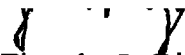


AN ABSTRACT OF THE THESIS OF

Michael D. Halbleib for the degree of Master of Science in Soil Science presented
on August 28, 2001 Title: Using Advanced Spreadsheet Features for Agricultural
GIS Applications

Redacted for privacy

Abstract approved: _____


Timothy L. Righetti

A GIS analysis procedure was developed to explore relationships between imagery, yield data, soil information, and other assessments of a field or orchard. A set of conversion utilities, a spreadsheet, and an inexpensive shape file viewer were used to manipulate, plot, and display data. Specific features described include procedures used to: 1) display automated yield monitoring and aerial imagery data as surface maps for visual analysis, 2) generate maps from gridded soil sampling schemes that display either the collected soil data values or management information derived from further manipulation of the sample values, 3) evaluate relationships among data layers such as yield monitor, imagery, and soil data, 4) conduct an upper boundary line evaluation of potential yield-limiting factors. The analysis process is demonstrated on wheat, meadowfoam, and hazelnut data, from crops grown in Oregon.

Using Advanced Spreadsheet Features for

Agricultural GIS Applications

by
Michael D. Halbleib

A THESIS
Submitted to
Oregon State University

In partial fulfillment of
the requirements for the
degree of

Master of Science

Presented August 28, 2001
Commencement June, 2002

Master of Science thesis of Michael D. Halbleib presented on August 28, 2001

APPROVED:

Redacted for privacy

Major Professor, representing Soil Science

Redacted for privacy

Chair of Department of Crop and Soil Science

Redacted for privacy

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for privacy

Michael D. Halbleib, Author

ACKNOWLEDGEMENTS

Many people have helped in the completion of this task in so many ways, through their insightful contributions, constant encouragement, and patient endurance.

Special mention and thanks goes to Dr. Tim Righetti for his time spent in teaching, no matter how basic the question, his willingness to share no matter what the hour, and his never ending enthusiasm for learning. My life has been deeply affected and I will always be grateful for his support and direction, which made this endeavor possible.

Special thanks goes to my partner Mary, who contributed in many ways, from basic editing tasks to threats of physical violence to help motivate me. She has been supportive in this final task.

As always, the list is long for those who contributed help and encouragement, thanks goes to all the members of my committee, my parents and family, and all the friends who have offered encouragement, love, and support along the way.

TABLE OF CONTENTS

	<u>Page</u>
Abstract.....	1
Introduction	2
Materials and Methods.....	7
Data Collection.....	8
Wheat.....	8
Meadowfoam.....	9
Hazelnuts.....	10
Data Visualization and Analysis.....	12
Overview of procedures.....	12
Pivot tables.....	17
Identification of boundary line conditions.....	20
Results and Discussions.....	22
Displaying automated yield monitoring and imagery data that are cropped to a field boundary.....	22
Generating maps from soil grid sampling schemes.....	26
Evaluating relationships for spatial data.....	33
Boundary line evaluations.....	37
Conclusion.....	43
References.....	45

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1a. Kriged yield data, as a surface map, depicting the wheat field yields as contours. The soil-sampling regions are overlaid as blue polygons that are representative of the area from which each soil sample was taken.	14
1b. A enlarged view showing the corner of one soil-sampling polygon. Within the polygon are red points that are set to a regular 2 m grid interval; these points are assigned the same attributes as the soil-sampling polygon and have the same spacing as the yield data points.	15
1c. An enlarged close up of the same wheat field that is depicted in Fig 1a and b. showing the spatial relationship that exists between data layers by displaying all of the data layers as overlapping different sized points. The largest points that are shown are underneath as the bottom layer and are representative of the yield data, with the medium sized points representing the image data, and the smallest points that are on top represent the soils data. This helps illustrate the spatial relationship that exists between the data layers and shows how the information, when set to a uniform grid spacing, can be managed using the location as the common link.	17
1d. A page data of results that were generated from a query using the Excel pivot table features.	19
2a. Image data converted to a set of points on a 2 m interval with the composite band digital values inverted (0=255 and 255=0) so that the values are representative of increasing darkness. It is displayed as a shape file and classified into 10 categories.	23

LIST OF FIGURES CONTINUED

<u>Figure</u>	<u>Page</u>
2b. Is the same image data file as Fig. 2a, by using the pivot table feature, we were able to link the image data to a set of points derived from a GPS field boundary cropping the edge of the image data, limiting the view to only those points that fall within the field boundary.	23
2c. Yield data displayed in a shape file as 2 m gridded data, classified to 10 categories, and cropped to those points that fall only within the field boundary.	24
3a. Kriged 2 m yield data for wheat as a surface plot, using the pivot table and Excels surface-plotting feature. The x and y coordinates were grouped to make a 10 m cell size. The darkest color is the highest yielding area (Purple color is the highest yield category).	25
3b. The “snapped” to a 2 m grid yield data for wheat as a surface plot, using the pivot table and Excels surface-plotting feature. The x and y coordinates were grouped to make a 10 m cell size. The darkest color is the highest yielding area (Purple color is the highest yield category).	26
4a. A surface plot created from soil pH data for the wheat field. Point data was assigned to soil-sampling polygons. Regularly spaced (2 m interval) points within soil-sampling polygons were evaluated with an Excel pivot table. The x and y coordinates were grouped to make a 20 m cell size.	27
4b. A surface plot of soil pH data from the wheat field. It is the same point data that was processed and displayed in Fig. 4a but for this process Golden Software’s Surfer for Windows was used instead of Excel. The data was kriged to 20m spacing and then displayed as a contour map.	28

LIST OF FIGURES CONTINUED

<u>Figure</u>	<u>Page</u>
5a. A map of the wheat field that displays the areas of the field that had a pH that was lower than 5.8 (the dark areas) or that had a pH that was greater than 5.8 (the light colored areas).	29
5b. A map that displays the areas of the wheat field that had a yield of less than 3298 kg ha ⁻¹ [(3000 lbs ac ⁻¹) the light colored area] or the areas of the field that were higher than 3298 kg ha ⁻¹ [(3000 lbs ac ⁻¹) the darker colored areas].	30
5c. A map that displays the areas of the wheat field that have both a higher yield [yield greater than 3298 kg ha ⁻¹ (3000 lbs ac ⁻¹)] and a lower pH (pH less than 5.8). This intersection of the data sets is represented by the areas in the map that are the darker colored areas.	31
5d. A surface plot of soil buffer pH (SMP) data from wheat. Point data was assigned to sampling polygons. Regularly spaced (2 m interval) points within sampling polygons were evaluated with an Excel pivot table. The x and y coordinates were grouped to make a 20 m cell size. The SMP pH values were converted to lime rates using the formula (lime rate in tons/acre=(6.2-SMP)*0.5).	32
6a. Snapped yield plotted against kriged yield for the wheat field.	33
6b. Snapped yield plotted against kriged yield for the meadowfoam field.	34
6c. Snapped yield plotted against kriged yield for the hazelnut orchard.	34

LIST OF FIGURES CONTINUED

<u>Figure</u>	<u>Page</u>
7. The average cross sectional area CSA by soil-sampling polygon is plotted against the average yield for the same soil-sampling polygon. The upper boundary line (red square) is marked by a line that runs through the set of points that is derived from the maximum yield in each of ten CSA classes.	38
8a. The graph of yield plotted against image darkness for the entire wheat field. The upper yield boundary line is marked by the regression line that runs through the set of points that is derived from the maximum yield across ten image darkness categories.	39
8b. The graph of yield plotted against image darkness for the entire meadowfoam field. The upper boundary line is marked by the curve that runs through the set of points that is derived from the maximum yield in 11 image darkness categories.	40

Using Advanced Spreadsheet Features for Agricultural GIS Applications

Key index words

Spreadsheets, geostatistics, boundary condition, hazelnuts, wheat, meadowfoam, soil pH

Abstract

A GIS analysis procedure was developed to explore relationships between imagery, yield data, soil information, and other assessments of a field or orchard. A set of conversion utilities, a spreadsheet, and an inexpensive shape file viewer were used to manipulate, plot, and display data. Specific features described include procedures used to: 1) display automated yield monitoring and aerial imagery data as surface maps for visual analysis, 2) generate maps from gridded soil sampling schemes that display either the collected soil data values or management information derived from further manipulation of the sample values, 3) evaluate relationships among data layers such as yield monitor, imagery, and soil data, 4) conduct an upper boundary line evaluation of potential yield-limiting factors. The analysis process is demonstrated on wheat, meadowfoam, and hazelnut data, from crops grown in Oregon.

