Using Spatial Statistics and Geostatistical Analyst as Educational Tools

By Konrad Dramowicz

Centre of Geographic Sciences

Lawrencetown, Nova Scotia, Canada

ESRI User Conference, San Diego, California

July 13, 2010

Presentation Outline

- Centre of Geographic Sciences
- Spatial Analysis and Geostatistics Course
- Learned skills
- Evolution of software tools used
- Student work samples
  - Prediction, error and probability surface maps
  - Point pattern analysis
  - Spatial autocorrelation
  - Hot spot analysis
  - Measuring spatial distribution
  - Geographically weighted regression

Centre of Geographic Sciences

- Located in Lawrencetown, Nova Scotia, Canada
- Population: ~ 300
- Number of students: 170 – 180
- Number of COGS alumni currently working for ESRI: 50
- Number of faculty: 18
- ESRI software user since 1983
- GIS program since 1985
- New Advanced Diploma in Geographic Sciences with concentration in GIS, Remote Sensing or GIS for Business will be launched in 2010/2011

Spatial Analysis and Geostatistics Course

- The *Spatial Analysis and Geostatistics* course has been offered at COGS under different names since 1991
- The course has been delivered as an elective for students from GIS, GIS for Business, and Remote Sensing Advanced Diploma Programs
- The number of course participants has varied from 3 to 12

Spatial Data Analysis Module

- Data Analysis module is mandatory for all students (30-45) taking the Advanced Diploma in Geographic Sciences program in their first term
- This module includes exploratory data analysis, geographically weighted regression, and some fundamental multivariate statistics in a spatial context

Necessary Basic Statistical Skills

- Using descriptive statistics
- Applying regression and curve fitting
- Understanding normal and *t* distributions
- Doing cluster analysis
- Comprehending statistical decision theory
  - Significance and confidence
  - Critical values
  - Testing statistical hypotheses

Using Statistical Packages with ArcGIS

- The following links were the most common for ESRI users who wanted to add more statistical functionality to GIS
  - SAS Bridge (ArcGIS)
  - S-Plus for ArcView (ver. 3.0 – 3.2)
- In this course, IBM PASW software package (formerly known as SPSS), was used and data were passed to and from ArcGIS using an intermediate file format (*.dbf)

- Also, students learn how to perform statistical analysis using MS Excel (low cost solution)

Course Learning Outcomes

- The learners completing the Spatial Analysis and Geostatistics course demonstrate the ability to:
  - Select the most appropriate tool for solving a given problem
  - Map and interpret results
  - Think spatially while setting model parameters and interpreting results
  - Properly use fundamentals of statistics as a foundation for spatial analysis and geostatistics

Course Textbooks

- Mitchell Andy, 2005. *The ESRI Guide to GIS Analysis*; *Volume 2: Spatial Measurements & Statistics*, ESRI
- De Smith Michael J., Michael F. Goodchild, Paul A. Longley, 2009 (3rd edition). *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, Matador

- Johnston Kevin *et al*, 2001. *Using ArcGIS Geostatistical Analyst*, ESRI

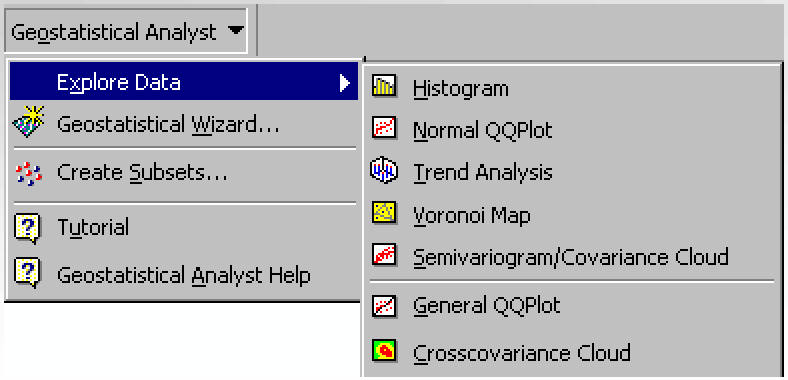
Changes in Software Used

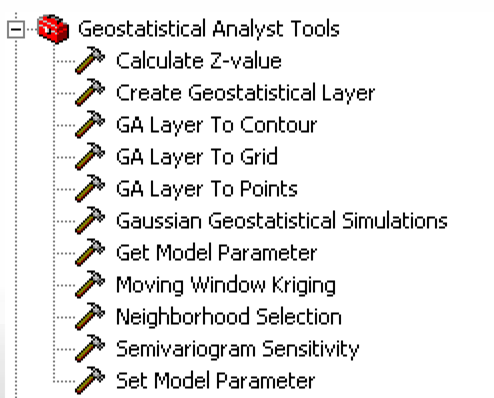
- ArcInfo Workstation → ArcGIS
- GS+ → Geostatistical Analyst
- IDRISI → Spatial Statistics Toolset
- AML → Python (with Tkinter)
- SpaceStat → GeoDa / Spatial Statistics Toolset
- ArcView / Avenue (Wong & Lee scripts) → ArcGIS / Python

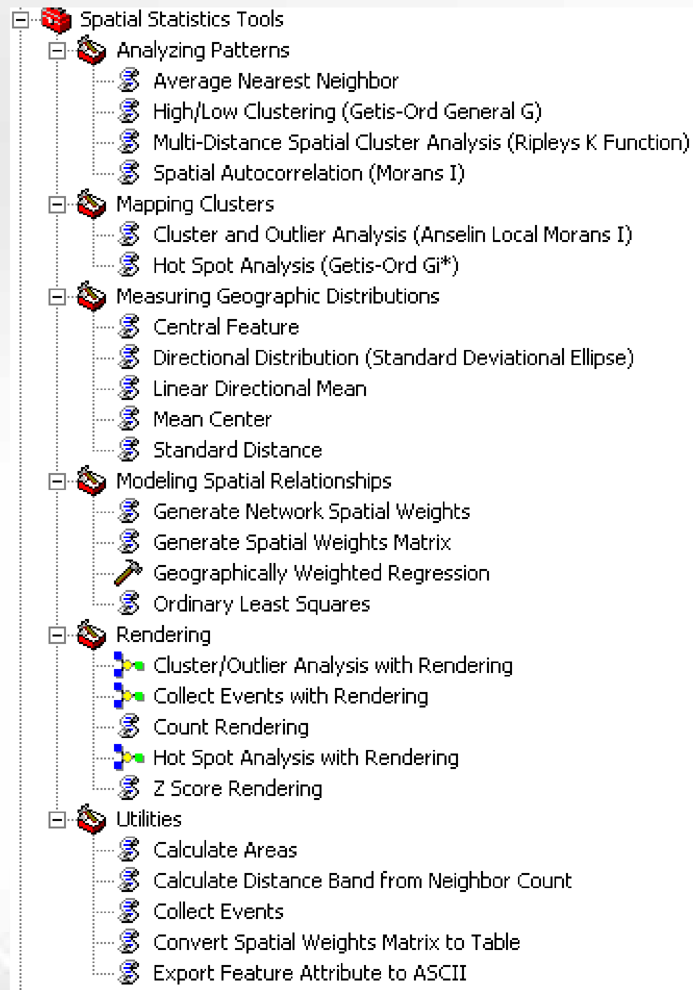
- MapInfo / MapBasic → ArcGIS / Python
- Hawth's Tools
- Fragstats

Main ArcGIS Tools Used in the Course

- Geostatistical Analyst and tools
- Spatial Statistics Tools







Recent Evolution of Geostatistical Tools

- Geostatistical Analyst: ArcGIS 8.1, April 2001
- Geostatistical Wizard enhancements, nine new geostatistical tools: ArcGIS 9.2, September 2006
- Geostatistical Analyst supported by ArcGIS Server, animated tutorial: ArcGIS 9.3, June 2008
- Eleven new geoprocessing tools, improvements to the Geostatistical Analyst Wizard: ArcGIS 10.0, Fall 2010

Recent Evolution of Spatial Statistics Tools

- ArcToolbox Window with eighteen Spatial Statistics tools (10.7% of all available tools: ArcGIS 9.0, May 2004
- Two more tools (including *Ripley's K Function)*: ArcGIS 9.2, September 2006
- New *Geographically Weighted Regression* tool: ArcGIS 9.3, June 2007

- New *Ordinary Least Squares* tool: ArcGIS 9.3.1, May 2009
- One new and eight improved tools: ArcGIS 10.0, Fall 2010

Geostatistics: Basic Topics Taught

- Regionalized variable
- Deterministic vs. geostatistical interpolation techniques (algorithms, examples, advantages and disadvantages)
- Global and local trends, neighborhood search, anisotropy
- Kriging methods, semivariance models, cokriging
- Transformations, cross-validation, bivariate distribution, detrending, declustering
- Gaussian simulations
- Prediction, error, and probability maps

Basic Properties of Regionalized Variable

- Applied for continuous spatial distribution
- Its nature is mixed: somewhere between deterministic and stochastic
- Used for modeling very complex processes

Major Advantages of Using Kriging

- Exact interpolation
- Standard error maps come as a derivative
- Allows the user to control numerous parameters

