Resource materials for a GIS spatial analysis course

By Gary L. Raines[1]

Open-File Report 01-221
Version 1.0

2001

This report is preliminary and has not been reviewed for conformity with U.S. Geological Survey editorial standards or with the North American Stratigraphic Code. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Manuscript approved ###, 2001

U.S. DEPARTMENT OF THE INTERIOR
U.S. GEOLOGICAL SURVEY

1. U.S. Geological Survey, Reno, NV 89557

CHAPTERS

Introduction

This report consists of materials prepared for a GIS spatial analysis course offered as part of the Geography curriculum at the University of Nevada, Reno and the University of California at Santa Barbara in the spring of 2000. The report is intended to share information with instructors preparing spatial-modeling training and scientists with advanced GIS expertise. The students taking this class had completed each universities GIS curriculum and had a foundation in statistics as part of a science major. This report is organized into chapters that contain the following:

- Slides used during lectures,
- Guidance on the use of Arcview,
- Introduction to filtering in Arcview,
- Conventional and spatial correlation in Arcview,
- Tools for fuzzification in Arcview,
- Data and instructions for creating using ArcSDM for simple weights-of-evidence, fuzzy logic, and neural network models for Carlin-type gold deposits in central Nevada,
- Reading list on spatial modeling, and
- Selected student spatial-modeling posters from the laboratory exercises.

Avenue scripts and date useful for the exercises are included in zip files for each chapter. The textbook used for this class is "Geographic Information Systems for Geoscientists –Modeling" (See Syllabus for complete reference). The lecture slides were designed to enhance the information in the textbooks; so, the slides are not for self-guided education. The remaining chapters were prepared to assist in the self-guided laboratory exercises. Because the students were expected to be GIS experts, the laboratory problems were designed to state a problem for which the students were expected to find and implement a solution. Thus, chapters following the lecture slides give specific guidance and Arcview scripts or extensions for spatial-modeling tasks of filtering, correlation, fuzzification, and spatial modeling using ArcSDM.

The last three chapters of Graeme Bonham-Carter's Geographic Information Systems for Geoscientists –Modeling in GIS were used as the textbook, specifically: Chapter 7 (Tools for Map Analysis: Single Maps), Chapter 8 (Tools for Map Analysis: Map Pairs), and Chapter 9 (Tools for Map Analysis: Multiple Maps). The emphasis was on Chapter 9 and the weights-of-evidence method of spatial analysis. Data for the laboratory exercises were provided by the Tahoe Regional Planning Authority (TRPA) and included goshawk, osprey, and spotted owl nesting sites that were used as training sites for predictive modeling of animal habitat. The URL for the TRPA is http://ceres.ca.gov/trpa/. Software used for class exercises included ArcView 3 GIS with the Spatial Analyst extension, and a Weights-of Evidence extension ArcWofE available at http://gis.nrcan.gc.ca/software/arcview/wofe. The ArcWofE extension has been replaced by an enhanced extension ArcSDM, available at http://ntserv.gis.nrcan.gc.ca/sdm/. In addition to the weights-of-evidence method, ArcSDM has logistic regression, fuzzy logic, and two neural network tools.

This report is an evolving document. Since the class was presented, changes have been made to improve some chapters and to document application issues. Some of the chapters, such as the filtering chapter, are simply notes that are incomplete but complement aspects of the lectures. These chapters may be expanded in future versions to document the subject more completely.

Spring 2000 - Geography Dept., UCSB
Professor: Gary Raines
Lab Instructors: David Jones and Ethan Inlander

Geography 176C
Spatial Analysis in GIS

Course Description

The goal of this class is to introduce the concepts of modeling in which multiple spatial data sets are combined to predict the distribution or occurrence of some complex process. Examples of the types of applications addressed might be predictive models of animal habitat, occurrence of infectious diseases, or undiscovered mineral resources. These types of models all have the characteristic that the processes involved are complex and sometimes poorly understood, that is the models are not prescriptive, but are often fuzzy or probabilistic in nature.

We will use Arcview 3.x, the Spatial Analysis, and modeling extensions (WofE and SDM). This will require the student to be familiar with Arcview 3.x and Spatial Analysis. Students will create simple to complex models using this software to gain experience in the process of modeling complex natural science processes. Exercises will work toward the types of multi-disciplinary problems that are common in land management or natural resources organizations. Self directed exercises using available data will be utilized.

The class will be a combination of lectures and student-lead discussions. Besides the textbook, journal articles will be read and discussed in student-lead discussions. Also, results of exercises will be presented to the class by students.

Schedule

Lecture: Thursday evening (6-9 pm)
Laboratory: 3 hours per week

Office Hours

Friday morning or by appointment, (Phone: Reno – 775-784-5596, email: graines@usgs.unr.edu). I maintain an open door policy. When I am in, the door is open. You are welcome to come in anytime if you need help.

Assignments

Spatial Analyst exercise – general use of spatial analysis
Assignment 1 – Reclassify a map to create a derivative map
Assignment 2 – Prepare a prescriptive model such as erosion potential
Assignment 3 - Prepare a probabilistic model such as animal habitat or mineral potential

Textbook

Bonham-Carter, G.F., 1996, Geographic information systems for geoscientists – modeling in GIS: New York, Elsevier Science Inc, 398 p. ISBN 0 08 042420 1

Course Grade

Assignment 1	10%
Assignment 2	20%
Assignment 3	45%
Class participation	10%
Examinations	15%

The grade for the assignments will be based on the originality of the model prepared, the cartographic quality of the presentation, and the oral presentation.

Schedule of Lectures and Assignments

Topic	**Assignment**
Course Introduction	Read King and Kramer, Chap. 7, Assignment 1
Overview of Spatial Analysis and Background	Spatial analyst exercise
Tools for Map Analysis: Single Maps	
Reclassification and Spatial Modeling	
Assignment Presentations	Assignment 1 due
Tools for Map Analysis: Map Pairs	Chap. 8, Assignment 2
Overlay and map correlation	
Map correlation	
Assignment Presentations	Assignment 2 due
Tools for Map Analysis: Multiple Maps	Chap. 9, Assignment 3
Boolean	
Index overlay	
Fuzzy logic	
Bayesian methods	
Logistic Regression methods	
Fuzzy logic	
Neural networks	
Assignment Presentations	

Approximate Schedule.

Date	USCB Lecture	Subject	Reading	Labs
Apr 6-7	1	Intro/Overview	King and Kramer	#1
Apr 13-14	2	Intro-Reclass/Filtering/Open	Chap 7	
Apr 20-21	3	Presentations/Open/Overlay	Chap 8	#1 Due, #2
Apr 27-28	4	Correlation1/Correlation2/Open		
May 4-5	5	Presentations/Open		#2 Due
May 11-12	6	Tahoe Problem/Multimap Overview	Chap 9	#3
May 18-19	7	Boolean-Index/Fuzzy1	Nova Scotia	
May 25-26	8	Arc WofE/WofE1/WofE2		
Jun 1-2	9	Nature Evidence/Logistic/Fuzzy2		
Jun 8-9	10	Open/Neural Networks/SMD Dem		
June 15 -16	11	Presentations		#3 Due

Spring 2000 – Geography Dept., UNR
Professor: Gary Raines
Lab Instructor: Wade Epperson

Geography 491/691 Section 2
Spatial Analysis in GIS

Course Description

The goal of this class is to introduce the concepts of modeling in which multiple spatial data sets are combined to predict the distribution or occurrence of some complex process. Examples of the types of applications addressed might be predictive models of animal habitat, occurrence of infectious diseases, or undiscovered mineral resources. These types of models all have the characteristic that the processes involved are complex and sometimes poorly understood, that is the models are not prescriptive, but are often fuzzy or probabilistic in nature.

We will use Arcview 3.x, the Spatial Analysis, and modeling extensions (WofE and SDM). This will require the student to be familiar with Arcview 3.x and Spatial Analysis. Students will create simple to complex models using this software to gain experience in the process of modeling complex natural science processes. Exercises will work toward the types of multi-disciplinary problems that are common in land management or natural resources organizations. Self directed exercises using available data will be utilized.

The class will be a combination of lectures and student-lead discussions. Besides the textbook, journal articles will be read and discussed in student-lead discussions. Also, results of exercises will be presented to the class by students.

Schedule

Lecture: Mon. & Wed. 1p
Laboratory: 222 Mackay Science Hall. Three hours per week minimum. Schedule is on students time during each week.

Office Hours

Mon. and Wed. or by appointment, 271 Laxalt Mineral Research (Phone: 784-5596, email: graines@usgs.unr.edu). I maintain an open door policy. When I am in, the door is open. You are welcome to come in anytime if you need help.

Assignments

Spatial Analyst exercise – general use of spatial analysis
Assignment 1 – Reclassify a map to create a derivative map
Assignment 2 – Prepare a prescriptive model such as erosion potential
Assignment 3 - Prepare a probabilistic model such as animal habitat or mineral potential

Textbook

Bonham-Carter, G.F., 1996, Geographic information systems for geoscientists – modeling in GIS: New York, Elsevier Science Inc, 398 p. ISBN 0 08 042420 1

Course Grade	491	691
Assignment 1	20%	10%
Assignment 2	20%	10%
Assignment 3	40%	50%
Class participation	10%	10%
Examinations	10%	10%
Discussions		10%

The grade for the assignments will be based on the originality of themodel prepared, the cartographic quality of the presentation, and the oral presentation.

Schedule of Lectures and Assignments

Week	Topic	Assignment
1	Course Introduction	Read King and Kramer
2	Overview of Spatial Analysis and Background	Spatial analyst exercise
3	Tools for Map Analysis: Single Maps Reclassification and Spatial Modeling	Chapter 7, Assignment 1
4	Assignment Presentations	Assignment 1 due
5	Tools for Map Analysis: Map Pairs Overlay and map correlation	Chapter 8, Assignment 2
6	Map correlation	
7	Assignment Presentations	Assignment 2 due
8	Tools for Map Analysis: Multiple Maps	Chapter 9, Assignment 3
9	Boolean and index overlay	
10	Spring Break	
11	Bayesian methods	
12	Application of Bayesian and Logistic Regression methods	
13	Fuzzy logic	
14	Neural networks	
15	Assignment Presentations	
16	Summary (Mon. only)	Assignment 3 due

Lectures

Spatial Analysis in GIS
Dr. Gary Raines

- Geography 491/691

USGS science for a changing world

Version 1, January 2000

Gary Raines

- USGS Research Geologist
- Remote Sensing applications to mineral exploration
- Development of techniques for spatial modeling in mineral and environmental applications
- Focus on large areas

Version 1, January 2000

Class Syllabus

- Lecture schedule
- Three Laboratory Assignments
- Examinations
- Reading
  - Geographic Information systems for geoscientists - modeling in GIS: Chapters 7, 8, and 9
  - Additional reading - student lead discussion

Version 1, January 2000

Grading

Task	491	691
Assignment 1	20%	10%
Assignment 2	20%	10%
Assignment 3	40%	50%
Participation	10%	10%
Examinations	10%	10%
Discussions		10%

Version 1, January 2000

Laboratory Assignments

- 1 - Demonstrate knowledge of Arcview and Spatial Analyst.
- 2 - Comparison of maps.
- 3 - Create a probabilistic model working with an interdisciplinary team.
  - Familiarity with Arc WofE
  - Graduate students are expected to demonstrate a leadership role in this assignment.

Version 1, January 2000

Class Participation

- What I know is obvious!
- Your job is to ask questions!

Version 1, January 2000

Examinations

- Take home
- Short essays
- Probably will be one at the end of each section.

Discussions

- Journal articles will be assigned to enhance material in book.
  - Discussion of these articles will be lead by graduate students.
- Laboratory assignments will be presented and discussed in class by students.

Goals and Expectations

- To introduce the concepts and process of spatial modeling in GIS.
- Emphasis on probability and favorability models, that is nondeterministic models.
- Students are GIS experts!

What is a model?

- A simplification of nature.
- A representation of a set of objects and their relationships.
- A model is a way of describing something that cannot be directly observed.
- A model is a way of communicating complex ideas.

Why Model?

- "…when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of science, whatever the matter may be." Lord Kelvin
- GIGO "Garbage In, Garbage Out"

Reading Assignment

- Read King and Kramer
  - Why models?
  - Volunteer to present?
- Chapter 7
  - Reclassification
  - Filtering

Laboratory Assignment

- Objective - Demonstrate knowledge of Arcview and Spatial Analyst.
  – Show logic and creativity in the use of the data.
- Data - Lake Tahoe
  – Watershed10.shp and Dem10m

Version 1, January 2000

Laboratory Exercise

- For each basin, calculate mean elevation, mean aspect, and relief.
- Rank basins by potential snowfall using only information from Dem10m and Watershed10. shp.
- For a portion of the basin, create an interesting elevation residual.
- Present your results as a short, 8.5x11 page-size report. Include a concise summary of processing steps for a knowledgeable user.

Version 1, January 2000

Challenges in this exercise

- How to process the data in Arcview and to report the results elegantly.
- How to concisely summarize the processing steps.
  - Assume a knowledgeable Arcview users, such as yourself.
  - Example: View/Menu/Analysis/Reclassify
    - Classify type - equal intervals, 5 classes

Laboratory Grading

- Elegant solution
- Logical thinking
- Quality of writing and cartography
  - Concise writing is a virtue.
  - Tell the reader only what is important.

Laboratory Hint

- Zonal Statistics
  - Look up help/index/zones.
  - Look in help/contents/extensions/spatial analyst/performing analysis/calculate summary attributes using a grid theme
- Neighborhood Statistics
  - Look up help/index/neighborhoods
  - Look in help/contents/extensions/spatial analyst/performing analysis/calculating neighborhood statistics.

Background

- Measurement Scales
- Precision
  - Integers versus Real numbers
- Map Scale and Resolution
- Guidelines for modeling

Measurement Scales

- Nominal (Categorical)
  - An unordered label of categories or classes.
- Ordinal (Rank)
  - Measurements ordered (ranked) according to relative position on a scale with unequal intervals between classes.
- Interval
  - Measurements that can be labeled and ordered with an equal interval between classes but without a true zero.
- Ratio
  - Measurements that can be labeled and ordered, with an equal interval between classes, and with a true zero.

Version 1, January 2000

Examples of Measurement Scales

Scale Type	Examples	Operations	Means
Nominal	Rock type	=	Mode
Ordinal	Relative age	> <	Median
Interval	Temperature	+-*/	Mean
Ratio	Distance	+-*/	Mean

Version 1, January 2000

Precision = a measure of ability to distinguish between nearly equal numbers.

- The number of significant figures determines how maps can be reclassified and symbolized.
- Integers versus real numbers in Arcview
  - Real numbers can be reclassified by equal intervals and standard deviations
  - Integers can be also be reclassified byquantiles, equal areas, and natural breaks
  - Integer grids have VAT

Version 1, January 2000

Map Scale and Resolution

Map Scale	Base Resolution	Information	Buffer?
1:2,500,000	1250	2500	5000
1:500,000	250	500	1000
1:250,000	125	250	500
1:100,000	50	100	200

Units - Meters

Base Resolution ~ Scale denominator / 2000

Version 1, January 2000

Guidelines for Modeling

- Formal statement of the problem.
- Define the user of the model.
- Specification - preprocess the data to provide useful information, that is evidence.
  - Data exploration
    - Reclassification, filtering, transformation, and scaling
  - Reduce the dimensionality by eliminating redundant or correlated information
  - Use the minimum information necessary
- Prediction - combine the evidence to create the model.
- Testing - evaluate the model and it's properties.

Properties of Evidence

- Selected attributes must discriminate between one or more classes of objects.
- Selected attributes must not be correlated with other attributes to any moderately strong extent.
- Selected attributes must have meaning for humans.

Scientific Method

- Define a problem
- Gather pertinent data
- Form a working hypothesis or explanation
- Do experiments to test the hypothesis
- Interpret the results
- Draw a conclusion and modify the hypothesis as needed.

Version 1, January 2000

Occam's Razor

Occam's razor is a logical principle attributed to William of Occam, although it was used by some scholastic philosophers prior to him. The principle states that a person should not increase, beyond what is necessary, the number of entities required to explain anything, or that the person should not make more assumptions than the minimum needed. This principle is often called the principle of parsimony. Since the Middle Ages it has played an important role in eliminating fictitious or unnecessary elements from explanations. In the development of logic, logicians such as Bertrand Russell removed traditional metaphysical concepts by applying Occam's razor. Questions have been raised, however, as to whether a person can determine without any doubt that given entities or assumptions are not needed in an explanation. Unless this determination can be made, it is impossible to tell with complete certainty when the principle can be applied

Note: the following has been abstracted from the Grolier Encyclopedia.

Version 1, January 2000

Spatial Analysis in GIS Overview

- Overview
- Why Model? - King and Kramer

USGS science for a changing world

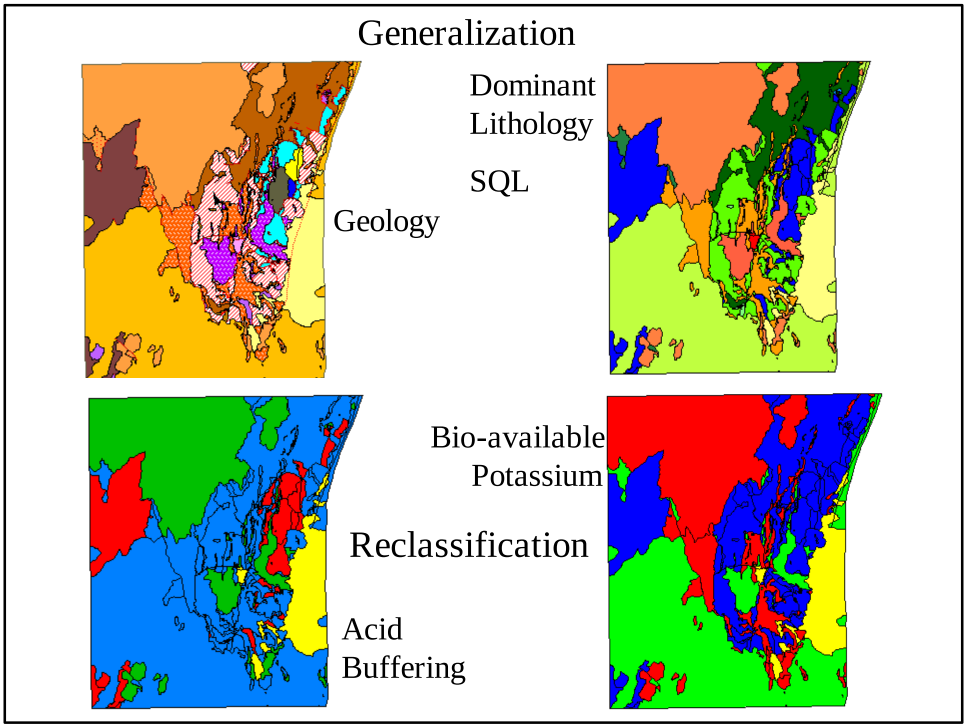
Version 1, January 2000

Overview

- Lake Tahoe Poster
- King and Kramer Discussion
- Weights of Evidence Models

USGS science for a changing world

Version 1, January 2000



King and Kramer

Modeling Continuum

Engineering ⟷ Science ⟷ Public Policy Decision Making

Engineering	Science	Public Policy Decision Making
How to build a bridge at location X?	Which location is best?	Do we need a bridge somewhere?

Deterministic ⟷ Probabilistic
Fuzzy
?

Version 1, January 2000

King and Kramer

- Models are most useful when the right answer is not clear.
- Modeling clarifies the issues of debate in evaluation of an answer.
- Modeling enforces a discipline of analysis, discourse, and consistency.
- Models provide a powerful form of “advice”, that is not “truth”, but a refined result of a particular viewpoint.

USGS
science for a changing world

Version 1, January 2000

Spatial Analysis in GIS
Single Maps

- Modeling - Pattern Recognition
- Reclassification
- Filtering

Version 1, January 2000

Guidelines for Modeling

- Formal statement of the problem.
- Define the user of the model.
- Specification - preprocess the data to provide useful information, that is evidence.
  - Data exploration
    - Reclassification, filtering, transformation, and scaling
  - Reduce the dimensionality by eliminating redundant or correlated information
  - Use the minimum information necessary
- Prediction - combine the evidence to create the model.
- Testing - evaluate the model and it’s properties.

Version 1, January 2000

Data Exploration

- Process of seeking patterns on maps that help predict spatial phenomena.
  - Visualization leads to recognition of a pattern and the association of the pattern with something of interest.
  - A model is proposed that describes the association.

Data Exploration

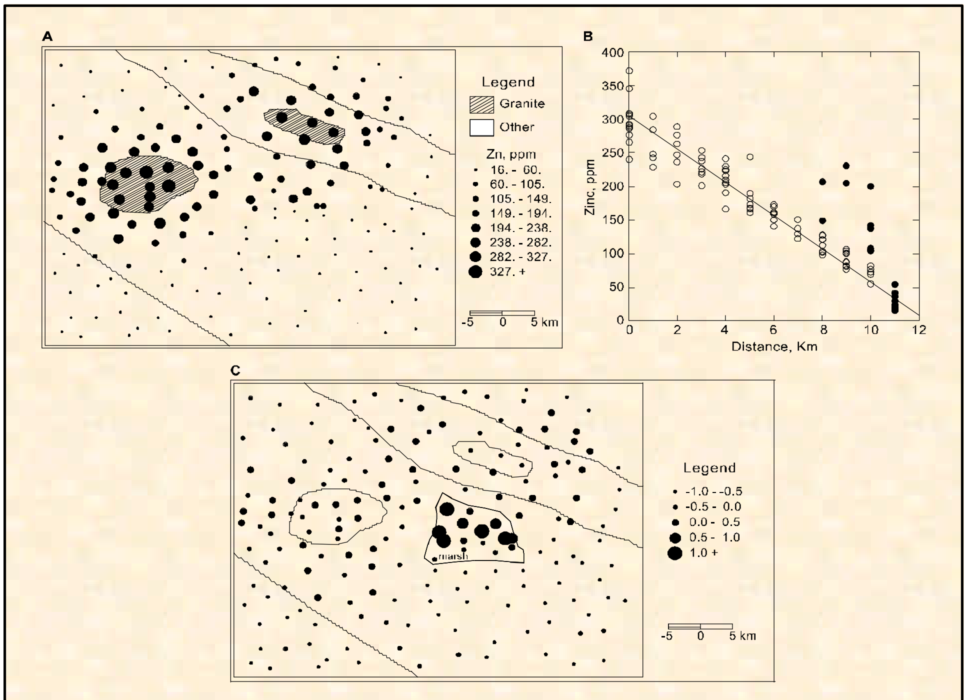
- Seeking patterns involves:
  - Measurement
  - Statistical Summary
  - Visualization
  - Description
  - Understanding of processes causing pattern
- Foundation is data model.

Pattern

- An area having a consistent, recognizable characteristics associated with some object or process.
  - A pattern is something that deviates from the norm.
  - A pattern is associated with a particular scale of observation!
  - It is a primitive.
- Association of patterns and their causes are the bricks of scientific knowledge.

Types of Recognition

- Classification is the process of grouping objects together in classes according to perceived similarities.
- Identification is the recognition of an individual object as a unique singleton class.



Bonham-Carter, 1996

Recognition of a Pattern

- Task - Determine what the appropriate level of aggregation and simplification is for the problem at hand, a problem of reclassification.
  - Aggregation and simplification are tied to scale of observation.
  - The is no single scale at which to view a system.
  - Does not mean that all scales serve equally well or there are not scaling laws.
- Description of patterns is the starting point.
- Spatial models start with an assemblage of patterns and associated processes.

Version 1, January 2000

Measurement Scales

- Nominal (Categorical)
  - An unordered label of categories or classes.
- Ordinal (Rank)
  - Measurements ordered (ranked) according to relative position on a scale with unequal intervals between classes.
- Interval
  - Measurements that can be labeled and ordered with an equal interval between classes but without a true zero.
- Ratio
  - Measurements that can be labeled and ordered, with an equal interval between classes, and with a true zero.

Reclassification

- Reclassification Methods - Continuous measurement scales
  - Natural breaks
  - Quantile, Equal area
  - Equal intervals
  - Standard deviation
- Semantic Reclassification - Categorical measurement scales

Data Transformations

- Transform to common range

$$X_i^* = \frac{X_i - X_{\min}}{X_{\max} - X_{\max}}$$

Skewed Distributions

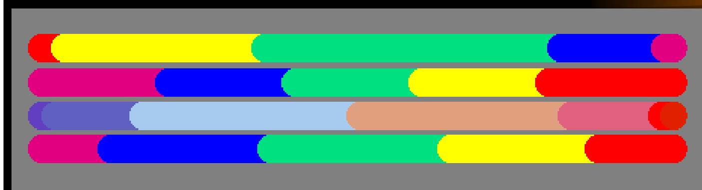
$$X_i^* = \frac{X_i - \bar{X}}{STD(X)} \qquad X_i^* = \log(X_i)$$

Floating to Integer Transform

$$X_i^* = (X_i + 0.5.AsGrid).\text{int}$$

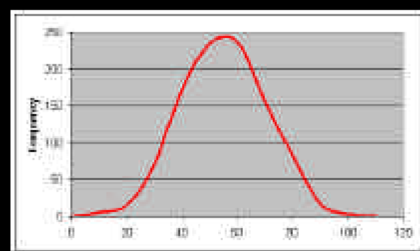
where
X_i^* is an integer value
X_i is a floating value

Reclassification Continuous Measurements



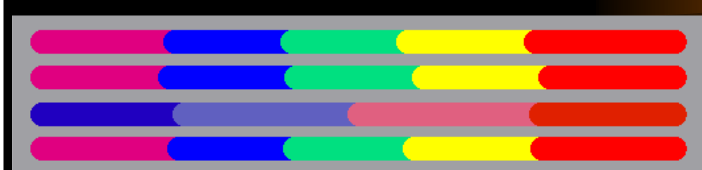
Equal Interval
Quantile
Standard Deviation
Natural Breaks

Input - 1000 points of sorted values from a normal distribution with a mean of 50 and standard deviation of 15.



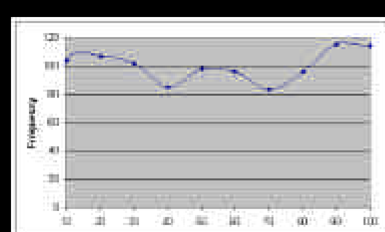
Version 1, January 2000

Reclassification Continuous Measurements



Equal Interval
Quantile
Standard Deviation
Natural Breaks

Input - 800 points of sorted values from a uniform distribution.



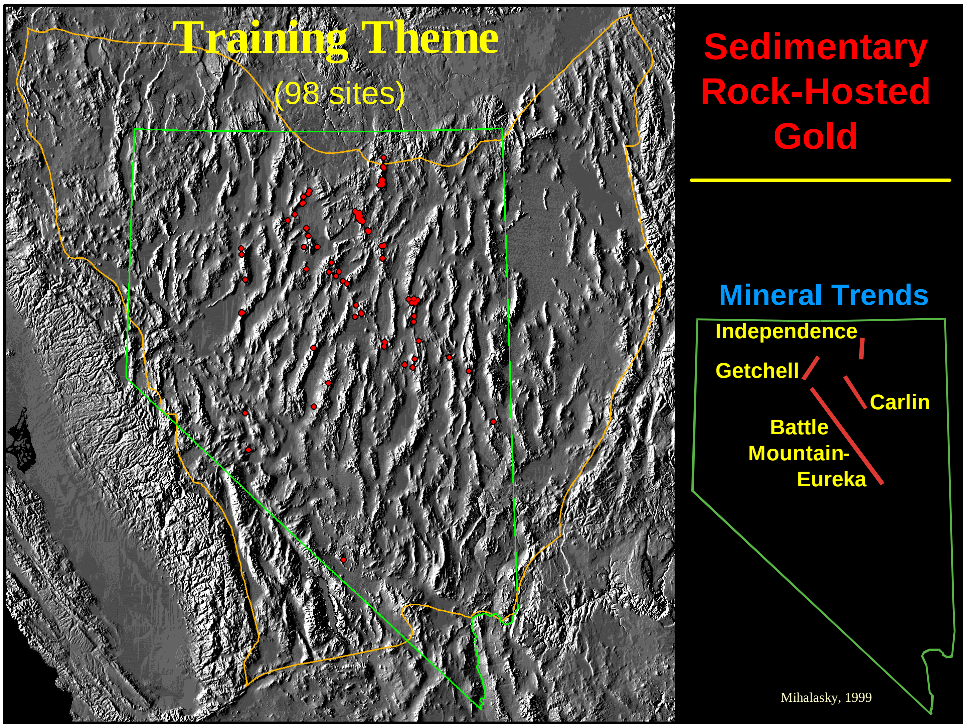
Version 1, January 2000

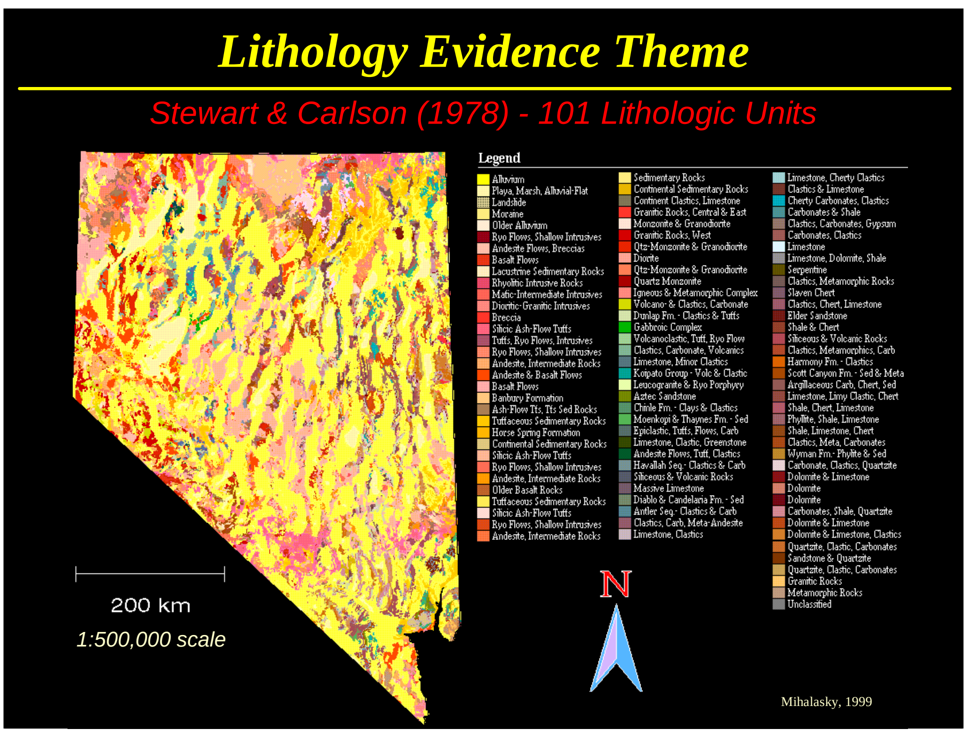
Guidelines Continuous Measurements

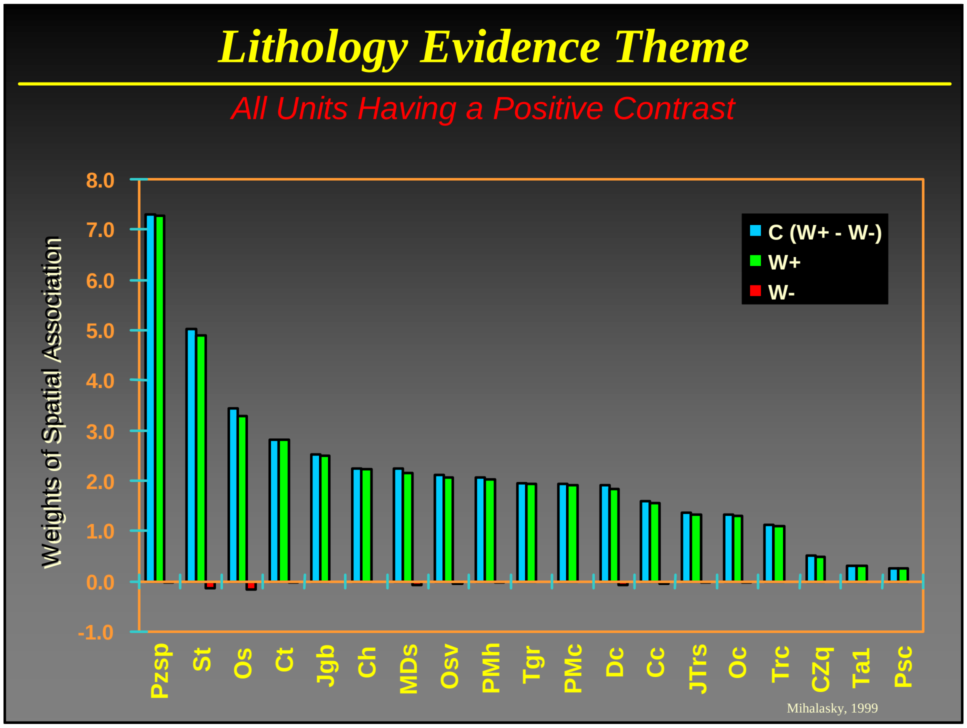
- Histograms are useful.
- Quantile is distribution free.
- Standard deviation, Natural Breaks, and Equal Intervals are not distribution free.
- If interested in tails, use standard deviation.
- If interested in middle, use quantile.
- If interested in minimizing class variance, use natural breaks.

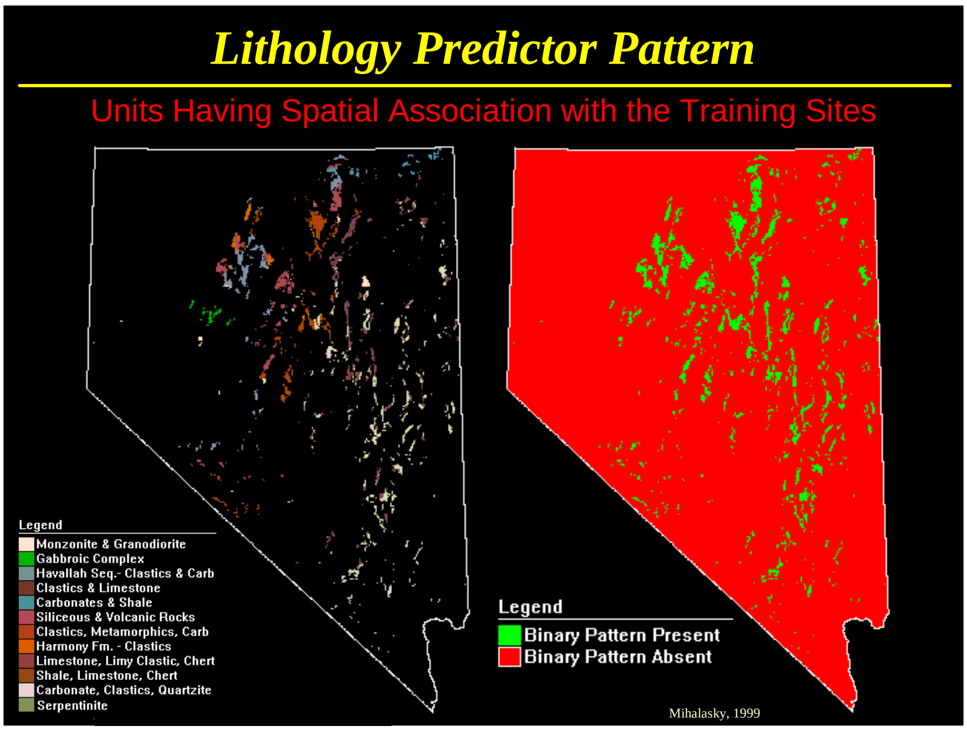
Semantic Reclassification Categorical Measurements

- This is an important problem!
- Expert Systems
  - GeoGen
  - http://geology.usgs.gov/dm/
- Spatial Association - How to define?
  - Expert decision
  - Measurement such as Arc WofE Contrast

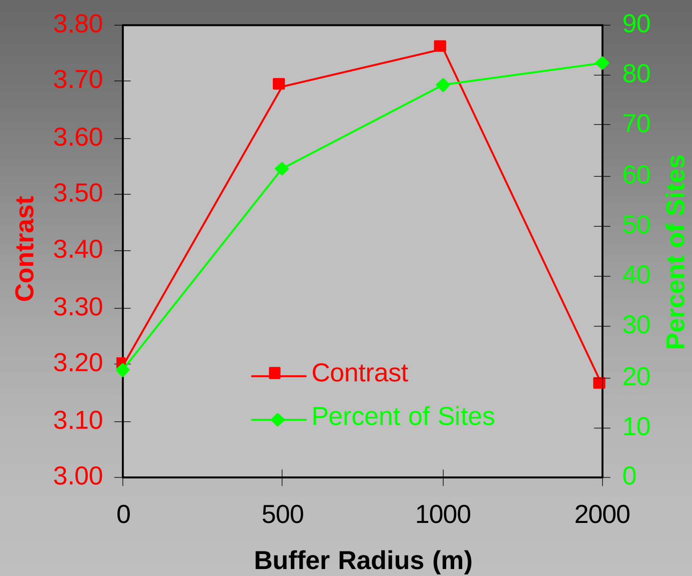

Training Theme
(98 sites)
Sedimentary Rock-Hosted Gold
Mineral Trends
Independence
Getchell
Carlin
Battle Mountain-Eureka
Mihalasky, 1999


Lithology Evidence Theme
Stewart & Carlson (1978) - 101 Lithologic Units
Legend
Alluvium
Playa, Marsh, Alluvial-Flat
Landslide
Moraine
Older Alluvium
Ryo Flows, Shallow Intrusives
Andesite Flows, Breccias
Basalt Flows
Lacustrine Sedimentary Rocks
Rhyolitic Intrusive Rocks
Mafic-Intermediate Intrusives
Dioritic-Granitic Intrusives
Breccia
Silicic Ash-Flow Tuffs
Tuffs, Ryo Flows, Intrusives
Ryo Flows, Shallow Intrusives
Andesite, Intermediate Rocks
Andesite & Basalt Flows
Basalt Flows
Banbury Formation
Ash-Flow Tfs, Tfs Sed Rocks
Tuffaceous Sedimentary Rocks
Horse Spring Formation
Continental Sedimentary Rocks
Silicic Ash-Flow Tuffs
Ryo Flows, Shallow Intrusives
Andesite, Intermediate Rocks
Older Basalt Rocks
Tuffaceous Sedimentary Rocks
Silicic Ash-Flow Tuffs
Ryo Flows, Shallow Intrusives
Andesite, Intermediate Rocks
Sedimentary Rocks
Continental Sedimentary Rocks
Continent Clastics, Limestone
Granitic Rocks, Central & East
Monzonite & Granodiorite
Granitic Rocks, West
Qtz-Monzonite & Granodiorite
Diorite
Qtz-Monzonite & Granodiorite
Quartz Monzonite
Igneous & Metamorphic Complex
Volcano- & Clastics, Carbonate
Dunlap Fm. - Clastics & Tuffs
Gabbroic Complex
Volcanoclastic, Tuff, Ryo Flow
Clastics, Carbonate, Volcanics
Limestone, Minor Clastics
Koipato Group - Volc & Clastic
Leucogranite & Ryo Porphyry
Aztec Sandstone
Chinle Fm. - Clays & Clastics
Moenkopi & Thaynes Fm. - Sed
Epiclastic, Tuffs, Flows, Carb
Limestone, Clastic, Greenstone
Andesite Flows, Tuff, Clastics
Havallah Seq.- Clastics & Carb
Siliceous & Volcanic Rocks
Massive Limestone
Diablo & Candelaria Fm. - Sed
Antler Seq.- Clastics & Carb
Clastics, Carb, Meta-Andesite
Limestone, Clastics
Limestone, Cherty Clastics
Clastics & Limestone
Cherty Carbonates, Clastics
Carbonates & Shale
Clastics, Carbonates, Gypsum
Carbonates, Clastics
Limestone
Limestone, Dolomite, Shale
Serpentine
Clastics, Metamorphic Rocks
Slaven Chert
Clastics, Chert, Limestone
Elder Sandstone
Shale & Chert
Siliceous & Volcanic Rocks
Clastics, Metamorphics, Carb
Harmony Fm. - Clastics
Scott Canyon Fm. - Sed & Meta
Argillaceous Carb, Chert, Sed
Limestone, Limy Clastic, Chert
Shale, Chert, Limestone
Phyllite, Shale, Limestone
Shale, Limestone, Chert
Clastics, Meta, Carbonates
Wyman Fm.- Phylite & Sed
Carbonate, Clastics, Quartzite
Dolomite & Limestone
Dolomite
Dolomite
Carbonates, Shale, Quartzite
Dolomite & Limestone
Dolomite & Limestone, Clastics
Quartzite, Clastic, Carbonates
Sandstone & Quartzite
Quartzite, Clastic, Carbonates
Granitic Rocks
Metamorphic Rocks
Unclassified
200 km
1:500,000 scale
N
Mihalasky, 1999


Lithology Evidence Theme
All Units Having a Positive Contrast
Weights of Spatial Association
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
C (W+ - W-)
W+
W-
Pzsp
St
Os
Ct
Jgb
Ch
MDs
Osv
PMh
Tgr
PMc
Dc
Cc
JTrs
Oc
Trc
CZq
Ta1
Psc
Mihalasky, 1999


Lithology Predictor Pattern
Units Having Spatial Association with the Training Sites
Legend
Monzonite & Granodiorite
Gabbroic Complex
Havallah Seq.- Clastics & Carb
Clastics & Limestone
Carbonates & Shale
Siliceous & Volcanic Rocks
Clastics, Metamorphics, Carb
Harmony Fm. - Clastics
Limestone, Limy Clastic, Chert
Shale, Limestone, Chert
Carbonate, Clastics, Quartzite
Serpentinite
Legend
Binary Pattern Present
Binary Pattern Absent
Mihalasky, 1999

Near - Proximity To Alteration



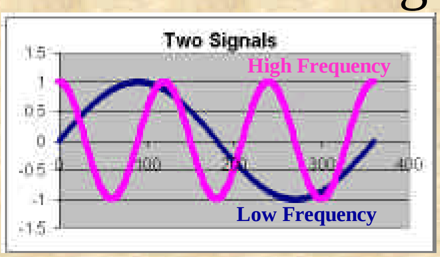
Summary

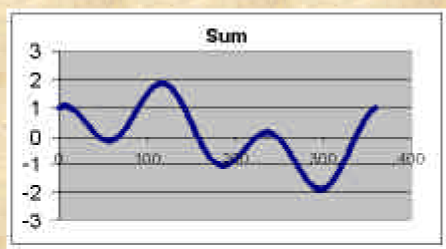
- Concept of a pattern.
- Reclassification of continuous measurement scales.
  - Many tools
- Reclassification of categorical measurement scales.
  - Few tools

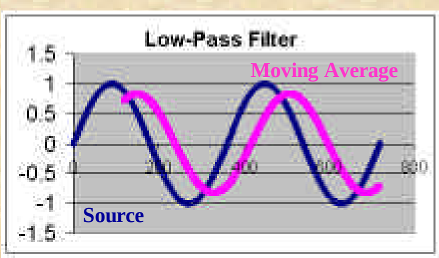
Spatial Analysis in GIS
Single Maps

- Modeling - Pattern Recognition
- Reclassification
- Filtering

Filtering Overview







Moving Average

$$F_{(t+1)} = 1/N \sum_{j=1}^{N} A_{t-j+1}$$

N = Number of prior periods to include in average

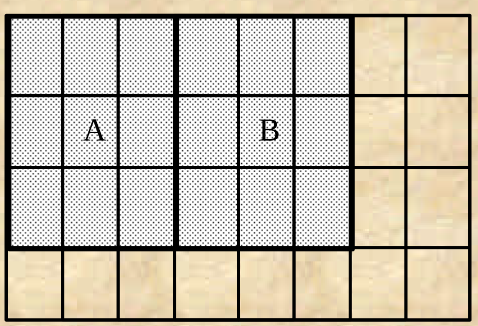
A_j = Actual value at time j

F_j = Forcasted value at time j

Filters

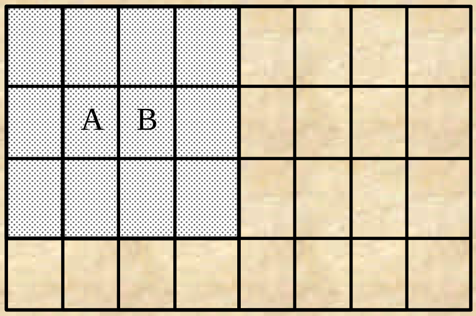
- Interpolate a surface
  - Inverse distance weighting (IDW)
  - Spline
- Block statistics and Focal statistics
  - Neighborhood Statistics
- Zonal Statistics
- Hillshade, slope, and aspect
- Convolution Filters

3x3 BlockStats Function



- No overlap of neighborhoods
- All cells in neighborhood receive same value
- A way to decrease the resolution

3x3 FocalStats Function



- Overlapping neighborhoods
- Only the central value receives the new value
- Loose the outside of the theme.

Types of Neighborhoods or Filters

Interval and Ratio Scales

*Mean (Low Pass)

Standard Deviation

Ordinal Scales

*Median (Low Pass)

Nominal Scales

*Majority (Low Pass)

*Variety (Diversity)

Maximum (High Pass?)

Minimum (Low Pass?)

Kernal Properties

Height and Width - 3x3

Type of neighborhood

Weights

1/9 1/9 1/9

1/9 1/9 1/9 Mean Filter weights

1/9 1/9 1/9

Others

Minority?