Introduction

Our intent was to develop spatial analysis procedures to be used as tools to evaluate relationships between soil data, yield data, or other plant growth measures in both field and orchard crop systems. We wanted to explore and develop an inexpensive approach to data analysis that would have broad application to most spatial data and information. The analysis software needed to be widely available with an open and compatible format. Since many computers come with spreadsheet software and most users are familiar with them, spreadsheets became a logical choice. In a previous paper, we described preliminary development efforts on an inexpensive shape file viewer and spreadsheet data visualization and analysis system (Righetti and Halbleib, 2000). This paper presents advanced features not previously described, and illustrates their application with data from three crops (wheat, meadowfoam, and hazelnuts) grown in Oregon.

Our system uses a software utilities that converts spatial data files into a form that can be utilized in a spreadsheet. Once data conversion has been completed, one can view and query the information with a simple shape file viewer and Microsoft® Excel® (Microsoft Corp., Redmond, WA). The key to our package is to convert all spatial information (data associated with polygons, soil test and yield data, images, and other geographic information) into regularly spaced data points or grids.

The software utility previously described (Righetti and Halbleib, 2000) that perform these processes are as follows:

1. A program that takes ASCII (text) files consisting of spatial coordinates (points) and other associated information and converts them to a shape file. The shape file contains the original data, a positioncode field (x.y or Easting.Northing), and empty fields for future calculations and analysis for each grid point.
2. A program that takes any shape file polygon(s) and creates another shape file that consists of a set of regularly spaced points that falls within the boundary of the polygon(s) with the grid spacing output (distance between points) selected by the user. The data originally assigned to the polygon(s) are preserved for each point and additional database fields are added. One of the additional database fields is called positioncode (x.y or Easting.Northing).
3. A program that takes a geo-referenced TIFF image file (UTM coordinates, Universal Transverse Mercator) and creates a shape file that consists of a set of grid points with the output grid spacing set by the user. The RGB image (Red image band, Green image band, Blue image band) digital values for each band are preserved as a point database file with additional fields being added that consist of an Easting/Northing field (position code), extra blank

fields and calculated file fields which are preserved for each point location in the shape file database.

4. A program that takes ASCII (text) files of spatial information (grid soil samples, yield monitor data, etc.) krigs the appropriate values and places the output into a shape file. This utility also has an option that creates gridded shape files that are calculated from point density rather than a kriging routine.

Additional file conversion and data processing procedures have been developed.

Some of these procedures can already be accomplished using the tools in a spreadsheet but we wanted to automate the process for less experienced users.

5. "Snap to a grid" is a feature where data points that have spatial coordinates (data must be projected in UTM coordinate system) can be rounded to the interval of choice and moved to the closest corresponding, regularly spaced grid point.
6. "Point data to polygons" is a utility that converts an ASCII text file that defines the vertices of a polygon (s) creating a shape file that can be viewed with most GIS software. An option in this procedure allows one to generate a series of rectangular shape files from ASCII text files (point data). The

original point coordinates define the centroid of the rectangular polygon.

The user specifies the size of rectangles output (in meters from the centroid). All attributes from the point data are associated with the new polygons.

7. A data export feature that converts Microsoft Excel summary tables (pivot tables used to develop and display surface charts) to a shape file format of points consisting of the x and y coordinates and a third value (z value) representing the surface feature described.
8. A conversion utility that transforms data from Latitude and Longitude to UTM projections and vice versa.

Most of the utilities and file conversion features described are available with an easy to use Windows interface. When these procedures are combined with the analytical features of Microsoft® Excel®, most routine precision agriculture analysis tasks (yield display, soil factors, and variable rate dosage maps; data visualization; statistical correlations; and other analyses) can be accomplished. Boundary line analyses to define the upper limits of crop production or plant growth (Webb, 1972, Lark, 1997, Schung et. Al., 1996) can also be conducted. Spreadsheet tools can be used to display information and explore relationships between the data in linked files in a way that is analogous to the manipulation of

layered data in GIS software programs. We feel that our process may be easier to use than many of the GIS software offerings on the market today, and in some cases Excel's advanced analysis tools support evaluation procedures that cannot be accomplished with some of the more expensive software.

Materials and Methods

Data collected from the three crops, included information on yield data, soil pH, buffer pH (SMP), imagery, and other measured crop responses. The specific data utilized were different for each site, but the spreadsheet-based analysis procedures were similar. Computer CD's with the raw data, conversion utilities, spreadsheet files, and a tutorial are available from the authors.

Data Collection

Wheat

Yield data were collected September 1997 using a flow-based yield monitor (RDS Technology Ltd., Sewell, NJ) on spring wheat (*Triticum aestivum*) planted in the Willamette Valley, Oregon. The Trimble (Trimble Corp., Sunnyvale, CA; <http://www.trimble.com/trimble.htm>) Ag 132 real-time, differentially corrected, GPS unit and yield monitor were used to record position as a latitude and longitude, while simultaneously recording the yield estimate data from the combine at five-second intervals.

Soil samples were collected in the fall after harvest. The 17.25 ha field was “gridded” (128 samples) using a cell size of 1349 m² (one third of an acre per sampling area). The center of each sampling cell location was recorded with a differentially corrected GPS receiver. Within each of these cells, 12 cores were collected to a depth of 30 cm (twelve inches) and combined to form a representative sample. The soil samples were oven dried, ground, and analyzed for pH (5:1, deionized water:soil) and soil SMP buffer pH.

Early season three color RGB aerial imagery of the field was acquired from the WAC Corporation (Eugene, OR; <http://www.waccorp.com/>). The WAC Corporation routinely collects aerial photography of the Willamette Valley and

provides imagery for the public. The image was scanned on a flatbed scanner, digitized and stored as a TIFF file. The image was geo-referenced using targets that had been placed in the field and located with a GPS receiver. The resolution in the digital image was set at one pixel per meter of ground resolution and a registration file (*.tfw) was created.

Preliminary evaluations of the relationships between the image and yield data revealed that various expressions of pixel intensity (red band, green band, blue band, composite [$\{\text{red} + \text{green} + \text{blue}\} / 3 = \text{composite}\}$) and image indices produced similar evaluation results. Ratios among different bands of the imagery (red, green, and blue bands) or the index $\{(r-g)/(r+g)\}$ did not improve the relationship between imagery and yield (data not shown). In the interest of brevity, only one evaluation is presented here. The original composite brightness values (pixel values from 0-255) were first inverted ($0 = 255$; $255 = 0$) to create a scale where increasing values represent increasing darkness.

Meadowfoam

Yield data were collected in August 1997 with a flow-based yield monitor on meadowfoam (*Limnanthes alba*) grown in the Willamette Valley of Oregon, using the same procedures as described above (wheat).

Soil samples were collected in the fall just after harvest. The 23.5 ha field was “gridded” (174 samples) using a cell size of 1349 m² (one third of an acre per sampling area). Soil samples were processed, and analyzed as described above (wheat).

Imagery of the meadowfoam field was acquired from Emerge Inc. (Greeley, CO), a company that captures digital images for general land use and analysis. They provided false color NIR (near infra-red) imagery at one meter pixel ground resolution, in a TIFF file with an accompanying reference file that provides the coordinates for geo-referencing. The soil was bare with no cover growing at the time of image and soil data collection.

Hazelnuts

Yield data were collected in September of 1999 in a 15-year-old hazelnut (*Corylus avellana* L.) orchard near Albany, Oregon, using a weight-based yield monitor. The yield monitor consisted of load cells linked to a computer that recorded on a 5 second interval both accumulating weight and position as the nuts were swept from the ground and transferred into a storage bin on a trailer.

Soil data were collected in August before harvest. The 4.3 ha field was “gridded” (32 samples) using a cell size of approximately 1011 m² (one quarter of an acre per

sampling area). Samples were collected, processed, and analyzed as described above (wheat).

Two caliper measurements of the trunk were taken at a height 30-cm above the ground in both a North-South direction and an East-West direction. The average of the two measurements was used to calculate an estimate of the tree trunk cross sectional area (CSA).

Data Visualization and Analysis

Overview of procedures

To explore data relationships among the different layers requires that all files have a spatial component assigned to the data and the data must be converted to gridded point files. Once the files are converted to grids, individual points are set to the same interval (user defined grid spacing) and assigned position codes. This can be accomplished by either a geo-statistical approach such as kriging (Mulla, 1991; Isaaks and Srivastava, 1987) or a snapping to a grid approach (Kitchen et al., 1999). We recognize that moving data points could introduce some positional inaccuracy, however, if the areas evaluated and the scale of variation in the field are much larger than the grid spacing, the positional errors introduced by moving a point to the closest grid node should be minimal in its impact. In an effort to compare the kriging and snapping to a grid approach, we visually compared maps prepared from 2m spaced yield data derived from both procedures. Linear regressions between data produced by both approaches were also conducted.

