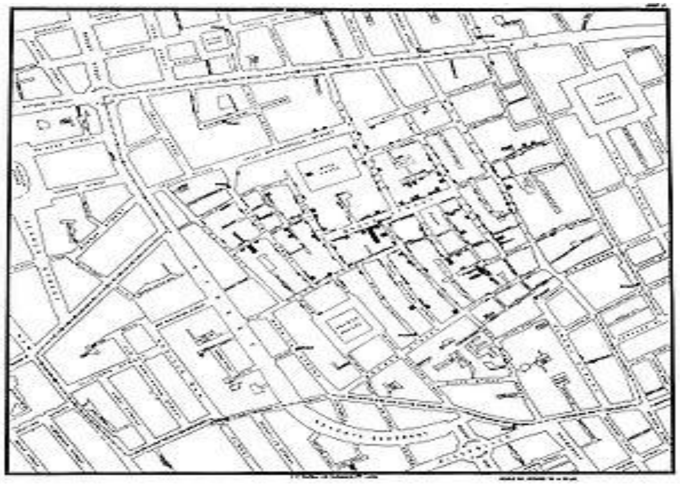
Geographic Information System
for Ocean and Coastal Science

P. S. Roy
Honorary Professor
National Academy of Sciences-
Platinum Jubilee Senior Scientist Fellow
University Center for Earth and Space Sciences
University of Hyderabad,
Hyderabad - 500046 India
psroy13@gmail.com

January, 2016

John Snow's London Map showing the clusters of cholera cases (1854)



Pump (Well) without a Handle

What is GIS? Definition

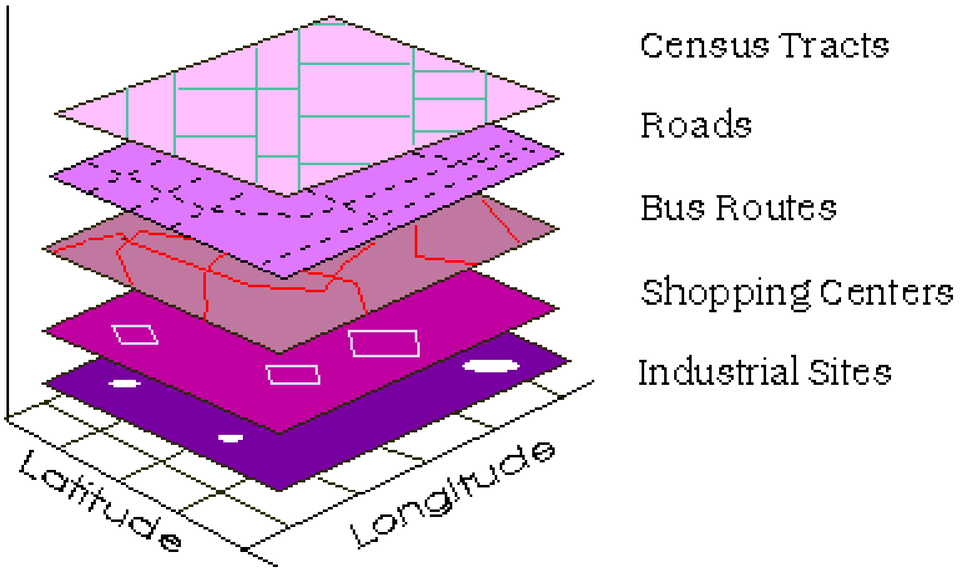
- A powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world (Burrough, 1986)

  Capturing, storing, checking, manipulation, analysing and displaying data, spatially Referenced

What can you do with a GIS?

- Collect & organise data
  - Normally data comes in data layers
    - Railroad dataset (data layer)
    - Roads dataset
    - River dataset
    - City, land use, soil type, population datasets etc. etc. etc.

Point
Line
Shape



Indispensable tool

- Spatial analysis:

  Keep an eye on the state of earth systems using satellites and monitoring stations (water, pollution, ecosystems, urban development,…)

- Modeling and simulation:

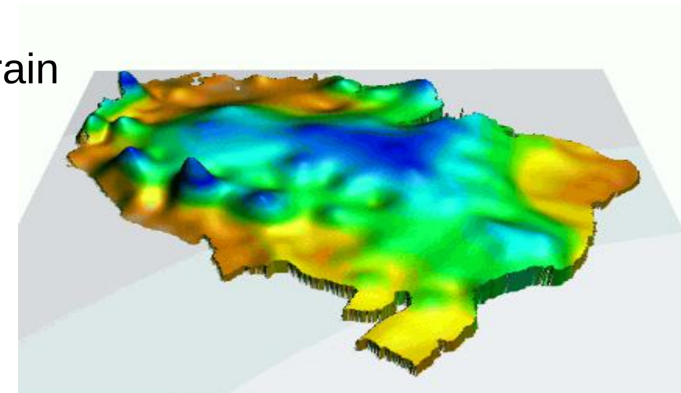
  Predict consequences of human actions and natural processes

  Analysis and risk assessment:
  find the problem areas and analyse the possible causes (soil erosion, groundwater pollution,..)

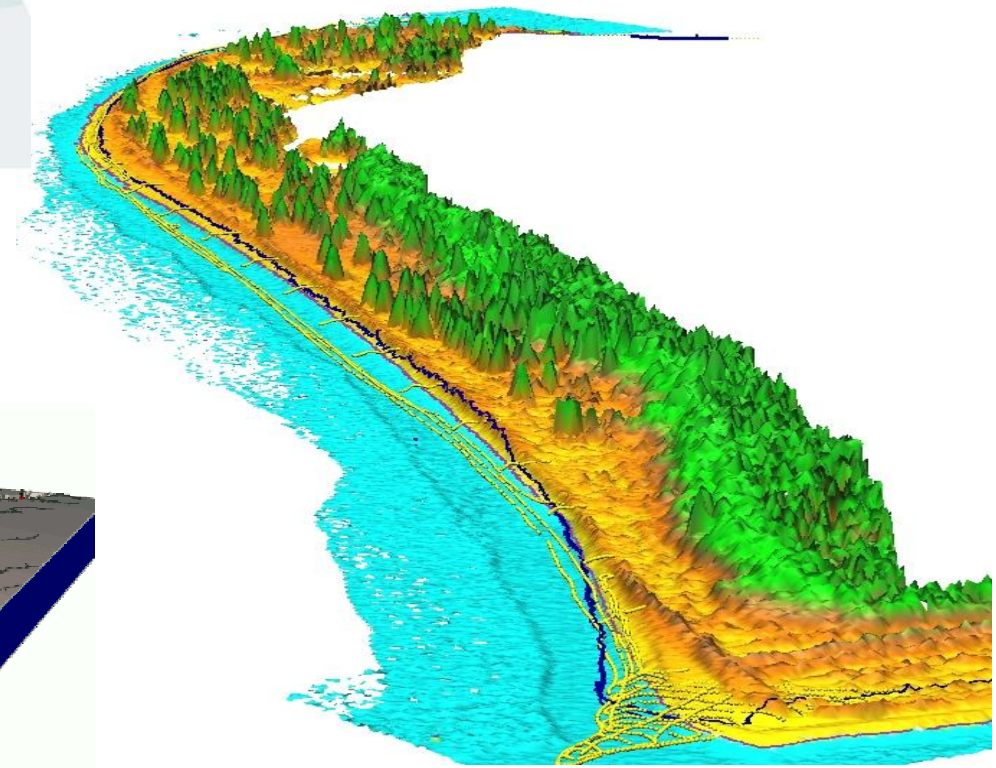
  Planning and decision support:
  provide information and tools for better management of resources

Examples

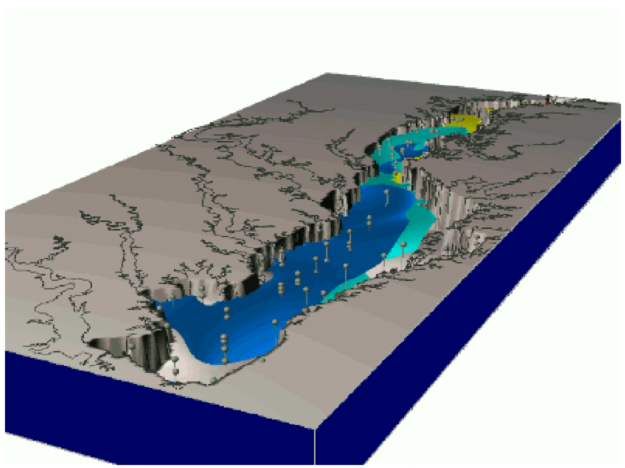
Lots of rain
Dry



Coastal simulations



High nitrogen concentrations



Types of Geographical data available

- Topographic maps
- Satellite images and photographs
- Administrative data
- Census data
- Statistical data on people, land cover and land use at a wide range of levels
- Data from marketing surveys
- Data on utilities (gas, water, electricity) and their locations
- Data on rocks, water, soil, atmosphere, biological activity, natural disaster and hazards collected for a wide range of spatial and temporal resolutions

Sources/Producers of Data

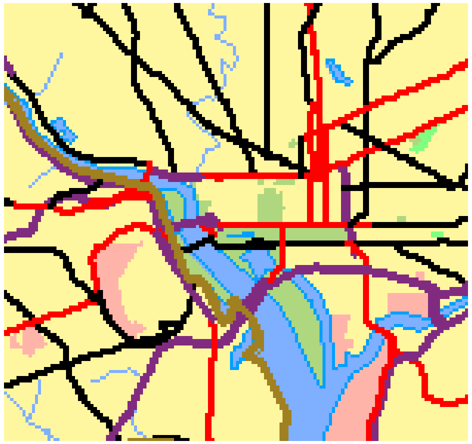
- National Mapping Agencies (SOI….); private mapping companies (Open Street Map..)
- Land registration and Cadastre
- Military organizations
- Hydrographic mapping
- Remote sensing companies and satellite agencies
- Natural resource surveys
  - Geologists, hydrologists, physical geographers
  - Land evaluators, ecologists, meteorologists
  - Oceanographers

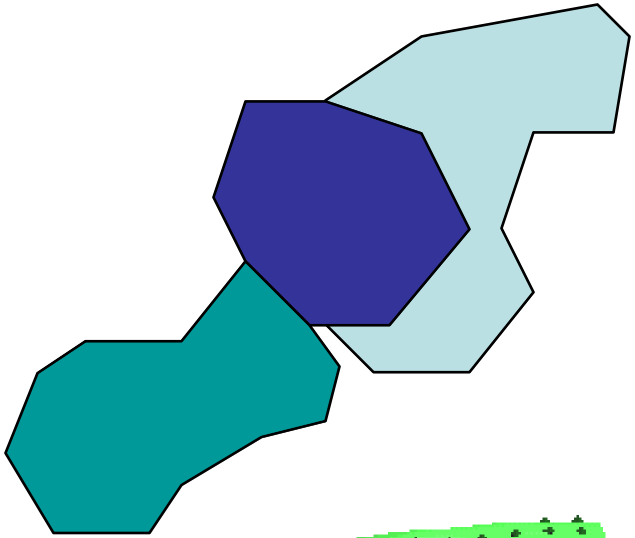
GIS Software

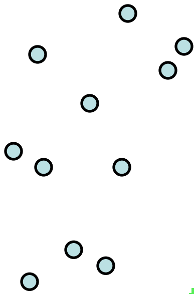
- ArcInfo (ESRI)
  - ArcView, ArcIMS, ArcExplorer, MapObjects, ArcLogistics, ArcCAD, NetEngine …
- InterGraph
- MapInfo
- Spatial DBMS (Oracle, Sybase, Microsoft SQL …)
- QGIS, GRASS (Open Source GIS)
- Customized software used by companies
- Research groups at universities

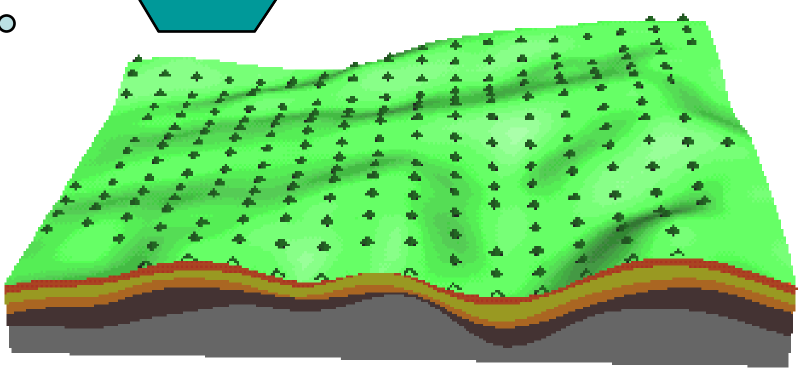
GIS and Computer Science

– Problems for Computer Science
  - databases
  - graphics
  - graph theory
  - computational geometry
  - visualization
  - combinatorial optimization
  - interfaces









Spatial analysis

The process of examining the locations, attributes, and relationships of features in spatial data through overlay and other analytical techniques in order to address a question or gain useful knowledge. Spatial analysis extracts or creates new information from spatial data.

