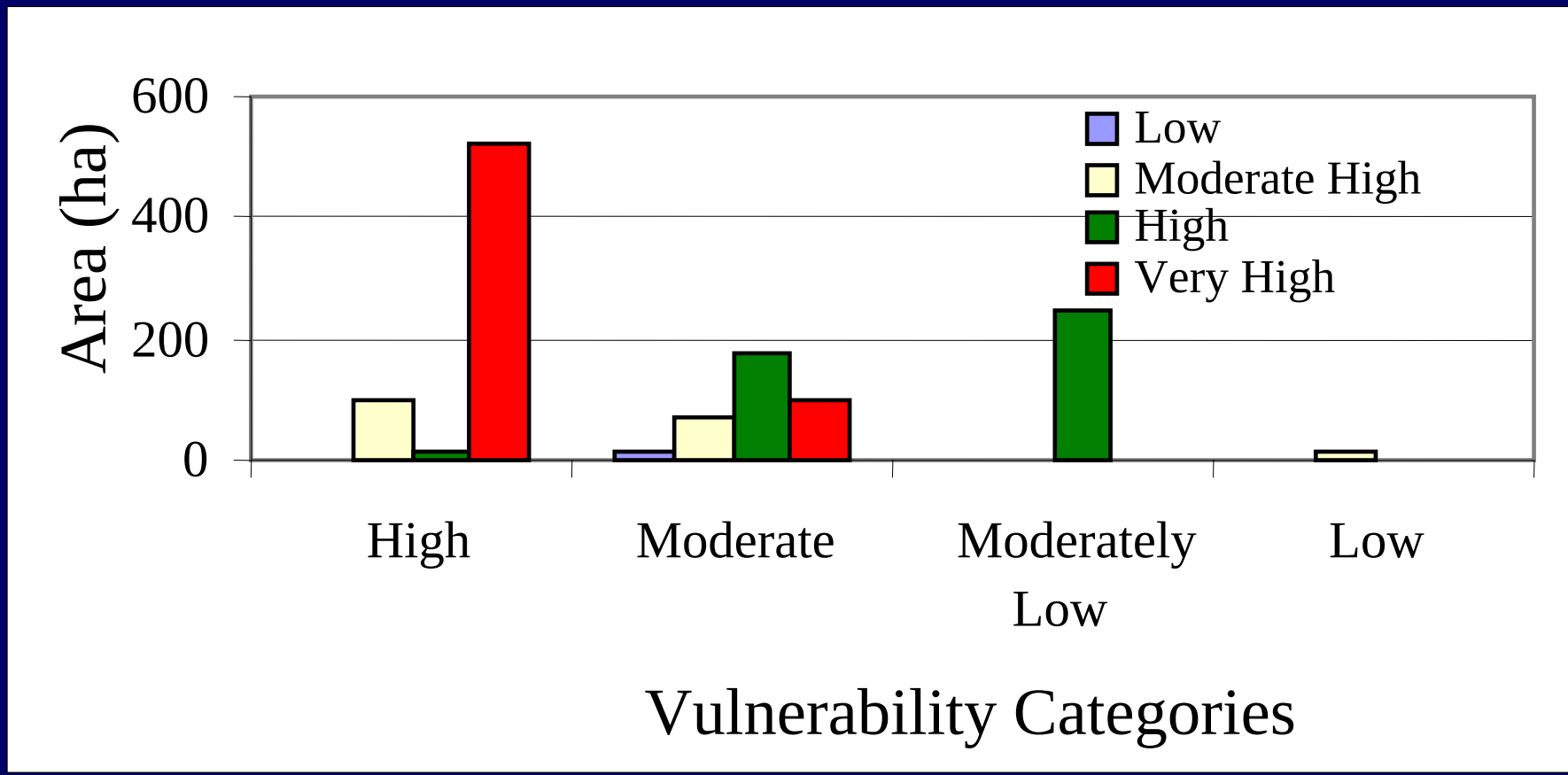
Areal Coverage of Vulnerability Categories