Spatial data that is uniformly arranged can be easily evaluated in a spreadsheet. For example, relationships between yield monitor data and imagery can be easily examined by kriging or snapping both data sets to the same grid spacing. However,

it is not appropriate to statistically evaluate kriged or otherwise manipulated data that has a grid spacing much less than the original data (Isaaks and Srivastava, 1989). To investigate relationships between soil grid samples or other spatially sparse data sets relative to the spatially dense yield monitor or imagery data, the following procedure was devised. In Fig. 1a, soil sampling polygons (viewed as a shape file) that represent soil-sampling locations at the wheat site are presented in a shape file viewer. Procedure #6 was used to convert the original soil point data to the polygons displayed. These soil-sampling polygons have been placed on top of another layer that represents kriged (2m grid interval) wheat yield. This makes it easy to visualize where the soil-sampling units are spatially positioned relative to the wheat yield map.

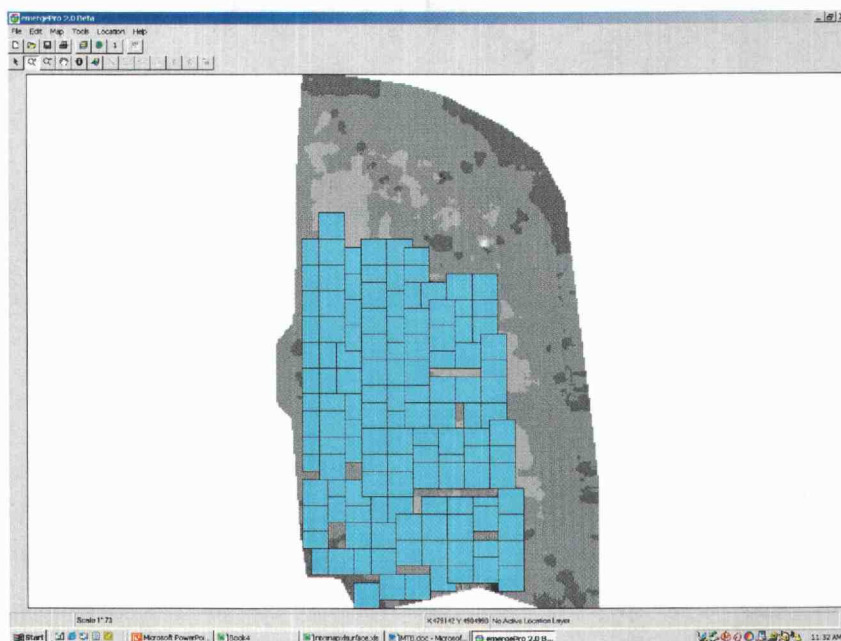


Figure 1a. Kriged yield data, as a surface map, depicting the wheat field yields as contours. The soil-sampling regions are overlaid as blue polygons that are representative of the area from which each soil sample was taken.

We have used EmergePro® (Emerge Inc., Greeley, CO) as the shape file viewer in this example but other inexpensive or free products (such as Arc Explorer; ESRI, Redlands, CA), can also be used.

Figure 1b shows a closer view (zoomed and enlarged) of a small area of the field displayed in Fig 1a, revealing a portion of one of the soil-sampling polygons. The kriged yield points now appear as small squares on 2 m spacing. The series of points (small dots in Fig. 1b) on a regular spaced 2 m grid that were derived from the original soil-sampling polygon shape file are also displayed. These points fall

within the boundaries of the soil-sampling polygon. Each point that falls within the soil polygon is assigned the same informational data (attributes) that were assigned to the original soil polygon from which it was derived. Therefore, each point within the boundary of the original soil-sampling polygon has the same pH as the entire polygon. Each derived point was also assigned the sample number associated with the original polygon (such as number one for the first sampling area, number two for the second sampling area, etc.).

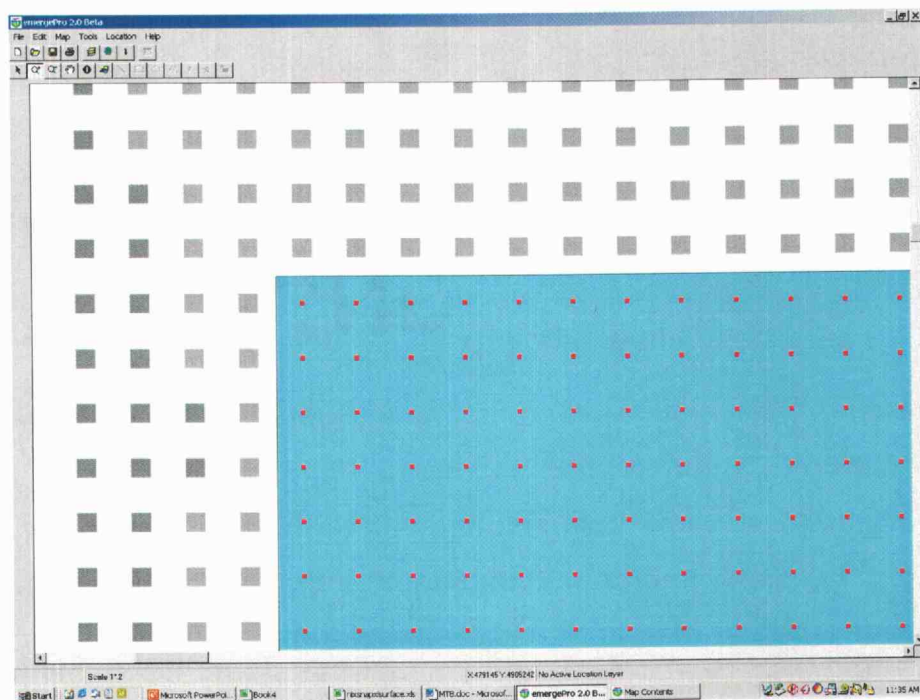


Figure 1b. An enlarged view showing the corner of one soil-sampling polygon. Within the polygon are red points that are set to a regular 2 m grid interval; these points are assigned the same attributes as the soil-sampling polygon and have the same spacing as the yield data points.

In Fig. 1c, points for the soil-sampling polygon (small points) are shown with image data (intermediate sized squares) and the yield data (large sized squares) overlaid as they appear when displayed in a shape file viewer. The position code of each data point becomes the common identifying feature between layers of information (data sets) in Excel. Excel's pivot table feature uses the position code numbers to link different data layers. When the position code numbers match between data layers, a join can be performed that either includes, excludes, or selects only matching points from each data layer being compared. All the associated information for individual data layers (image data, yield data, soil data, etc.) can be linked together for each unique position code identifier. Once the linking task is accomplished, it is relatively simple to use the charting or summary capabilities in Excel to plot or summarize data.

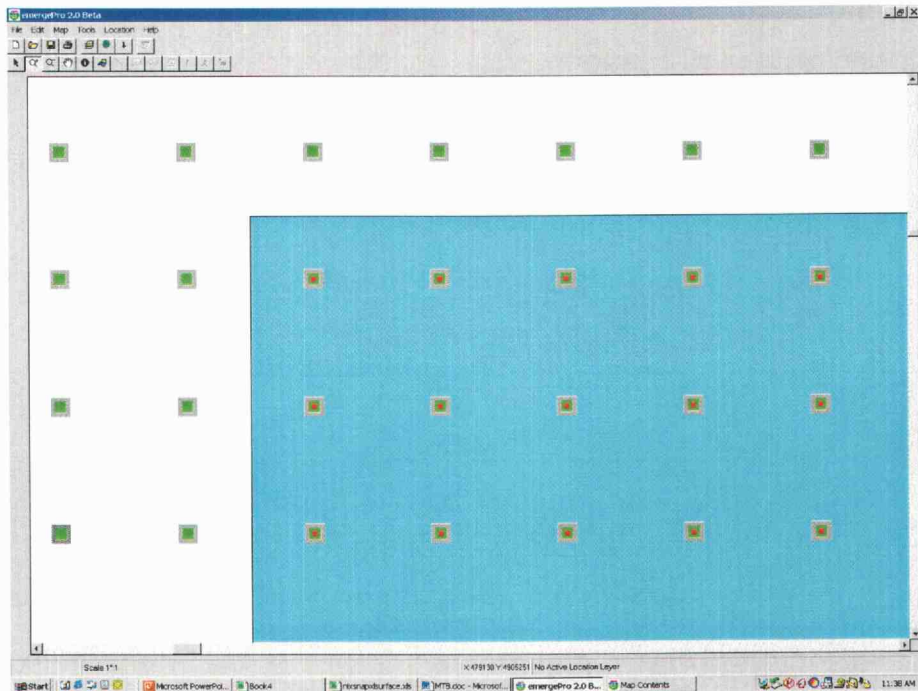


Figure 1c. An enlarged close up of the same wheat field that is depicted in Fig 1a and b. showing the spatial relationship that exists between data layers by displaying all of the data layers as overlapping different sized points. The largest points that are shown are underneath as the bottom layer and are representative of the yield data, with the medium sized points representing the image data, and the smallest points that are on top represent the soils data. This helps illustrate the spatial relationship that exists between the data layers and shows how the information, when set to a uniform grid spacing, can be managed using the location as the common link.