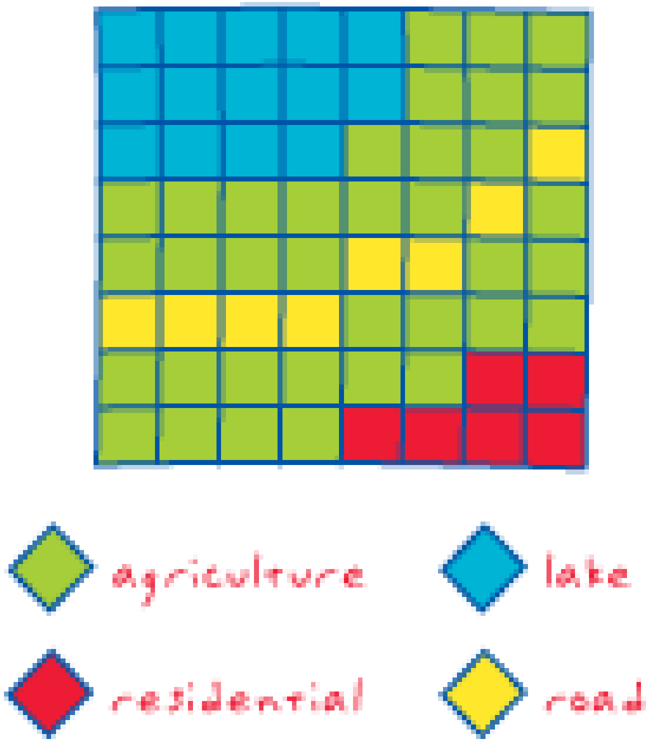
Spatial modeling

A methodology or set of analytical procedures used to derive information about spatial relationships between geographic phenomena.

Topology is the mathematical study of the properties that are preserved through deformations, twisting, and stretching of objects. Tearing, however, is not allowed. A circle is topologically equivalent to an ellipse (into which it can be deformed by stretching) and a sphere is equivalent to an ellipsoid.

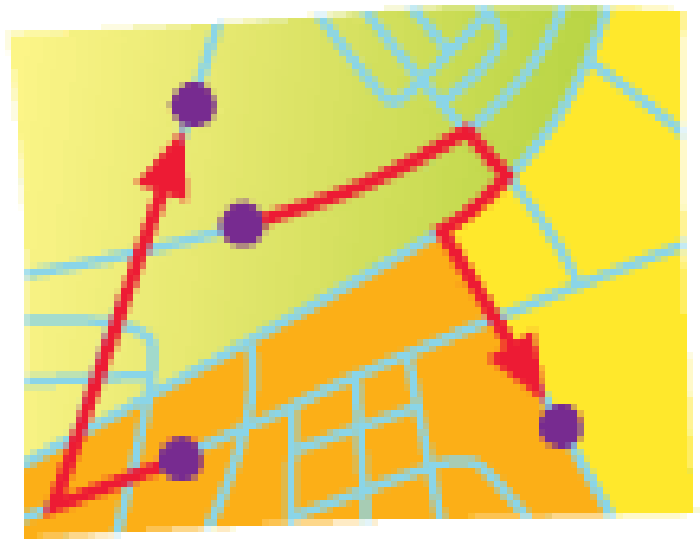
Raster

A spatial data model that defines space as an array of equally sized cells arranged in rows and columns, and composed of single or multiple bands. Each cell contains an attribute value and location coordinates. Unlike a vector structure, which stores coordinates explicitly, raster coordinates are contained in the ordering of the matrix. Groups of cells that share the same value represent the same type of geographic feature.



Vector

A coordinate-based data model that represents geographic features as points, lines, and polygons. Each point feature is represented as a single coordinate pair, while line and polygon features are represented as ordered lists of vertices. Attributes are associated with each vector feature, as opposed to a raster data model, which associates attributes with grid cells.



Network dataset

A collection of topologically connected network elements (edges, junctions, and turns) that are derived from network sources, typically used to represent a linear network, such as a road or subway system. Each network element is associated with a collection of network attributes. Network datasets are typically used to model undirected flow systems.

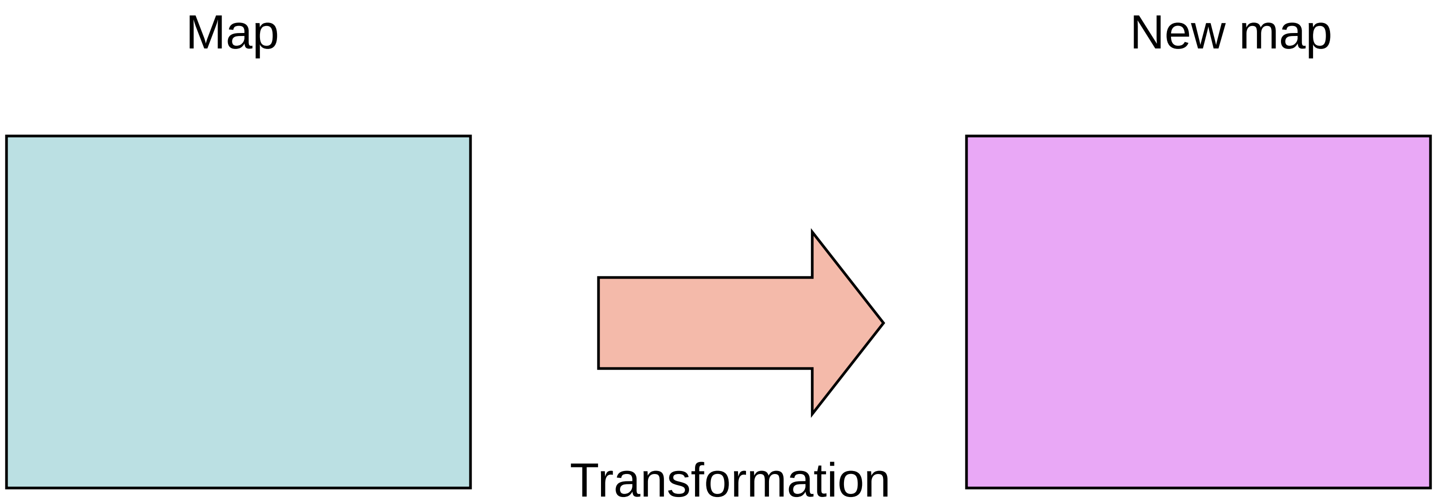
Runoff estimation using GIS

GIS Layers LULC, Drainage, Topography, drainage (direction of flow, amount of flow), Soil (depth, texture, impermeability and infiltration) precipitation

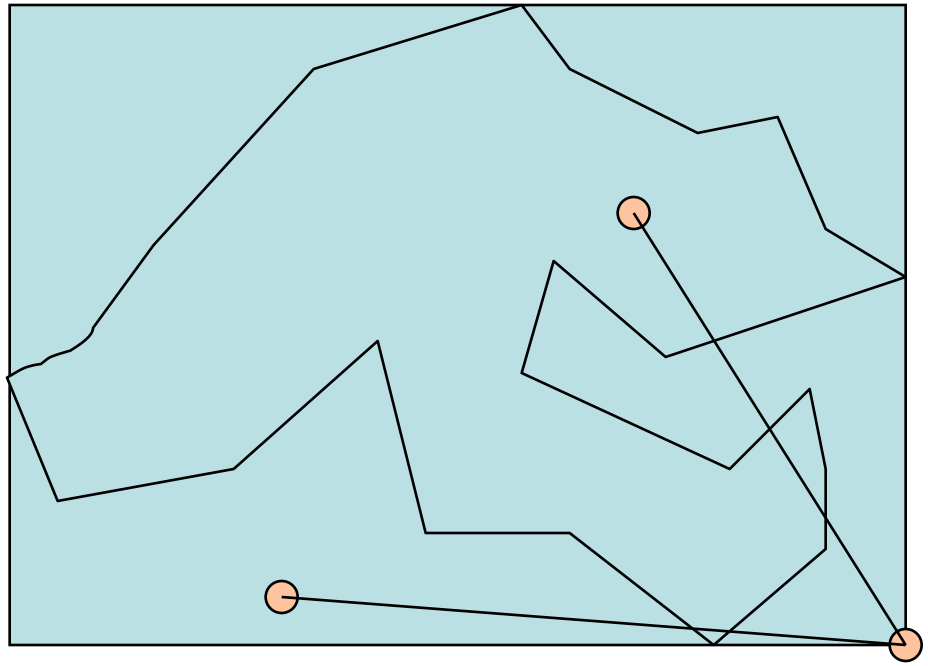
Transformations

- Buffering (Point, Line, Area)
- Point-in-polygon
- Polygon Overlay
- Spatial Interpolation
  - Theissen polygons
  - Inverse-distance weighting
  - Kriging
  - Density estimation

Basic Approach

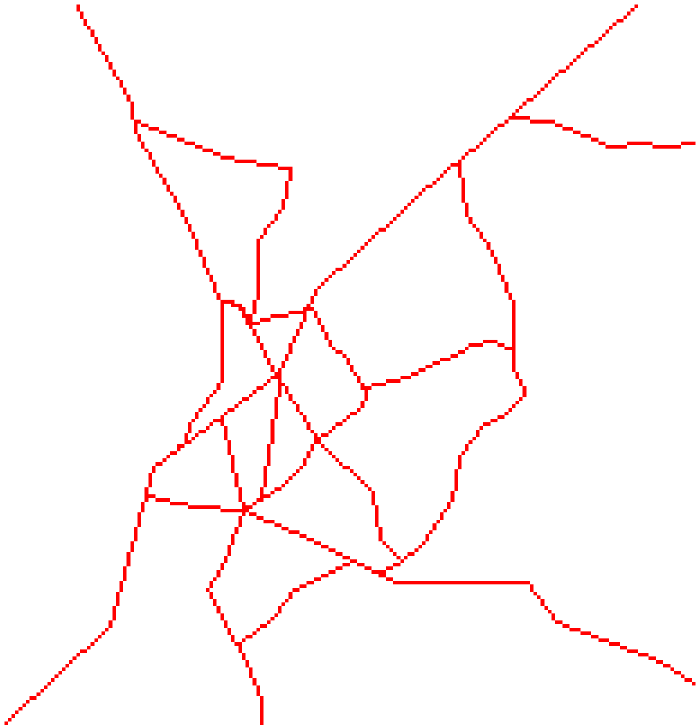


Point-in-polygon

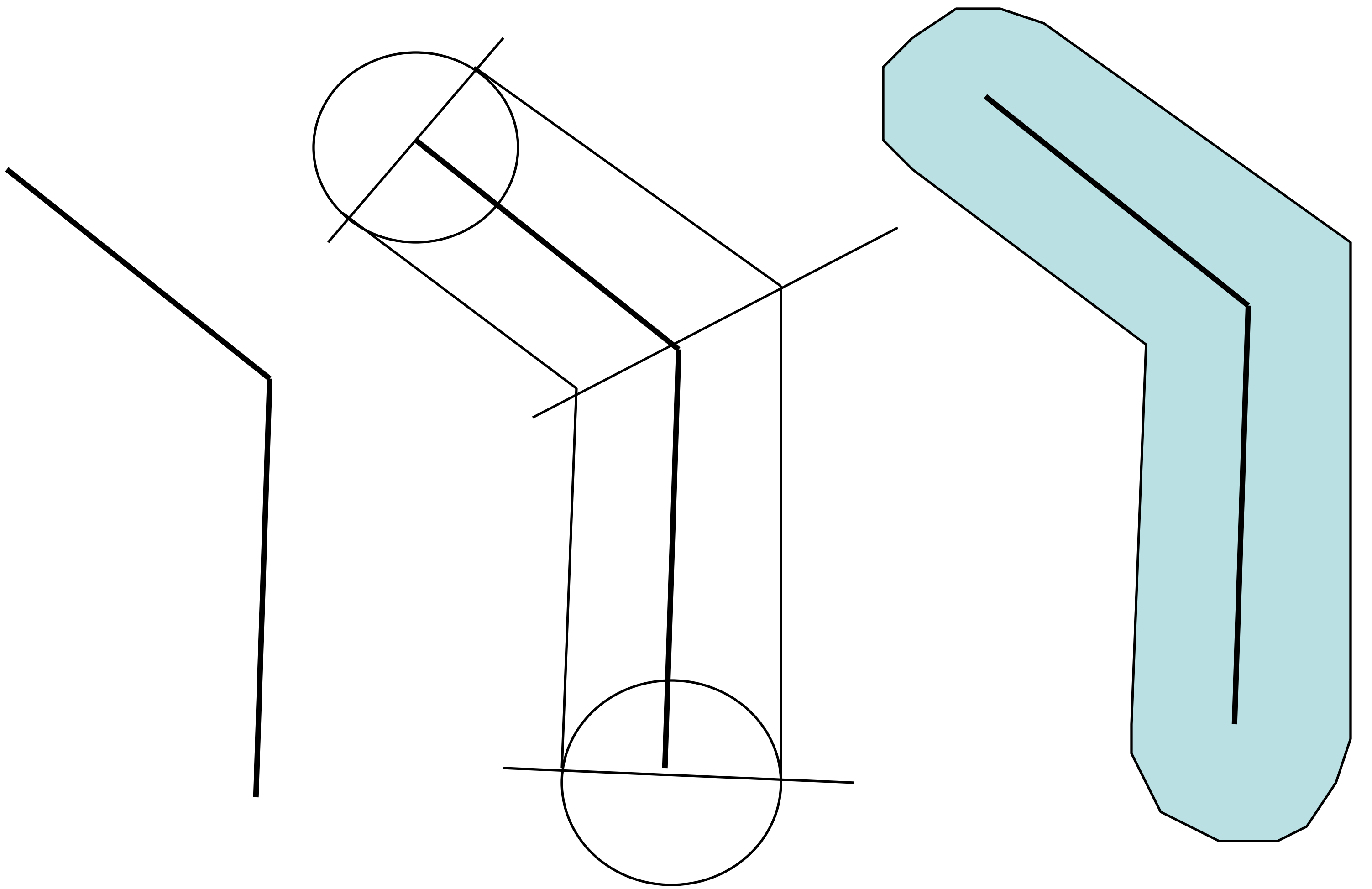


Select point known to be outside
Select point to be tested
Create line segment
Intersect with all boundary
  segments
Count intersections
EVEN=OUTSIDE
ODD=INSIDE

