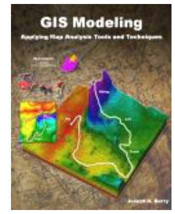
Beyond Mapping IV

Topic 9 – A Math/Stat Framework for Map Analysis



GIS Modeling book

<*Click here*> *for a printer-friendly version of this topic (.pdf).*

(*Back to the Table of Contents*)

SpatialSTEM Has Deep Mathematical Roots

(GeoWorld, January 2012)

(return to top of Topic)

Recently my interest has been captured by a new arena and expression for the contention that "maps are data"—*spatial*STEM (or *s*STEM for short)—as a means for redirecting education in general, and GIS education in particular. I suspect you have heard of STEM (Science, Technology, Engineering and Mathematics) and the educational crisis that puts U.S. students well behind many other nations in these quantitatively-based disciplines.

While Googling around the globe makes for great homework in cultural geography, it doesn't advance quantitative proficiency, nor does it stimulate the spatial reasoning skills needed for problem solving. Lots of folks from Freed Zakaria to Bill Gates to President Obama are looking for ways that we can recapture our leadership in the quantitative fields. That's the premise of *spatial*STEM– that "maps are numbers first, pictures later" and we do mathematical things to mapped data for insight and better understanding of spatial patterns and relationships within decision-making contexts.

This contention suggests that there is a map-*ematics* that can be employed to solve problems that go beyond mapping, geo-query, visualization and GPS navigation. This column's discussion about the quantitative nature of maps is the first part of a three-part series that sets the stage to fully develop this thesis— that grid-based <u>*Spatial Analysis Operations* are extensions of traditional mathematics</u> (Part 2 investigating map math, algebra, calculus, plane and solid geometry, etc.) and that grid-based <u>*Spatial Statistics Operations* are extensions of traditional statistics</u> (Part 3 looking at map descriptive statistics, normalization, comparison, classification, surface modeling, predictive statistics, etc.).

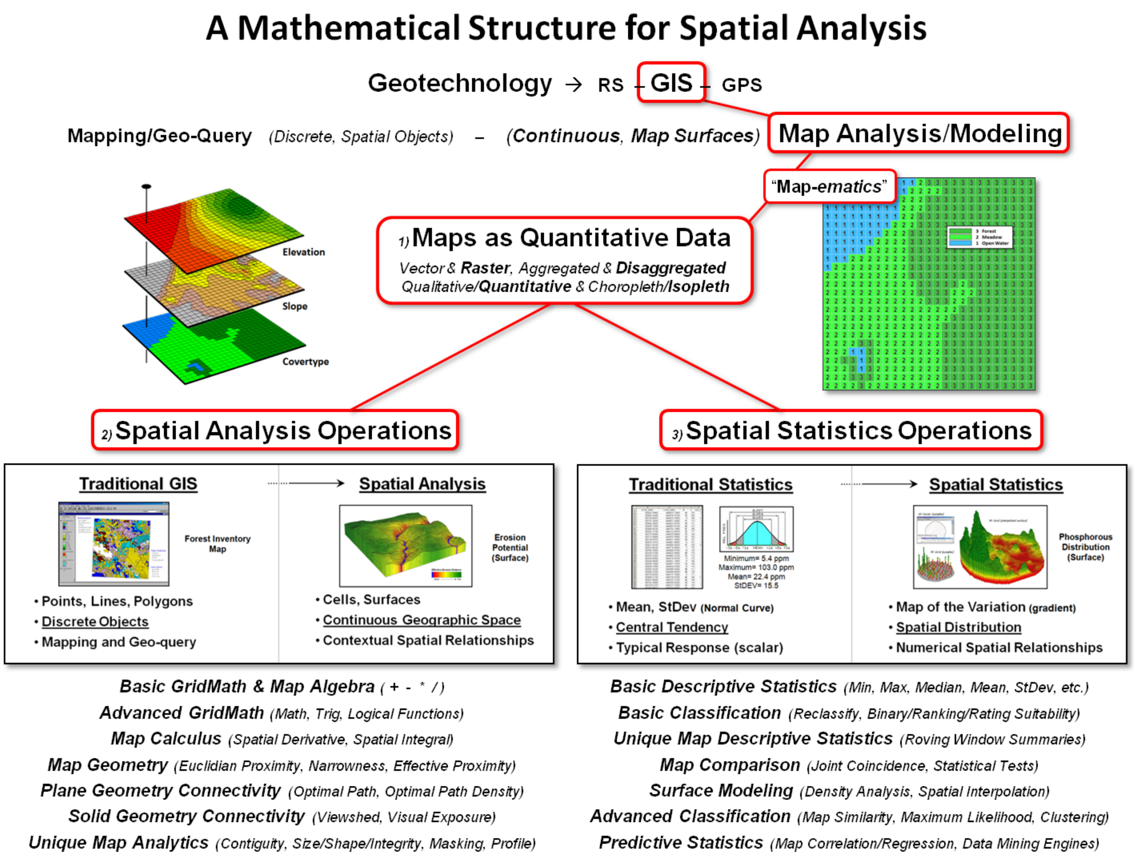


Figure 1. Conceptual overview of the SpatialSTEM framework.

Figure 1 outlines the important components of map analysis and modeling within a mathematical structure that has been in play since the 1980s (see author's note). Of the three disciplines forming Geotechnology (Remote Sensing, Geographic Information Systems and Global Positioning System), GIS is at the heart of converting mapped data into spatial information. There are two primary approaches used in generating this information—*Mapping/Geo-query* and *Map Analysis/Modeling*.

The major difference between the two approaches lies in the structuring of mapped data and their intended use. Mapping and geo-query utilizes a data structure akin to manual mapping in which discrete spatial objects (*points*, *lines* and *polygons*) form a collection of independent, irregular features to characterize geographic space. For example, a Water map might contain categories of Spring (points), Stream (lines) and Lake (polygons) with the features scattered throughout a landscape.

Map analysis and modeling procedures, on the other hand, operate on continuous map variables (termed map *surfaces*) composed of thousands upon thousands of map values stored in geo-registered matrices. Within this context, a Water map no longer contains separate and distinct features but is a collection of adjoining grid cells with a map value indicating the characteristic at each location (e.g., Spring=1, Stream= 2 and Lake= 3).

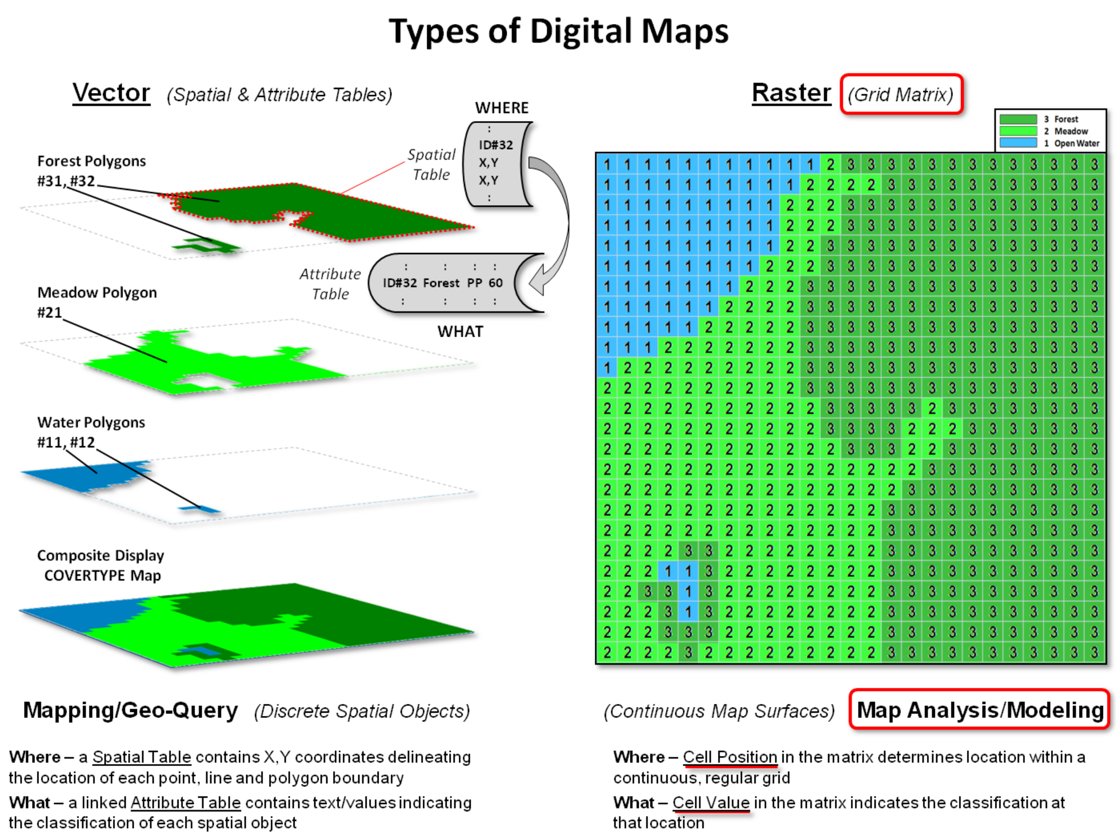


Figure 2. Basic data structure for Vector and Raster map types.

Figure 2 illustrates two broad types of digital maps, formally termed *Vector* for storing discrete spatial objects and *Raster* for storing continuous map surfaces. In vector format, spatial data is stored as two linked data tables. A "spatial table" contains all of the X,Y coordinates defining a set of spatial objects that are grouped by object identification numbers. For example, the location of the Forest polygon identified on the left side of the figure is stored as ID#32 followed by an ordered series of X,Y coordinate pairs delineating its border (connect-the-dots).

In a similar manner, the ID#s and X,Y coordinates defining the other cover type polygons are sequentially listed in the table. The ID#s link the spatial table (Where) to a corresponding "attribute table" (What) containing information about each spatial object as a separate record. For example, polygon ID#31 is characterized as a mature 60 year old Ponderosa Pine (PP) Forest stand.

The right side of figure 2 depicts raster storage of the same cover type information. Each grid space is assigned a number corresponding to the dominant cover type present— the "cell position" in the matrix determines the location (Where) and the "cell value" determines the characteristic/condition (What). It is important to note that the raster representation stores information about the interior of polygons and "pre-conditions geographic space" for analysis by applying a consistent grid configuration to each grid map. Since each map's underlying data structure is the same, the computer simply "hits disk" to get information and does not have to calculate whether irregular sets of points, lines or polygons on different maps intersect.