Sum- Program other filters

Convolution Filters

Laplacian weights

0	-1	0
-1	4	-1
0	-1	0

High Frequency

-1	-1	-1
-1	9	-1
-1	-1	-1

<u>Directional</u> West

1	1	-1
1	-2	-1
1	1	-1

High Frequency

0.25	0.50	0.25
0.50	1.00	0.50
0.25	0.50	0.25

<u>Directional</u> South

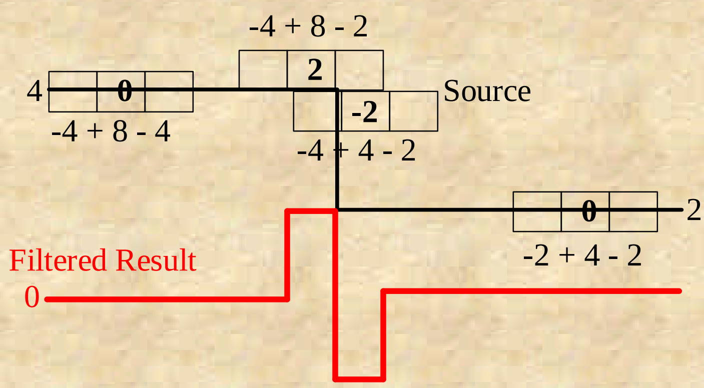
-1	-1	-1
1	-2	1
1	1	1

There are a large number of other filters for many applications.

Available as Arcview extensions with problems.

Version 1, January 2000

Laplacian Filter



One-Dimensional Laplacian weigths: -1 2 -1

Version 1, January 2000

Cascade Programming

Problem: How do you define the weights? Neighborhoods can only be defined as including or not including a cell (0,1).

Fragment of Cascading Avenue Code

```
firstLine = {0,1,0}
secondLine = {1,0,1}
thirdLine = {0,1,0}
theKernal = {firstLine,secondLine,thirdLine}
aNbrHood = NbrHood.MakeIrregurlar (theKernal)
theResult = sourceGrid*4.AsGrid -
   sourceGrid.FocalStats(#GRID_STATYPE_SUM,
   theNbrHood, True)
```

Version 1, January 2000

Cascade Neighborhoods

Laplacian 3x3

0	1	0
1	0	1
0	1	0

Laplacian 9x9

0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
1	1	1	0	0	0	1	1	1
1	1	1	0	0	0	1	1	1
1	1	1	0	0	0	1	1	1
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0

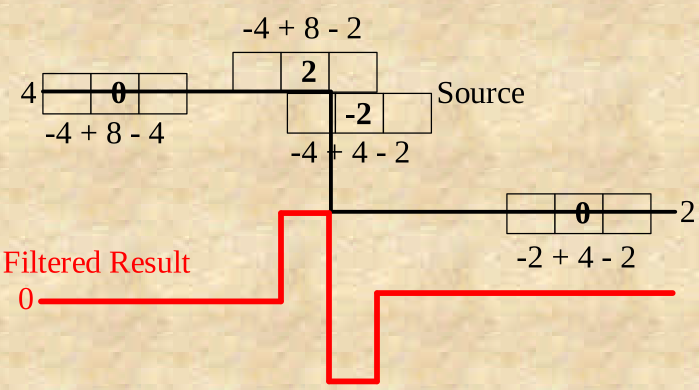
Odd number of rows and columns!

Version 1, January 2000

Recursive Filtering

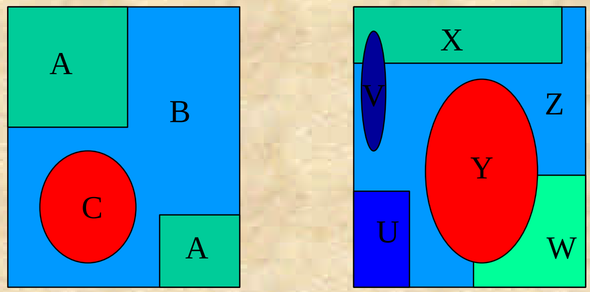
- Often necessary to filter the filtered grid to remove artifacts.
  - For example on the Laplacian, may only want the high and not the low.
  - May wish to eliminate isolated cells.
- Often human interpretation necessary to remove various types of artifacts.

Laplacian Filter



One-Dimensional Laplacian weigths: -1 2 -1

ZonalStats



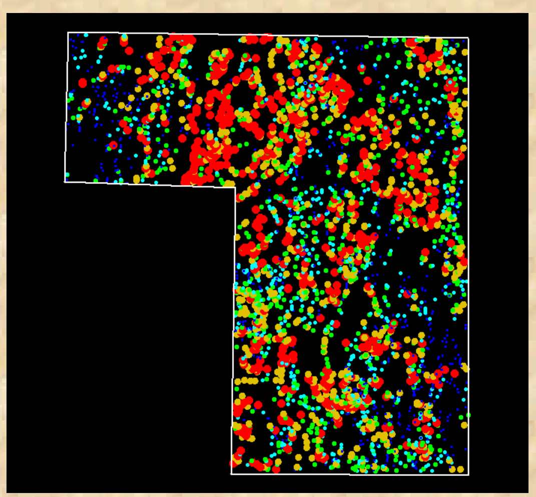
Shape or Grid Theme — Grid Theme

The Zones — Measurement to Summarize

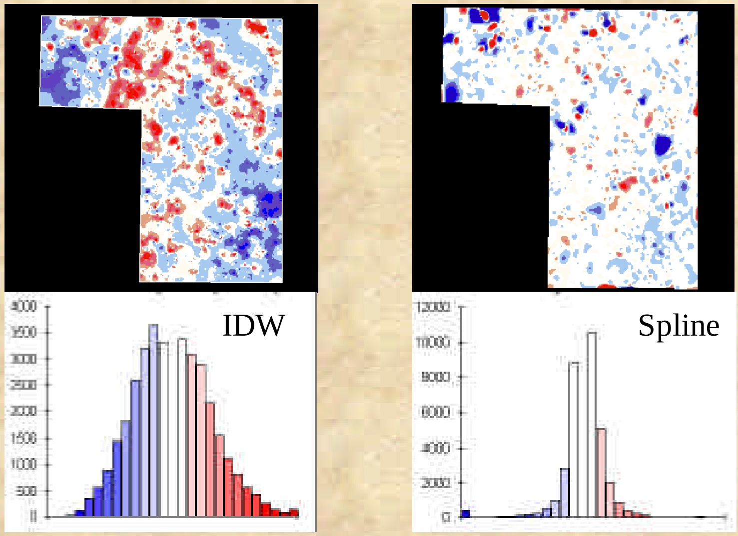
Table from ZonalStats

	Mean	STD	Min	Max
A|	n1	n2	n3	n4
B|	n5	n6	n7	n8
C|	n9	n10	n11	n12

Antimony Point Samples

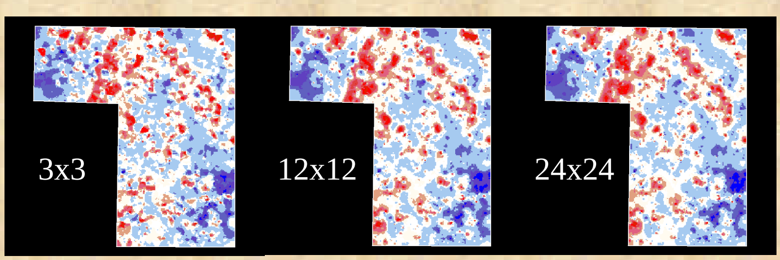


Interpolation Methods

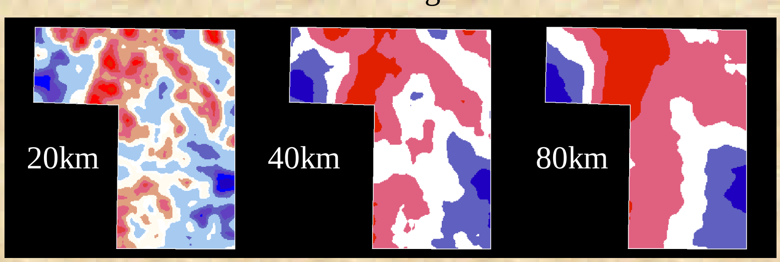


Symbolized by 1/4 standard deviations

Filtering Antimony

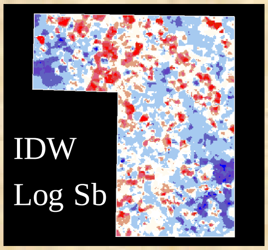


IDW Log Sb



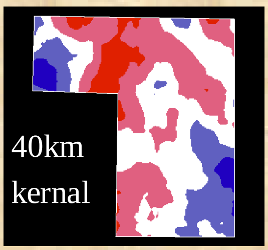
Neighborhood mean from 3x3 surface at various kernal sizes

A pattern?

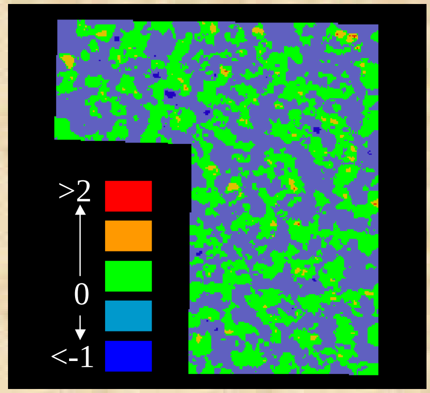


Original

Minus →



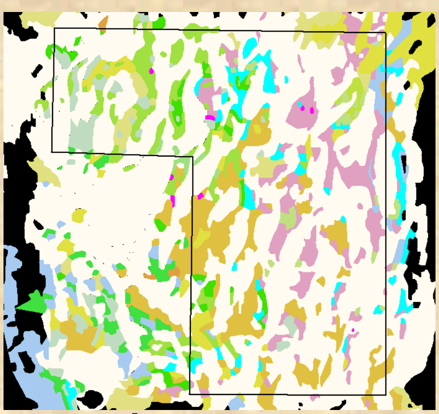
"Low Pass"



High-Pass Residual

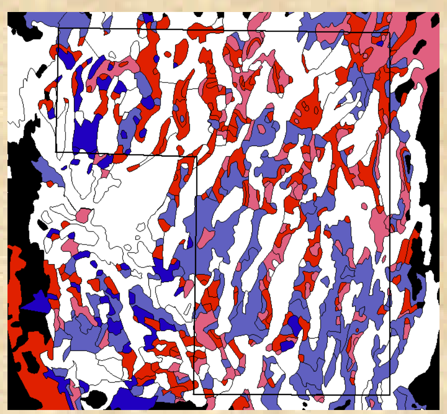
(IDW 3x3) - (Mean 40km)

Zonal Statistics



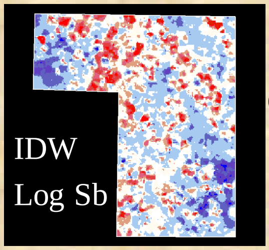
Geologic Units

Central Nevada



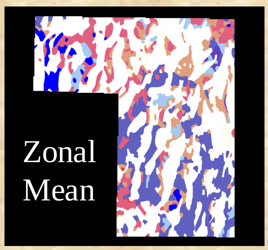
Mean Antimony

Another Pattern?

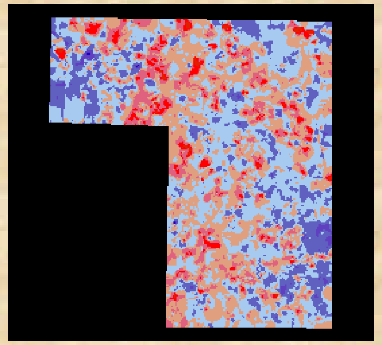


Original

Minus



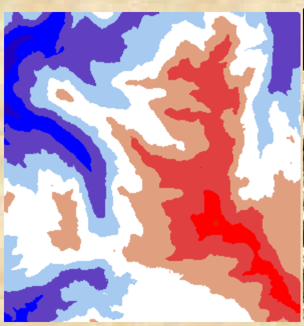
“Low Pass”



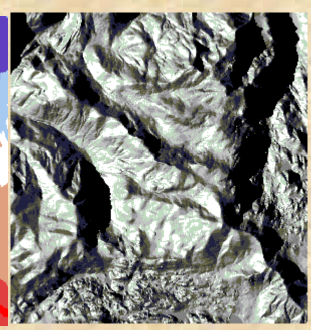
High-Pass Residual
IDW 3x3 - Zonal Stat

Version 1, January 2000

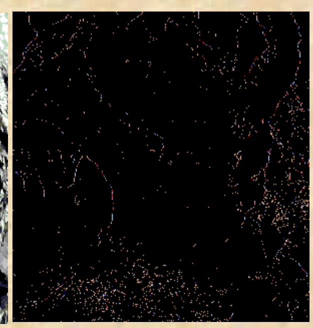
3 x 3 Laplacian Filter



Source: DEM



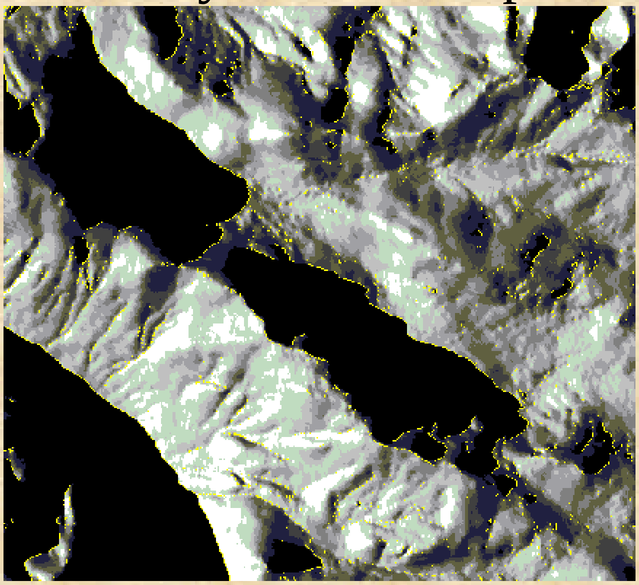
Shaded Relief
HillShade



Laplacian of
Shaded Relief

Version 1, January 2000

Recursively Filtered Laplacian



Version 1, January 2000

Filtering Summary

- Objective of filtering is to define a pattern that may not be obvious in the original data.
  - Edges of homogeneous areas are often important.
- Filtering is an art!
  - May require recursive filtering or interpretation to remove artifacts.
- Powerful tool for data exploration!

Version 1, January 2000

Spatial Analysis in GIS
Map Pairs

- Overlay
- Map Correlation

USGS science for a changing world

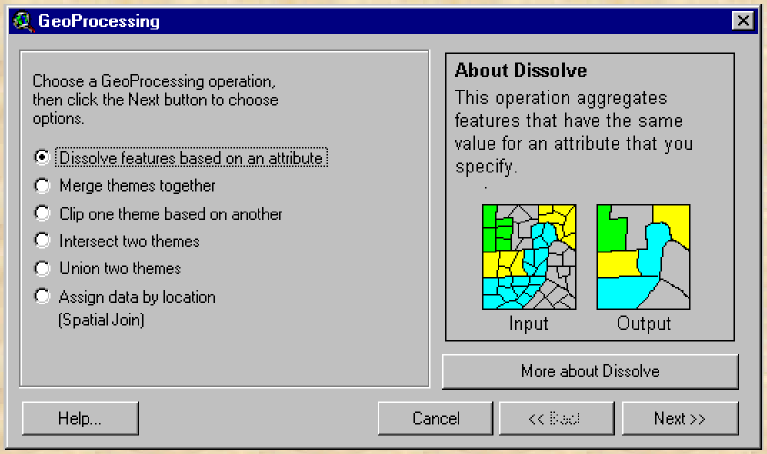
Version 1, January 2000

Guidelines for Modeling

- Formal statement of the problem.
- Define the user of the model.
- Specification - preprocess the data to provide useful information, that is evidence.
  - Data exploration
    - Reclassification, filtering, transformation, and scaling
  - Reduce the dimensionality by eliminating redundant or correlated information
  - Use the minimum information necessary
- Prediction - combine the evidence to create the model.
- Testing - evaluate the model and it’s properties.

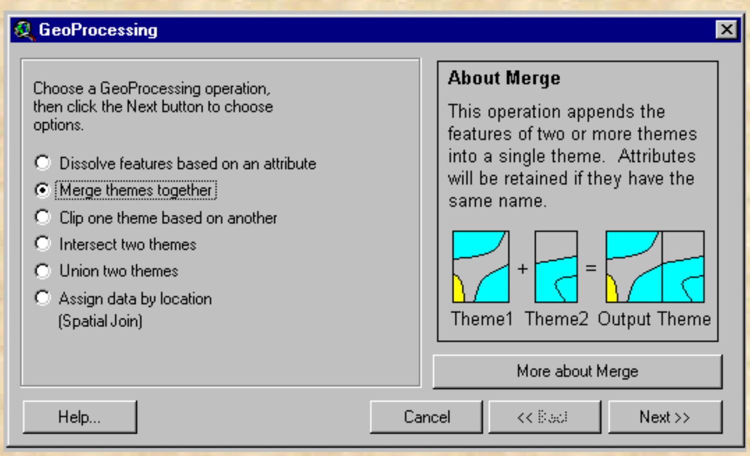
Version 1, January 2000

Shape - Dissolve



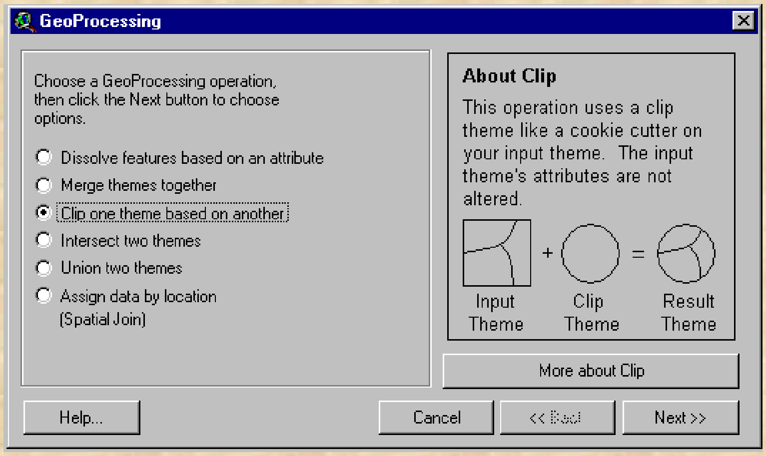
Version 1, January 2000

Shape - Merge



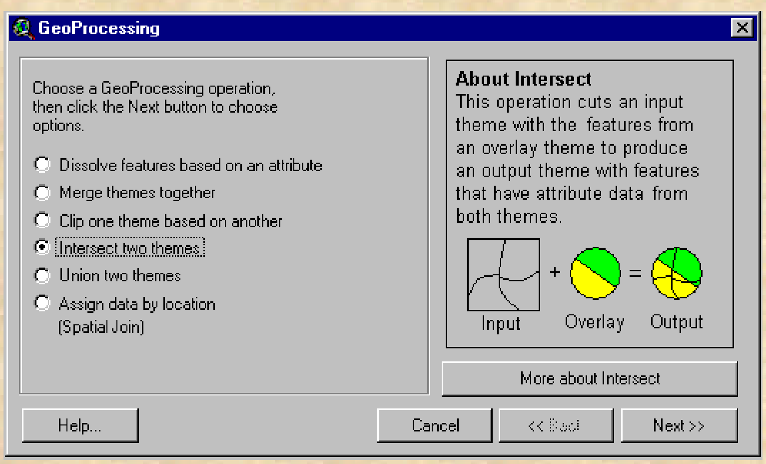
Version 1, January 2000

Shape - Clip



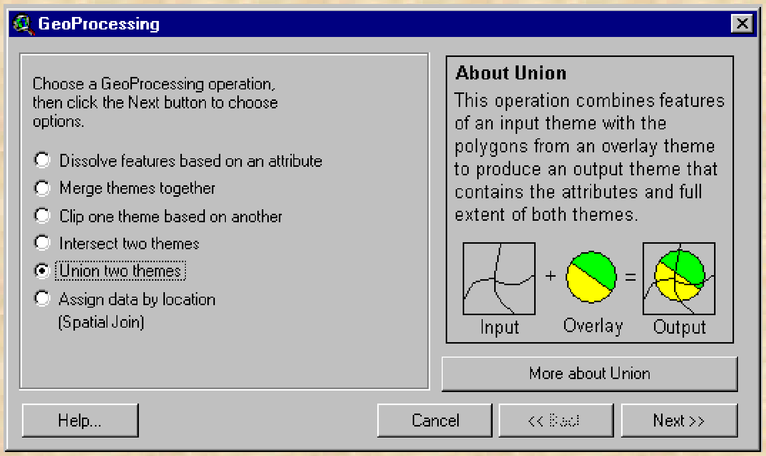
Version 1, January 2000

Shape - Intersect



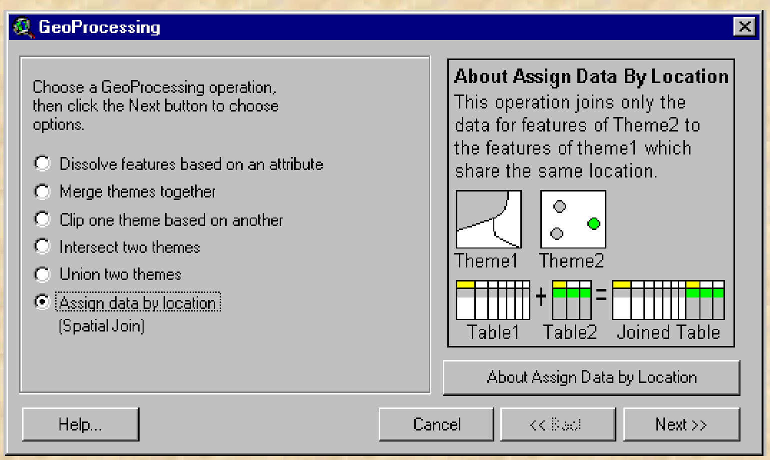
Version 1, January 2000

Shape - Union



Version 1, January 2000

Spatial Join



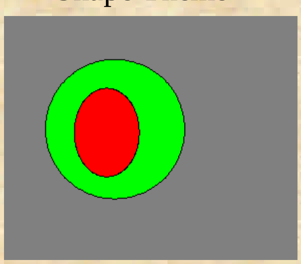
Version 1, January 2000

Combining Grids

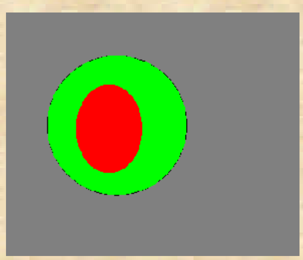
- Zonal Statistics - summarize one grid for zones in another grid or shape file
- Map Calculator - some sort of map algebra
- Merge grids
  - Unique polygons
  - Unique conditions

Stamped Overlay

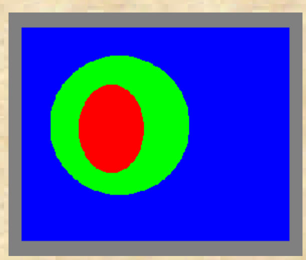
Shape Theme



Grid Theme



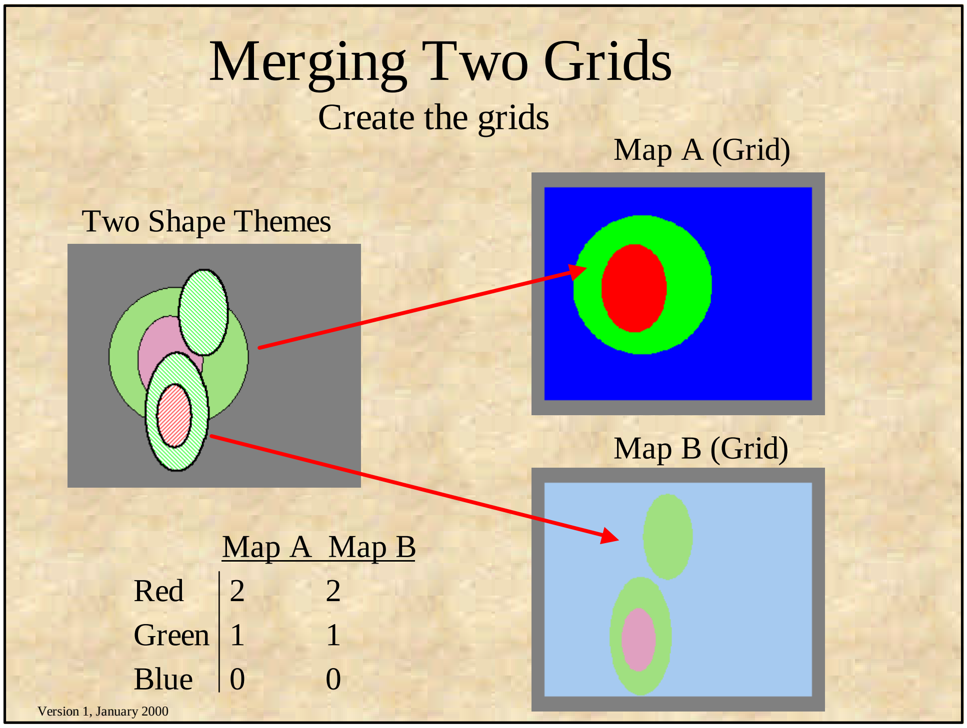
Reclassified Grid

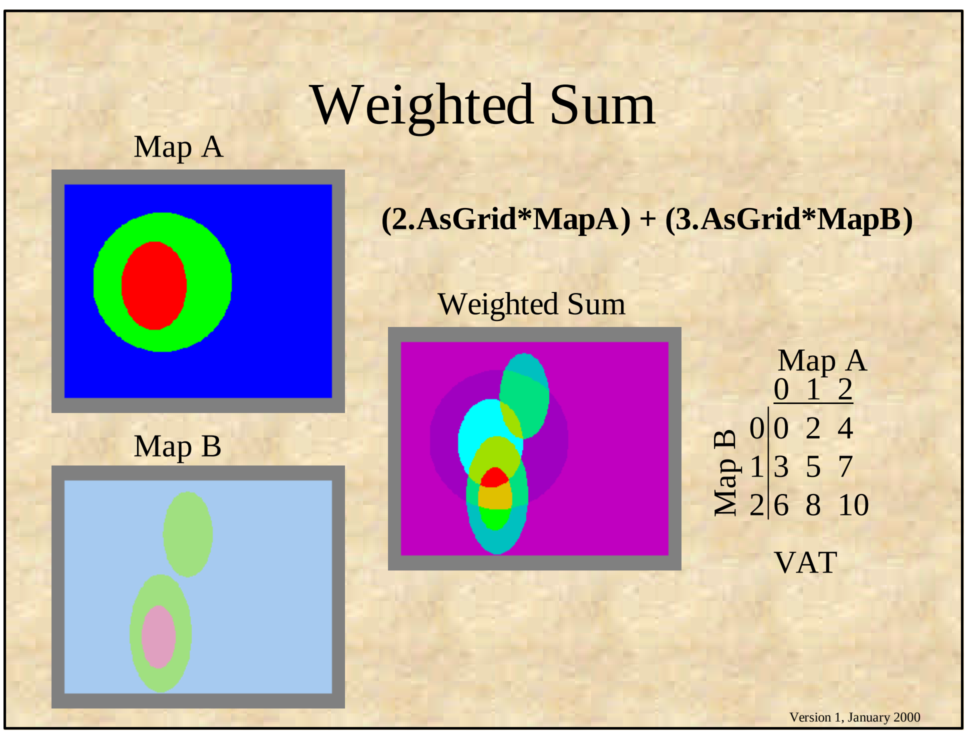


Two Polygons
Red has ID = 2
Green has ID = 1
Gray = No Data

Analysis Parameters/Analysis Extent = View

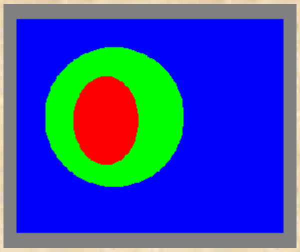
Reclassify/ 2 = 2, 1= 1, No Data = 0


Merging Two Grids
Create the grids
Two Shape Themes
Map A (Grid)
Map B (Grid)
Map A Map B
Red 2 2
Green 1 1
Blue 0 0
Version 1, January 2000


Weighted Sum
Map A
(2.AsGrid*MapA) + (3.AsGrid*MapB)
Weighted Sum
Map B
Map A
0 1 2
Map B
0 0 2 4
1 3 5 7
2 6 8 10
VAT
Version 1, January 2000

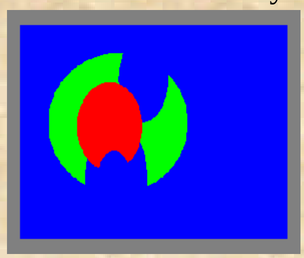
Conditional Overlay

Map A

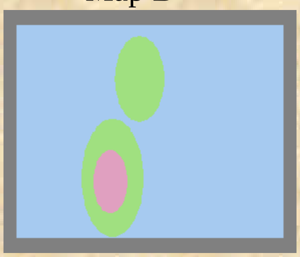


(MapA > MapB).Con(MapA,0.AsGrid)

Conditional Overlay



Map B

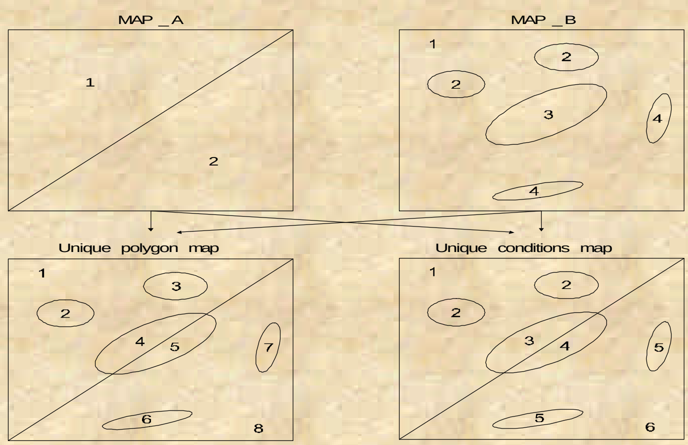


Map B \ Map A	0	1	2
0	0	1	2
1	0	0	2
2	0	0	0

VAT values

Stamped, Joined, Compare

Version 1, January 2000



ATTRIBUTE TABLE OF MAP_C
'UNIQUE POLYGONS'

MAP_C	MAP_A	MAP_B
1	1	1
2	1	2
3	1	2
4	1	3
5	2	3
6	2	4
7	2	4
8	2	1

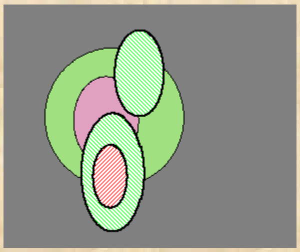
ATTRIBUTE TABLE OF MAP_D
'UNIQUE CONDITIONS'

MAP_D	MAP _A	MAP _B
1	1	1
2	1	2
3	1	3
4	2	3
5	2	4
6	2	1

Bonham-Carter, 1996 8-13

Unique Polygons vs. Unique Conditions

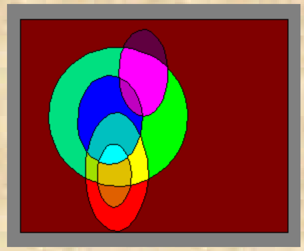
Two Shape Themes



Unique Polygons

14 polygons

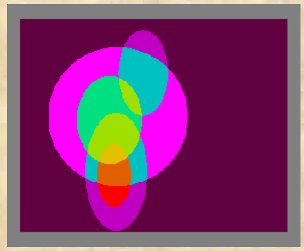
Shape file or Grid



Unique Conditions

9 Classes

Grid theme



Transform Grid/Combine

sptrnfrm.avx

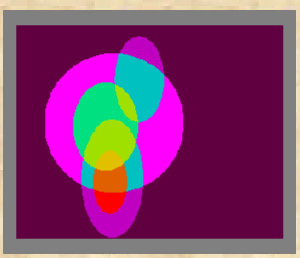
Version 1, January 2000

Problem with the VAT

Unique Conditions

9 Classes

Grid theme



VALUE	COUNT	Map B	Map A
1	5517	0	0
2	3291	1	0
3	9556	0	1
4	4139	1	1
5	2971	0	2
6	2642	1	2
7	545	2	2
8	1071	2	1
9	718	2	0

Value not sorted with regards to Map B and Map A values.

Version 1, January 2000

Frequency for Arcview

VAT with Case added

VALUE	COUNT	Map B	Map A	CASE
1	5517	0	0	1
3	956	0	1	2
5	2971	0	2	3
2	3291	1	0	4
4	4139	1	1	5
6	2642	1	2	6
9	718	2	0	7
8	1071	2	1	8
7	545	2	2	9

Consistent numbering of the matrix or VAT.

Tables/Xtools/Table Frequency

Xtoolsmh.avx

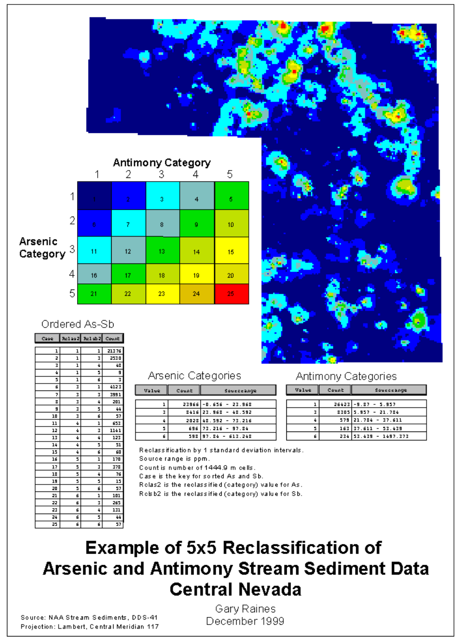
Frequency Table

CASE	FREQUENCY	Map B	Map A	COUNT
1	1	0	0	5517
2	1	0	1	956
3	1	0	2	2971
4	1	1	0	3291
5	1	1	1	4139
6	1	1	2	2642
7	1	2	0	718
8	1	2	1	1071
9	1	2	2	545

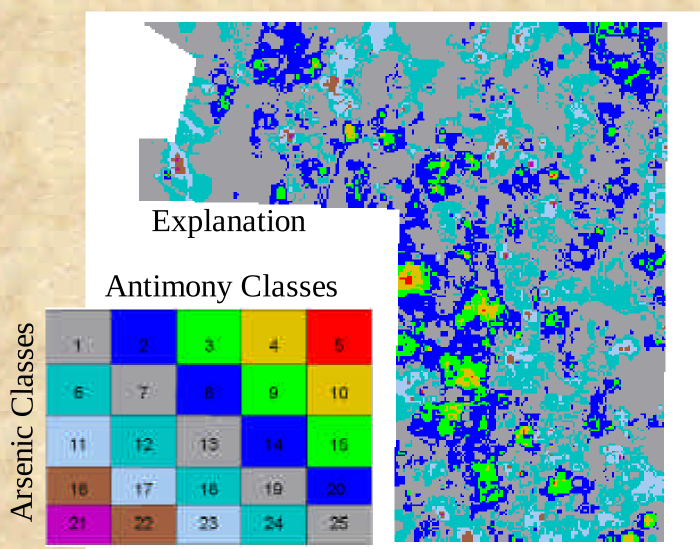
Frq1.dbf

Grid Overlay Application

- Reclassification
- Shape file Legend
- Transform Grid/Combine
- Xtools/Table Frequency



Correlation Analysis



Summary

- Shape files - several tools
  - Computations can be slow
- Grid overlay offers great flexibility
  - Numerical and logical combinations
  - Ordered VAT or table of combinations opens the door for many types of modeling
  - Unique conditions table shortens the ordered matrix and simplifies programming in modeling
  - Computations are very fast

Exercise 2 - Data Sources

- Three shape files (geology, NURE, and soils) and one Arc/Info coverage (timber types)
- Using the NURE shape file in the table calculate log As and log Sb.
  - As is arsenic stream sediment geochemistry in part per million. Sb is antimony stream sediment geochemistry in part per million. The log transformation makes these data more normally distributed.
  - Create grids of each of these log values.

Exercise 2 - NURE (Ratio Data)

- Using the tools in Merge5a.zip, reclassify and merge the LogAs and LogSb NURE grids and spatially evaluate the correlation between those grids.
  - Five classes in each grid would be appropriate as explained in Merge3.
  - Based on the spatial display, how would you describe this spatial correlation?
  - Calculate the Pearson's Correlation for for the LogAs to LogSb grids.
  - Is your spatial evaluation of the correlation different than the calculated Pearson's coefficient?
- For the point NURE data, calculate a Pearson's correlation coefficient.
  - Why are the Pearson's correlation coefficient for the points different for the points and the grids?

Exercise Two - Categorical Correlation

- Use the categorical Tahoe themes to investigate correlation between nominal grids.
  - For example does soil type correlation with geologic unit?

Exercise 2 -Report

- Prepare a short report documenting the procedure and the results.
- Show the inputs maps and the output spatial correlation maps.
- Show your calculations of the categorical correlation measure.
- Discuss your results.

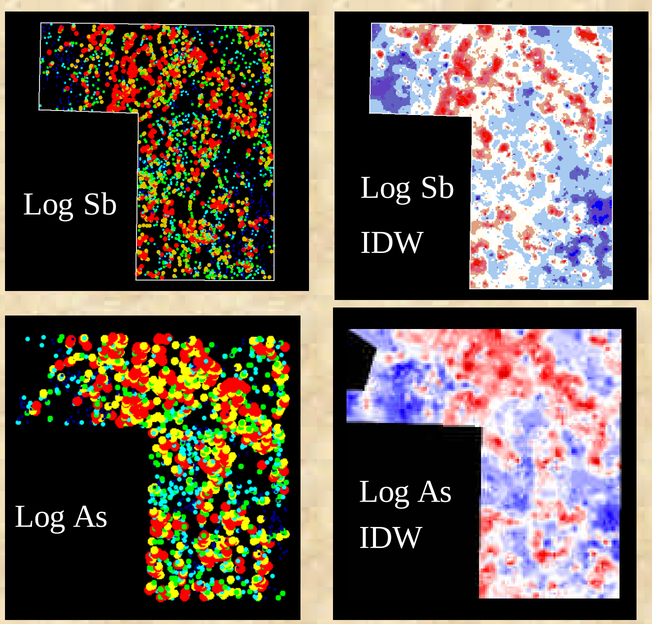
Version 1, January 2000

Spatial Analysis in GIS
Map Pairs

- Overlay
- Map Correlation

Version 1, January 2000

Source Ratio Data



Version 1, January 2000

Correlation Analysis

Points (ratio scale)

Correlation Coefficient

Log As:Log Sb = 0.738

As:Sb = 0.3000

Grid (ratio scale, IDW)

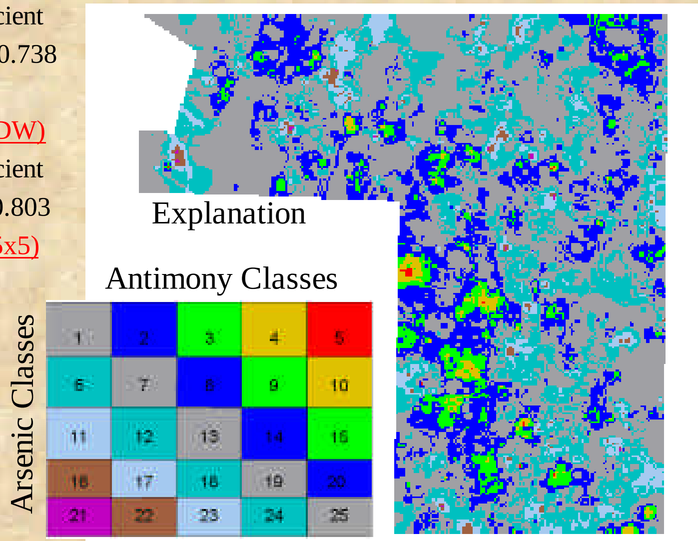
Correlation Coefficient

Log As:Log Sb = 0.803

Grid (reclassified 5x5)

Agreement = 68%

Nominal-Scale Representation



Interval and Ratio Scale

- Pearson's product moment correlation coefficient - measure of linear correlation
  - Varies from -1 to 1
    - -1 - perfect negative correlation
    - 0 - no correlation
    - 1 - perfect positive correlation
  - Use for ratio and interval measurement scales.
  - Not appropriate for nominal and ordinal measurement scales.

Pearson's Correlation Coefficient

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2}}$$

where
X_i and Y_j = values of Map X and Y respectively
$\overline{X}$ and $\overline{Y}$ = Average of Map X and Map Y respectively
n = number of cells in intersection of two Maps.

USGS science for a changing world

Version 1, January 2000

Correlation Coefficient Tools

- Correlation Coefficient Tool
  - For tables and shape files. Need to have both attributes in one place.
  - Corrcoef.avx (Charlie Frye, 1998)
- Grid Covariance
  - For covariance and correlation between grids.
  - Covarian.ave (Kenneth R. McVay, 1998)
  - Works on separate grid themes.
    - Use this on real or floating point grids

USGS science for a changing world

Version 1, January 2000

Covarian.ave

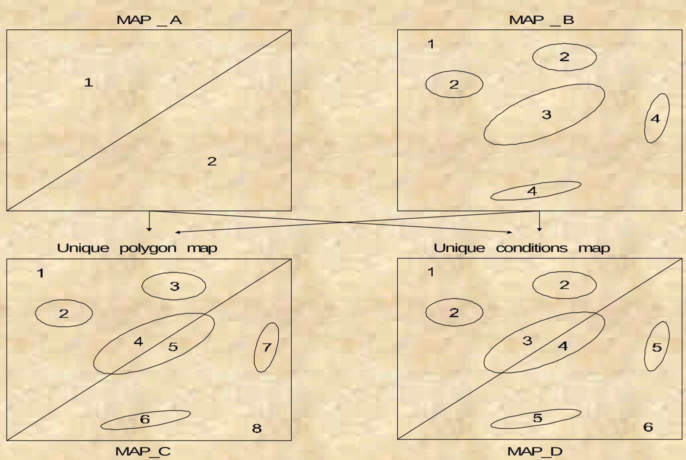
```
Univariate Statistics
              Min            Max            Mean           Stdv

LogAs         0.338654       2.63371        1.16411        0.271111
LogSb         -0.424419      2.35446        0.55231        0.302415

Covariance Matrix
          LogAs          LogSb
LogAs     0.0735009      0.0658513
LogSb     0.0658513      0.091455

Correlation Matrix
          LogAs          LogSb
LogAs     1              0.803182
LogSb     0.803182       1
```

Version 1, January 2000



ATTRIBUTE TABLE OF MAP_C
'UNIQUE POLYGONS'

MAP_C	MAP_A	MAP_B
1	1	1
2	1	2
3	1	2
4	1	3
5	2	3
6	2	4
7	2	4
8	2	1

ATTRIBUTE TABLE OF MAP_D
'UNIQUE CONDITIONS'

MAP_D	MAP _A	MAP _B
1	1	1
2	1	2
3	1	3
4	2	3
5	2	4
6	2	1

Bonham-Carter, 1996 Fig. 8-13

Tabulate Areas or Unique Conditions

Reclassified Antimony

Reclassified Arsenic

VALUE	VALUE_1	VALUE_2	VALUE_3	VALUE_4	VALUE_5
1	881207640.750	411230232.350	2098113.430	0.000	0.000
2	3212211661.900	26041783897.000	354581169.730	0.000	0.000
3	39864155.177	9426823642.600	2979321071.100	10490567.152	0.000
4	0.000	117494352.100	448996274.100	46158495.468	0.000
5	0.000	0.000	23079247.734	4196226.861	2098113.430
	Units = Area			Correlation Coefficient	
					0.803

VALUE	VALUE_1	VALUE_2	VALUE_3	VALUE_4	VALUE_5
1	2.00	0.93	0.00	0.00	0.00
2	7.30	59.18	0.81	0.00	0.00
3	0.09	21.42	6.77	0.02	0.00
4	0.00	0.27	1.02	0.10	0.00
5	0.00	0.00	0.05	0.01	0.00
	Units = Percent of Area				Agreement
					68.07

Agreement = 100*(Sum of Diagonal (gray cells)/Total).

Also called area cross tabulation or confusion matrix.

USGS

Version 1, January 2000

Nominal Scale Data

Chi-square statistic

Area Cross-Tabulation Table

Map A

Map B

T_{11}	T_{12}	T_{13}	…	$T_{1.}$
T_{21}	T_{22}	T_{23}	…	$T_{2.}$
T_{31}	T_{32}	T_{32}	…	$T_{3.}$
…	…	…	…	…
$T_{.1}$	$T_{.2}$	$T_{.3}$	…	$T_{..}$

$$T_{ij}^{*} = \frac{T_{i\bullet} * T_{\bullet j}}{T_{\bullet\bullet}}$$

$$c^{2} = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{\left(T_{ij} - T_{ij}^{*}\right)^{2}}{T_{ij}^{*}}$$

Where

T_{ij}, where there are I = 1, 2, 3, …, N classes on Map B (rows of the table) and j = 1, 2, 3, …, M classes on Map A (columns of the table).

Ti. is the sum of the i^{th} row,

T.j is the sum of the j^{th} column, and

T.. is grand sum over rows and columns.