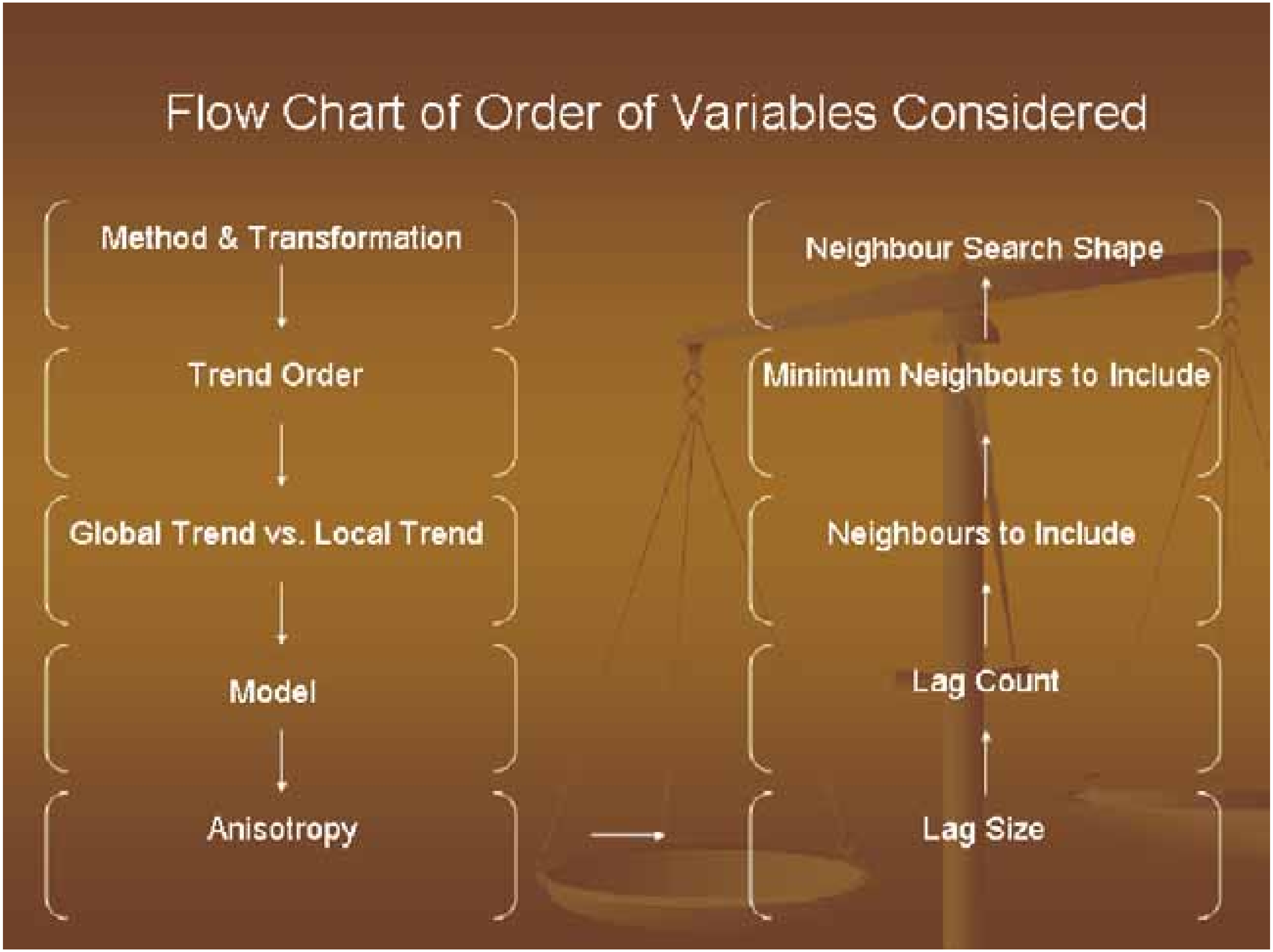
Controlling Kriging Parameters

- There are many kriging parameters to control and students are asked to control one parameter at time
- The optimal settings are those which minimize standard errors (root-mean-square, average standard error, or mean standardized)

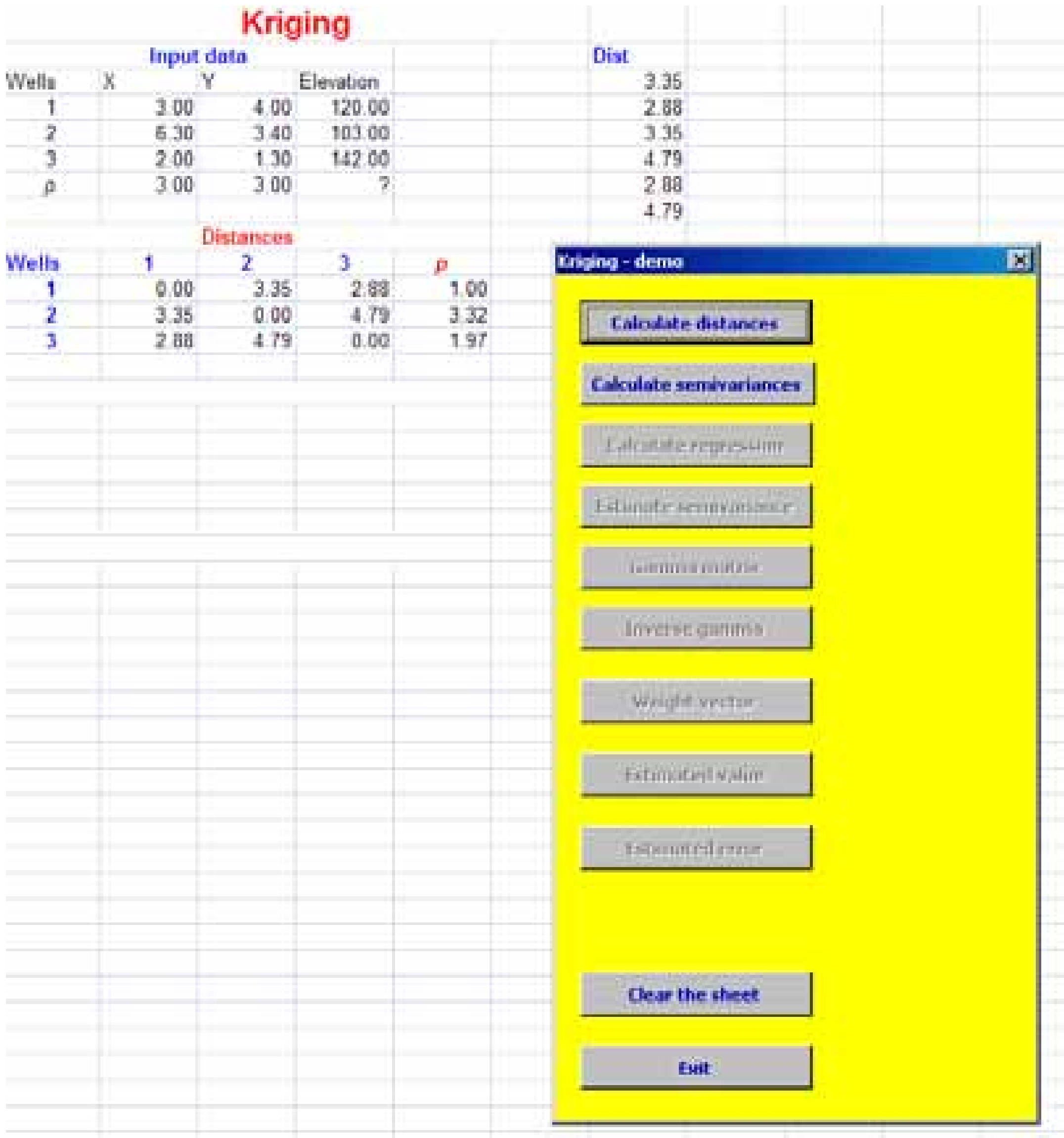
Some Kriging Parameters to Control

- Kriging methods (7)
- Transformations (4)
- Semivariance models (7)
- Trend order (5)
- Anisotropy (individual values)
- Nugget (individual values)

- Number of histogram lags and their size
- Neighborhood search (number of neighbors, minimal number, four search shapes)


Flow Chart of Order of Variables Considered
Method & Transformation
Trend Order
Global Trend vs. Local Trend
Model
Anisotropy
Lag Size
Lag Count
Neighbours to Include
Minimum Neighbours to Include
Neighbour Search Shape

Excel Tool for Teaching Kriging



Critical Opinions About Geostatistics (1)

GEOSTATSCAM.com

Geostatistics: From human error to scientific fraud

Home

What's wrong with geostatistics?

Sampling paradox
Kriging game
Kriged estimate
Missing variance
Kriging Variance

Geostatistical factoids

Spatial dependence
Functional dependence
Degrees of freedom
Interpolation or extrapolation
Tolstoy Syndrome
Silent pundits
Who's what and where?

Statistical facts

Variance of a set
Variance of a central value
Degrees of freedom
Counting degrees of freedom
Test for spatial dependence

The Bre-X fraud

Salted boreholes
Salted boreholes and kriged boreholes
Intrinsic variance

Dr Frits Agterberg, IAMG's Past President and NRCan's Emeritus Scientist, has not explained why he stripped the variance off his distance-weighted average point grade in 1970 and in 1974.

Dr Roussos Dimitrakopoulos, McGill's miracle man and IMG's Editor-in-Chief, talked about stochastic mine planning optimization at APCOM 2009. He should apply it to Gemcom's gold resource for Busang and see what sort of mining plan his voodoo variances cook up.

