7-1-2007

Spreadsheet-based GIS Models

Roger Blake
University of Massachusetts Boston, roger.blake@umb.edu

Jeffrey Keisler
University of Massachusetts Boston, jeff.keisler@umb.edu

Janet Wagner

Spreadsheet-based GIS Models

Jeffrey M. Keisler
Roger Blake
Janet Wagner

UMBCMWP 1029

July 2007

COLLEGE OF Management
UMASS BOSTON

Spreadsheet-based GIS models

Jeffrey M. Keisler, Roger Blake and Janet M. Wagner

Abstract: *Geographic Information Systems (GIS), which are computer-based systems that allow decision makers to incorporate geographically based data into their analyses, are widespread and powerful tools in many business and scientific settings today. In this paper, we present a tutorial discussing ways in which GIS functionality can be implemented within the spreadsheet environment. We show the straightforward and natural analogy between several GIS functions with spreadsheet functions, particularly for raster based data. We present two realistic examples meshing OR/MS methods with GIS methods --- an integration which is greatly enhanced by the "remarkable development platform" provided by spreadsheets. We discuss the many benefits of the spreadsheet enabled seamless integration of geographical data, analysis, and display.*

1. INTRODUCTION

Geographic Information Systems (GIS), which are computer-based systems that allow decision makers to incorporate geographically based data into their analyses, are widespread and powerful tools in many business and scientific settings today. Geographic information systems --- broadly defined as systems that can store, retrieve, map, and analyze geographic data --- have grown dramatically in the past decade, helped in large part by the advent of affordable applications for the desktop. The field has also benefited from the increased availability of free and low cost data distributed easily on the internet. GIS have spread from their traditional domains of military applications, utility management, environmental and resource management to fields such as marketing (Sohovich[1], 2002), insurance and real estate assessment (see Longley & Clarke[2], 1995), PDA applications for fieldwork, and even human rights work (O' Sullivan[3]). Many, if not most, U.S. and Canadian government agencies as well as states in the U.S. now have GIS departments and publicly available GIS data on the web. Organizations are making use of exciting new interactive web-based packages that allow for easy deployment of maps and spatial data. However, there is still a need to expand the use of GIS within organizations, particularly corporations, and to allow for more interaction between GIS experts and other departments.

The idea that will be explored in this paper is that GIS analyses, particularly those based on raster data, can --- in fact --- be done in spreadsheets. This tutorial will then investigate what GIS applications can be done within spreadsheets, why one might use the spreadsheet platform for GIS functionality, and how to do it. The applications in this paper are developed with Excel 2003, although the ideas should be easily implemented in any spreadsheet program. We note that Microsoft has a mapping program, MapPoint, which can be linked to Excel. However, we are suggesting a different kind of interaction, where raster data is analyzed and displayed in the spreadsheet itself, providing seamless integration of geographical data, analysis, and display.

Exactly as the call for papers for this issue says "spreadsheets, because of their ubiquity, power for rapid development, and transparency to non-technical managers who might be the ultimate users, are remarkable development platforms for MS/OR applications". Having GIS in the spreadsheet modeling toolkit will:

- increase the functionality of spreadsheet based systems,
- extend the reach of possible applications for spreadsheet prototyping,
- allow creative merging of OR/MS and GIS methods and technologies, and
- enhance the ability of non-technical spreadsheet end-users to understand and accept analytical work and results involving geographical data.

Seamlessly integrating GIS into spreadsheets is motivated by several strengths of the most common spreadsheet, Microsoft Excel. Excel has a very large user base --- in the hundreds of millions compared to around one million users for a widely used GIS product suite (Caravallo[4], 2002). The Excel user base extends deeply within and broadly across organizations, and there are many "power users" able to conduct sophisticated analyses or develop sophisticated applications. Numerical models for DSS are easily developed within Excel. Furthermore, because of its large commercial use, Microsoft has invested in developing a large number of features supporting quantitative analysis and information project management for use with Excel.

Of course, this idea is only attractive if it can be implemented. In this paper we will describe a number of challenges we foresee and discuss possible solutions. We also identify limits to the functionality of Excel, in order to better define the appropriate relationship between the spreadsheet and the GIS.

2. TIES TO PREVIOUS WORK ("WHAT")

GIS initially developed as an interdisciplinary field combining elements from the field of computer science with geography and mapmaking. The use of computers for mapping applications was initially developed during the 1960's for a survey of land use and planning in Canada in an effort headed by Roger Tomlinson (sometimes called the "Father of GIS") a geographer in an aerial survey company who had dabbled with the use of computers for mapping (see GeoWorld, 2004[5] for an interesting interview with Roger Tomlinson). GIS has now become its own specialty, with numerous stand-alone GIS departments and programs. Resources providing an overview of the field of GIS and its capabilities include a nice layperson introduction to GIS by the U.S. Geological Service[6] and numerous GIS textbooks, for example Lo and Yeung[7] (2006) and Longley[8], et. al. (2005).

GIS data itself has become an area of interest on its own. GIS data is now stored in a number of standard data formats or protocols, including those for the systems from ESRI (shapefiles), and GRASS (run-length encoding for rasters). GIS systems increasingly read, store and create standardized metadata and make these data files available and accessible over the web to the general public. Many GIS systems use a database such as Oracle to store enterprise-wide spatial attribute data. A DBMS, such as ESRI's SDE, allows for full integration of data and map elements.

There are also existing books exploring areas that merge business and public sector applications, including OR/MS techniques, with GIS. Longley and Clarke[9] (1995) provide an overview of GIS concepts and technology for business practitioners and academics and a discussion of a number of GIS applications in business, including a number of case studies. Malczewski[10] (1999) explores GIS applications for Multi-attribute Decision Analysis, which includes discussion of a number of applications that link GIS systems with software for multi-attribute decision analysis, linear programming, interactive programming, and even mentions linking GIS systems to spreadsheet based optimization. Although never quite using maps *per se*, Klosterman[11], et. al. (1993) developed a number of spreadsheet models for urban and regional analysis some of which include demographic, economic, and other geographically based data sets.