USGS

Version 1, January 2000

Nominal Scale Data

Information Statistic

Area-Proportions Cross-Tabulation Table

Map A (columns); **Map B** (rows)

P_{11}	P_{12}	P_{13}	…	$P_{1.}$
P_{21}	P_{22}	P_{23}	…	$P_{2.}$
P_{31}	P_{32}	P_{32}	…	$P_{3.}$
…	…	…	…	…
$P_{\bullet 1}$	$P_{\bullet 2}$	$P_{\bullet 3}$	…	$P_{..}$

Where
$P_{ij} = T_{ij}/T_{..}$
$P_{i.} = T_{i.}/T_{..}$
$P_{.j} = T_{.j}/T_{..}$

Information Statistics

$$H(A) = -\sum_{j=1}^{m} p_{\bullet i} \ln p_{\bullet j}$$

$$H(B) = -\sum_{j=1}^{n} p_{i\bullet} \ln p_{i\bullet}$$

Joint Entropy

$$H(A,B) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \ln p_{ij}$$

Joint Information Uncertainty

$$U(A,B) = 2\left[\frac{H(A)+H(B)-H(A,B)}{H(A)+H(B)}\right]$$

USGS science for a changing world

Version 1, January 2000

Nominal Scale Data

Coefficient of Agreement, kappa

Area-Proportions Cross-Tabulation Table

Map A (columns); **Map B** (rows)

P_{11}	P_{12}	P_{13}	…	$P_{1.}$
P_{21}	P_{22}	P_{23}	…	$P_{2.}$
P_{31}	P_{32}	P_{32}	…	$P_{3.}$
…	…	…	…	…
$P_{\bullet 1}$	$P_{\bullet 2}$	$P_{\bullet 3}$	…	$P_{..}$

Where
$P_{ij} = T_{ij}/T_{..}$
$P_{i.} = T_{i.}/T_{..}$
$P_{.j} = T_{.j}/T_{..}$

$$k = \frac{\sum_{i=1}^{n} p_{ii} - \sum_{i=1}^{n} q_{ii}}{1 - \sum_{i=1}^{n} q_{ii}}$$

where

$q_{ij} = p_{i\bullet} * p_{\bullet j}$

n = number of classes, which is the same in both maps.

Conditional kappa for the i-th class

$$k_i = \frac{p_{ii} - q_{ii}}{p_{i.} - q_{ii}}$$

USGS science for a changing world

Version 1, January 2000

Weighted Pearson's Correlation Coefficient
Modified for Cross-Tabulation Table

$$r_s = \frac{\sum_{i=1}^{columns}\sum_{j=1}^{rows} T_{ij} * \left(X_i - \overline{X}\right)\left(Y_j - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{columns}\sum_{j=1}^{rows} T_i * (X_i - \overline{X})^2 \sum_{i=1}^{columns}\sum_{j=1}^{rows} T_i * (Y_j - \overline{Y})^2}}$$

where
X_i and Y_j = Values of Map X and Map Y respectively
$\overline{X}$ and $\overline{Y}$ = Area-weighted mean respectively
T_{ij} = area or count in the i^{th} - j^{th} pair or cell of the cross-tabulation matrix.

Version 1, January 2000

Ordinal Data
Weighted Spearman's Rank Correlation

$$r_s = 1 - \frac{6 * \sum_{i=1}^{rows}\sum_{j=1}^{columns} T_{ij} * \left(R_{xi} - R_{yj}\right)^2}{n\left(n^2 - 1\right)}$$

where
R_x and R_y are ranks for Maps X and Y
n = sum of cells in cross-tabulation matrix

Use this formula:

where ranks are numbered 1, 2, 3, …,n and

where there are no ties.

Version 1, January 2000

Bonham-Carter's Modification

Weighted Spearman's Correlation Coefficient

$$r_{sw} = \frac{\sum_{i=1}^{columns}\sum_{j=1}^{rows} T_{ij} * \left(R_{xi} - \overline{R}_x\right)\left(R_{yj} - \overline{R}_y\right)}{\sqrt{\sum_{i=1}^{columns}\sum_{j=1}^{rows} T_{ij} * (R_{xi} - \overline{R}_x)^2 \sum_{i=1}^{columns}\sum_{j=1}^{rows} T_{ij} * (R_{yj} - \overline{R}_y)^2}}$$

where

R_{xi} and R_{yj} = ranks of Map X and Y respectively

$\overline{R}_x$ and $\overline{R}_y$ = Area-weighted average rank respectively

T_{ij} = area or count in the i^{th} - j^{th} pair.

USGS

Version 1, January 2000

Spearman's - Ordinal Data

Table 8-9B

Tabulate Area Table

T_{ij}

	Map A (x)					Sum	Cum.	Rank Y
Map B (y)	80	0	18	0	0	98	98	49
	37	71	0	0	0	108	206	152
	0	0	0	10	0	10	216	211
	0	0	0	22	30	52	268	242
	0	0	17	3	51	71	339	303.5
Sum	117	71	35	35	81		169.5	
Cum.	117	188	223	258	339	169.5		
Rank X	58.5	152.5	205.5	240.5	298.5			

Sum = row or column sum

Cum. = Cumulative row or column sum

$$r_{swG} = 0.826$$

$$r_{sw} = 0.874$$

$$RankX_i = CumX_{i-1} + \frac{SumX_i}{2} = 223 + \frac{35}{2} = 240.5$$

$$\overline{RankX} = \frac{Max(CumX_i)}{2} = \frac{339}{2} = 169.5$$

USGS

Version 1, January 2000

Summary

- Quantitative comparison between two maps can be done several ways!
  - Chap. 8 provides a brief overview and a starting point for further investigation.
- Area tabulation or cross-tabulation table is a fundamental input to most of the correlation measures.
- Major measures are Pearson's, Confusion Matrix, and Rank Correlation.
- Next week - Correlation measures for binary map pairs.

Version 1, January 2000

Spatial Analysis in GIS
Map Pairs - Part 2

- Overlay
- Map Correlation - Binary Maps

USGS science for a changing world

Version 1, January 2000

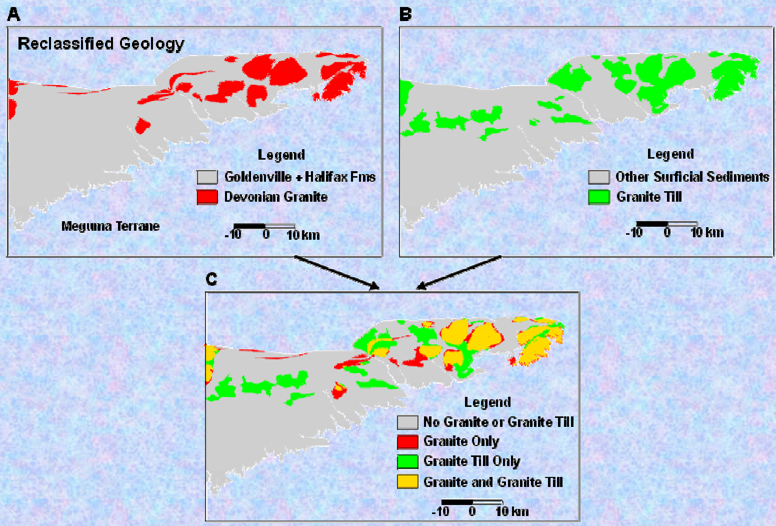
Guidelines for Modeling

- Formal statement of the problem.
- Define the user of the model.
- Specification - preprocess the data to provide useful information, that is evidence.
  - Data exploration
    - Reclassification, filtering, transformation, and scaling
  - Reduce the dimensionality by eliminating redundant or correlated information
  - Use the minimum information necessary
- Prediction - combine the evidence to create the model.
- Testing - evaluate the model and it's properties.

USGS science for a changing world

Version 1, January 2000

Combining Two Binary Grids



Bonham-Carter, Fig. 8-14

Counts to Probability

	A	$\bar{A}$	
B	345 (T_{11})	382 (T_{12})	727 $(T_{1\bullet})$
$\bar{B}$	141 (T_{21})	2077 (T_{22})	2218 $(T_{2\bullet})$
	486 $(T_{\bullet 1})$	2459 $(T_{\bullet 2})$	2945 $(T_{\bullet\bullet})$

	A	$\bar{A}$	
B	0.117 (p_{11})	0.130 (p_{12})	0.247 $(p_{1\bullet})$
$\bar{B}$	0.049 (p_{21})	0.705 (p_{22})	0.753 $(p_{2\bullet})$
	0.165 $(p_{\bullet 1})$	0.835 $(p_{\bullet 2})$	$(p_{\bullet\bullet})$

Area Tabulation $\longrightarrow$ **Probability Tabulation**

A = Granite B = Granite Till

USGS

Version 1, January 2000

Cross-Tabulation Table

	A	$\overline{A}$	
B	345 (T_{11})	382 (T_{12})	727 $(T_{1\bullet})$
$\overline{B}$	141 (T_{21})	2077 (T_{22})	2218 $(T_{2\bullet})$
	486 $(T_{\bullet 1})$	2459 $(T_{\bullet 2})$	2945 $(T_{\bullet\bullet})$

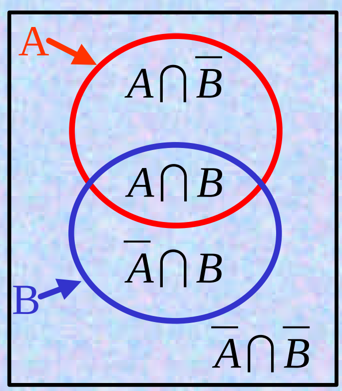
$$\text{Area}\,(\text{A} \cap \text{B}) = T_{11} = 345$$
$$\text{Area}(\text{A} \cap \overline{\text{B}}) = T_{21} = 141$$
$$\text{Area}(\overline{\text{A}} \cap \text{B}) = T_{12} = 382$$
$$\text{Area}(\overline{\text{A}} \cap \overline{B}) = T_{22} = 2077$$

Area Tabulation

Version 1, January 2000

Venn Diagram

$$\text{Area}\,(\text{A} \cap \text{B}) = T_{11} = 345$$
$$\text{Area}(\text{A} \cap \overline{\text{B}}) = T_{21} = 141$$
$$\text{Area}(\overline{\text{A}} \cap \text{B}) = T_{12} = 382$$
$$\text{Area}(\overline{\text{A}} \cap \overline{B}) = T_{22} = 2077$$



Version 1, January 2000

Proportional-Area Table

	A	$\overline{A}$	
B	0.117 (p_{11})	0.130 (p_{12})	0.247 $(p_{1\bullet})$
$\overline{B}$	0.049 (p_{21})	0.705 (p_{22})	0.753 $(p_{2\bullet})$
	0.049 $(p_{\bullet 1})$	0.705 $(p_{\bullet 2})$	$(p_{\bullet\bullet})$

Probability Tabulation

$$p_{ij} = \frac{T_{ij}}{T_{\bullet\bullet}}$$

$$P\{A\} = p_{\bullet 1}$$

$$P\{B\} = p_{1\bullet}$$

$$P\{A \cap B\} = p_{11}$$

$$P\{A \cap \overline{B}\} = p_{21}$$

$$P\{\overline{A} \cap B\} = p_{21}$$

$$P\{\overline{A} \cap \overline{B}\} = p_{22}$$

USGS science for a changing world

Version 1, January 2000

Conditional Probability

$$P\{B \mid A\} = \frac{P\{B \cap A\}}{P\{A\}} = \frac{p_{11}}{p_{\bullet 1}} = \frac{T_{11}}{T_{\bullet 1}}$$

$$P\{\text{Granite Till} \mid \text{Granite}\} = \frac{345}{486} = 0.7098$$

$$P\{\text{Granite Till}\} = p_{1\bullet} = \frac{T_{1\bullet}}{T_{\bullet\bullet}} = 0.247$$

USGS science for a changing world

Version 1, January 2000

Probability and Odds

P = probability

O = odds

$$O = \frac{P}{1-P}$$

P	O	lnO
0.0	0	-∞
.1	1/9	-2.20
.2	1/4	-1.39
.4	2/3	-0.41
.5	1/1	0.00
.6	3/2	0.41
.8	4/1	1.39
.9	9/1	2.20
1.0	∞	∞

USGS science for a changing world

Version 1, January 2000

Conditional Odds

$$O\{B\} = \frac{P\{B\}}{1-P\{B\}} = \frac{\dfrac{T_{1\bullet}}{T_{\bullet\bullet}}}{1-\dfrac{T_{1\bullet}}{T_{\bullet\bullet}}} = \frac{T_{1\bullet}}{T_{\bullet\bullet}-T_{1\bullet}}$$

$$O\{B \mid A\} = \frac{P\{B \mid A\}}{1-P\{B \mid A\}} = \frac{P\{B \mid A\}}{P\{\overline{B} \mid A\}}$$

$$O\{B \mid A\} = \frac{p_{11}/p_{\bullet 1}}{p_{21}/p_{\bullet 1}} = \frac{p_{11}}{p_{21}} = \frac{T_{11}}{T_{21}}$$

USGS science for a changing world

Version 1, January 2000

Conditional Odds Example

$$O\{GraniteTill\} = \frac{727}{2945 - 727} = 0.328$$

or 3 to 10

$$O\{GraniteTill \mid Granite\} = \frac{345}{141} = 2.45$$

or 25 to 10

If Granite is present, then the odds of Granite Till also being present is 25 to 10

Version 1, January 2000

Odds Ratio - Binary Maps

$$O_R = \frac{O\{B \mid A\}}{O\{B \mid \overline{A}\}} = \frac{T_{11}T_{22}}{T_{12}T_{21}}$$

$$O_R = \frac{345 * 2077}{382 * 141} = 13.3$$

$$O_R = \frac{\text{Measure of Agreement}}{\text{Measure of Disagreement}}$$

Version 1, January 2000

Contrast

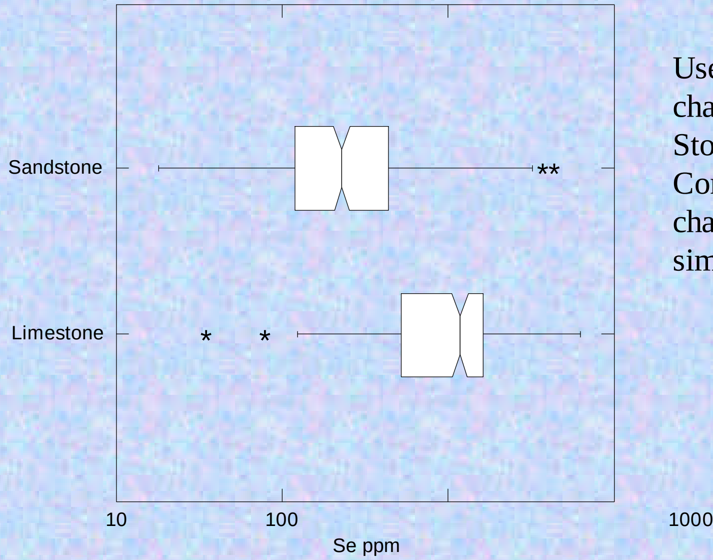
$$C_W = \ln(O_R)$$

$$C_W = \ln(13.3) = 2.59$$

USGS science for a changing world

Version 1, January 2000

Mixed Scales - Box Plots



Use Excel chart types Stock and Combination charts to get similar plots.

USGS science for a changing world

Bonham-Carter, 1996 Fig. 8-16

Mixed Scales

	Nominal	Ordinal	Interval/ Ratio
Nominal	Chi-square, O_r, C_w, etc.	Median by nominal class	Mean by nominal class
Ordinal		Rank correlation coefficient	Rank correlation coefficient
Interval/ Ratio			Covariance Correlation coefficient

Version 1, January 2000

Summary - Correlation

- Ratio and Interval
  - Pearson's correlation coefficient
- Ordinal
  - Spearman's rank correlation coefficient

Version 1, January 2000

Categorical Correlation Summary

α and κ	Useful, nice results between –1 and +1.
κ	Where number of classes match, useful for binary and multi-class maps.
O_R and C_W	Useful, comparable results to κ and α and are easy to compute.
C_j	Useful test if positive agreement is more important than negative agreement.
χ^2, C, and U	Avoid for binary maps. Does not distinguish large interactions due to agreement or disagreement.
C_A	Use with care because does not account for chance associations.
<u>Qualification</u>	Choice of counting region (study area) influences the correlation measured.

Version 1, January 2000

Arc/Info Statistical Tools

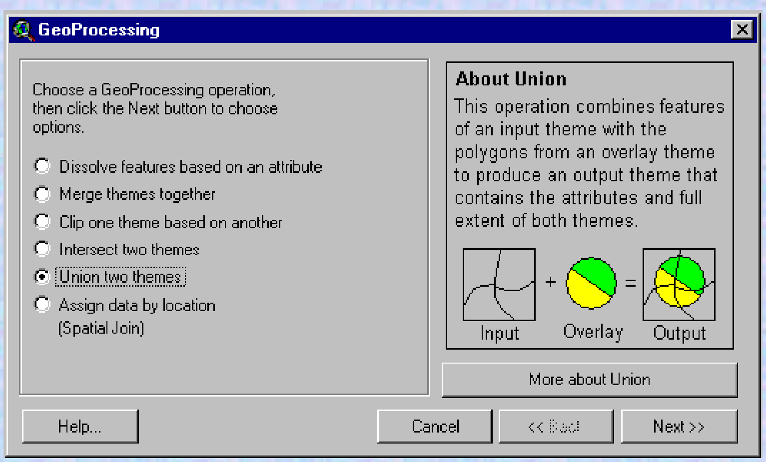
- Grid: Autocorrelation tools
  - Correlation - calculates cross correlation
  - Geary and Moran spatial autocorrelation index
- Grid: Multi-variant clustering
  - Isocluster( ) - natural clustering of attributes in attribute space
  - Mlclassify( ) - maximum-likelihood classification in attribute space
  - Princomp ( ) - principal components classification in attribute space
  - Regression - linear or logistic regression coefficients
- Stackstats - standard statistics for a stack of grids

Version 1, January 2000

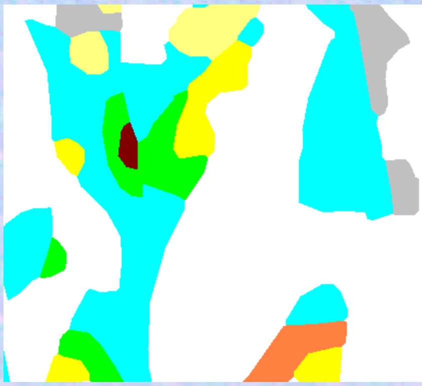
Miscellaneous - Chaps. 7 & 8

- Shape file processing to reduce dimensionality
  - Unions & Intersections
- Topological operations
- Derived properties - calculations
  - DEM
  - Shapes

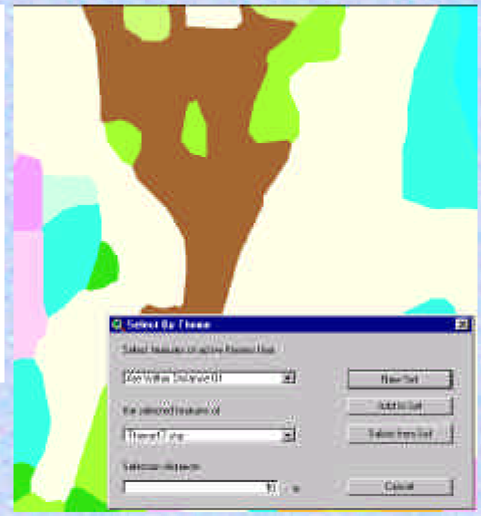
Shape - Union



Topological Operations



Source with selected polygon



Select by Theme without selected polygon.

Version 1, January 2000

Calculations from DEM

- Arcview - Hydrologic modeling extension
  - Watersheds
- Arc/Info DrawZoneShape and ZonalGeometry - major and minor axis of ellipsoid and orientation
- Large number of other calculations
  - Wetness Index = ln(catchment area/tan(slope))
  - Thinness Ratio = 4πArea/Perimeter^2
- Shape analysis
  - Circularity = (4 * Area)/(perimeter)^2
  - Elongation = w/l

Version 1, January 2000

Probability

Put 3 red balls and 7 blue balls in a bag.

What is the probability of drawing a blue ball from the bag?

What is the probability of drawing a red ball from the bag?

Probability of drawing a blue ball is 7/10 = 0.7 $=P_b$

Probability of drawing a red ball is 3/10 = 0.3 = $P_r = 1 - P_b$

Version 1, January 2000

Probability

Probability
0.3
0.2
0.1
0.4

Put the following balls in a bag:

3 red-blue balls ----Red-Blue (RB)
2 red-green balls ----Red-Not Blue (RG)
1 blue-green ball ----Blue-Not Red (BG)
4 green balls ----Not Red-Not Blue (G)

What is the probability of drawing each type?

What is probability of drawing a blue ball?

Marginal Probability of a blue ball = 0.4

<u>Area Tabulation Table</u>

	R	BG	
B	$3=T_{11}$ $0.3=P_{11}$	$1=T_{12}$ $0.1=P_{12}$	$4=T_{1.}$ $0.4=P_{1.}$
RG	$2=T_{21}$ $0.2=P_{21}$	$4=T_{22}$ $0.4=P_{22}$	$6=T_{2.}$ $0.6=P_{2.}$
	$5=T_{.1}$ $0.5=P_{.1}$	$5=T_{.2}$ $0.5=P_{.2}$	$10=T_{..}$

Version 1, January 2000

Tools for Map Analysis

Multiple Maps

Boolean Logic

Index Overlay

Fuzzy Logic

Weights of Evidence

Logistic Regression

Neural Networks

Version 1, January 2000

Reading Assignment

- Chapter 9
  - Look over whole
  - Boolean Logic
  - Index Overlay
  - Fuzzy Logic
  - Bayesian Models (Weights of Evidence)
  - Logistic Regression
  - Fuzzy knowledge representation
  - Neural Networks (Radial Bias and Fuzzy Self Organizing)

Version 1, January 2000

Laboratory Exercise

- Training sets
  - California Spotted Owl habitat
  - Goshawk nesting habitat
  - Osprey nesting habitat
  - IPES (building-permit evaluation)
- Form groups
- Define problem and explore data
  - Primarily TRPA CD-ROM
- Create model
- Poster Presentations (June 8)

Guidelines for Modeling

- Formal statement of the problem.
- Define the user of the model.
- Specification - preprocess the data to provide useful information, that is evidence.
  - Data exploration
  - Data transformation, filtering, and scaling
  - Reduce the dimensionality by eliminating redundant or correlated information
  - Use the minimum information necessary
- Prediction - combine the evidence to create the model.
- Testing - evaluate the model and it's properties.

Version 1, January 2000

Purpose of GIS Projects

- Combine data from diverse sources
- To describe and analyze interactions
- To make predictions, that is models
- To provide support for decision makers

Version 1, January 2000

Properties of Evidence

- Selected attributes must discriminate between one or more classes of objects.
- Selected attributes must not be correlated with other attributes to any moderately strong extent.
- Selected attributes must have meaning for humans.

Version 1, January 2000

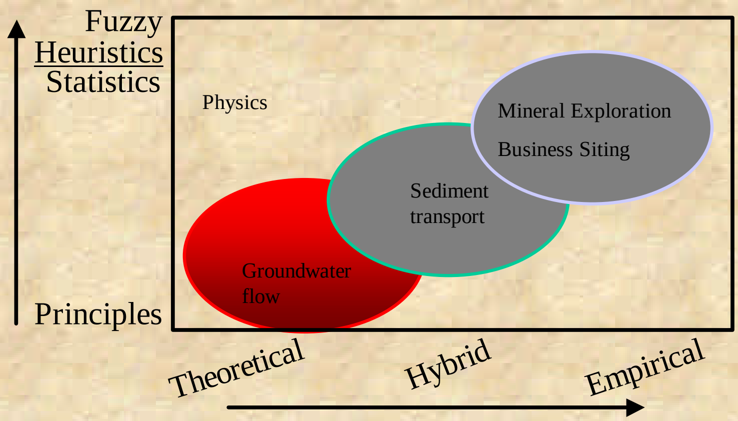
Scientific Method

- Define a problem
- Gather pertinent data
- Form a working hypothesis or explanation
- Do experiments to test the hypothesis
- Interpret the results
- Draw a conclusion and modify the hypothesis as needed.

Types of Models

- Prescriptive or Deterministic
  - Application of good technical practices
  - Process: Boolean rules, Equations
  - Output: Binary map (yes or no)
- Predictive
  - Application of mathematics to represent how people think about the evidence but cannot represent as equations.
  - Process: weighting of evidence and combination of weights
  - Output: Favorability, probability, or fuzzy map (0 to 1)

Types of Models



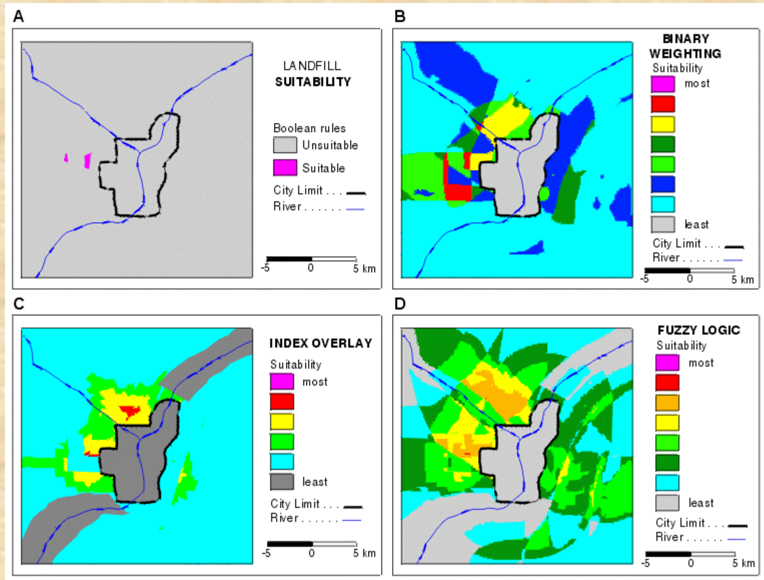
Version 1, January 2000

Knowledge Driven Methods

- **Boolean Logic** - True/False representation of maps with all maps rated equally. Simple method with True/False answer.
- **Index Overlay with Binary Maps** - Maps are given different weights. Linear combination of maps.
- **Index Overlay with Multi-Class Maps** - Maps are given different weights as well as the classes of the maps are given different weights. Linear combination of maps.
- **Fuzzy Logic** - More flexible weighting of maps and map classes. Nonlinear combination of maps.

Version 1, January 2000

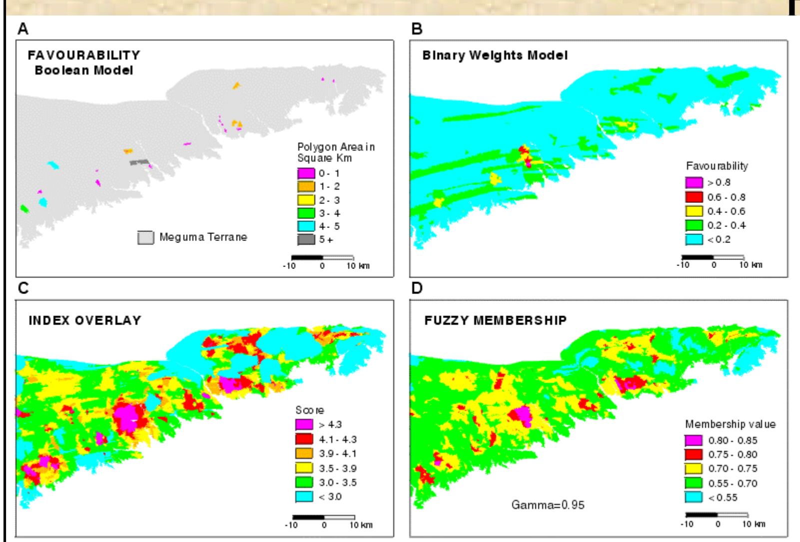
Fig 9-2



A: Boolean B: Binary Index Overlay C: Multi-class Index Overlay D: Fuzzy logic

Version 1, January 2000

Fig 9-4



A: Boolean B: Binary Index Overlay C: Multi-class Index Overlay D: Fuzzy logic

Version 1, January 2000

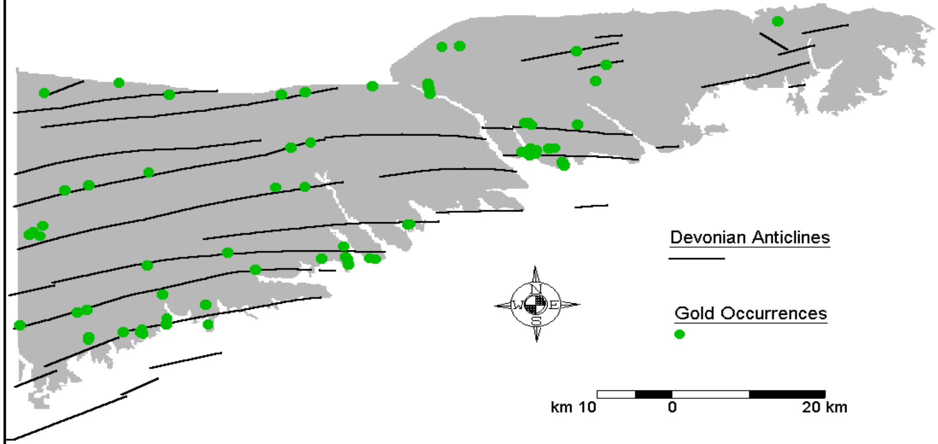
Data Driven Methods

- Weights of Evidence
  - log linear combination of binary maps.
  - Classifies areas by probability or favorability of occurrence of a training site.
  - Model parameters easy to understand.
- Logistic Regression
  - log regression combination of binary maps
  - Classifies areas by probability of occurrence of a training site.
  - Model parameters complex.
- Neural networks
  - Experimental, nonlinear combination of fuzzy or rescaled maps
  - Classifies areas by fuzzy membership in training set.
  - Can also be self organizing to produce fuzzy membership.
  - Model parameters complex.

Fuzzy Membership

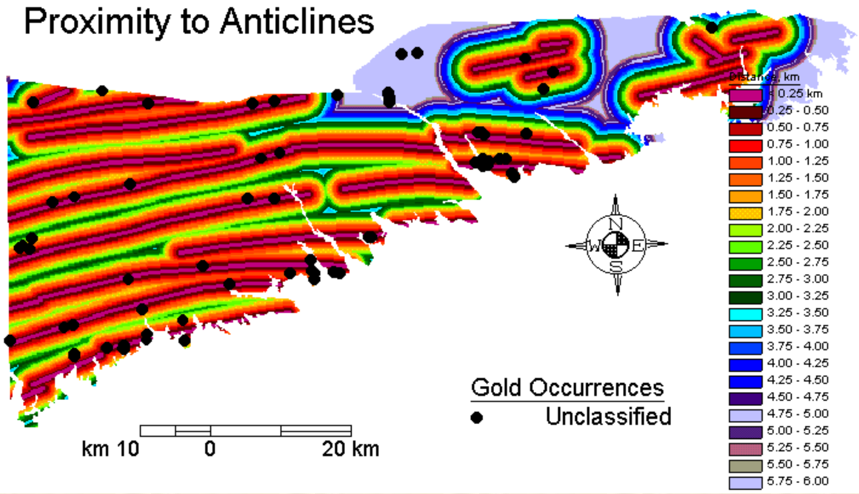
Nominal Measurement Scale

Application of Contrast Proximity

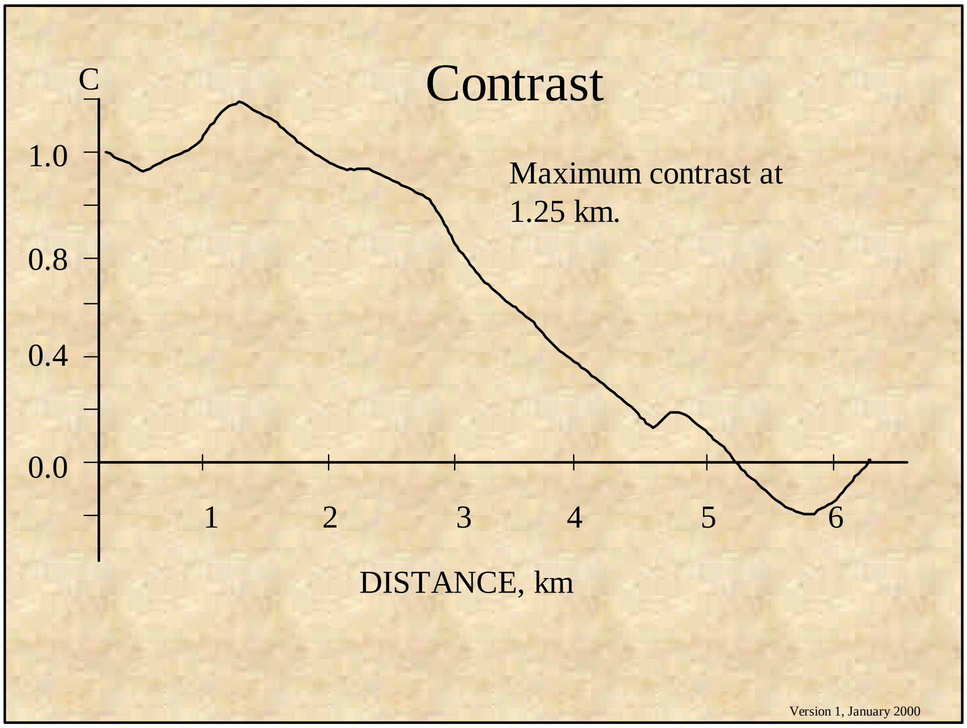


Version 1, January 2000

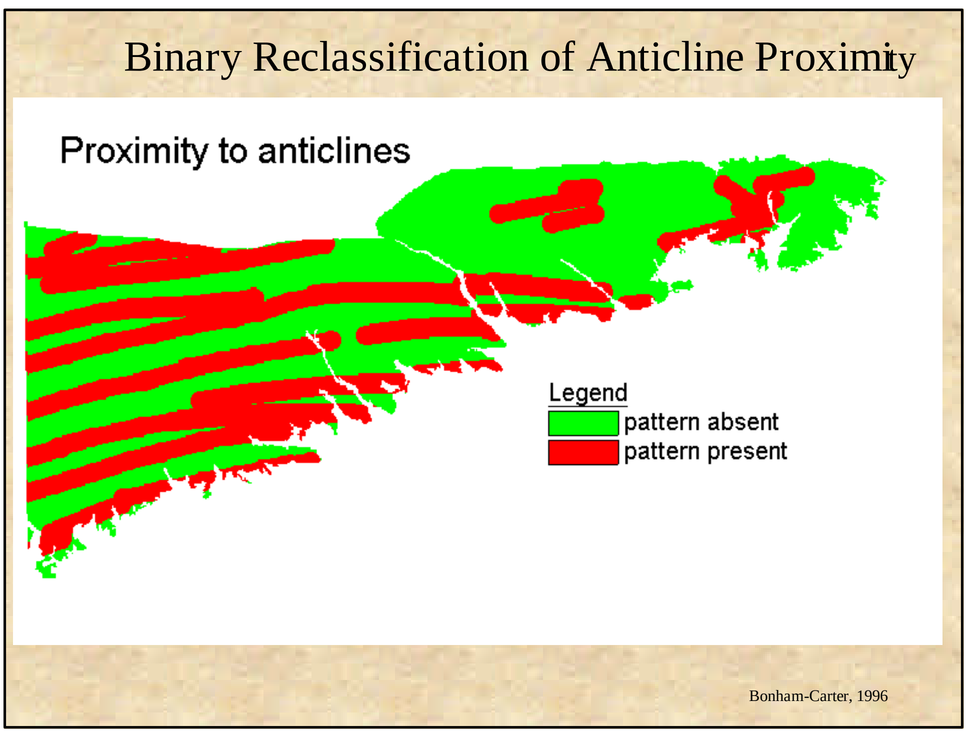
Proximity to Anticlines



Version 1, January 2000



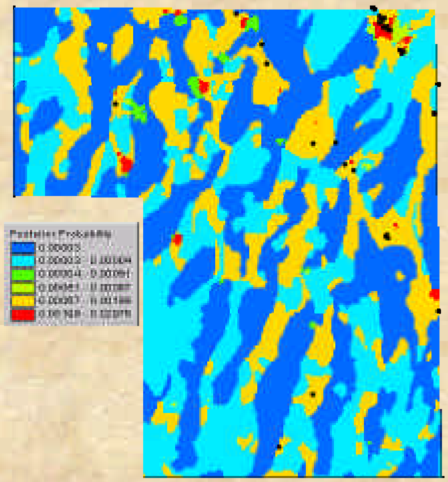
Contrast
C
1.0
0.8
0.4
0.0
1
2
3
4
5
6
DISTANCE, km
Maximum contrast at
1.25 km.
Version 1, January 2000



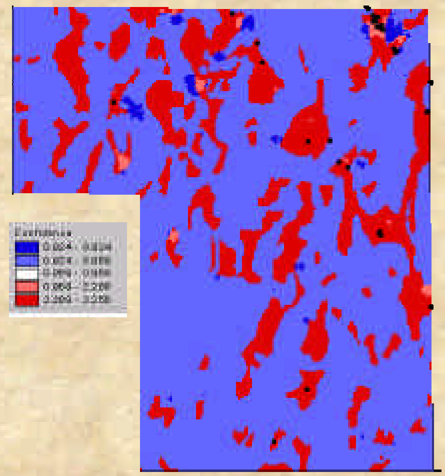
Binary Reclassification of Anticline Proximity
Proximity to anticlines
Legend
pattern absent
pattern present
Bonham-Carter, 1996

Variance and Uncertainty

Posterior Probability

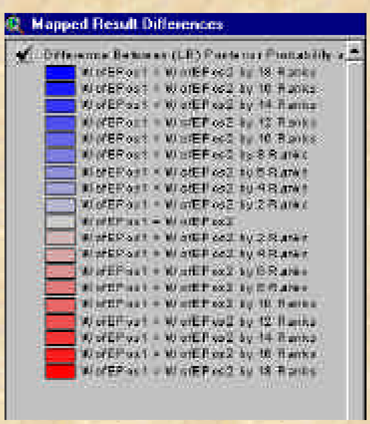


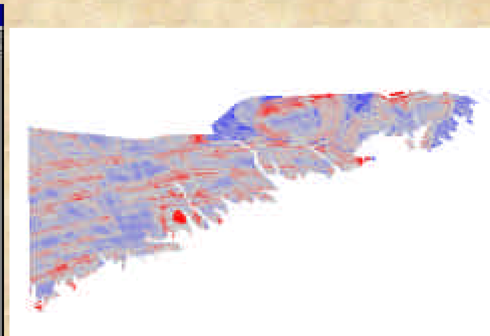
Studentized Posterior Probability



Compare Results

Map of Rank Differences

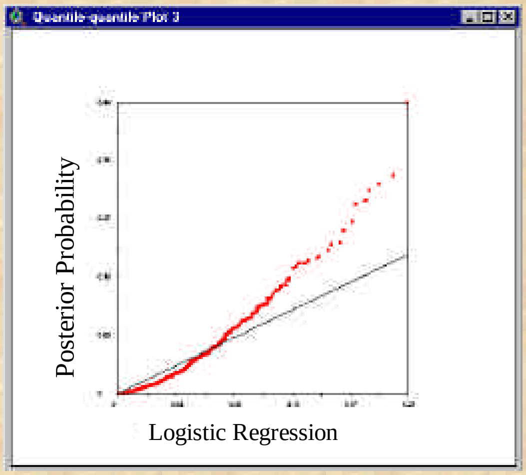




Bonham-Carter, 1999

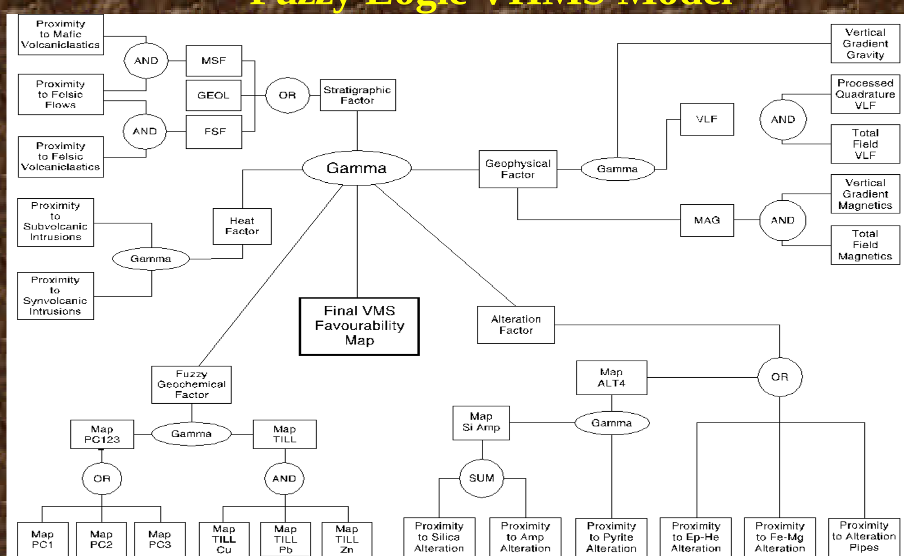
Compare Results

Quantile-quantile Plot



Bonham-Carter, 1999

Fuzzy Logic VHMS Model



From Wright, 1996

Laboratory Exercise

- Carl Hasty/David Atkins – TRPA (March 8)
  - Possible training sets for modeling
- Form groups (March 13)
- Define problem and data (March 27)
- Model (March 27-April 30)
- Presentations (May 1)

Version 1, January 2000

Tools for Map Analysis
Multiple Maps

Boolean Logic

Index Overlay

Fuzzy Logic

Weights of Evidence

Logistic Regression

Neural Networks

Boolean Operators

- And - Returns True (= 1) only if all are true
  - Logical intersection
- Or - Returns False (= 0) if all are false, otherwise returns True (= 1)
  - Logical union
- Xor - Returns True (1) if one and only one is true.
- Not - Negates the operation

Examples

- 1 and 1 = 1
- 1 and 0 = 0
- 0 and 0 = 0
- 1 or 1 = 1
- 1 xor 1 = 0
- 1 or 0 = 1
- 1 xor 0 = 1
- 0 or 0 = 0
- 0 xor 0 = 0
- 0 or (not 0) = 1

- For Boolean operators, an input of zero (0) equals False.
- Any other number is True.
- -3 and 2 = 1
- 2 and 0 = 0
- -3 and 2 and 12 = 1
- -3 or 2 or 12 = 1
- -3 or 2 or 0 = 1
- -3 xor 0 xor 0 = 1

Version 1, January 2000

Landsite Selection Statement of the Problem

1. Be in an area where unconsolidated surficial material is more than a minimum thickness, AND
2. Be in material that has a low permeability, AND

etc.

Example on page 272 of text.

Version 1, January 2000

Boolean Map Algebraic Statement of the Problem

: At current location, determine if conditions for each input ar e satisfied
: The conditions, C1 to C2 are either TRUE (=1) or FALSE (=0)
: See Table 9-5 for a summary of the map classes

C1 = class('OVERTHIK')>4
etc.
C10 = class('ECOLOG') = = 1

:Combine conditions with Boolean "AND" operator
: The variable OUTPUT is either TRUE (=1) or False (=0)

OUTPUT = C1 AND C2 AND … AND C10

: Map results as a binary 2-class map

RESULTS(OUTPUT)

Portion of calculation on page 273.

Version 1, January 2000

Translate class into Arcview

- C1 = class('OVERTHIK') > 4
  - OVERTHIK is an integer grid
  - Returns TRUE (= 1) if OVERTHIK > 4; otherwise returns FALSE (= 0)
  - Arcview:
    - Analysis/Map Query ([OVERTHIK] > 4.AsGrid)
    - Analysis/Map Calculator ([OVERTHIK] > 4.AsGrid).Con(1.AsGrid,0.AsGrid)

Version 1, January 2000

Inference Net for Landfill Site Boolean Logic

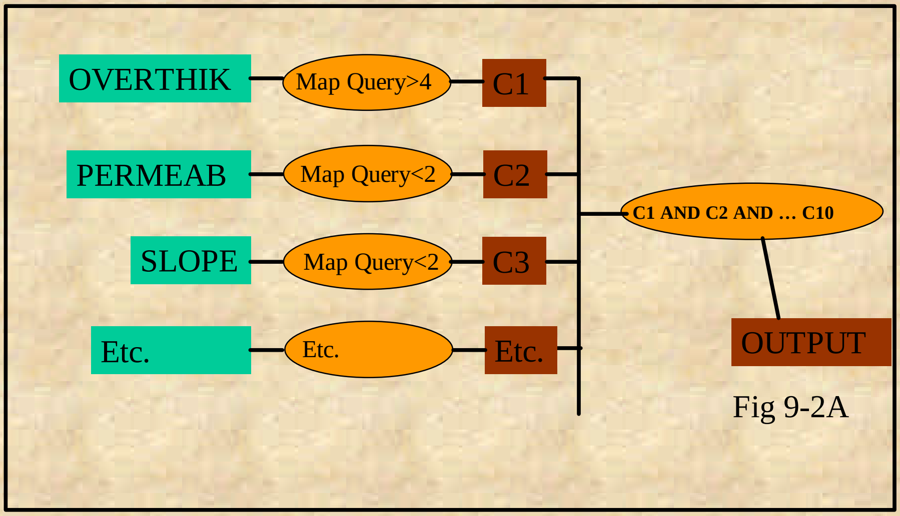
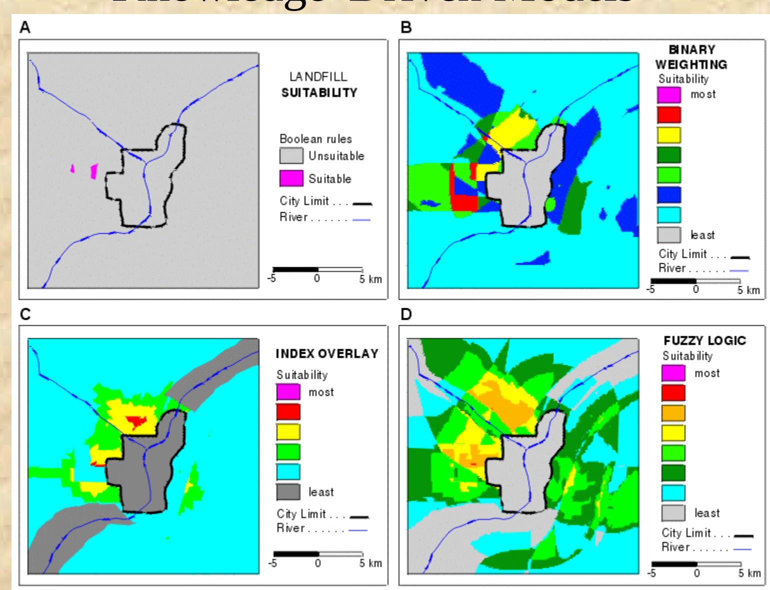


Fig 9-2A

Version 1, January 2000

Knowledge-Driven Models



A: Boolean B: Binary Index Overlay C: Multi-class Index Overlay D: Fuzzy logic

Bonham-Carter, 1996, Fig.9-2

Decisions for Boolean Logic

Reclassify Attributes and Map Interactions

- Thresholds
  - Greater than some value
    - Distance from some feature
    - Some high measured value (e.g. slope > 20)
  - Less than some value
    - Some measured low value (e.g. thickness < 4)
- Equal or Not Equal to some named class
- How the criteria (maps) interact
  - AND, OR, XOR, NOT

Version 1, January 2000

Boolean Logic Summary

- Advantages
  - Models are simple.
  - Where prescriptive guidelines from law, Boolean combinations are practical and easily applied.
- Disadvantages
  - All evidence (Maps) are treated equally. A weak representation of how people think about spatial problems
  - Output is binary, either Suitable or Not Suitable.