Create a buffer: Raster



Create a Buffer: vector

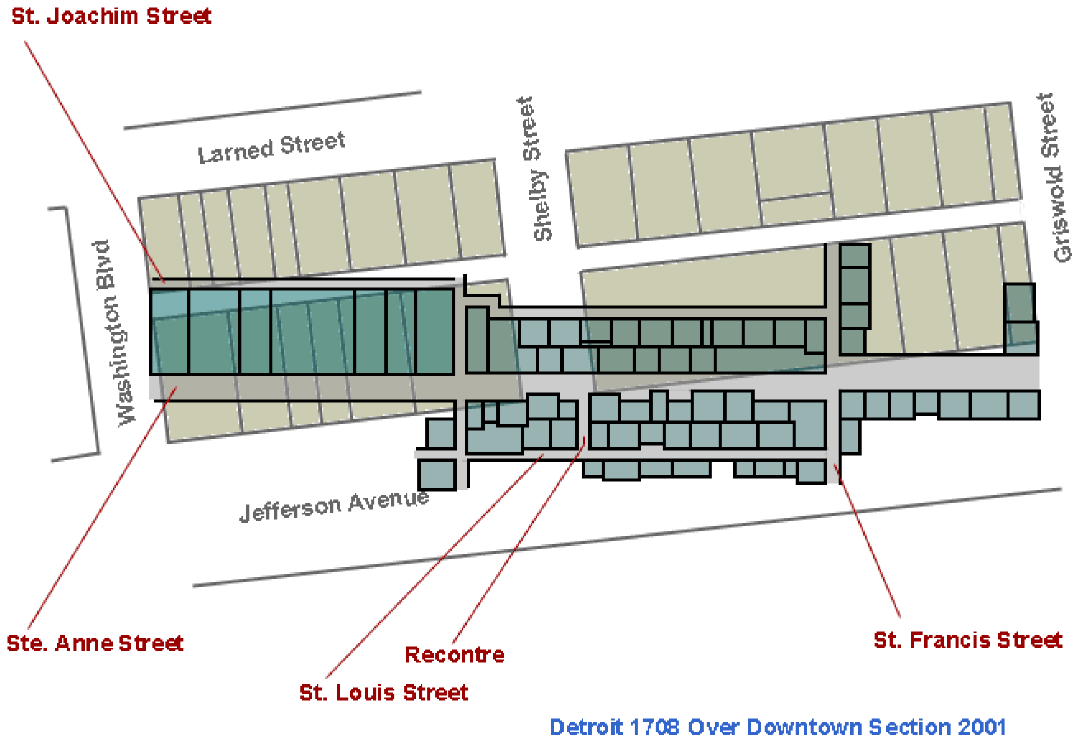


Combining maps

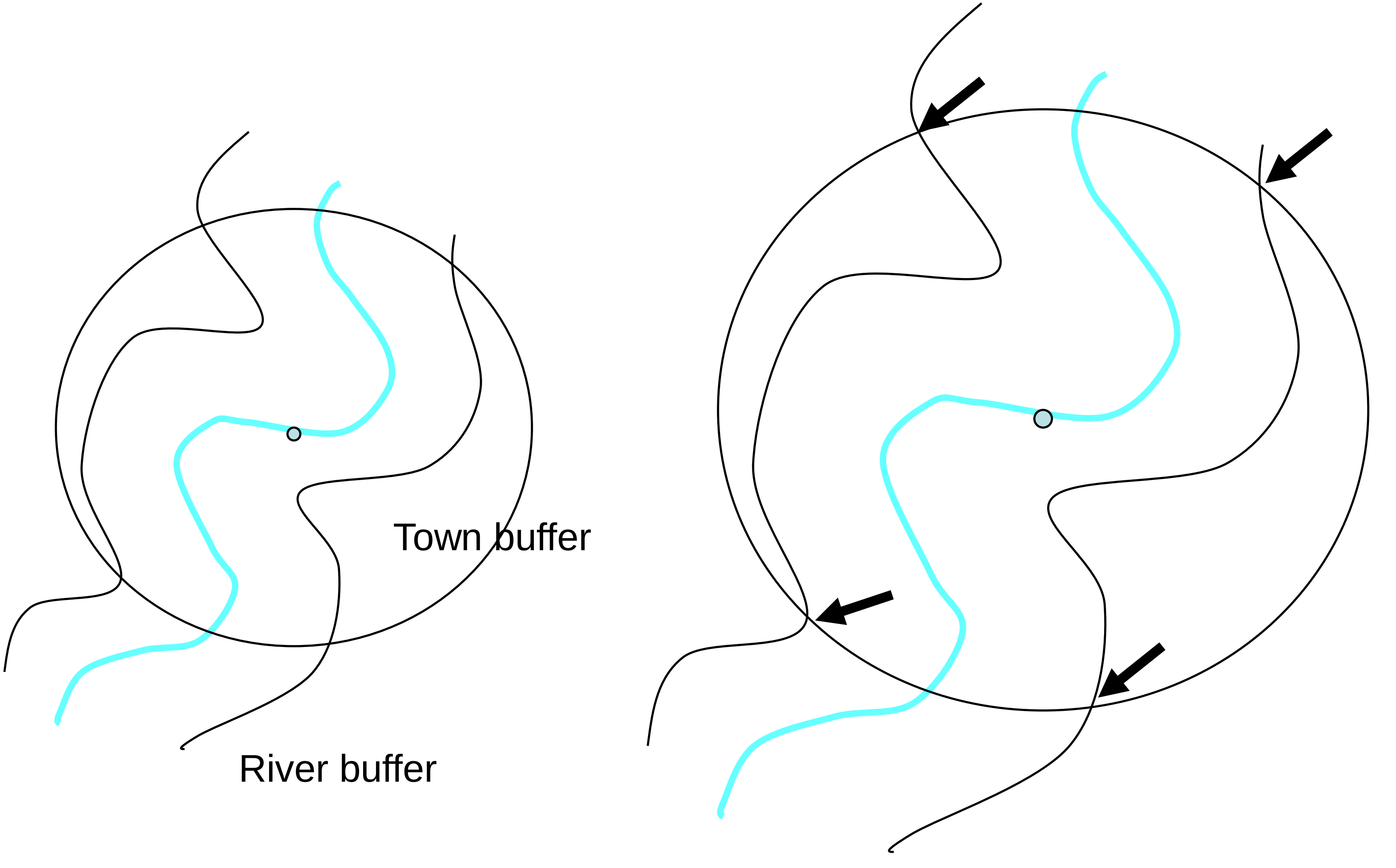
- RASTER
- As long as maps have same extent, resolution, etc, overlay is direct (pixel-to-pixel)
- Otherwise, needs interpolation
- Use map algebra (Tomlin)
- Tomlin’s operators
  - Focal, Local, Zonal

Combining Maps

- VECTOR
- A problem



Creating new zones



Other spatial analysis methods

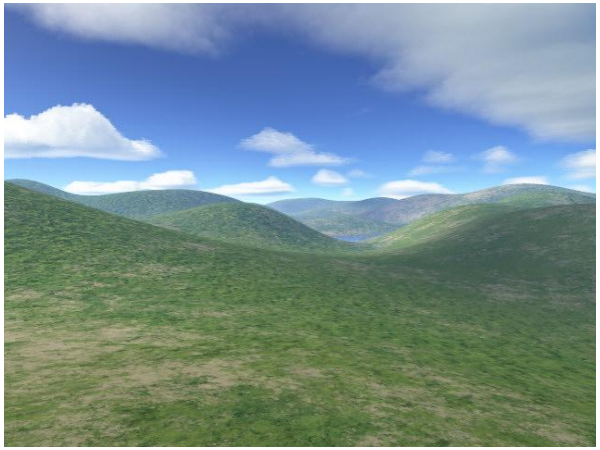
- Centrographic analysis (mean center)
- Dispersion measures (stand. Dist)
- Point clustering measures (NNS)
- Moran's I: Spatial autocorrelation (Clustering of neighboring values)
- Fragmentation and fractional dimension
- Spatial optimization
  - Point
  - Route
- Spatial interpolation

Spatial autocorrelation

[**spatial** statistics (use for geostatistics)] A measure of the degree to which a set of **spatial** features and their associated data values tend to be clustered together in space (positive **spatial autocorrelation**) or dispersed (negative **spatial autocorrelation**).