Pivot tables

Excel's pivot table feature can be used to combine, condense, and evaluate data files. As described above, the first analysis step is to link data files of interest (layers) using the joining capability of Excel's pivot table feature.

Very large datasets (millions of points) can be linked . Once the data are joined, Excel's pivot table can be used to query, summarize, group, categorize, and display data. For example, the pivot table can be used to extract data enabling one to evaluate the pH of soil-sampling polygons and the average yield or image data associated with the same polygons as shown in Fig. 1d. At this point in the process, the extracted data can easily be statistically evaluated.

The pivot table in Excel also lets the user summarize spatial information by creating rows and columns. Summarizing the data in this way allows the user to select the appropriate UTM grid coordinates as a row (x or Easting number) and column (y or Northing number), with a third value being displayed in the grid as the z value. Once the data are organized in this fashion, Excel can be used to plot the information as a surface map. Yield and imagery data can be manipulated in a pivot table and directly displayed as a surface map. Data in rows and columns can be grouped to define the cell size desired (reduce the resolution) and then replotted. It is also possible to change the number of contour levels displayed.

Another GIS task that the pivot table can be used to accomplish is to interpolation between values. Soil data associated with sample area polygons (such as the smallest points in Fig. 1c) can also be evaluated with a pivot table and presented as a surface map. This procedure works best when the polygons created from the original sampling points slightly overlap.

Soil sampling polygon	pH	Average yield within soil sampling polygon
1	5.6	3491.618421
2	5.9	3979.828947
3	5.9	3983.746753
4	5.8	4256.870968
5	6.3	3616.825
6	6	3754.058442
7	5.9	4414.206107
8	5.7	3457.913793
9	5.7	3886.034188
10	5.8	3122.788889
11	5.8	2716.390244
12	5.7	2602.860656
13	5.7	3243.449612
14	5.6	4074.272
15	6	3523.606383
16	5.7	4077.322222
17	5.5	3942.597826
18	6.2	3375.392857
20	5.7	4249.659574
21	5.7	4085.031579
22	5.8	4142.084906
23	5.6	3407.018349
24	5.3	3258.076923
25	5.6	3263.79375
26	5.7	3684.955556
27	5.7	4129.369369
28	5.7	4187.607041

Figure 1d. A page of data results that were generated from a query using the Excel pivot table features.

Surface plots for soil data produced in Excel appear to be similar to the contour maps produced in many geostatistical and precision agriculture software packages. To compare maps produced in Excel to maps produced with more sophisticated mapping programs, we produced kriged maps from the original soil grid sample data with Surfer for Windows (Golden Software, Greeley CO). We visually compared maps prepared in Excel from 2m soil polygon data with the kriged maps produced in Surfer.

Identification of boundary line conditions

The data summarized using the pivot table analysis can be further evaluated with a second pivot table analysis applied to the first pivot table results. This manipulation allows the user to select the highest yield from the set of average yield values for each level of an independent factor. For example, maximum yield values for different levels of CSA or image darkness can be obtained. Excel can then be used to generate a regression equation that fits the maximum points. If the number of observations in an independent variable category is insufficient, it is possible to reduce the number of grouped categories to more reliably detect a set of maximum values. The maximum value for each category can be plotted against the median value for the category interval to generate a response curve. This application works well for identification of points on the upper edge of the data set when trying to identify an upper boundary.

The curves that are shown later in this report were produced using this procedure. We chose not to use a more sophisticated approach to identifying boundary lines (Evanylo and Sumner, 1987; Heym and Schnug, 1995; Walworth et al., 1986; Webb, 1972). However, more advanced approaches could be incorporated into a spreadsheet environment.

All data presented in this paper were manipulated and displayed using only our data conversion utilities, the Excel procedures, and a shape file viewer.

Results and Discussions

Displaying automated yield monitoring and imagery data that are cropped to a field boundary

In Fig. 2a, evenly spaced 2 m points derived from the image data for the wheat field are shown as displayed in the shape file viewer. The utilities and procedures described above were used to create data files from the digital image (*.tif format). In this display the data points that represent the image values are classified into 10 categories that reveal increasing image darkness, respectively.

A GPS-derived field boundary was converted to a set of points on a 2m interval, that fall only within the field boundary. This set of points can be joined in Excel with other data sets such as the image or yield data in this example. This process is used to limit the data points to only those yield or image points lying within the field borders. In figure 2b, the original image derived file has been cropped.

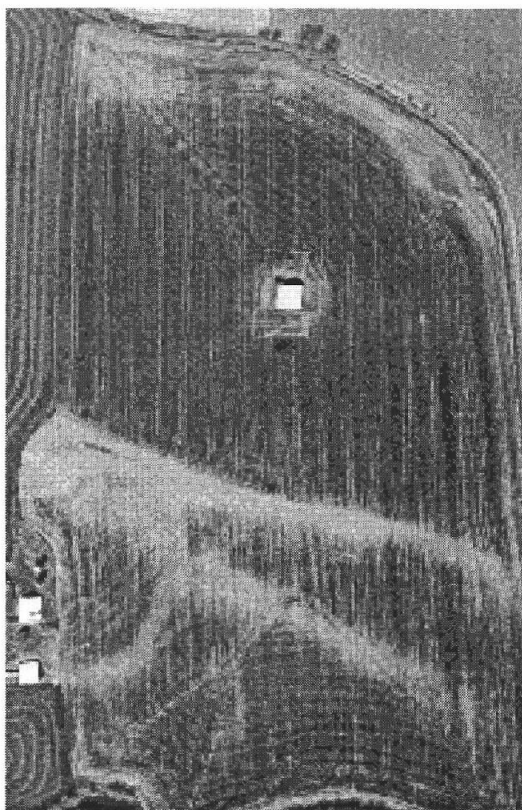


Figure 2a. Image data converted to a set of points on a 2 m interval with the composite band digital values inverted ($0=255$ and $255=0$) so that the values are representative of increasing darkness. It is displayed as a shape file and classified into 10 categories.

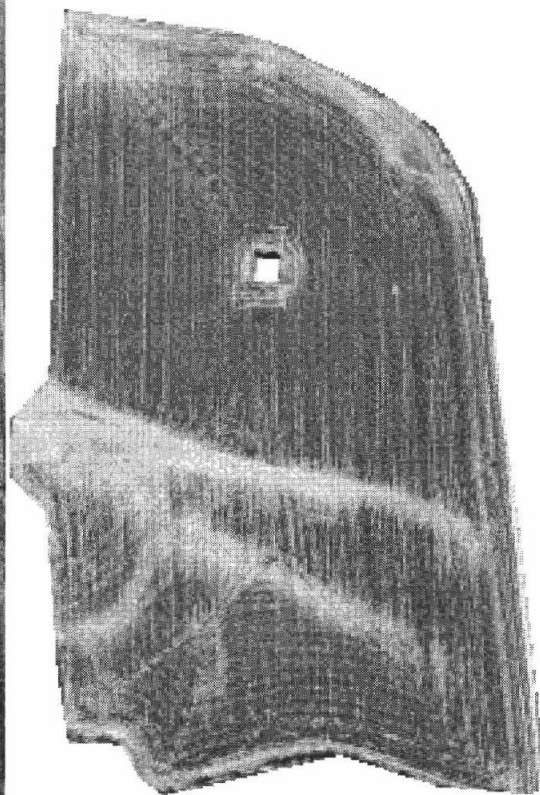


Figure 2b. Is the same image data file as Fig. 2a, by using the pivot table feature, we were able to link the image data to a set of points derived from a GPS field boundary cropping the edge of the image data, limiting the view to only those points that fall within the field boundary.

In Fig. 2c, the kriged 2 m yield data is presented. The yield data is divided into ten categories with increasing darkness associated with increased yield. Although the kriging routine produces a rectangular set of points (data not shown), the kriged

data displayed here has also been cropped to the actual field boundary using Excel's joining capability.