What kept Bre-X humming along was to assume gold mineralization between salted boreholes, to krige and smooth a little, and to rig the rules of statistics a lot. The National Securities Regulator should rule that unbiased confidence limits for metal contents be reported.

Matheron, the self-made wizard of odd statistics, lost the variance of the lenght-weighted grade of core samples in 1954, and the variance of the length-weighted block grade in 1960!

Photo by Nick Didlick

Jan W Merks
metrologist, author, lecturer, consultant, whistleblower, 'gadfly', 'pariah', 'iconoclast', CIM Life Member

Dr F P Agterberg lost the variance of his distance-weighted average point grade first in his 1970 paper and once more in his 1974 *Geomathematics*. He did refer to Fisher's F-test elsewhere in this textbook but didn't know how to test for spatial dependence between core samples in a borehole, or between boreholes on a line. He ought to delete *Chapter 10 Stationary Random Variables and Kriging*. How long does it take to right a wrong? Why does he assume that his distance-weighted average point grade needs no variance? It's high time to explain why!

In November 1989, we applied Fisher's F-test in Precision Estimates for Ore Reserves to confirm spatial dependence between gold grades of ordered rounds in a decline. Testing for spatial dependence troubled Professor Dr M David, CIM Bulletin's most dedicated enforcer of Matheronian geostatistics. Scores of similarly gifted geostatistocrats postulated spatial dependence may be inferred unless proven otherwise. All sorts of degrees of freedom fighters were troubled when "classical Fischerian [sic] statistics" proved spatial dependence between gold grades of ordered rounds. Fisher's F-test also proved the intrinsic variance of Bre-X's bogus gold in Busang's barren rock to be statistically identical to zero, as it ought to be. Sound statistics did so several months before Bre-X's boss salter vanished! Bre-X's original and duplicate bogus assays for a few early boreholes could have proved early on that a salting scam was in progress at the Busang project!

Geostatistics or voodoo science?

by Jan Merks
Special to The Northern Miner

Guest Column

The objective of mineral exploration is to find ore deposits, and to estimate grades and contents in an unbiased manner and with a reliable, realistic and affordable degree of precision. The term "unbiased" implies that the estimated grade and content of an ore deposit are statistically identical to its unknown true grade and content. A bias or systematic error in the measurement chain distorts the metal grade and content of an ore deposit.

Kriged variances are mathematically invalid and are useless for porphyry deposits, meaningless for stratabound deposits and even dangerous for gold deposits.

Author M. David cautions that his book, "Geostatistical Ore Reserve Estimation," is not for professional statisticians and correctly predicts that they will find many unqualified statements.

Kriging attempts to make a few drill holes go a long way. It creates ore grade values in domains where grades and variances cannot possibly be estimated with a realistic degree of confidence, and where discontinuities in mineralization are bound to occur. Thus kriging tends to inflate expectations for the continuity of mineralization between measured data points with often devastating consequences.

Covariances and kriged variances dominate geostatistics. Yet dependencies and degrees of freedom, concepts even more fundamental than variances, seem irrelevant to the geostatistical theorist. Perhaps, not surprisingly, the geostatistical theorist likes to fit spherical, exponential or linear models to the variance terms for ordered data sets. After all, continuous functions that obscure mundane matters such as dependencies and degrees of freedom most effectively, make a fitting choice!

Some geostatisticians seem cautiously concerned. For example, M. Armstrong and N. Champigny, (*CIM Bulletin*, March 1989), display smooth plots of kriged variances plunging to zero in chaotic domains. The authors caution in the abstract that "the oversmoothed estimates should not be used for calculating recoverable reserves" which is quite understandable.

Kriged variances converge to zero exponentially so that only a few fabricated data points are required to generate suspiciously low variances in small blocks. Zero variances, which permit perfectly precise grade predictions, even geostatisticians find hard to believe.

In another paper, David, too, displays his fondness for smoothing in "Grade control problem dilution and geostatistics: choosing the required quality and number of samples for grade control" he shows how to predict recoveries and grades by entering differences between estimated variances and kriged variances into smoothing relationships that some computer programs so conveniently provide.

The estimated and kriged variances turned out to be statistically identical which implies that their differences are random numbers. Applying mathematical analysis, or just a little smoothing, to that class of random numbers is an exercise in futility and a typical example of voodoo statistics of the worst kind. Nonetheless, recoveries are predicted with one decimal point and grades with two.

Kriged variances inflate expectations for the continuity of mineralization between measured data points, and cannot possibly deliver the reliable and realistic precision estimates that ore reserves demand. However, a technique that replaces a bulldozer's crooked blade with a surgeon's scalpel has already been developed and published.

The "grade-squared partition technique" merges Matheronian basics with the conventional block design and proven statistical techniques that take full advantage of the additive property of variances for grades and contents. Not only does it provide reliable and realistic precision estimates but it is also an effective tool to optimize exploration programs. It would allow geologists to apply statistical tests and techniques in a spreadsheet environment on laptops and portables in the field, and combine geological interpretations and applied statistics most effectively.

— *Jan Merks is president of Matrix Consultants in Vancouver.*

The Northern Miner April 20, 1992

Critical Opinions About Geostatistics (2)

Geostatistics or voodoo science?

by Jan Merks
Special to *The Northern Miner*

Guest Column

The objective of mineral exploration is to find ore deposits, and to estimate grades and contents in an unbiased manner and with a reliable, realistic and affordable degree of precision. The term "unbiased" implies that the estimated grade and content of an ore deposit are statistically identical to its unknown true grade and content. A bias or systematic error in the measurement chain distorts the metal grade and content of an ore deposit.

Kriged variances are mathematically invalid and are useless for porphyry deposits, meaningless for stratabound deposits and even dangerous for gold deposits.

Author M. David cautions that his book, "Geostatistical Ore Reserve Estimation," is not for professional statisticians and correctly predicts that they will find many unqualified statements.

Kriging attempts to make a few drill holes go a long way. It creates ore grade values in domains where grades and variances cannot possibly be estimated with a realistic degree of confidence, and where discontinuities in mineralization are bound to occur. Thus kriging tends to inflate expectations for the continuity of mineralization between measured data points with often devastating consequences.

Covariances and kriged variances dominate geostatistics. Yet dependencies and degrees of freedom, concepts even more fundamental than variances, seem irrelevant to the geostatistical theorist. Perhaps, not surprisingly, the geostatistical theorist likes to fit spherical, exponential or linear models to the variance terms for ordered data sets. After all, continuous functions that obscure mundane matters such as dependencies and degrees of freedom most effectively, make a fitting choice!

Some geostatisticians seem cautiously concerned. For example, M. Armstrong and N. Champigny, (*CIM Bulletin*, March 1989), display smooth plots of kriged variances plunging to zero in chaotic domains. The authors caution in the abstract that "the over-smoothed estimates should not be used for calculating recoverable reserves" which is quite understandable. Kriged variances converge to zero exponentially so that only a few fabricated data points are required to generate suspiciously low variances in small blocks. Zero variances, which permit perfectly precise grade predictions, even geostatisticians find hard to believe.

In another paper, David, too, displays his fondness for smoothing in "Grade control problem dilution and geostatistics: choosing the required quality and number of samples for grade control" he shows how to predict recoveries and grades by entering differences between estimated variances and kriged variances into smoothing relationships that some computer programs so conveniently provide.

The estimated and kriged variances turned out to be statistically identical which implies that their differences are random numbers. Applying mathematical analysis, or just a little smoothing, to that class of random numbers is an exercise in futility and a typical example of voodoo statistics of the worst kind. Nonetheless, recoveries are predicted with one decimal point and grades with two.

Kriged variances inflate expectations for the continuity of mineralization between measured data points, and cannot possibly deliver the reliable and realistic precision estimates that ore reserves demand. However, a technique that replaces a bulldozer's crooked blade with a surgeon's scalpel has already been developed and published.

The "grade-squared partition technique" merges Matheronian basics with the conventional block design and proven statistical techniques that take full advantage of the additive property of variances for grades and contents. Not only does it provide reliable and realistic precision estimates but it is also an effective tool to optimize exploration programs. It would allow geologists to apply statistical tests and techniques in a spreadsheet environment on laptops and portables in the field, and combine geological interpretations and applied statistics most effectively.