Journal articles reporting on applications incorporating OR/MS models with GIS models (usually by linking two or more disparate computer systems) are also quite numerous. From **Interfaces** alone, in the decade (or so):

- Blakely[12], et. al. (2003) developed a GIS linked to OR models to schedule preventative maintenance for Schindler Elevator Corp.,
- Weigel and Cao[13] (1999) linked GIS and vehicle routing algorithms to help Sears with their technician dispatching and home delivery routing,

- Fletcher and Alden[14] (1999) used GIS, a database resource capability model, a policy-alternative model, and linear programming for long-term forest ecosystem planning at Pacific Lumber,
- Begur[15], et. al. (1997) combined a PC-based GIS with scheduling databases and heuristics for scheduling and routing home health care nurses,
- Bucciarelli and Brown[16] (1995) developed a desktop decision support system combining GIS and traveling salesman heuristics to support the U.S. Coast Guard buoy tending operations, and
- Kuby[17] at. al. developed a mixed integer programming model linked to a GIS to assist with planning China's coal and electricity delivery system.

There have been a few previous discussions in the literature specifically exploring the concept of implementing GIS in spreadsheets. In the early 1990's, Raubal[18] *et al.* (1997) demonstrated the concept of importing GIS raster data into Excel as a pedagogical device, and their students were able to develop models with it, although it was not intended as an actual application. Charles Ehlschlaeger[19] developed an application involving a promising method of piping linear programming functionality from Mathematica, an Excel Plug-in, into a GIS.

Cole[20] (1998) wrote a helpful paper focusing on the technical aspects of using spreadsheets to produce maps "from scratch" including drawing map objects and coloring them so as to display spatial data. Supporting the overall theme of this paper, Cole writes that spreadsheets:

> *"Can be used to prepare acceptable maps rather quickly and gives more direct links between data, analysis, and mapping, enabling more effective GIA (geographic information analysis), and can be used for quite large-scale applications."*

Later in his paper Cole concludes:

> *"Spreadsheets provide a tool to explore ideas for novel interfaces or operations, and avoiding some of the continuing frustrations of mainstream GIS, but which might subsequently be implemented within GIS as part of students' own research*

or professional kit bag. Important here is the direct link to the other facilities of spreadsheets for data processing and model construction."

Although it didn't use spreadsheets, we also note a Decision Support System (DSS) involving OR/MS techniques used on geographical data by Keisler & Sundell[21] (1997). It was the recognition that the ability to operate on GIS data directly from Excel was possible and would have greatly simplified the development of the Keisler & Sundell DSS that provided the impetus for this paper. In that application several additional features could have been incorporated had we been operating in the faster DSS prototyping environment provided by Excel.

The contribution of this paper then is to extend these previous concepts into realistic GIS-all-within-a-spreadsheet applications to provide the seamless integration of geographic data, OR/MS analysis, and mapping technology. Our applications demonstrate how to tie the analytical power of spreadsheets to geographically defined data. Our examples that will illustrate the "what", will discuss the details of "how" and we hope will motivate the "why".

3. IMPLEMENTATION OF SPREADSHEET BASED GIS ("HOW")

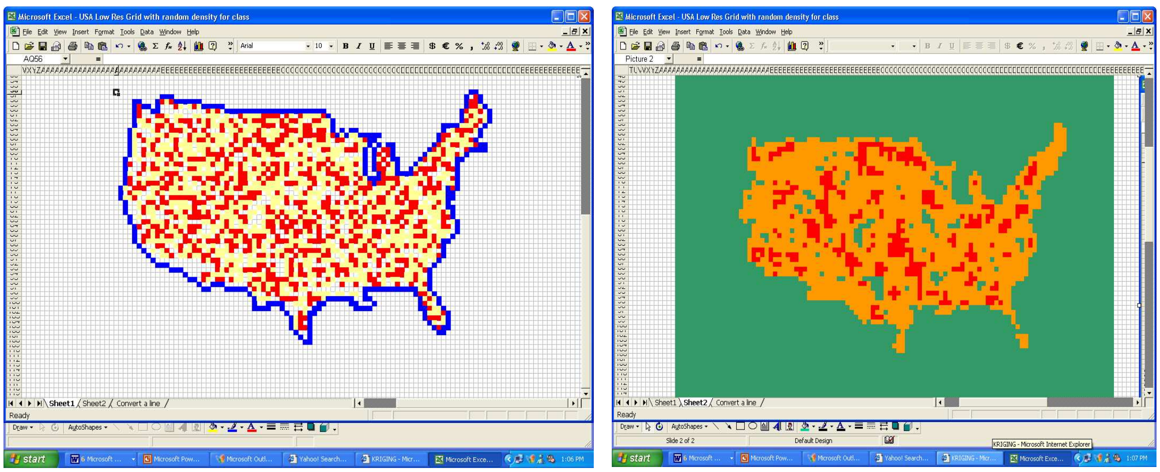
In this section we will describe the technical details of how several basic GIS functions can be implemented in spreadsheets and how to incorporate common GIS based data types into spreadsheets. These technical ideas will be the building blocks used to develop the Section 4 applications incorporating these GIS functions into OR/MS analyses.

3.1 Basic Mapping Functions

One of the basic GIS data types for storing mapping information is known as *raster data*, involving square or rectangular *pixels* (also even in the GIS world called *cells*) arranged in rows and columns where the logical position in the data array corresponds to a physical position. In raster data formats, each cell contains a single value. Often this value is a color, but it can also be a numerical value indicating values such as land use or elevation above sea level. Assuming the pixels contain colors, a raster display will the produce what we would recognize as a *map*.

The analogy between a raster display and a spreadsheet is then straightforward. By treating each cell of the spreadsheet as a pixel, sizing the cells as squares or small rectangles, and using the cell contents to specify a color property (or other appearance), spreadsheets can be used to produce maps. The spreadsheet function of *conditional formatting* of cells, which changes the appearance of a cell depending on its contents, is a critical capability for this application. With conditional formatting (available from "Format" on Excel's toolbar) it is possible to change the color of the background, the border, and the font and color of cell values shown; for our purposes, we simply change the background color and make the cell border invisible by having it appear in the same color. Conditional formatting is limited to three possible different conditions per cell (plus a default format) --- although different cells can use different conditional formats. Figure 1 shows a spreadsheet based representation of a map of the U.S., the map on the left is filled with random integers (using the equation "*=INT(RAND()*12)+2)*"), between 2 and 13, and the cells representing border of the region contain a value of 1. The cells are sized to a height and width of 8 pixels, and the format applied makes a cell blue if it contains a 1, yellow if its value is in the range from 4 to 8, red if its value is in the range from 9 to 12, and white otherwise (i.e., if its value is 2 or 3).