Version 1, January 2000

Index Overlay

$$Score = \frac{\sum_{i=1}^{n} w_i * s_{ij}}{\sum_{i=1}^{n} w_i}$$

Where
w_i = weight of Map I
For binary-class maps, s_{ij} is either 1 for true or present or 0 for false or absent
For multi-class maps, s_{ij} is the score or weight assigned to a particular attribute.
Score will vary between 0, extremely unfavorable, and 1, extremely favorable.

Version 1, January 2000

Index Overlay Algebraic Statement of the Problem

: Calculate normalization sum
SUMW = 3 + 4 + 5 + 3 + 2 + 4 + 5 + 4 + 2 + 1
: Define a variable to name the row
ROW = class('BASIN')
: For current location, determine map weights
M1 = 3 * (class('GEOL') = = 1 OR class('GEOL') = = 2)
M2 = 4 * table('BASIN', ROW,'AS') > 30
M3 = 5 * table('BASIN', ROW, 'SB') > 0.8
etc.
: Calculate normalized sum of weight factors
NEW = (M1 + M2 + M3 … + M10)/SUMW
: Classify and map output
NEWMAP = CLASSIFY(NEW,'BINWT')
RESULTS(OUTPUT)

Portion of calculation on page 287, Mineral model.

Version 1, January 2000

Translate table into Arcview

- M1 = 4 * table('BASIN', ROW,'AS') > 30
  - Basin is an integer grid with multiple attributes
  - Returns TRUE (= 4) if AS > 30; otherwise returns False (= 0)
  - Arcview
    - Analysis/Map Query ([BASIN.AS] > 30.AsGrid)*4.AsGrid
    - Returns 4 if TRUE and 0 if FALSE, but will be labeled TRUE(1) and FALSE (0), respectively.
    - If had a real or float grid, that is only one attribute (Value), can use the same procedure in Arcview. If want an integer result, may ha ve to appropriately use .int in the equation.
    - Can also use the longer form in the Map Calculator of the Boolea n example (con statement).

From Mineral Model page 287.

Version 1, January 2000

Translate table into Arcview

- M1 = 4 * table('BASIN',class('BASIN'),'Score')
  - Basin is an integer grid with multiple attributes
  - Returns 4 * Score
  - Arcview
    - Analysis/Map Calculator ([BASIN.Score]*4.AsGrid)
    - If Score is a real (floating) number, returns a floating grid
    - If Score is an integer number, returns an integer grid
    - If multiplier is a real number, returns a real grid
    - Use .int and appropriate round off to convert real -valued grids to integer grids in the equation.

Version 1, January 2000

Inference Net for Landfill Site Binary Index Overlay

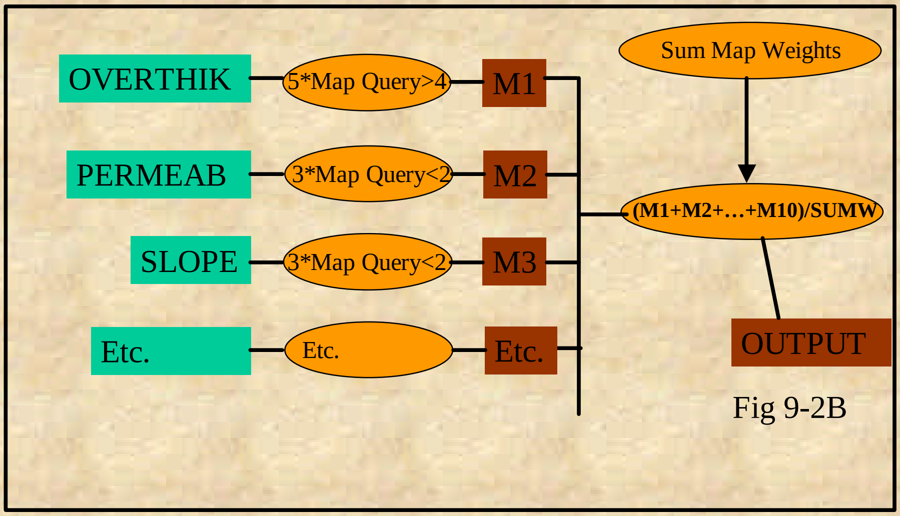


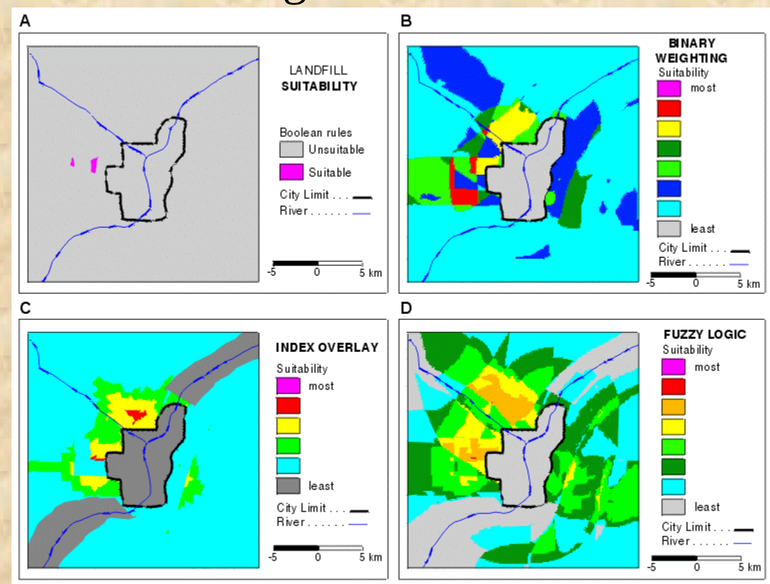
Fig 9-2B

Decisions for Index Overlay

Weights for Attributes and Maps

- Thresholds
  - Greater than some value
    - Distance from some feature
    - Some high measured value (e.g. slope > 20)
  - Less than some value
    - Some measured low value (e.g. thickness < 4)
- Equal or Not Equal to some named class
- How the criteria (maps) interact
  - Weight individual maps. What is the value of each criteria (map)?
  - Summation

Knowledge-Driven Models



A: Boolean B: Binary Index Overlay C: Multi-class Index Overlay D: Fuzzy logic

Bonham-Carter, 1996, Fig.9-2

Index Overlay Summary

- Advantages
  - Weights for individual maps and attribute values allows for better representation of experts opinion of the data.
  - By adjusting weights of maps and attributes can evaluate many different scenarios.
  - Output is a ranking of suitability, which gives decision makers more flexibility.
  - Scaling of Output is by reclassification, an expert decision.
- Disadvantages
  - Linear additive nature is greatest disadvantage.

Version 1, January 2000

Model Complexity

- Boolean Logic does binary, logical reclassification of evidential layers (maps).
- Binary Index Overlay adds relative weighting of evidential layers (maps).
- Multi-Class Index Overlay adds relative weighting of an attribute or attributes of each evidential layer (map).

USGS

Version 1, January 2000

Multiple Maps
Fuzzy Logic

Modified from Graeme Bonham-Carter

Bonham-Carter, Oct. 1999

Version 1, January 2000

OUTLINE

- Crisp vs. fuzzy logic
- Fuzzy membership functions
- Fuzzy combination operators
- Application

Bonham-Carter, Oct. 1999

Version 1, January 2000

Crisp Logic

- Membership of crisp set defined as either 1 or 0, True or False
  - (1) Truth(This location is close to a lineament) = 1
  - (2) Truth(This location is on a geochemical anomaly)= 0
- Combination of (1) and (2) by AND, OR, NOT Boolean operators.
  - Truth(1 AND 2) = 0
  - Truth(1 OR 2) = 1

Bonham-Carter, Oct. 1999

Version 1, January 2000

Fuzzy logic

- Fuzzy membership defined in the range [0,1], allowing for gradational membership
  - (1) Truth(This location is close to a lineament) = 0.6
  - (2) Truth(This location is on a soil geochemical anomaly) = 0.9
- Fuzzy operators
  - fuzzy AND, fuzzy OR, fuzzy NOT, fuzzy algebraic SUM, fuzzy algebraic PRODUCT, fuzzy GAMMA, etc

Bonham-Carter, Oct. 1999

Version 1, January 2000

Fuzzy Membership Functions

- Membership defined by a functional relationship, or by a table of ordered pairs
- Membership reflects degree of truth of some proposition or hypothesis (often a linguistic statement)

Non-spatial example

- Truth of proposition (Person X is Tall)
- Degree of tallness depends on height
- Need a fuzzy membership function relating height to degree of tallness
- In range [0,1], similar to probability, but not satisfying probability laws
- Sometimes termed “possibility”

Person	Height	Tallness
Fred	3'2"	0.00
Mike	5'5"	0.21
Sally	5'9"	0.38
Marg	5'10"	0.42
John	6'1"	0.54
Sue	7'2"	1.00

Tallness = 0 if height < 5',
Tallness = (height-5)/2;
if 5 < =height<=7'; or
Tallness = 1 if height > 7'

Truth(Marg is tall) = 0.42

Bonham-Carter, Oct. 1999

Version 1, January 2000

Person	Age	Oldness
Sally	27	0.21
Mike	30	0.29
Marg	32	0.33
John	31	0.54
Sue	45	0.64
Fred	65	1.00

Oldness = 0 if age < 18;
Oldness = (age-18)/42
if 18 <= age <= 60; or
Oldness = 1 if age > 60

Truth(Fred is old)=1.00

Bonham-Carter, Oct. 1999

Version 1, January 2000

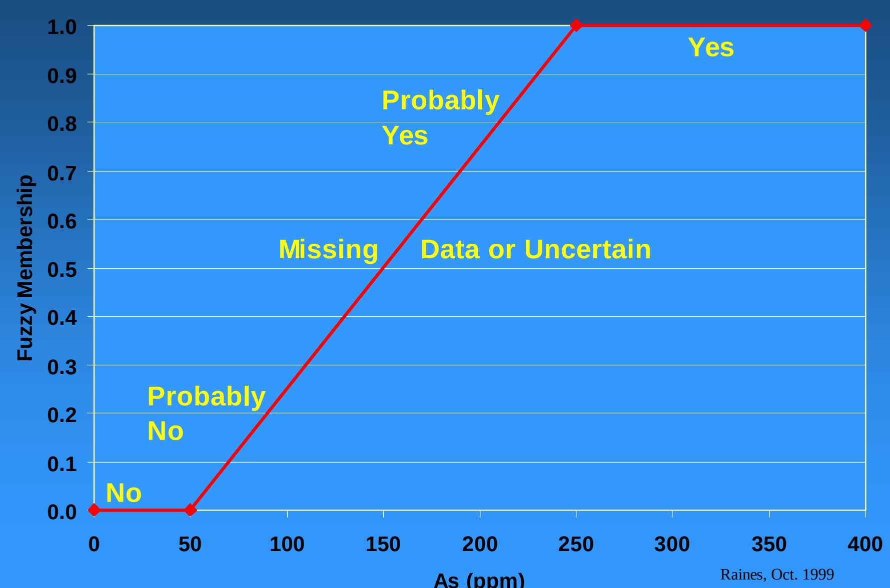
Person	Height	Tallness	Age	Oldness	Tall and old	Tall or old
Fred	3'2"	0.00	65	1.00	0.00	1.00
Mike	5'5"	0.21	30	0.29	0.21	0.29
Sally	5'9"	0.38	27	0.21	0.21	0.38
Marg	5'10"	0.42	32	0.33	0.33	0.42
John	6'1"	0.54	31	0.54	0.54	0.54
Sue	7'2"	1.00	45	0.64	0.64	1.00

Truth(Sally is tall AND old) = min(0.38, 0.21) = 0.21

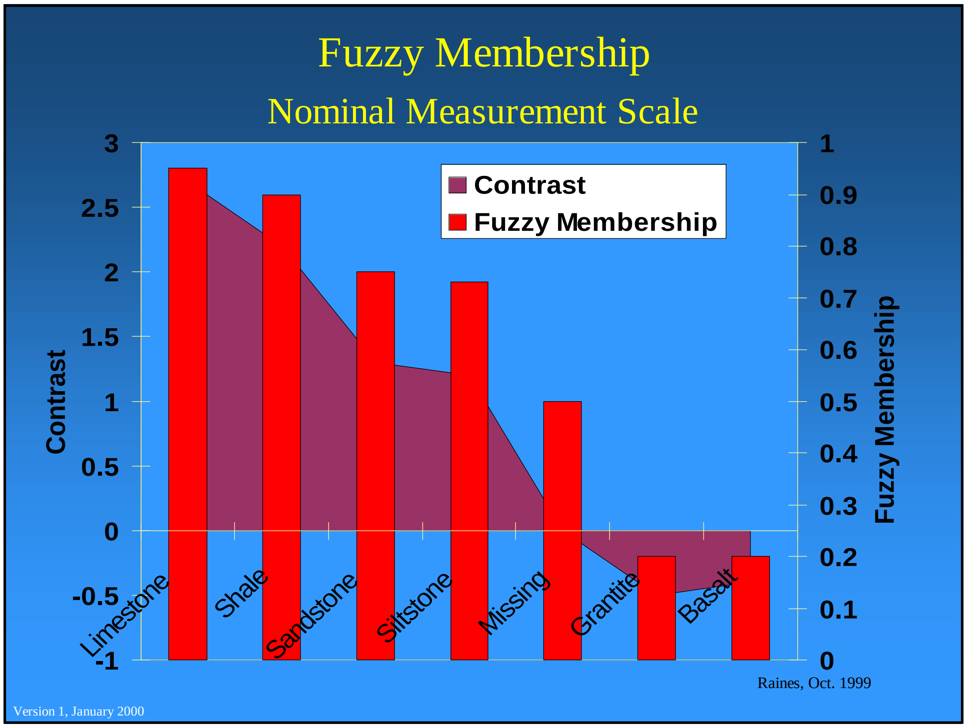
Truth(John is tall OR old) = max(0.54, 0.54) = 0.54

Bonham-Carter, Oct. 1999

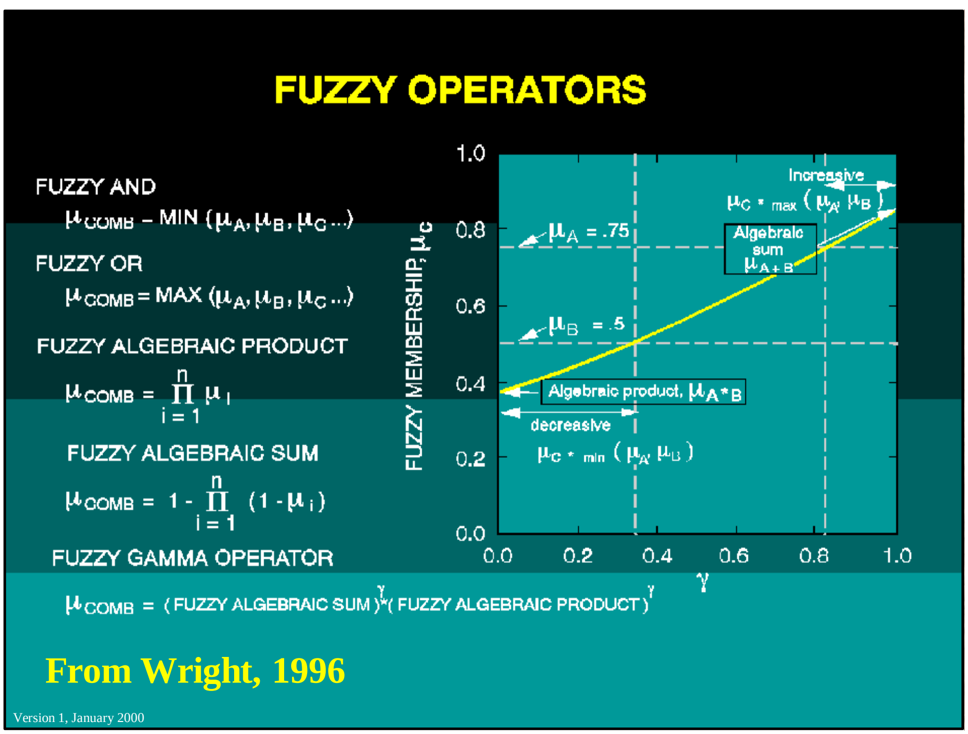
Functional Fuzzy Membership



Raines, Oct. 1999



Fuzzy Membership
Nominal Measurement Scale
Contrast
Fuzzy Membership
Limestone
Shale
Sandstone
Siltstone
Missing
Granite
Basalt
Raines, Oct. 1999
Version 1, January 2000



FUZZY OPERATORS
FUZZY AND
$\mu_{COMB} = MIN\ (\mu_A, \mu_B, \mu_C ...)$
FUZZY OR
$\mu_{COMB} = MAX\ (\mu_A, \mu_B, \mu_C ...)$
FUZZY ALGEBRAIC PRODUCT
$\mu_{COMB} = \prod_{i=1}^{n} \mu_i$
FUZZY ALGEBRAIC SUM
$\mu_{COMB} = 1 - \prod_{i=1}^{n} (1 - \mu_i)$
FUZZY GAMMA OPERATOR
$\mu_{COMB} = (\text{FUZZY ALGEBRAIC SUM})^{\gamma} * (\text{FUZZY ALGEBRAIC PRODUCT})^{\gamma}$
FUZZY MEMBERSHIP, μ_C
γ
Increasive
Algebraic sum μ_{A+B}
$\mu_A = .75$
$\mu_B = .5$
Algebraic product, μ_{A*B}
decreasive
From Wright, 1996
Version 1, January 2000

FUZZY OPERATORS

FUZZY AND

$\mu_{COMB} = MIN(\mu_A, \mu_B, \mu_C \ldots)$

FUZZY OR

$\mu_{COMB} = MAX(\mu_A, \mu_B, \mu_C \ldots)$

FUZZY ALGEBRAIC PRODUCT

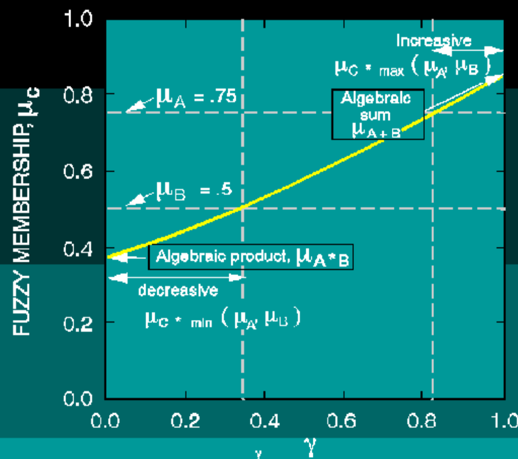
$$\mu_{COMB} = \prod_{i=1}^{n} \mu_i$$

FUZZY ALGEBRAIC SUM

$$\mu_{COMB} = 1 - \prod_{i=1}^{n} (1 - \mu_i)$$

FUZZY GAMMA OPERATOR

$$\mu_{COMB} = (\text{FUZZY ALGEBRAIC SUM})^{\gamma} * (\text{FUZZY ALGEBRAIC PRODUCT})^{1-\gamma}$$



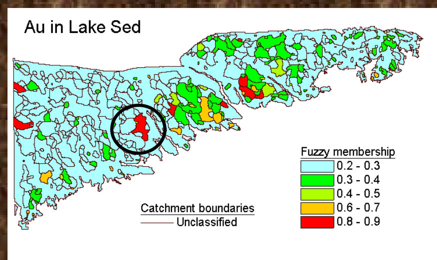
From Wright, 1996

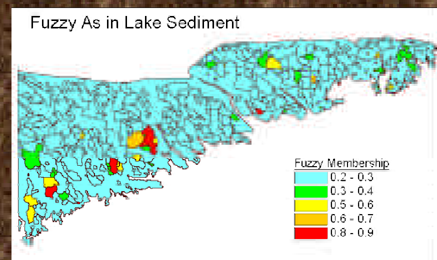
Fuzzy Membership Table for As in Lake Sediment

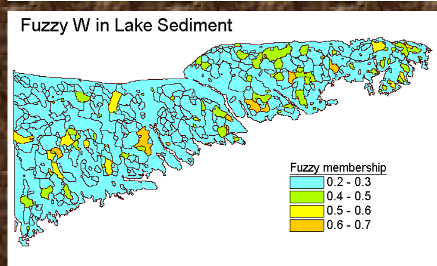
Class	Membership	Source Intervals
1	0.8	'142 - 166 ppm As'
2	0.7	'112 - 142 ppm As'
3	0.3	'28 - 52 ppm As'
5	0.2	'17 - 28 ppm As'
6	0.2	'12 - 17 ppm As'
7	0.2	'7 - 12 ppm As'
8	0.2	'2-7 ppm As'
9	0.2	'No data'

Bonham-Carter, Oct. 1999

Comparison of Fuzzy Evidence



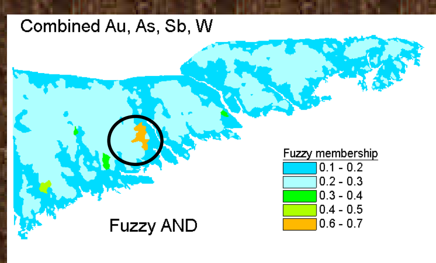


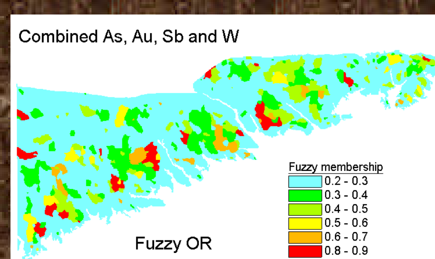


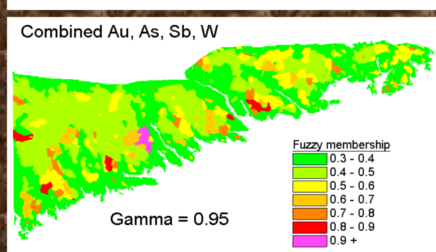
Bonham-Carter, Oct. 1999

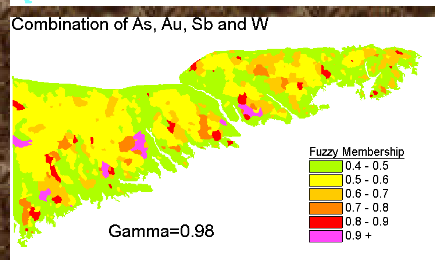
Version 1, January 2000

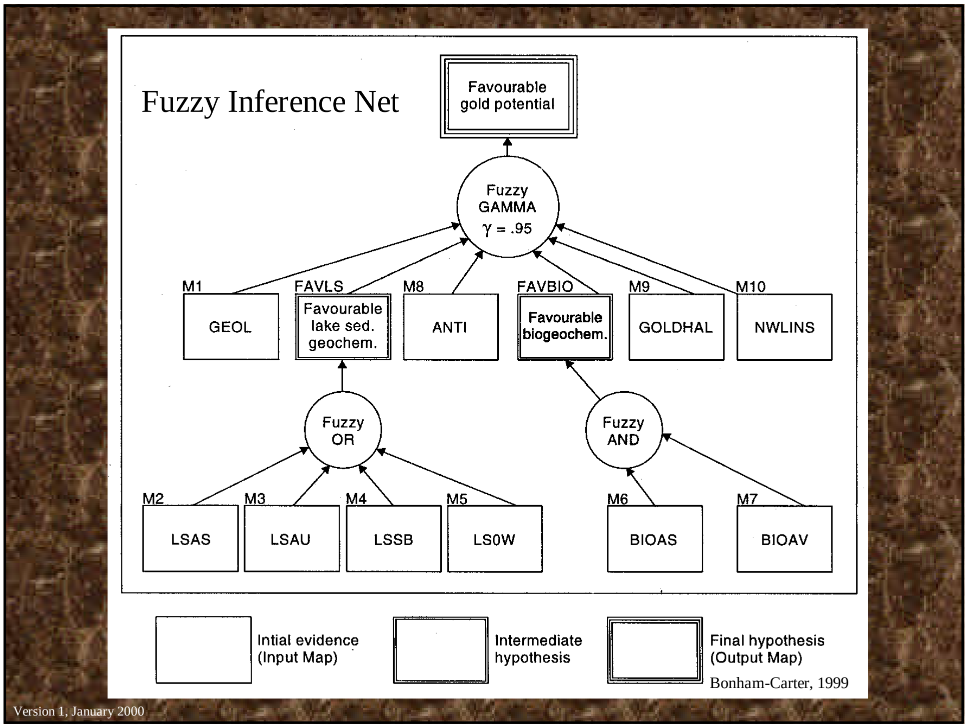
Operator Comparison



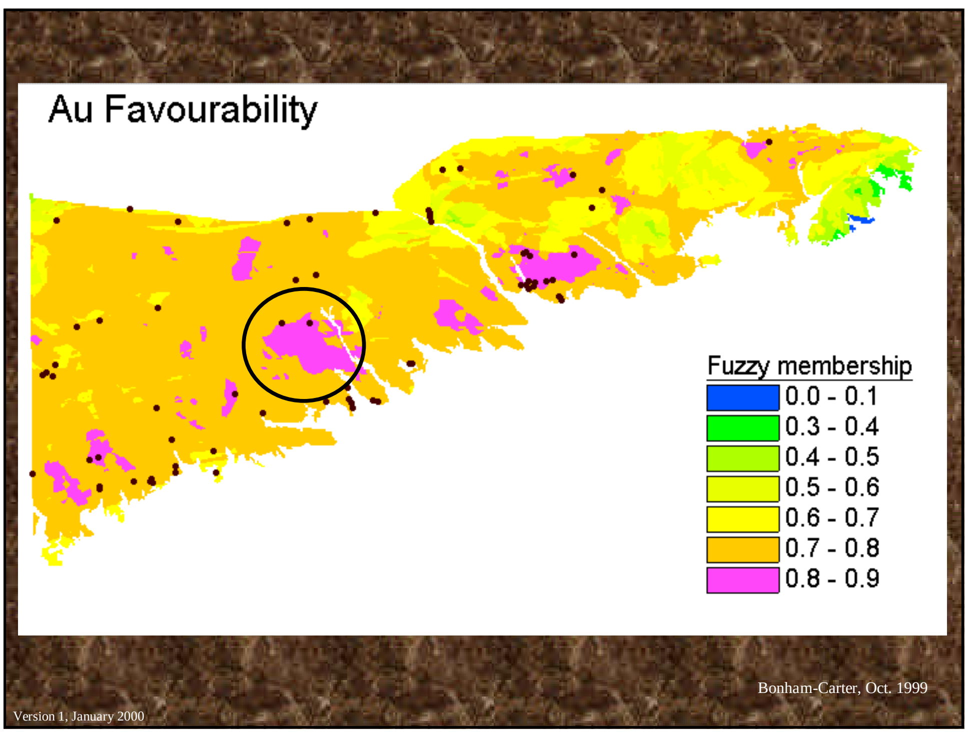






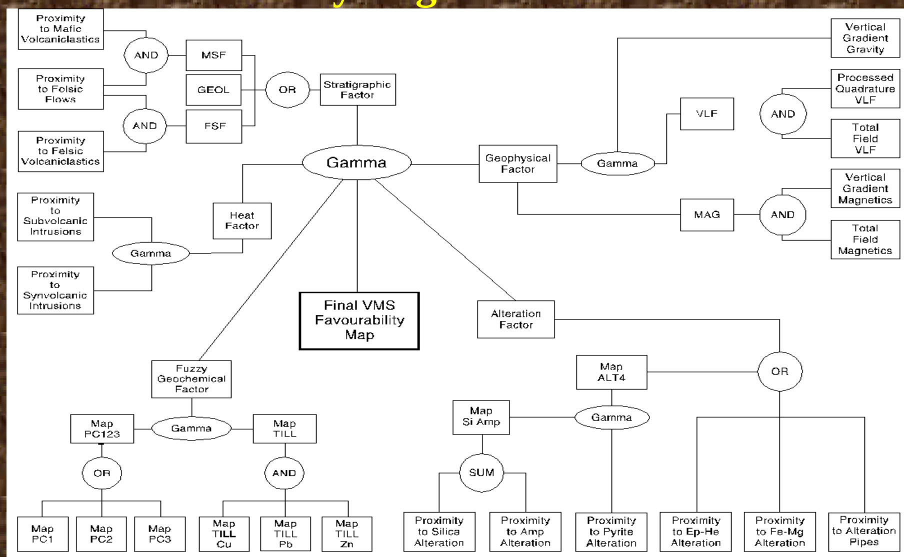


Fuzzy Inference Net
Favourable gold potential
Fuzzy GAMMA γ = .95
M1
FAVLS
M8
FAVBIO
M9
M10
GEOL
Favourable lake sed. geochem.
ANTI
Favourable biogeochem.
GOLDHAL
NWLINS
Fuzzy OR
Fuzzy AND
M2
M3
M4
M5
M6
M7
LSAS
LSAU
LSSB
LS0W
BIOAS
BIOAV
Intial evidence (Input Map)
Intermediate hypothesis
Final hypothesis (Output Map)
Bonham-Carter, 1999
Version 1, January 2000



Au Favourability
Fuzzy membership
0.0 - 0.1
0.3 - 0.4
0.4 - 0.5
0.5 - 0.6
0.6 - 0.7
0.7 - 0.8
0.8 - 0.9
Bonham-Carter, Oct. 1999
Version 1, January 2000

Fuzzy Logic VHMS Model



From Wright, 1996

Decisions for Fuzzy Logic

- Fuzzy Memberships
  - Thresholds can be gradational, potentially many values to assign
  - Named classes can be fuzzy, potentially a value for each class
- How the criteria (maps) interact
  - Fuzzy AND, OR, and GAMMA
  - Fuzzy SUM and PRODUCT - not used often
  - Gamma value to define fuzzy relationships of criteria

Fuzzy Logic Summary

- Advantages
  - Flexibility of assigning fuzzy memberships
  - Choice of combination operators
  - Mimic decision making by expert
  - Can deal with "maybe"
  - Not limited to binary criteria
  - Easy to understand
- Disadvantages
  - Problem of missing data
  - Confusion between fuzzy membership and uncertainty
  - Potentially many fuzzy membership values to assign

Version 1, January 2000

Modified from Bonham-Carter, Oct. 1999; Wright, 1996

Tools for Map Analysis

Multiple Maps

Boolean Logic

Index Overlay

Fuzzy Logic

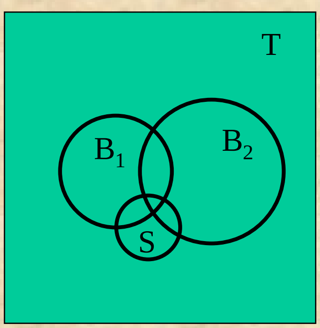
Weights of Evidence (Part 1)

Logistic Regression

Neural Networks

Version 1, January 2000

Numerical Example



Venn diagram of point and grid intersections, not draw to scale.

$N\{T\} = 10{,}000$	$W^+_1 = 0.9474$
$N\{B_1\} = 3600$	$W^-_1 = -1.8734$
$N\{B_2\} = 5000$	$W^+_2 = 0.3447$
$N\{S\} = 200$	$W^-_2 = -0.5189$
$N\{B_1 \cap S\} = 180$	$C_1 = 2.8208$
$N\{B_2 \cap S\} = 140$	$C_2 = 0.8636$

$C_1 > C_2$:Therefore Pattern B_1 is a better predictor that Pattern B_2!

Fig. 9-9

Version 1, January 2000

Weights

- Define the area to be studied
  - Count its area in unit cells = N{Study Area}
- Count the number of training sites in the study area = N{Training Sites}= N{S}
- Count the area of the pattern B = N{B}
- Prior probability = P{S} = N{Training Sites}/N{Study Area}
- Conditional Probability: Posterior Probability of a training site given the presence of a binary pattern B and absence of B.

$$P\{S|B\}=\frac{P\{S\cap B\}}{P\{B\}}=\frac{N\{S\cap B\}}{N\{B\}}=P\{S\}*\frac{P\{B|S\}}{P\{B\}}$$

$$P\{S|\bar{B}\}=P\{S\}*\frac{P\{\bar{B}|S\}}{P\{\bar{B}\}}$$

Version 1, January 2000

Odds Formulation

$$P\{S\,|\,B\}=\frac{P\{S\cap B\}}{P\{B\}}=\frac{N\{S\cap B\}}{N\{B\}}=P\{S\}*\frac{P\{B\,|\,S\}}{P\{B\}}$$

$$O\{S\,|\,B\}=O\{S\}\frac{P\{B\,|\,S\}}{P\{B\,|\,\bar{S}\}}$$

$$\ln O\{S\,|\,B\}=\ln O\{S\}+\ln\left\{\frac{P\{B\,|\,S\}}{P\{B\,|\,\bar{S}\}}\right\}$$

$$\mathrm{logit}\{S\,|\,B\}=\mathrm{logit}\{S\}+\mathrm{logit}\left\{\frac{P\{B\,|\,S\}}{P\{B\,|\,\bar{S}\}}\right\}=\mathrm{logit}\{S\}+W^{+}$$

$$P\{S\,|\,\bar{B}\}=P\{S\}*\frac{P\{\bar{B}\,|\,S\}}{P\{\bar{B}\}}$$

$$\mathrm{logit}\{S\,|\,\bar{B}\}=\mathrm{logit}\{S\}+\mathrm{logit}\left\{\frac{P\{\bar{B}\,|\,S\}}{P\{\bar{B}\,|\,\bar{S}\}}\right\}=\mathrm{logit}\{S\}+W^{-}$$

Version 1, January 2000

Weights Calculation Formula

$$W^{+} = \ln\left[\frac{T_{11} * T_{2\bullet}}{T_{21} * T_{1\bullet}}\right] \quad \text{eq. } 8\text{-}20$$

$$W^{-} = \ln\left[\frac{T_{12} * T_{2\bullet}}{T_{22} * T_{1\bullet}}\right] \quad \text{eq. } 8\text{-}21$$

<u>Binary Patterns!</u>

Bayes' Theorem

P{Rain|Time-of-Year} = P{Rain} * Time-of-Year Factor

P{Rain|Evidence} = P{Rain} * Evidence1 * Evidence 2 etc.

P{Rain} = <u>Prior Probability</u>, the probability before considering the evidence

P{Rain|Evidence} = <u>Posterior Probability</u>, the probability after considering the evidence.

- The evidence can increase or decrease the prior probability
- Applied to maps, the evidence is a pattern!

Bayes' Theorem and Training Sites

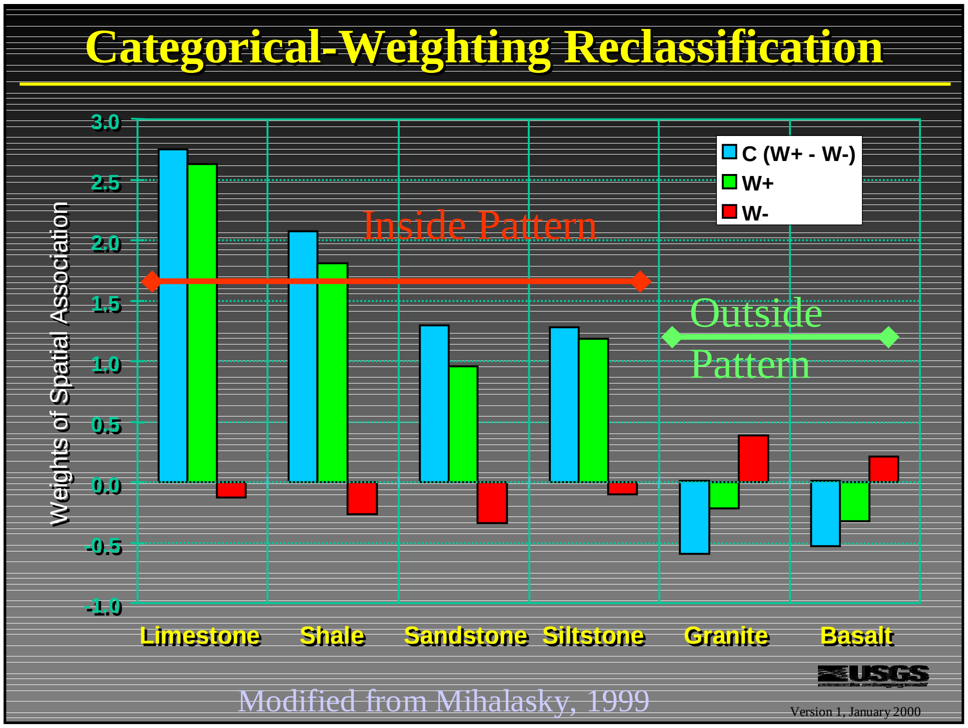
- Used here to predict the presence or absence of a set of point objects.
  - Points objects used include mineral deposits, animal habitat, human disease, etc.
  - Points represent a small unit of area, the unit cell, relative to the area studied and the resolution of the evidence.
  - Points are the training sites.
  - Assumes one training site per unit cell.
- Assumes conditional independence of evidence with regards to training sites.

Version 1, January 2000

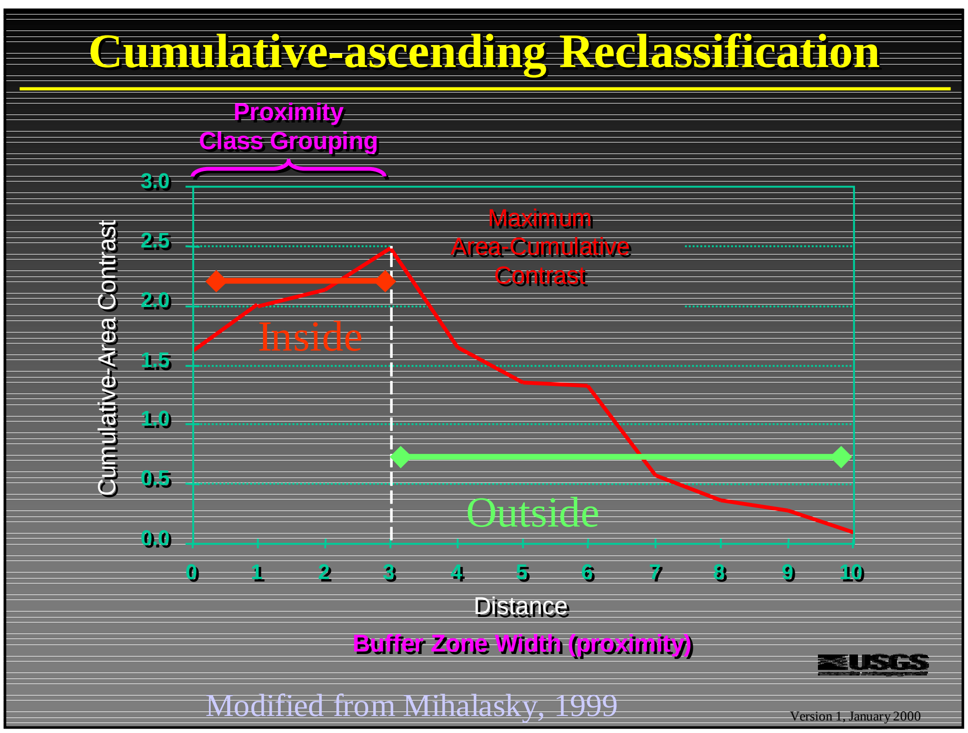
Logit Form of Baye's Theorem

- This allows for summation of the weights for all patterns as opposed to products
- W^+ is weight for inside the pattern, B
- W^- is weight for outside the pattern, not B
- Positive W^+ and negative W^- indicates a positive correlation between training sites and the pattern
- Contrast = W^+ - W^-
  - Relative measure of correlation - larger the contrast the greater the correlation
  - Can use contrast to help define best pattern!

Version 1, January 2000

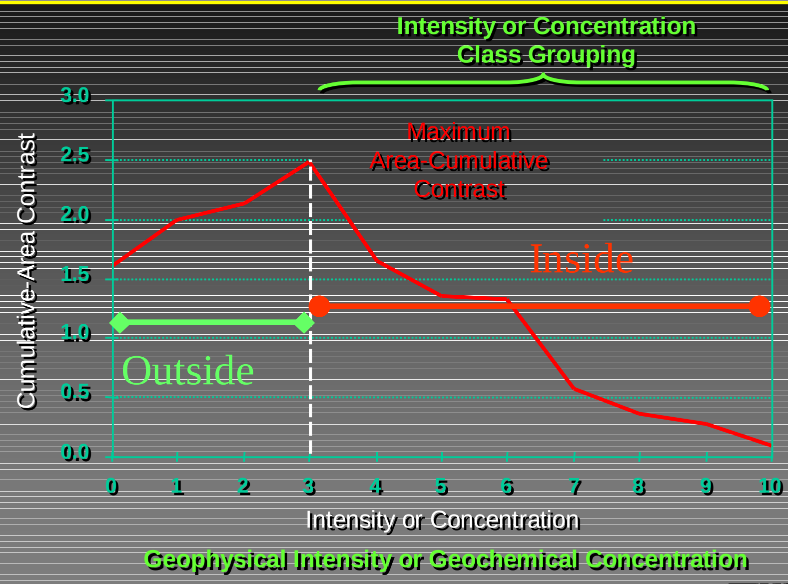


Categorical-Weighting Reclassification
Weights of Spatial Association
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
C (W+ - W-)
W+
W-
Inside Pattern
Outside Pattern
Limestone
Shale
Sandstone
Siltstone
Granite
Basalt
USGS
Modified from Mihalasky, 1999
Version 1, January 2000



Cumulative-ascending Reclassification
Proximity Class Grouping
Cumulative-Area Contrast
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Maximum Area-Cumulative Contrast
Inside
Outside
0
1
2
3
4
5
6
7
8
9
10
Distance
Buffer Zone Width (proximity)
USGS
Modified from Mihalasky, 1999
Version 1, January 2000

Cumulative-descending Reclassification



Modified from Mihalasky, 1999

Version 1, January 2000

Multiple Patterns = Multiple Weights

- Objective is to combine all the evidence to obtain a combined posterior probability.
  - Use Bayes' Theorem to combine patterns
  - Assumes conditional independence of patterns with regards to the training sites.

Conditional independence implies

$$P\{B_1 \cap B_2 \mid S\} = P\{B_1 \mid S\} * P\{B_2 \mid S\}$$

This allows

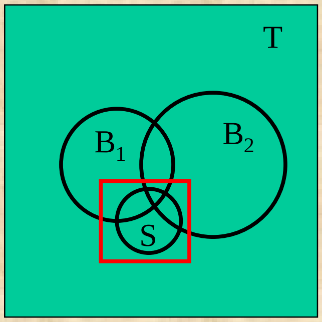
$$P\{S \mid B_1 \cap B_2\} = P\{S\} * \frac{P\{B_1 \mid S\}}{P\{B_1\}} * \frac{P\{B_2 \mid S\}}{P\{B_2\}}$$

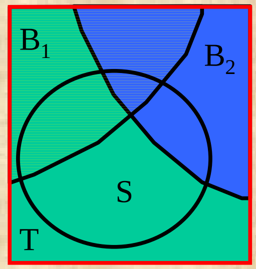
or

$$\text{logit}\{S \mid B_1 \cap B_2\} = \text{logit}\{S\} + W_1^+ + W_2^+$$

Version 1, January 2000

Conditional Independence





Conditional Independence is satisfied if :

$$\frac{N\{B_1 \cap S\}}{N\{S\}} * \frac{N\{B_2 \cap S\}}{N\{S\}} = \frac{\mathrm{N}\{\mathrm{B}_1 \cap \mathrm{B}_2 \cap S\}}{N\{S\}}$$

Using numbers from Fig 9 - 9

$$\frac{180}{200} * \frac{140}{200} = 0.63 \Rightarrow \text{for CI } \frac{\mathrm{N}\{\mathrm{B}_1 \cap \mathrm{B}_2 \cap S\}}{N\{S\}} = \frac{\boxed{126}}{200} = 0.63$$

What if there were three patterns?