Figure 3 depicts the fundamental concepts supporting raster data. As a comparison between vector and raster data structures consider how the two approaches represent an Elevation surface. In vector, contour lines are used to identify lines of constant elevation and contour interval polygons are used to identify specified ranges of elevation. While contour lines are exacting, they fail to describe the intervening surface configuration.

Contour intervals describe the interiors but overly generalize the actual "ups and downs" of the terrain into broad ranges that form an unrealistic stair-step configuration (center-left portion of figure 3). As depicted in the figure, rock climbers would need to summit each of the contour interval "200-foot cliffs" rising from presumed flat mesas. Similarly, surface water flow presumably would cascade like waterfalls from each contour interval "lake" like a Spanish multi-tiered fountain.

The upshot is that within a mathematical context, vector maps are ineffective representations of real-world gradients and actual movements and flows over these surfaces— while contour line/interval maps have formed colorful and comfortable visualizations for generations, the data structure is too limited for modern map analysis and modeling.

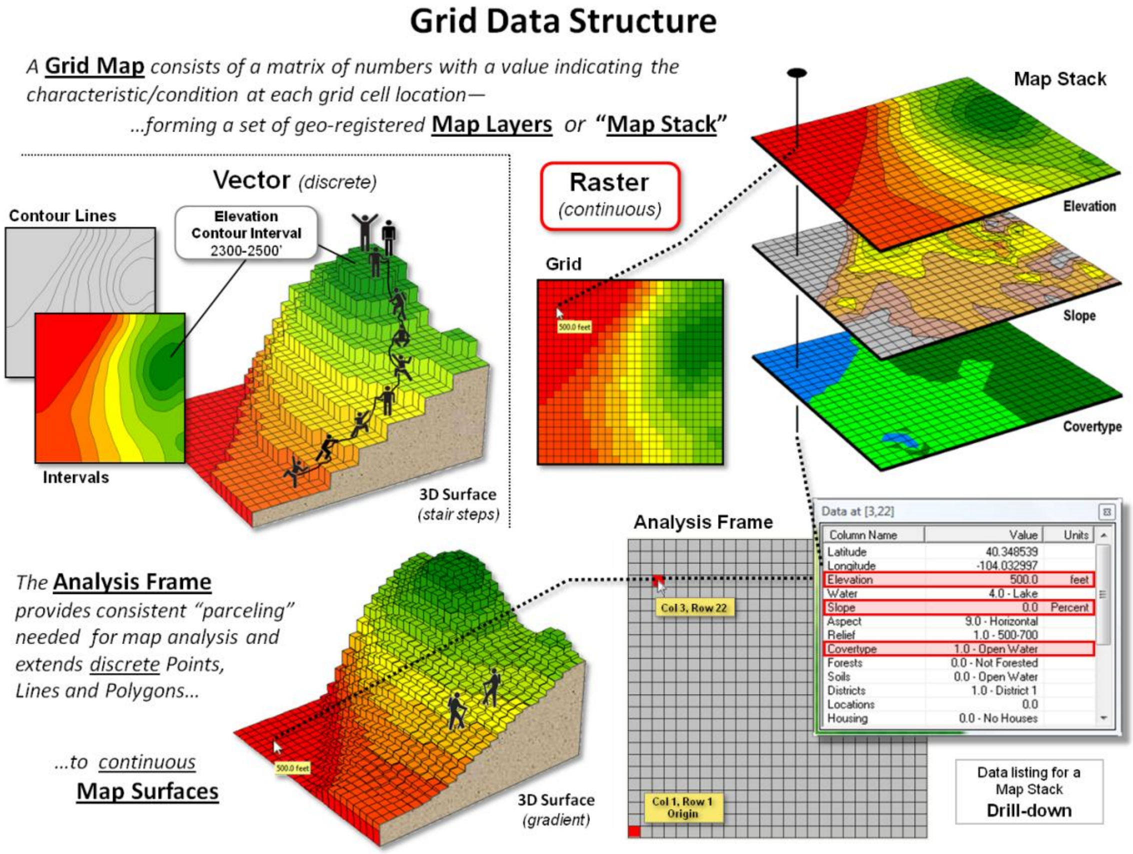


Figure 3. Organizational considerations and terminology for grid-based mapped data.

The remainder of figure 3 depicts the basic Raster/Grid organizational structure. Each grid map is termed a *Map Layer* and a set of geo-registered layers constitutes a *Map Stack*. All of the map layers in a project conform to a common *Analysis Frame* with a fixed number of rows and columns at a specified cell size that can be positioned anywhere in geographic space. As in the case of the Elevation surface in the lower-left portion of figure 3, a continuous gradient is formed with subtle elevation differences that allow hikers to step from cell to cell while considering relative steepness. Or surface water to sequentially stream from a location to its steepest downhill neighbor thereby identifying a flow-path.

The underlying concept of this data structure is that <u>grid cells for all of the map layers precisely coincide</u>, and by simply accessing map values at a row, column location a computer can "drill" down through the map layers noting their characteristics. Similarly, noting the map values of surrounding cells identifies the characteristics within a location's vicinity on a given map layer, or set of map layers.

Keep in mind that while terrain elevation is the most common example of a map surface, it is by no means the only one. In natural systems, temperature, barometric pressure, air pollution concentration, soil chemistry and water turbidity are but a few examples of continuous mapped data gradients. In human systems, population density, income level, life style concentration, crime occurrence, disease incidence rate all form continuous map surfaces. In economic systems, home values, sales activity and travel-time to/from stores form map variables that that track spatial patterns.

In fact the preponderance of spatial data is easily and best represented as grid-based continuous map surfaces that are preconditioned for use in map analysis and modeling. The computer does the heavy-lifting of the computation …what is needed is a new generation of creative minds that goes beyond mapping to "thinking with maps" within this less familiar, quantitative framework— a *Spatial*STEM environment.

Author's Notes: *My involvement in map analysis/modeling began in the 1970s with doctoral work in computer-assisted analysis of remotely sensed data a couple of years before we had civilian satellites. The extension from digital imagery classification using multivariate statistics and pattern recognition algorithms in the 70s to a comprehensive grid-based mathematical structure for all forms of mapped data in the 80s was a natural evolution. See www.innovativegis.com, select "Online Papers" for a link to a 1986 paper on "A Mathematical Structure for Analyzing Maps" that serves as an early introduction to a comprehensive framework for map analysis/modeling.*

Simultaneously Trivializing and Complicating GIS

(GeoWorld, April 2012)

(return to top of Topic)

Several things seem to be coalescing in my mind (or maybe colliding is a better word). GIS has moved up the technology adoption curve from *Innovators* in the 1970s to *Early Adopters* in the 80s, to *Early Majority* in the 90s, to *Late Majority* in the 00s and is poised to capture the *Laggards* this decade. Somewhere along this progression, however, the field seems to have bifurcated along technical and analytical lines.

The lion's share of this growth has been GIS's ever expanding capabilities as a "*technical tool*" for corralling vast amounts of spatial data and providing near instantaneous access to remote sensing images, GPS navigation, interactive maps, asset management records, geo-queries and awesome displays. In just forty years GIS has morphed from boxes of cards passed through a window to a megabuck mainframe that generated page-printer maps, to today's sizzle of a 3D fly-through rendering of terrain anywhere in the world with back-dropped imagery and semi-transparent map layers draped on top—all pushed from the cloud to a GPS enabled tablet or smart phone. What a ride!

However, GIS as an "*analytical tool*" hasn't experienced the same meteoric rise—in fact it might be argued that the analytic side of GIS has somewhat stalled over the last decade. I suspect that in large part this is due to the interests, backgrounds, education and excitement of the ever enlarging GIS tent. Several years ago (see figure 1 and author's note 1) I described the changes in breadth and depth of the community as flattening from the 1970s through the 2000s. By sheer numbers, the balance point has been shifting to the right toward general and public users with commercial systems responding to market demand for more technological advancements.

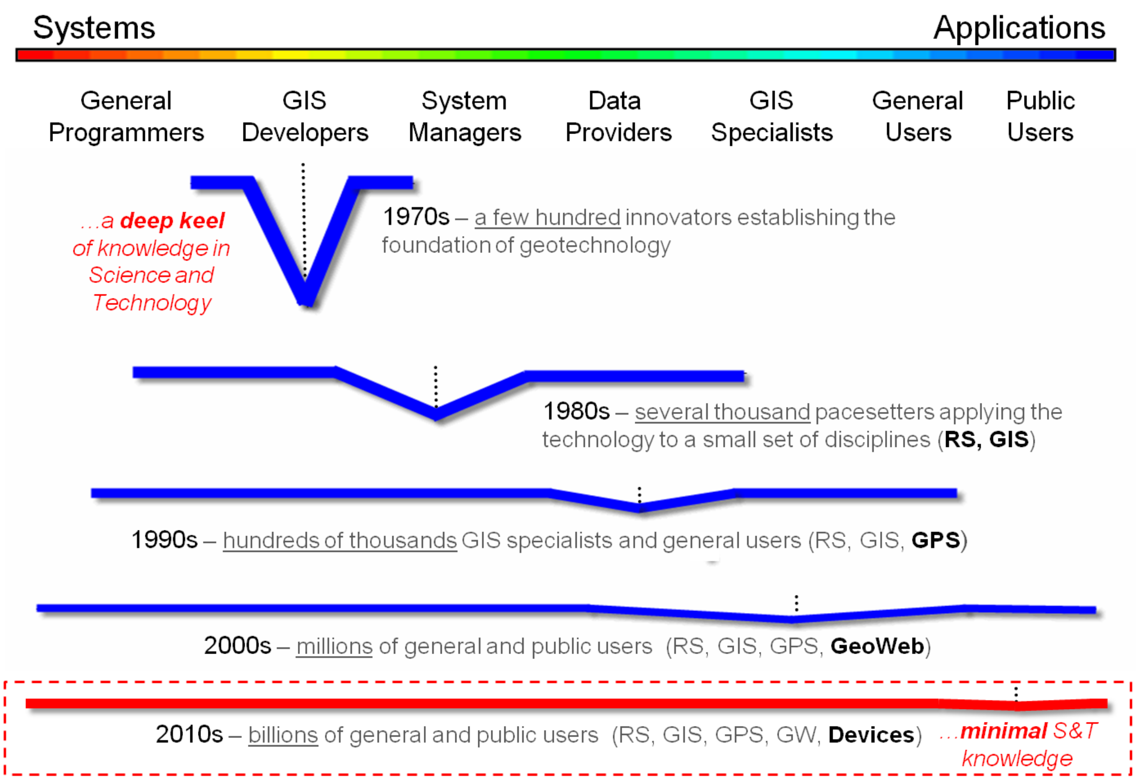


Figure 1. Changes in breadth and depth of the community.