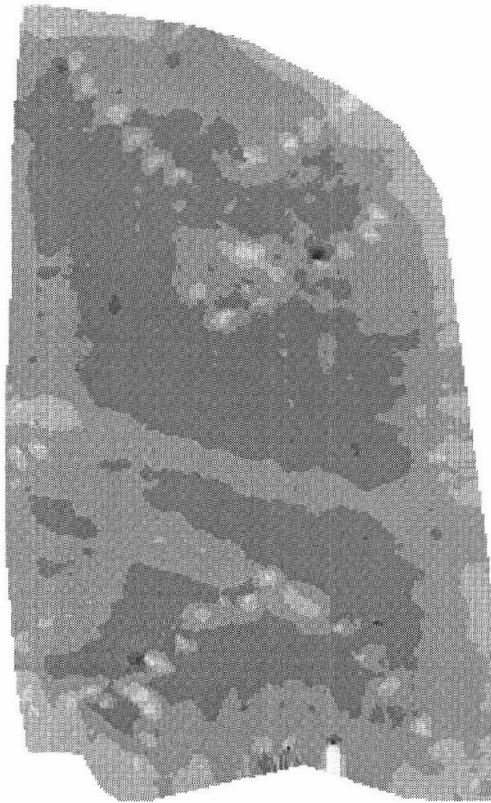


Figure 2c. Yield data displayed in a shape file as 2 m gridded data, classified to 10 categories, and cropped to those points that fall only within the field boundary.

Evenly spaced data points can also be displayed as a surface plot in Excel. This manipulation may eliminate some of the need for a shape file viewer. In Fig. 3 Excel's surface plot feature was used to present two-meter kriged (Fig. 3a) and

snapped (Fig. 3b) yield for the wheat field. In the display shown, all points were grouped to a ten-meter interval and then averaged using the pivot table analysis. In this example, yields were classified into five categories. Both methods produce similar visual maps. Both procedures also produced similar maps for meadowfoam and hazelnuts (data not shown).

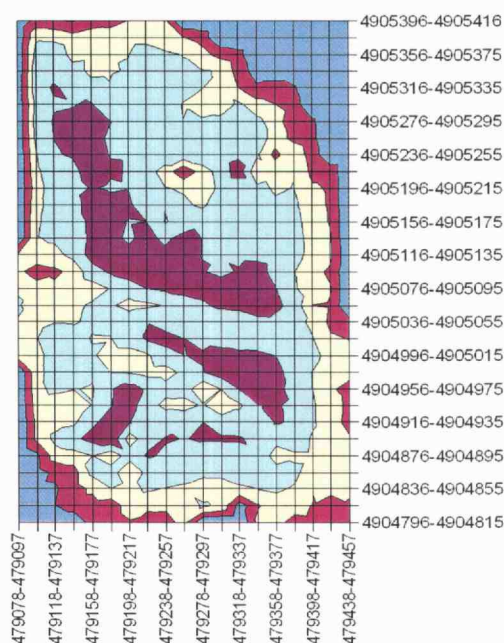


Figure 3a. Kriged 2 m yield data for wheat as a surface plot, using the pivot table and Excels surface-plotting feature. The x and y coordinates were grouped to make a 10 m cell size. The darkest color is the highest yielding area (Purple color is the highest yield category).

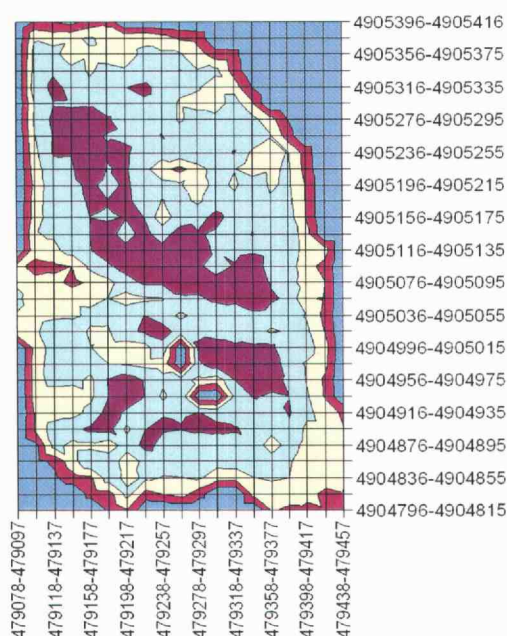


Figure 3b. The “snapped” to a 2 m grid yield data for wheat as a surface plot, using the pivot table and Excel’s surface-plotting feature. The x and y coordinates were grouped to make a 10 m cell size. The darkest color is the highest yielding area (Purple color is the highest yield category).

Generating maps from soil grid sampling schemes.

In Fig. 4a, Excel’s surface plot feature was used to map the 2 m pH data from the soil polygons for the wheat field. Excel’s grouping capability was used to define a grid size of 20 m. The data was placed into four categories ranging from a pH of 5.1 to a pH of 6.3. The dark blue area of the map represents locations that were outside the field boundary and were not sampled.

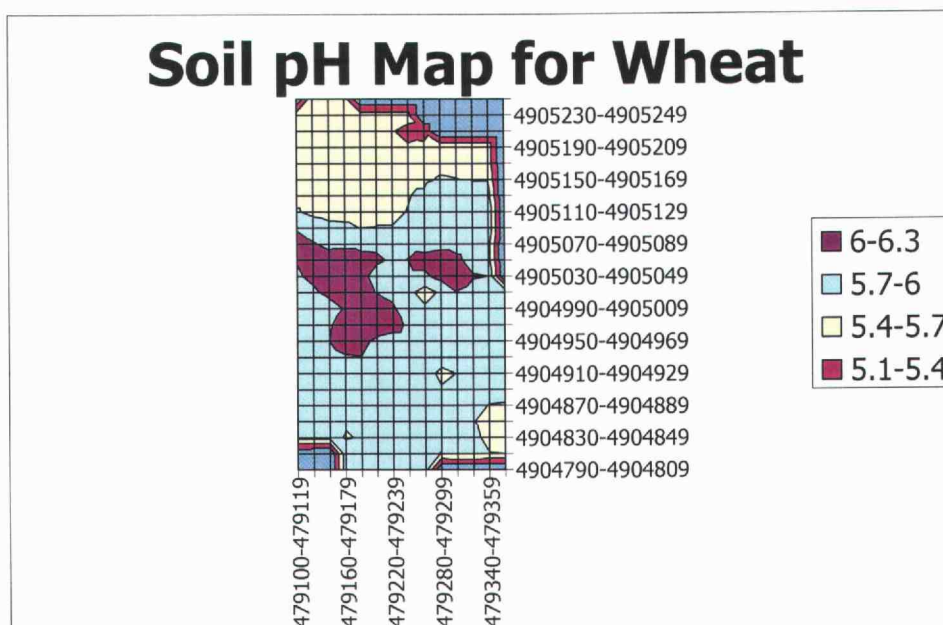


Figure 4a. A surface plot created from soil pH data for the wheat field. Point data was assigned to soil-sampling polygons. Regularly spaced (2 m interval) points within soil-sampling polygons were evaluated with an Excel pivot table. The x and y coordinates were grouped to make a 20 m cell size.

The map in Fig. 4a is very similar to a map created from a kriging analysis with a more complex geostatistical (Golden Software, Surfer for Windows) program that is shown in Fig. 4b. The kriged map was created by using the original point data and setting the predicted interval to 20 m. The kriged surface data was then placed into the same four categories described above.

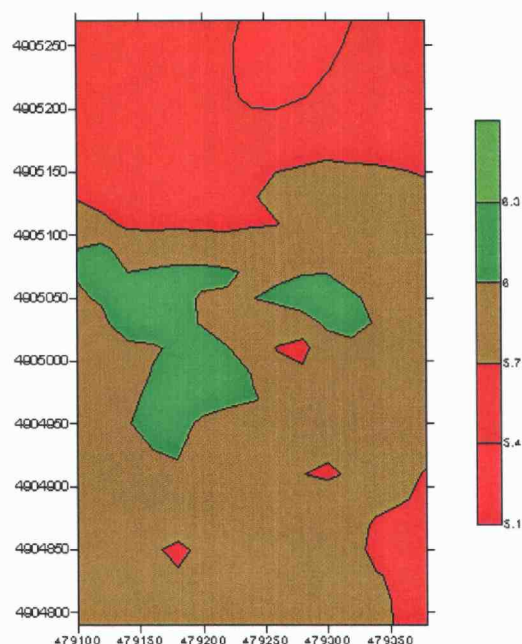


Figure 4b. A surface plot of soil pH data from the wheat field. It is the same point data that was processed and displayed in Fig. 4a but for this process Golden Software's Surfer for Windows was used instead of Excel. The data was kriged to 20m spacing and then displayed as a contour map.

Similar pH maps were obtained when either the contour mapping capabilities in Excel or when Surfer's kriging analysis was applied to the pH data at the meadowfoam or hazelnut sites (data not shown). This data suggests that a simple pivot table data averaging method produces similar maps to approaches that use more sophisticated geostatistical procedures. This may allow processing of data in a simple spreadsheet. If more sophisticated approaches are desired the geostatistical procedures can be completed with an additional software utility that utilizes a more advanced process while still allowing for visualization in Excel.