Low

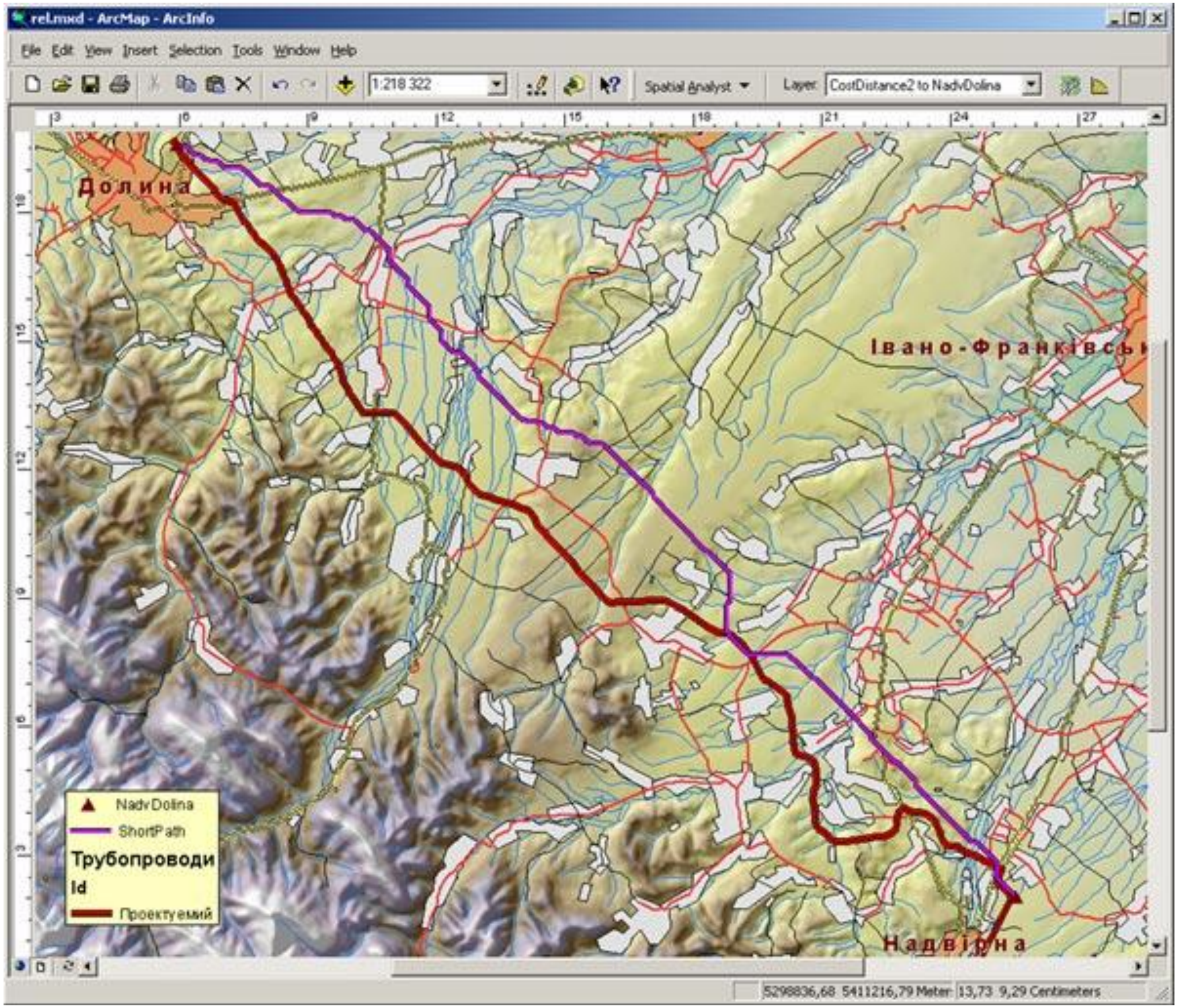
Correlation of a field with itself



High

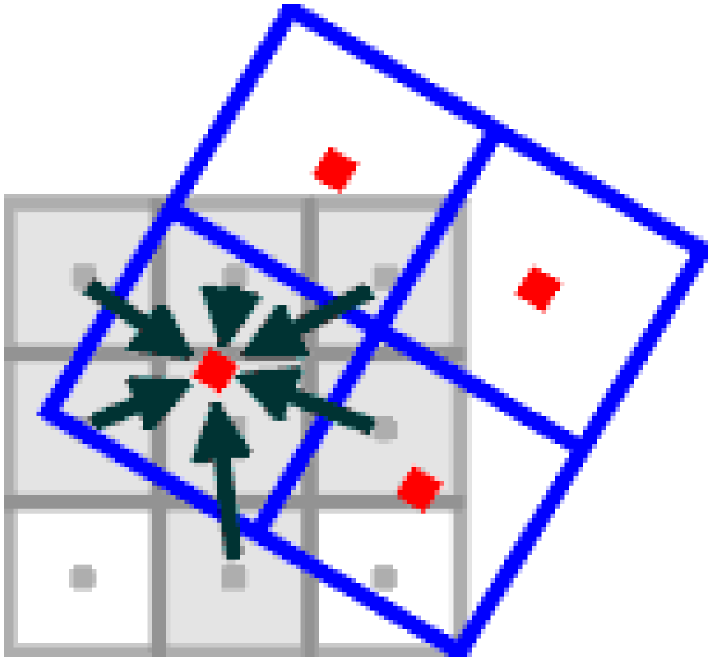
Maximum

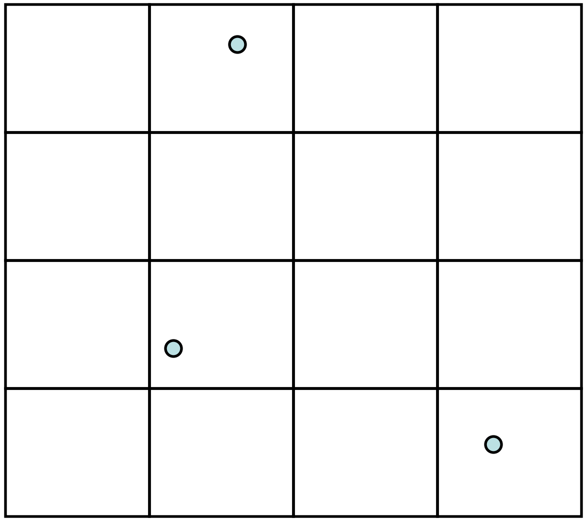
Spatial optimization



www.giscenter.net/eng/work_03_e.html

Spatial interpolation

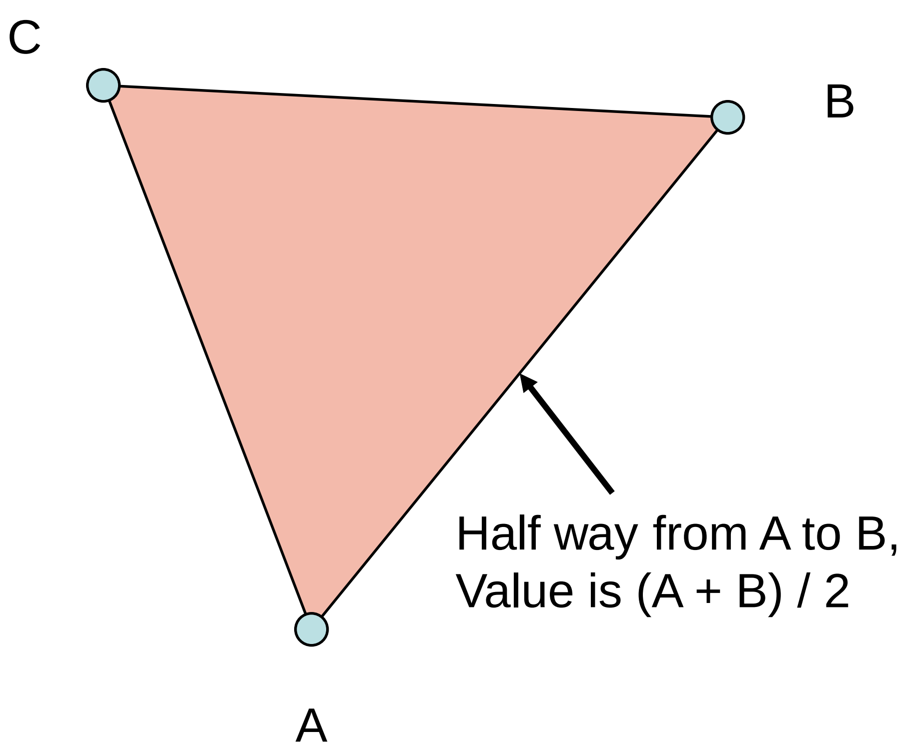




Spatial interpolation is the procedure of estimating the value of properties at unsampled sites within the area covered by existing observations. in almost all cases the property must be interval or ratio scaled.

Linear interpolation

In mathematics, **linear interpolation** is a method of curve fitting using linear polynomials.



Nonlinear Interpolation

- In the mathematical field of numerical analysis, **interpolation** is a method of constructing new data points within the range of a discrete set of known data points.
- **When things aren't or *shouldn't* be so simple**
- **Values computed by piecewise "moving window"**
- **Basic types:**

  1. Trend surface analysis / Polynomial

  2. Minimum Curvature Spline

  3. Inverse Distance Weighted

  4. Kriging

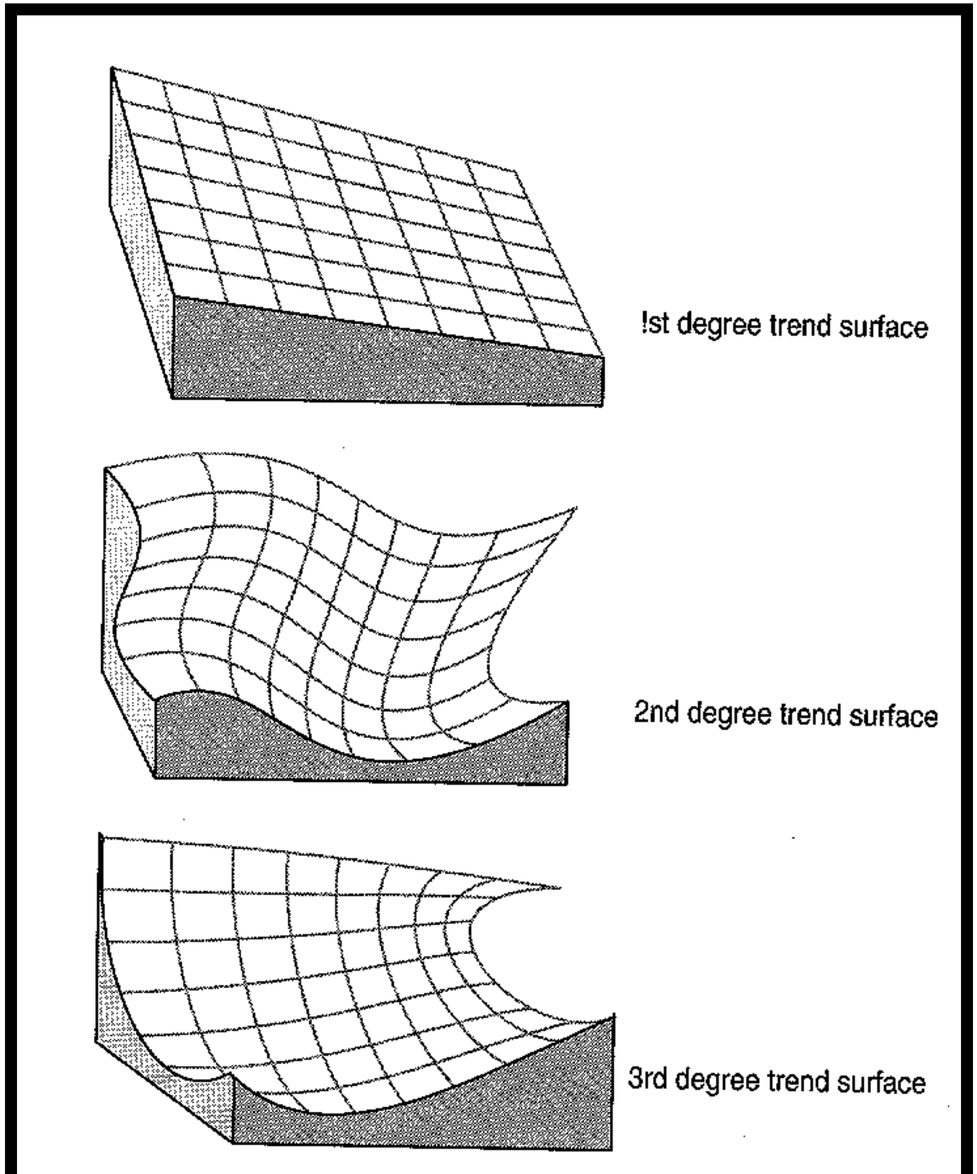
Trend Surface/Polynomial

- **point-based**
- **Fits a polynomial to input points**
- **When calculating function that will describe surface, uses least-square regression fit**
- **approximate interpolator**
  - Resulting surface doesn’t pass through all data points
  - global trend in data, varying slowly overlain by local but rapid fluctuations

Trend Surface cont.

- **flat but TILTED plane to fit data**
  - surface is approximated by linear equation (polynomial degree 1)
  - $z = a + bx + cy$
- **tilted but WARPED plane to fit data**
  - surface is approximated by quadratic equation (polynomial degree 2)
  - $z = a + bx + cy + dx^2 + exy + fy^2$

Trend Surfaces



IDW

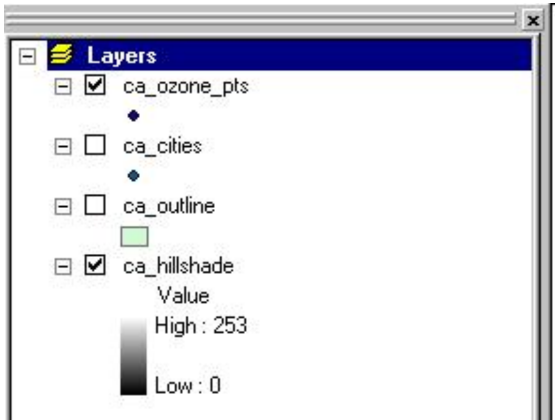
- IDW is popular, easy, but problematic
- Interpolated values limited by the range of the data
- No interpolated value will be outside the observed range of z values
- How many points should be included in the averaging?
- What about irregularly distributed points?
- What about the map edges?

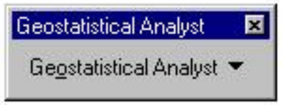
IDW Example

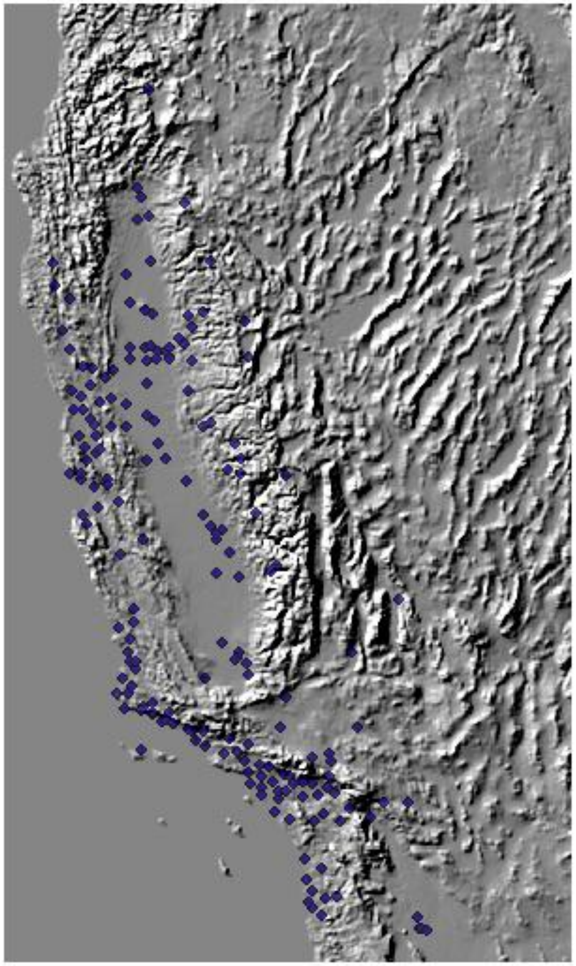
- ozone concentrations at NCR measurement stations

1. estimate a complete field, make a map
2. estimate ozone concentrations at specific locations (e.g., New Delhi)

