*Spatial*STEM: A New Perspective and Conceptual Framework for Grid-based Map Analysis and Modeling

Joseph K. Berry

Adjunct Faculty in Geosciences, University of Denver
Adjunct Faculty in Natural Resources, Colorado State University
Principal, Berry and Associates // Spatial Information Systems, Fort Collins, CO
Email: jberry@innovativegis.com — Website: http://www.innovativegis.com/basis/

Note: this white paper is posted at http://www.innovativegis.com/basis/Papers/Other/sSTEM_overview/ (.html). Click here for a printer-friendly version (.pdf)

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.

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." 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.

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. Figure 1 outlines the important components of map analysis and modeling within a mathematical/statistical structure. Of the three disciplines forming Geotechnology (the "spatial triad" of 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*.

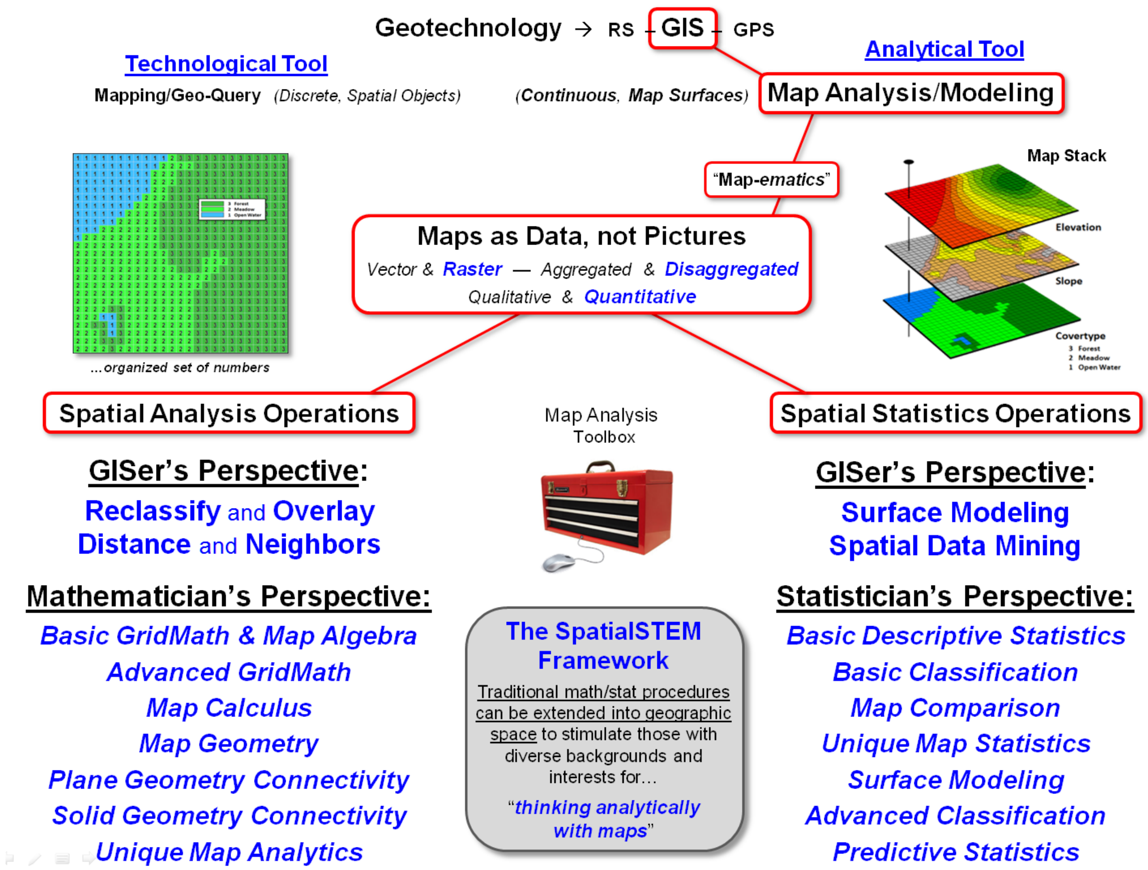

Figure 1. Conceptual overview of the SpatialSTEM framework.