Surface plots can be further combined and manipulated in Excel. For example in Fig. 5a, pH data is presented as a surface map that shows areas of the field that had pH values that were in the upper 50% of the pH range (light color) and the areas that were in the lower half of the pH range (dark color).

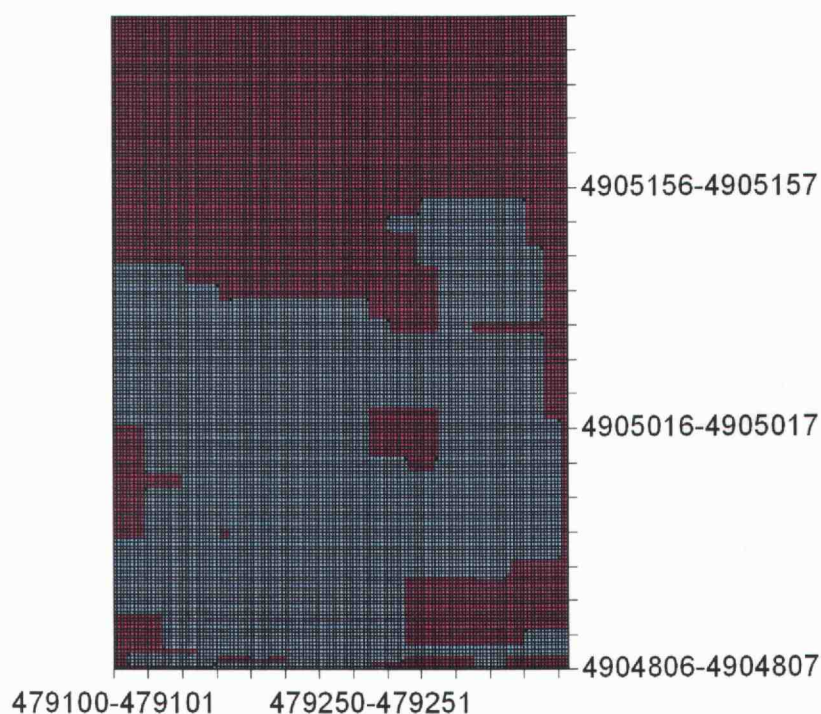


Figure 5a. A map of the wheat field that displays the areas of the field that had a pH that was lower than 5.8 (the dark areas) or that had a pH that was greater than 5.8 (the light colored areas).

In Fig. 5b, the wheat yield data from the same area of the field that overlaps with the pH testing areas is displayed. This map delineates the areas of the field that are

either in the upper 50% of the yield values (dark color) or the lower 50% of the yield values (light color).

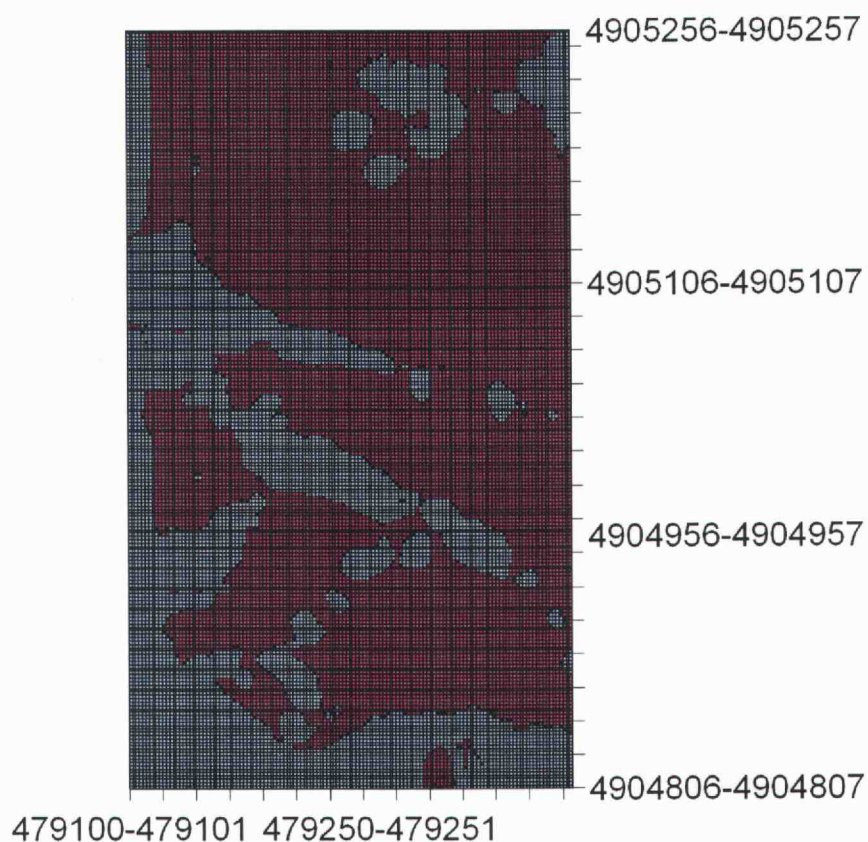


Figure 5b. A map that displays the areas of the wheat field that had a yield of less than 3298 kg ha^{-1} [$(3000 \text{ lbs ac}^{-1})$ the light colored area] or the areas of the field that were higher than 3298 kg ha^{-1} [$(3000 \text{ lbs ac}^{-1})$ the darker colored areas].

By combining the data from the two maps (Figs. 5a and 5b) we can produced a map Fig. 5c that shows the areas that are in the upper half of the yield values and also fall within the areas of the field that are in the lower pH range. This map displayed in Fig. 5c delineates those areas that have both a low pH and a higher yield (dark

color). This map may represent the areas of the field that will most likely respond to lime additions and could be used as the basis for a variable rate lime application.

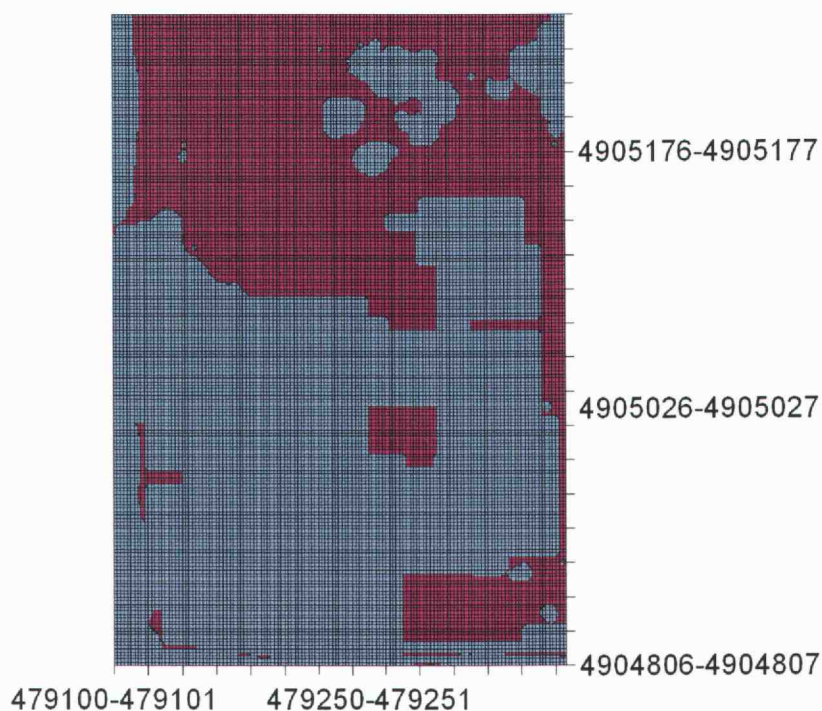


Figure 5c. A map that displays the areas of the wheat field that have both a higher yield [yield greater than 3298 kg ha^{-1} (3000 lbs ac^{-1})] and a lower pH (pH less than 5.8). This intersection of the data sets is represented by the areas in the map that are the darker colored areas.

Dosage maps where application rates are determined from many approaches can also be created and exported as an ASCII file (x, y, application rate) that is compatible with many variable rate controllers. An example is the lime dosage map presented in Fig 5d. Standard Excel procedures can be used to combine spatial data or convert spatial information into suggested application rates.