Ozone in S. Cal: Text Example

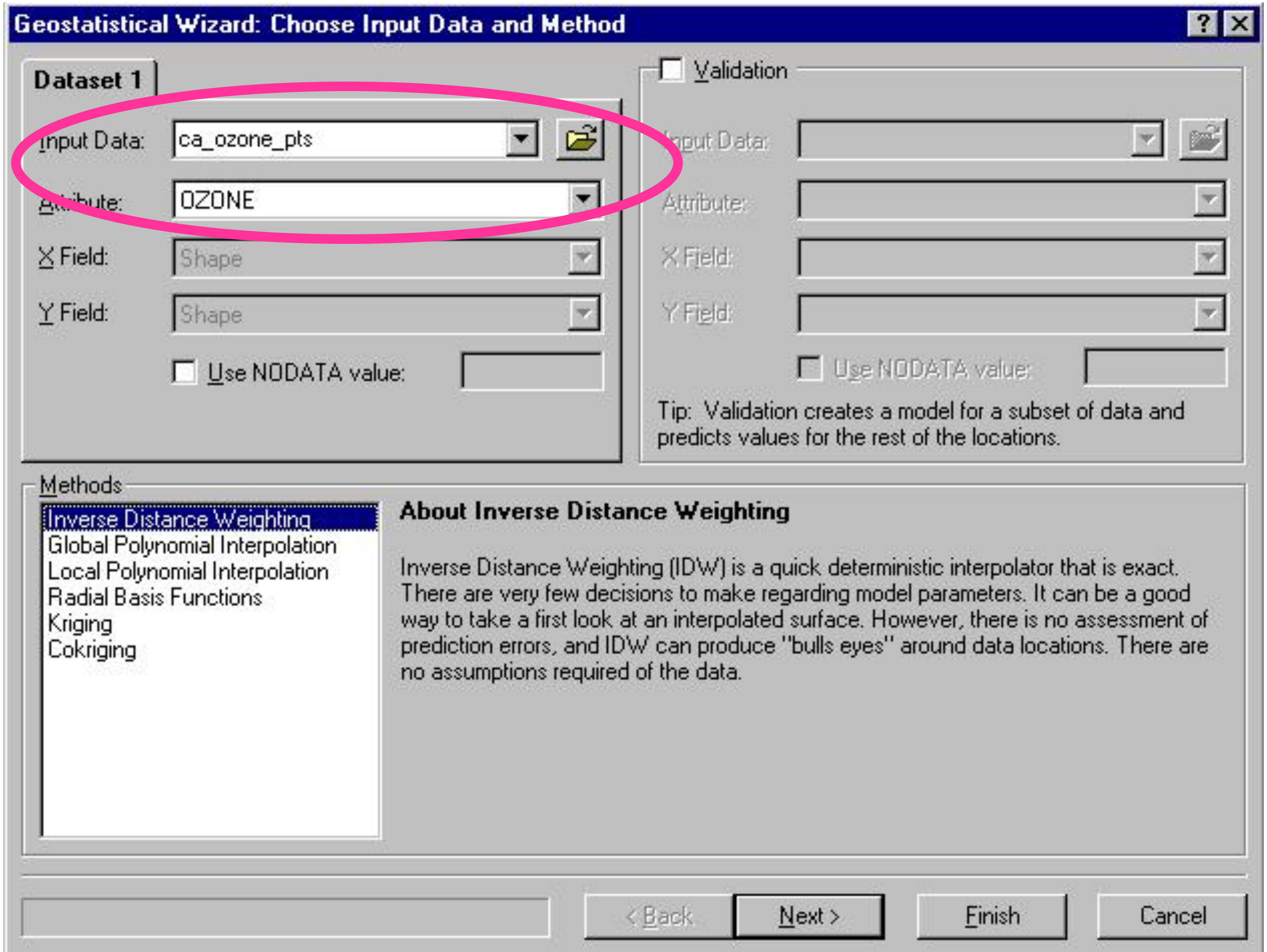




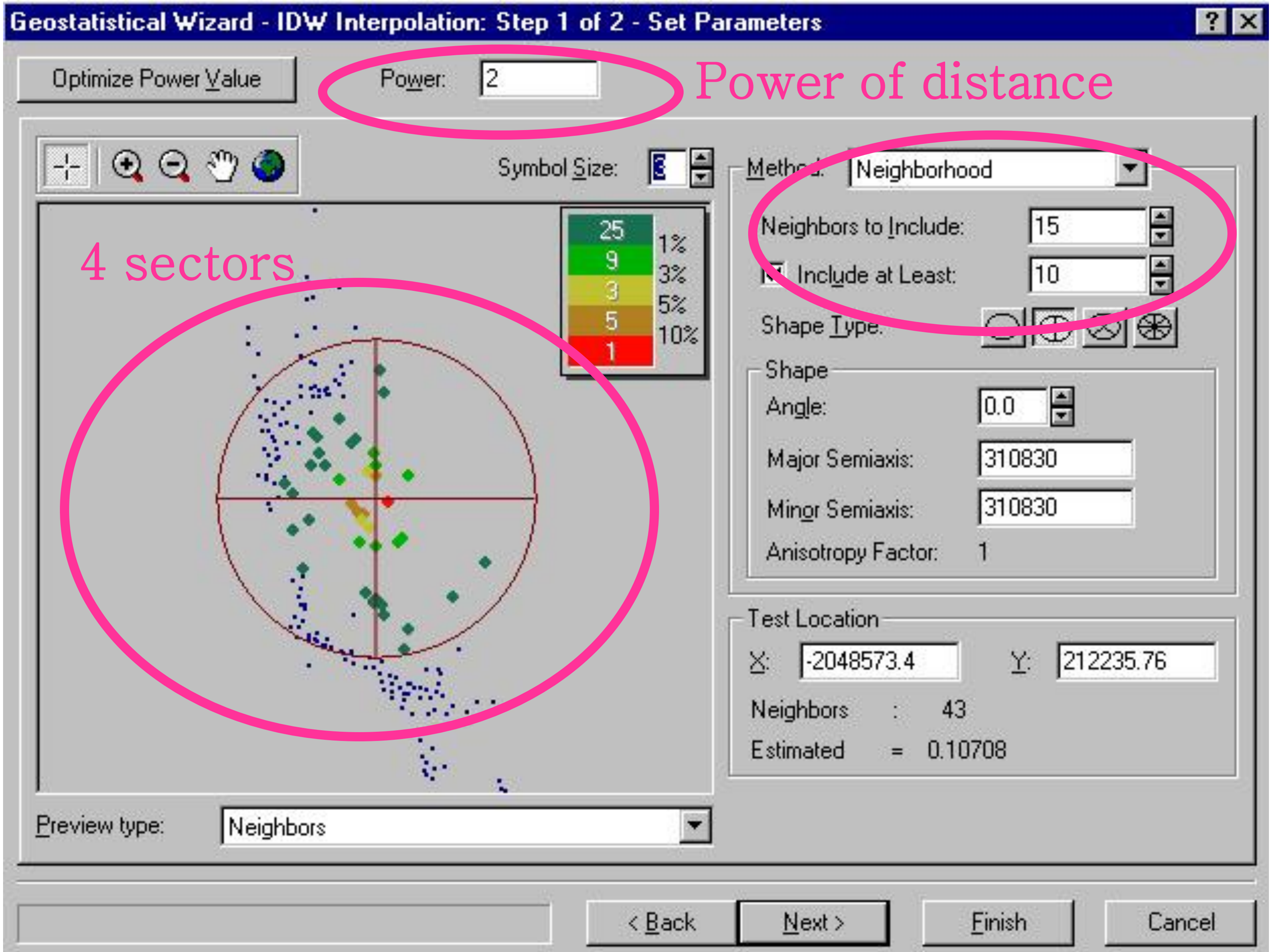


measuring stations and concentrations
(point shapefile)
NCR cities (point shapefile)
NCRoutline (polygon shapefile)
DEM (raster)

IDW Wizard in Geostatistical Analyst define data source

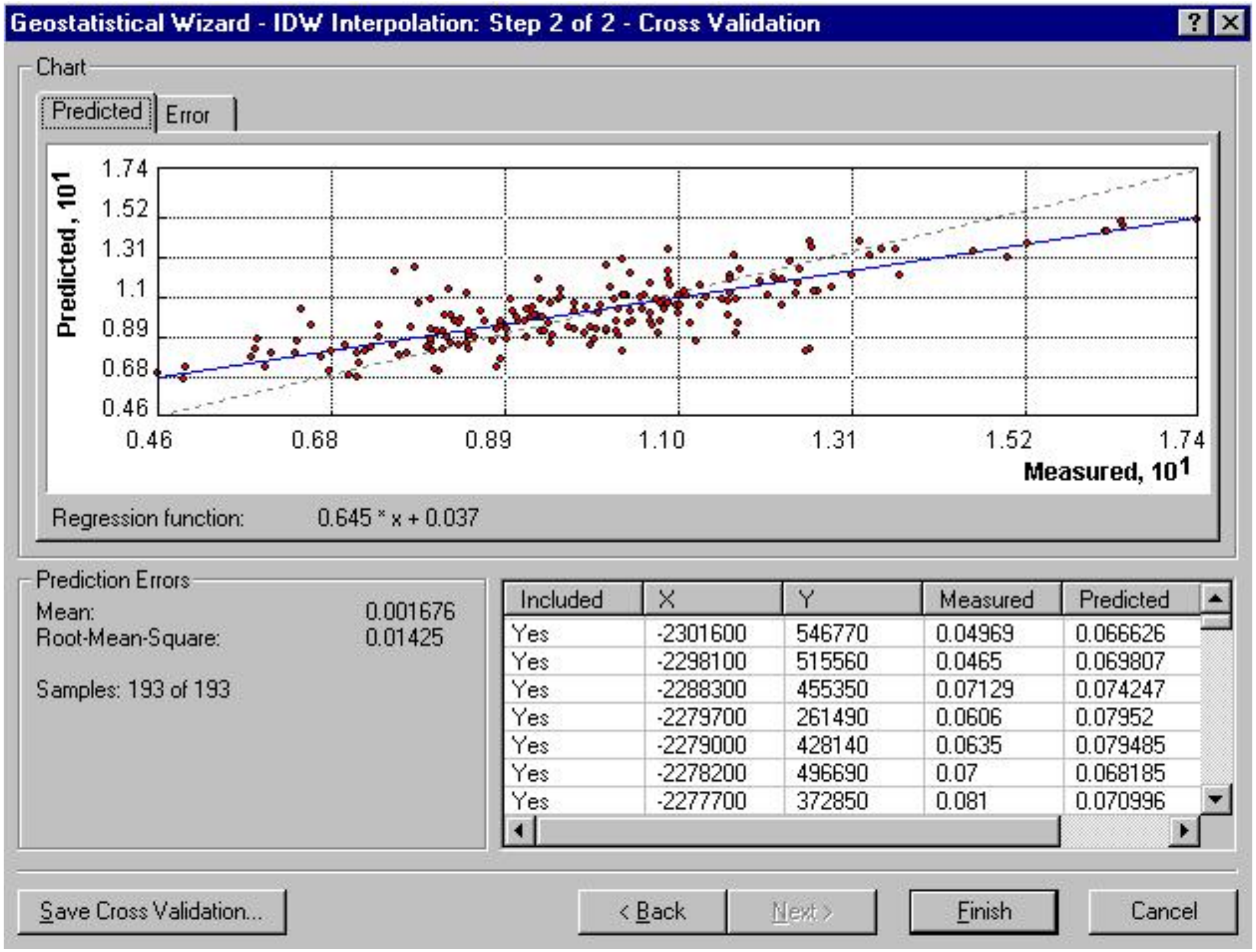


Further define interpolation method

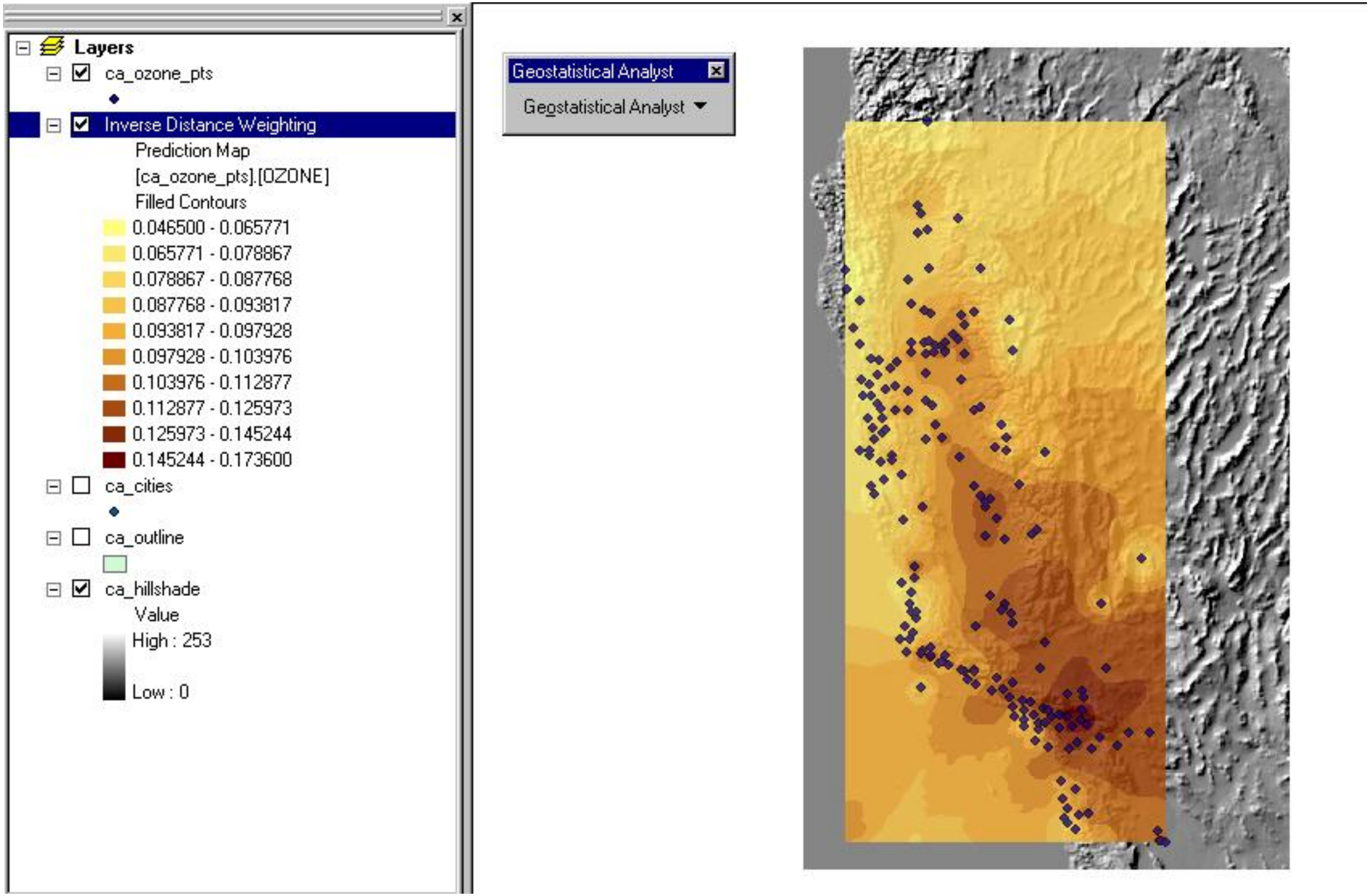


Cross validation

- removing one of the n observation points and using the remaining n-1 points to predict its value.
- Error = observed - predicted



Result


Layers
ca_ozone_pts
Inverse Distance Weighting
Prediction Map
[ca_ozone_pts].[OZONE]
Filled Contours
0.046500 - 0.065771
0.065771 - 0.078867
0.078867 - 0.087768
0.087768 - 0.093817
0.093817 - 0.097928
0.097928 - 0.103976
0.103976 - 0.112877
0.112877 - 0.125973
0.125973 - 0.145244
0.145244 - 0.173600
ca_cities
ca_outline
ca_hillshade
Value
High : 253
Low : 0
Geostatistical Analyst
Geostatistical Analyst

In spatial statistics the theoretical **variogram** is a function describing the degree of spatial dependence of a spatial random field or stochastic process .