Figure 1. Conditional formatting allows spreadsheet cells to function as GIS cells, with original map (filled with random integers) on left and smoothed version on right.



In the GIS world, the standard approach is to store different elements of geographical information in layers; example layers might be elevations, town borders, roads, and population. Again, there is a natural analogy between raster layers and spreadsheet worksheets.

In Microsoft Excel, each worksheet can be up to 256 columns and 65536 rows. For most applications 256 columns would be inadequate for professional quality graphics, but certainly detailed enough for some purposes. .Other spreadsheet products such as Borland's Quattro Pro and Lotus 1-2-3 contain 16556 columns and 16556 rows per worksheet, allowing higher resolutions. With some additional ingenuity, such as treating each worksheet as a vertical strip of a map and then piecing them together, larger maps could be produced. Alternatively cells could be sized to be horizontal rectangles, so that maps would have less horizontal definition than the maximum allowable vertical definition.

Another basic mapping function is known in the GIS world as *map algebra*, and again there is a natural analogy with spreadsheet cell functions. Spreadsheet cell formulas provide considerable flexibility in defining new map layers. With a few keystrokes users can create a new worksheet (layer) with *formulas* involving values from cells of other worksheets (layers). Using formulas such as AVERAGE or SUM allows map smoothing and aggregation. For example, in Figure 1 in the screen on the right, we applied a map layer that smoothes out the pattern in the first sheet by summing values over small regions. Cell AE9 on Sheet2 contains the formula "*=SUM(Sheet1!AD8:AF10)*", and this cell was copied to all cells up to EC 150. The conditional formatting applied makes the cell green if its value is less than 50, orange if its value is between 50 and 75 and red if its value is above 75.

Additionally, IF statements can be used for filtering map layers. For example, if Sheet1 contains buildings and Sheet2 contains roads, we can calculate the places available for building (not on top of buildings or roads) in Sheet3 with the formula =IF(Sheet1!a1=0,IF(Sheet2!a1=0,1,0),0).

Thus the combination of using spreadsheet cells as pixels, conditional formatting to modify cell appearance, worksheets as data layers, and cell formulas as map algebra already gives considerable GIS functionality to spreadsheet models. Cole (1998) used some similar ideas for using spreadsheets to produce maps, and invoked the metaphor of spreadsheet mapping as “word processing in colors”.

It is important to note that due to the large size of typical GIS raster displays, efficient calculation strategies within the spreadsheet become increasingly important. Depending on the size of the spreadsheet and the parameters of the hardware on which it is running, some of the approaches described here could lead to slow calculations. The main factor driving calculation speed is the large number formulas that must be computed for each cell. Simple formulas, such as the conditional formatting take only a few seconds for a large area. Complex formulas, as would be used in layers of maps, will take longer. To facilitate efficient calculation it is possible to keep an equation in one cell and use Excel’s copy-paste-values command to freeze the remaining values when they are not in use. This also reduces memory requirements. Another way to speed calculation is to shift the spreadsheet to manual calculation mode, and calculate one sheet at a time. Of course, smaller raster displays will use less memory and compute faster; if a grid of 100x100 suffices for a particular application the smaller/coarser representation should be used.

3.2 Incorporating Existing GIS Data Sets into GIS Spreadsheets

In this internet age, there is a wealth of existing and publicly available GIS data out there. Merely brushing the surface, government agencies such as the U.S. Geological Survey (USGS) and the Canadian Geospatial Data Infrastructure (CGDI) have extensive sets of geographic information and analyses of their own countries and the world. There are also major GIS open source and user groups (such as GRASS), and GIS companies (such as ESRI) which create and make available extensive libraries of GIS data. With a search engine and an internet connection, there's a good chance the geographic information needed for pretty much any OR/MS application is available. The issue, however, is how to get it into a spreadsheet.

The technique we develop in this paper, which is only one of several that could be used, is based on the fact that bitmap images are made of pixels. Thus any map display can be imported, pixel by pixel into a spreadsheet simply by "picking up" the color of the pixel as a cell value. The source map can then be reproduced within a worksheet. Appendix I provides the VBA code for this process, linking to a user-specified bitmap file (*.bmp) and reproducing the map within a worksheet. For the examples later in this paper, this VBA subroutine was used to import our campus map (from our campus website) and a population density map of the eastern portion of Massachusetts (from the U.S. Census website).

This "bitmap conversion" procedure has limitations however. The first, as mentioned before as an issue with spreadsheet based GIS, is dimensionality. The Excel

2003 limit of 256 columns (resulting in a width of 256 pixels) limits the size and resolution of data that can be imported (in one of our applications, for example, the text for building names on the campus map is pretty hard to read). This method can also make building layers a bit difficult, as the pixel reading process just takes what it sees, and so for example doesn't differentiate between a pixel containing a population density value from a pixel that's black indicating a road (which we'd prefer on a different layer). Also, the pixel by pixel method can make it hard to align data layers (e.g. from different sources). However this alignment issue can be circumvented in some cases. For example the census site allows a user to set up a particular map and then display various values on it (population density, % of different races, income data) which, if brought in with this method, would already be aligned.

A deeper issue is that as well as raster based displays, GIS data often comes as vector data which defines objects using various coordinate schemes. The population density in the census data set is, in fact, vector based data. Population in the census is not counted per square mile (which would be a raster representation) but by irregularly-shaped census blocks (a vector representation). The population density map is then a rasterized display of vector data, and is then only an approximation. In general though, since vector data sets all refer back to some kind of grid-based coordinate system, converting vector based data sets to raster displays should be possible.

There are numerous existing standard *GIS data formats* (e.g. a specific protocol or procedure used to store and manage data (Gardels[22])). These data formats include ESRI shapefiles, ArcInfo coverages, MapInfo files, GeoTiff, GRASS run-length encoding for rasters, various digital elevation formats (DEMS, SDTS, etc), and various GRID and image formats. Some of these are binary raster formats in which data items accompany the image data, which can solve some of the alignment issues noted above. Although beyond the scope of this paper, it would certainly be quite possible to develop VBA macros to import data from common schemes for raster and even vector based geographic data.