— *Jan Merks is president of Matrix Consultants in Vancouver.*

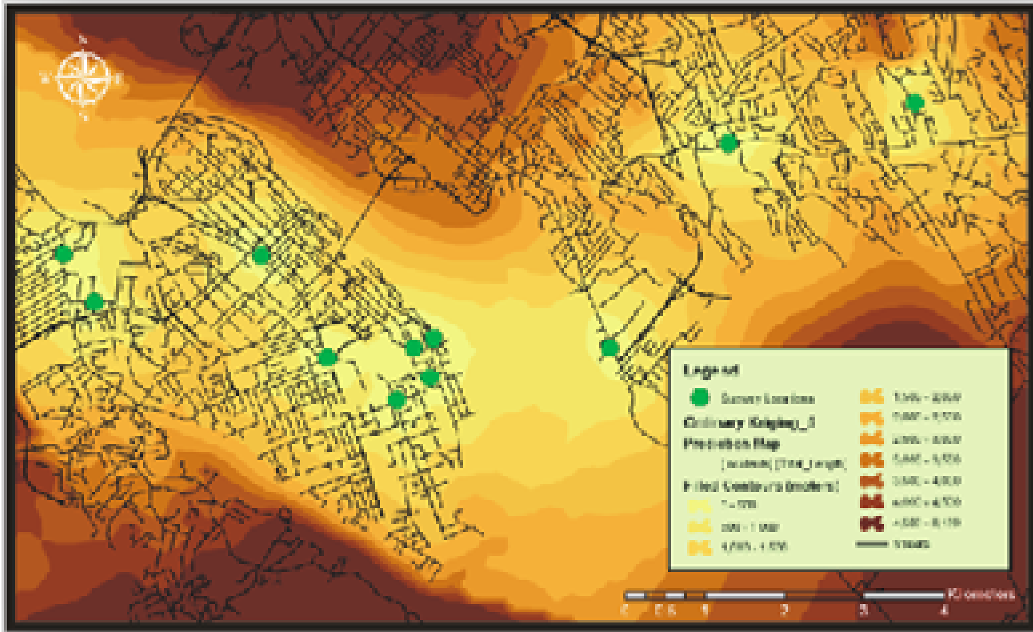
The Northern Miner April 20, 1992

Student Work Samples: Prediction Surface Maps

- Driving time to Subways restaurants in Halifax [1]
- *Driving time to Nova Scotia Breast Screening Program's clinics*
- *Prediction and standard error surfaces: number of houses sold per km sq in Halifax Regional Municipality in 2009*
- *Spending on unhealthy products*
- *Number of recorded GPS locations of tourists in Kejimkujik National Park per one grid cell*

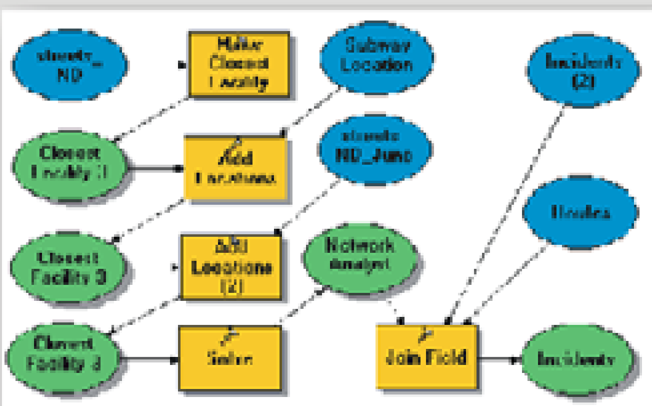
[1] Only one item from the list of student work samples is provided with this document. However, all of them they will be demonstrated at a time of presentation

Isochrones Mapping

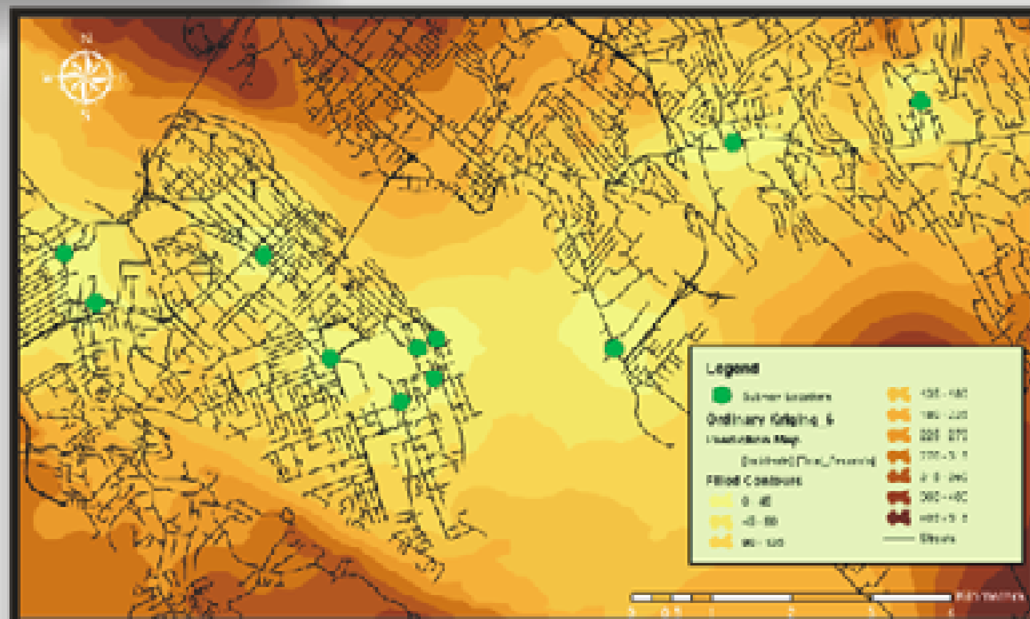


This assignment draws on Network Analyst, Geostatistical Analyst and geoprocessing using model builder to derive to distance models for Subway locations in Halifax. Through the Halifax street network, routes were calculated from every junction to the closest Subway location within the network.

By calculating fields based on the solved network analyst layer, appropriate time and length values were derived. Ordinary kriging was then used to calculate the drive time and length isochrone maps shown below.



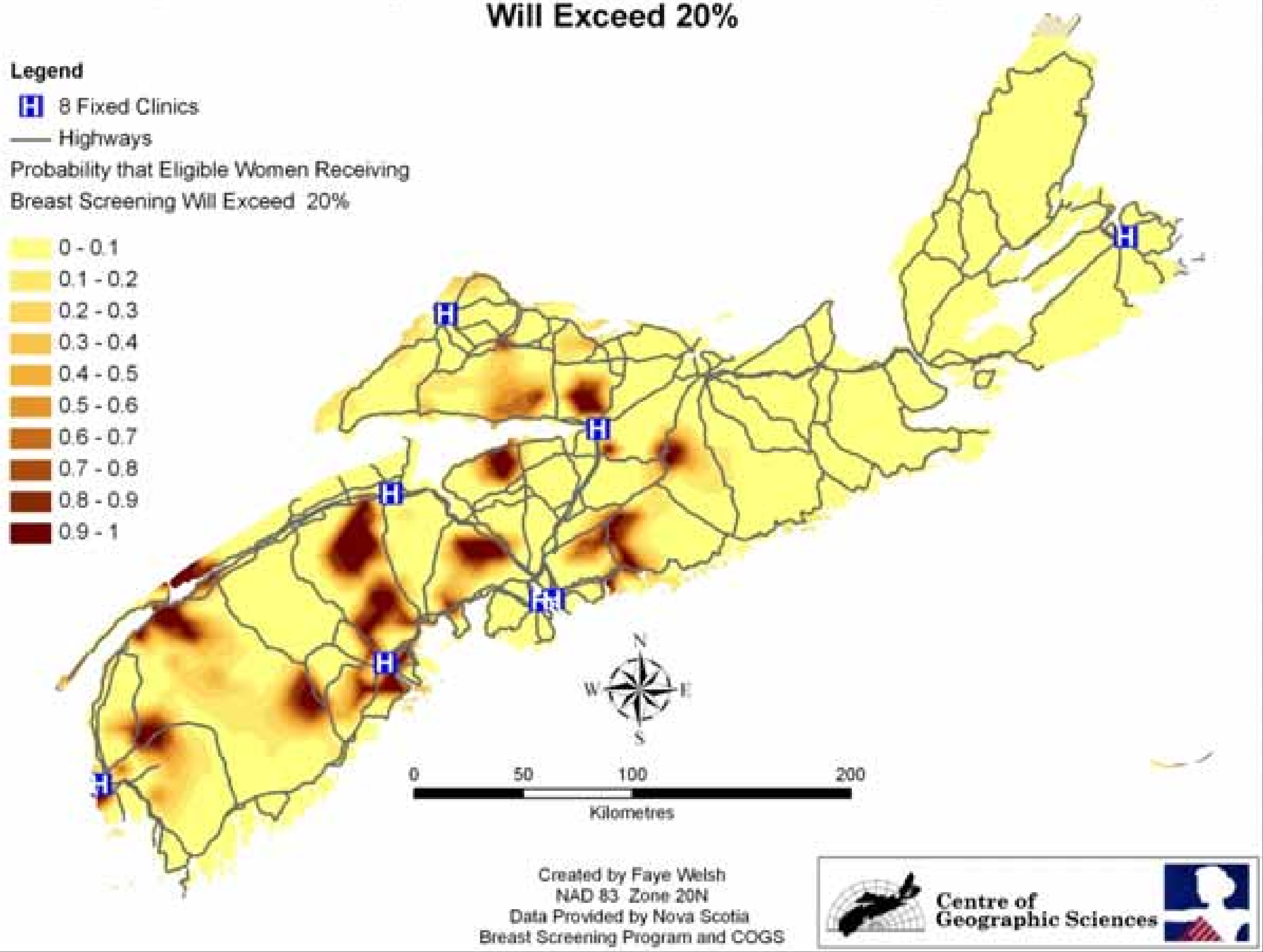
Model to find closest layer



Student Work Samples: Probability Surface Maps

- Percentage of woman aged 40-60 receiving breast screening exceeds 20%
- *Percentage of spending on unhealthy products exceeds 17%*
- *House prices exceeding $150,000 in Halifax Regional Municipality 1982-1986*
- *House prices exceeding $150,000 in Halifax Regional Municipality 1997-2001*