For instance in mining a variogram will give a measure of how much two samples taken from the mining area will vary in gold percentage depending on the distance between those samples. Samples taken far apart will vary more than samples taken close to each other.

The variogram is defined as the variance of the difference between field values at two locations (x and y) across realizations of the field (Cressie 1993):

$$2\gamma(x,y) = \text{var}\left(Z(x) - Z(y)\right) = E\left[\left((Z(x) - \mu(x)) - (Z(y) - \mu(y))\right)^2\right].$$

If the spatial random field has constant mean μ, this is equivalent to the expectation for the squared increment of the values between locations x and y (Wackernagel 2003) (where x and y are not coordinates but points in space):

$$2\gamma(x,y) = E\left[\left(Z(x) - Z(y)\right)^2\right],$$

where $\gamma(x,y)$ itself is called the **semivariogram**. In the case of a stationary process, the variogram and semivariogram can be represented as a function $\gamma_s(h) = \gamma(0, 0+h)$ of the difference $h = y - x$ between locations only, by the following relation (Cressie 1993):

$$\gamma(x,y) = \gamma_s(y - x).$$

If the process is furthermore isotropic, then the variogram and semivariogram can be represented by a function $\gamma_i(h) := \gamma_s(he_1)$ of the distance $h = \|y - x\|$ only (Cressie 1993):

$$\gamma(x,y) = \gamma_i(h).$$

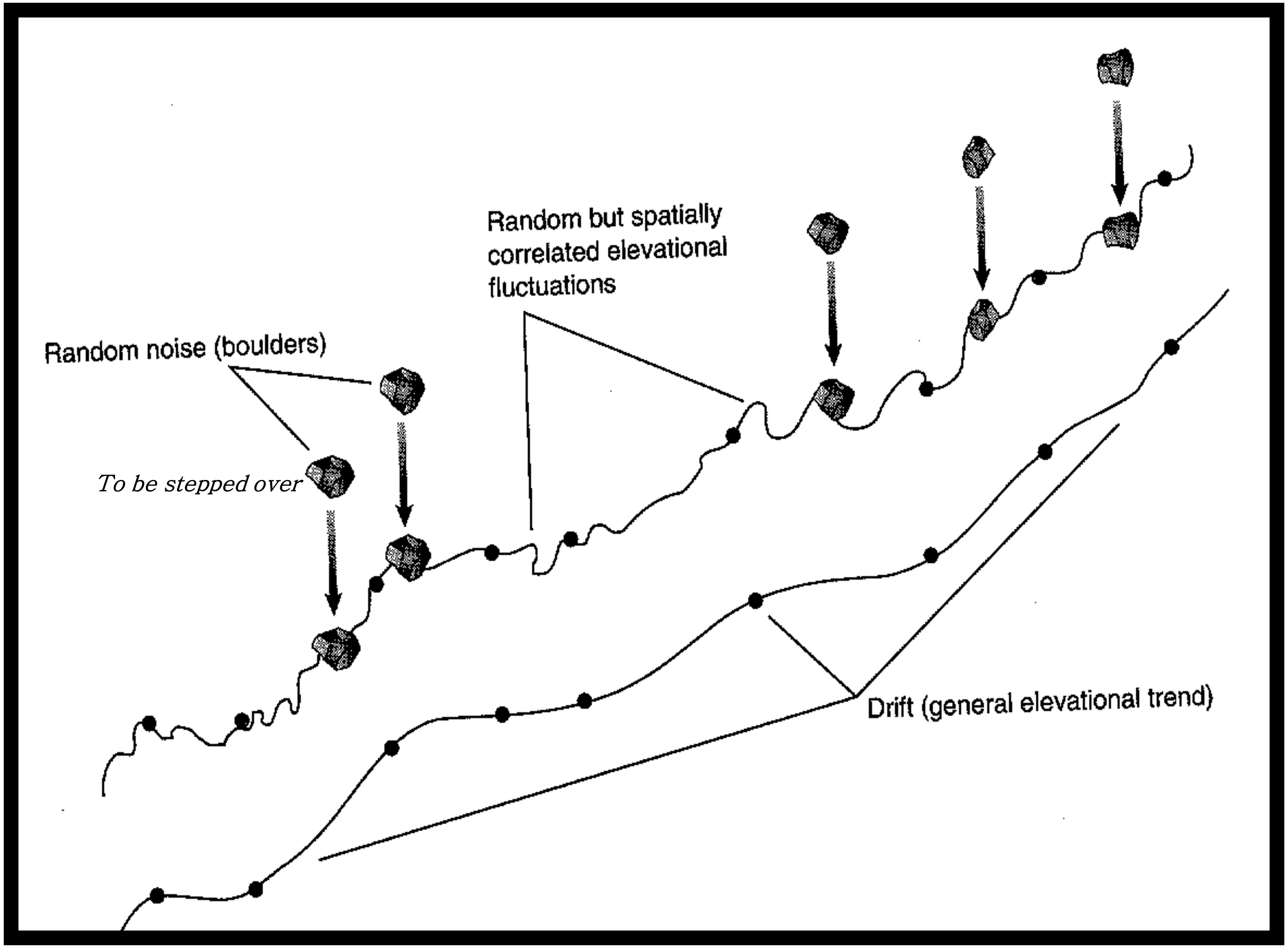
The indexes i or s are typically not written. The terms are used for all three forms of the function. Moreover, the term "variogram" is sometimes used to denote the semivariogram, and the symbol γ is sometimes used for the variogram, which brings some confusion.

Kriging

- Assumes distance or direction between sample points shows a spatial correlation that help describe the surface
- Fits function to
  - Specified number of points OR
  - All points within a window of specified radius
- Based on an analysis of the data, then an application of the results of this analysis to interpolation
- Most appropriate when you already know about spatially correlated distance or directional bias in data
- **Involves several steps**
  - **Exploratory statistical analysis of data**
  - **Variogram modeling**
  - **Creating the surface based on variogram**

Kriging

- Breaks up topography into 3 elements: Drift (general trend), small deviations from the drift and random noise.

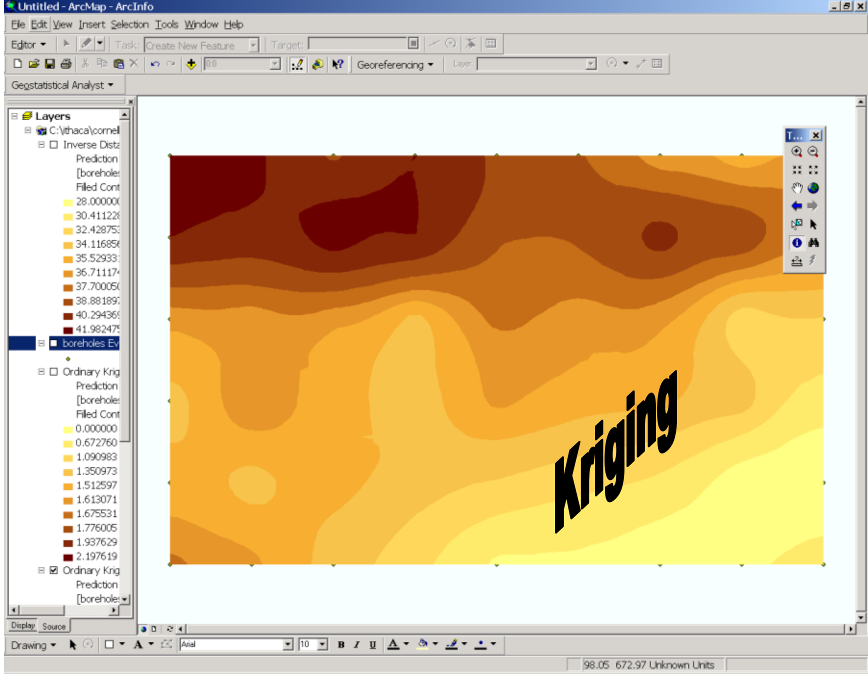


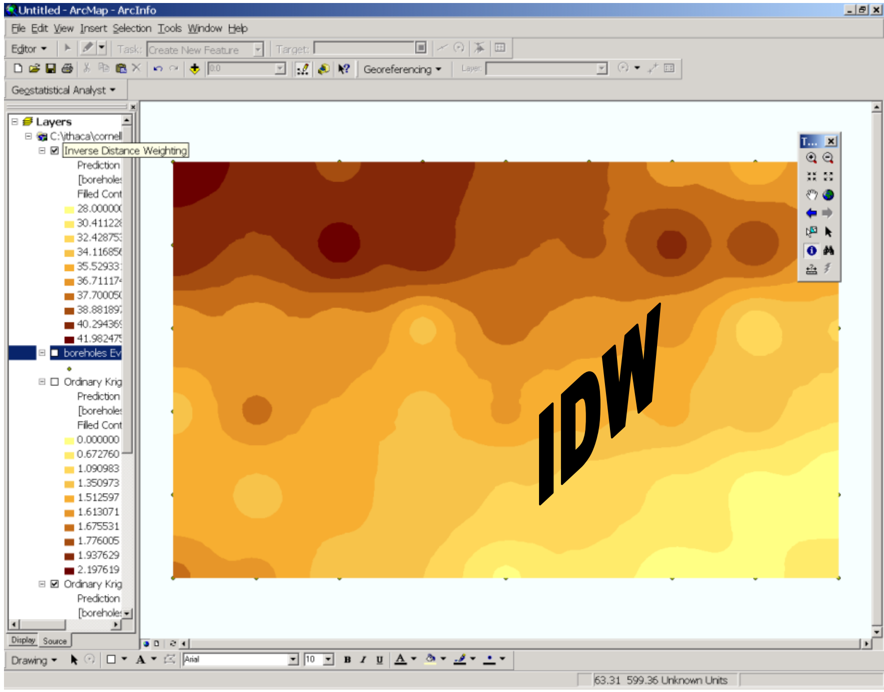
Kriging Results

- Once the variogram has been developed, it is used to estimate distance weights for interpolation
- Computationally very intensive w/ lots of data points
- Estimation of the variogram complex
  - No one method is absolute best
  - Results never absolute, assumptions about distance, directional bias

IDW vs. Kriging

- Kriging appears to give a more “smooth” look to the data
- Kriging avoids the “bulls eye” effect
- Kriging gives us a standard error





Spatial Modeling:
Thinking About Models

- Models can be:
  - Descriptive or Prescriptive
  - Deterministic or Stochastic
  - Static or Dynamic
  - Deductive or Inductive

Spatial Modeling

- General Types of (Spatial) Models
  - Descriptive: characterization of the distribution of spatial phenomena
  - Explanatory: deal with the variables impacting the distribution of a phenomena
  - Predictive: once explanatory variables are identified, predictive models can be constructed
  - Normative: models that provide optimal solutions to problems with quantifiable *objective functions* and *constraints*

Spatial Modeling

- More specific types of spatial models:
  - Binary models (descriptive): use logical expressions to identify or select map features that *do or do not meet certain criteria…*How?
  - Index models (descriptive): use index values calculated for variables to produce a ranked spatial surface…How?
    - Weighted Linear Combination Model
  - Regression models (explanatory or predictive): a dependent variable is related or explained by independent variables in an equation…How?
    - Linear and logistic regression
  - Process (explanatory or predictive): integrate existing knowledge about environmental processes into a set of relationships and equations for quantifying those processes…How?

Spatial Modeling

- Steps in the Modeling Process
  - Define the goals of the model
  - Break down the model into elements
  - Implementation and calibration of the model
  - Model validation
    - Sometimes difficult or not feasible

Spatial Modeling

- Important Issues in Conducting Spatial Analysis:
  - Delineation of geographic units of analysis
    - How do you choose geographic units of analysis so that spatial analyses are valid?
  - Identification of structural and spatial factors that impact spatial analysis
    - Structural – impact site
    - Spatial – impact situation (absolute and relative location, neighborhood effects)

The Role of GIS in Spatial Modeling

- How can GIS enable spatial modeling?
  - GIS is a tool that can integrate a myriad of data sources
  - GIS can incorporate raster and/or vector data into modeling schemes
  - Modeling may take place within a GIS, or require linking to other computer programs
    - Loose coupling
    - Tight coupling
    - Embedded System

Different approaches of coupling GIS with models

Three Types of Coupling Approaches:

- Coupling GIS with a stand-alone modeling package by exchanging files:
  - GIS is used mainly for pre-processing (data preparation) and post-processing (display/visualization).
  - Requires an exchange format that is understood by both the GIS and the modeling packages (e.g., ArcView shapefile format).
  - This is the easiest approach.

(Source: Goodchild, 2005)

- Integrating GIS with modeling package using standards such as Microsoft’s COM and .NET that allow a single script to invoke commands from both packages:
  - This is now a common approach (e.g., ArcObjects with COM-compliant programming languages).

(Source: Goodchild, 2005)

- The entire model is executed by calling functions of the GIS using a single script.
  - In other words, the model is embedded in the GIS (e.g., Model Builder in ArcGIS).