Version 1, January 2000

Overall Test for Conditional Independence

$$N\{S_{Calc}\} = \sum_{k=1}^{m} P_k * (\text{unit cell})$$

where m = total number of unit cells.

$$\text{CI Ratio} = \frac{N\{S\}}{N\{S_{Calc}\}}$$

- **Unit cell is a constant in the grid implementation of Weights of Evidence.**
- **CI Ratio is typically less than 1.**
- **If CI Ratio is less than .90 to .85 then a serious CI problem has occurred.**

Version 1, January 2000

Pairwise CI Test

Chi-square statistic

Area Cross-Tabulation Table

Pattern	Training Sites		
Present	T_{11}	T_{12}	$T_{1.}$
Absent	T_{21}	T_{22}	$T_{2.}$
	$T_{\cdot 1}$	$T_{\cdot 2}$	$T_{..}$

$$T_{ij}^{*} = \frac{T_{i\bullet} * T_{\bullet j}}{T_{\bullet\bullet}}$$

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(T_{ij} - T_{ij}^{*}\right)^2}{T_{ij}^{*}}$$

Where
T_{ij}, where there are 2 rows and 2 columns.
Ti. is the sum of the i^{th} row,
T.j is the sum of the j^{th} column, and
T.. is grand sum over rows and columns.
T^{*}_{ij} is the expected value.

Identify pairs of patterns contributing to conditional dependency.

USGS

Version 1, January 2000

Solutions to CI Problems

- Combine pairs in some logical fashion and recalculate the model.
- If still have CI problem, cannot use the "posterior probability" as probability, a ratio measurement-scale number.
  – Treat the "posterior probability" as favorability, an ordinal measurement-scale number.
  – Call it favorability even though the software might label it posterior probability.
  – Limits what you can do with the model.

USGS

Version 1, January 2000

Variance of Weights and Contrast

$$s^2(W^+) = \frac{1}{N\{B \cap S\}} + \frac{1}{N\{B \cap \overline{S}\}}$$

$$s^2(W^-) = \frac{1}{N\{\overline{B} \cap S\}} + \frac{1}{N\{\overline{B} \cap \overline{S}\}}$$

$$s^2(\text{Contrast}) = s^2(W^+) + s^2(W^-)$$

Version 1, January 2000

Total Variance of Posterior Probability

$$s^2(P_{Posterior}) = \left[\frac{1}{N(S)} + \sum_{j=1}^{n} s^2(W_j^k)\right] * P_{Posterior}^2$$

where
k is + and - and
n is the number of patterns

$$s_j^2(\text{missing}) = \{P(S \mid B_j) - P(S)\}^2 * P(B_j) + \{P(S \mid B_j) - P(S)\}^2 * P(\overline{B}_j)$$

where j is a pattern with missing data

$$s^2(\text{total}) = s^2(P_{Posterior}) + \sum_{j=1}^{m} s_j^2(\text{missing})$$

Version 1, January 2000

Studentized Value

- Studentized Contrast = Contrast/s(C)
- Studentized Posterior Probability = Post. Prob./s(total Post. Prob.)
- An informal test of the hypothesis that value tested is zero. If studentized value greater than 1.5 to 2 then can assume that the value tested is not equal to zero.
- Use in a relative sense and to structure decision making.

USGS

Version 1, January 2000

Student T Values

Confidence	Test Value
99.5%	2.576
99%	2.326
97.5%	1.96
95%	1.645
90%	1.282
80%	0.842
70%	0.542
60%	0.253

Because Studentized test applied here is only approximate, use these values as a guide. If you can accept more risk, then you can use lower confidence values!

USGS

Version 1, January 2000

Decisions for Weights of Evidence

- Define the study area
- Define the training set
- Select confidence level for contrast
- Select the evidential maps
  - Use Contrast and Studentized Contrast to evaluate.
  - Binary Reclassification
  - Thresholds maximum, minimum, or grouping of nominal classes
- These decisions define objective, binary reclassification
  - Needed measurements: Area of study, Area of the pattern, Number of training sites, Number of training sites inside the pattern

USGS

Version 1, January 2000

Weights of Evidence

- Advantages
  - Objective assignment of weights
  - Multiple patterns combined simply
  - Binary reclassification to optimize contrast gives insights into spatial relationships
  - Deals with missing data
  - Measures aspects of uncertainty that can be mapped
- Disadvantages
  - Assumption of conditional independence
  - Requires a training set of sufficient size.

USGS

Version 1, January 2000

Tools for Map Analysis

Multiple Maps

Boolean Logic

Index Overlay

Fuzzy Logic

Weights of Evidence (Part 2)

Logistic Regression

Neural Networks

USGS

Version 1, January 2000

Weights of Evidence Method

- Originally developed as a medical diagnosis system
  - relationships between symptoms and disease evaluated from a large patient database
  - each symptom either present/absent
  - weight for present/weight for absent (W+/W-)
- Apply weighting scheme to new patient
  - add the weights together to get result

USGS

Version 1, January 2000

Weights of Evidence

- Data driven technique
  - Requires training sites
- Statistical calculations are used to derive the weights based upon training sites.
- Evidence (maps) are generally reclassified into binary patterns.

Weights of Evidence Terms

- Weights for patterns
  - W+ - weight for inside the pattern
  - W- - Weight for outside the pattern
  - 0 - Weights for areas of no data
- Contrast - a measure of the spatial association of pattern with sites
- Studentized Contrast - a measure of the significance of the contrast

Weights of Evidence

- Binary maps to define favorable areas
  - Can use multi-layer patterns
- Measurements
  - Area of study
  - Area of Pattern
  - Number of training sites
  - Number of training sites inside the pattern

USGS

Version 1, January 2000

Preprocessing

Nominal Measurement Scale

- For example - Geological map
  - select particular stratigraphic units or class
  - generalize by reclassification
  - extract and buffer boundaries between units

USGS

Version 1, January 2000

Preprocessing
Continuous Measurement Scale

- Histogram transformations
- Physical properties processing
- Filter
  - separate anomaly/background
- Spatial interpolation (e.g. surfaces, krige)
- Logical combinations (merging, boolean, fuzzy logic)
- Summarize by zonal statistics
  - separate anomaly/background
  - define a residual
  - multivariate analysis
    - principal components analysis and others

USGS

Version 1, January 2000

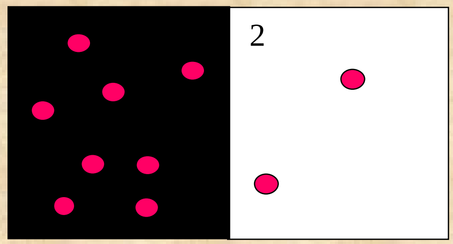
Overlay combination

- In vector
  - create polygon overlay and associated PAT
  - create unique conditions overlay and associated PAT
  - Topological selections
- In raster
  - superimpose grids

USGS

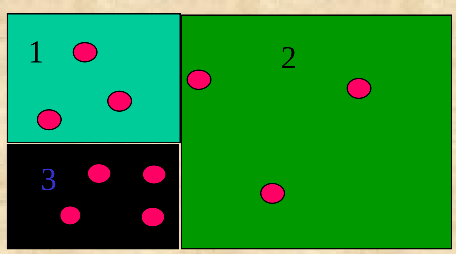
Version 1, January 2000

Application to Binary Map



Class	Area	#sites	Relative density	Weight
1	50	8	0.8/0.5=1.6	ln(1.6)= + 0.47
2	50	2	0.2/0.5=0.4	ln(0.4)= - 0.92
Total	100	10		

Application to Ternary Map



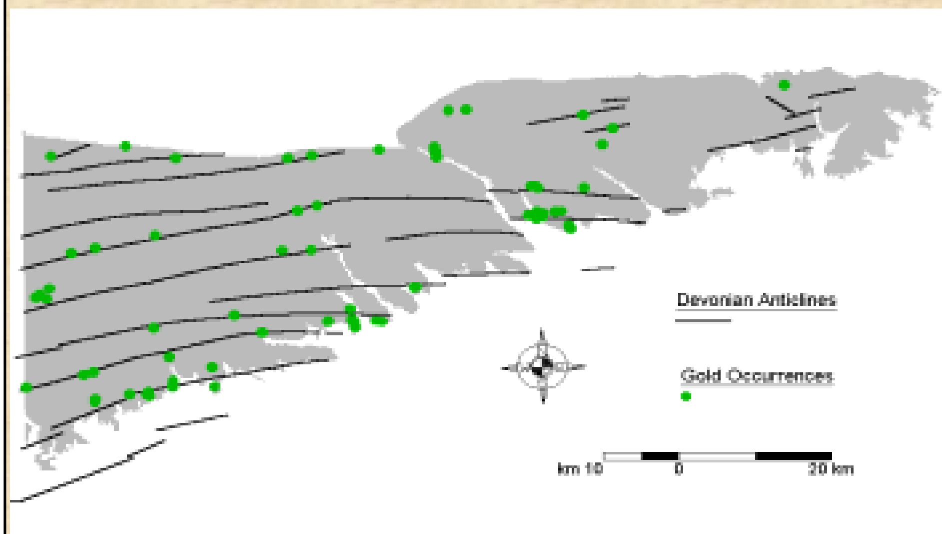
Class	Area	#sites	Relative density	Weight
1	25	3	0.3/0.25=1.2	ln(1.0)= 0.18
2	60	3	0.3/0.60=0.5	ln(0.5)= - 0.69
3	15	4	0.4/0.15=2.7	ln(2.7)= + 0.98
Total	100	10		

Expected Values of Weights

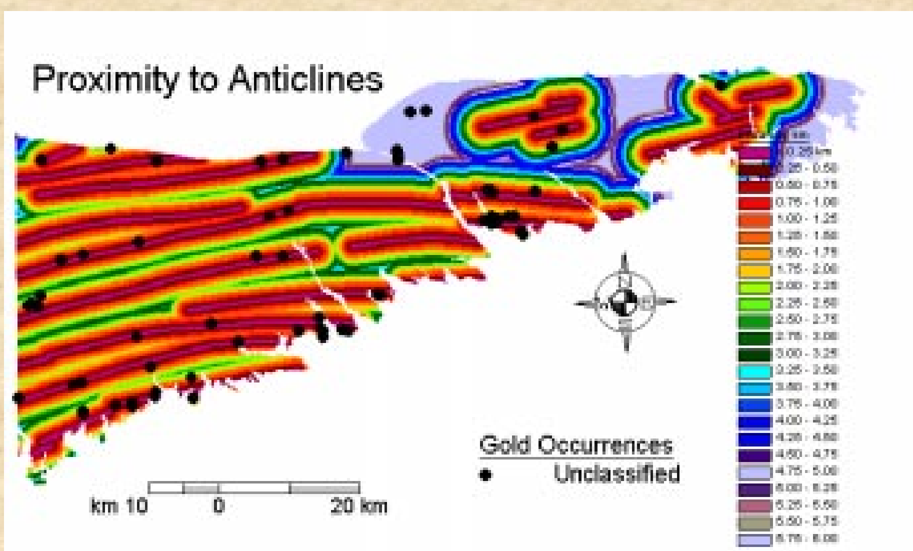
- If sites occur randomly,
  - Relative density (RD)=1.0
  - Weight (W) = ln(RD) =0.0
- If sites occur more frequently than chance
  - RD > 1.0, W is positive
- If sites occur less frequently than chance
  - RD < 0.0, W is negative

Weights Calculations

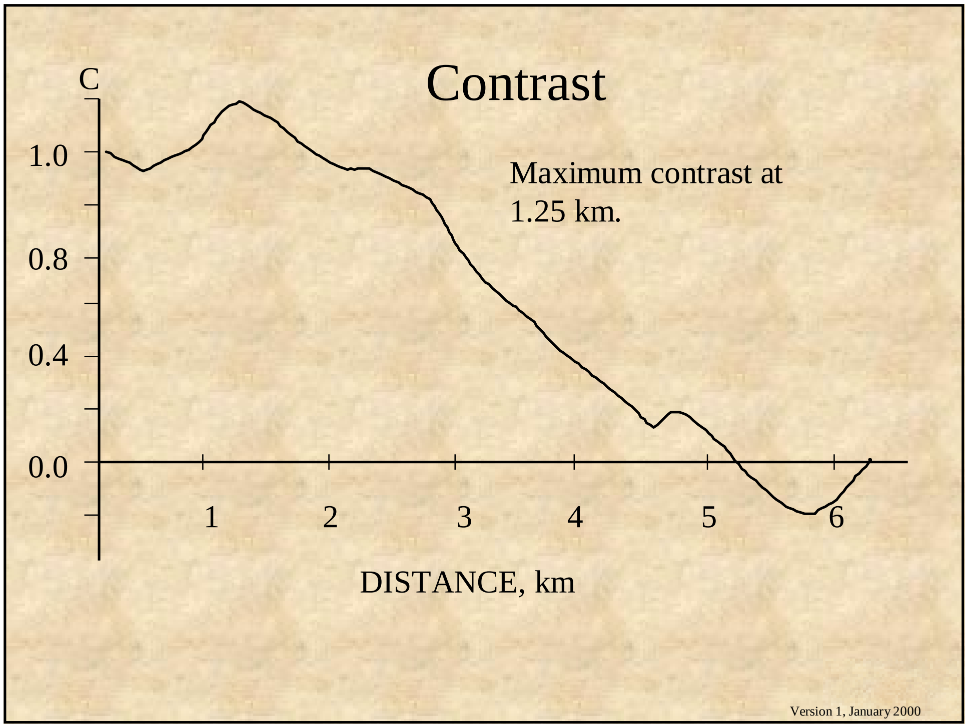
- Choose a small unit cell – affects the prior probability but only a little on the weights
- Can have multi-class maps but often not enough training points to get stable weights.
  - Use Studentized contrast to evaluate stability of weights.
- Contrast can be used to define optimal thresholds.
  - Use Studentized contrast to evaluate stability of contrast.

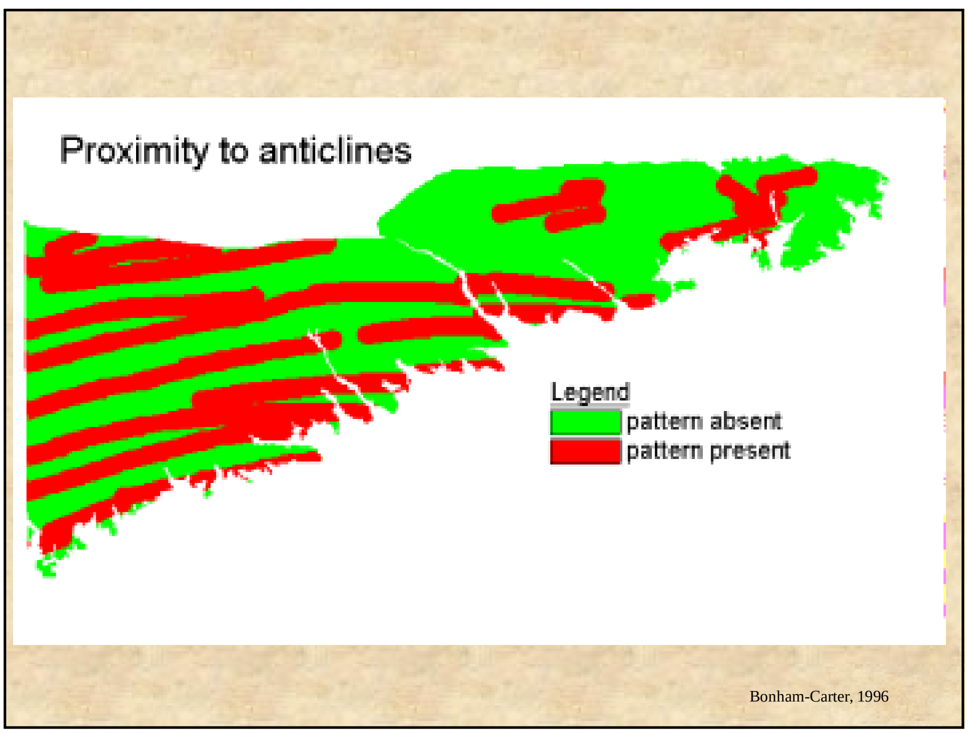


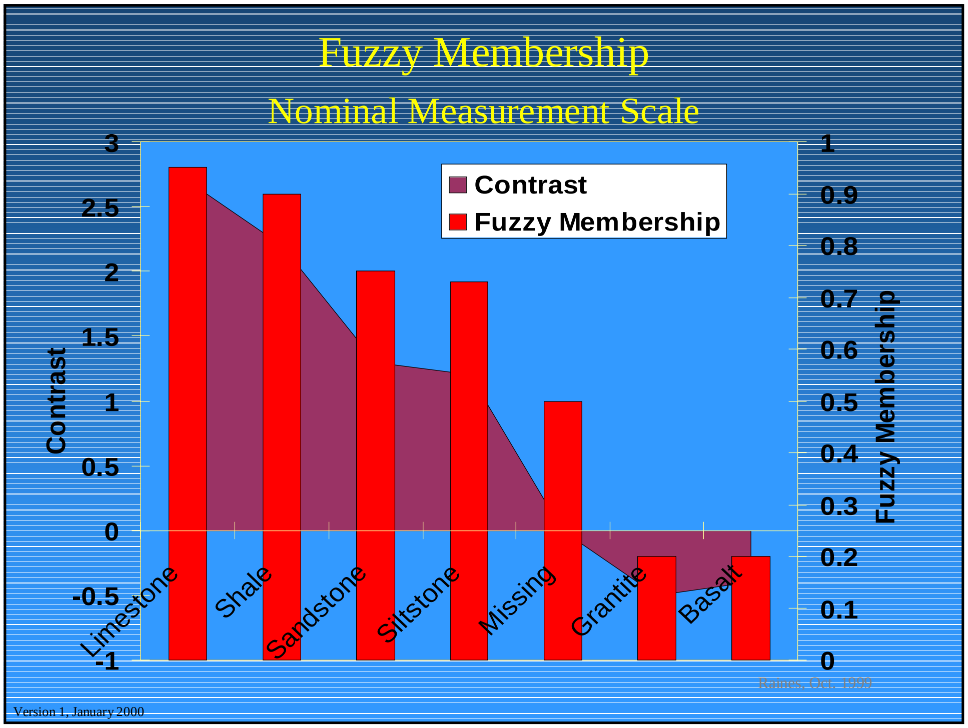
Version 1, January 2000



Version 1, January 2000


Contrast
C
1.0
0.8
0.4
0.0
1
2
3
4
5
6
Maximum contrast at
1.25 km.
DISTANCE, km
Version 1, January 2000


Proximity to anticlines
Legend
pattern absent
pattern present
Bonham-Carter, 1996



Handling Uncertainty

- Uncertainty due to weights – variance of weights.
- Uncertainty due to missing data – estimate of variance due to missing data
- Other measures of uncertainty?
- For Response Map can combine the various uncertainty measures to obtain a total variance.
- Studentized posterior probabilty (PP/s(PP) can provide a useful measure of confidence.

USGS

Version 1, January 2000

Decisions for Weights of Evidence

- Define the study area
- Define the training set
- Select confidence level for contrast
- Select the evidential maps
  - Use Contrast and Studentized Contrast to evaluate.
  - Binary Reclassification
  - Thresholds maximum, minimum, or grouping of nominal classes
- These decisions define objective, binary reclassification
  - Needed measurements: Area of study, Area of the pattern, Number of training sites, Number of training sites inside the pattern

USGS

Version 1, January 2000

Weights of Evidence

- Advantages
  - Objective assignment of weights
  - Multiple patterns combined simply
  - Binary reclassification to optimize contrast gives insights into spatial relationships
  - Deals with missing data
  - Measures aspects of uncertainty that can be mapped
- Disadvantages
  - Assumption of conditional independence
  - Requires a training set of sufficient size.

USGS

Version 1, January 2000

Miscellany

Fuzzy Membership

Nature of Evidence

Comparison of Results

Testing of Predictions

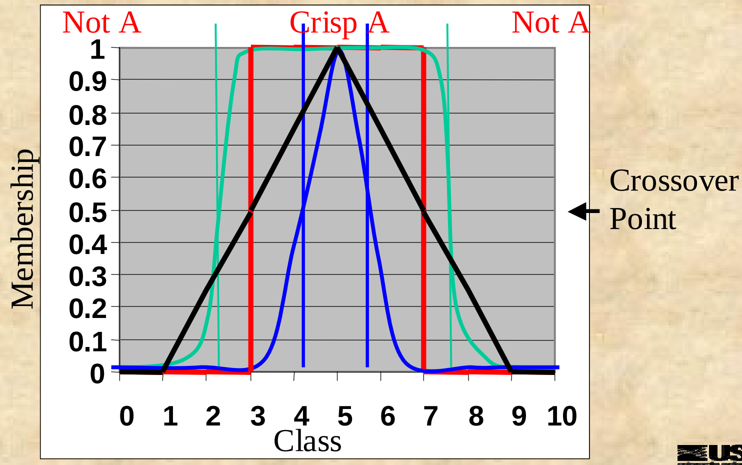
Version 1, January 2000

Fuzzy Membership

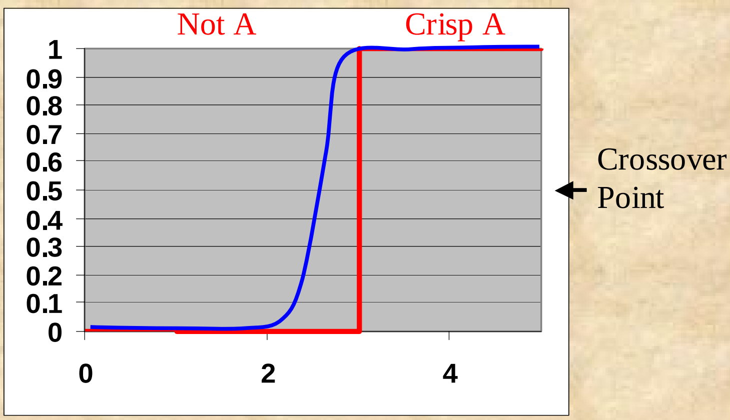
Semantic Approach

Version 1, January 2000

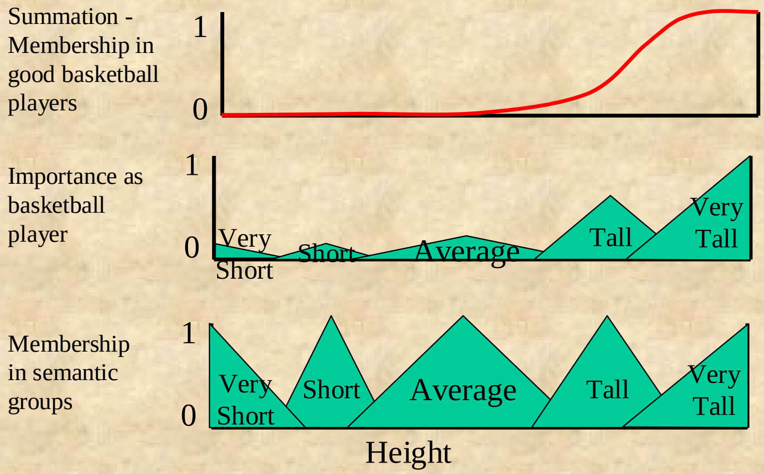
Membership Functions



Membership Function



Semantic Summation



Version 1, January 2000

Fuzzy Semantic Rules

T has two fuzzy membership values, T_{Tall} and $T_{Very\ Tall}$

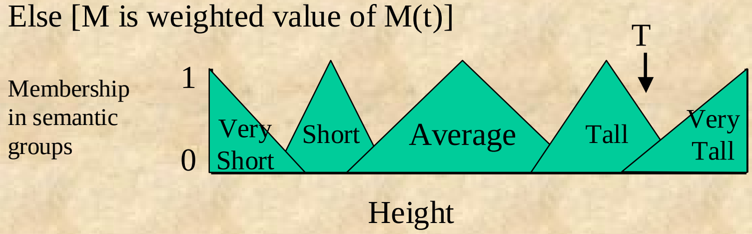
Then a fuzzy rule to define the membership (M) might be

If [T < Tall] Then [M = 0]

ElseIf [T is Tall and T is Very Tall]

Then [M is function of T_{tall} and $T_{Very\ Tall}$]

Else [M is weighted value of M(t)]



Version 1, January 2000

Speculations on the Nature of Evidence

Positive and Negative Evidence

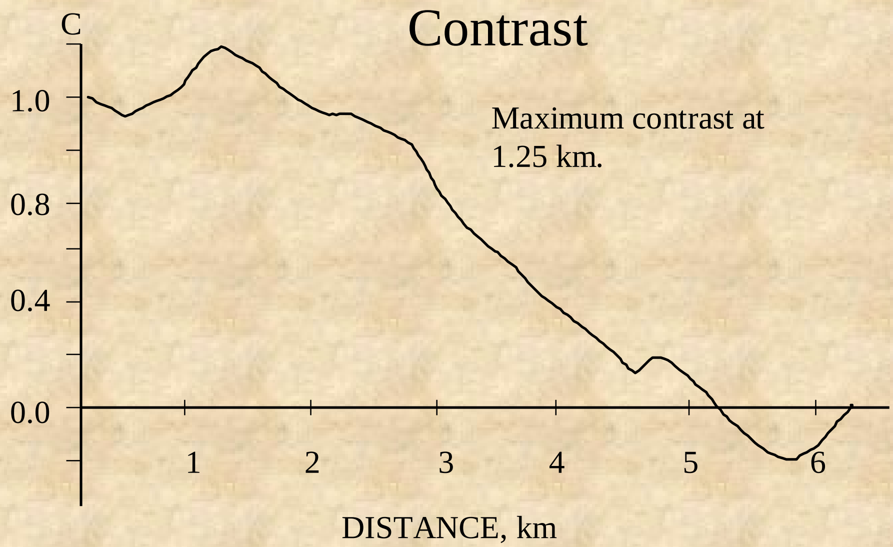
Spatial Resolution of Evidence

Contrast and Map Scale

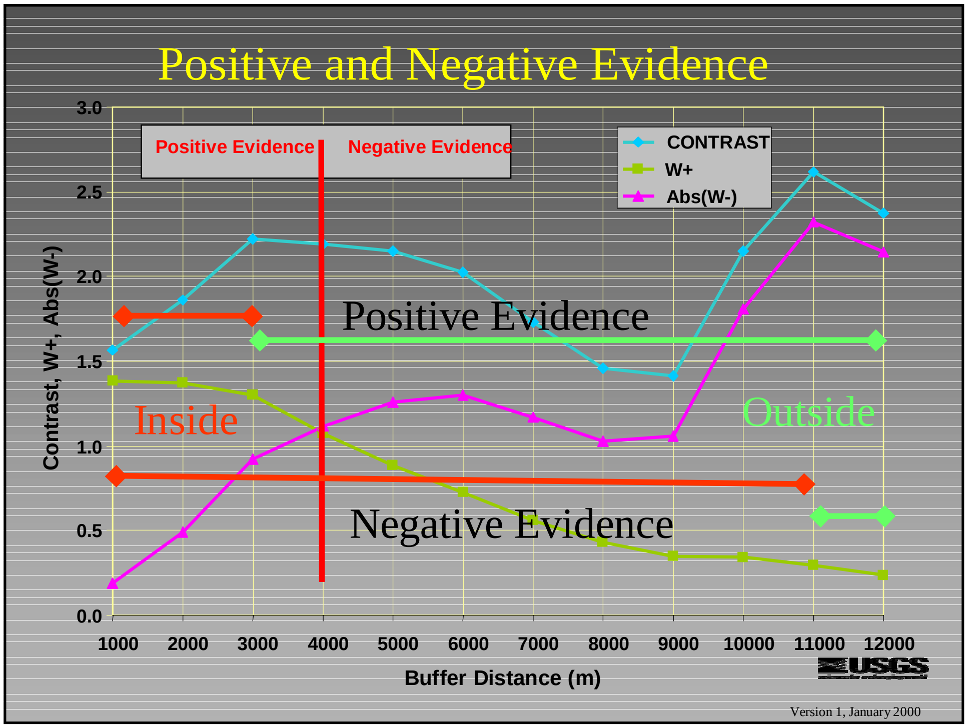
Surrogates

Version 1, January 2000

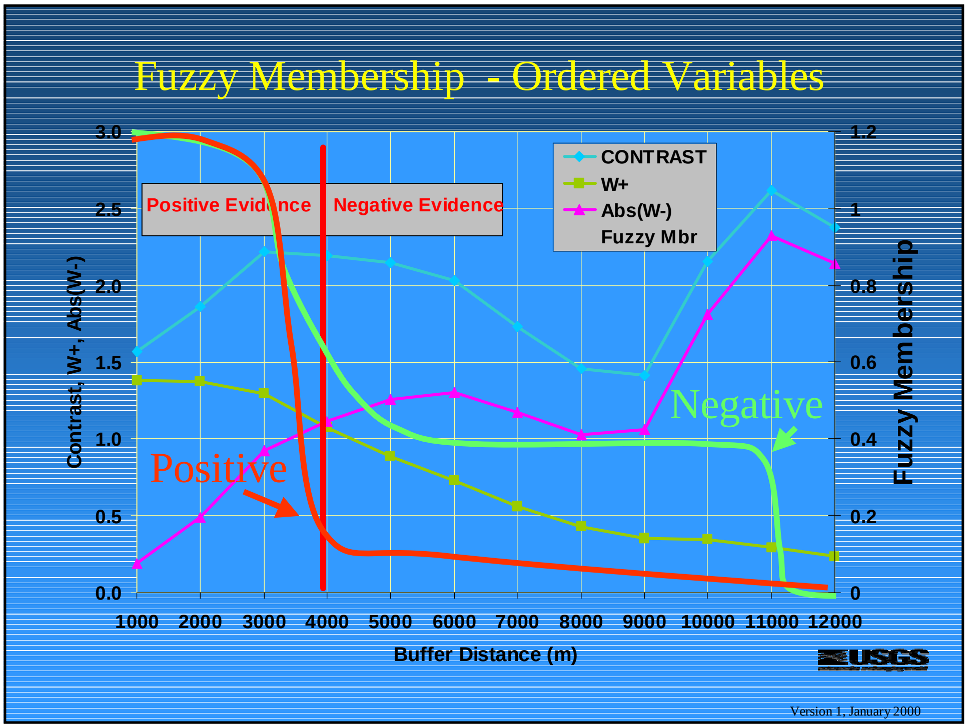
Contrast



Version 1, January 2000



Positive and Negative Evidence
Positive Evidence
Negative Evidence
CONTRAST
W+
Abs(W-)
Positive Evidence
Inside
Outside
Negative Evidence
Contrast, W+, Abs(W-)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000
Buffer Distance (m)
USGS
Version 1, January 2000



Fuzzy Membership - Ordered Variables
Positive Evidence
Negative Evidence
CONTRAST
W+
Abs(W-)
Fuzzy Mbr
Negative
Positive
Contrast, W+, Abs(W-)
Fuzzy Membership
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.2
1
0.8
0.6
0.4
0.2
0
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000
Buffer Distance (m)
USGS
Version 1, January 2000

Buffer Resolution

Threshold Weighting Reclassification

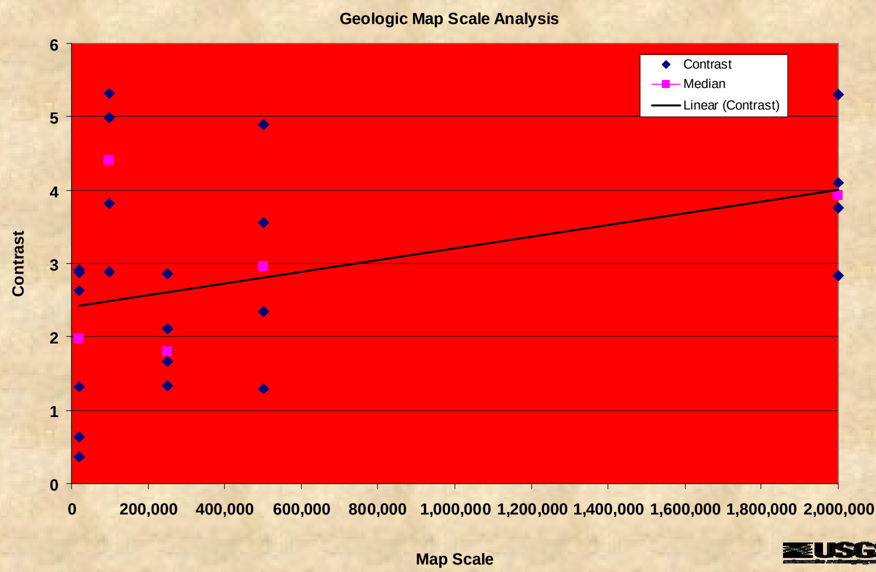
Map Scale	Map Resolution	Geologic Resolution	Buffer Resolution
1:2,500,000	1250	2500	5000
1:500,000	250	500	1000
1:250,000	125	250	500
1:100,000	50	100	200

Units - Meters

Map Resolution = (Scale denominator)/2000

Version 1, January 2000

Contrast and Map Scale



Version 1, January 2000

Comparison of Results

Boolean Logic

Index Overlay

Fuzzy Logic

Weights of Evidence

USGS

Version 1, January 2000

Model Measurement Scales

- Boolean Logic produces a binary, categorical model
- Index Overlay produces a interval-scale model that could be scaled [0,1]
- Fuzzy Logic produces a ratio-scale model in the interval [0,1]
- Weights of Evidence produces a ratio-scale model in the interval [0,1]

USGS

Version 1, January 2000

Mixed Scales

	Nominal	Ordinal	Interval/ Ratio
Nominal	Chi-square, O_r, C_w, etc.	Median by nominal class	Mean by nominal class
Ordinal		Rank correlation coefficient	Rank correlation coefficient
Interval/ Ratio			Covariance Correlation coefficient

Correlation Measures

	WofE	Fuzzy Logic	Index Overlay
Boolean Logic	Chi-square, O_r, C_w, etc.	Chi-square, O_r, C_w, etc.	Chi-square, O_r, C_w, etc.
Index Overlay	Rank correlation coefficient	Rank correlation coefficient	Rank correlation coefficient
Fuzzy Logic	Pearson's Correlation coefficient	Pearson's Correlation coefficient	Rank correlation coefficient

Other Correlation Measures

- For comparison between WofE, Fuzzy Logic, and Index Overlay
  - Reclassify by quantiles, combine grids, and symbolize or make quantile-quantile plot
  - Subtract one from the other
  - Inspect cross-tabulation table
- For comparison of any with Boolean Logic
  - Reclassify into binary classes, combine grids, and symbolize.
  - Inspect cross-tabulation table

USGS

Version 1, January 2000

Surrogates

- There is often need for an attribute or property that is not directly mapped.
  - Bat Habitat
    - Geology - limestones and lava tubes
    - Roads crossing rivers
  - Bioavailable Nitrogen
    - Important control of water clarity at Lake Tahoe
    - Timber capacity - measure of maximum timber growth

USGS

Version 1, January 2000

Surrogate for Bioavailable Nitrogen

- Red areas indicate areas of maximum forest-growth potential.
- Plants are primarily responsible for moving nitrogen from the atmosphere to the soil.
- Therefore, where there is a large volume of plant growth, there is high nitrogen.
- Disturbance in any of these red areas will release nitrogen into the lake.

USGS

Version 1, January 2000

Testing Data-driven Methods

- Use points not included in training set to test the model
  - Implementation - use a random subset of training set to develop the weights and use the remainder to evaluate the model.
  - Problem - for many models there may only be a small number of training points to start with.

USGS

Version 1, January 2000

Logistic Regression Method

Graeme Bonham-Carter

Bonham-Carter, 1999

Introduction

- "Data-driven" method applicable where training set of mineral sites is available
- The response variable is dichotomous (binary), e.g. presence/absence of mineral site
- The explanatory variables (evidential themes) are ordered or dichotomous (not multi-class categorical).

Bonham-Carter, 1999

In ordinary regression, the response variable is continuous, unbounded and measured on an interval or ratio scale

In situations where the response variable is binary (present/absent) this causes a problem, because the predicted response must be in the interval [0,1].

The response variable can be assumed to be $P(Y=1)$, from which we also know $P(Y=0)=1-P(Y=1)$

Bonham-Carter, 1999

The solution to the problem of forcing the response variable to be in the range [0,1] is to use the logit transform.

Logits = natural logs of odds

Odds = Probability/(1-Probability)

$$\text{Logit}(Y) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

Where the b's are unknown coefficients and the X's are the explanatory variables

Bonham-Carter, 1999

Logistic Regression Vs. Weights of Evidence

$$\text{Logit}(Y) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

(simultaneous solution of b's)

$$\text{Logit}(Y) = \text{Prior Logit} + W_1 + W_2 + W_3 + \ldots + W_k$$

(solution for W's theme by theme, not simultaneous)

Note that the b_0 term in LR is comparable to the prior logit in WofE, and the b's are comparable to the W's. However, instead of 1 coefficient, there are 2 (or more) weights, depending on the number of classes. Therefore, the b's are more comparable to the contrast values

Bonham-Carter, 1999

Solution to Logistic Regression Equation

- The coefficients cannot be solved by ordinary least squares (a direct matrix inversion), because the equation is non-linear
- The method of maximum likelihood is used to maximize the value of a log-likelihood function
  - This requires an iterative solution
- So coefficients are obtained simultaneously without an assumption of conditional independence.

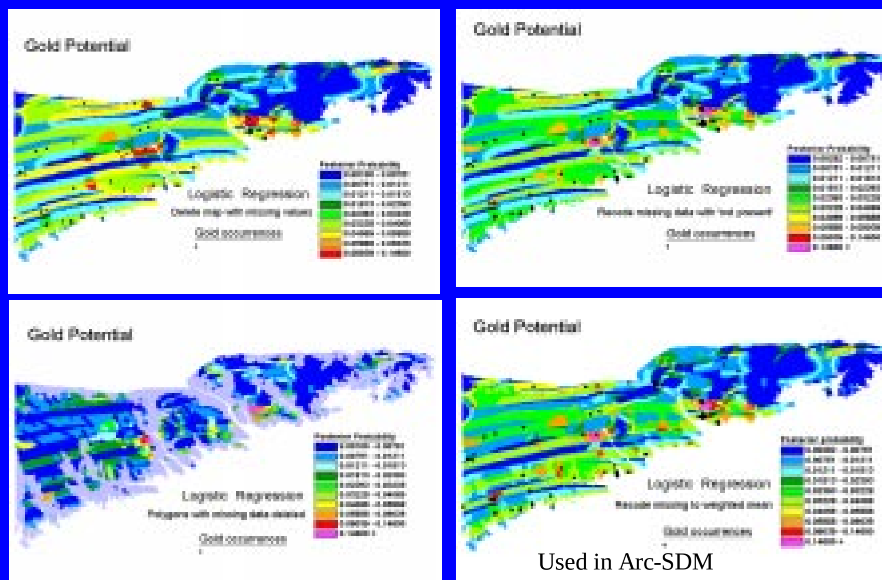
Practicalities

- Can calculate the logistic regression coefficients using the same unique conditions table as for WofE
  - Muti-class themes must be split into binary themes in unique conditions table.
- In Arc SDM deal with missing data and multi-class problem automatically.
- In Arc/Info does not deal with missing data and has another input format.

Problem of Missing Data

- Deleting all unique conditions with missing values in any of the evidential themes.
- Deleting themes that have missing data totally.
- Replacing missing values with zero, or some other constant.
- Replacing missing values with an expected value, e.g. area weighted mean

“Missing Data” Approaches



Can then compare the results from weights of evidence to logistic regression

This is then a check on the effect of conditional dependence on the results of weights of evidence, although if missing data and multi-class categorical evidential themes have been used, then one cannot be absolutely sure what effect the recoding in logistic regression has on the results.