Probability that Percentage of Women Aged 40 - 69 Receiving Breast Screening Will Exceed 20%



Spatial Analysis: Basic Topics (1)

- Point pattern analysis
  - Nearest neighbor analysis (Spatial Statistics tool)
  - Quadrat analysis (implemented by students using geoprocessing and scripting)
  - *K* and *L* functions (Spatial Statistics tool)
- Spatial autocorrelation
  - Moran (Spatial Statistics tool)
  - Geary (ArcInfo Grid)
  - Local Moran (Spatial Statistics tool)

Spatial Analysis: Basic Topics (2)

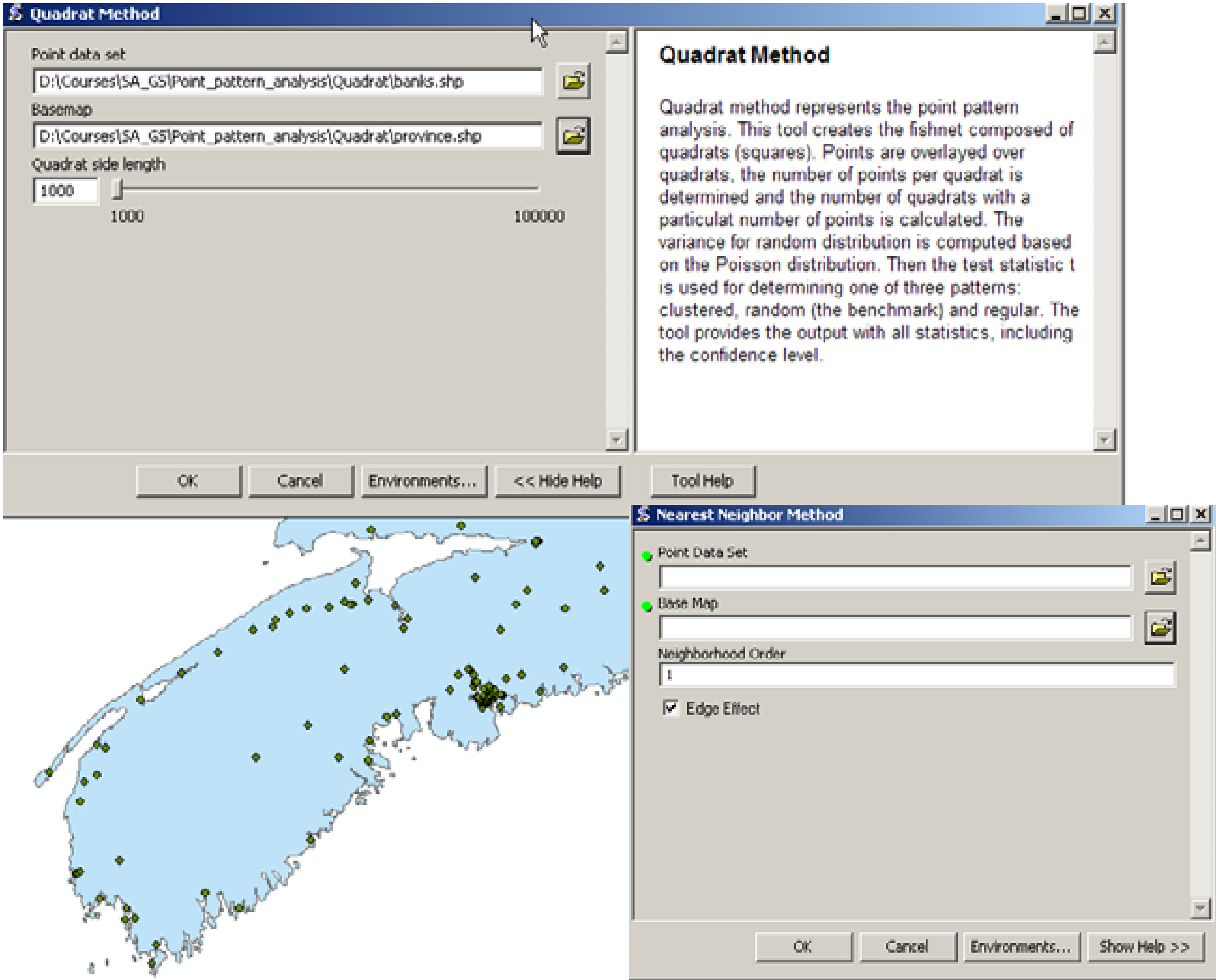
- G statistic
  - General G statistic
  - Hot spot analysis
- Measuring geographic distributions
  - Mean center and central feature
  - Standard distance and standard deviational ellipse
  - Linear directional mean

Implementing Quadrat Analysis

- Use multiple point data sets
- Determine optimal quadrat size
- Calculate lambda (average number of points per quadrat), variance
- Use *t* statistic, various significance levels, determine point pattern
- Analyze relationships between quadrat size and point pattern

Enhancing Nearest Neighbor Analysis

- Use multiple point data sets
- Apply various neighborhood orders (1-6)
- Apply edge effect correction factor
- Analyze relationships between neighborhood order and *z* statistic
- Determine pattern for various significance levels


Quadrat Method
Point data set
D:\Courses\SA_GS\Point_pattern_analysis\Quadrat\banks.shp
Basemap
D:\Courses\SA_GS\Point_pattern_analysis\Quadrat\province.shp
Quadrat side length
1000
1000
100000
Quadrat Method
Quadrat method represents the point pattern analysis. This tool creates the fishnet composed of quadrats (squares). Points are overlayed over quadrats, the number of points per quadrat is determined and the number of quadrats with a particular number of points is calculated. The variance for random distribution is computed based on the Poisson distribution. Then the test statistic t is used for determining one of three patterns: clustered, random (the benchmark) and regular. The tool provides the output with all statistics, including the confidence level.
OK
Cancel
Environments...
<< Hide Help
Tool Help
Nearest Neighbor Method
Point Data Set
Base Map
Neighborhood Order
1
Edge Effect
OK
Cancel
Environments...
Show Help >>

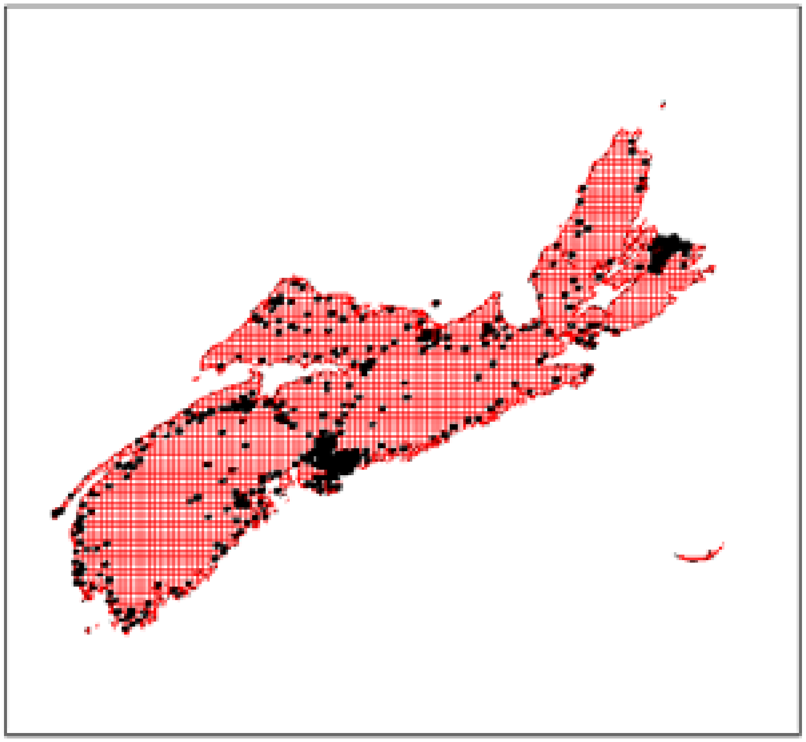
Student Work Samples: Point Pattern Analysis

- Spatial distribution of Nova Scotia Community College students and campuses
- *Spatial distribution of GPS locations representing location of tourists in Kejimkujik National Park*

Distribution of Nova Scotia Community College Students

RESULTS OF QUADRAT ANALYSIS

Point coverage: students
Number of points: 7937
Applied quadrat size (m): 4829.43
Number of quadrats: 3201
Average number of points per quadrat: 2.48
Variance of the number of points per quadrat: 246.89
Confidence level: 0.95
Test t statistic: 9776.47
Critical t value: 1.96
Point pattern: CLUSTERED



NEAREST NEIGHBOR ANALYSIS

NEAREST NEIGHBOR ANALYSIS

Neighborhood order: 1
Mean distance (meters): 3334.1
Expected mean distance (meters): 1374.0
Variance: 69.80
Z-value: 234.6
Pattern: REGULAR

Continue

Spatial Autocorrelation

- ArcInfo Workstation Grid commands Moran and Geary have two drawbacks:
  - Work only for grid data
  - Provide only values of coefficients which does not allow the user to make conclusion about the pattern (test statistics or significance levels are missing)
- ArcGIS does not have a tool for calculating Geary's coefficient of spatial autocorrelation *c.*