(Source: Goodchild, 2005)

Linking GIS and Modeling Systems:

- Maguire (2005) suggests three similar approaches:

  - *Loose coupling* employs common file structures, file translators, and more recently, Web services messaging.

  - *Moderate integration* uses techniques such as remote procedure calls and shared database access.

  - *Tight integration* can be achieved by, for example, object-component calls, or function calls.

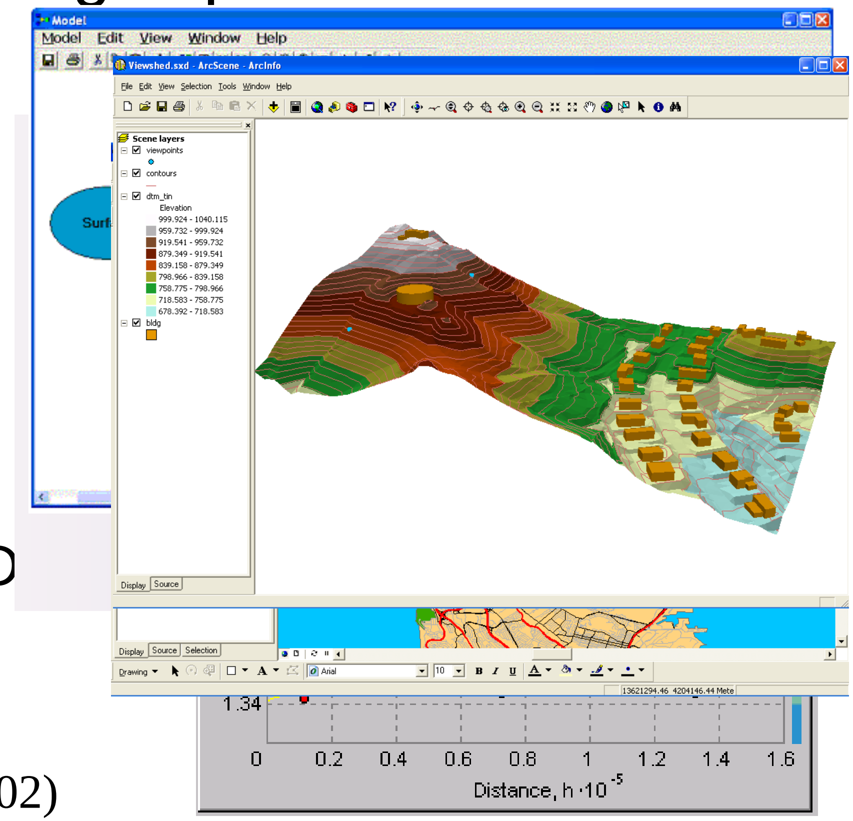
(Source: Maguire, 2005)

	Loose	Moderate	Tight
Time to integrate	Fast	Medium	Slow
Programmer expertise	Low	High	Medium
Execution speed	Slow	Medium	Fast
Simultaneous execution capability	Low	Low	High
Debugging	Easy	Moderate	Hard

(Source: Westervelt, 2002; Maguire, 2005)

GIS for Spatial Analysis and Modeling:

- Current leading commercial GIS software offers decent analysis and modeling capabilities in areas such as:
  - Vector overlay
  - Raster analysis
  - Cartographic modeling
  - Geostatistical estimation
  - Network analysis
  - Visualization of 2D and 2.5D



the probability of polygon C in Map 3 having clay soil and a scrubland cover is 0.8 × 0.6 = 0.48.

(Source: Maguire, 2005; Lo & Yeung, 2002)

- Current commercial GIS software need further development in areas such as:
  - Exploratory spatial data analysis (EDSA)
  - Dynamic system simulation
  - Operations research optimization
  - Spatial statistics
  - Visualization of multidimensional data

(Source: Maguire, 2005)

Software Architecture	Characteristics	Examples
GIS-centric Systems	- Extend GIS software system - GIS look and feel - Single UI, data model, data management, etc. - Geographic focus is explicit	- NatureServe Vista (Ecosystem management) - CATS (Emergency response) - MIKE BASIN (Hydrological basins)
Linked GIS-Modeling Systems	- Loose- and close-coupled options - Two UIs, data models, data management, etc. - Custom file translation or software linkage	- SLEUTH (Land-use estimation) - RAMAS GIS (Ecological population simulation) - TRANUS-ArcGIS (Transport planning)
Modeling-centric Systems	- Extend modeling systems - Modeling system look and feel - Single UI, data model, data management, etc. - Limited tools for visualization, data transformation, etc.	- SWARM (Wildfire management) - RePast (Infectious diseases, civil violence, etc.)

(Source: Maguire, 2005)

Model Sharing:

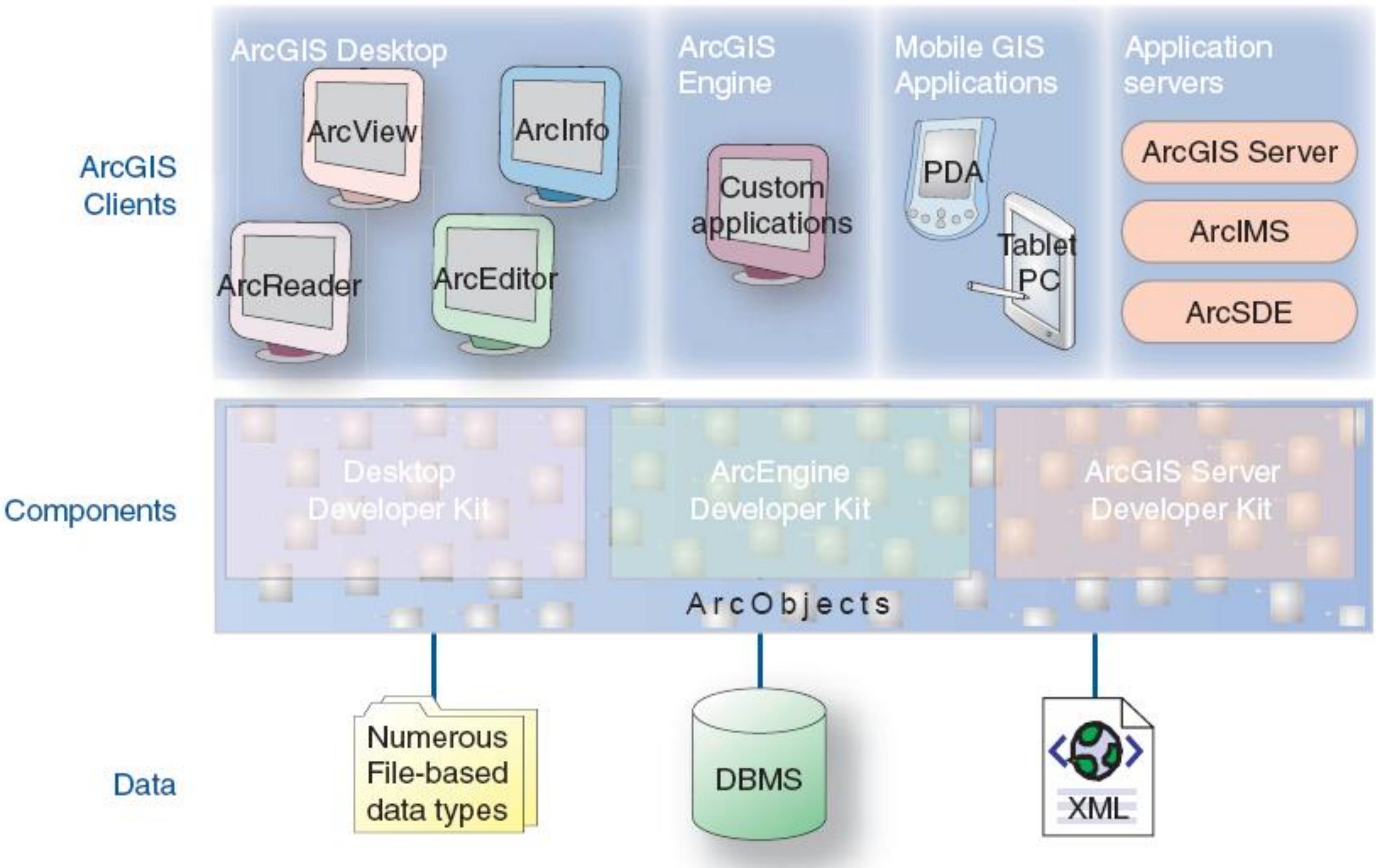
- There have been many efforts of promoting data sharing (e.g., data repositories, digital libraries, data warehouses). But, limited efforts are put into the creation of equivalent infrastructure for sharing methods and models.
  - "Model and method sharing, or more generally the sharing of process objects, is a core concept of the emerging Grid, the high-performance worldwide network of research computers, ….. In the next few years, dramatic improvements are expected in the availability of techniques for sharing methods and models." (Goodchild, 2005, p. 15)

(Source: Goodchild, 2005)

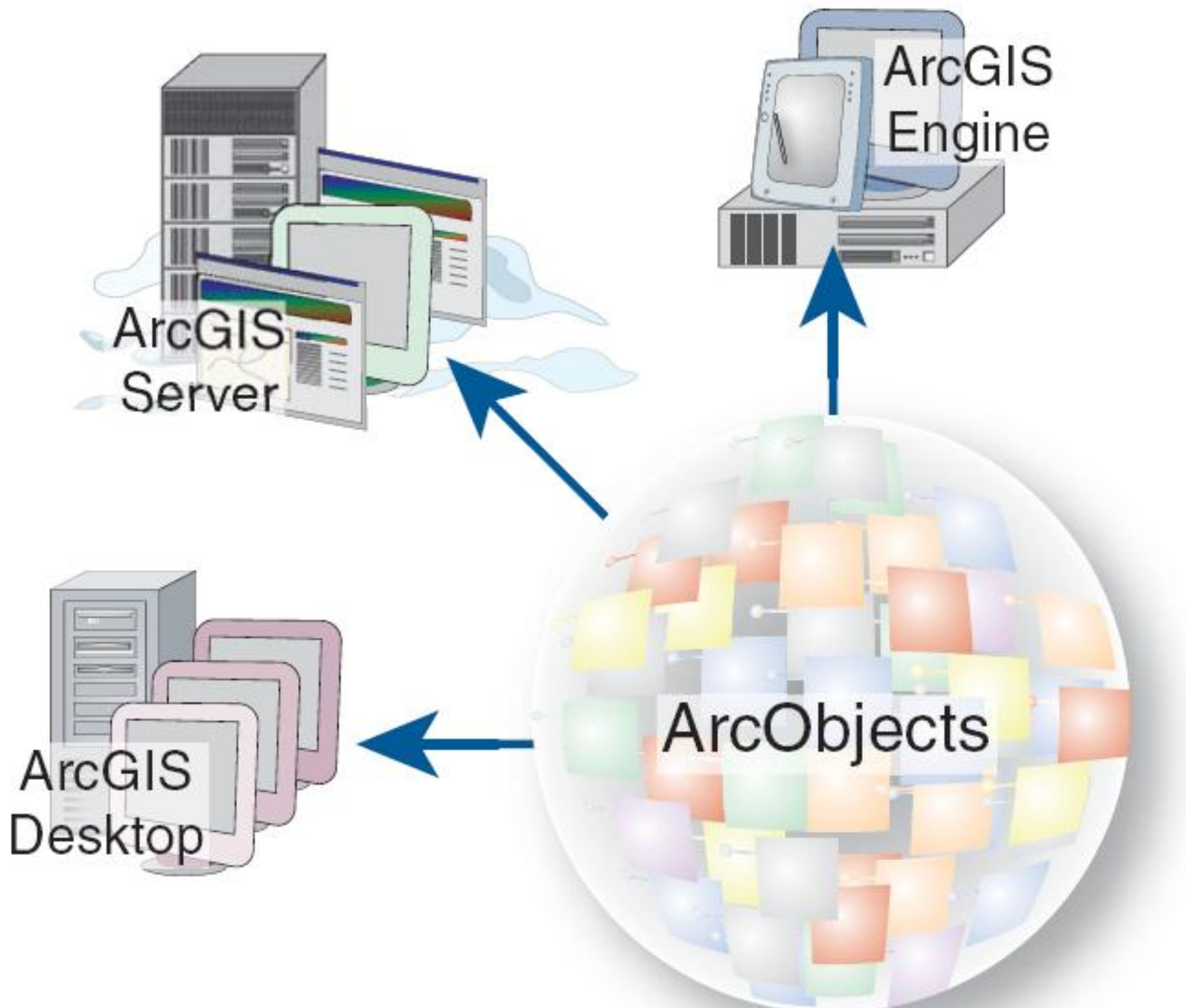
Four key parts of GIS software:

- *Information Model:* defines the classes of objects that can be represented from the domain of interest and how they behave and interact (e.g., vector, raster, TIN, image, etc.)
- *Data Manager:* commonly implemented as a database management system (DBMS).
- *Process Functions (Tools):* e.g., data input, management, query, analysis, visualization functions.
- *Interfaces:* include user interfaces (UI) and application programming interfaces (API)

(Source: Maguire, 2005)



ArcGIS Clients
Components
Data
ArcGIS Desktop
ArcView
ArcInfo
ArcReader
ArcEditor
ArcGIS Engine
Custom applications
Mobile GIS Applications
PDA
Tablet PC
Application servers
ArcGIS Server
ArcIMS
ArcSDE
Desktop Developer Kit
ArcEngine Developer Kit
ArcGIS Server Developer Kit
ArcObjects
Numerous File-based data types
DBMS
XML


ArcGIS
Engine
ArcGIS
Server
ArcGIS
Desktop
ArcObjects

What is ArcObjects?