3.3 Additional Spreadsheet Based GIS Functionality

Within the raster display, *regions* (sets of cells) can be created, named, and referred to in formulas. The region is selected by holding the control key while dragging the mouse over cells or clicking on cells, and when the region is selected, the user chooses Insert -- Name -- Define from Excel's menu. Rectangular regions can be defined by dragging the mouse, and can include border lines, and are named the same way. This feature could be useful for customizing maps – for example, it would be possible to select a region (using the "go to" command) and then assign a different conditional formatting scheme to that region. It is also possible to convert parametrically defined shapes into raster representations, for example, after defining a rectangular area in terms of its upper left corner (xorigin, yorigin), length and width, we could populate the cells within the rectangle with 1's by using the formula:

```
=IF(AND(row()>y,row()<y+length,column()>x,column()<x+width),1,0)
```

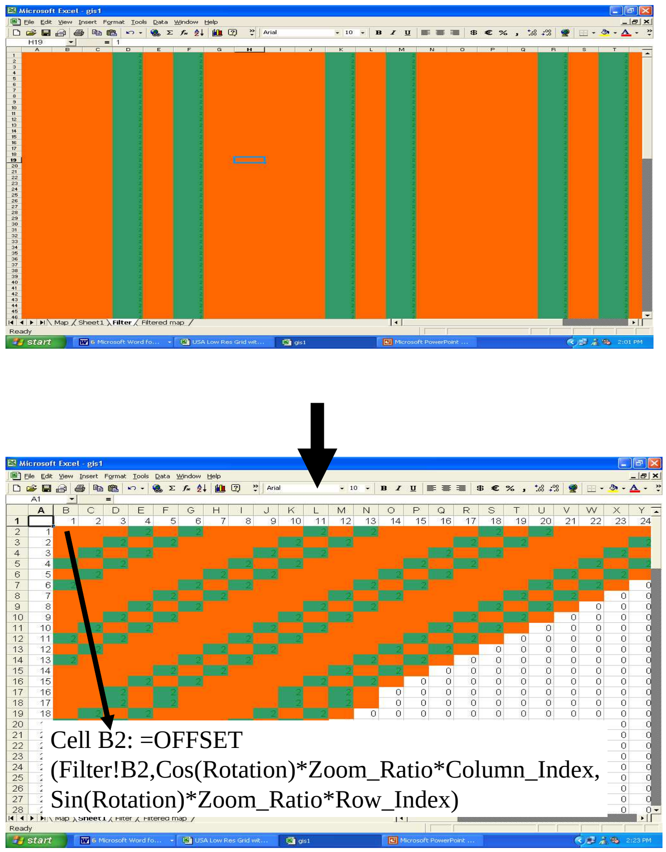
We could define other shapes similarly, e.g., circles.

Regions could also be the basis for an implementation of *vector based data*. Assuming the boundary issues can be worked out, for example, a data layer could be produced which gives the town, or census block, or other vector based identification of each pixel. Such a data layer would then allow considerations of *topology*, namely determining relationships of how different vectors are contiguous or connected, since adjacent vector blocks could be determined by looking for a difference in value between adjacent cells. Another idea is that shapes stored as vector data could be converted to a pixel representation using a map layer that uses *algebraic formulas to define vector objects*. For example, a circle is defined by a center and a radius, and the formula identifies a point as being in the circle if its distance from the center is less than or equal to the radius. In addition *Excel drawing objects* have associated parameters that may be manipulated and converted to parameters for generating pixel shapes using VBA. Objects themselves may be stretched or moved by hand or using VBA, although it would be inelegant.

Another common GIS function, *transformations* between coordinate systems and *projections* by rotating and stretching map regions is also straightforward to implement in a spreadsheet. The row and column numbers are treated as Cartesian coordinates. To transform the coordinates, we select the appropriate cells in a source worksheet using the offset function. Rotation is achieved using sine and cosine of coordinates rounded to the nearest integer to generate the new coordinates. Figure 2 illustrates this concept. In figure 2 a segment of the "filter" sheet described above has been rotated by 70 degrees and transformed using this technique. Here, we filled the cells in the top row of a sheet with the numbers 1, 2, 3… and also filled the first column of the sheet with the numbers 1, 2, 3… Then in the cell B2, we entered the formula:

=OFFSET(Filter!A1,xfactor*B$1,yfactor*$A2), where xfactor is the sine of the angle and yfactor is the cosine of the angle.

Figure 2: Sines and cosines are used as arguments of an offset function, to rotate a display.



Again the limited dimensionality of spreadsheets also limits the application of this particular technique; rotated maps are going to look quite “unsmooth”. Spreadsheets also seem ill suited to global projections (e.g. mapping a globe or sphere onto a rectangle); although with a macro it would certainly be possible to generate a new map by transforming addresses cell by cell. A simple way to stretch or compress a map is to select all cells and modify the cell height and width to the desired number of pixels. Stretching or compressing could also be done using formulas; the technique would be similar to that for rotation, though simpler. A rectangular range of cells can be transposed by 90 degrees simply using Excel’s Copy -- Paste -- Transpose command.

3.4 Spreadsheet GIS for Communication Support

For GIS to be used effectively within organizations, particularly corporations, there needs to be easy interactions between GIS experts and other departments. We believe that spreadsheet based GIS systems, even just as “think pieces” or prototypes, are exactly the solution to the organizational challenge of maximizing the sharing and use of GIS data and information throughout an agency. Historically, GIS has been separated from the rest of the organization because of its complexity and its special technology needs (large workstations, plotters, etc.). Many have described GIS as a “back-office” technology (e.g., Castle[23], 2002, Smurfit[24], 1995). We propose that spreadsheet based GIS may provide exactly the means to bring GIS to OR/MS analysts, financial analysts, data mining experts, project managers, and even the corner office. The language of spreadsheets facilitates cross-functional communication and sharing of expertise. Broader acceptance could then increase overall understanding of the benefits available from analyzing spatial data.