The 2010s will likely see billions of general and public users with the average depth of science and technology knowledge supporting GIS nearly "flatlining." Success stories in quantitative map analysis and modeling applications have been all but lost in the glitz n' flash of the technological whirlwind. The vast potential of GIS to change how society perceives maps, mapped data and their use in spatial reasoning and problem solving seems relatively derailed.

In a recent editorial in Science entitled *Trivializing Science Education*, Editor-in-Chief Bruce Alberts laments that "Tragically, we have managed to simultaneously trivialize and complicate science education" (author's note 2). A similar assessment might be made for GIS education. For most students and faculty on campus, GIS technology is simply a set of highly useful apps on their smart phone that can direct them to the cheapest gas for tomorrow's ski trip and locate the nearest pizza pub when they arrive. Or it is a Google fly-by of the beaches around Cancun. Or a means to screen grab a map for a paper on community-based conservation of howler monkeys in Belize.

To a smaller contingent on campus, it is career path that requires mastery of the mechanics, procedures and buttons of extremely complex commercial software systems for acquiring, storage, processing, and display spatial information. Both perspectives are valid. However neither fully grasps the radical nature of the digital map and how it can drastically change how we perceive and infuse spatial information and reasoning into science, policy formation and decision-making—in essence, how we can "think with maps."

A large part of missing the mark on GIS's full potential is our lack of "reaching" out to the larger science, technology, engineering and math (STEM) communities on campus by insisting 1) that non-GIS students interested in understanding map analysis and modeling must be tracked into general GIS courses that are designed for GIS specialists, and 2) that the material presented primarily focuses on commercial GIS software mechanics that GIS-specialists need to know to function in the workplace.

Map Analysis Toolbox

Grid Map Layers

GIS Perspective:

Reclassify *(Position, Value, Size, Shape, Contiguity)*
Overlay *(Location-specific, Region-wide)*
Distance *(Distance, Proximity, Movement, Optimal Path, Visual Exposure)*
Neighbors *(Characterizing Surface Configuration, Summarizing Values)*

Mathematical Perspective:

Basic GridMath & Map Algebra *(+ - * /)*
Advanced GridMath *(Math, Trig, Logical Functions)*
Map Calculus *(Spatial Derivative, Spatial Integral)*
Map Geometry *(Euclidian Proximity, Narrowness, Effective Proximity)*
Plane Geometry Connectivity *(Optimal Path, Optimal Path Density)*
Solid Geometry Connectivity *(Viewshed, Visual Exposure)*
Unique Map Analytics *(Contiguity, Size/Shape/Integrity, Masking, Profile)*

Figure 2. Alternative frameworks for quantitative map analysis.

Much of the earlier efforts in structuring a framework for quantitative map analysis has focused on how the analytical operations work within the context of *Focal*, *Local* and *Zonal* classification by Tomlin, or even my own the *Reclassify*, *Overlay*, *Distance* and *Neighbors* classification scheme (see top portion of figure 2 and author's note 3). The problem with these structuring approaches is that most STEM folks just want to understand and use the analytical operations properly—not appreciate the theoretical geographic-related elegance, or code the algorithm.

The bottom portion of figure 2 outlines restructuring of the basic spatial analysis operations to align with traditional mathematical concepts and operations (author's note 4). This provides a means for the STEM community to jump right into map analysis without learning a whole new lexicon or an alternative GIS-centric mindset. For example, the GIS concept/operation of *Slope*= spatial "derivative", *Zonal functions*= spatial "integral", *Eucdistance*= extension of "planimetric distance" and the Pythagorean Theorem to proximity, *Costdistance*= extension of distance to effective proximity considering absolute and relative barriers that is not possible in non-spatial mathematics, and *Viewshed*= "solid geometry connectivity".

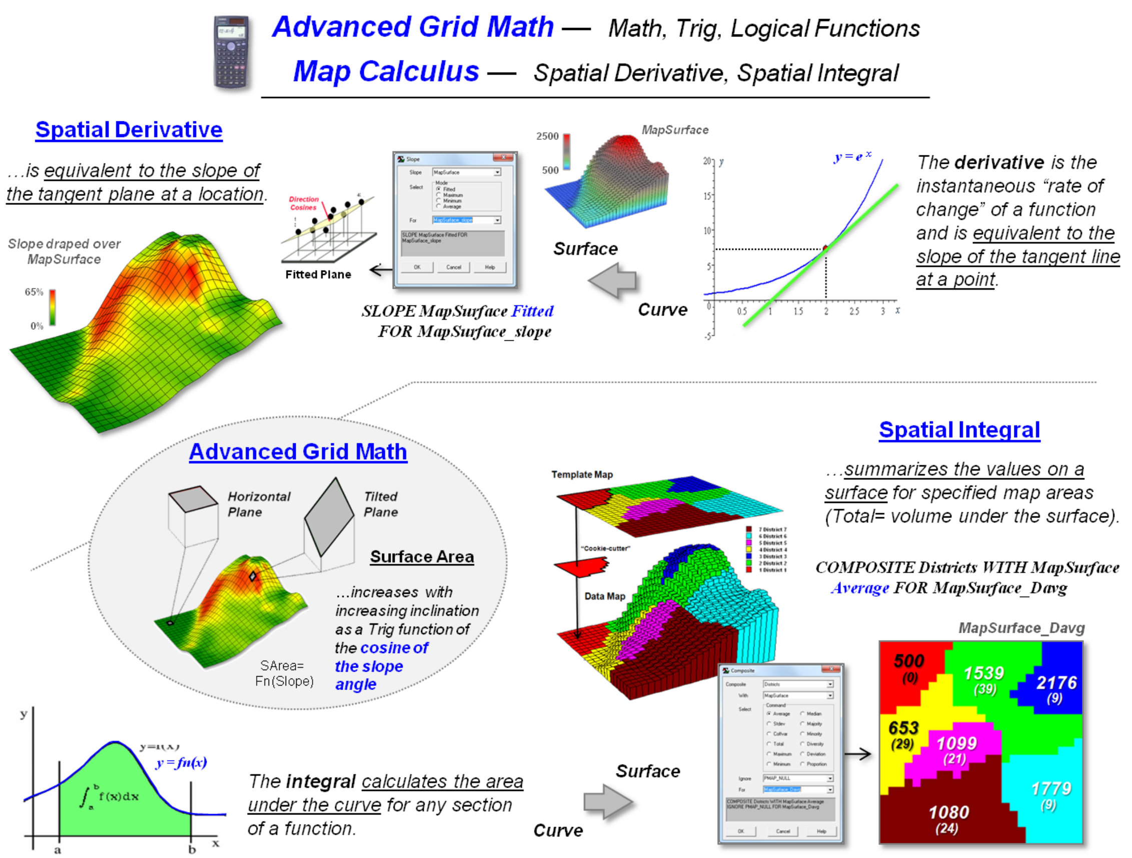


Figure 3. Conceptual extension of derivative, trigonometric functions and integral to mapped data and map analysis operations.

Figure 3 outlines the conceptual development of three of these operations. The top set of graphics identifies the *Calculus Derivative* as a measure of how a mathematical function changes as its input changes by assessing the slope along a curve in 2-dimensional abstract space—

calculated as the "slope of the tangent line" at any location along the curve. In an equivalent manner the *Spatial Derivative* creates a slope map depicting the rate of change of a continuous map variable in 3-dimensional geographic space—calculated as the slope of the "best fitted plane" at any location along the map surface.

Advanced Grid Math includes most of the buttons on a scientific calculator to include trigonometric functions. For example, calculating the "cosine of the slope values" along a terrain surface and then multiplying times the planimetric surface area of a grid cell will solve for the increased real-world surface area of the "inclined plane" at each grid location.

The *Calculus Integral* is identified as the "area of a region under a curve" expressing a mathematical function. The *Spatial Integral* counterpart "summarizes map surface values within specified geographic regions." The data summaries are not limited to just a total but can be extended to most statistical metrics. For example, the average map surface value can be calculated for each district in a project area. Similarly, the coefficient of variation ((Stdev / Average) * 100) can be calculated to assess data dispersion about the average for each of the regions.

By recasting GIS concepts and operations of map analysis within the general scientific language of math/stat we can more easily educate tomorrow's movers and shakers in other fields in "spatial reasoning"—to think of maps as "mapped data" and express the wealth of quantitative analysis thinking they already understand on spatial variables.

Innovation and creativity in spatial problem solving is being held hostage to a trivial mindset of maps as pictures and a non-spatial mathematics that presuppose mapped data can be collapsed to a single central tendency value that ignores the spatial variability inherent in the data. Simultaneously, the "build it (GIS) and they will come (and take our existing courses)" educational paradigm is not working as it requires potential users to become a GIS'perts in complicated software systems.

GIS must take an active leadership role in "leading" the STEM community to the similarities/differences and advantages/disadvantages in the quantitative analysis of mapped data—there is little hope that the STEM folks will make the move on their own. Next month we'll consider recasting spatial statistics concepts and operations into a traditional statistics framework.

Author's Notes: *1) See "A Multifaceted GIS Community" in Beyond mapping Compilation Series book III, Epilog section 2 posted at www.innovativegis.com. 2) Bruce Alberts in Science, 20 January 2012:Vol. 335 no. 6066 p. 263. 3) see "An Analytical Framework for GIS Modeling" posted at www.innovativegis.com/basis/Papers/Other/GISmodelingFramework/. 4) See "SpatialSTEM: Extending Traditional Mathematics and Statistics to Grid-based Map Analysis and Modeling" posted at www.innovativegis.com/basis/Papers/Other/SpatialSTEM/.*

Infusing Spatial Character into Statistics

(GeoWorld, May 2012)

(return to top of Topic)

The previous section discussed the assertion that we might be simultaneously trivializing and complicating GIS. At the root of the argument was the contention that "innovation and creativity in spatial problem solving is being held hostage to a trivial mindset of maps as pictures and a non-spatial mathematics that presuppose mapped data can be collapsed into a single central-tendency value that ignores the spatial variability inherent in data."