The bottom-left portion of figure 1 restructures the basic *Spatial Analysis Operations* in grid-based GIS systems to align with traditional mathematical concepts and procedures. The bottom-right listing in the figure does a similar translation to identify the *Spatial Statistics Operations* in the map analysis toolbox within more commonly understood non-spatial statistical concepts and procedures.

Spatial analysis can be thought of as an extension of traditional mathematics involving "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.

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 2 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.

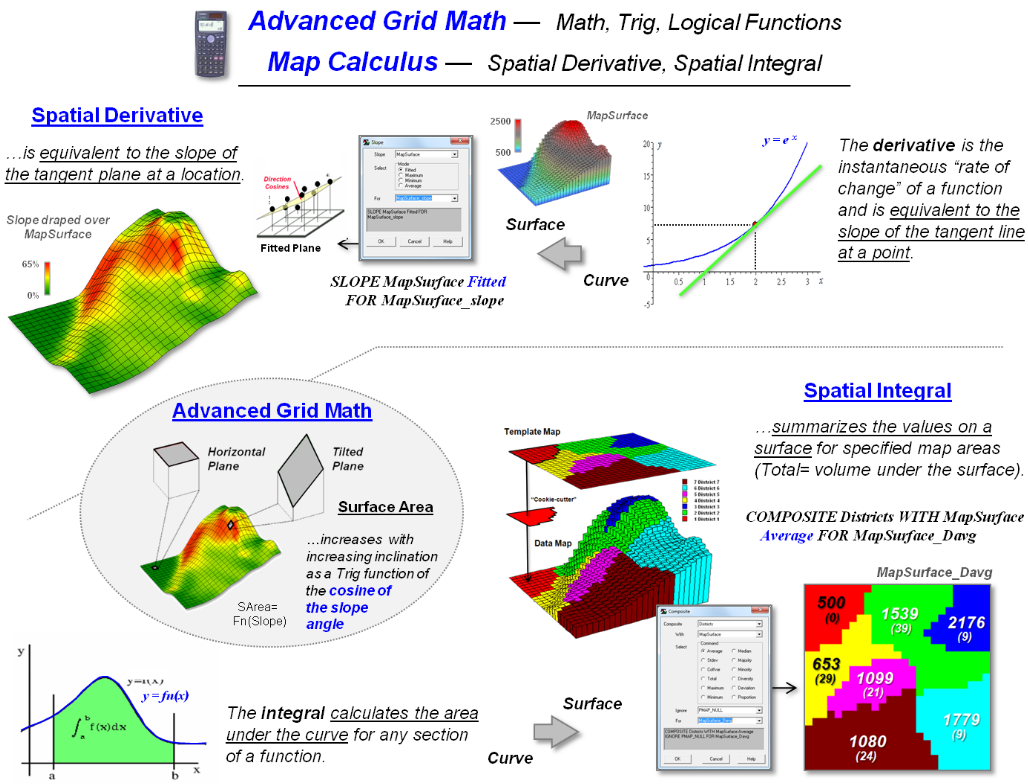

Figure 2. Conceptual extension of the derivative, trigonometric functions and integral to mapped data and map analysis operations.

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.

Spatial Statistics, in contrast to *Spatial Analysis* (a mathematician's perspective), can be thought of as an extension of traditional statistics involving "numerical" relationships within and among mapped data layers (a statistician's perspective). 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.

Figure 3 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."