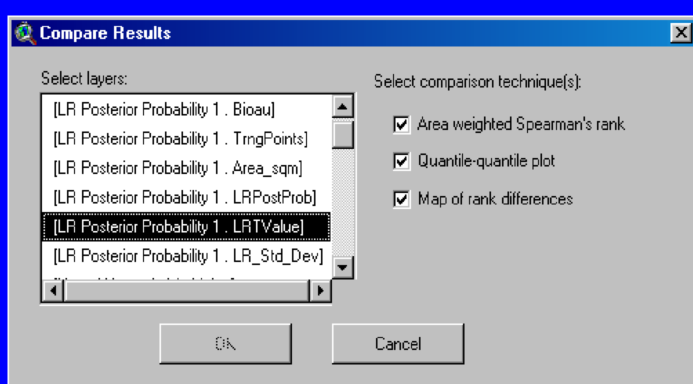
Bonham-Carter, 1999

Compare Results

- Arc-SDM includes three techniques for comparing the results of different techniques:
  - Spearman's Area Weighted Rank Correlation
  - Quantile-quantile plot
  - Map of rank differences

Bonham-Carter, 1999

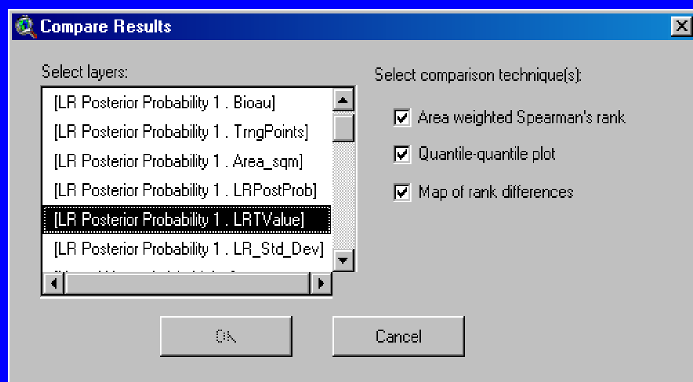
Compare Results



Bonham-Carter, 1999

Compare Results

- Possible inputs:
  - integer grid theme with numeric field(s)
  - floating point grid theme



Bonham-Carter, 1999

Compare Results

Spearman's Rank Correlation and Rank Mapping

- Classifies both variables into 20 quantiles (ranks)
- Spearman's Area Weighted Rank Correlation is calculated and written to a dBase file
- Map of rank differences generates a difference map, classifies and symbolizes it to show where the two input evidential themes are similar or dissimilar

Bonham-Carter, 1999

Compare Results

Spearman's Rank Correlation

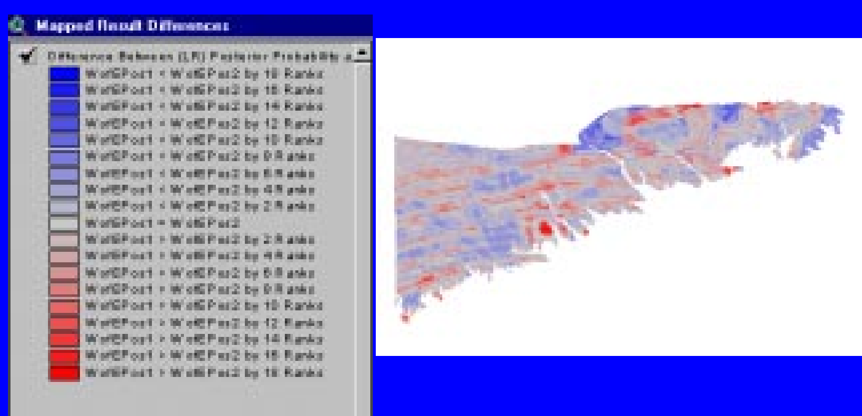
Area Weighted Spearman's Rank 1

Theme, Field	*WofE Posterior Probability 1, Post_prob*
WofE Posterior Probability 1, (LR) Posterior Probability	0.755

Bonham-Carter, 1999

Compare Results

Map of Rank Differences



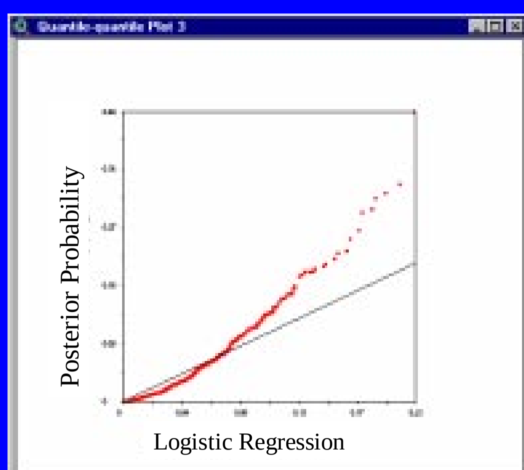
Bonham-Carter, 1999

Compare Results
Quantile-quantile plot

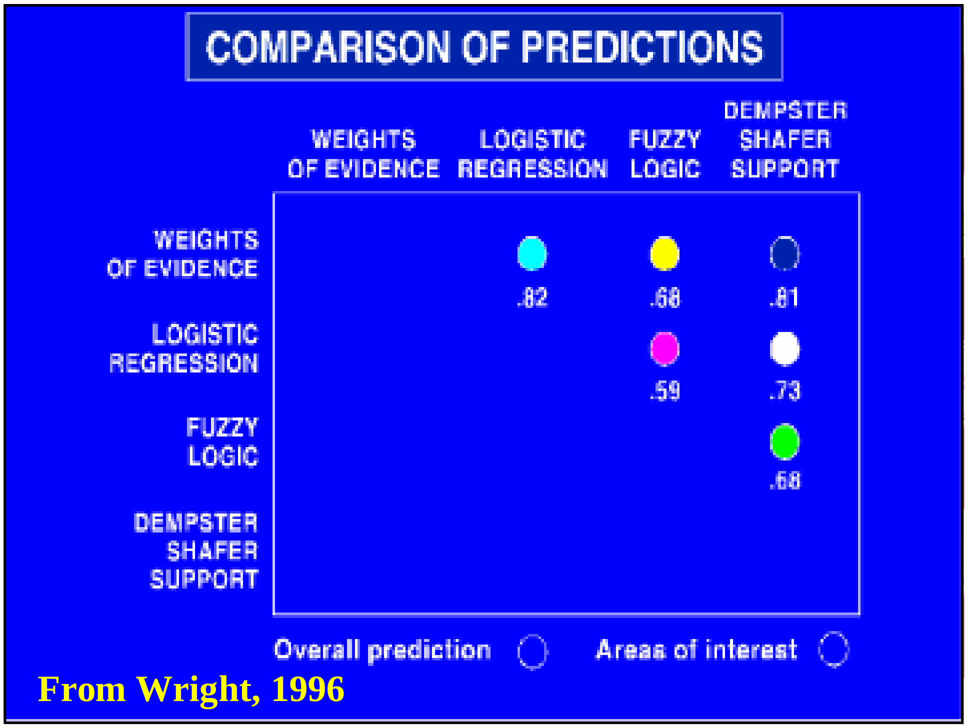
- Sorts the values in each field or theme in ascending order
- if one variable has more observations than the other, its values are interpolated so that there are equal number of values
- values are plotted as x and y coordinates

Bonham-Carter, 1999

Compare Results
Quantile-quantile Plot



Bonham-Carter, 1999



SUMMARY

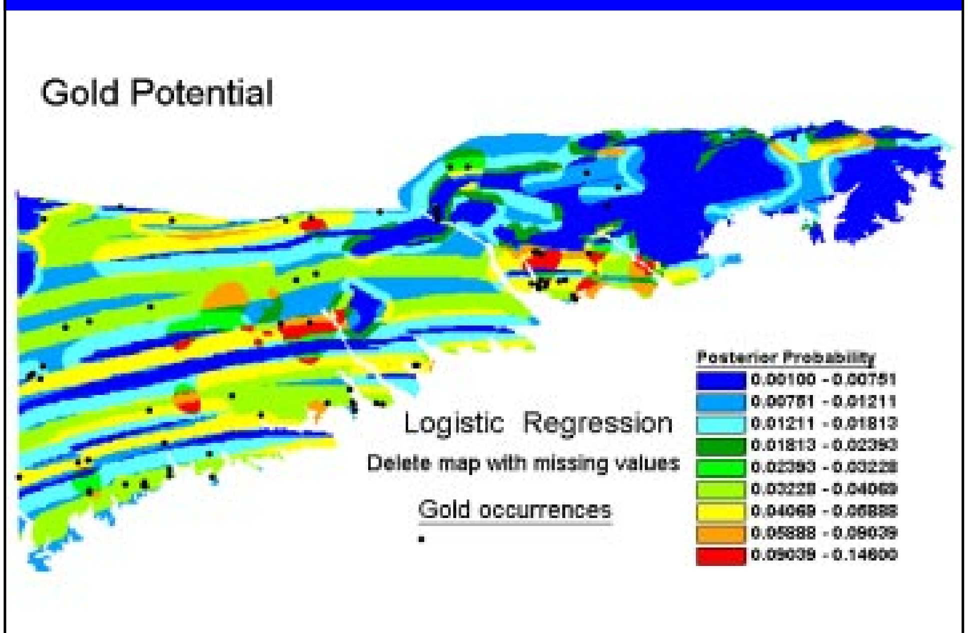
- Logistic regression can be compared to weights of evidence to check CI assumption
- The total expected number of deposits is usually slightly underestimated by LR (rounding?)
- In general the results of the two methods are similar in terms of ranks, except the WofE probabilities are usually higher than LR probabilities because of CI

Bonham-Carter, 1999

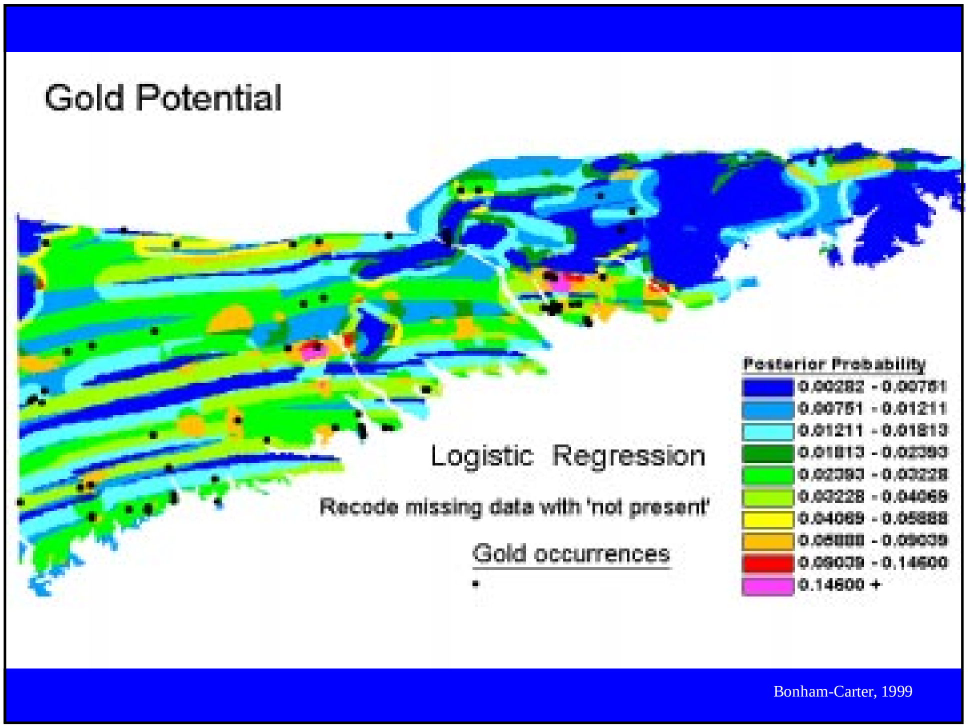
SUMMARY (2)

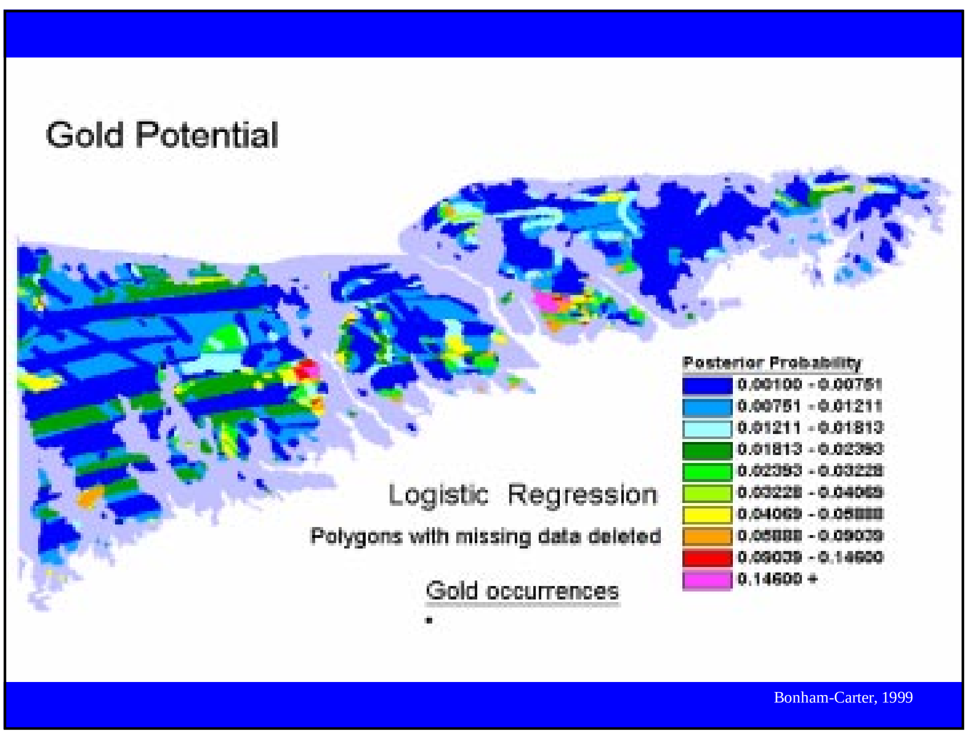
- Arc-SDM will generate LR automatically (expanding the UC table for categorical themes and substituting area-weighted mean values for missing data) at the same time as running WofE, if desired
- Tools for comparing maps are provided

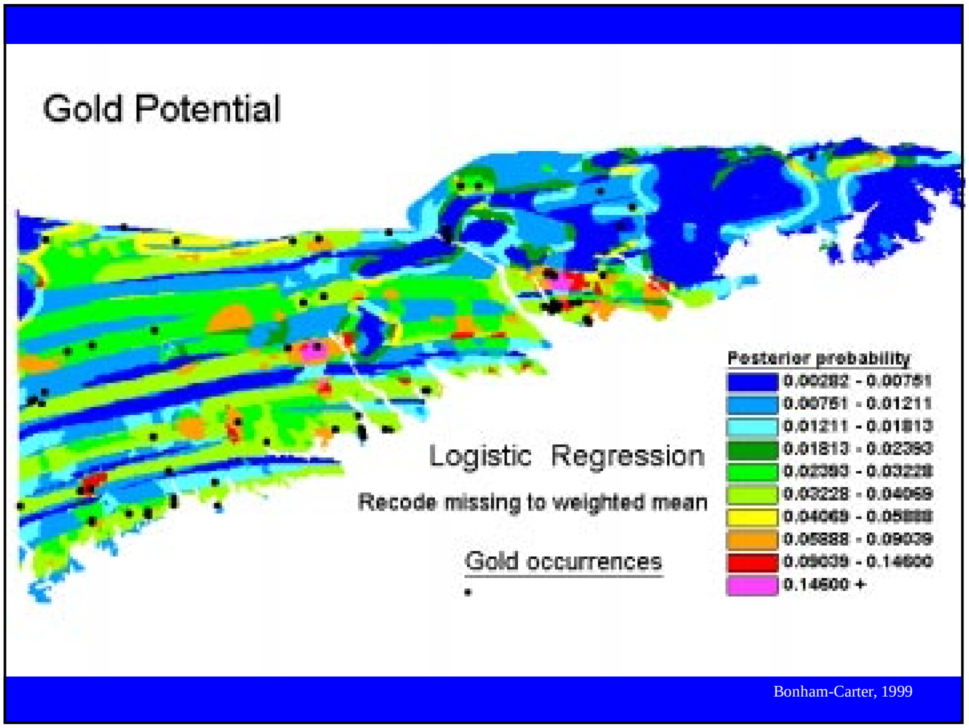
Bonham-Carter, 1999



Bonham-Carter, 1999


Gold Potential
Posterior Probability
0.00282 - 0.00751
0.00751 - 0.01211
0.01211 - 0.01813
0.01813 - 0.02393
0.02393 - 0.03228
0.03228 - 0.04069
0.04069 - 0.05888
0.05888 - 0.09039
0.09039 - 0.14600
0.14600 +
Logistic Regression
Recode missing data with 'not present'
Gold occurrences
Bonham-Carter, 1999


Gold Potential
Posterior Probability
0.00100 - 0.00751
0.00751 - 0.01211
0.01211 - 0.01813
0.01813 - 0.02393
0.02393 - 0.03228
0.03228 - 0.04069
0.04069 - 0.05888
0.05888 - 0.09039
0.09039 - 0.14600
0.14600 +
Logistic Regression
Polygons with missing data deleted
Gold occurrences
Bonham-Carter, 1999



Gold Potential
Posterior probability
0.00282 - 0.00751
0.00751 - 0.01211
0.01211 - 0.01813
0.01813 - 0.02393
0.02393 - 0.03228
0.03228 - 0.04069
0.04069 - 0.05888
0.05888 - 0.09039
0.09039 - 0.14600
0.14600 +
Logistic Regression
Recode missing to weighted mean
Gold occurrences
Bonham-Carter, 1999

Neural Networks

Fuzzy Clustering (Unsupervised)

Radial Basis Functional Link Net (Supervised)

Fuzzy Clustering

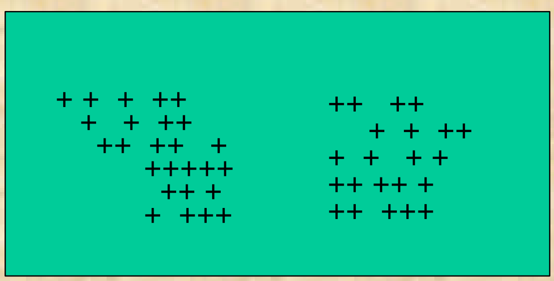
Unsupervised Method

No Training Sites Needed

Unique Conditions Table VAT

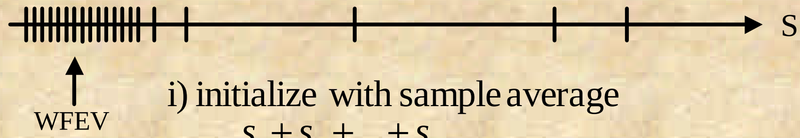
- Each row can be thought of as a feature vector, $\mathbf{x} = (x_1, x_2, \ldots x_N)$ where each x_n is the value or attribute of the feature.
  - There are N attributes for any object in a population of objects.
- There are Q rows or feature vectors
- Goal is to partition the population of feature vectors in classes of objects by partitioning the feature vectors.

Classification



Each + represents vectors (s) in the plane, includes error and measurement noise, but on average they fall into two subpopulations (classes).

Weighted Fuzzy Expected Value



i) initialize with sample average

$$\mu = \frac{s_1 + s_2 + \ldots + s_p}{P}$$

ii) compute fuzzy weights

$$\mathrm{w_p} = \exp[-(s_p - \mu)^2 / (2\sigma^2)] \text{ for } \mathrm{p} = [1, \ldots, \mathrm{P}]$$

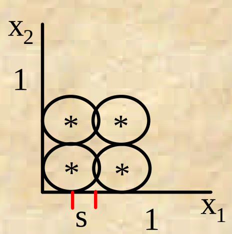
$$W_p = w_p / \sum_{r=1}^{P} w_r$$

iii) compute WFEV

$$\mu = \sum_{p=1}^{P} \mathrm{W_p} \bullet s_p$$

iv) if (stop_criteria) then stop

Estimation of Variance



Xs are standardized between [0,1]

N = 2 = number of evidential layers

M = 4 = number of clusters, experience indicates if want 2 final clusters start with M = 10

Where M is large enough, then can initially estimate the variance by

$$\sigma = \frac{1}{4} * \left(\frac{1}{M}\right)^{\frac{1}{n}} = \frac{1}{4} * \left(\frac{1}{4}\right)^{\frac{1}{2}} = \left(\frac{1}{8}\right) = 0.125$$

Fuzzy Clustering Algorithm

- Input a number K of classes that is larger than the expected number of classes
- Assign first K of the Q vectors as cluster centers $z^{(1)}, \ldots, z^{(K)}$
- For q = 1 to Q
  - Assign $x^{(q)}$ to closest $z^{(K)}$ by c[q]=k
  - Find WFEV for each cluster to obtain a new center $\{z^{(K)}\}$
  - If(any center changes more than e) start over
  - Else Compute weighted fuzzy variance for each cluster and WFEV d_{WFEV} of distances between centers
- for k =1 to K-1
  - for kk = k+1 to K
    - if distance($z^{(k)}, z^{(kk)}$)<$ß d_{WFEV}$ then merge (k, kk)

Radial Basis Function Link Net

Supervised

Training Sites Required

Radial Basis Functional Link Nets

- A radial basis functional link net (neural network, NN) transforms each N-dimensional input ***feature vector*** into an output ***target vector***
  - $\mathbf{x} = (x_1,\ldots,x_n) \rightarrow \mathbf{NN} \rightarrow \mathbf{t} = (t_1,\ldots,t_n)$
- **Target vector t** is a code word that represents a class. This is called ***supervised learning*** because the network must be told the class for each input feature vector x.
- NNs have a relatively large number of parameters that can be thought of as dials. The parameters are also known as ***weights***.
- During training a set of feature vectors are presented to the network and the dials are adjusted until each feature vector is mapped to its known target vector
  - These feature vectors are called ***training vectors*** when used to train the network.

Diagram of Process

$$x = (x_1,\ldots,x_n) \rightarrow NN \rightarrow z = (z_1,\ldots,z_n) \rightarrow e \leftarrow t = (t_1,\ldots,t_n)$$

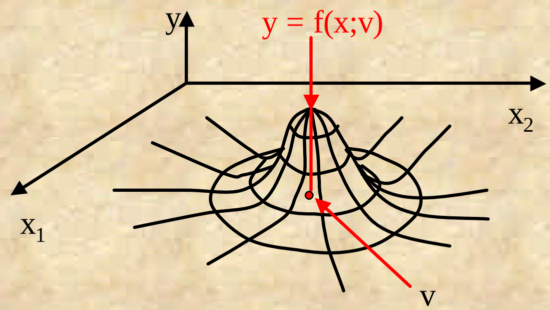
The error to be minimized over all Q input feature vectors is

$$\mathrm{E} = \sum_{q=1}^{Q}\sum_{j=1}^{J}\left(t_j^q - z_j^q\right)^2$$

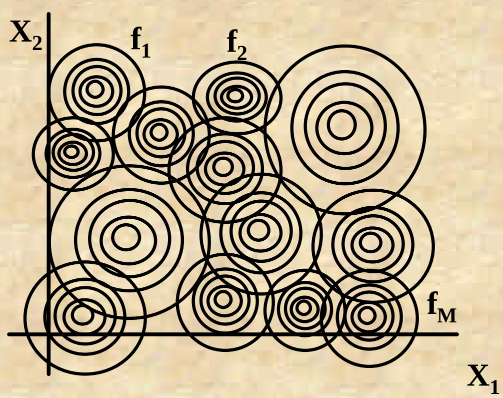
In our case j=1 because only one target value.

Radial Basis Function

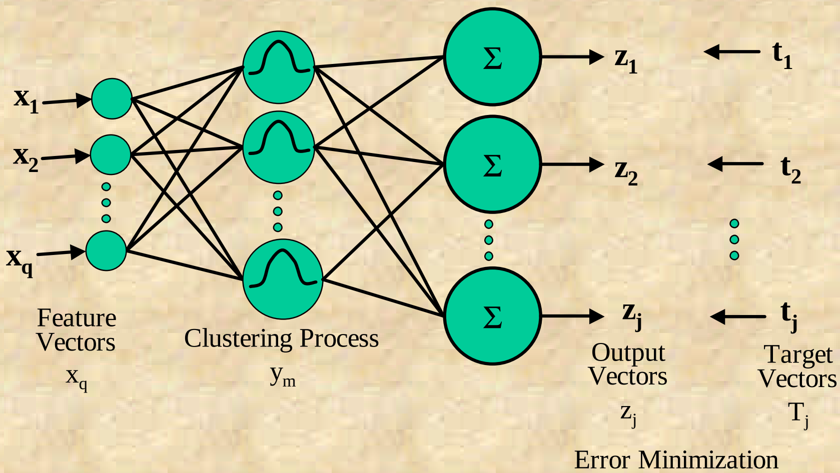
- RBF is a Gaussian function. It has a center vector v and processes any input vector x via $y = f(x;v) = \exp[(x-v)^2/(2s^2)]$ $(0<y\leq 1)$
- Each middle-layer node in RBFN or RBFLN contains a RBF whose output fans out to each node in the output layer.



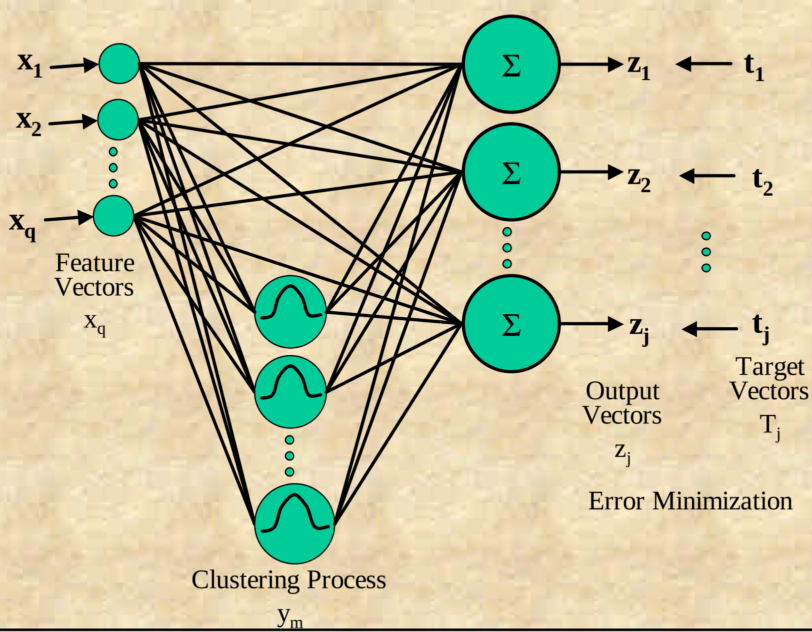
RBF Contour Curves in the Plane



Radial Basis Functional Net



Radial Basis Functional Link Net



Radial Basis Functional Link Net

For an input vector x, the outputs of m^{th} node in the middle layer and the j^{th} node in the output layer

$$y_m = \exp\left[-\left(x - v^m\right)^2 / \left(2s^{\ 2}\right)\right]$$

$$z_j = [1/(M+N)]\left\{\sum_{m=1}^{MJ} u_{mj} y_m + \sum_{n=1}^{N} v_{mj} x_n + b_j\right\}$$

We adjust the weights u_{mj}, v_{mj}, and b_j (the dials) by steepest descent (with gain b) to minimize the total sum-squared error

$$E = \sum_{q=1}^{Q}\sum_{j=1}^{J}\left(t_j^q - z_j^q\right)^2, \quad b_j = b_j - b\left(\partial E / \partial b_j\right)$$

$$u_{mj} = u_{mj} - b\left(\partial E / \partial u_{mj}\right), \quad v_{nj} = v_{nj} - b\left(\partial E / \partial v_{nj}\right)$$

RBFLN Algorithm

$s = \left(1/4\right)\left(1/M\right)^{1/N}$ //set s for RBF width
draw_weights() //draw random weights, -0.5 to 0.5
E_{old} = evalnet() // evalnet() updates NN, gets E

do
{for j = 1 to J do //adjust weights on input node lines
for n = 1 to N do $v_{nj} = v_{nj} - b_1\left(\partial E / \partial v_{nj}\right)$
E_{new} = evalnet();
$if\left(E_{new} < E_{old}\right)$ then $b_1 = b_1 * 1.24$;
$else\ \ b_1 = b_1 * 0.96$
$E_{old} = E_{new}$;

for j = 1 to J do //adjust weights on hidden node lines
for m = 1 to M do $u_{mj} = u_{mj} - b_1\left(\partial E / \partial u_{mj}\right)$;
E_{new} = evalnet();
$if\left(E_{new} < E_{old}\right)$ then $b_2 = b_2 * 1.24$;
$else\ \ b_2 = b_2 * 0.96$
$E_{old} = E_{new}$;

Iterations = Iterations + 1
if (Interations > 1) then exit;
} while (E_{new} > 0.02);

Functions draw_weights() draws initial weights and the function evalnet() updates the actual $\{z^{(q)}\}$ and error E. The biases of β_j updated by steepest descent.

Input Data Format

N M J Q

$g_1^{(1)}, g_2^{(1)}, g_3^{(1)}, x_1^{(1)}, x_2^{(1)},\ldots, x_N^{(1)}, t_1^{(1)}$

$g_1^{(2)}, g_2^{(2)}, g_3^{(2)}, x_1^{(2)}, x_2^{(2)},\ldots, x_N^{(2)}, t_1^{(2)}$

……………………..

……………………..

$g_1^{(Q)}, g_2^{(Q)}, g_3^{(Q)}, x_1^{(Q)}, x_2^{(Q)},\ldots, x_N^{(Q)}, t_1^{(Q)}$

The t_1 values are the training -set fuzzy-membership output values. This allows for ranking of training sets.

- N = number of evidential layers
- M = number of nodes (RBFs) in middle layer
- J = number of output classes = 1
- Q = number of feature vector/target vector pairs, that is number of unique conditions
- $g_1^{(1)}, g_2^{(1)}, g_3^{(1)}, x_1^{(1)}, x_2^{(1)},\ldots, x_N^{(1)}$ = first input feature vector, $g_1^{(1)}$ is the key field to join with unique conditions table.
- $t_1^{(1)}$ = first target output value in [0,1], where
  - **1 = yes and 0 = no**
  - **0.9 = strong indication of yes**
  - **0.1 = strong indication of no**
  - **Can use to say "kind of like" a training site!**

Output Results File

$g_1^{(1)}, c^{(1)}, f_1^{(1)}, f_2^{(1)},\ldots, f_K^{(1)}$

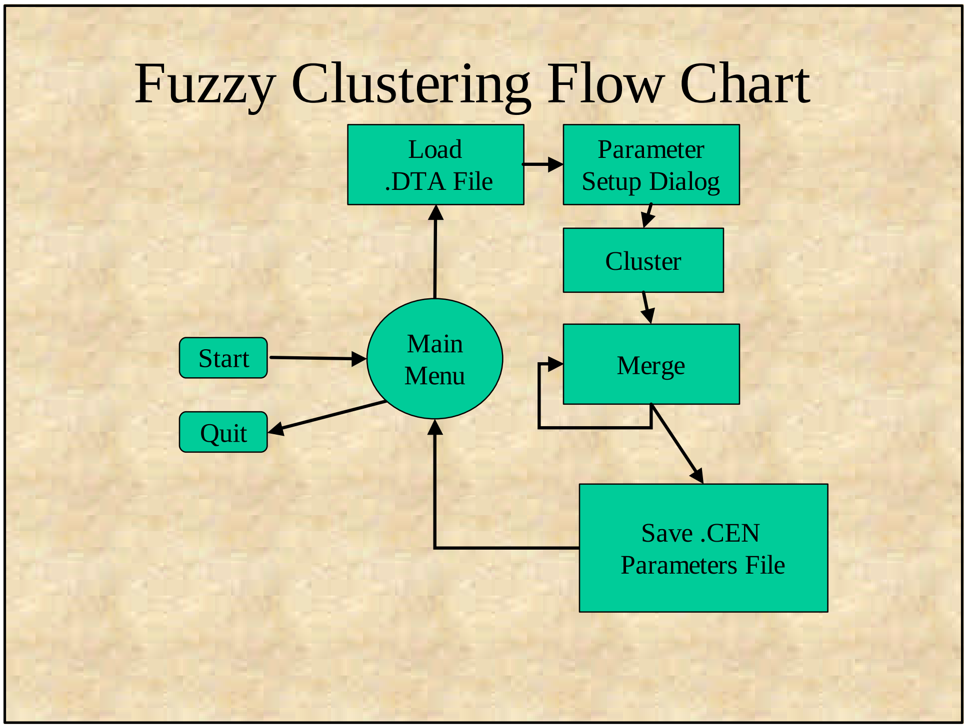
$g_1^{(2)}, c^{(2)}, f_1^{(2)}, f_2^{(2)},\ldots, x_K^{(2)}$

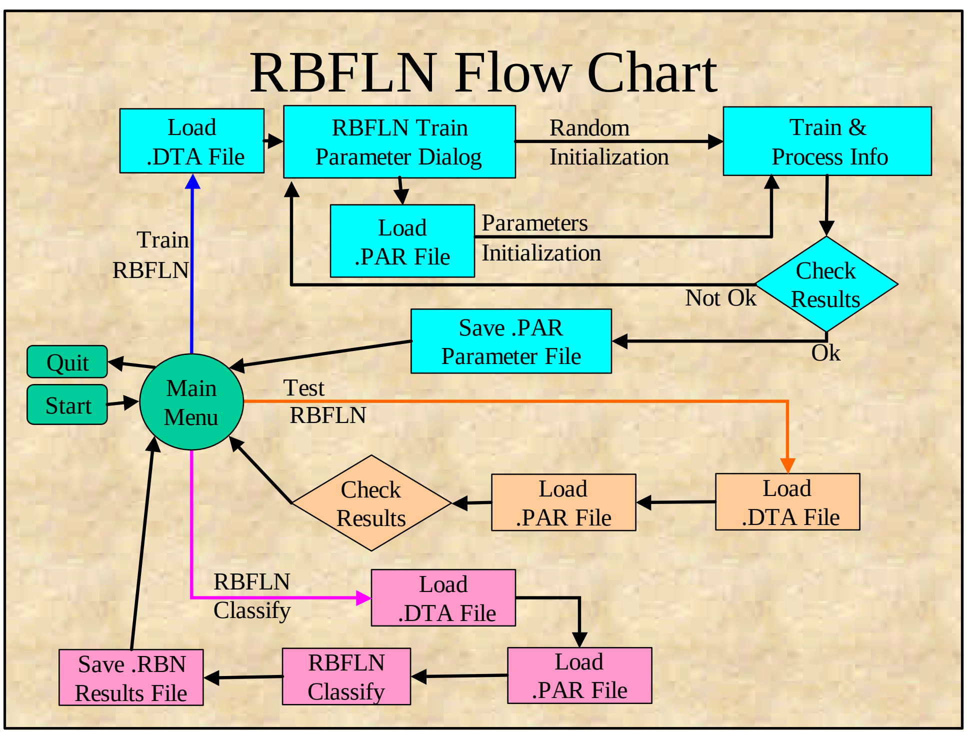
……………………..

……………………..

$g_1^{(Q)},c^{(Q)}, f_1^{(Q)}, x_2^{(Q)},\ldots, x_K^{(Q)}$

- $g_1^{(q)}$ is the key field to join with unique conditions table
- $c^{(q)}$ is the fuzzy class number
- $f_1^{(q)}$ fuzzy membership values, respectively for input vector q belonging to class k = 1, …,K.


Fuzzy Clustering Flow Chart
Load .DTA File
Parameter Setup Dialog
Cluster
Merge
Start
Main Menu
Quit
Save .CEN Parameters File


RBFLN Flow Chart
Load .DTA File
RBFLN Train Parameter Dialog
Random Initialization
Train & Process Info
Load .PAR File
Parameters Initialization
Train RBFLN
Check Results
Not Ok
Save .PAR Parameter File
Ok
Quit
Start
Main Menu
Test RBFLN
Check Results
Load .PAR File
Load .DTA File
RBFLN Classify
Load .DTA File
Save .RBN Results File
RBFLN Classify
Load .PAR File

Decision with Neural Networks

- Transform evidential values into range [0,1]
  - Can use fuzzy membership values as inputs
  - Possibly can use value field
- Ranking of training sites
- Evaluation of reported measures of classification

Summary

- Advantages
  - Can rank training sites
  - Non-linear mathematics
  - Unsupervised and Supervised method
- Disadvantages
  - Model parameters are difficult to understand
  - Need training sites for occurrence and non - occurrence
  - Approaches to ranking of training sites not well understood
  - Overall use is poorly understood

Miscellaneous Guidance

Sharing HTML Files

HTML files are readily read on many computers, and many word processing programs will create HTML files. If the Word document contains inserted pictures such as bit-maps files (BMP files) or some other format; it is necessary to edit the HTML file after it is created. This editing can be done in Notebook or a word processing.

The path to the pictures is, however, hard-coded in an absolute format in the HTML file; so a user of the HTML file on another computer will not see the pictures. For example, if I save the HTML file in the d:\temp folder, this would create the appropriate HTML file and an image file in the GIF format for each picture in the word processing document. These image files would be named Image1.gif, Image2.gif, … up to the number of pictures in the original document.

In the HTML file, these images would be referenced in a line of code beginning as follows:
<P><IMG SRC="d:\temp\Image3.gif" WIDTH =

The d:\temp is the absolute path in the computer where the HTML file was created. This path needs to be changed to a relative path. In addition, it often helps exchange of the file if the images are all collected in some appropriately named folder, for example for a HTML file named Exer1, the images could all be stored in a folder called Exer1Img. Then the line of code above would be edited as follows:

Find: SRC="d:\temp\Image
Replace with: SRC="Exer1Img\Image

When this is completed for each image, that is there is a separate line of code for each image, the code will be as follows:
<P><IMG SRC="Exer1Img\Image3.gif" WIDTH =

Packaging the HTML File

To share the HTML file with an associate, send the modified HTML file and the folder with the included image files. Storing everything in a ZIP file can facilitate the exchange. Do the following so the paths are relative in the ZIP file:

1. Create a folder with the HTML file in it and a subfolder with all of the pictures.
2. Edit the HTML file to reflect the relative path to the subfolder.
3. Create a ZIP file above the folder with the HTML file.
4. In WinZip, open the folder with the HTML file.
5. Check the "recurse folders" or "includes subfolders" checkbox. Depending on the version of WinZip being used.
6. Click the Add with Wildcard button. This will add the HTML file with no path information and the pictures with a relative path.
7. The Zip file can be shared.

To open and use the Zip file
1. Create a folder to store the contents of the ZIP file.
2. Extract everything in the ZIP file into this folder.

Geographic Information Systems for Geoscientists: Modeling in GIS by Graeme Bonham-Carter

Errors in Text

1. P. 215, middle right of page. Text reads "…**W**, as shown in Table 7-12." It should read "…**W**, as shown in Table 7-11."
2. P. 215, bottom right of page. Test reads " 1s in W, Table 7-12." It should read " 1s in W, Table 7-12."
3. P. 260, Table 8-9B, the 253.5 in the lower right corner of the table should be 303.5. This correction causes the Spearman's correlation coefficient on p. 259 to change from 0.822 to 0.826 (lower right edge of p. 259).

Layout Problems

1. On pages 226 to 235, the gray algebraic expressions are often quite a few pages from the figure referenced. For example, the gray box at the top of page 226 refers to figure 8.2 on page 223. So be careful. The following shows the correlation: Page 227 top refers to Fig 8.3 (p. 225), Page 227 middle refers to Fig 8.5 (p. 226), Page 227 bottom refers to Fig 8.6 (p. 228), Page 230 refers to Fig 8.7 (p. 229), Page 231 top refers to Fig 8.9 (p. 230), Page 231 middle refers to Fig 8.11 (p. 232), Page 234 refers to Fig 8.12 (p. 233).

ArcPress

- ArcPress is a powerful tool for preparing maps for printing. One of its major advantages is to rasterize your file before it goes to the printer. This saves printer time, the printer is not tied up rasterizing the file, and your map is printed faster.
- There are many other features of ArcPress that are useful such as allowing you to print files too large for printer memory, paneling of large maps, and color control. Print options include page handling, page layout, clip, scale, autofit, and rotation. Autopanel slices the map to fit the paper.
- Often map colors are too saturated. If you reduce the saturation in ArcPress with the saturation control, the blacks become gray. ESRI recommends that you adjust the CMYK values in ArcPress instead of the saturation. The recommended CMYK values are 55, 55, 55, 90. This will uniformly reduce the saturation but leave the blacks alone.

Useful Extensions

Xtools.avx

A set of useful table and shapefile tools available from the ESRI User Supplied ArcScripts (http://gis.esri.com/arcscripts/scripts.cfm)

Quoted from Mike DeLaune
XTools is a package of tools useful in vector spatial analysis. Included are various overlay, shape conversion and table tools. Go to http://www.odf.state.or.us/StateForests/sfgis/document/Xtools.htm *for a more complete description of XTools.*

Go to http://www.odf.state.or.us/StateForests/sfgis/default.htm *for a Frequently Asked Questions (FAQ) page, documentation, and a comparison of XTools overlay operations to ArcInfo overlay operations.*

Patch.avx

Provides spatial statistics to describe maps and Grid Transform tools. The source code is available from Elkie and others (1999) and a critical users manual is available from McGarigal and others (1994). This extension is distributed by the authors in a controlled fashion; consequently each user has to download for their own use.

Elkie, P.C., Rempel, R.S., and Carr, A.P., 1999, Patch analyst user’s manual – a tool for quantifying landscape structure: Northwest Science and Technology Thunder Bay, Ontario TM-002, 16p + appendix, http://flash.lakeheadu.ca/~rrempel/patch/.

McGarigal, Kevin, and Marks, B.J., 1994, FRAGSTATS – spatial pattern analysis program for quantifying landscape structure, version 2: For. Sci. Dept. Oregon State Univ., 67p. + 3 appendices, ftp://ftp.fsl.orst.edu/pub/fragstats.2.0/.

AVPrimed

A useful collection of Arcview tools for management of Arcview projects. Nice enhancements of some tools such as Spatial Analysis/Analysis/Map Calculator. I have found some conflicts between this extension and others; so I use it only when needed.

Quoted from Dave Theobald, Natural Resource Ecology Lab
http://www.ndis.nrel.colostate.edu/davet
AVPrimed is an ArcView v3.x extension that enhances your productivity by automatically documenting both your work session and the changes you have made to shapefiles. Many standard ArcView dialogs and operations have been optimized for speed and usability. While the standard ArcView interface works well when working with 1 or 2 maps at a time, it is limited when processing multiple themes.
Key functions of AVPrimed include:
- *Automatic generation of log files for ArcView projects and shapefiles*
- *Restore previous queries in the Query Builder and calculate strings in the Field Calculator*
- *Simple queries with one click*
- *Copy documents within and between projects*
- *Display detailed information about documents, such as table source file names and field name, type, and width*
- *Repeat last action at the click of a button to reduce tedious, repetitive work*
- *Select features based on adjacency and common value*
- *and many others!*

29 Feb 2000 - fixed bug in clip themes...
4 April 2000 - added "is within" theme dissolve
9 April 2000 - fixed bug in "is within" theme dissolve
14 April 2000 - added "Convert shapefile to..." with CLEAN option and RE-ORDER option
6 June 2000 - added View quick query tool
9 June 2000 - modified Field Statistics to calculate on multiple fields
22 June 2000 - preview button in the Field Calculator
31 July 2000 - calculate adjacency (topology) and store adjacent features in a .dbf
27 September 2000 - calculate the nearest distance from the selected features in the first active theme and the selected features in the second active theme.

Compiled_table_tools.avx

Quoted from Charles Herbold
This Extension Adds Several Tools To Your Table Tool Bar . Most are from ESRI's web site. I only compiled them into one extension to save room on my tool bar. Import / Export to Excel ,Export to Word Processor,Rename Fields,Multi-Field Sort,Multi-Field Delete,Field & Table Properties,Add Increment Field, Auto ID with Prefix/Suffix,Make Joins Permanent,Append Tables,Concatenate Fields,Find Common Attributes,Import From Txt File,Find Duplicate in a Table and Tag,Pad a Field,Print Tables, Zoom to Selected,Sum Col and Row's, Convert Text Data To/From {Uppercase/Lowercase/Proper},Break Apart String Field by Word,Merge Any Fields Together,Hide/unhide Fields,Add Formulas to Fields,Calculate with Formulas Stored in a Field, Cut/Paste A Fields Contents,Cut/Paste a Records Contents And more

CorrCoef.avx

Quoted from Charlie Frye
This extension adds a button to your projects View and Table DocGUIs. When you have an active theme or an open table that contains two or more numeric fields, this button will be enabled. You just pick two fields in the list and then click Run Correlation. The correlation coefficient will be shown at the bottom.

Useful Script

GridCorr.ave

Calculates Pearson's correlation coefficient between multiple grids. The calculation uses the grids value attribute and the grid must have a nine-character or less name. Modified from covarian.ave by Kenneth R. McVay. Available from the ESRI User Supplied ArcScripts (http://gis.esri.com/arcscripts/scripts.cfm)

ArcView Technical Tips

This section provides various tips on the use of Arcview.

Sharing Arcview Projects

It is often useful to share an Arcview project with an association. The easiest way to do this is to put the shapefiles, grids, and tables into one folder. In the Arcview project file (APR), the full paths to these files, grids, and tables are absolute paths. It is necessary to edit these paths. For example if you are using a shapefile (carlin.shp) in a folder called carlin on disk drive e, there will be a line in the APR that looks like the following:

Path: "e:/carlin/carlin.shp"

The format of this line in a text editor, such as MS Word is Path:[tab]"e:/carlin/carlin.shp where [tab] indicates a tab. The [tab] is searched for in MS Word with the search string Path:^t"e:/carlin/carlin.shp.

To share the APR these absolute paths should be changed to relative paths. To do this search and replace all occurrences of "e:/carlin/ with /carlin/. Then share the carlin folder and all of its subdirectories. The shared files should keep the same folder name. Then when the shared APR is started, it will find all of the files and will add absolute paths when the APR is saved.

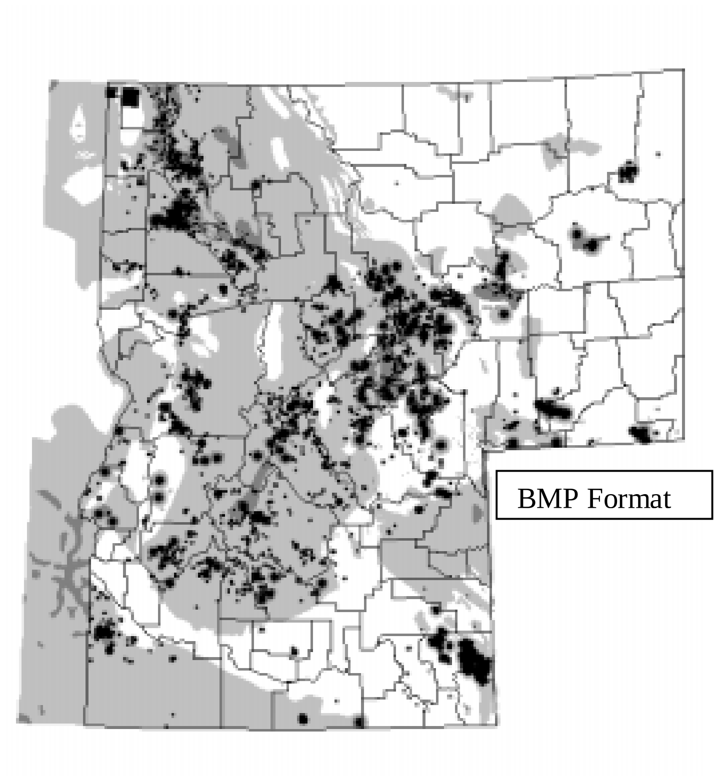
Selecting blank Fields

With ArcSDM, an Arcview extension, it often occurs that the contrast cannot be calculated because there are zero training points in the class. This causes the weight and contrast fields to be blank. If it is necessary to select the records containing these blank numerical fields, a query of the contrast field should use the following format, **([contrast]).IsNull**.

If the blank field is a character field the form of the query would be **([Name] = “”)**.

Report Generation

When incorporating maps from Arcview into reports consider file size. Figure 1 shows a comparison between two common export formats of a view from Arcview. The sizes of the Windows bitmap (bmp) and Windows metafile (wmf) files are 329kb and 72,502kb, respectively. The larger wmf file takes much longer to load into your document and increases the processing speed and file size considerably. The wmf file gives better resolution in the printed version of the vector data in the view and a slightly different contrast. The resolution and contrast are also a function of the printer or the screen. The differences in resolution between the various formats are difficult to see on the computer screen or in PDF formatted reports. In the PDF format, for example, it is necessary to enlarge the image greatly by zooming it to see the differences in the format.



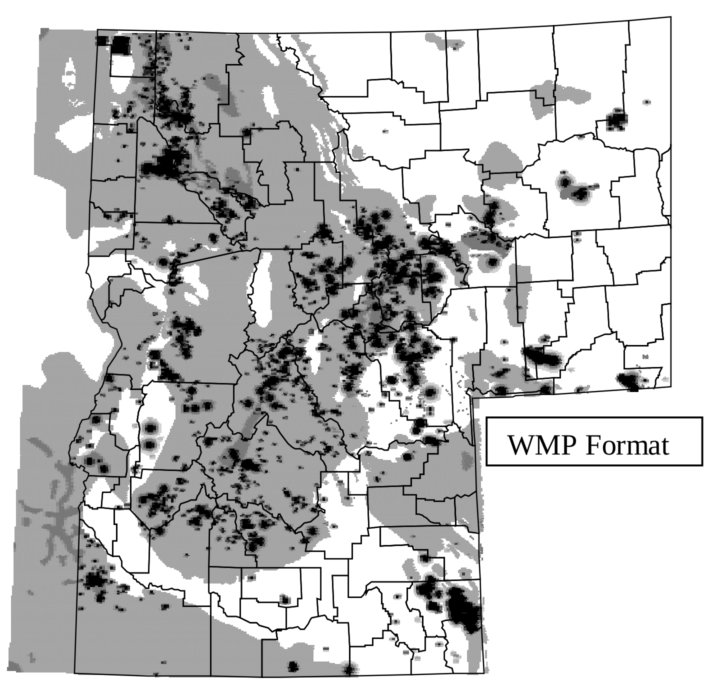


Figure 1: Comparison of Windows bitmap (bmp) versus Windows metafile (wmf) formats.

Arcview-Excel Connection

The Xtools and Patch extensions both have tools that add a menu selection File/Export to Excel when a table is active. These tools will start Excel and then ask for a selection of column headings to transfer. It will transfer all records or only the selected records from Arcview to an Excel table. This is a convenient way to get publication quality tables, charts, or graphs from Arcview tables. It is also a useful way to do various types of statistical analysis using Excel. More information on this subject can be found in the Arcview Help/Index tab, search for Excel or ODBC.

Arcview- Database Connection (ODBC)

ODBC is a PC function for establishing communication between two programs, such as Arcview and Access. ODBC is Microsoft's open interface for accessing data in a heterogeneous environment of relational and non-relational database management systems. For more information about ODBC, in the Arcview Help/Index tab search for ODBC. With an ODBC connection between Arcview and Access, you can run SQL queries in the Access database that will return a table from Access to Arcview. The GeoGen ArcView extension at http://geology.usgs.gov/dm/ is an example of a geologic map with the map attributes in a database.