The discussion described a mathematical framework that organizes the spatial analysis toolbox into commonly understood mathematical concepts and procedures. For example, the GIS concept/operation of *Slope*= spatial "derivative," *Zonal functions*= spatial "integral," *Eucdistance*= extension of "planimetric distance" and the Pythagorean Theorem to proximity, *Costdistance*= extension of distance to effective proximity considering absolute and relative barriers that is not possible in non-spatial mathematics, and *Viewshed*= "solid geometry connectivity."

This section does a similar translation to describe a statistical framework for organizing the spatial statistics toolbox into commonly understood statistical concepts and procedures. But first we need to clarify the differences between spatial analysis and spatial statistics. *Spatial analysis* can be thought of as an extension of traditional mathematics involving the "contextual" relationships within and among mapped data layers. It focuses on geographic associations and connections, such as relative positioning, configurations and patterns among map locations.

Spatial statistics, on the other hand, can be thought of as an extension of traditional statistics involving the "numerical" relationships within and among mapped data layers. It focuses on mapping the variation inherent in a data set rather than characterizing its central tendency (e.g., average, standard deviation) and then summarizing the coincidence and correlation of the spatial distributions.

The top portion of figure 1 identifies the two dominant GIS perspectives of spatial statistics—*Surface Modeling* that derives a continuous spatial distribution of a map variable from point sampled data and *Spatial Data Mining* that investigates numerical relationships of map variables.

The bottom portion of the figure outlines restructuring of the basic spatial statistic operations to align with traditional non-spatial statistical concepts and operations (see author's note). The first three groupings are associated with general descriptive statistics, the middle two involve unique spatial statistics operations and the final two identify classification and predictive statistics.

GIS Perspective:

Surface Modeling *(Density Analysis, Spatial Interpolation, Map Generalization)*
Spatial Data Mining *(Descriptive, Predictive, Prescriptive)*

Statistical Perspective:

Basic Descriptive Statistics *(Min, Max, Median, Mean, StDev, etc.)*
Basic Classification *(Reclassify, Binary/Ranking/Rating Suitability)*
Map Comparison *(Joint Coincidence, Statistical Tests)*
Unique Map Descriptive Statistics *(Roving Window Summaries)*
Surface Modeling *(Density Analysis, Spatial Interpolation)*
Advanced Classification *(Map Similarity, Maximum Likelihood, Clustering)*
Predictive Statistics *(Map Correlation/Regression, Data Mining Engines)*

Figure 1. Alternative frameworks for quantitative map analysis.

Figure 2 depicts the non-spatial and spatial approaches for characterizing the distribution of mapped data and the direct link between the two representations. The left side of the figure illustrates non-spatial statistics analysis of an example set of data as fitting a standard normal curve in “data space” to assess the central tendency of the data as its average and standard deviation. In processing, the geographic coordinates are ignored and the typical value and its dispersion are assumed to be uniformly (or randomly) distributed in “geographic space.”

The top portion of figure 2 illustrates the derivation of a continuous map surface from geo-registered point data involving spatial autocorrelation. The discrete point map locates each sample point on the XY coordinate plane and extends these points to their relative values (higher values in the NE; lowest in the NW). A roving window is moved throughout the area that weight-averages the point data as an inverse function of distance—closer samples are more influential than distant samples. The effect is to fit a surface that represents the geographic distribution of the data in a manner that is analogous to fitting a SNV curve to characterize the data’s numeric distribution. Underlying this process is the nature of the sampled data which must be numerically quantitative (measurable as continuous numbers) and geographically isopleth (numbers form continuous gradients in space).

The lower-right portion of figure 2 shows the direct linkage between the numerical distribution and the geographic distribution views of the data. In geographic space, the “typical value” (average) forms a horizontal plane implying that the average is everywhere. In reality, the average is hardly anywhere and the geographic distribution denotes where values tend to be higher or lower than the average.

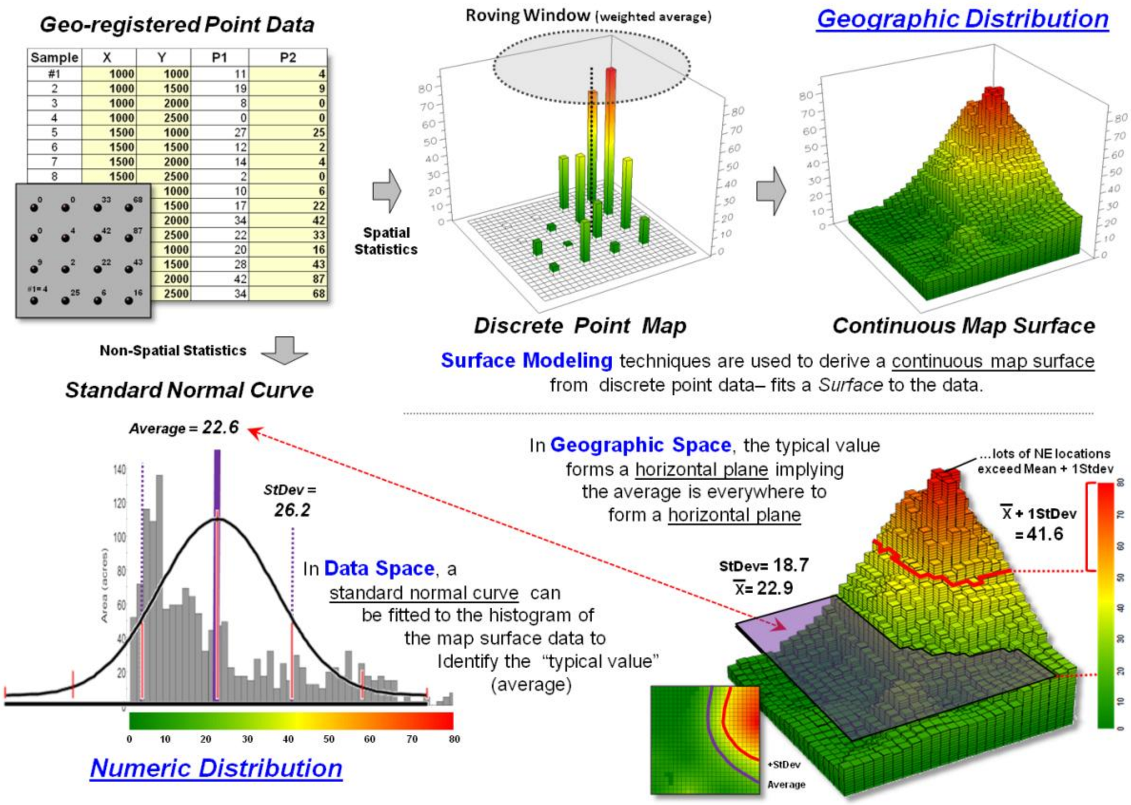


Figure 2. Comparison and linkage between spatial and non-spatial statistics.

In data space, a histogram represents the relative occurrence of each map value. By clicking anywhere on the map, the corresponding histogram interval is highlighted; conversely, clicking anywhere on the histogram highlights all of the corresponding map values within the interval. By selecting all locations with values greater than + 1SD, areas of unusually high values are located—a technique requiring the direct linkage of both numerical and geographic distributions.

Figure 3 outlines two of the advance spatial statistics operations involving spatial correlation among two or more map layers. The top portion of the figure uses *map clustering* to identify the location of inherent groupings of elevation and slope data by assigning pairs of values into groups (called clusters) so that the value pairs in the same cluster are more similar to each other than to those in other clusters.

The bottom portion of the figure assesses map correlation by calculating the degree of dependency among the same maps of elevation and slope. Spatially "aggregated" correlation involves solving the standard correlation equation for the entire set of paired values to represent the overall relationship as a single metric. Like the statistical average, this value is assumed to be uniformly (or randomly) distributed in "geographic space" forming a horizontal plane.

"Localized" correlation, on the other hand, maps the degree of dependency between the two map variables by successively solving the standard correlation equation within a roving window to generate a continuous map surface. The result is a map representing the geographic distribution of the spatial dependency throughout a project area indicating where the two map variables are highly correlated (both positively, red tones; and negatively, green tones) and where they have minimal correlation (yellow tones).

With the exception of unique Map Descriptive Statistics and Surface Modeling classes of operations, the grid-based map analysis/modeling system simply acts as a mechanism to spatially organize the data. The alignment of the geo-registered grid cells is used to partition and arrange the map values into a format amenable for executing commonly used statistical equations. The critical difference is that the answer often is in map form indicating where the statistical relationship is more or less than typical.

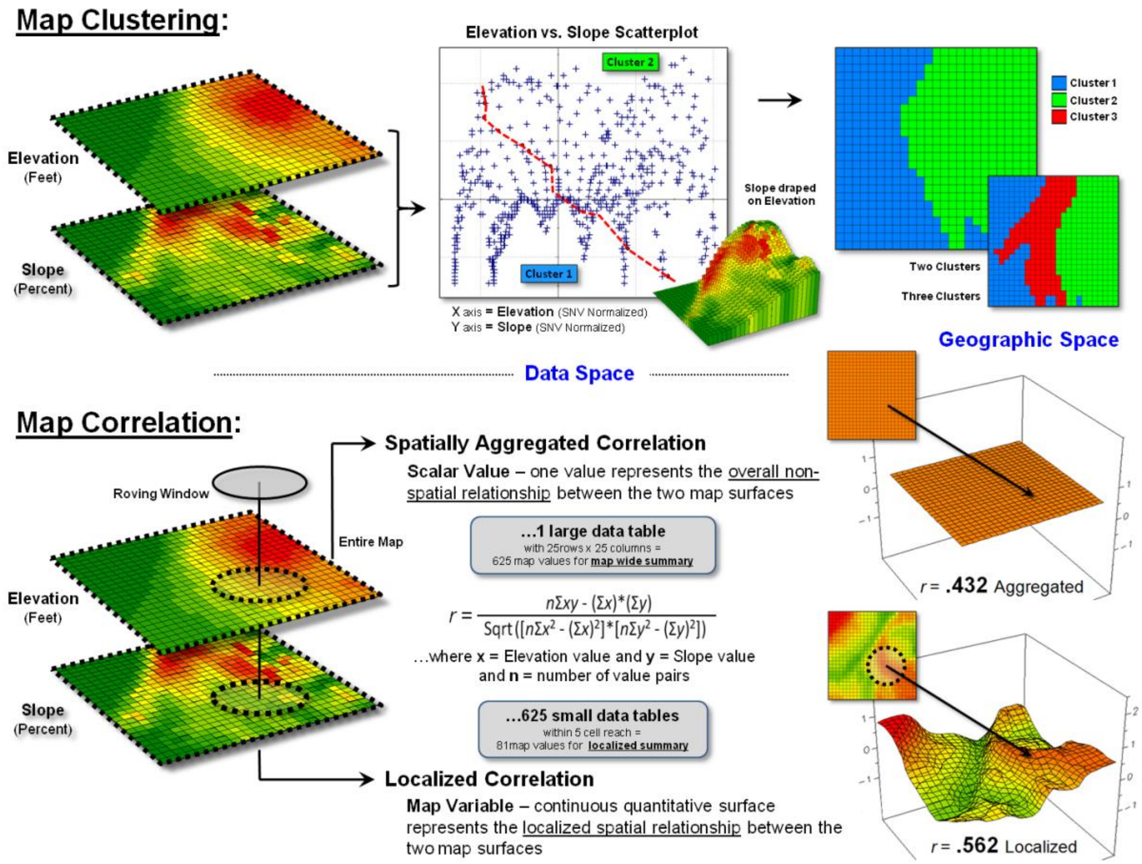


Figure 3. Conceptual extension of clustering and correlation to mapped data and analysis.