- ArcGIS system is built and extended using software components called *ArcObjects*, which includes a large number of programmable components.
- Users can work with ArcObjects using standard programming frameworks to extend ArcGIS Desktop, build custom applications with ArcGIS Engine, and implement enterprise GIS applications using ArcGIS Server.

- ArcObjects is a set of platform-independent software components, written in C++, that provides services to support GIS applications on the desktop, in the form of thick and thin clients, and on the server.

- ArcObjects makes use of the Microsoft Component Object Model (COM). For operating systems other than Microsoft Windows, this infrastructure must be provided for the ArcObjects system to function.

Software Development Kit (SDK):

- *ArcGIS Desktop* software developer kit (SDK) supports the COM and .NET programming frameworks. Developers can apply the ArcGIS Desktop SDK to add extended functions, new GIS tools, custom user interfaces, and full extensions for improving professional GIS productivity of the ArcGIS Desktop applications.

- *ArcGIS Server* developer kit enables developers to build central GIS servers to host GIS functions that are accessed by many users, perform back office processing on large central GIS databases, build and deliver GIS Web applications, and perform distributed GIS computing.

- *ArcGIS Engine* is a simple, application-neutral programming environment for ArcObjects.
  - Its SDK provides a series of embeddable ArcGIS components that are *used outside the ArcGIS Desktop application framework*.
  - Using the ArcGIS Engine Developer Kit, developers can build focused GIS solutions with simple interfaces to access GIS functions or embed GIS logic in existing user applications to deploy GIS to broad groups of users.
  - ArcGIS Engine is *supported on Windows, Solaris, and Linux (Intel),* developers can create cross-platform custom solutions.

Choosing API and Development Env.:

- ArcGIS Engine Developer Kit provides four developer APIs—COM, .NET, Java, and C++.
- ESRI recommends and supports the following integrated development environments (IDEs) or compilers when working with ArcGIS Engine.

COM:

- Visual Basic 6 sp3 or later
- Visual C++ 6 sp3 or later
- Visual C++ (Visual Studio .NET 2003)

(Source: ESRI, 2004)

.NET:

- C# (Visual Studio .NET 2003 with .NET Framework 1.1)
- VB.NET (Visual Studio .NET 2003 with .NET Framework 1.1)

Java:

- Eclipse v. 3.0 or 3.0.1
- JBuilder X
- NetBeans 3.6

(Source: ESRI, 2004)

C++ (Compilers):

– Visual C++ sp3 or later for Windows
– Visual C++ (VS.NET 2003) for Windows
– GCC 3.2 C++ for Linux (Intel)
– WorkShop 6 Update 2 for Sun Solaris

- The COM and .NET APIs are only supported on the Microsoft Windows platform, while the Java and C++ APIs are supported on all the platforms supported by ArcGIS Engine.

(Source: ESRI, 2004)

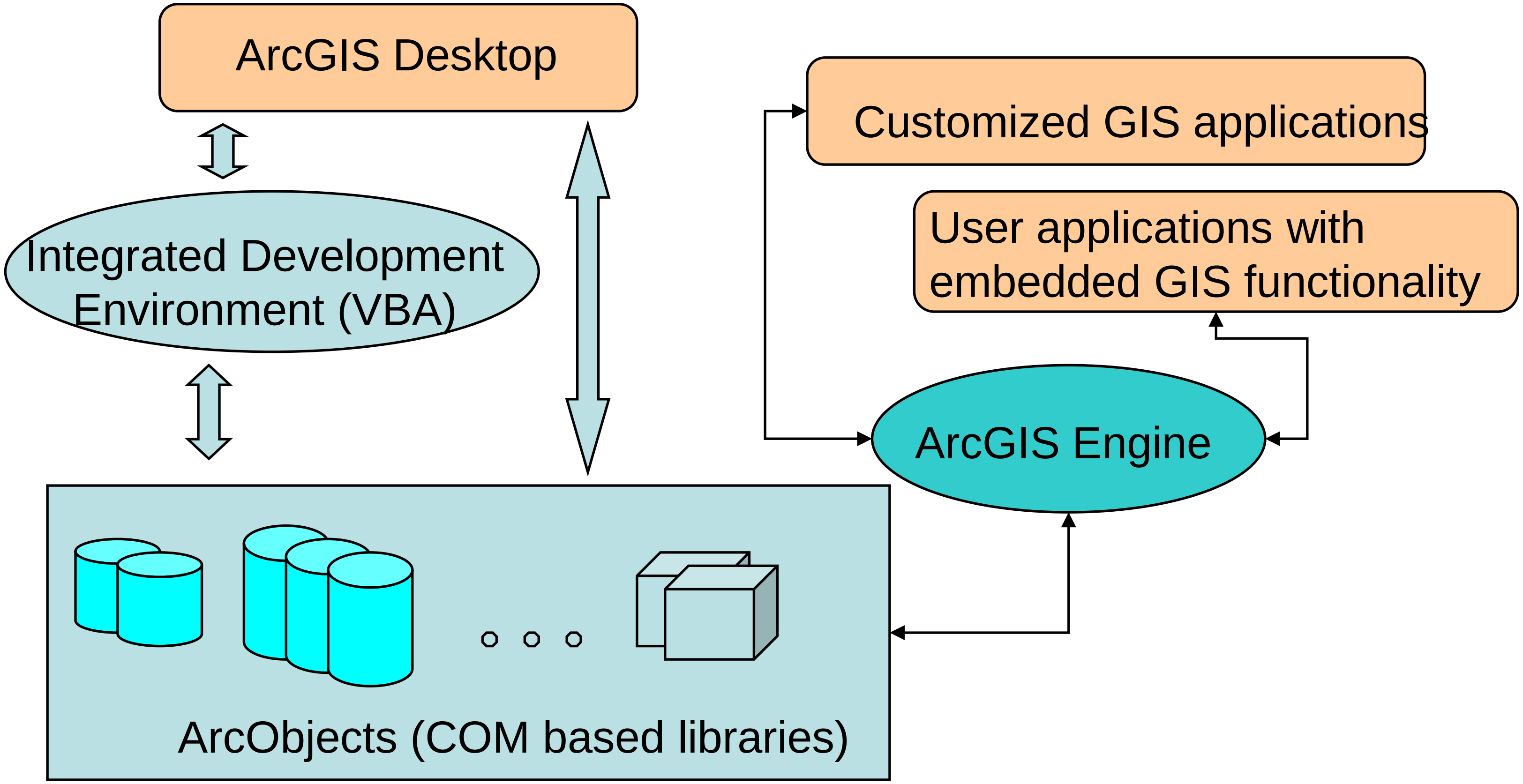
Integration of ArcGIS with Dynamic modeling

GIS (software design)

- Advantage of GIS systems (in general):
  - Provides user-friendly interfaces, interactive and intuitive methods
  - Supports multiple concurrent data accesses and inquires (Database)

- Limitations:
  - Designed or optimized for spatial component only
  - Large software system design overheads, performance is one of the key concerns at many cases

Software structure of two ESRI products

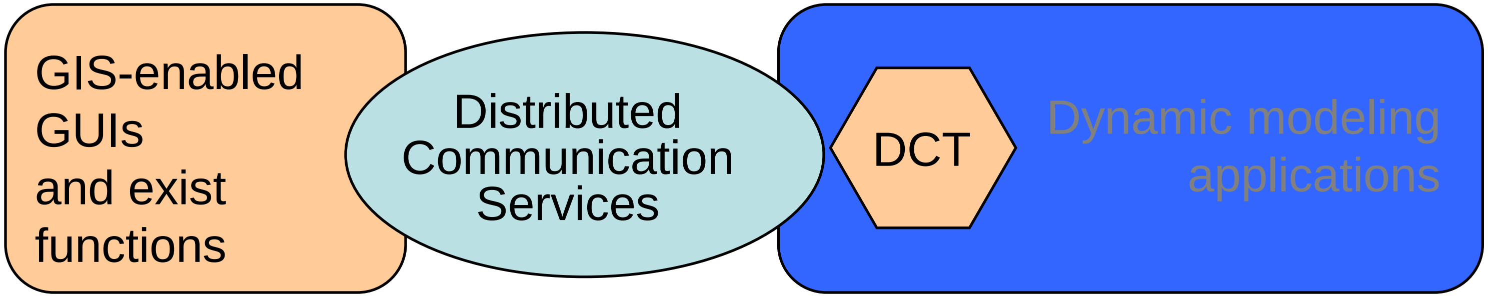


Dynamic modeling (software design)

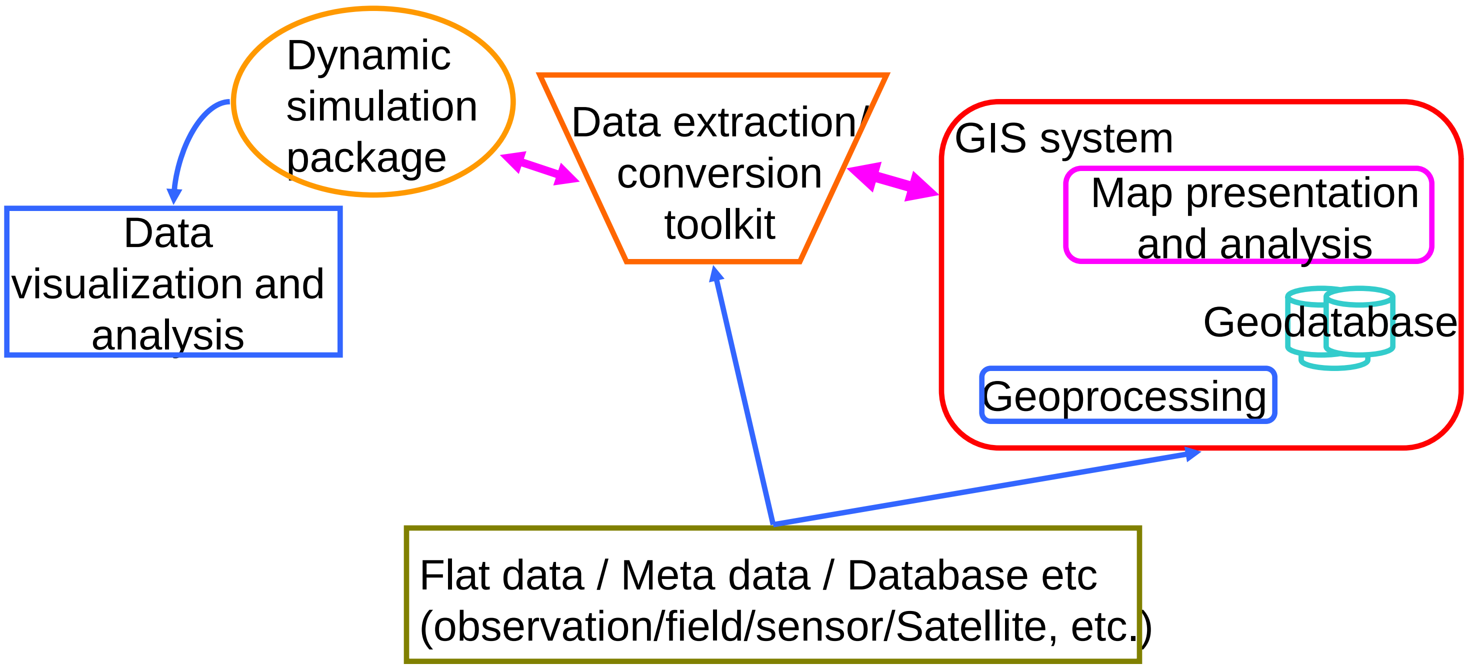
- Strong temporal component
- Users may already have hundreds of thousands in-house code
- Complex data-structures to expedite the modeling of species dynamics
- Maybe need high-performance oriented software design
- Limitations - user-friendly interfaces for control and analysis

An Integration Model

- Side-by-side deployment -- key concept of component based architecture
- A Solution : Dedicated Distributed Simulation Framework with Data Conversion Toolkits (DCT)



Another Integrated Model



Advantages

- Time to solution (coding + testing)
- No (or very few) extra trainings needed for users (managers)
- Flexible for further developments, either on the GIS side and dynamic modeling side
- Easier software maintenance (no need to catch up so closely with new ArcGIS products). The lifetime of a library (i.e., ArcObject) is likely much longer than that of a ArcGIS product (Integrated Development Environment VBA)
- Supports simulations on multiple computational platforms (Linux/Windows/Unix – even HPCs)

FOSS4G is now a worldwide phenomenon

FOSS4G NORTH AMERICA • 2015 •

FOSS4G SRI LANKA - 2014

FOSS4G EUROPE BREMEN2014

FOSS4G Sydney 2009
20 - 23 October
Free and Open Source Software for Geospatial

CAPE TOWN SOUTH AFRICA
FOSS4G 2008
2008 FREE AND OPEN SOURCE SOFTWARE FOR GEOSPATIAL CONFERENCE

FOSS4G NOTTINGHAM 2013

FOSS4G KOREA 2014

FOSS4G DENVER 2011

FOSS4G-CE Bucharest 2013

10th Anniversary FOSS4G PDX·2014

FOSS4G
Free&Open Source Software for Geospatial
2013 Tokyo / Osaka

MAY THE FOSS BE WITH YOU

FOSS4G-ASIA
2-5 dec 2014 - BANGKOK

FOSS4G-FR

FOSS4G EUROPE COMO 2015

FOSS4G SEOUL 2015

Home Brewed at

Japan, Korea, India, Malaysia, The Philippines, Sri Lanka…and Asia

October 2006 OSGeo - India formed

Your Open Source Compass

20

O   op at

G

In

RSA

>2

Thank you for your attention!!