- The tool for Moran coefficient *I* provides an appropriate output for making conclusion about pattern

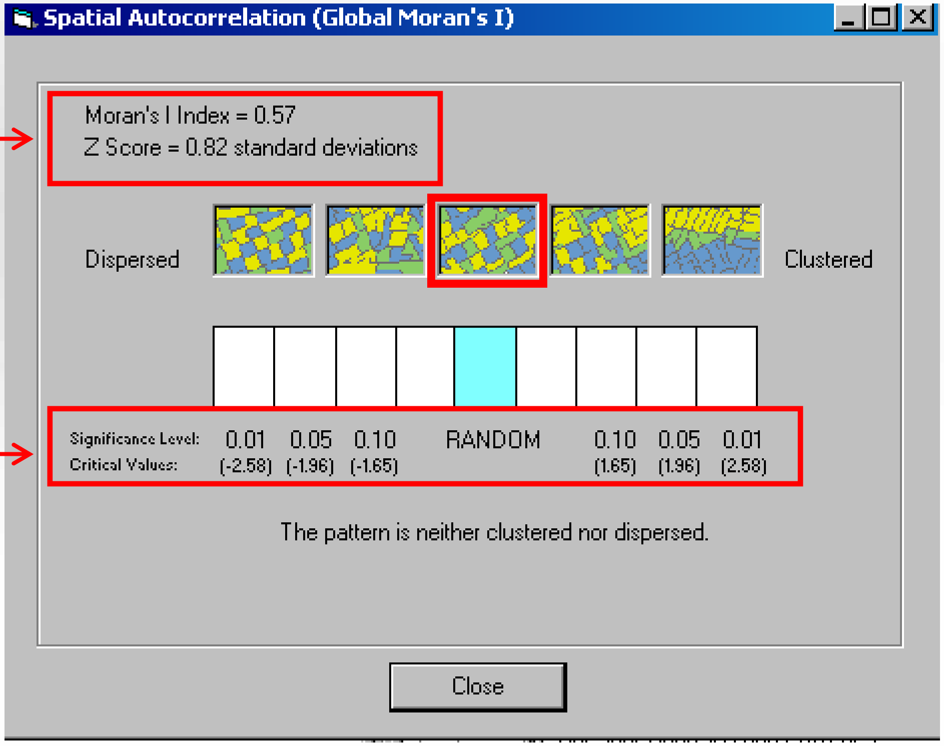
ArcInfo Workstation Grid Commands Moran and Geary

```
 Available GRIDs
 -------------------
  GRID100            GRID200            GRID300            GRID400
  GRID50
Grid: moran grid100
Running... 100%
Moran's Coefficient (I): 0.8225915149909
Grid: geary grid100
Running... 100%
Geary's Coefficient (c): 0.1323510913329
Grid: moran grid200
Running... 100%
Moran's Coefficient (I): 0.6867757219644
Grid: geary grid200
Running... 100%
Geary's Coefficient (c): 0.2319805102707
Grid: moran grid300
Running... 100%
Moran's Coefficient (I): 0.5987939746177
Grid: geary grid300
Running... 100%
Geary's Coefficient (c): 0.2877525626295
Grid: _
```

Spatial Autocorrelation (Global Moran's *I*)

Provided values of a coefficient and test statistic

Provided significance levels and corresponding critical values



Student Work Samples: Spatial Autocorrelation

- Hurricane statistics (scale intensity, damage, injuries, fatalities, mean path width) for continental USA, central USA, and Kansas
- *Patterns on spending on eleven food categories across Nova Scotia*
- *Patterns on fishing five species (cod, haddock, yellowtail, skate, lobster) in three fishing areas (entire bank, western portions and selected areas of Georges Bank) near Nova Scotia*

Spatial Autocorrelation of Hurricanes for Three Study Areas and Five Variables

Study Area	Variable Name	Moran's I Statistic	Variable Variance	Test Statistic Z	Spatial Autocorrelation Pattern
Continental USA	Fujita Scale Intensity	0.348765	0.000304	20.019246	Positive
Continental USA	Damage in Dollars	0.390424	0.0003	22.544983	Positive
Continental USA	Injuries	0.020646	0.000235	1.347858	Positive
Continental USA	Fatalities	0.055808	0.000235	3.646061	Positive
Continental USA	Mean path width of tornado, expressed in tenths of feet	0.135205	0.000455	6.342105	Positive

Study Area	Variable Name	Moran's I Statistic	Variable Variance	Test Statistic Z	Spatial Autocorrelation Pattern
Central USA	Fujita Scale Intensity	0.26338	0.000775	9.46343	Positive
Central USA	Damage in Dollars	0.376513	0.000085	12.935778	Positive
Central USA	Injuries	0.005448	0.00031	0.313422	No Autocorrelation
Central USA	Fatalities	0.000032	0.000401	0.005183	No Autocorrelation
Central USA	Mean path width of tornado, expressed in tenths of feet	0.058775	0.001356	1.600326	Positive

Study Area	Variable Name	Moran's I Statistic	Variable Variance	Test Statistic Z	Spatial Autocorrelation Pattern
Kansas	Fujita Scale Intensity	0.2324	0.0108	2.24	Positive
Kansas	Damage in Dollars	0.3721	0.0125	3.34	Positive
Kansas	Injuries	0.004566	0.003576	0.085184	No Autocorrelation
Kansas	Fatalities	0.001556	0.001571	0.052579	No Autocorrelation
Kansas	Mean path width of tornado, expressed in tenths of feet	0.081955	0.052537	0.363351	No Autocorrelation

Student Work Samples: Mapping Spatial Autocorrelation Clusters

- *Clusters of Nova Scotia Community College students by Dissemination Areas*
- Clusters of GPS locations of tourists in Kejimkujik National Park

Map 15: Local Moran *I* and Z-Score

Examination of Local Moran *I* and Z-Score for Visitor Reception Centre

This map displays the Local Moran *I* value given to each cell which displays cells with similar values.

Moran *I*

- -26558.40 - -714.57
- -714.56 - 7.26
- 7.27 - 554.97
- 554.98 - 1114.22
- 1114.23 - 519683.00

This map displays the Local Moran *I* z-score which shows cells from the Local Moran *I* analysis that are statistically significant.