While the technological applications of GIS have soared over the last decade, the analytical applications seem to have flat-lined. The seduction of near instantaneous geo-queries and awesome graphics seem to be masking the underlying character of mapped data— that maps are numbers first, pictures later. However, grid-based map analysis and modeling involving Spatial Analysis and Spatial Statistics is, for the larger part, simply extensions of traditional mathematics and statistics. The recognition by the GIS community that quantitative analysis of maps is a reality and the recognition by the STEM community that spatial relationships exist and are quantifiable should be the glue that binds the two perspectives. That reminds me of a very wise observation about technology evolution—

"Once a new technology rolls over you, if you're not part of the steamroller, you're part of the road." *~Stewart Brand, editor of the Whole Earth Catalog*

____

Author's Notes: *for a more detailed discussion, see "SpatialSTEM: Extending Traditional Mathematics and Statistics to Grid-based Map Analysis and Modeling" posted at www.innovativegis.com/basis/Papers/Other/SpatialSTEM/.*

To Boldly Go Where No Map Has Gone Before

(GeoWorld, October 2012)

(return to top of Topic)

Previous sections have described a mathematical framework (dare I say a "map-*ematical*" framework?) for quantitative analysis of mapped data. Recall that *Spatial Analysis* operations investigate the "contextual" relationships within and among maps, such as variable-width buffers that account for intervening conditions. *Spatial Statistics* operations, on the other hand, examine the "numerical" relationships, such as map clustering to uncover inherent geographic patterns in the data.

The cornerstone of these capabilities lies in the grid-based nature of the data that treats geographic space as continuous map surfaces composed of thousands upon thousands of cells with each containing data values that identify the characteristics/conditions occurring at each location. This simple matrix structure provides a detailed account of the unique spatial distribution of each map variable and a geo-registered stack of map layers provides the foothold to quantitatively explore their spatial patterns and relationships.

The most fundamental and ubiquitous grid form is the Latitude/Longitude coordinate system that enables every location on the Earth to be specified by a pair of numbers. The upper portion of figure 1, depicts a 2.5^0 Lat/Lon grid forming a matrix of 73 rows by 144 columns= 10,512 cells in total with each cell having an area of about $18{,}735 mi^2$.

The lower portion of the figure shows that the data could be stored in Excel with each spreadsheet cell directly corresponding to a geographic grid cell. In turn, additional map layers could be stored as separate spreadsheet pages to form a map stack for analysis.

Of course this resolution is far too coarse for most map analysis applications, but it doesn't have to be. Using the standard single precision floating point storage of Lat/Long coordinates expressed in decimal degrees, the precision tightens to less than half a foot anywhere in the world (365214 ft/degree * 0.000001= .365214 ft *12 = 4.38257 inches or 0.11132 meters). However, current grid-based technology limits the practical resolution to about 1m (e.g., Ikonos satellite images) to 10m (e.g., Google Earth) due to the massive amounts of data storage required.

For example, to store a 10m grid for the state of Colorado it would take over two and half billion grid cells (26,960km²= 269,601,000,000m² / 100m² per cell= 2,696,010,000 cells). To store the entire earth surface it would take nearly a trillion and a half cells (148,300,000km^2 = 148,000,000,000,000m^2 / 100m² per cell= 1,483,000,000,000 cells).

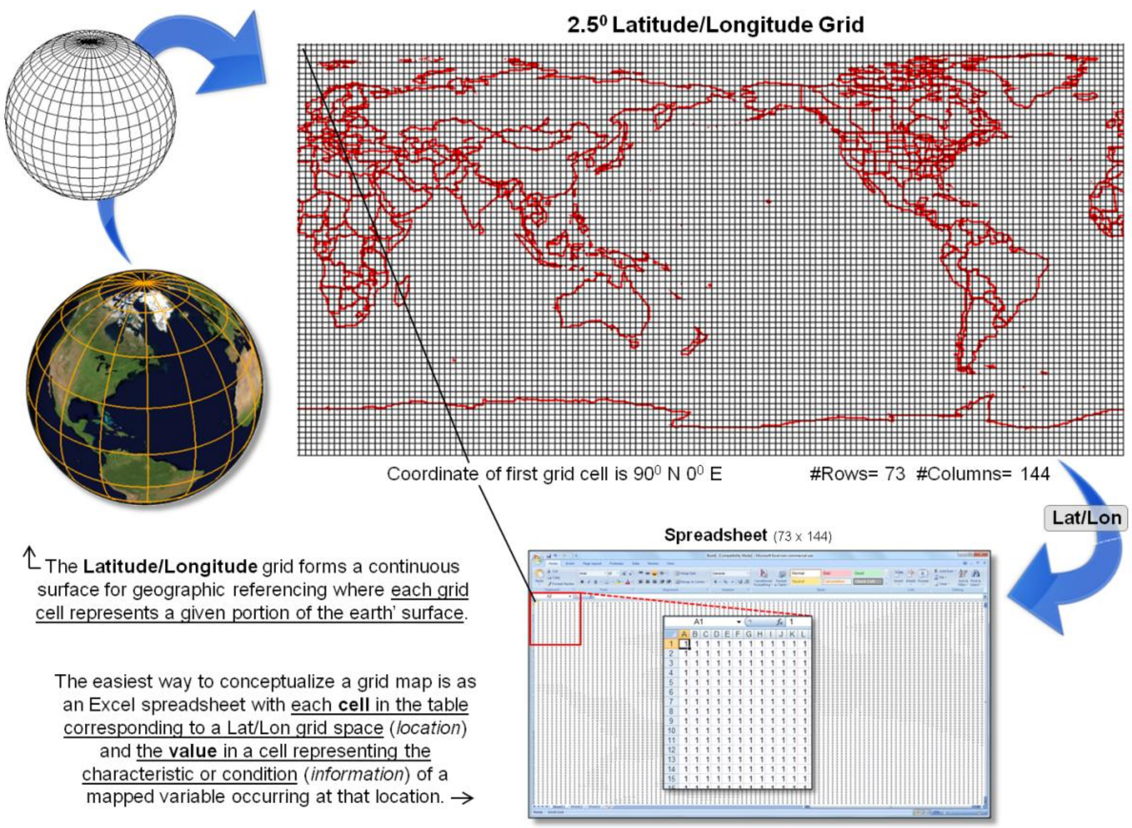


Figure 1. Latitude and Longitude coordinates provide a universal framework for parsing the earth's surface into a standardized set of grid cells.

At first these storage loads seem outrageous but with distributed cloud computing the massive grid can be "easily" broken into manageable mouthfuls. A user selects an area of interest and data for that area is downloaded and stitched together. For example, Google Earth responds to your screen interactions to nearly instantaneously download millions of pixels, allowing you to pan/zoom and turn on/off map layers that are just a drop in the bucket of the trillions upon trillions of pixels and grid data available in the cloud.

Figure 2 identifies another, more practical mechanism for storage using a relational database. In essence, each of the conceptual grid map spreadsheets can be converted to an interlaced format with a long string of numbers forming the columns (data fields); the rows (records) identify the information available each of the individual grid cells that form the reference grid.