To manually connect Arcview to another program, you can use Project/SQL Connect when the ArcView Project window is active. If you have established an ODBC connection in your operating system Control Panel/ODBC Data Source, then this connection will be available in the connection menu of Project/SQL Connect. Otherwise, you can select the appropriate standard connection, such as MS Access Connection, and then select the appropriate Access database file. Then you can select SQL queries that are stored in the database or create SQL queries interactively to return a table. This is particularly useful for complex databases, such as geologic maps.

You can also use the Database Access extension available with ArcView. This adds a menu item Project/Add Data Table in the Project window menu. This extension is explained in the Help Contents Extensions/Database Access.

Long file and directory names

Be careful when moving or creating data for use with ArcView. ArcView does not support long file names. These include file names that contain spaces or are long (greater than 8 characters) in the path directories or name of the data source.

The naming convention in Arcview can be confusing because there are two names for a theme, theme name and source. These two names do not have to be the same. The theme name can be edited in View menu Theme/Properties. For many analytical functions, such as map calculator, the theme name will reflect the analytical function to the source, resulting in a very long name with many spaces. The source file for such an operation will have a short name, such as calc3. Similarly in ArcWofE and ArcSDM operations such as buffering create very long theme names. These theme names are then used as column headings in DBF tables, so it is necessary to shorten these long theme names to a maximum of 9 characters with no spaces.

In the Windows operating system, there is a 256-character limit to absolute path names. Be careful, if you are used to creating multi-layered folders.

Filtering

Types of Filters

Name: Laplacian
Form:

0	-1	0
-1	4	-1
0	-1	0

Application: Detection of edges.

Programming and testing filters

Running the Script

Load the filter script and compile it. Then in an active grid, make a grid theme active and then make the filter-script window active. Run the script. This will create a new filtered grid theme at the top of the Active View Window. Symbolization with values near zero set to clear or transparent is then useful.

Known Problems

1. If the range of values in the filtered grid is larger than –100,000 to 100,000 the grid will not display. The solution is to reclassify using Analysis/Reclassify into some smaller range where the extremes are compressed. The interesting values are above or below zero but not close to zero.

Laplacian Filter Demonstration

The Avenue script laplac.ave is a demonstration of how to write a filter in Avenue.

Laplac.ave

```
' Demonstration filtering script
' Calculates a 3x3 Laplacian filter
' Laplacian filter has the following form
' 0 -1 0
' -1 4 -1
' 0 -1 0
' There may be a reclassification problem with the output.
' Values somewhat above and below zero define the highs and lows.
' If the range gets larger than -100,000 to 100,000 Arcview cannot
'   display. Then need to reclassify grid to make the extreme values
'   closer to slightly above and below zero. The problem seems to be
'   with how the neighborhood deals with the edges of the grid.
' ***** The above problem is mostly fixed by making the last
' ***** argument of focalstat TRUE.


'Gary Raines, January 2000

theView = av.GetActiveDoc
```

```
' Use first active GTheme
theTheme = theView.GetActiveThemes.Get(0)

' Obtain grid from theme
theGrid = theTheme.GetGrid

' Make the neighborhood
Line1 = {0,1,0}
Line2 = {1,0,1}
Line3 = {0,1,0}
theNbrHood = NbrHood.MakeIrregular ({Line1,Line2,Line3})
' theNbrHood = NbrHood.MakeIrregular ({{0,1,0},{1,0,1},{0,1,0}})

' Make the filtered grid
' Works best with floating grid
' Multiply the source grid by 4 and subtract the FocalStats sum to get the filter
' More complex filter will require a more FocalStats sums

if(theGrid.isInteger) then
  floatGrid = theGrid.Float
  filtGrid = floatGrid*4.AsGrid.Float - floatGrid.FocalStats(#GRID_STATYPE_SUM, theNbrHood,
TRUE)
else
  filtGrid = theGrid*4.AsGrid.Float - theGrid.FocalStats(#GRID_STATYPE_SUM, theNbrHood, TRUE)
end

' Create a theme
theGTheme = GTheme.Make(filtGrid)
' Check if output is ok
if (filtGrid.HasError) then
  return NIL
end
' Add theme to the View
theView.AddTheme(theGTheme)
```

Normal Grid

The Avenue script normalgrid.ave is a script to generate a grid with normally or randomly distributed data. No edges are expected from this grid.

```
' Demonstration creating Normal or Random Grid
' Need to set cell size, extent, etc. in Analysis Parameters
'   for this script to work.
'   Or could code these into the script with setCellSize

'Gary Raines, January 2000

theView = av.GetActiveDoc
' Use first active GTheme
theTheme = theView.GetActiveThemes.Get(0)

' Make a grid with a normal (0,1) distribution
' Histogram will be a normal distribution
outGrid = Grid.MakeNormal
```

```
'Make a grid with random values between 0 and 1
' Histogram will be flat
'outGrid = Grid.MakeRandom

' Create a theme
theGTheme = GTheme.Make(outGrid)
' Check if output is ok
if (outGrid.HasError) then
  return NIL
end
' Add theme to the View
theView.AddTheme(theGTheme)
```

Test Grid

A useful simple grid to filter and gain an understanding of how filters work is a hillshade grid made from a DEM or other appropriate grid theme. The hillshade is useful because it has simple properties and very sharp edges.

Another Random or Normal Grid

Creating a random or normal grid using the Spatial Analyst/Map Calculator can easily be done with the Analysis/Map Calculator. First, set the properties of the grid using Analysis/Properties menu to establish the desired grid extent and cell size. To activate the Analysis/Properties menu, you first need to add or create a grid. This grid can then be used to define the extent and cell size. To create a random or normal grid, use the Analysis/Map Calculator with one of the following expressions:

Grid.MakeRandom < 0.1

MakeRandom creates a Grid with uniformly distributed random values. Random values are between 0 and 1. With the above expression, approximately 10% of the grid cells will have a value of 1 and the remainder will have a value of 0.

Grid.MakeNormal < 0

MakeNormal creates a Grid with normally distributed random values. The normal distribution uses a mean of 0 and a standard deviation of 1 to create the random values. With the above expression approximately 50% of the grid cells will have a value of 1 and the remainder will have a value of 0.

Correlation

Purpose

The ArcView procedures described here are used to display the spatial variation of two attributes. Such a display can be derived from the combination of two grids, that is one grid with two attributes or one shapefile with two attributes. This can be a useful data exploration approach, particularly to understand the spatial correlation of two variables. This procedure has been developed because in many projects, there has been a need to reclassify grids with two variables. Because this can be a tedious process, a procedure has been developed to begin to automate this task.

The organization of the symbolization of two attributes is not easily understood in the linear display used for legends in ArcView. A solution to this problem is to create a polygon shapefile like a 2-dimensional matrix to serve as the legend or explanation of the variation of the attributes displayed. For example, if two attributes are reclassified into five classes each, a 5x5 matrix (a polygon shapefile with 5 rows and columns of square polygons) can serve as the legend for the symbolization. An example application of this matrix legend in a map layout is shown in Figure 1. This particular example is designed to emphasize the spatial correlation between antimony and arsenic and to identify areas where the variables are not correlated. Other types of displays might emphasize other aspects of spatial variation and are useful for exploring relationships between two maps. The intent of this document is to explain how to create these types of displays primarily for grids in ArcView using the Spatial Analyst and two extensions included here.

Details - Grids

The processing is summarized diagrammatically in Figure 2. The initial grids can be made by several means in ArcView; often the grids might be made from some sort of point observations.

Rescale and Reclassify

These grids are typically floating grids of ordinal, interval, or ratio measurement scales. Floating grids will need to be reclassified into five integer classes to use the reclassification tables provided here. Because ArcView provides a limited selection of reclassification methods for floating grids, it is often useful to convert a floating grid to an integer grid. Conversion of a floating grid is often better if it is rescaled to account for significant figures. The rescaling and conversion to integer can be done with the Map Calculator. For example, if a GridA had three significant figures after the decimal point, the Map Calculator equation would be ([GridA]*1000).int. Then the rescaled grids can be reclassified using the Reclassification tool and selecting quantile, which often provides better results for normal-frequency distribution data. For the shapefile legend and reclassification table provided with this report, the grid should be reclassified into five classes. The number five is an arbitrary value; using the same process but with a different reclassification table (explained below) would allow for 3x3 or 9x9 or any other combination. The use of quantiles assumes the frequency distributions of the grids are approximately normally distributed. If this is not the case, adjustments can be made as described below.

Combine and Sort

After the two grids have been reclassified into 5-class integer grids, i.e. there are only five values in the VAT for each grid; these grids can be combined using the Combine option in the Transform Menu Selection. Installing the extension, Sptrnfrm.avx creates this menu. Two or multiple grids that are adjacent in the View Legend can be selected and combined. Unfortunately, the combinations from the two grids in the combined grid are randomly ordered and there is no key attribute to order them into some consistent or standard order.

To obtain a standardized ordering, install the Xtools extension, Xtoolsmh.avx (Pyle and DeLaune, 1998). Then activate the VAT for the combined grid (GridC in the flowchart) and select Table Frequency in the Xtools menu. This menu selection asks the following questions:

- Select one or more frequency fields to be summarized? – for this application select the fields associated with the combined grids
- Select zero or many summary fields? – this can be ignored for this application
- Do you want to add a case item? – this is the key field that will sort the two attributes, answer yes and then enter the name of the case field (Case is a good choice).
- Name of the output file? This summary DBF table is not used here. Answer OK and the process will be completed. If you CANCEL, the Case item will not be added to the VAT of the Combined grid.

Edit VAT

Sorting the combined grid on Case should give a table similar to the first three columns in Table 2. If the maximum value of Case is not 25, some combinations do not occur in your data set. In order to be able to complete the next step, it is necessary that the VAT have twenty-five entries. Inspection of the VAT sorted by Case should quickly indicate which combinations of classes did not occur in your data set. This deficiency can be quickly fixed by editing the table and adding additional records for the needed pairs. **While editing, do not edit the Value and Count attributes.** Simply enter the appropriate pairs in their columns and then edit the Case attribute so the combined-pairs attributes are properly ordered from 1 to 25. When done the VAT should be like the first three columns in Table 2, except the heading on the columns will reflect your grids. This process always leads to this standard numbering of the 5x5 combinations and 25 records in the VAT of the combined grid (GridC in the flowchart).

Join Reclassification and Symbolize

With a standardized numbering system for the combined grid, a reclassification table can be made to use with standard or customized color ramps to symbolize the combined grid. In order to visualize more easily this symbolization, a shapefile with 25 polygons can be used to display the variation in a type of matrix. This 25-polygon shapefile has the polygons identified with the attribute Case as shown in Table 1. By joining the Table 2 to the combined grid VAT by join item Case, and using the ArcView standard color ramps the shapefile legend and the combined grid can be symbolized in several different ways. Examples of various approaches to symbolization and different color ramps are shown in Figures 3 to 8. There are undoubtedly many other useful ways to symbolize this 5x5 matrix for various types of data exploration. The ones shown here are designed to utilize the standard color ramps that come with ArcView. The symbolization can be further enhanced by increasing the number of intervals in the Legend Editor Classify option to obtain larger color steps between each class. Other useful effects can be obtained by inverting the color ramp to emphasize different aspects of the range of classes.

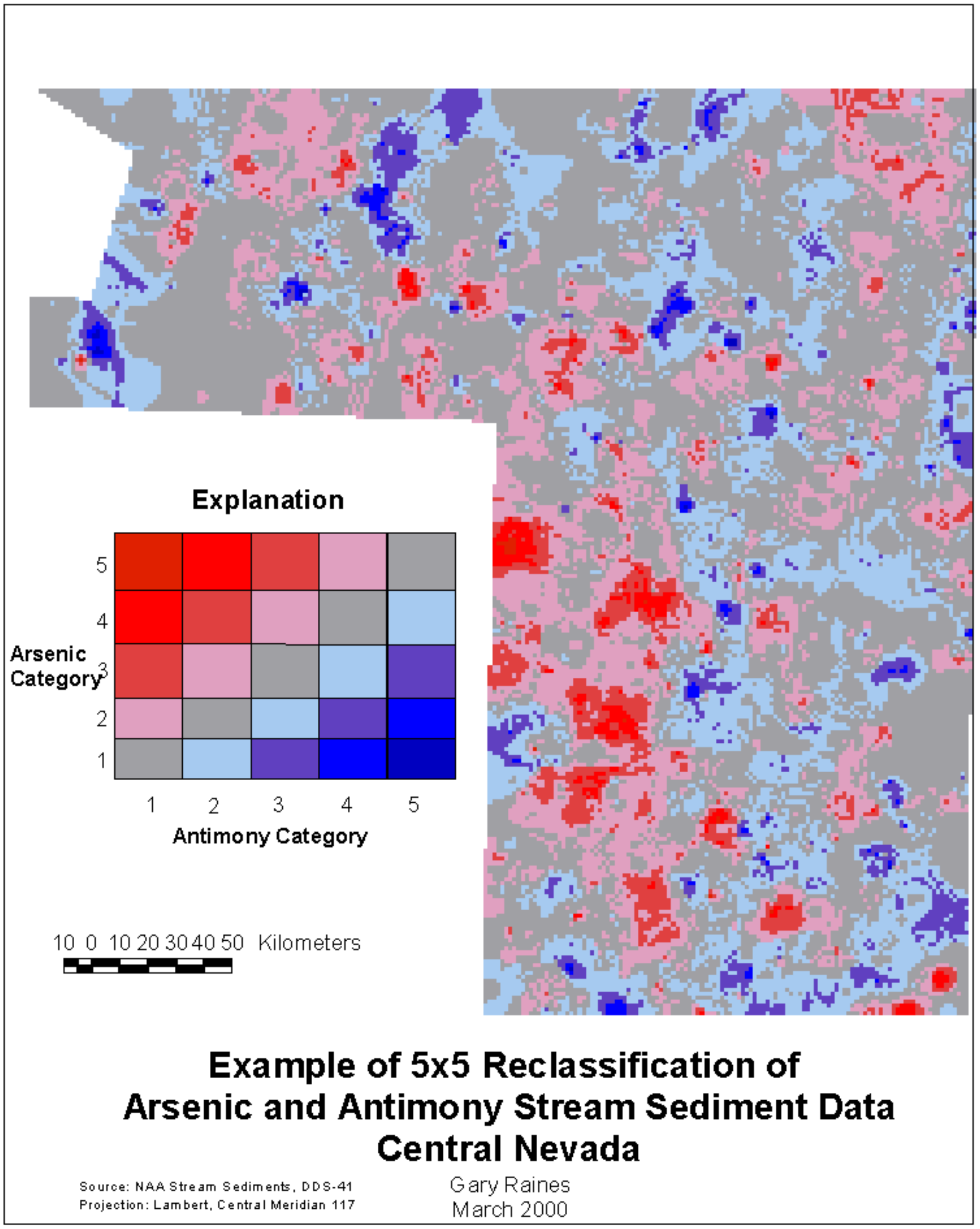


Figure 1: Spatial variation of arsenic and antimony. White areas are where arsenic and antimony are correla ted. As the degree of redness increases, arsenic is higher than antimony. As the degree of blueness increases, antimony is higher than arsenic.

Generalization of Procedure

The procedure outlined here applies only to a 5x5 reclassification of grids. In order to do this for a 3x3, 9x9, or any other combination, it is only necessary to make a new table like Table 2 and the appropriate shapefile for the legend. Editing the table is easily done by displaying the shapefile in ArcView while entering the numbers in the table.

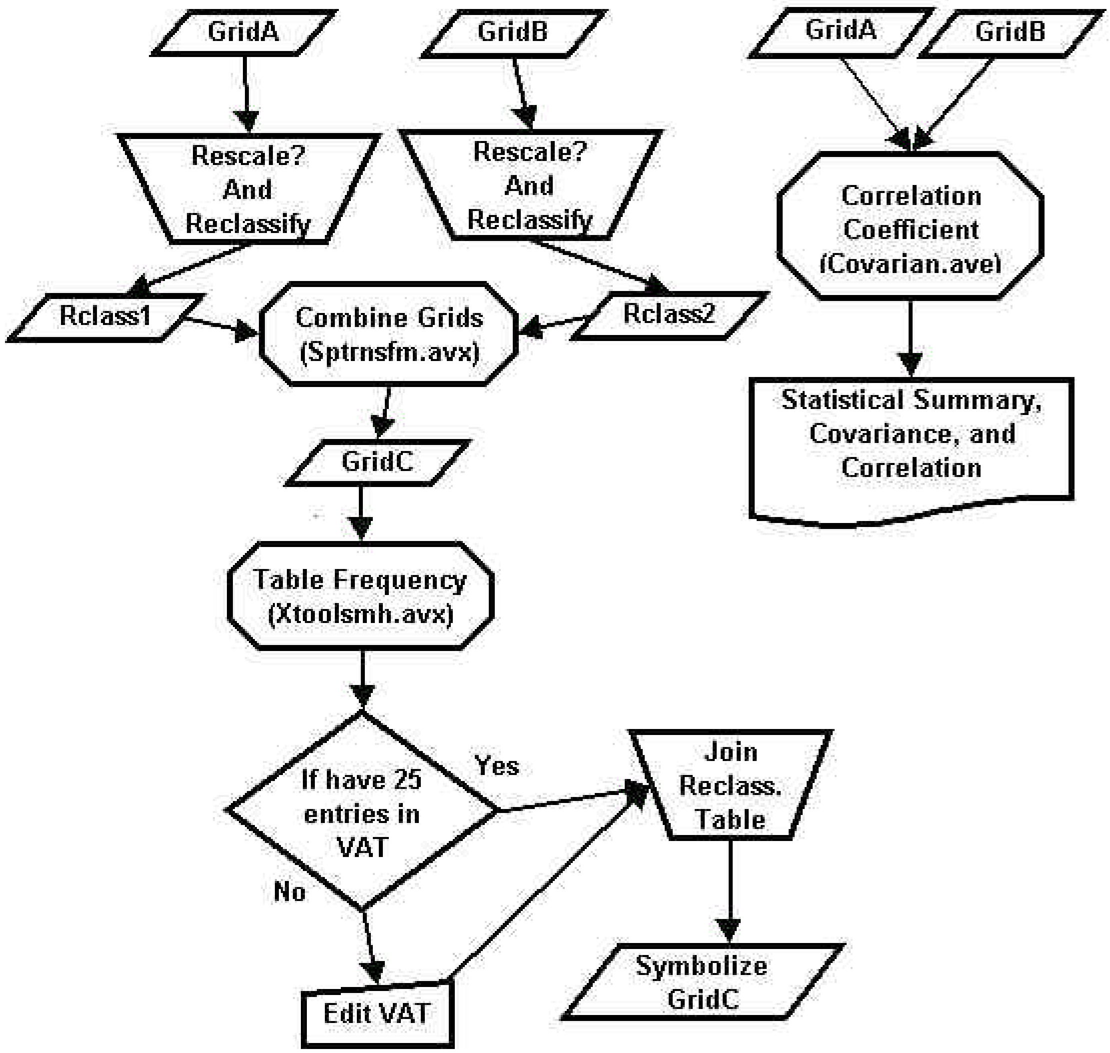


Figure 2: Flowchart of the processing steps. The combined grid (GridC) is symbolized to display the joint spatial variation of grids A and B. Using the Covarian script, a correlation coefficient can be calculated f or the two grids.

Table 1: 5x5 matrix numbering. Grid 1 is first and has 5 categories. Grid 2 is second and has 5 categories. Numbering here reflects the numbering that results from the Table Frequency menu.

Grid1\Grid2	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
Cat 5	21	22	23	24	25
Cat 4	16	17	18	19	20
Cat 3	11	12	13	14	15
Cat 2	6	7	8	9	10
Cat 1	1	2	3	4	5

Table 2: Reclassification table for case to be used with standard Arcview color palettes. This is table rcls5x5c.dbf.

Case	Row	Column	Complex1	Complex2	Simple	Cor_Pos	Cor_Pos3	Cor_Neg	Cor_Neg3
1	1	1	1	1	1	5	2	1	1
2	1	2	2	2	2	6	3	2	1
3	1	3	4	6	3	7	3	3	1
4	1	4	7	7	4	8	3	4	1
5	1	5	11	15	5	9	3	5	2
6	2	1	3	3	2	4	1	2	1
7	2	2	5	5	3	5	2	3	1
8	2	3	8	8	4	6	3	4	1
9	2	4	12	14	5	7	3	5	2
10	2	5	16	16	6	8	3	6	3
11	3	1	6	4	3	3	1	3	1
12	3	2	9	9	4	4	1	4	1
13	3	3	13	13	5	5	2	5	2
14	3	4	17	17	6	6	3	6	3
15	3	5	20	22	7	7	3	7	3
16	4	1	10	10	4	2	1	4	1
17	4	2	14	12	5	3	1	5	2
18	4	3	18	18	6	4	1	6	3
19	4	4	21	21	7	5	2	7	3
20	4	5	23	23	8	6	3	8	3
21	5	1	15	11	5	1	1	5	2
22	5	2	19	19	6	2	1	6	3
23	5	3	22	20	7	3	1	7	3
24	5	4	24	24	8	4	1	8	3
25	5	5	25	25	9	5	2	9	3

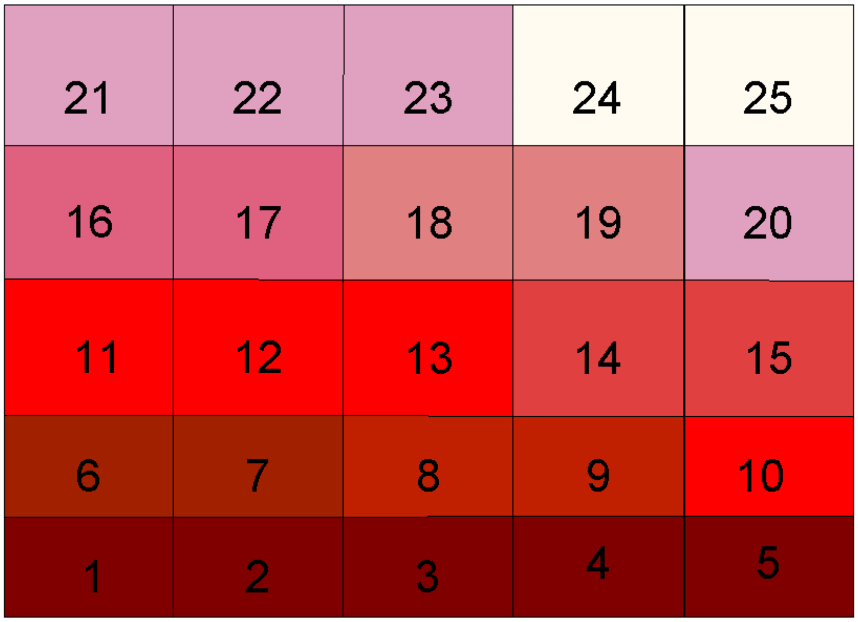


Figure 3: Symbolization of 5x5 matrix of the Case attribute using the standard Red monochromatic selection with 25 equal intervals. The cell numbering shows the numbering sequence that is us ed with the Xtools/Tables Frequency process.

Figure 4: Symbolization of 5x5 matrix using the Simple attribute with the standard Full - Color color ramp with 9 equal intervals. The cell numbering shows the numbering sequence for Simple.

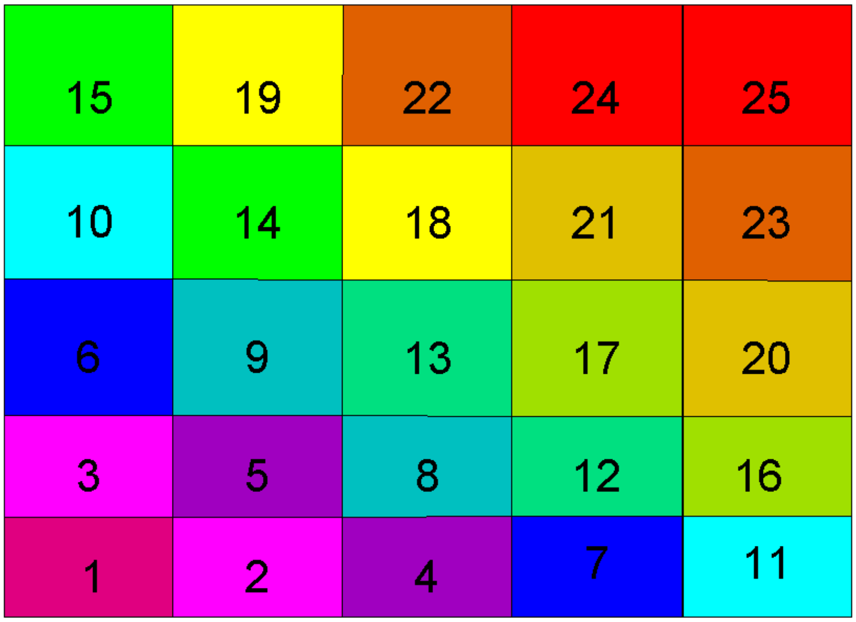


Figure 5: Symbolization of 5x5 matrix using the Complex1 attributes with the Full -Color color ramp with 25 equal intervals. The cell numbering shows the numbering sequence for Complex1.

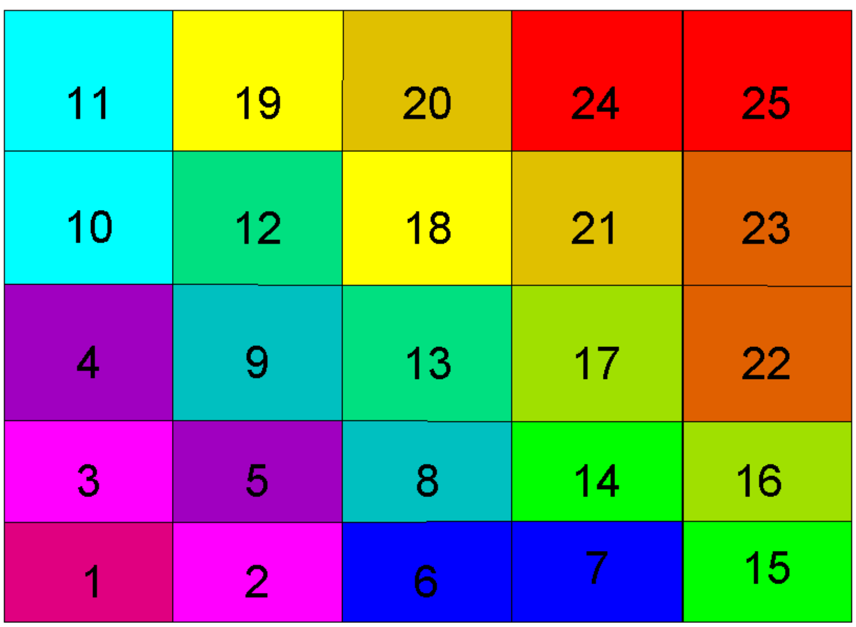


Figure 6: Symbolization of 5x5 matrix using the Complex2 attribute with the standard Red Monochromatic color ramp with 25 equal intervals. The cell numbering shows the numbering sequence for Complex2.

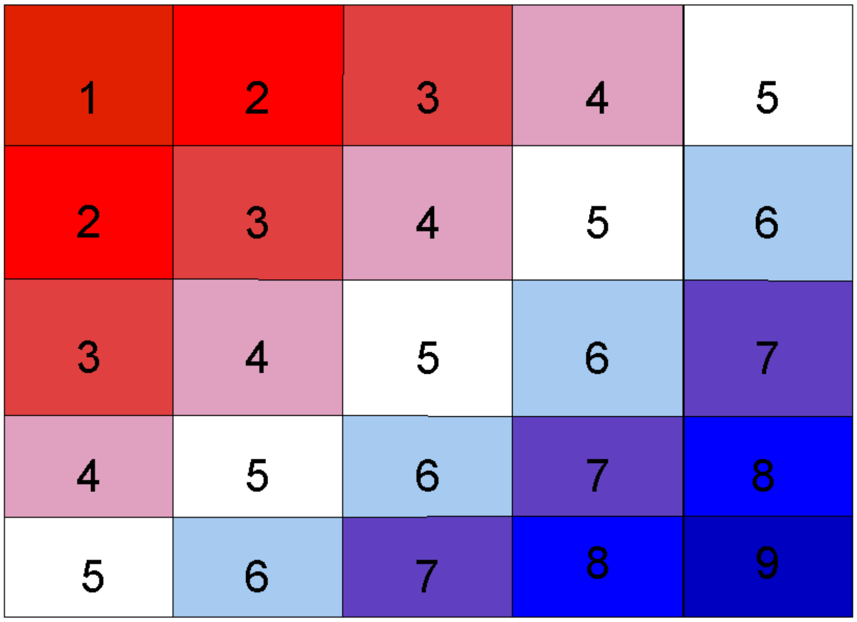


Figure 7: Symbolization of 5x5 matrix using the Positive Correlation (Cor_Pos) attributes with the standard Blue -Red Dichromatic color ramp with 9 equal intervals. The cells labeled five are where the classes are positively correlated. Cells at lower and larger numbers are farther showing depa rture from positive correlation. The cell numbering shows the numbering sequence for Correlation.

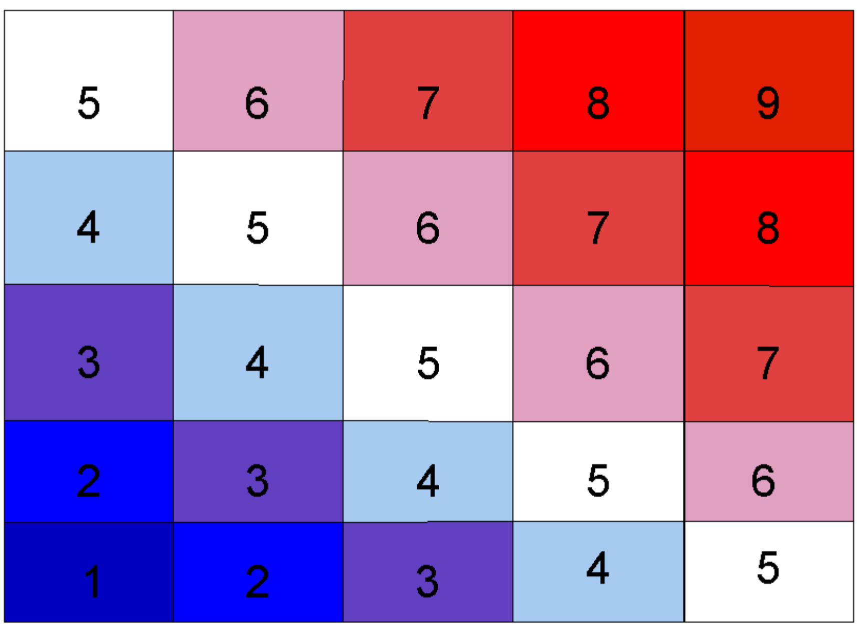


Figure 8: Symbolization of 5x5 matrix using the Negative Correlation (Cor_Neg) attributes with the standard Blue -Red Dichromatic c olor ramp with 9 equal intervals. The cells labeled five are where the classes are negatively correlated.

Correlation Measures

The procedure described above provides a spatial display of the joint variation of two grids. The correlation coefficient can provide a numerical measure of this variation. The Covarian Avenue script (McVay, 1998) calculates the weighted Pearson's Correlation Coefficient between two or more grids. This correlation coefficient is appropriate for interval and ratio measurement scale but not for ordinal-scale data. This script requires that your grid-file names be 9 characters or less. A trap for this problem in Covarian have been made in a revised script called Grid Correlation (GridCorr.ave). **To get correct results with Grid Correla tion, consider the following:**

- The script creates a table with an extension TXT for the output. This table can be directly printed.
- The grids used for this extension are required to have a name with a maximum of 9 characters.
- The script is programmed to use the value attribute for the calculations. Consequently, it may be necessary to create grids with the attributes to be correlated in the value field.
- **If your objective is to measure the Pearson's Correlation Coefficient between posterior probability, favo rability, or membership from an ArcWofE or ArcSDM response grid, there are several steps to create the necessary grids.** The response grid is an integer grid with the model calculations in a table. The table is joined to the VAT of the response grid. Maybe because the calculated values are so small and in a joined table, the Map Calculator will not properly deal with these numbers. Repeat the following process for each grid that is to be included in the correlation measure.
  - Add a new field to the response map VAT to hold the desired model value. Be sure that the new field has sufficient significant figures.
  - Calculate the model value into the new field. This new field is part of the response grid VAT.
  - Create a new floating-point grid from the new field added above with the Map Calculator. For example, a response grid named SDMuc1 with an new attribute of wofePP, the calculation would be **([SDMuc1.wofePP].float)** .
  - Rename this new grid to a meaningful name with a maximum of 9 characters.
- To calculate the correlation coefficient(s), make active all of the floating-valued grids to be used. Then run the extension. The results will be in the working directory in a file **stats#.txt**, where # will be a unique number for each time the script is run.

For tables and shapefiles, the Correlation Coefficient extension (Corrcoef.avx by Frye, 1998) is useful. This extension calculates the Pearson's Correlation Coefficient between two shapefiles or a table.

Details – Shapefiles

The same process can be used for shapefiles except that shapefiles are combined with the union or intersection options in the Geoprocessing Wizard. The shapefile attributes will still need to be reclassified into five classes to provide the 5x5 array. The Xtools/Table Frequency and the editing to have twenty-five records will also be the same for the shapefile table. The Correlation Coefficient extension (Corrcoef.avx by Frye, 1998) is useful to calculate the Pearson's Correlation Coefficient for both tables and shapefiles.

References Cited

ESRI, 1998, Grid transformation tools (transform.avx) sample extension for Spatial Analyst: transform.zip, http://gis.esri.com/arcscripts/scripts.cfm.

Pyle, Vince, and DeLaune, Mike, 1998, Guide to Xtools Arcview Extension: xtoolsmh.zip, http://gis.esri.com/arcscripts/scripts.cfm.

Raines, G.L., Sawatzky, D.S., and Connors, K.A., 1996, Great Basin geoscience data base: U.S. Geological Survey, Digital Data Series 41 (DDS-41), CD-ROM.

Appendix A: Extensions included in zip file

All of these files were obtained from the ESRI page, Welcome to Arcscripts (http://gis/esri.com/arcscripts/scripts.cfm). They are provided by ESRI as freeware to be used with no guarantees.

- Transform.zip – sptrnfrm.avx, no user guide was provided.
- Xtoolsmh.avx and xtools.pdf (Users guide)
- GridAnalyst.zip – GridAnalyst.avx, sptrnsfrm.avx, spgenrlze.avx, and GRID ANALYST EXTENION.doc (Users guide). This is redundant with Transform.zip but provides a users guide and a single useful combination of tools.
- Grdreclsfromleg.ave – an Avenue script to reclassify grids based on the classification in the legend.
- Correlation.zip –
    - covarian.ave – calculates Pearson's Correlation Coefficient for floating grids, no users guide was provided. This script can sometimes not operate correctly and has corrupted the APR. Use with care.
    - tablesstatsmultifield.ave - no users guide was provided
    - Corrcoef.zip – corrcoef.avx – calculates Pearson's Correlation Coefficient for tables and shape files.

Fuzzification – An Arcview Script To Be Used With Fuzzy Logic And Neural Network Applications

Introduction

The fuzzy logic method of spatial analysis requires that the crisp data be scaled into fuzzy membership values, ranging from zero to one. This is a process called fuzzification (Tsoukalas and Uhrig, 1997). Fuzzification is also useful for pre-processing the data for analysis in neural networks. An Avenue script for Arcview called fuzzy.ave that augments the tools in ArcSDM (Kemp and others, 2001) is documented here. Fuzzy.ave implements several algorithms that are in common use in fuzzy-logic applications. The advantage of using an algorithm to transform the crisp measurements into fuzzy membership values is that it makes the transformation repeatable and easy to report. For reporting, it is only necessary to identify the algorithm used and the parameters selected for that algorithm. Additional insights into fuzzy logic and fuzzification can be found in Tsoukalas and Uhrig (1997), Burrough and McDonnell (1998), and Masters (1993).

Fuzzification Algorithms

The algorithms implemented in Fuzzy.ave are the following: Small (Tsoukalas and Uhrig, 1997), Near (Tsoukalas and Uhrig, 1997), Gaussian (Masters, 1993), and Large (Tsoukalas and Uhrig, 1997). These fuzzification algorithms can also be modified, such as very small, with an additional set of algorithms referred to as hedges (Tsoukalas and Uhrig, 1997, Zadeh, 1993).

Fuzzy.ave adds a new attribute containing the attribute to be fuzzified into the active integer grid. The menu for the selection of the parameters is shown in Figure 1. The user selects the algorithm and hedge by typing one of the names in square brackets and the spread and mid values. Spread is a parameter of the fuzzification algorithm that determines how rapidly the fuzzy membership values decrease from one to zero. Mid is the parameter that defines the crisp value. That value will have a membership of 0.5 for small or large fuzzification algorithms or the middle value having the maximum fuzzy membership value for the near and Gaussian fuzzification algorithms. If it is desired to scale the fuzzy membership values from a maximum of less than one, then the fuzzy membership values from these algorithms can always be rescaled, for example multiplication by 0.75 would reduce the maximum fuzzy membership value to 0.75.

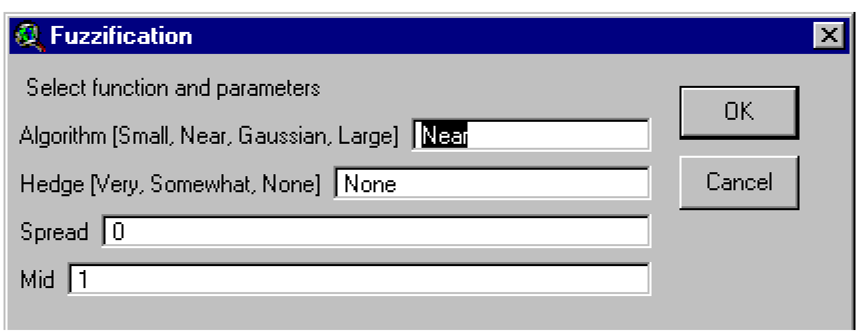


Figure 1: Fuzzification menu for input of fuzzification parameters.

Hedges

The two hedges implemented are *very* and *somewhat* (Tsoukalas and Uhrig, 1997). *Very* is also known as concentration. *Very* is defined as the fuzzy membership function squared. *Somewhat* is also known as dilation or the linguistic term “More or Less”. *Somewhat* is the square root of the membership function. The *very* and *somewhat* hedges decrease and increase, respectively, the fuzzy membership functions.

Small and Large

The fuzzification algorithms small and large are used to indicate that small or large values of the crisp set are members of the fuzzy set. The spread and mid parameters are subjectively defined to reflect the expert opinion. Examples of the small and large functions and hedges are shown in Figure 2. The small fuzzification algorithm is defined as

$$\mu(x) = \frac{1}{1+\left(\frac{x}{f_2}\right)^{f_1}} \qquad \text{(Equation 1: Fuzzy Membership Small)}$$

Where f_1 is the spread of the transition from a membership value of 1 to 0 and f_2 is the midpoint where the membership value is 0.5 (Tsoukalas and Uhrig, 1997).

The large fuzzification algorithm is defined as

$$\mu(x) = \frac{1}{1+\left(\frac{x}{f_2}\right)^{-f_1}} \qquad \text{(Equation 2: Fuzzy Membership Large)}$$

Where f_1 is the spread of the transition from a membership value of 1 to 0 and f_2 is the midpoint where the membership value is 0.5 (Tsoukalas and Uhrig, 1997).

Note this function works improperly for negative crisp values. To apply these functions to negative numbers, the crisp values need to be transformed to positive numbers before fuzzification.

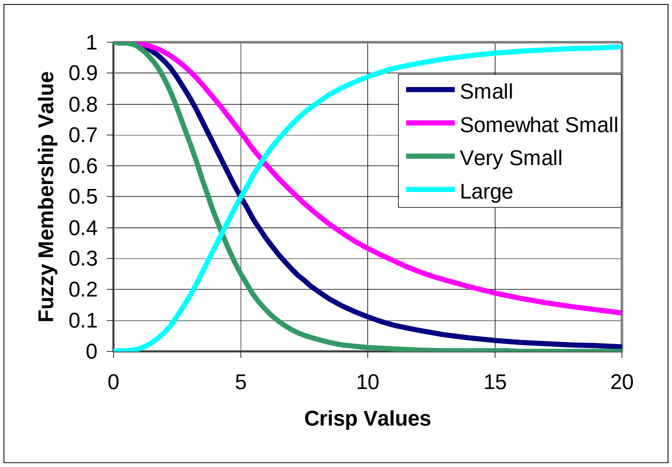


Figure 2: Examples of small and large fuzzification using a mid value of 5 and a spread of 3.

Near

The fuzzification function near is used when some intermediate crisp value is the member of the fuzzy set. The spread and mid parameters are subjectively defined to reflect the expert opinion. An example of the near algorithm is given in Figure 3. The near function is also known as a sinusoidal membership function (Burrough and McDonnell, 1998). The near fuzzification algorithm is defined as

$$\mu(x) = \frac{1}{1 + f_1 (x - f_2)^2} \qquad \text{(Equation 3 : Fuzzy Membership Near)}$$

Where f_1 is the spread of the transition from a membership value of 1 to 0 and f_2 is the midpoint where the membership value is 0.5 (Tsoukalas and Uhrig, 1997).

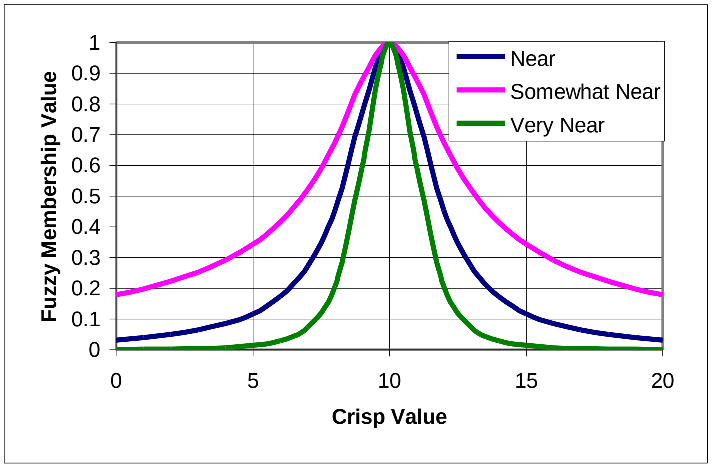


Figure 3: Example of near fuzzification using a mid value of 10 and a spread of 0.3.

Gaussian

The fuzzification function Gaussian is similar to the near function but has a more narrow spread. The near fuzzification algorithm is defined as

$$\mu(x) = e^{-f_1 (x - f_2)^2} \qquad \text{(Equation 4: Fuzzy Membership Gaussian)}$$

Where f_1 is the spread of the transition from a membership value of 1 to 0 and f_2 is the midpoint where the membership value is 0.5 (Tsoukalas and Uhrig, 1997).

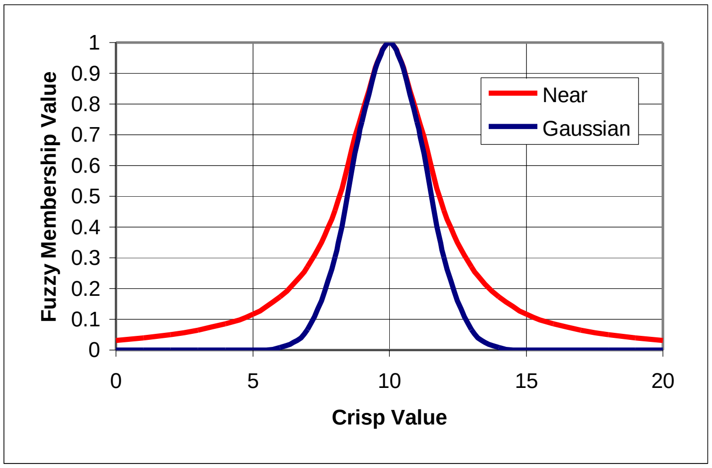


Figure 4: Example of the Gaussian fuzzification function compared to the near function using a spread of 0.3 and a mid value of 10.

Combinations

Combination fuzzification functions can be made by applying multiple fuzzification functions to an integer-grid table and then editing the table to piecewise combine the different functions. An example of such a process using the two near functions is shown in Figure 5

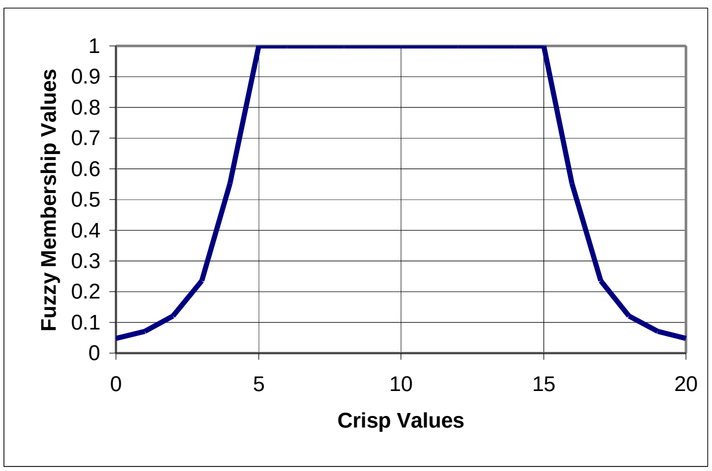


Figure 5: Example of a combination of two near functions with spread of 0.9, a mid value of 5 for less than 5, a mid value of 15 for greater than 15, and a membership value of 1 between 5 and 15.

Spread

The selection of the appropriate spread value is a subjective process that is dependent on the range of the crisp values. A useful way to experiment with different spread values is to use a spreadsheet program with graphs. Then a picture of the effects of different spread functions can be quickly developed. Note, as shown in Figure 6 and Figure 7, as the spread gets smaller the fuzzy memberships approach zero more slowly.