There are a number of spreadsheet functions that should support GIS integration and communication. For example:

- **Exporting graphics:** A screen can be converted to a bitmap image using the Control-Shift-PrintScreen keys to copy the screen image and then paste it into Powerpoint, Paint, or other compatible graphics programs (including Microsoft Word as was used in this paper).
- **Web-publishing:** Spreadsheets may be saved as web pages that can be viewed using Microsoft Internet Explorer. When the spreadsheet is saved, the user can specify the level of access (values, formulas, modifiable formulas) available to viewers. Similarly, it is easy to embed web links within a spreadsheet.
- **Use of real time data:** Excel supports live web queries, and in Office XP, also provides rich support for the use of real-time data on the world-wide web. Earlier versions of Excel connect to real-time financial data. One illustrative possibility would be applying real-time regional weather information to a map stored within the spreadsheet, by incorporating weather parameters in the cell formulas.
- **Collaboration:** Excel has reasonable functionality supporting sharing and distributing of workbooks. These allow multiple users to access and modify the same sheet. The versioning support is built in, as are personalized views, annotation and access rights. Security features such as protection and hiding of sheets are also available. Auditing tools help individual or multiple users trace a model's logic, which can aid in debugging.
- **Security functions:** Within a spreadsheet, individual cells, ranges or worksheets can be protected. This could be useful if certain data is not to be modified by

some operation (e.g., only edges are affected), as well as for public data and collaboration (if different people are allowed to modify different data or scenarios). Similarly, specified cells, ranges, columns, and rows or whole worksheets can be hidden (and locked). Notes attached to cells may also be hidden along with the indicator showing the presence of a note, which could be useful for private annotation of sensitive information.

4. APPLICATIONS (HOW AND WHY)

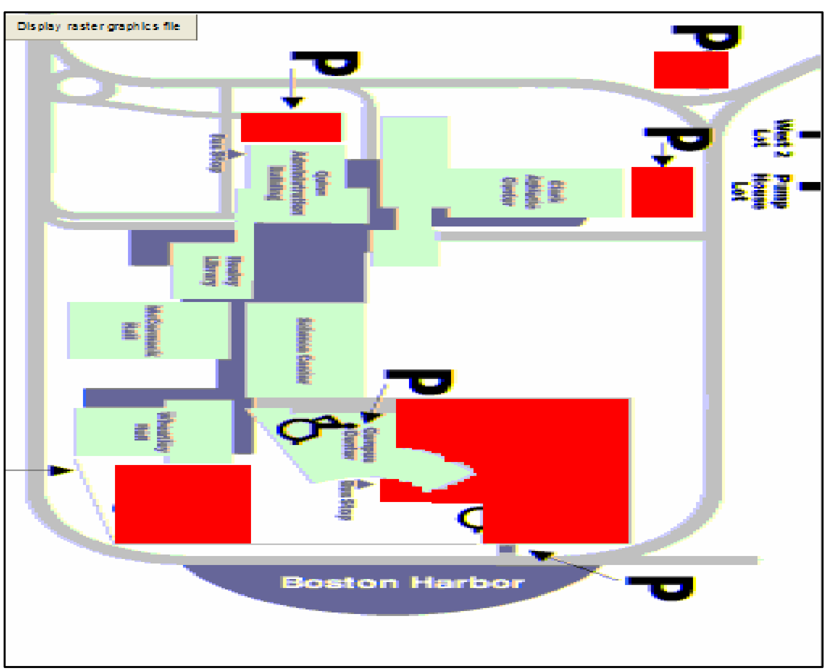
This section provides two examples of how spreadsheets can be used to integrate analytical calculations and GIS data. Each of these short examples demonstrates techniques discussed above. Both are based on importing a raster image into Excel, and then applying various worksheets as layers onto that imported image. Both use combinations of Excel functions such as ROW, COLUMN, and INDIRECT to access and manipulate data from that image. Other than the VBA subroutine which imports the image, neither example is dependent upon any additional macros.

4.1 Parking Lot Analysis

The first example is intended to present a GIS based spreadsheet designed to facilitate easy data input and analysis. This example was motivated by an issue faced by many organizations, namely how to cope with demands for parking. In our university, the primary parking facility was a 1,500 space parking garage which was found to be structurally unsound and therefore closed, putting a tremendous premium on parking on our campus. Our parking squeeze, while exacerbated by a unique situation, is not especially unique among colleges and universities.

This example presents a rapid prototype of a tool to assist facilities and similar managers to size and place new parking facilities. Figure 3 shows the raster data representing a map of the campus that was imported into a worksheet. Although the 256 pixel width limitation renders the names buildings difficult to read, the overall map and building locations are clear. On a separate worksheet (Figure 4), the user enters a value for the size of a parking space (usually a value larger than the square feet covered by a parked car to allow for lot circulation), the cell size (based on the scale of the map), and a location for the campus "center". The user then interacts "live" with the map by constructing shapes directly on the worksheet representing existing and proposed parking facilities, either by directly shading parking areas to a specified color, or by inserting a predefined value which conditional formatting then uses to set the background shade of cells to that same color.

Figure 3: A bitmap file of campus map can be imported as a raster display.



Showing the advantage of using a spreadsheet (instead of a paper map and a pencil), useful characteristics of a given parking configuration are then easily and dynamically calculated and displayed. Our example interactively calculates and displays (see Figure 4) total parking coverage, expected parking spots, and average distance to the center of campus for any given configuration of parking areas. We envision this application enabling “charrette” type sessions, where groups of people participate in a meeting where various alternative parking schemes are discussed and refined.

Figure 4: Based on user inputs and parking areas specified, information about a given parking configuration is interactively calculated.

	C	D	E	F	G	H	I	J
1								
2								
3								
4								
5								
6								
7								
8	Parameters		Square feet per parking space	325				
9			Square feet per pixel	36				
10			Weighted distance exponent	2				
11			Cost per sq feet for parking	$ 1.50				
12								
13								
14			Building	Location	Weight			
15			Quinn Hall	CS122	5%			
16			Healey Library	BU167	10%			
17			McCormack Hall	BF208	20%			
18			Wheatley Hall	BD277	20%			
19			Science Building	CN224	10%			
20			Clark gym	ER119	5%			
21			Student Center	DH295	30%			
22					100%			
23								
24								
25								
26								
27	Outputs		Total parking spaces on campus	1,389				
28			Percentage parking of total area	12%				
29			Weighted average value function	17,570				
30			Total cost function	14,069				
31								
32								
33								
34								
35								

Technically, each of the shaded cells (pixels) represents a specific square footage for parking based on the scale of the imported bitmap, and therefore a portion of a single parking space. A second worksheet uses the map scale to calculate the distance from each cell to the cell represented the center of campus, and a third worksheet has a formula for each cell which produces a 0 if the cell is not part of parking space, or 1 if it is. The average distance to the center of campus is therefore proportional to the average value of product of the cells in the second and third worksheets. The total number of parking spaces, total parking square feet, and parking lot coverage is based on the sum of the cells in the third worksheet.