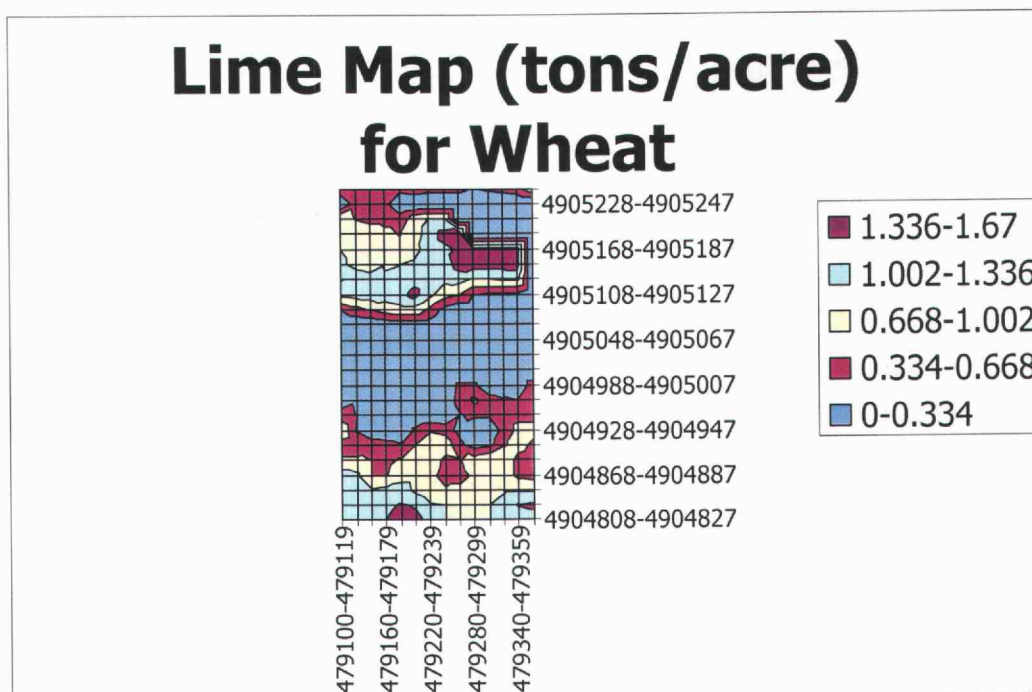


Figure 5d. A surface plot of soil buffer pH (SMP) data from wheat. Point data was assigned to sampling polygons. Regularly spaced (2 m interval) points within sampling polygons were evaluated with an Excel pivot table. The x and y coordinates were grouped to make a 20 m cell size. The SMP pH values were converted to lime rates using the formula (lime rate in tons/acre= $(6.2 - \text{SMP}) \times 0.5$).

In the example shown, SMP buffer pH values were converted to lime requirement using a procedure in Excel {lime rate in tons/acre= $(6.2 - \text{SMP}) \times 5$ } with all negative values treated as zero. This corresponds to an application rate of 0.5 tons lime per acre for every 0.1 buffer pH unit below 6.2. The blue area corresponds to areas

where no lime is required. It is interesting that some of the areas identified as having both low pH and high yield (Fig. 5d) were not identified as areas needing lime with the SMP soil test. This demonstrates the value of combining multiple approaches when developing dosage maps. Creating dosage maps likely requires consideration of many variables, coupled with thought and judgment, and may not be an easy process to simplify.

Evaluating relationships for spatial data

A sample of spatial data evaluation is presented in Fig. 6. The average values for 2 m kriged yield data is plotted against 2 m snapped yield data for Figs. 6a, b, and c, for wheat, meadowfoam, and hazelnuts, respectively.

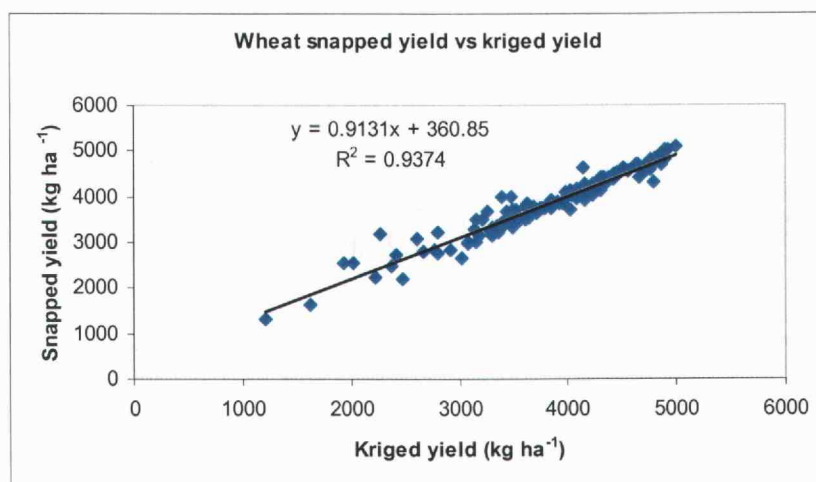


Figure 6a. Snapped yield plotted against kriged yield for the wheat field.

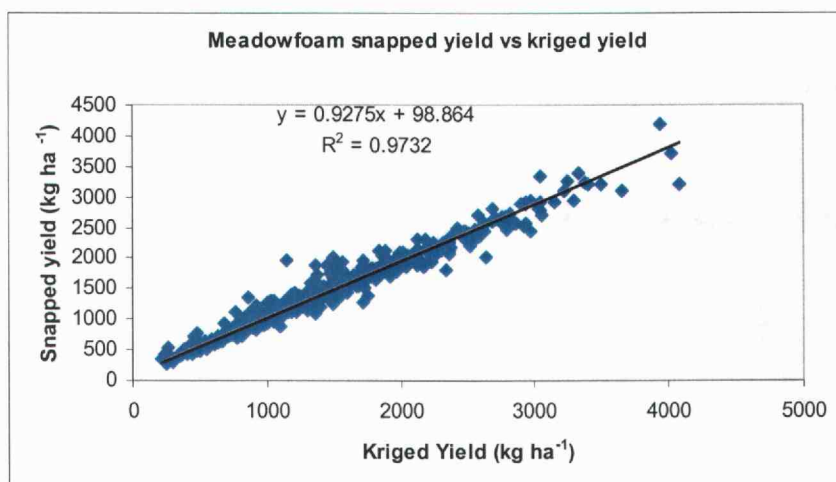


Figure 6b. Snapped yield plotted against kriged yield for the meadowfoam field.

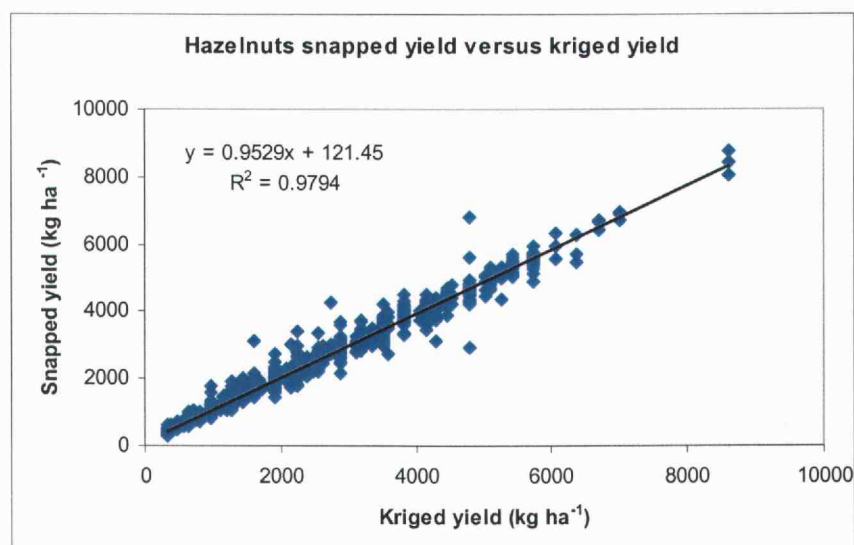


Figure 6c. Snapped yield plotted against kriged yield for the hazelnut orchard.

Data points shown were derived by averaging all yield values within each soil sample polygon. If the 10 m kriged yield data is plotted against the 10 m snapped data for the entire field or orchard, results are similar (data not shown). A strong linear relationship between the two approaches was expressed with r^2 values of .9374, .9732, and .9794 for wheat, meadowfoam, and hazelnuts, respectively. The data suggest that snapping to the nearest grid point instead of kriging is a viable method of analysis. The snapping procedure requires less sophisticated software and allows for the processing of data in a simple spreadsheet. The use of snapped data has the added advantage of not raising the concerns associated with a statistical analysis of kriged results. Automated contouring to a regular grid can create artifacts, thus data in contour maps are helpful in qualitative displays, but may be of questionable quantitative significance (Isaaks and Srivastava, 1989). In all analyses that follow, where yield was used as a dependent variable, the snap to a grid procedure was applied. Although not shown, an evaluation of kriged data produced similar results.

Examples of linear regression coefficients for relationships among yield monitor, imagery, and soil data at the three sites is shown in Table 1.