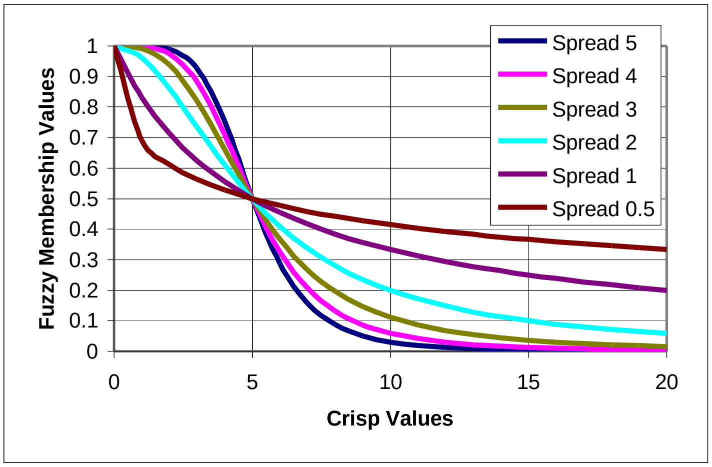


Figure 6: Examples of a range of spreads for a small function with a constant mid value of 5.

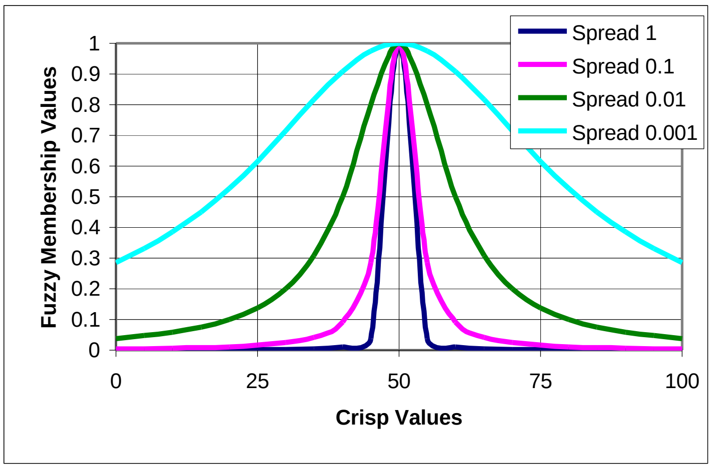


Figure 7: Examples of a range of spreads for a near function with a constant mid value of 50.

References

Burrough, P.A., and McDonnell, R.A., 1998, Principles of geographical information systems: New York, Oxford University Press, 333 p.

Kemp, L., Bonham-Carter, G.F., Raines, G.L., and Looney, C.G., 2001, ArcSDM and DataXplore – spatial data modeler for Arcview and Spatial Analyst: http://ntserv.gis.nrcan.gc.ca/sdm/.

Masters, Timothy, 1993, Practical neural network recipes in C++: New York, Academic Press, p. 285.

Tsoukalas, L.H., and Uhrig, R.E., 1997, Fuzzy and neural approaches in engineering: New York, John Wiley and Sons, Inc., 587 p.

Zadeh, L.A. A computational approach to fuzzy quantifiers in natural languages: Computers and Mathematics, v.9, p. 149-184.

Carlin Exercise

Introduction

This exercise is intended to guide a user through the process of creating a weights-of-evidence (WofE), fuzzy-logic, and two neural-network models using ArcSDM, which is available at http://ntserv.gis.nrcan.gc.ca/sdm/. The user is assumed to have a working knowledge of Arcview, the Spatial Analyst extension, and ArcSDM. All of the materials needed for this exercise are contained in the Carlin zip file. The files in this zip file will extract into a folder called Carlin. If you are copying the Carlin folder from a CD-ROM source, it may be necessary to change the read/write permissions after copying the folder to your disk. To do this, copy the Carlin folder from the source to a root folder. To have appropriate read-write permissions on the files and grids, find the clear.bat file in the carlin folder and double click it to run it. This will change the permissions so you can use the files.

The data that is provided in the Carlin folder is to be used for modeling of Carlin deposits of central Nevada. These data are purposely selected to provide simple evidential layers for learning about the ArcSDM tools, not necessarily to provide the best model of these deposits. This document summarizes the Arcview themes and APR in the Carlin folder. The processing steps to create a WofE model are discussed in detail. Guidance for fuzzy-logic and neural-network models are provided for use after completion of the WofE model. The data source for this exercise is Raines, Sawatzky, and Connors (1996). The user should review the users manual provided with the ArcSDM software to better understand the various menus.

The WofE model is discussed in detail as it provides a foundation for many of the decisions necessary to complete a fuzzy-logic or neural-network model. Fuzzy membership values are often a useful approach to reclassification of categorical data in the neural-network model, as well as - for controlling the number of classes that the neural network has to deal with. The number of classes can significantly influence the time it takes for the neural network to complete classification.

The models are primarily built using geology and antimony evidence. For the WofE model, guidance is given for using proximity to faults as evidence. The following additional data sets are provided for creating models that are more complex: multi-element stream sediment geochemistry, gravity, magnetics, and gamma ray (uranium, thorium, and potassium).

It may sometimes be necessary to change the paths within the carlin.apr. When first used the paths should be Path:"/carlin/. Once you have saved the copied APR file, the paths should be Path: "e:/carlin, if you copied the APR to the e: drive. The path Path:"/carlin/ is a generic path name that will work on any location in the directory structure. If you desire to edit these paths, open the carlin.apr file in a text editor. Search for the string Path: See what path is after the quote and change all occurrences of this string to your desired path. This is an easily way to share APRs.

Arcview project

Carlin.apr – an Arcview project with the data sets loaded and symbolized.

Summarized Metadata

Expert Assessment

An example of a Carlin model made by experts using analog methods.
Expert2 – grid file
This 3-unit grid is provided to give an example of a mineral assessment for Carlin deposits in the study area. It classifies the area into three categories, favorable, permissive, and nonpermissive. Nonpermissive areas are areas where the probability of a deposit is so low that deposits are not expected to occur. Permissive areas are areas where the age and lithology of the rocks are of the character associated with this deposit type. Favorable areas are areas where processes associated with the formation of the deposit type are known to occur. This grid is derived from the USGS National Assessment (Ludington and Cox, 1996).

Study area

The area to be studied and the analysis mask.
Studygrd3 – grid file
Studyarea.shp – shapefile

Training Sites

Defines the locations of known Carlin deposits in the study area. These are used by the supervised methods to make a model. These points are locations of deposits and occurrences that were classified by a group of experts as sediment hosted gold deposits (Carlin deposits).
Train2.shp – shapefile

Evidential themes

These themes are used to predict Carlin Deposits.

Geology

Kbgeol – grid file
This data is 1:2,500,000-scale geology polygons from the King and Beikman map of the United States

Kbgeoltbl.dbf - DBF table of attributes describing some aspects of the geologic map units.
Rockdesc – The name of the geologic map units.
Carlin – this attribute has the value T or F. T indicates that the unit is as older or older than the Carlin deposits. F indicates that the unit is younger than the Carlin deposits. This is used to define which map units might be covering deposits.

Stream Sediment Geochemistry

Naa.shp – shapefile
Source point file for antimony evidential theme. This is part of the NURE stream-sediment geochemistry data. These data are normally considered 1:250,000 scale and the units are parts per million (ppm). The theme consists of a suite of element analyses by neutron activation. A value of zero (0) in this file indicates that the element was not analyzed in the particular sample. The antimony (naa_sb) measurements were used to create sbface1 using inverse distance weighting and system default parameters. Many additional themes for use in models could be created from this shapefile.

Sb Surface (STD)

Sbface1 – grid file

The surface created from the antimony data in Naa.shp. This is a real-number grid that must be reclassified to an integer grid (reclassb2) for use in ArcSDM. The grid is symbolized using ¼ standard deviation classes.

Sb Surface (INT)

Rclassb2 – grid file

The integer grid reclassification of the antimony surface. The reclassification was done using ¼ standard deviation intervals. The 15 values in this grid represent the ¼ standard deviation intervals from 1 to 16, low to high values.

Faults

Gbfaults3.shp – shapefile

This file contains faults shown on the 1:500,000-scale Geologic map of Nevada (Stewart and Carlson, 1978). This digital representation of the faults was created by digitization of the end points of straight-line sections of the faults. The attribute Nhem_az gives the northern-hemisphere azimuth of the faults.

Faults with a northern-hemisphere azimuth near 330 can be buffered with 1000m-wide buffers to define areas proximal to Carlin deposits. Additional azimuthal groupings of faults might be used to define additional evidential themes.

Geophysics

Bouguer – grid file

Bouguer gravity anomaly at 20 milligals contour interval. This file is from Raines, Sawatzky, and Connors (1996). The source gravity data was widely spaced regional measurements.

Aeromag – grid file

Aeromagnetic data from the NURE program. The file is derived from Raines, Sawatzky, and Connors (1996). The source magnetic data were flown with 3-mile line spacing.

Gamma Ray

Uranium – grid file

Uranium gamma-ray data from the NURE program. The file is derived from Raines, Sawatzky, and Connors (1996). The source gamma-ray data were flown with 3-mile line spacing. The units are equivalent uranium.

Thorium – grid file

Thorium gamma-ray data from the NURE program. The file is derived from Raines, Sawatzky, and Connors (1996). The source gamma-ray data were flown with 3-mile line spacing. The units are equivalent uranium.

Potassium – grid file

Potassium gamma-ray data from the NURE program. The file is derived from Raines, Sawatzky, and Connors (1996). The source gamma-ray data were flown with 3-mile line spacing. The units are equivalent uranium.

Instructions for Weights-of-Evidence Model

The user should review the ArcSDM Users Manual to fully understand the menus and functions. The user is assumed to be familiar with the Spatial Analyst functions.

1. Start the Spatial Data Modeler Extension

- In View/Properties set the Map Units to meters and Distance Units to meters or kilometers.

2. SDM/Set Analysis Parameters – use this menu to set up the Analysis Properties and set the modeling parameters.
   - Study Area Grid Theme – select Study Area Mask
   - Training Point Theme – select Training Sites
   - Define Unit Area – select 1 km
   - Missing Data – select -99
   - Select OK
3. There are three evidential data sets provided.
   - Geology – There are two reclassifications of this grid for modeling (Value2 and Fmemship1) Value2 and S_value2 are examples of the reclassification used for ArcSDM. Fmemship1 is an example of fuzzy membership values, which is discussed in the section on fuzzy-logic modeling.
     - There is a data table (kbgeoltbl.dbf) associated with the geology grid that will be used to define map units that are younger than the deposits and therefore potentially covering the map units containing deposits. This table is used to define areas of missing data.
   - Faults – This line theme contains faults and northern hemisphere azimuths so the faults can be selected by azimuth for proximity analysis.
   - Sb Sample Sites – Two grids have been derived from these points, Sb Surface (STD) and Sb Surface (Int).
     - The grid Sb Surface (STD) was made with Surface/Interpolate a Grid using the Inverse Distance Weighting (IDW) and the default parameters to make a floating (real) valued grid.
     - Sb Surface (Int) is a reclassification of the floating Sb grid into integer classes. There are two reclassifications of this grid for modeling (value 5 and Fmemship1). Value 5 and S_value5 are examples of the reclassification used for SDM. Fmemship1 is an example of fuzzy membership values, which is discussed below in the section on fuzzy-logic modeling.
     - **Because Arcview does not fully support long names, the grid Sb Surface (INT) should be renamed. A suggested name is Sbint. Use the Theme/Properties menu to do this.**
4. Analysis of categorical evidential theme (Geology) – the objective is to reclassify the geology into a binary map of areas associated with training sites (inside the pattern) and areas not associated with training sites (outside the pattern). Additionally areas of missing data will be defined using the table kbgeoltbl.dbf.
   - Check the Geology Theme so it is the active theme
   - SDM/Calculate Theme Weights – use this menu selection to explore the association of geologic map units with the training points.
     - Select Evidential Theme – Geology
     - Select Class Field – Value
     - Select Class Descriptor Field – None
     - Check Type of Data – Free
     - Check Write Results to a text file – if desired
     - Calculate Weights, Categorical should be the only option available – check it and the calculation will begin. Save the table in some appropriate place.
     - Respond Calculations of weights for Geology completed. This creates a weights table geology-ct in the tables.

- Open geology-ct (meaning geology categorical weights table) to inspect the contents. It is useful to sort the table on the #points, the number of training points in that map unit.
  - Those map units with contrast greater than zero include more points than expected by chance and are associated with the training sites. Those units with contrast less than or equal to zero are not associated with training sites. Those units that contain no deposits lack a contrast value because contrast cannot be calculated.
- SDM/Generalize Evidential Theme – use this menu to reclassify (generalize) the geology theme to a binary theme based on this contrast information.
  - Select Evidential Theme – Geology
  - Select Class Field – Value
  - Select Class descriptor field – None
  - Select Generalization Method – Define Groups. This method used the query tool to generalize based on information in the geology-ct table.
  - Select Generalize – opens the Group Classes Dialog Box.
  - Group Dialog Box
    - Select Table to Join – geology-ct. This table will be joined to the VAT for Geology and used in the query.
    - Enter New Class Field Name – Value10 (enter a field name not yet used to store an integer value for the binary reclassification). Hit Tab to move to the next field.
    - Enter Class Descriptor Field Name – S_Value10 (a field name not yet used to store a description of what Value10 means. Hit Tab to move to the next field.
    - Enter 1 or 0 in New Class. This is the value for outside the pattern. Hit Tab to move to the next field.
    - Enter Outside in New Class Descriptor. This is a short description defining what the New Class integer value means. Hit Tab to move to the next field.
    - In the Group Definition, select the query builder (hammer symbol) to construct the query. This brings up the standard Query Builder Menu.
      - Create the query [#Points] = 0 and select OK. This will enter this query into Group Definition.
    - Select the Plus button to do this query. In the large box below the Plus button, this query will be listed and #Records = 17. At the bottom of the box Number of records remaining should be 8. The cursor should now be in the New Class box.
    - In the New Class Box enter a 1. Hit Tab to move to the next field.
    - In the New Class Descriptor enter Outside. Hit Tab to move to the next field.
    - In the Group Definition, create the query [Contrast] <= 0 and select the Plus Button. As before, this will enter this new query into the large box. #Records should equal 3 and the Number of records remaining should equal 5. The cursor should be in the New Class box.
    - In the New Class Box enter a 2. Hit Tab to move to the next field.
    - In the New Class Descriptor enter Inside. Hit Tab to move to the next field.
    - In the Group Definition, create the query [Contrast] > 0 and select the Plus Button. #Records should equal 5 and the Number of Records remaining should equal 0. So the reclassification for all of the records has been defined.
    - If you make an entry mistake in any of the queries, highlight that row in the large box. This will activate the X in the lower-left bottom of the Group

Classes Dialog box. Selecting this X will remove this query, which can then be reentered properly.
- Select the Generalize Button to do the reclassification. This will add two new fields to the Geology theme, Value10 and S_Value10, with the generalization information.
- To view the results, use the legend editor to symbolize Geology with S_Value10. Inside the pattern might be colored red and outside the pattern might be colored green.

5. Some of the geologic units are younger than the deposits in Training Set; so these map units should be treated as missing data.
   - Open the kbgeoltbl.dbf file and highlight the Unit field.
   - Open the attribute table for the Geology Theme and highlight the S_Value field.
   - With the attribute table of Geology active, join the kbgeoltbl.dbf.
   - Edit the Attribute Table of Geology, Value10 and S_Value10 fields.
     - Select those records with Carlin = F.
     - For the selected records, calculate Value10 = -99 and S_value10 = "Missing".
     - Stop editing and save the edits. Remove the joins when done.
     - To view the results, use the legend editor to symbolize Geology with S_Value10. Inside the pattern might be colored red, outside the pattern might be colored green, and missing data might be colored blue.

- The results of this reclassification are shown in Figure 1.

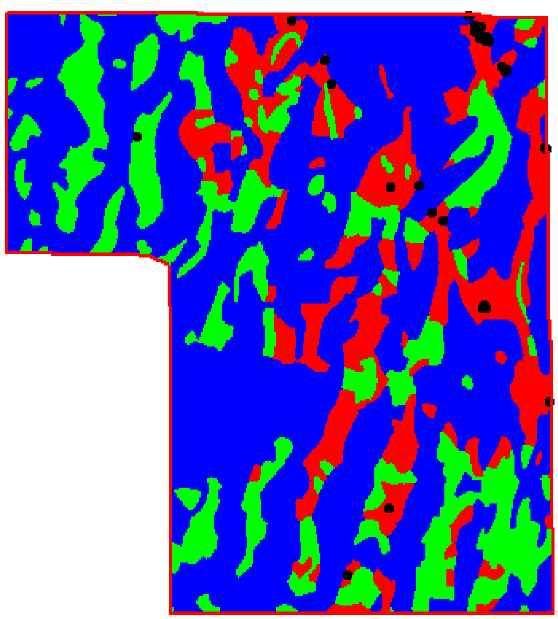


Figure 1: Generalized Geology theme with Training Points.

6. Analysis of ratio data (Antimony, Sb Surface (Int)) – the objective is to reclassify the antimony into a binary map of areas associated with training sites (inside the pattern) and areas not associated with training sites (outside the pattern).

- The Sb Surface(INT) grid can be recreated using the Surface/Interpolate Grid and then Analysis/Reclassify menu selections. An integer Grid is required for the modeling to provide a VAT file to store the generalized binary attributes.
- Check the Sb Surface (INT) grid to make it active.
- SDM/Calculate Theme Weights – use this menu selection to explore the association of Sb Surface classes with the training points.
  - Follow the same procedure as for the Geology Theme except select Type of Data as Ordered and select the Cumulative Descending button. Use this button? because the objective is to define a cutoff of the high values.
  - This will create a table Sb_surface_(INT)-cd.dbf.
  - To inspect the results open the Sb_surface_(INT)-cd.dbf table or better, create a chart. Select SDM/Create Charts. This will create a chart of descending values. Inspect this chart or the table to find the maximum contrast. For the Sb Surface (INT) provide this will be class 10 with a contrast of 3.2. Note the studentized contrast (Stud(C)) value is much larger than 2 so the contrast is significant.
    - In the Charting Parameters Dialog Box, select Table/Class Field sb_surface_(int)-cd, Class, Chart Type Line, What to Plot select Contrast.
- To reclassify the Sb Surface (INT) into binary classes, proceed with SDM/Generalize Evidential Theme as before, except for Generalization Method select Define threshold/Chart, select the table sb_surface_(INT)_cd, and select the Generalize Button. This will bring up the chart previously created and a Generalize Evidential Theme Dialog box. It should select the value and value descriptor fields defined in the Generalize Evidential Themes Dialog Box and have one line in the large box with 1 and 1-16.
- Select the Threshold Selection Tool (the Arrow) and point at the highest value on the graph (Class 10). This will enter a second line into the Generalize Evidential Theme Dialog large box.
- To edit the blank descriptions, highlight the value 1 line. The whole line should be black.
- Put the cursor in the Edit Descrip box, type *Outside*, and then hit enter. This will add the word *Outside* to the Descrip field.
- Now highlight the value 2 line and enter *Inside* to the Descrip field as above.
- Select Generalize to add the generalized attributes to Sb Suface (INT).
- Inspect the generalization by symbolizing Sb Surface (INT) with the descriptor field created by the Generalization. The result is shown in Figure 2.

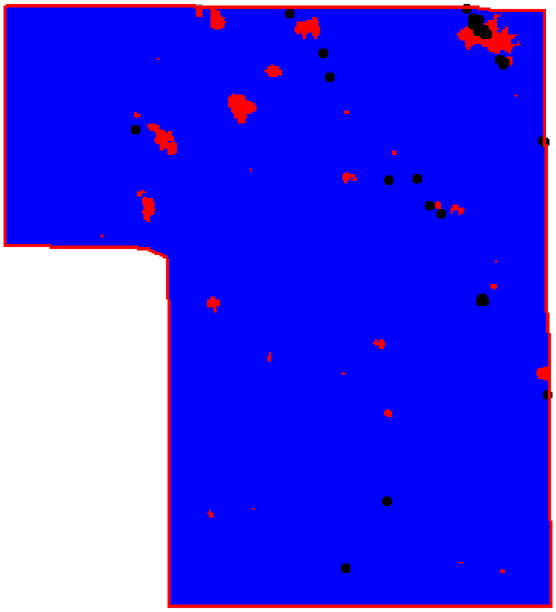


Figure 2: Generalized Antimony surface with Training Points.

7. If a third evidential theme is desired, the faults can be used.
   - Using the query builder, select a subset such as northern hemisphere azimuth (320-360) and make a new theme.
   - Use the SDM/Buffer Features menu selection to create buffers around the faults. A buffer distance of 500 meters, 20 buffers, and checking both options is a good starting place.
   - Rename the buffered grid to a short name, such as fltnw for northwest faults. Do this with Theme/Properties menu.
   - Use SDM/Calculate Theme Weights selecting the Cumulative Ascending Method, SDM/Create Chart, and SDM Generalize Evidential theme as before.

8. To integrate the evidential themes, use SDM/Calculate Response Theme menu. This produces the model shown in Figure 3.
   - **If the Sb Surface (INT) grid has not already been renamed, it must be renamed at this point to a short name, such as Sbint because Arcview does not deal with long names. Use the Theme/Properties menu.**
   - In the Inputs to Weights of Evidence Model Themes Dialog Box, select the evidential themes by highlighting them in the left box and adding them to the right box with the add button.
   - Then select the Specify Fields buttons to select the reclassification attributes desired. If you use the generalized fields already provided, for Sb Surface (INT) select Value5 and for Geology select Value2. Then select OK. This will activate the Calculate Weights button.

- Select the Calculate Weights button. This will create a series of tables as dbf files and the response map grid, which will be named woeuc1 if this is the first you have created. This name means weights of evidence unique conditions #1.
- To the question Do you want to create a table of probabilities to assess Conditional Independence now, select Yes. This will create tables of chi squared values for a pair-wise tests of conditional independence.
- To the question, Do you want to associated conditional probabilities in the response theme with the training points, select Yes. This will ask a question about overwriting RecordID, say Yes.
- A box will then come up with an Assessment of Conditional Independence. If you used only the Geology and Sb Surface generalized as provided, the CI ratio will be 0.97. Select OK to complete this box.
- A box will then inform you that the Calculations are complete for Posterior Probability. Select OK and the symbolized Posterior Probability Map will be added to your view and symbolized.
- It is often necessary to change the number of decimal places for the symbolization of Posterior Probability because these numbers are often very small. Use the Legend Editor and increase the number of decimal places to the maximum using the Classify button.
- The default number of classes is more than are appropriate for this particular model. A smaller number of classes give a more appropriate representation of this model.

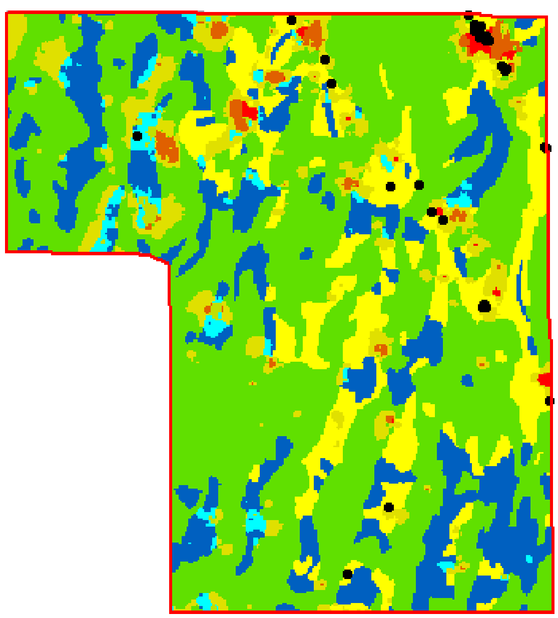


Figure 3: Posterior Probability map with Training points. The training points are shown as black dots. The highest to lowest values are symbolized red, yellow, green, cyan, through blue symbolized by natural breaks.

Guidance for a Fuzzy-Logic Model

The primary decisions when making a fuzzy-logic model are to assign fuzzy memberships to the attributes of the model and to decide which fuzzy operators to apply. ArcSDM provides a tool to help with creation of fuzzy membership values. The fuzzification functions implemented in the fuzzy.ave script provided in the Fuzzification chapter are an alternative approach. The advantage of the fuzzification functions is the fuzzy membership values are exactly reproducible and the process is easily reported.

To activate the fuzzy membership section of SDM, select Set Analysis Parameters, check the Fuzzy Logic box, and select the study area grid. Now the Define Fuzzy Membership menu selection will be active. Selection of this menu leads to a table or graphic tool that assists you to enter the fuzzy membership values.

For gaining experience in selection of fuzzy membership values, the contrast values from the WofE analysis, discussed above, provide useful guidance. For example, a contrast of zero is logically a fuzzy membership value of 0.5. Positive and negative contrast can be rescaled between 0 and 1. For those categorical variables that contain no training points and thus cannot have a contrast value, it is necessary to define a membership value. These categories might be assigned a membership value of zero or 0.5 if the category is a younger map unit that might cover a deposit that is a missing value in the WofE analysis. To select records containing blank numerical fields, a query of the contrast field should use the following format, **([contrast]).IsNull**.

Fuzzy membership values entered manually are included with the geology and reclassified antimony grids. These fuzzy membership values can be used with a fuzzy Or to create the model shown in Figure 4. This fuzzy model is by design similar to the WofE posterior probability (Figure 3). Alternatively, the application of the fuzzification script is described below.

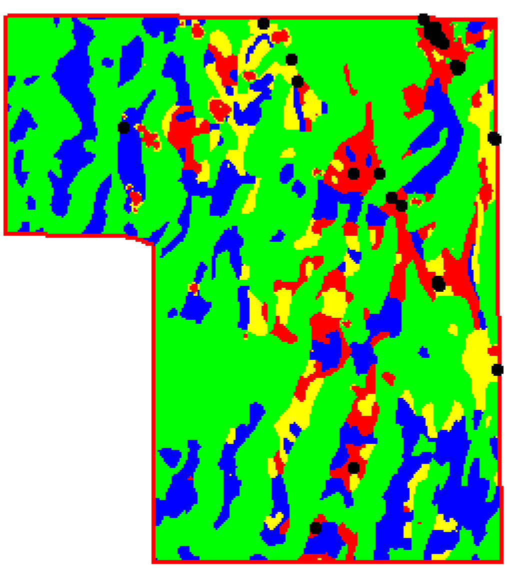


Figure 4: Fuzzy Or model using geology and antimony. The fuzzy membership valu es are Fmembr1 for geology and antimony derived using the SDM manual definition of fuzzy membership. The training points used for the WofE model are shown as black dots. The

highest to lowest values are symbolized red, yellow, green, cyan, through blue usi ng equal intervals.

Fuzzification of Geology Evidence

- Using WofE weights for kbgeol in table kbgeol-ct, join kbgeol-ct to kbgeol with the class and value attributes. Table 1 shows selected columns from these joined tables and the resulting fuzzification. If this table is not available, go back to the WofE modeling exercise above and recreate this table using the calculate weights menu selection.
- In order to use the Large fuzzification function calculated from contrast, it is necessary to have positive numbers and to deal with the classes that contain no training points, that is those classes for which contrast cannot be calculated.
  - Add attribute Rescale to kbgeol to hold the rescale and reclassified Contrast values.
  - Calculate Contrast into Rescale plus the minimum contrast value, 3.2548. Where contrast is blank (null), rescale will be blank. This results in a contrast of zero being rescaled to 3.2548, which will be used for the mid value in fuzzification to give a fuzzy membership value of 0.5. The minimum contrast will be rescaled to 0.
  - These null values are most of the classes that were reclassified as Outside and Missing in the weights-of-evidence analysis.
  - Select those records with the following function **([S_Value2] = “Outside”) and ([Rescale].IsNull)**. Calculate a number near zero into these records. In order to get the fuzzy membership just above zero, I have selected arbitrarily a rescaled value of 0.5.
  - Select those remaining null records with the following function **([S_Value2] = “Missing”) and ([Rescale].IsNull)** . The fuzzy membership value for the classes treated as missing will have a value of 0.5; so calculate into these selected records a value of 3.2548.
- Run the fuzzification script using the Large function with no hedge and with a spread of 3 and mid of 3.2548. The resulting fuzzy membership values are similar to the values manually defined in Fmemshp1. The intent is to calculate fuzzy membership values that reflect how the experts value the geologic map units.

Table 1: Attribute table for kbgeol showing fuzzification based on contrast. The contrast must first be rescaled to positive numbers and the blank contrasts (those classes that have zero training points) must be assigned some rescaled value. Fuz zification parameters for attribute Mbr1 are the following: function = Large, spread = 3, and mid = 3.2548 (equivalent to a contrast of zero). This table is sorted on Mbr1 and S_Value2.

Attributes Of kbgeol joined with kbgeol-ct									
Value	S_value	Value2	S_value2	Fmemshp1	Rescale	Mbr1	Class	No_Points	Contrast
7	TRPE	1	Outside	0.2	0.5	0.004	7	0	
10	LTV	1	Outside	0.2	0.5	0.004	10	0	
12	UPZ	1	Outside	0.2	0.5	0.004	12	0	
13	KG	1	Outside	0.2	0.5	0.004	13	0	
15	P	1	Outside	0.2	0.5	0.004	15	0	
16	JG	1	Outside	0.2	0.5	0.004	16	0	
18	TI	1	Outside	0.2	0.5	0.004	18	0	
19	LMZV	1	Outside	0.2	0.5	0.004	19	0	
21	TRG	1	Outside	0.2	0.5	0.004	21	0	
22	KC	1	Outside	0.2	0.5	0.004	22	0	
23	JMI	1	Outside	0.2	0.5	0.004	23	0	

24	KG2	1	Outside	0.2	0.5	0.004	24	0	
9	LMZ	1	Outside	0.2	2.4473	0.298	9	1	-0.8075
1	Q	-99	Missing	0.53	3.2548	0.5	1	1	-3.2548
3	TPC	-99	Missing	0.53	3.2548	0.5	3	1	-0.19
2	TPF	-99	Missing	0.53	3.2548	0.5	2	0	
6	TMV	-99	Missing	0.53	3.2548	0.5	6	0	
8	TPV	-99	Missing	0.53	3.2548	0.5	8	0	
20	TMF	-99	Missing	0.53	3.2548	0.5	20	0	
25	QV	-99	Missing	0.53	3.2548	0.5	25	0	
14	UPZE	2	Inside	0.7	3.3566	0.523	14	1	0.1018
11	LPZ	2	Inside	0.7	4.6249	0.742	11	4	1.3701
4	C	2	Inside	0.7	4.9122	0.775	4	3	1.6574
17	UPZC	2	Inside	0.7	5.4408	0.824	17	2	2.186
5	LPZE	2	Inside	0.95	6.1754	0.872	5	22	2.9206

Fuzzification of Antimony Evidence

The objective is to calculate fuzzy membership values by fuzzification similar to those manually defined in Fmemshp1, assuming these represent the opinion of the experts.

- Using the reclassified (integer) grid of the antimony evidence, rclssb2, run the fuzzification script with the Large function, no hedge, and a mid value of 9.5.
- Select the Value attribute for the fuzzification. The Value attribute is the reclassification of the antimony by quarter standard deviation classes. So Value 3 is the mean and 16 is more than 3 standard deviations above the mean.
- Mbr1, Mbr2, Mbr3, and Mbr4 are fuzzification for spreads of 3, 6, 12, and 24, respectively.

Table 2: Fuzzification of antimony evidence. Fmemshp1 is an example of fuzzy membership values defined manually. Mbr1, Mbr2, Mbr3, and Mbr4 show examples of different fuzzification

Attributes Of rclssb2							
Value	**Value5**	**S_value5**	**Fmemshp1**	**Mbr1**	**Mbr2**	**Mbr3**	**Mbr4**
1	1	Outside	0.06	0.001	0	0	0
2	1	Outside	0.08	0.009	0	0	0
4	1	Outside	0.12	0.069	0.006	0	0
5	1	Outside	0.13	0.127	0.021	0	0
6	1	Outside	0.16	0.201	0.06	0.004	0
7	1	Outside	0.17	0.286	0.138	0.025	0.001
8	1	Outside	0.19	0.374	0.263	0.113	0.016
9	1	Outside	0.21	0.46	0.42	0.343	0.215
10	2	Inside	0.81	0.538	0.576	0.649	0.774
11	2	Inside	0.84	0.608	0.707	0.853	0.971
12	2	Inside	0.87	0.668	0.802	0.943	0.996
13	2	Inside	0.9	0.719	0.868	0.977	0.999
14	2	Inside	0.94	0.762	0.911	0.991	1
15	2	Inside	0.97	0.797	0.939	0.996	1
16	2	Inside	1	0.827	0.958	0.998	1

Fuzzification Model

To create the fuzzy logic model shown in Figure 5 using geology (kbgeol with Mbr) and antimony (rclssb2 with Mbr4), use the Fuzzy Logic menu selection with an Or operator. Additional evidential layers provided with the exercise could be used to create a more complex model that could involve other types of fuzzification and fuzzy operators. This model is purposely designed to take advantage of what was learned in the WofE model, but in real applications, a fuzzy-logic model would be considered when no training sites are available to develop a WofE model.

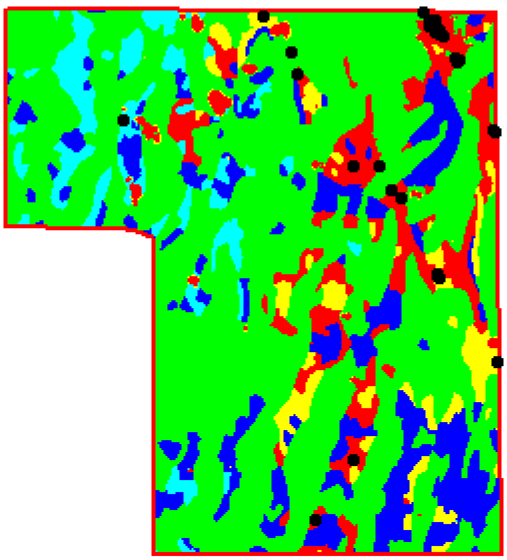


Figure 5: Fuzzy Or model using geology and antimony. The fuzzy membership values ar e Mbr1 for geology and Mbr4 for antimony derived using the fuzzification process. The training points used for the WofE model are shown as black dots. The highest to lowest values are symbolized red, yellow, green, cyan, through blue using equal intervals.

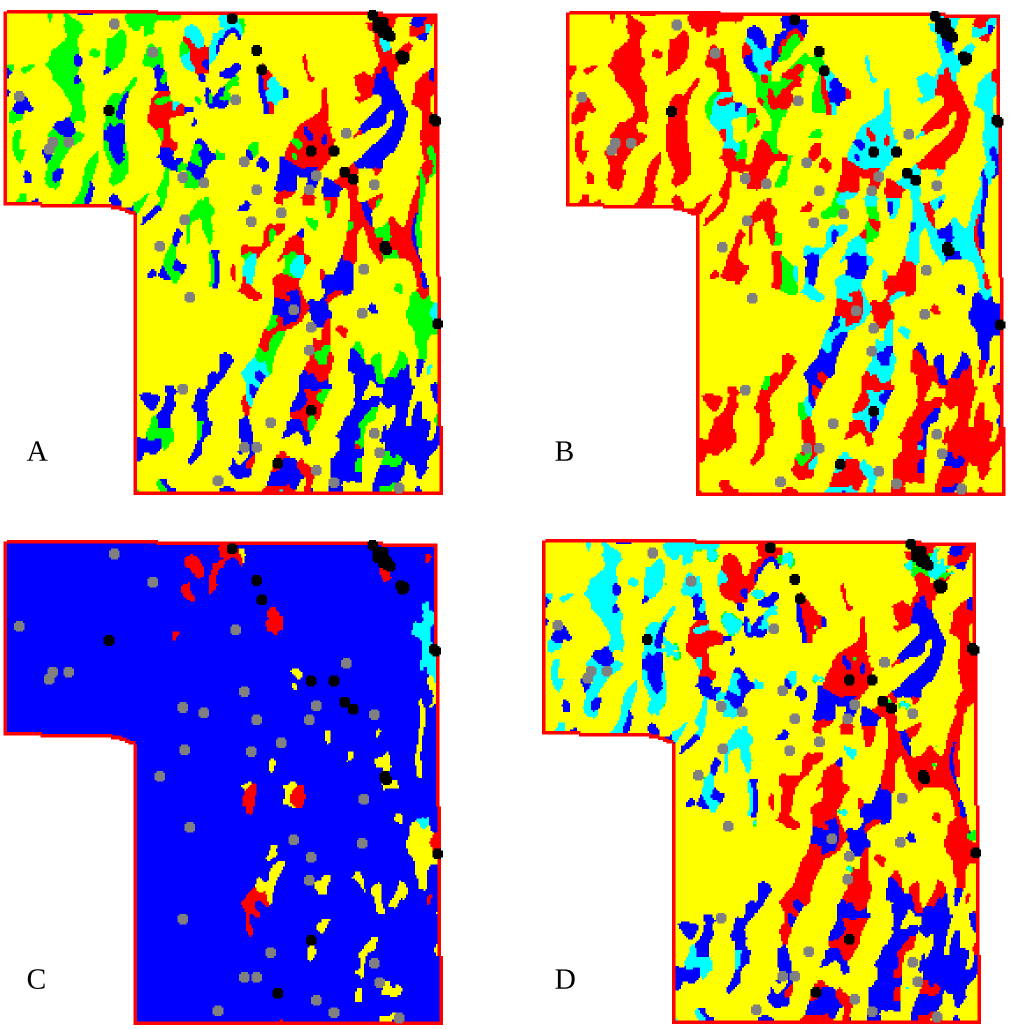
Guidance for a Neural Network Model

Interesting results can be obtained with the neural network by using fuzzy membership values as the inputs. For the neural network tool it is necessary to create a new integer grid with the value field as the input to the neural network. One way to do this is to create a grid from the fuzzy membership values using the Analysis/Map Calculator. You can calculate an integer value from the fuzzy membership value with a calculation such as **([Fmemshp1] * 100.AsGrid).int** . This will create an integer grid with values between 0 and 100. If you want fewer categories, multiply by 10, instead of 100. Before running the model, the input grids should be renamed with short names, as these names will be used in the resulting unique conditions grid.

This use of fuzzy membership can lead to problems in proximity analysis where some of the categories do not contain any training sites. This problem produces a map with zebra stripes of alternating high and low values. One possible solution is to reclassify increasing intervals of the proximity grid into binary grids, where each grid becomes an evidential layer in the neural network analysis. For example, if you buffered faults with 1000m buffers out to 10,000m. You might make a series of binary evidential layers with buffer 1 (1000m) as 1 and everything larger than 1000m as zero. Then buffers 1 and 2 (out to 2000m) would be reclassified as one and everything larger than 2000m as zero; etc. The neural network may then treat these proximity interval binary grids in a more appropriate fashion.

A training set of "non-deposits" is needed for the supervised neural network. One way to do this is provided in ArcSDM. A set of random "non-deposit" training points can be generated with the Spatial Data Modeler/Generate Random Training Points menu selection. This method of create the "non-deposit" training points will create a set of random points within the area defined by some cutoff in the WofE or fuzzy models. For this demonstration, I selected the fuzzification fuzzy model using a threshold of 0.5 for the random "non-deposits. The results of the neural net models using this training set and grids from the fuzzification fuzzy model are shown in Figure 6.

Figure 6: Neural-network models. Models A, B, and C are the three patterns created with the unsupervised (Fuzzy) neural network. Model D is created from the supervised (RBFLN) neural network. All used rescaled fuzzy membership values from the fuzzification fuzzy model; so model A and D are similar to the fuzzy and WofE models. The black and brown points are the deposit and "non-deposit" training sites. The highest to lowest values are symbolized red, yellow, green, cyan, through blue using natural breaks.



References

King, P.B., and Beikman, H.M., 1974, Geologic Map of United States: U.S. Geological Survey, scale 1:2,500,000.

Ludington, S., Cox, D. (editors), and McCammon, D. (project chief), 1996, Data base for a national mineral-resource assessment of undiscovered deposits of gold, silver, copper, lead, and zinc in the conterminous United States by U.S.G.S. Minerals Team: U.S. Geological Survey OFR 96-96, 1 CD-ROM.

Raines, G.L., Sawatzky, D.L., and Connors, K.A, 1996, Great Basin Geoscience data base: U.S. Geological Survey Digital Data Series 41, 2 CD-ROM.

Stewart, J.H. and Carlson, J.E., 1978, Geologic map of Nevada: U.S. Geological Survey, scale 1:500,000.

Reading List

This reading list contains publications important to the rapidly evolving field of spatial analysis, and relevant to students preparing for Masters of Science and Doctor of Philosophy degrees in disciplines involved with spatial modeling problems. The papers are classified for reading as follows: MSc - * and PhD – all. References noted with “&” are not available in the UNR Library.

&An, P., Moon, W.M., and Rencz, A., 1991, Application of fuzzy set theory for integration of geological, geophysical, and remote sensing data: Canadian Journal of Exploration Geophysics, v. 27, p. 1-11.

*Bonham-Carter, G.F., 1994, Geographic information systems for geoscientists – modeling in GIS: New York, Elsevier Science Inc., 398p.

*Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F., 1988, Integration of geological datasets for gold exploration in Nova Scotia: Photogrammetric Engineering and Remote Sensings, v. 54, no. 11, p. 1585-1592.

&Bonham-Carter, G.F., Agterberg, F.P., and Wright, D.F., 1989, Weights of evidence modeling: a new approach to mapping mineral potential *in* Agterberg, F.P, and Bonham-Carter, G.F., (eds.) Statistical applications in the Earth Science: Geological Survey of Canada, Paper 89-9, p. 171-183.

*Burrough, P.A., and McDonnell, R.A., Principles of geographic information systems: New York, Oxford University Press Inc., 333p. (Chapter 11 (p. 265-291) addresses fuzzy logic.)

&Harris, DeVerle, and Pan, Guocheng, 1999, Mineral favorability mapping: a comparison of artificial neural networks, logistic regression, and discriminant analysis: Natural Resources Research, v. 8, no 2., p. 93-109.

*King, J.L., and Kramer, K.L., 1993, Models, facts, and the policy process - the political ecology of estimated truth in Goodchild, M.F., Parks, B.O., and Steyaert, L.T., Environmental modeling with GIS: New York, Oxford University Press, p. 353-360.

*Levin, S.A., 1992, The problem of pattern and scale in ecology: Ecology, v.73, no.6, p. 1943-1967.

*Mensing, S.A., Elston, R.G., Raines, G.L., Tausch, R.J., and Nowak, C.L., 2000, A GIS model to predict the location of fossil packrat (Noetoma) middens in central Nevada: Western North American Naturalist, v. 60, no. 2, p. 111-120.

*&Raines, G.L., 1999, Evaluation of weights of evidence to predict epithermal-gold deposits in the Great Basin of the western United States: Natural Resources Research, v. 8, no. 4, p. 257-276.

&Singer, D.A., and Kouda, Ryoichi, 1999, A comparison of the weights-of-evidence method and probabilistic neural networks: Natural Resources Research, v. 8, no 4., p. 287-293.

Toffoli, Tommaso, and Margolus, Norman, 1987, Cellular automata machines – a new environment for modeling: Cambridge, Mass., The MIT Press, 259p.

*Tsoukalas, L.H., and Uhrig, R.E., 1997, Fuzzy and neural approaches in engineering: New York, John Wiley and Sons, Inc., 587p.

*&Wright, D.F., and Bonham-Carter, G.F., 1996, VHMS favorability mapping with GIS-based integration models, Chisel Lake-Anderson Lake area *in* Bonham-Carter, G.F., Galley, A.G., and Hall, G.E.M., (eds.), EXTECH I: a multidisciplinary approach to massive sulfide research in the Rusty-Lake-Snow Lake greenstone belts, Manitoba: Geological Survey of Canada, Bulletin 426, p. 339-376, 387-401.

Student's Posters

For the major laboratory exercise in the Spatial Analysis class, the students were assigned to prepare a weights-of-evidence model. This exercise was designed to be a group effort and provided the students an opportunity to form and work with a team whose members had diverse expertise and perspectives. The Tahoe Regional Planning Authority (TRPA) provided data for use as evidence, and nesting sites for Spotted Owls, Osprey, and Goshawks for use as training sites. The students could also prepare a model using other data and two groups did this. These students modeled individual parcel evaluation scores (IPES Scores in the TRPA terminology) in the Tahoe Basin and Mayan habitat sites in Belize. All of the models were presented in a poster format that would be appropriate for a technical meeting.

The following three posters were selected and provided with this report as the most outstanding:

- Goshawk Habitat – a model predicting goshawk habitat in the Lake Tahoe Basin. The poster is file UNRgoshawk.rtl.
- Spotted Owl Habitat – a model predicting spotted owl habitat in the Lake Tahoe Basin. The poster is file UCSBspotowl.rtl.
- IPES Scores – a model predicting the IPES scores for individual land parcels in the Lake Tahoe Basin. IPES scores are used by TRPA to determine whether construction can occur on a particular parcel. The post is file IPES.rtl.

The RTL files are the native raster format for the HP large format plotters such as the HP650, HP750, and HP2200 series of plotters. These files are stored in the zip file provided with this report.

One of the most interesting aspects of the animal habitat models was the identification of a spatial association between nesting sites and roads. Nesting sites are preferentially known near roads adjacent to large road-less areas and the interior of the road-less areas were not sampled. This leads to the conclusion that the nesting sites used for training were biased, that is the sampling programs to locate nesting sites did not sample all environments in the basin.