To demonstrate how this prototype could be easily extended, we introduced some additional factors into this example. Instead of calculating a single average distance to the center of campus, we calculated the distance to each building on campus (shown in Figure 4). To accomplish this task, an additional overlay worksheet for each building was created. We then added parameters which were estimates of the percentage of traffic going to each building.

These parameters were used to calculate a value function for a proposed configuration of parking spaces, instead of an average distance to a single point. This value function, whose calculations were stored in another overlay worksheet, presumed the value of each parking space to be inversely proportional to the weighted distances from destination buildings, and was of the form:

$$V = \Sigma_i (p_i / d_i^{\lambda})$$

where:

V is the value of a parking space,

p_i is the proportion of traffic going to building i,

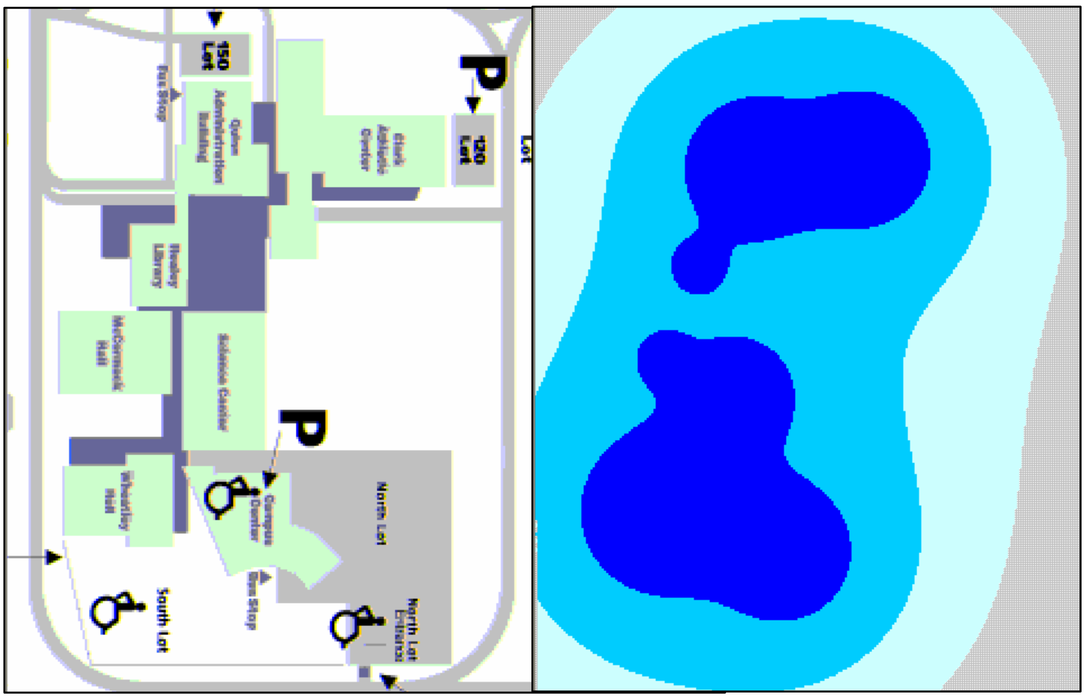
d_i is the distance from the parking space to building i, and

λ is a weighting exponent for distance.

The value for λ was entered as a parameter and for the example we used a value of 2.

The addition of this value function enabled us to use conditional formatting to produce a contour map of the potential parking value of each location on campus. We could then compare the original imported bitmap of the campus map with the plot of the value of having parking in each location. We termed this the “sweet spot analysis”, and a sample is shown as Figure 5.

Figure 5: GIS based spreadsheet enables the analysis of “sweet spots” for campus parking.



Finally we added a cost function intended to offset the value function, again represented in an overlay worksheet. This cost function was a straightforward calculation based on parameters for cost per square foot of parking spaces, the number of square feet in a parking space and pixel, and the total number of parking spaces for a given configuration.

4.2 Retail Store Location

This second example was motivated by the problem of determining retail store locations, which one of the authors has previously confronted for several retail chains. The scenario represented in this example is that of a “big box” retailer seeking to locate two stores in the Boston area where we postulated a competitor had already located one store.

Huff’s[25] model (1964) defined estimations of trading areas based on distances to stores and completeness of merchandise selection, and many variations of his model have been used over the years for many retail site selection analyses. (For example, see Stanley and Sewall[26] (1976) and Gautschi[27] (1981)). In common to these models is a determination of retail potential as an inverse function of distance, with an underlying assumption that the attraction of a retail store diminishes with distance. In our example, we used a logit-demand model which estimated the proportion of potential customers who will consider shopping at a retail store to be related to distance from that store as:

$$p_i = 1 / (1 + d_i^{\lambda}),$$

and the total proportion of potential customers shopping at the store as:

$$xj = p_i / \Sigma_J \, p_j$$

where:

p_i is the probability a customer will shop at retail store i,

d_i is the distance to retail store i,

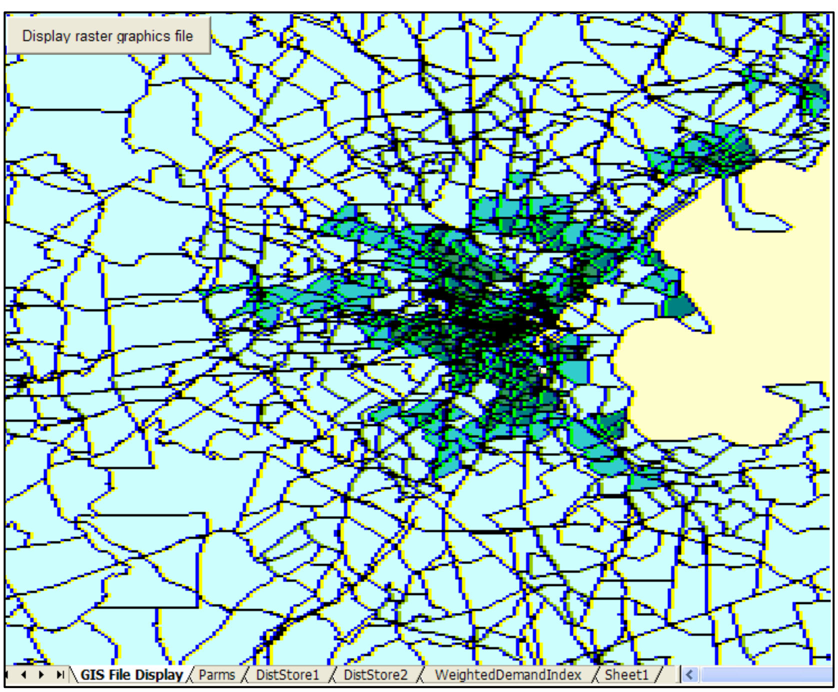
λ is an empirically determined constant. There have been previous attempts to estimate this constant; for our example, we chose a value for λ of 1.5, and

J is the set of store locations

Using this model, a share is calculated for each retail store, including competitive stores, and for a presumed store at a fixed distance from all customers. The distance of the presumed store is set at a figure somewhat beyond the expected trading areas of an actual store and its demand represents customers who find all of the stores too far away, and either turn to other channels such as mail-order, or constitute unfulfilled demand.

In order to apply this model, we needed to have population data. The most suitable population data we could find was from the U.S. Census Bureau[28], which was for population counts by 5-digit zip code for the year 2000. We downloaded a graphic of this for the Boston area and imported it into our spreadsheet; the result is shown as Figure 6. At the time this graphic was imported, a population density value was attached to each cell based on the image's legend. This was somewhat of a manual effort, and an example of a difficulty that could be overcome through use of the binary raster format. Unfortunately, we could find no readily accessible graphics for population in this format.

Figure 6: U.S. Census information, in this case population density, can be imported as a raster display



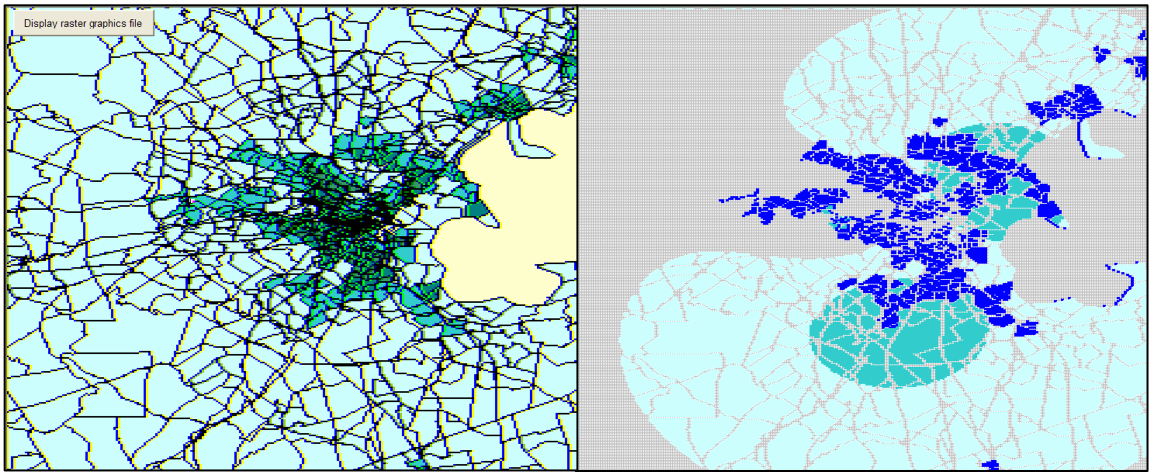
The overlay sheets for the store analysis included one for each retail store which contained a calculation of the distance from each point on the imported map to that store, and one for the competitor's store location. Another overlay sheet represented the probabilities of customers shopping at each store location (including the competitor's) and one was used to determine the proportion of market share that would accrue to the two stores in the configuration. The sum of the portion of market share for the two stores, from the last overlay sheet, was taken as an estimate of total sales. The input parameters and sample output of the potential sales estimate are displayed in Figure 7.

Figure 7: User- defined inputs enable interactive calculation of logit based sales potential

	A	B	C	D	E	F	G	H
1								
2								
3								
4								
5								
6				Location of store 1	FS113			
7				Location of store 2	EE204			
8				Location of competitor	DJ136			
9				Pixels per mile	3			
10				Distance parameter (miles)	25			
11				Market size (per capita)	$ 90			
12								
13								
14				Estimated total potential sales	$ 1,552,820			
15								
16								
17								
18								
19								
20								

Working with these parameters, the user could locate stores in any cells and determine an estimate of total sales. This interactive nature can be highly effective in practice, as often the user wishes to include considerations that are not likely to be part of a model, but still wishes to see the impact of various location scenarios on the model's output. Using conditional formatting, the sales estimate was displayed on an overlay sheet that could be compared to the original imported raster image. Figure 8 shows sample results, in which the candidate store locations were north and south of the center, and the competitor's location was west of the center. This figure shows the trading areas around each of the candidate locations, and demonstrates how the strength of those trading areas diminishes with distance, and also with proximity to the competitor's outlet located to the left of center in the map.

Figure 8: Spreadsheet GIS enables display of geographical sales potential



Reasonable extensions to this example would have been to use Solver iteratively find the store locations with global maximum projected sales, write a VBA macro to find the best locations through exhaustive enumeration, or implement a different technique for finding the best set of locations such as a genetic algorithm.

5. CONCLUSION AND DISCUSSION

We are certainly not claiming that spreadsheet based GIS implementations are going to replace the entire industry (proprietary and open-source) that currently exists for GIS technology. The dimensionality limits of spreadsheets, the relatively slow calculation rate for applications involving large number of cells, and the issue of forcing a general purpose tool (spreadsheets) to do a very specific purpose (GIS) will make spreadsheet based GIS systems useful only in certain circumstances.

However, to the non GIS expert, existing GIS technology is complex and daunting. Interacting with GIS systems implemented within spreadsheets opens up GIS functionality quickly and intuitively to the millions of business and other spreadsheet users across the globe. We also note the potential of this integrated application for classroom purposes; in our experience these visual displays are engaging to students and really help them grasp the importance of the underlying OR/MS models.