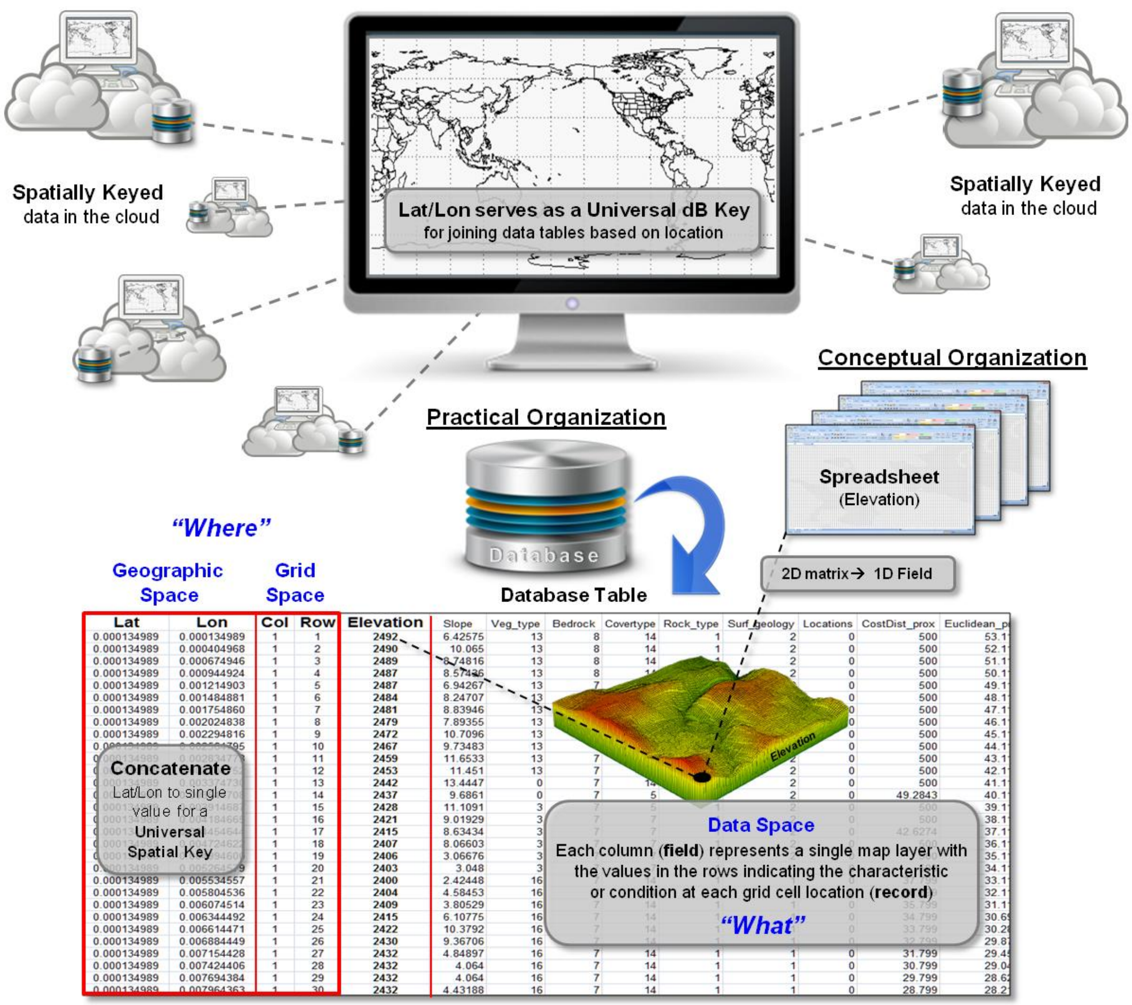


Figure 2. Within a relational database, Lat/Lon forms a Universal DBMS Key for joining tables.

For fairly small areas of up to a million or so cells this is an excellent way to store grid maps as their spatial coincidence is inherent in the organization and the robust standard set of database queries and processing operations is available. Larger grids use more advanced, specialized mechanisms of storage to facilitate data compression and virtual paging of fully configured grid layers.

But the move to a relational database structure is far more important than simply corralling mega-gulps of map values. It provides a "Universal DBMS Key" that can link seemingly otherwise disparate database tables (see Authors Note). The process is similar to a date/time stamp, except the "where information" provides a spatial context for joining data sets. Demographic records can be linked to resource records that in turn can be linked to business records, health records, etc— all sharing a common Lat/Lon address.

All that is necessary is to tag your data with its Lat/Lon coordinates ("where" it was collected) just as you do with the date/time ("when" it was collected) …not a problem with the ubiquitous availability and increasing precision of GPS that puts a real-time tool for handling detailed spatial data right in your pocket. In today's technology, most GPS-enabled smart phones are accurate to a few meters and specialized data collection devices precise to a few centimeters.

Once your data is stamped with its "spatial key," it can be linked to any other database table with spatially tagged records without the explicit storage of a fully expanded grid layer. All of the spatial relationships are implicit in the relative positioning of the Lat/Lon coordinates.

For example, a selection operation might be to identify of all health records jointly occurring within half a kilometer of locations that have high lead concentrations in the top soil. Or, locate all of the customer records within five miles of my store; better yet, within a ten-minute drive from a store.

Geotechnology is truly a mega-technology that will forever change how we perceive and process spatial information. Gone are the days of manual measurements and specialized data formats that have driven our mapping legacy. Lat/Lon coordinates move from cross-hairs for precise navigation (intersecting lines) to a continuous matrix of spaces covering the globe for consistent data storage (grid cells). The recognition of a universal spatial key coupled with spatial analysis/statistics procedures and GPS/RS technologies provides a firm foothold "to boldly go where no map has gone before."

Author's Note: *See "The Universal Key for Unlocking GIS's Full Potential," book IV, Topic 7, section 6 in the Beyond Mapping Compilation Series posted at www.innovativegis.com.*

Depending on Where is What

(GeoWorld, March 2013)

(return to top of Topic)

Early procedures in spatial statistics were largely focused on the characterization of spatial patterns formed by the relative positioning of discrete spatial objects—points, lines, and polygons. The "area, density, edge, shape, core-area, neighbors, diversity and arrangement" of map features are summarized by numerous landscape analysis indices, such as *Simpson's Diversity* and *Shannon's Evenness* diversity metrics; *Contagion* and *Interspersion/Juxtaposition* arrangement metrics; and *Convexity* and *Edge Contrast* shape metrics (see Author's Note 1). Most of these techniques are direct extensions of manual procedures using paper maps and subsequently coded for digital maps.

Grid-based map analysis, however, expands this classical view by the direct application of advanced statistical techniques in analyzing spatial relationships that consider continuous geographic space. Some of the earliest applications (circa 1960) were in climatology and used map surfaces to generate isotherms of temperature and isobars of barometric pressure.

In the 1970s, the analysis of remotely sensed data (raster images) began employing traditional statistical techniques, such as *Maximum Likelihood Classification*, *Principle Component Analysis* and *Clustering* that had been used in analyzing non-spatial data for decades. By the 1990s, these classification-oriented procedures operating on spectral bands were extended to include the full wealth of statistical operations, such as *Correlation* and *Regression*, utilizing diverse sets of geo-registered map variables (grid-based map layers).

It is the historical distinction between "*Spatial Pattern characterization* of discrete objects" and "*Spatial Relationship analysis* of continuous map surfaces" that identifies the primary conceptual branches in spatial statistics. The spatial relationship analysis branch can be further grouped by two types of spatial dependency driving the relationships— *Spatial Autocorrelation* involving spatial relationships within a single map layer, and *Spatial Correlation* involving spatial relationships among multiple map layers (see figure 1).

Spatial Autocorrelation follows Tobler's first law of geography— that "…near things are more alike than distant things." This condition provides the foundation for *Surface Modeling* used to identify the continuous spatial distribution implied in a set of discrete point data based on one of four fundamental approaches (see figure 2 and Author's Note 2). The first two approaches—*Map Generalization* and *Geometric Facets*—consider the entire set of point values in determining the "best fit" of a polynomial equation, or a set of 3-dimentional geographic shapes.

For example, a 1st order polynomial (tilted plane) fitted to a set of data points indicates its spatial trend with decreasing values aligning with the direction cosines of the plane. Or, a complex set of abutting tilted triangular planes can be fitted to the data values to capture significant changes in surface form (triangular tessellation).

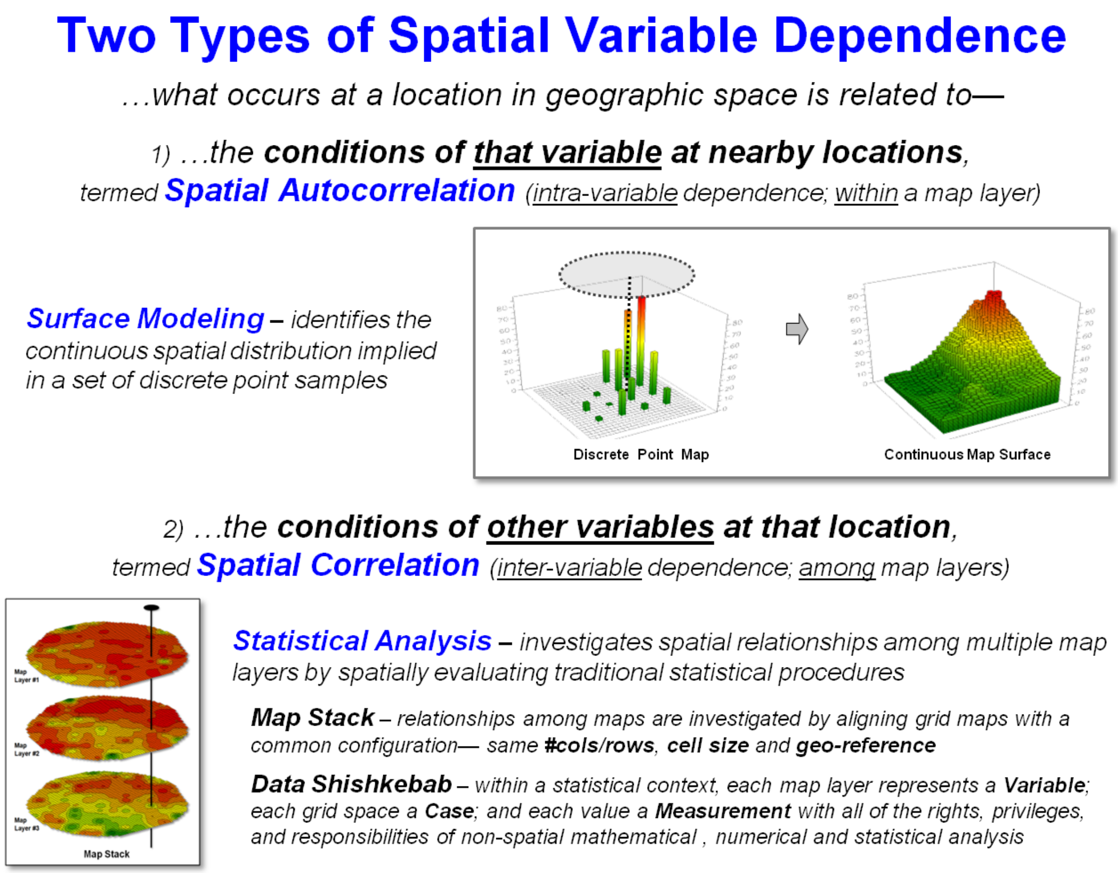


Figure 1. Spatial Dependency involves relationships within a single map layer (Spatial Autocorrelation) or among multiple map layers (Spatial Correlation).

The lower two approaches—*Density Analysis* and *Spatial Interpolation*—are based on localized summaries of the point data utilizing "roving windows." Density Analysis counts the number of data points in the window (e.g., number of crimes incidents within half a kilometer) or computes the sum of the values (e.g., total loan value within half a kilometer).

However, the most frequently used surface modeling approach is Spatial Interpolation that "weight-averages" data values within a roving window based on some function of distance. For example, Inverse Distance Weighting (IDW) interpolation uses the geometric equation $1/D^{\text{Power}}$ to greatly diminish the influence of distant data values in computing the weighted-average.