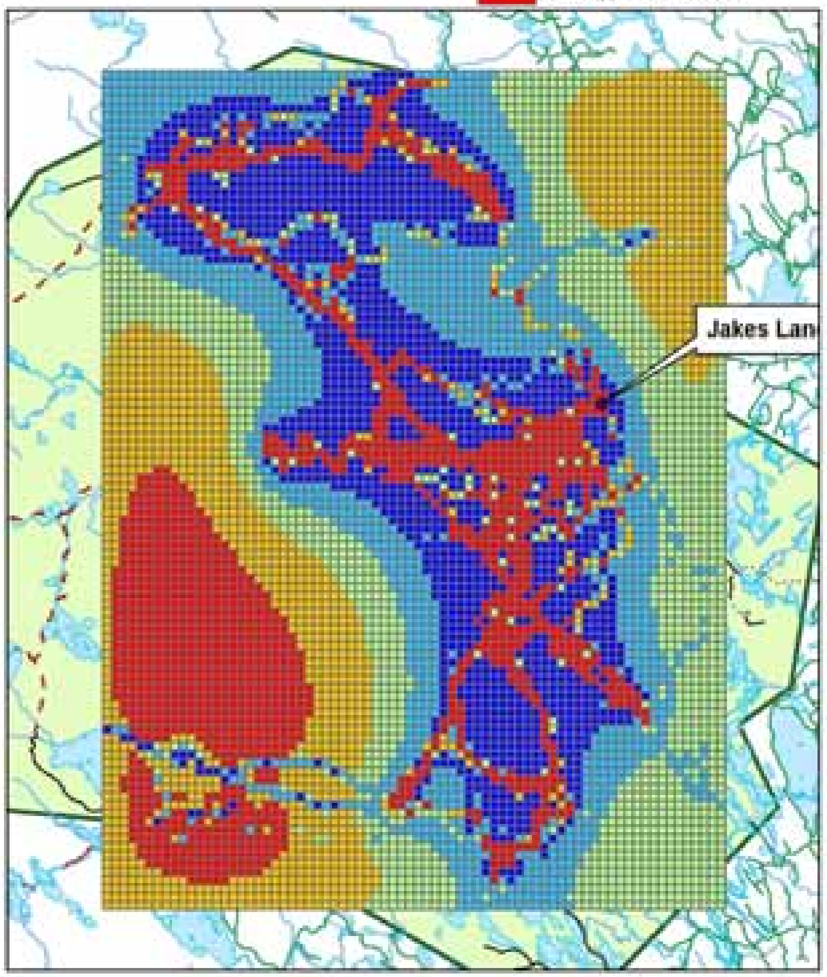
Z-Score

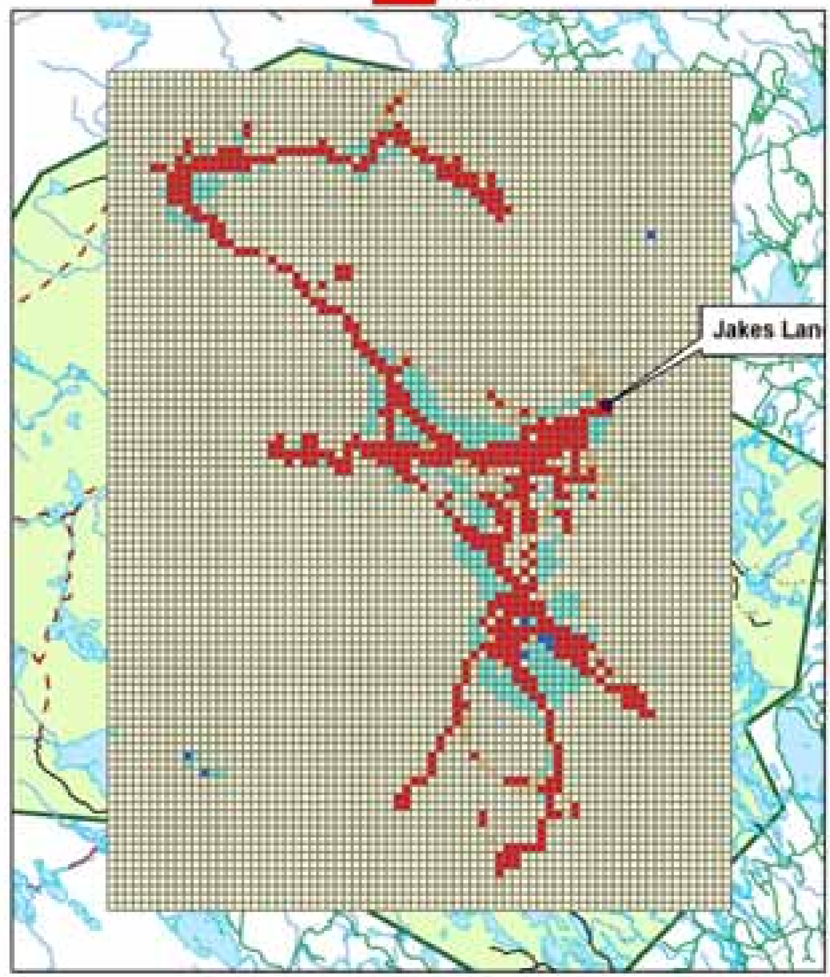
- < -2.0
- -2.0 to -1.0
- -1.0 to 1.0
- 1.0 to 2.0
- > 2.0

nscc
education that works for you

Centre of Geographic Sciences

Heather Marshall
GIS for Business Advanced Diploma Program
2007





Hot Spot Analysis

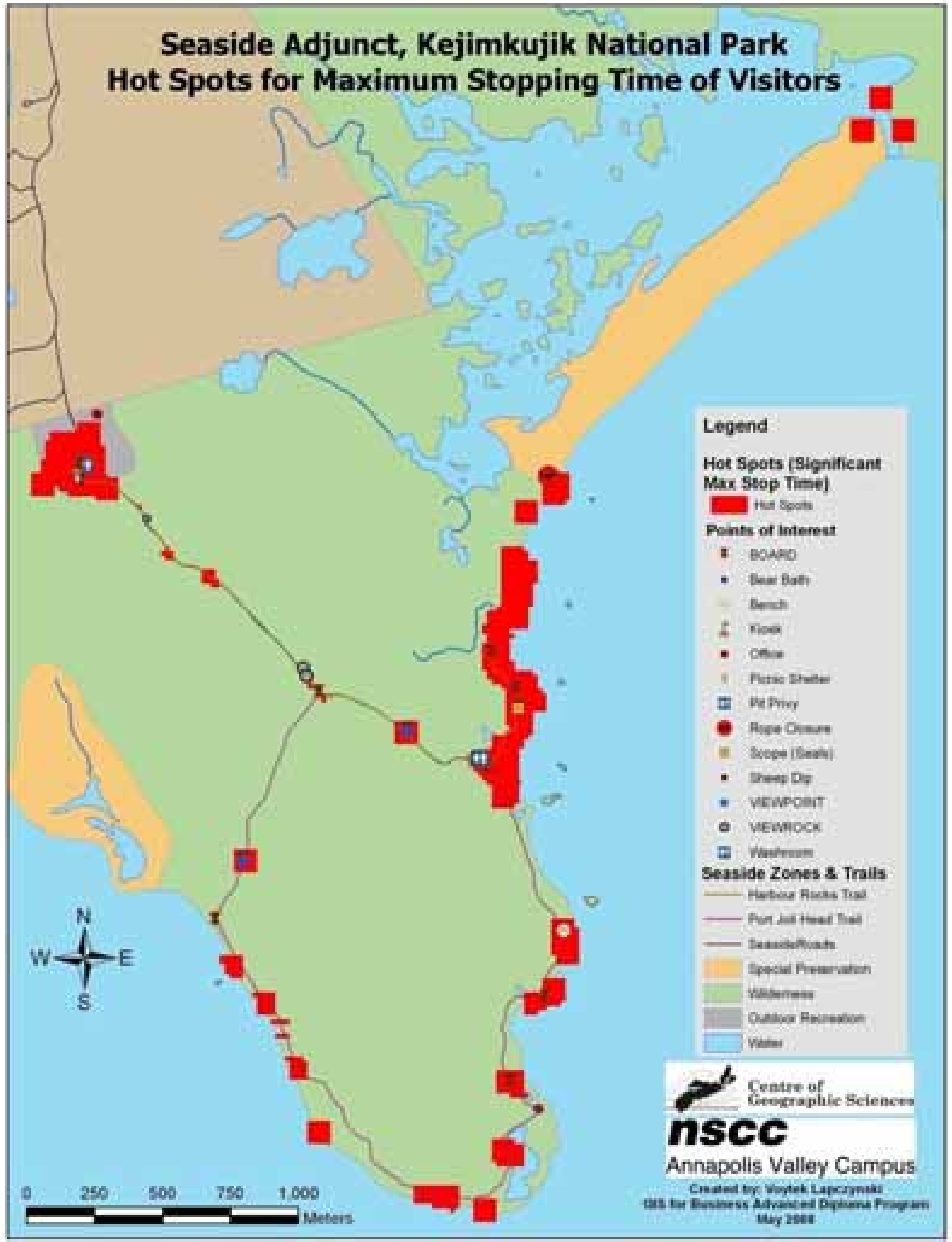
- Hot spot analysis is one of the most informative and visually appealed tools
- It is available trough such Spatial Statistics tools as
  - Hot Spot Analysis (Getis-Ord *Gi**)
  - Cluster/Outlier Analysis with Rendering
  - Hot Spot Analysis with Rendering
  - Z score Rendering

- Results can be visualized using thematic classified maps of:
  - Input variable
  - *Gi** values
  - *Z* values

Student Work Samples: Hot Spot Analysis

- Relationships between G statistic and distance threshold
- Clusters of tornado in Kansas
- Clusters of visitors in Kejimkujik National Park
- *Concentration of locations where tourists stopped in Kejimkujik Seaside Adjunct National Park*
- High and low participation rates for the Nova Scotia Breast Screening Program

- High and low spending on healthy food categories in Nova Scotia
- Clusters of high and low house prices in Halifax Regional Municipality in 2009


Seaside Adjunct, Kejimkujik National Park
Hot Spots for Maximum Stopping Time of Visitors
Legend
Hot Spots (Significant Max Stop Time)
Hot Spots
Points of Interest
BOARD
Bear Bath
Bench
Kiosk
Office
Picnic Shelter
Pit Privy
Rope Closure
Scope (Seals)
Sheep Dip
VIEWPORT
VIEWROCK
Washroom
Seaside Zones & Trails
Harbour Rocks Trail
Port Joli Head Trail
SeasideRoads
Special Preservation
Wilderness
Outdoor Recreation
Water
N
W
E
S
0 250 500 750 1,000
Meters
Centre of Geographic Sciences
nscc
Annapolis Valley Campus
Created by: Voytek Lapczynski
GIS for Business Advanced Diploma Program