In this paper, we've shown the straightforward and natural analogy between several GIS functions with spreadsheet functions, particularly for raster based data. We've discussed ways in which this GIS functionality can be implemented within the spreadsheet environment. The examples we've produced show the useful synergy that comes from meshing OR/MS methods with GIS methods --- an integration which is greatly enhanced by the "remarkable development platform" provided by spreadsheets.

Future work will focus on meshing advanced OR/MS techniques with GIS data. We've mentioned the possibility of using optimization techniques for the store location problem. Another intriguing possibility is to use the random number generation functions in Excel to set up geographically based stochastic simulations.

Perhaps the most promising aspect of the use of Excel for geographical analysis is its modeling capabilities, particularly dynamic and stochastic modeling. We see applications for time series modeling and for scenario analysis in the use of Excel in conjunction with data that has been processed in a traditional GIS environment. The opportunity for a larger population of modelers to develop shared spreadsheet-based geographic decision support tools is wide open. We hope this dynamic will draw more people to think of GIS and quantitative analysis together.

REFERENCES

[1] Sohovich, Martin. 2002. Business and IT Ramifications of Geocoding in Customer Data Integration. *Directions Magazine*. URL: http://www.directionsmag.com/article.asp?ArticleID=174.

[2] Longley, P., Clarke, G., eds. 1995. *GIS for Business and Service Planning*. John Wiley & Sons, New York.

[3] O'Sullivan, Gerald. *"Satellite Imagery and Human Rights."* Working paper. American University School of Communications.

[4] Caravallo, D. 2002. ESRI Corporation. Email Correspondence. March 4.

[5] GeoWorld, 2004. http://www.geoplace.com/gw/1996/0496/0496feat2.asp

[6] U.S. Geological Service, Geographic Information Systems (GIS) Poster, http://erg.usgs.gov/isb/pubs/gis_poster/

[7] Lo, Chor Pang and Albert K.W. Yeung. 2006. *Concepts and Techniques of Geographic Information Systems*, 2nd Edition. Prentice Hall, Englewood, N.J.

[8] Longley, Paul A. and Michael F. Goodchild, David J. Maquire, and David W. Rhind. 2005. *Geographic Information Systems and Science.* John Wiley & Sons, New York.

[9] Longley, P., Clarke, G., eds. 1995. *GIS for Business and Service Planning.* John Wiley & Sons, New York.

[10] Malczewski, Jacek. 1999. *GIS and Multicriteria Decision Analysis.* John Wiley & Sons, New York.

[11] Klosterman, Richard E., Richard K. Brail and Eric G. Bossard. (eds.) 1993. *Spreadsheet Models for Urban and Regional Analysis.* Center for Urban Policy Research, Rutgers University, New Brunswick, NJ.

[12] Blakely, Fred, Burcin Bozkaya, Buyang Cao, Wolfgang Hall and Joseph Knolmajer. 2003. Optimizing Periodic Maintenance Operations for Schindler Elevator Corporation. *Interfaces.* Vol. 33(1), pp. 67-79.

[13] Weigal, Don. and Buyang Cao. 1999. Applying GIS and OR Techniques to Solve Sears Technician-Dispatching and Home-Delivery Problems, *Interfaces.* Vol. 29(1), pp. 112-130.

[14] Fletcher, L. Russell and Henry Alden. 1999. Long-Term Forest Ecosystem Planning at Pacific Lumber. *Interfaces*. Vol. 29(1), pp. 90-112.

[15] Begur, Sachidanand V., David M. Miller and Jerry R. Weaver. An Integrated DSS for Scheduling and Routing Home-Health-Care Nurses. *Interfaces*. Vol. 27(4), pp. 35-48.

[16] Bucciarelli, Mark and Kip Brown. 1995. A Desktop-OR Success Story: Modeling Coast Guard Buoy Tenders. *Interfaces.* Vol. 25(4), pp. 1-11.

[17] Kudy, Michael, Shi Qingqi and Thawat Watanatada. Planning China's Coal and Electricity Delivery System. *Interfaces*. Vol. 25(1), pp. 41-68.

[18] Raubal, Martin, Gaupmann, Bernhard and Kuhn, Warner. 1997. Teaching Raster GIS Operations with Spreadsheets. *Journal of Geography*. 96, 5.

[19] Ehlschlaeger, Charles. *The GRASS/Mathematica Link: Developing Hydrological Models in Geographic Information Systems Interfaced with Algebra Systems.* Champaign, Illinois: U.S. Army Construction Engineering Research Lab.

[20] Cole, S. 1998. Of maps and macros: object-oriented spreadsheet GIS. *Environment and Planning B: Planning and Design*. Vol. 25, pp. 227-243.

[21] Keisler, Jeffrey M. and Sundell, Ronald C. 1997. Combining Multi-Attribute Utility and Geographic Information for Boundary Decisions: An Application to Park Planning *Journal of Geographic Information and Decision Analysis*, Vol 1.2, pp. 101- 118.

[22] Gardels, Ken. The Open GIS Approach to Distributed Geodata and Geoprocessing. URL: http://www.regis.berkeley.edu/gardels/envmodle.htm

[23] Castle, Gil. 2002. A Game of Information Arbitrage. *Directons Magazine*. URL: http://www.directionsmag.com/article.asp?ArticleID=142.

[24] Smurfit, Michael. 1995. *Using a GIS as a DSS Generator.* Working paper. Department of Management Information Systems, Graduate School of Business, University College, Dublin.

[25] Huff, D. 1964. Defining and Estimating a Trading Area. *Journal of Marketing*, Vol. 28 No. 3, pp. 34-38.

[26] Stanley, T. and Sewall, M. 1976. Image Inputs to a Probabilistic Model: Predicting Retail Potential, *Journal of Marketing*, Vol. 40, No. 3, pp. 48-53.

[27] Gautschi, D. 1981. Specification of Patronage Models for Retail Center Choice, *Journal of Marketing Research*, Vol. 18 No. 2, pp. 162-174.

[28] U.S. Census Bureau. 2006. American Fact Finder, Thematic Maps. URL: http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=DEC&_submenuId=&_lang=en&_ts=