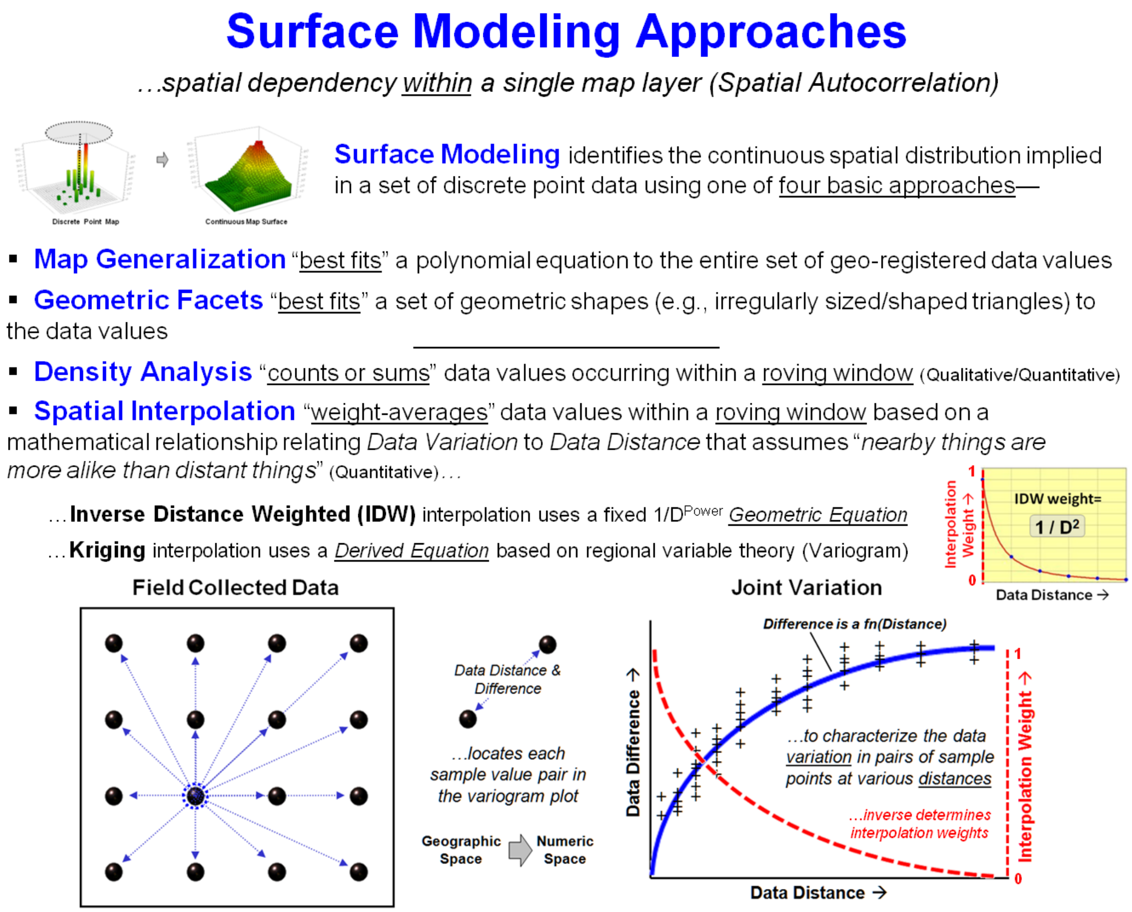


Figure 2. Surface Modeling involves generating map surfaces that portray the continuous spatial distribution implied in a set of discrete point data.

The bottom portion of figure 2 encapsulates the basis for Kriging which derives the weighting equation from the point data values themselves, instead of assuming a fixed geometric equation. A variogram plot of the joint variation among the data values (blue curve) shows the differences in the values as a function of distance. The inverse of this derived equation (red curve) is used to calculate the distance affected weights used in weight-averaging the data values.

The other type of spatial dependency—**Spatial Correlation**—provides the foundation for analyzing spatial relationships among map layers. It involves spatially evaluating traditional statistical procedures using one of four ways to access the geo-registered data— *Local*, *Focal*, *Zonal* and *Global* (see figure3 and Author's Notes 3 and 4). Once the spatially coincident data is collected and compatibly formatted, it can be directly passed to standard multivariate statistics packages or to more advanced statistical engines (CART, Induction or Neural Net). Also, a growing number of GIS systems have incorporated many of the most frequently used statistical operations.

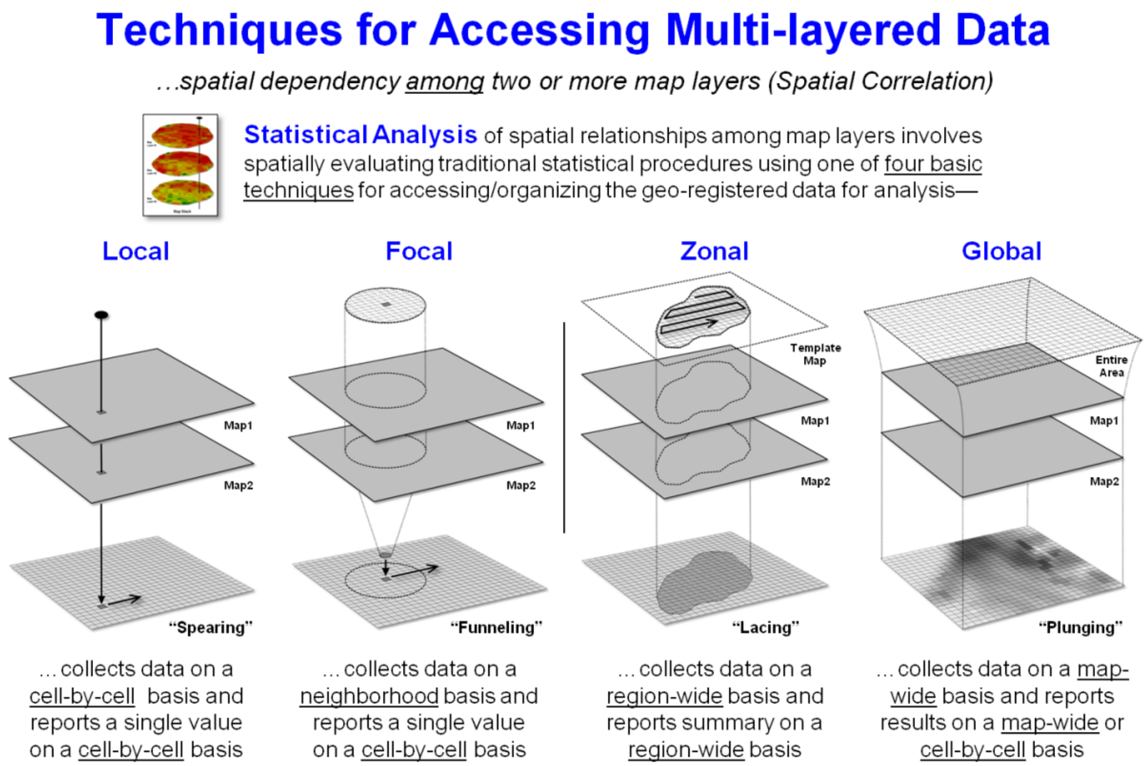


Figure 3. Statistical Analysis of mapped data involves repackaging mapped data for processing by standard multivariate statistics or more advanced statistical operations.

The majority of the *Statistical Analysis* operations simply "repackage" the map values for processing by traditional statistics procedures. For example, "Local" processing of map layers is analogous to what you see when two maps are overlaid on a light-table. As your eye moves around, you note the spatial coincidence at each spot. In grid-based map analysis, the cell-by-cell collection of data for two or more grid layers accomplishes the same thing by "spearing" the map values at a location, creating a summary (e.g., simple or weighted-average), storing the new value and repeating the process for the next location.

"Focal" processing, on the other hand, "funnels" the map layer data surrounding a location (roving window), creates a summary (e.g., correlation coefficient), stores the new value and then repeats the process. Note that both local and focal procedures store the results on a cell-by-cell basis.

The other two techniques (right side of figure 3) generate entirely different summary results. "Zonal" processing uses a predefined template (termed a map region) to "lace" together the map values for a region-wide summary. For example, a wildlife habitat unit might serve as a template map to retrieve slope values from a data map of terrain steepness, compute the average of the values, and then store the result for all of the locations defining the region. Or maps of animal activity for two time periods could be accessed and a paired t-test performed to determine if a significant difference exists within the habitat unit. The interpretation of the resultant map value assigned to all of the template locations is that each cell is an "element of a spatial entity having that overall summary statistic."

"Global" processing isn't much different from the other techniques in terms of mechanics, but is radically different in terms of the numerical rigor implied. In map-wide statistical analysis, the entire map is considered a variable, each cell a case and each value a measurement (or instance) in mathematical/statistical modeling terminology. Within this context, the processing has "all of the rights, privileges and responsibilities" afforded non-spatial quantitative analysis. For example, a regression could be spatially evaluated by "plunging" the equation through a set of independent map variables to generate a dependent variable map on cell-by-cell basis, or reported as an overall map-wide value.

So what's the take-home from all this discussion? It is that maps are "numbers first, pictures later" and we can spatially discover and subsequently evaluate the spatial relationships inherent in sets of grid-based mapped data as true map-*ematical* expressions. All that is needed is a new perspective of what a map is (and isn't).

Author's Notes: *1) in the Beyond Mapping Compilation Series posted at www.innovativegis.com see book III , Topic 6, sections 9 through 12 on Analyzing Landscape Patterns; 2) see book III, Topic 9 on Basic Techniques in Spatial Statistics; 3) refers to C. Dana Tomlin's four data acquisition classes; 4) for more discussion on data acquisition techniques, see book IV, Topic 5, Section 3 "Getting the Numbers Right."*

Laying the Foundation for SpatialSTEM:
Spatial Mathematics, Map Algebra and Map Analysis
(GeoWorld, October 2013)

(return to top of Topic)

Mathematics in general and geometry and trigonometry in particular have long been the keystone to mapping—from Spatial Mathematics that enables the development of mapped data; to a generalized Map Algebra for expressing math/stat relationships among map variables; to a comprehensive Map Analysis toolbox that extends traditional quantitative data analysis procedures by considering the spatial distribution and interaction of mapped data layers.