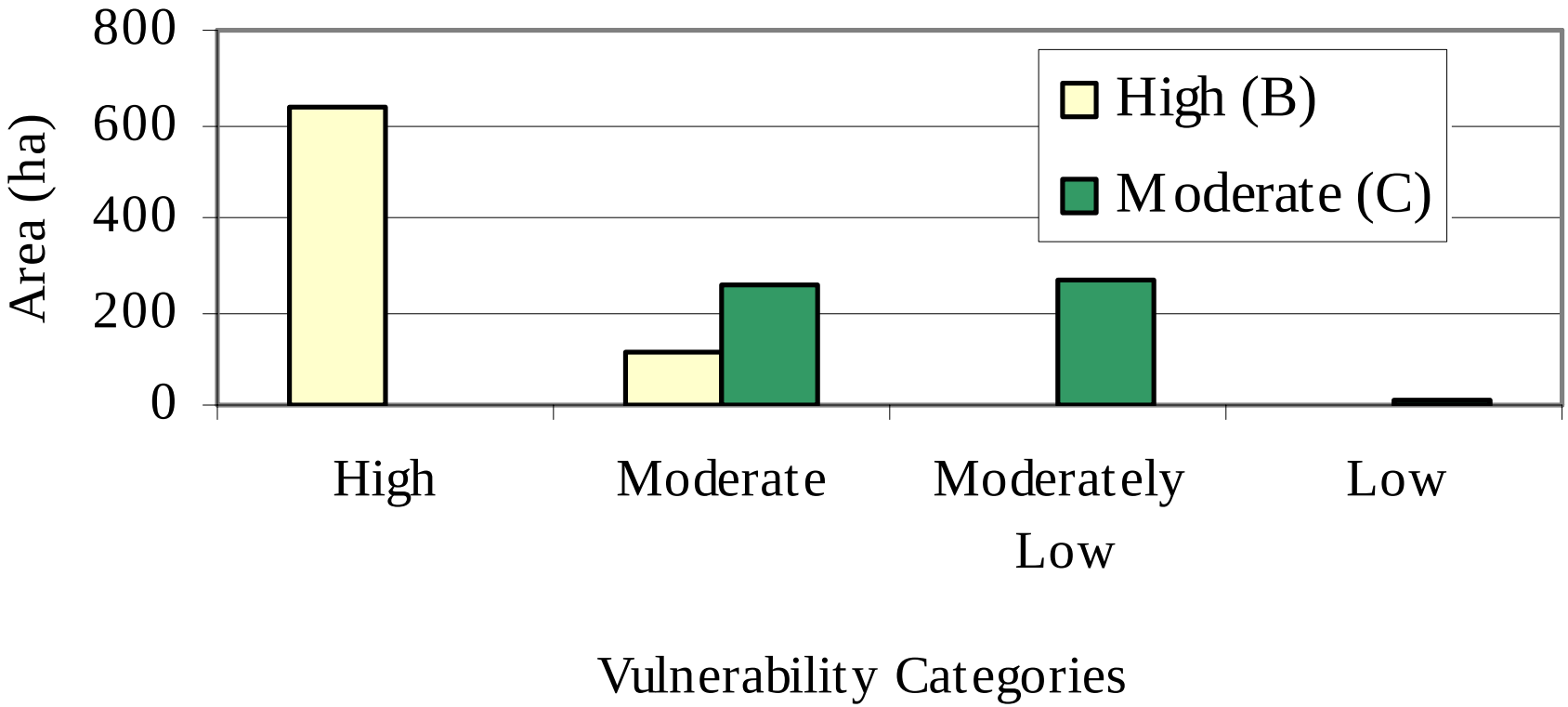
Soils vs. Vulnerability



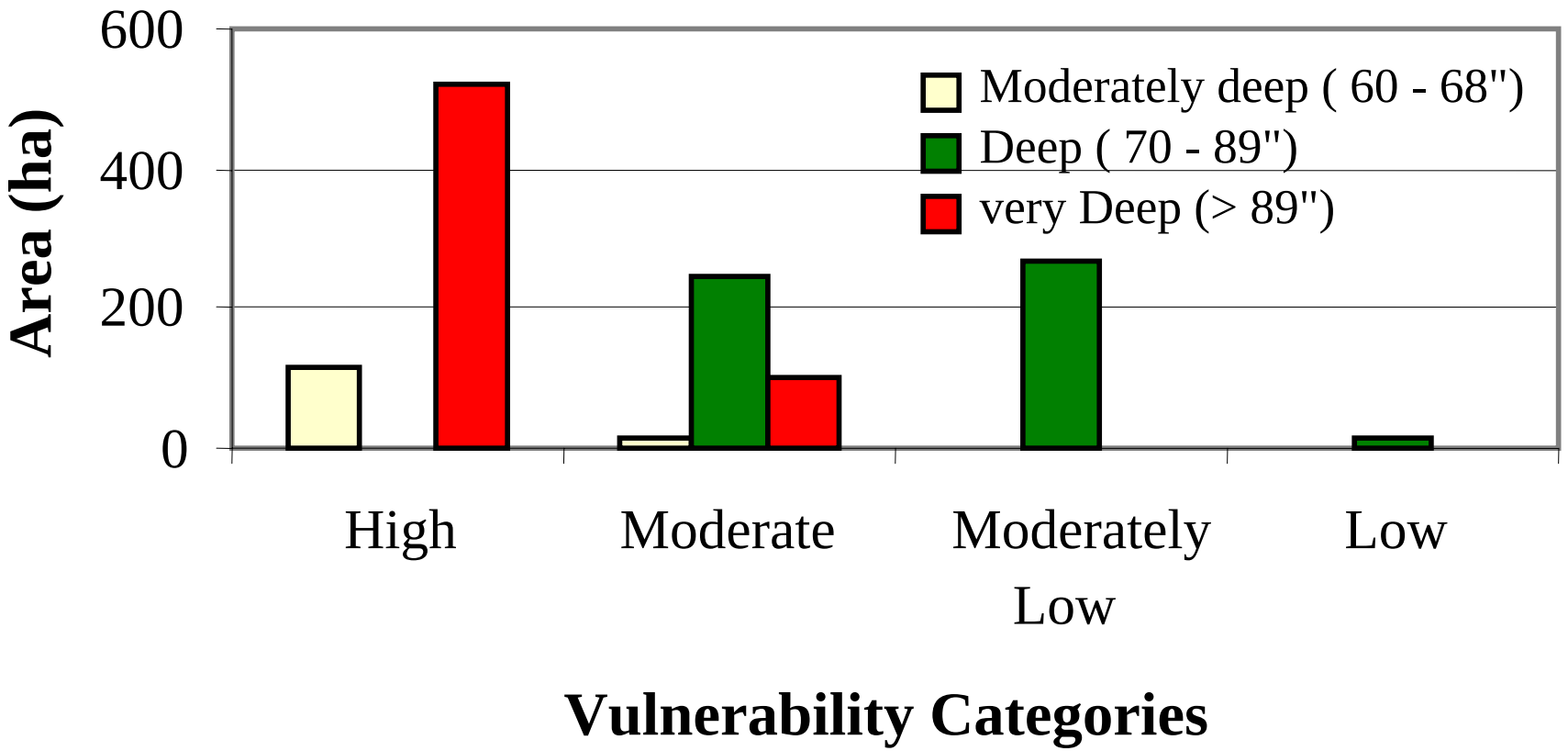
Soil Structure vs. Vulnerability



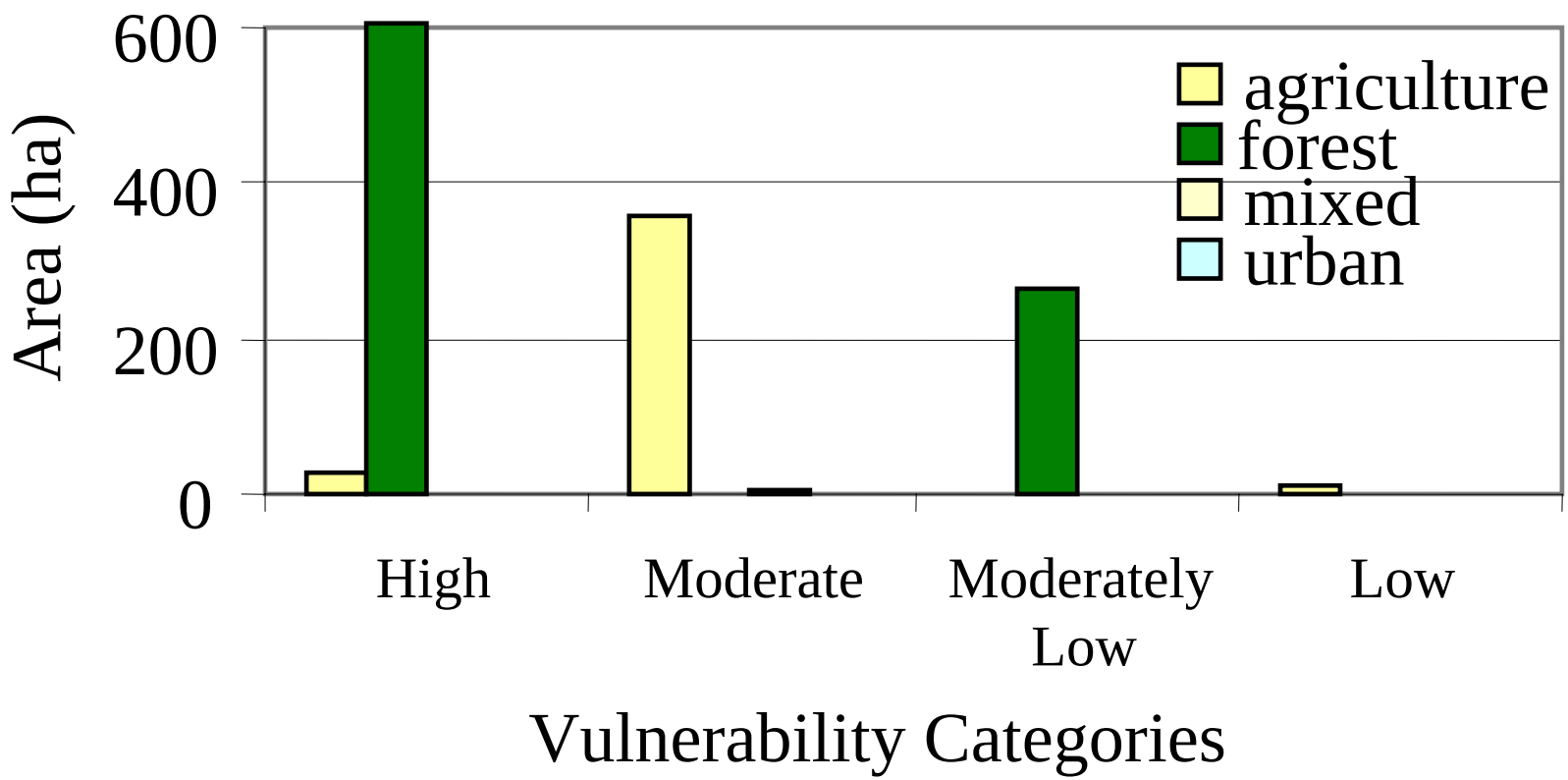
Hydrologic Group vs. Vulnerability



Depth vs. Vulnerability



LULC vs. Vulnerability



Summary

- When the watershed level training data are applied to field level application data (Model3_savoy), the entire data sets were classified by the net and no ‘non-classified’ category was found.
- This was due to the fact that the larger training data set (watershed) contained all possible combinations found in the smaller area (SEW).

Summary

- Transfer of SEW to the watershed scale models (model2_savoy) resulted in greater area in the non-classified category
- This indicated that the training data were not sufficient for the net to converge and apply the information acquired through the training processes to the unknown data set.

Summary

- Size of the training data and number of unique combinations represented in the training data set influenced the training and consequently, classification processes that classify the data to generate vulnerability maps with four vulnerability categories

Summary

- Training techniques used also influenced the prediction. Compared to Model1_savoy (SEW _ SEW), Model4_savoy (Watershed-watershed) showed more misclassification.
- This could be attributed to the difference in training strategies
- Size of the training data is important, so is training strategies.

Summary

- Neuro-fuzzy models are sensitive to the scale issues as they are related to the training data set
- The coincidence reports showed different association of input factors found in different models.
- Further study needed

Questions?