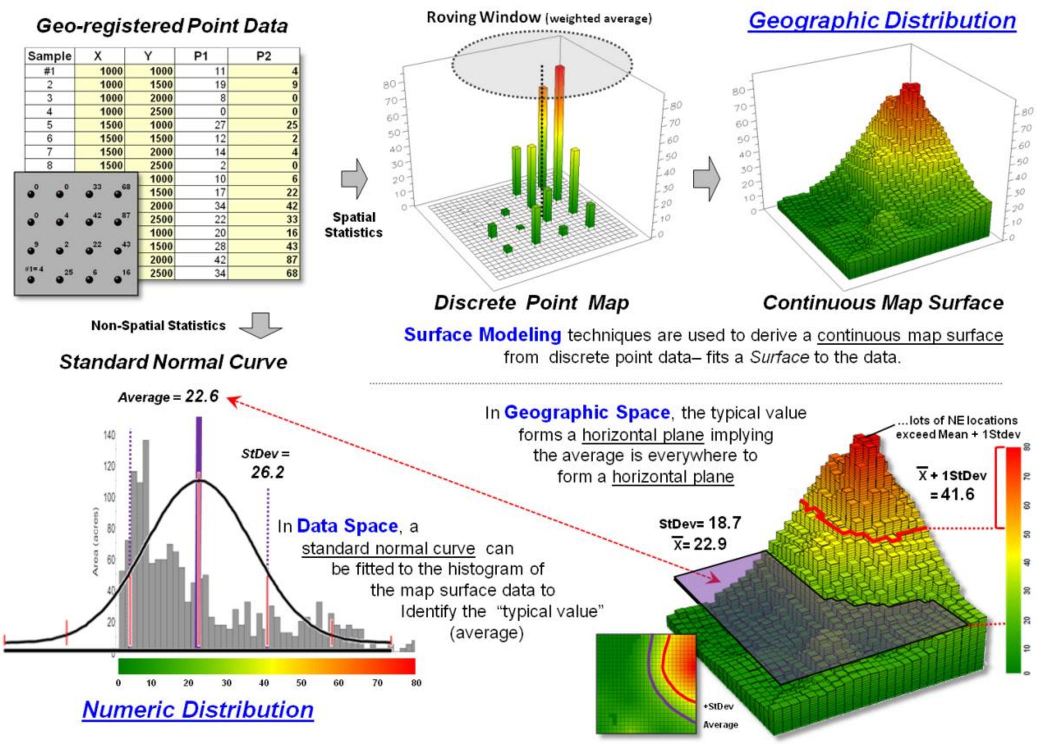

Figure 3. Comparison and linkage between spatial and non-spatial statistics

The top portion of figure 3 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 3 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. 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.

There are numerous direct extensions of data mining procedures such as *map clustering* to identify the location of inherent groupings of two map variables 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.

Another is direct extension is *map correlation* that calculates the degree of dependency among two map variables but instead of reporting a single "aggregated" correlation metric it generates a map of the "localized" correlation by successively solving the standard correlation equation within a roving window indicating where and how the two map variables are highly correlated and where they have minimal correlation. A similar extension is the derivation of a *Paired T-test map* that tells you precisely where two maps are significantly different and where they are not.

The grid-based data structure serves as the keystone supporting the SpatialSTEM approach. Mapping and geo-query utilize vector-based data in which <u>discrete spatial objects</u> (*points*, *lines* and *polygons*) form a collection of independent, irregular features to characterize geographic space. Quantitative analysis of mapped data, on the other hand, operates on grid-based <u>continuous map variables</u> (*surfaces*) composed of thousands upon thousands of map values stored in geo-registered matrices.

Figure 4 depicts a comparison between vector and raster data structures in representing 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 terrain configuration. 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. 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.

The remainder of figure 4 depicts the grid-based data 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.