Several years ago, Nigel Waters wrote a short synopsis on "The Most Beautiful Formulae in GIS" where he described the ten most useful Spatial Formulae and the ten most useful Attribute-related Formulae chosen for their elegance, simplicity, and generality, as well as their wide applicability and power (see author's note 1). More recently, the book "Spatial Mathematics:

Theory and Practice through Mapping" by Arlinghaus and Kerski further develops the wealth of enabling *Spatial Mathematics* equations and techniques (see author's note 2).

These and a host of similar treatises provide a comfortable conceptual springboard for STEM disciplines to extend traditional scalar mathematics into the spatial realm. The digital map expressed as an organized set of numbers fuels this transition— today "maps are numbers first, pictures later." The result is a generalized *Map Algebra* (see author's note 3) enabling a user to add, subtract, divide, raise to a power, root, log and even differentiate and integrate digital maps— all of the functionality of a pocket calculator (and then some) operating on geo-registered stacks of digital maps.

This algebraic framework provides a comprehensive toolbox of primitive mathematical operations transitioning traditional quantitative data analysis into *Map Analysis* that infuses the consideration of spatial patterns and relationships into the analysis. From this perspective, the spatial distribution of data is as important as its numerical distribution in analyzing map variables.

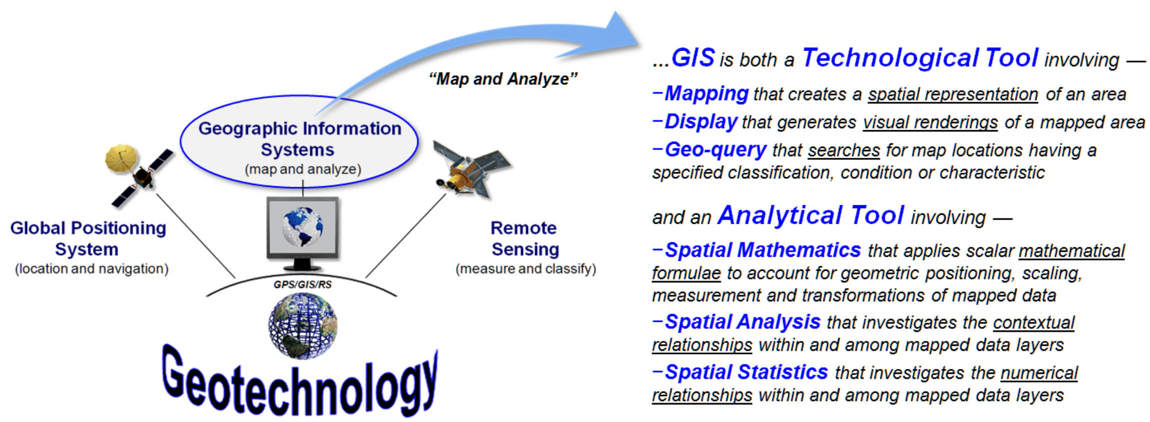


Figure 1. GIS can be viewed as both a "Technological Tool" and an "Analytical Tool."

Figure 1 provides a 40,000-foot overview of the evolving field of Geotechnology, one of the three mega-technologies for the 21st century as identified by the U.S. Department of Labor (the other two are Biotechnology and Nanotechnology). The left side of the figure depicts the "spatial triad" of technologies (GPS, GIS and RS) comprising Geotechnology that collects, stores, retrieves, processes, and displays digital mapped data. The mapping and analysis capabilities of GIS can be characterized as both a "Technological Tool" involving mapping, display and geo-query and an "Analytical Tool" involving spatial mathematics, analysis and statistics.

As a technological tool, GIS greatly extends traditional mapping and inventory techniques involving laborious, inefficient and generally ineffective manual procedures employed just a few decades ago. Today it is commonplace to get real-time routing directions, superimposed on an interactive map with a satellite image backdrop and a street view of your destination; all from a smartphone that rivals the computing power of a mainframe computer a few decades ago. For the most part, static paper maps have given way to dynamic digital mapped data that can be interactively viewed and processed in radically new ways—a revolution that is simply amazing for anyone over thirty, yet commonplace for those who are younger.

The meteoric rise in the technical expressions of Geotechnology is in large part due to its easily envisioned extension of its manual mapping and inventory legacies. Database systems replaced the walls of file cabinets (attribute data) and digital maps replaced paper maps (spatial data). Linking the two data set perspectives spawned a radically new paradigm of what a map is and isn't and catapulted mapping to "mega-technology" status.

Is a similar canonic step and radically changed paradigm in the future for traditional quantitative data analysis concepts, procedures and applications? What are the impediments holding back GIS as an analytical tool? What are the inducements needed for advancing spatially-aware quantitative data analysis?

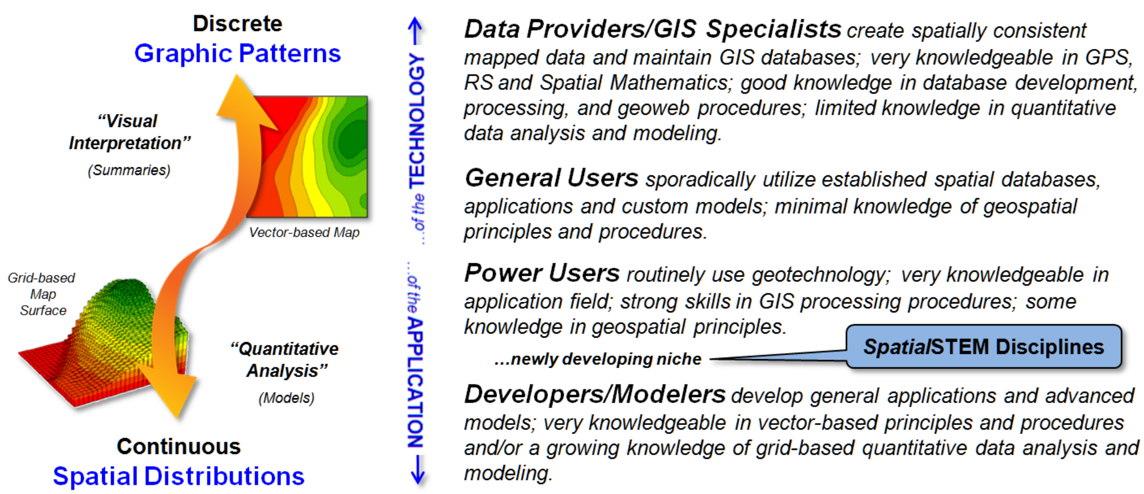


Figure 2. Types of GIS data, users and applications.

Figure 2 outlines the data, users and application approaches that is fueling this transformation. A major hurdle is the historical perspective of maps as being comprised of discrete spatial objects (point, line and areal patterns) as depicted in the 2D vector-based map in the upper-left portion of the figure. While this *vector data* format is comfortable and ideal for human visual interpretation, it lacks the spatial specificity and consistency required by advanced analysis procedures needed by most the STEM research and applications.

Alternatively, *raster data* depicted in the lower-left portion of the figure provides a continuous and consistent data form that is preconditioned for quantitative data analysis. A grid-based map surface tracks subtle spatial variations of a map variable as an uninterrupted gradient instead of aggregating the detailed data into discrete ranges (i.e., contour intervals).

In addition, the matrix structuring provides a consistent "analysis frame" for a geo-registered stack of map layers for a project area. Within this grid structure the row, column locators implicitly carry all of the necessary spatial topology relating each grid location to the positioning of all other locations within a single map layer and among multiple layers in a geo-registered map stack.

The right side of figure 2 identifies several types of GIS users. Currently, most of the GIS community is comprised of Data Providers, GIS Specialists, and General Users who are primarily involved with the technical aspects of GIS and their vector processing expressions—creating, maintaining and accessing mapped data and then executing standardized processing routines. These users can be thought of as "of the technology."

The Power Users, Developers and Modelers, on the other hand, are more "of the application." Within this context, domain expertise identifies the scope of a problem and the map variables involved and then map analysis capabilities are used to uncover spatial relationships that then forms a spatially-aware solution. It is in this arena that a "newly developing niche for SpatialSTEM" is poised to take-hold (see author's note 4).

Einstein noted that "we cannot solve our problems with the same level of thinking that created them" and that "the formulation of the problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill." This thinking suggests that the STEM disciplines need to be actively engaged and leading the search for spatially-aware solutions to today's complex spatial problems. Also, it recognizes that geospatial technologists need to fully recognize the quantitative nature mapped data and embrace its analytical potential, as well as its technical application.

However when it comes to Map Analysis (grid-based Spatial Analysis and Spatial Statistics operations), the old adage that "they who know not, know not they know not" takes center stage and the status quo paradigms of science and technology continue to dominate education, research and application development. As long as a conceptual chasm exists between the mapping and quantitative analysis communities, spatially-aware solutions to complex problems will continue to be mostly side-lined.

Author's Notes: *1) See "The Most Beautiful Formulae in GIS" by Nigel Waters (1995) posted at www.innovativegis.com/basis/MapAnalysis/Topic30/Beautiful_Formulae.pdf. 2) See "Spatial Mathematics: Theory and Practice through Mapping" by Sandra Arlinghaus and Joseph Kerski (2013, www.crcpress.com/product/isbn/9781466505322). 3) The concepts and procedures behind Spatial Mathematics was introduced by David Unwin with the University of London (Introductory Spatial Analysis, 1981, Methuen New York) and subsequently developed as a set-based Map Algebra for manipulating raster map layers by Dana Tomlin* *as a doctoral student at Yale University (Geographic Information Systems and Cartographic Modeling, 1990, Prentice-Hall, Englewood, New Jersey). 4) See "SpatialSTEM – Seminar, Workshop and Teaching Materials for Understanding Grid-based Map Analysis" posted at www.innovativegis.com/Basis/Courses/SpatialSTEM/.*

Further Online Reading*: (Chronological listing posted at* ***www.innovativegis.com/basis/BeyondMappingSeries/****)*

Map-ematically Messing with Mapped Data — *discusses the nature of grid-based mapped data and Spatial Analysis operations (February 2012)*

Paint by Numbers Outside the Traditional Statistics Box — *discusses the nature of Spatial Statistics operations (March 2012)*

The Spatial Key to Seeing the Big Picture — *describes a five step process for generating grid map layers from spatially tagged data (September 2013)*

Recasting Map Analysis Operations for General Consumption — *reorganizes ArcGIS's Spatial Analyst tools into the SpatialSTEM framework that extends traditional math/stat procedures (February 2013)*

(return to top of Topic)

(Back to the Table of Contents)