Measuring Geographic Distributions

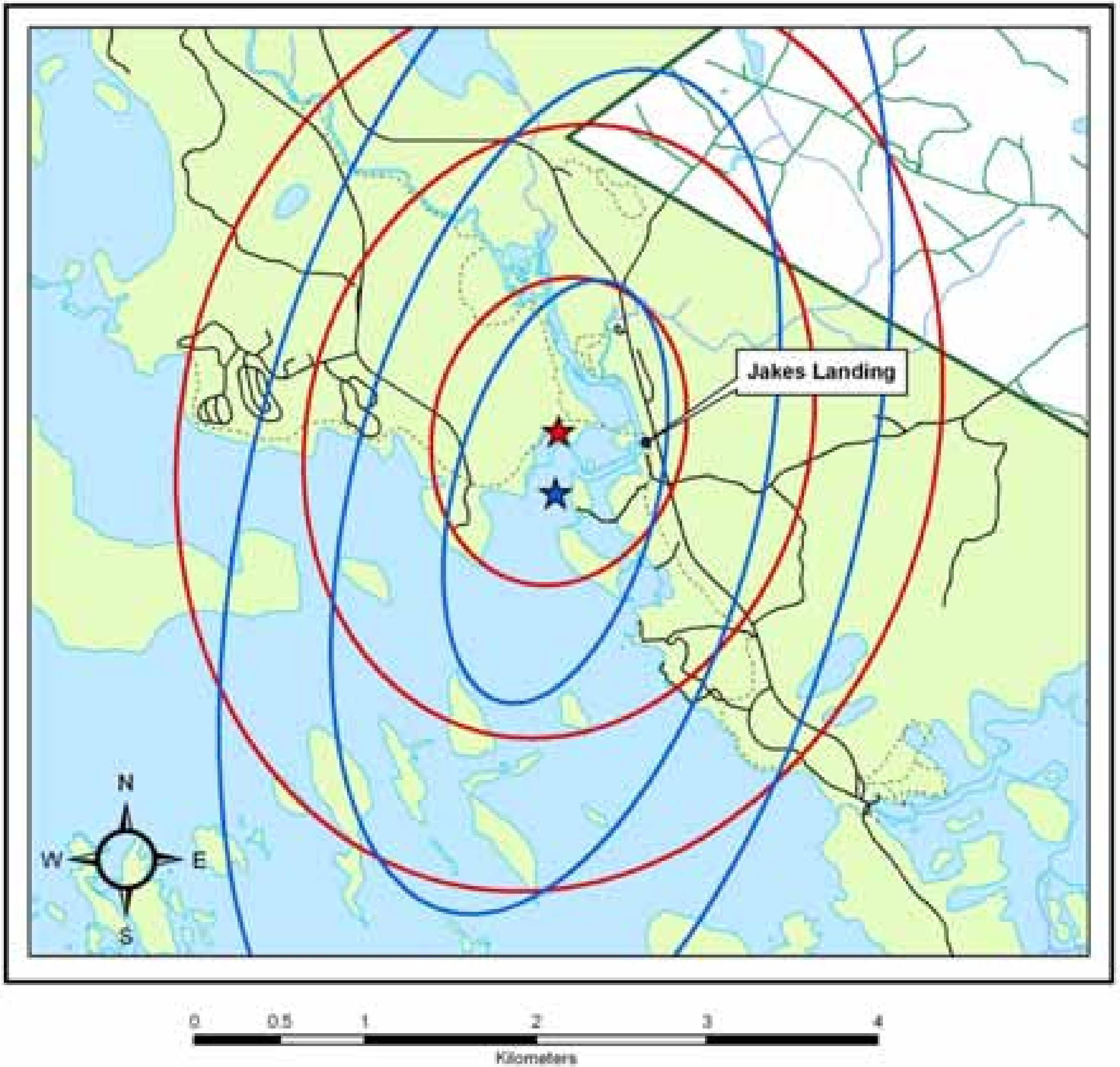
- Set of tool relatively easy to use and providing visually appealed results
- Extremely useful for such operations as:
  - Determining study area
  - Comparing spatial distributions and analyzing changes
  - Identifying the most accessible feature or location
  - Analyzing orientation and direction of linear features

Student Work Samples: Measuring Geographic Distributions

- *Determining study area based on weighted deviational ellipse for Air Miles cardholders*
- *Directional distribution of GIS jobs in North America*
- Mean centre and deviational ellipse: weekday vs. weekend tourists; morning, mid-day and evening tourists in Kejimkujik National Park
- *Succession of mean center and deviational ellipses for houses sold in Halifax Regional Municipality 1978-2006*

- *Geographic distribution of students and campuses of the Nova Scotia Community College*

Mean Centre of Activity and Standard Deviational Ellipses for Weekdays vs. Weekends using Jakes Landing Data



Mean Centre of Activity

- Weekday
- Weekend

Ellipsoid of Points

- Weekday (1, 2, 3 SD)
- Weekend (1, 2, 3 SD)

This map shows the mean centre of activity sub-divided into weekdays and weekends. This also shows standard deviational ellipses. The smallest of each colour represents 1SD, middle size 2SD and largest 3SD. These ellipses are orientated based on the distribution of activity for each category.

- Daytime Trails
- Trails
- Park Roads
- External Roads
- Portages
- Rivers
- Lakes
- National Park

nscc

Centre of Geographic Sciences

Heather Marshall
GIS for Business Advanced Diploma Program
2007

Geographically Weighted Regression

- Geographically Weighted Regression (GWR) developed in late 1970s, was added to ArcGIS as a new tool in 2007 and it is a great extension to the multiple linear regression added to ArcGIS two years later as the Ordinary Least Squares (OLS) tool
- OLS is a global-scale tool providing output for the entire study area, whereas GWR is a local-scale tool providing output for each GIS feature
- GWR has many input parameters and offers a lot of opportunities for mapping numerous output data

Comparing Results from ArcMap (OLS) and SPSS (Multiple Linear Regression)

ArcMap

SPSS

Estimated	Residual	StdResid	PRE_1	RES_1	ZRE_1
62094.40000000000	9359.57000000000	0.46264800000	62094.42766	9359.57234	0.46265
62075.00000000000	-62075.00000000000	-3.06840000000	62075.04437	-62075.04437	-3.06840
69164.00000000000	4829.97000000000	0.23874800000	69164.02794	4829.97206	0.23875
63603.10000000000	7310.93000000000	0.36138300000	63603.07268	7310.92732	0.36138
54370.90000000000	12036.10000000000	0.59495100000	54370.88642	12036.11358	0.59495
102495.00000000000	125025.00000000000	6.18003000000	102495.44337	125024.55663	6.18003
77754.10000000000	11521.90000000000	0.56953300000	77754.10527	11521.89473	0.56953
56124.40000000000	20428.60000000000	1.00979000000	56124.44101	20428.55899	1.00979
45699.20000000000	5511.78000000000	0.27245000000	45699.22475	5511.77525	0.27245
52594.40000000000	27485.60000000000	1.35863000000	52594.36524	27485.63476	1.35863
66351.70000000000	1396.33000000000	0.06902140000	66351.66782	1396.33218	0.06902
51774.40000000000	6917.64000000000	0.34194200000	51774.36189	6917.63811	0.34194
82810.20000000000	14541.80000000000	0.71880800000	82810.20622	14541.79378	0.71881
57318.20000000000	-3399.17000000000	-0.16802200000	57318.16535	-3399.16535	-0.16802
73462.40000000000	21169.60000000000	1.04643000000	73462.36951	21169.63049	1.04643
83895.70000000000	31318.30000000000	1.54808000000	83895.67461	31318.32539	1.54808
82938.90000000000	518.13200000000	0.02561150000	82938.86760	518.13240	0.02561
82991.00000000000	21640.00000000000	1.06968000000	82990.98000	21640.02000	1.06968
57334.00000000000	-4383.05000000000	-0.21665600000	57334.04623	-4383.04623	-0.21666
71167.30000000000	2275.74000000000	0.11249100000	71167.25526	2275.74474	0.11249

Possible Problems with Using GWR

- Relationships between predictors and target variable are not linear
- Correlations among predictors is detected (collinearity)
- Different patterns exist across study area (non-stationarity)
- Presence of data outliers can bias predicted values
- Model performance depends on the magnitude of the target variable value (heteroskedasticity)

- Spatial clusters of over-estimated and under-estimated predicted values exist (spatial autocorrelation of residuals)
- Residuals are not normally distributed

GWR Input Parameters to Control

- Set of independent variables (predictors)
- Distance or number of neighbors
- Optionally, two kernel types
  - Fixed
  - Adaptive
- Optionally, two bandwidth methods
  - Akaike Informative Criterion
  - Cross Validation
- Optionally, using other raster-related parameters

Comparison of GWR Scenarios: An Example

Scenario	Kernel type	Bandwidth method	Distance	Number of neighbors
A	Fixed	Parameter	4000 meters	N/A
B	Fixed	Parameter	2000 meters	N/A
C	Adaptive	Parameter	N/A	30
D	Fixed	AIC	N/A	N/A
E	Fixed	CV	N/A	N/A
F (weighted by area	Adaptive	Parameter	N/A	30
G (weighted by population)	Adaptive	Parameter	N/A	30

Preferred Values of GWR Input Parameters

Preferred values	
Low values	High values
Residual squares	R2 (close to 1.0)
Sigma	R2 adjusted (close to 1.0)
AICc	

Finding The Best Model: An Example

Scenario	Residual squares	Effective number	Sigma	AICc	R2	R2 adjusted
A	44,594,620,522	30.95	15,120	5,012	0.760	0.723
B	8,519,128,670	49.60	11,779	2,468	0.933	0.880
C	8,519,128,670	49.60	11,779	2,468	0.933	0.880
D	100,432,304,465	16.53	17,814	7,464	0.640	0.623
E	100,432,304,465	16.53	17,814	7,464	0.640	0.623
F	**814,654,793**	34.21	**4,128**	**1,649**	**0.949**	**0.898**
G	3,926,076,424,856	42.81	269,158	2,763	0.903	0.827

The best model is scenario F (Adaptive kernel, 30 neighbors, weighted by area)

Possible Output Maps From GWR

1. Observed values (column Observed)
2. Predicted values (column Predicted)
3. Raw residuals (column Residual)
4. Standardized residuals (column StResid)
5. Coefficients of determination (column Local R2)
6. Standard error (column StError)
7. Collinearity (column Cond)
8. Weighted factor (column Weight)

Final Thoughts

- Learners through practicing spatial analysis and geostatistics with ArcGIS acquire new analytical and research skills which can be used for
  - Enhancing students assignments in other courses
  - Completing major research projects and capstone projects
  - Continuing their education toward the degree of Masters in Geomatics, joined program between the Nova Scotia Community College and Acadia University (Wolfville, Nova Scotia)