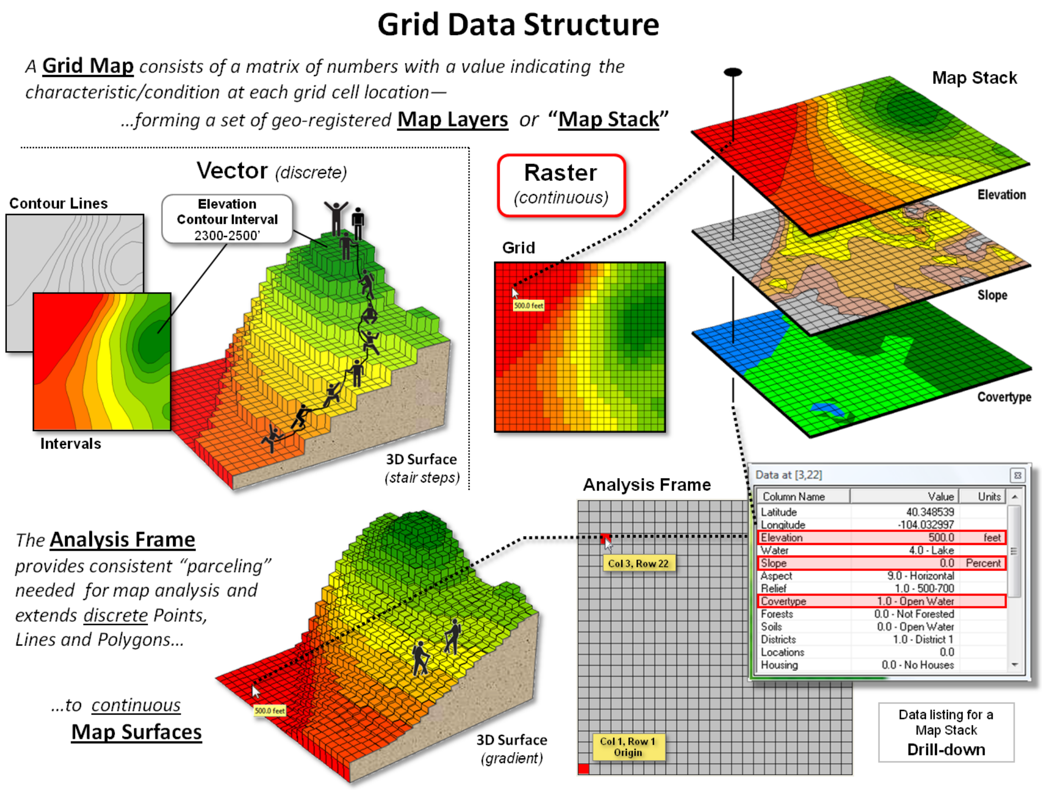

Figure 4. Organizational considerations and terminology for grid-based mapped data.

The underlying concept of this data structure is that grid cells for all of the map layers precisely coincide, 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 a 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. This simple matrix representation can be easily conceptualized by non-GIS users as a special type of spreadsheet. Extending this thinking to Latitude and Longitude coordinates as forming 4-inch grid cells throughout the globe sets the stage for a Universal Spatial Key that can be used to joint disparate data sets in a manner analogous for analysis.

The computer does the heavy-lifting of the data formatting and computation. What is needed is a new generation of creative minds that goes beyond mapping to "thinking with maps" within a quantitative framework— a *Spatial*STEM environment. By recasting GIS concepts and operations of map analysis within the general scientific language of math/stat we can more easily educate and communicate with tomorrow's movers and shakers to think of maps as "mapped data" and express the wealth of quantitative analysis thinking they already understand using grid-based map variables.

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.*

A twelve-part series of columns on SpatialSTEM is compiled into Topic 30, *"SpatialSTEM: A Math/Stat Framework for Grid-based Map Analysis and Modeling"* in the online book Beyond Mapping III posted at http://www.innovativegis.com/basis/MapAnalysis/Topic30/Topic30.htm		
Section 1	*January 2012*	*SpatialSTEM Has Deep Mathematical Roots — provides a conceptual framework for a map-ematical treatment of mapped data*
2	*February 2012*	*Map-ematically Messing with Mapped Data — discusses the nature of grid-based mapped data and Spatial Analysis operations*
3	*March 2012*	*Paint by Numbers Outside the Traditional Statistics Box — discusses the nature of Spatial Statistics operations*
4	*April 2012*	*Simultaneously Trivializing and Complicating GIS — describes a mathematical structure for spatial analysis operations*
5	*May 2012*	*Infusing Spatial Character into Statistics — describes a statistical structure for spatial statistics operations*
6	*March 2013*	*Depending on Where is What — develops an organizational structure for spatial statistics*
7	*April 2013*	*Spatially Evaluating the T-test — illustrates the expansion of traditional math/stat procedures to operate on map variables to spatially solve traditional non-spatial equations*
8	*September 2012*	Organizing Geographic Space for Effective Analysis *— an overview of data organization for grid-based map analysis*
9	*October 2012*	To Boldly Go Where No Map Has Gone Before *— identifies Lat/Lon as a Universal Spatial Key for joining database tables*
10	*September 2013*	*The Spatial Key to Seeing the Big Picture — describes a five step process for generating grid map layers from spatially tagged data*
11	*October 2013*	*Laying the Foundation for SpatialSTEM: Spatial Mathematics, Map Algebra and Map Analysis — discusses the conceptual foundation and intellectual shifts needed for SpatialSTEM*
12	*February 2013*	*Recasting Map Analysis Operations for General Consumption — reorganizes ArcGIS's Spatial Analyst tools into the SpatialSTEM framework that extends traditional math/stat procedures*