Table 1. Correlation coefficients for between spatial variables in wheat, meadowfoam and hazelnuts. Evaluations for Yield vs. Image Darkness and Yield vs. CSA were conducted on both the entire field or orchard and a smaller portion of the site that was evaluated for soil properties.

Crop	<u>Yield vs. pH</u>	<u>pH vs. Image Darkness^x</u>	<u>Yield vs. Image Darkness</u>	
			<u>Entire Field</u>	<u>Study Sites</u>
Wheat	0.0503	0.4138**	0.016	0.233**
Meadowfoam	0.0151	0.0405	0.0217	0.0289
	<u>Yield vs. pH</u>	<u>pH vs. CSA^y</u>	<u>Yield vs. CSA</u>	
			<u>Entire Field</u>	<u>Study Sites</u>
Hazelnuts	0.1081	0.2263	0.6324** ^z	0.861 **

^x Image darkness is defined as $255 - \{(r+g+b)/3\}$ where r, g, and b are the pixel values for red, green and blue bands in a digitized photograph.

^y CSA refers to the trunk cross sectional area 30 cm above the ground.

^z ** indicates statistical significance at $p < 0.05$.

Although imagery data is sometimes associated with yield (Merrill et. al., 1993), the relationship between image darkness (2m) and snapped yield (2m) within the 128 and 174 soil sample units respectively, was poor for both wheat and meadowfoam (r values = 0.233 and 0.029 respectively). Similar results were obtained whether data for the entire field or data just within soil sampling polygons were evaluated. The human eye may perceive similar patterns in surface plots derived from imagery and yield information such as the presented in Figs. 2b and 2c however; a strong statistical relationship does not exist between the imagery and the yield data.

Cross sectional area (CSA) of tree trunks has been consistently used to define maximum yield potential in tree fruits (Westwood, 1978). Our data verify this general observation between CSA and yield for hazelnuts when either the sample site or entire orchard data is used (Table 1). Relationships between yield and soil pH were weak for all three crops. Relationships between soil pH and either image darkness (wheat and meadowfoam) or CSA (hazelnuts) were also weak.

Boundary line evaluations

Boundary line analyses can be made for data based on soil sampling areas or from an entire field or orchard. An example of evaluations for individual sampling areas is shown for the relationship between hazelnut yield and CSA in Fig. 7.

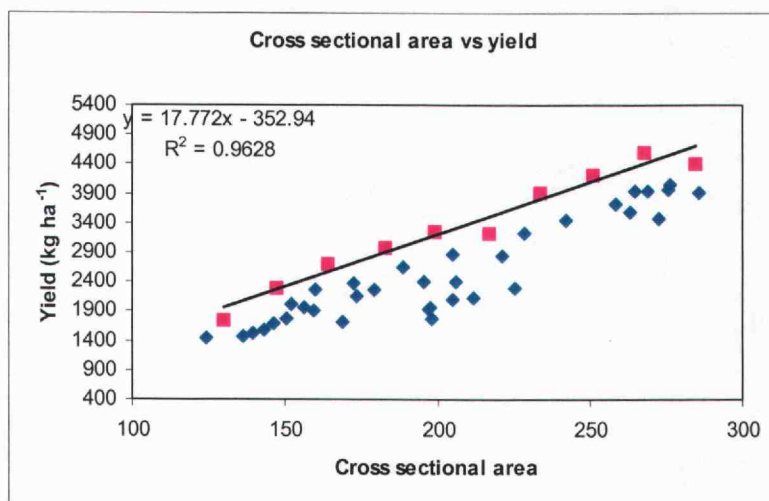


Figure 7. The average cross sectional area (CSA) by soil-sampling polygon is plotted against the average yield for the same soil-sampling polygon. The upper boundary line is marked by a line (red squares) that runs through the set of points that is derived from the maximum yield in each of ten CSA classes.

A clear boundary condition is apparent. Although many other factors can limit yield, the maximum yield obtainable for a given CSA is defined by the response curve. It is possible that the entire curve could be shifted upward or downward in high or low production years. However, the maximum relative return is probably limited by the average CSA in the individual sampling units. Many points in Fig. 7 are on or near the plotted boundary line. Locations represented by these points are performing at or near their maximum potential. Improving short-term performance at these locations is unlikely since points above the apparent boundary lines are rare and CSA would likely respond slowly to management changes. Emphasizing corrective management treatments for the areas represented by the points below the boundary line have the highest probability of success. The

concept that CSA can be used to define yield potential is well verified (Westwood, 1978). Linking spatial management practices to boundary conditions apparent in yield vs. CSA scatterplots has not been emphasized.

The relationship between CSA and yield in Fig. 7 could have been detected with conventional statistics (Table 1). In Fig. 8 the relationship between yield and image darkness is shown for wheat (Fig. 8a) and meadowfoam (Fig. 8b).

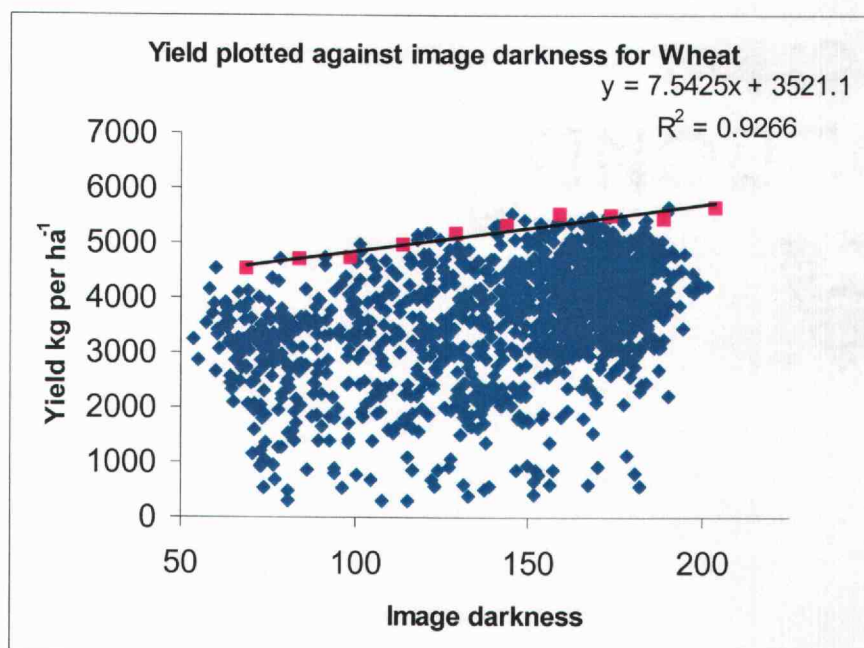


Figure 8a. The graph of yield plotted against image darkness for the entire wheat field. The upper yield boundary line is marked by the regression line that runs through the set of points that is derived from the maximum yield across ten image darkness categories.

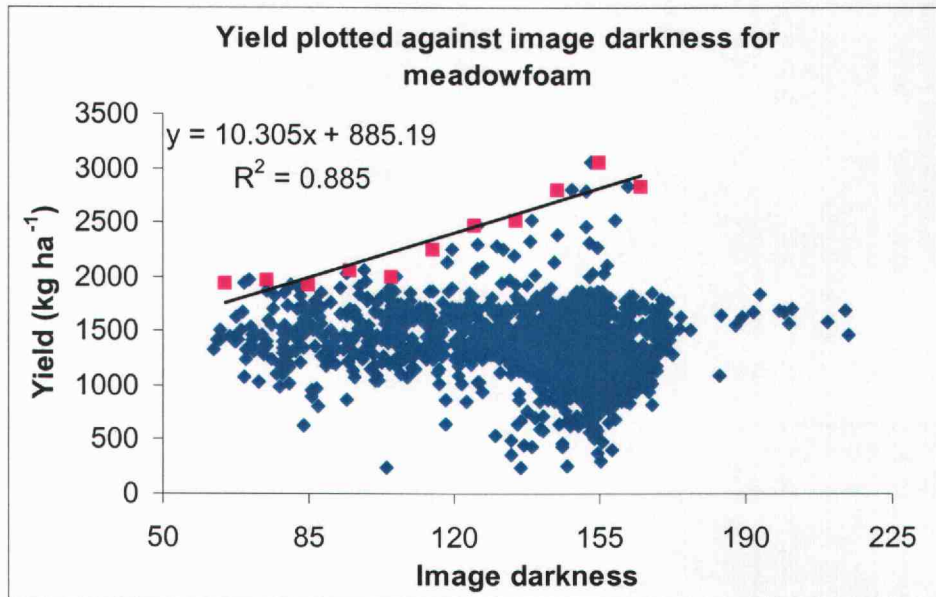


Figure 8b. The graph of yield plotted against image darkness for the entire meadowfoam field. The upper boundary line is marked by the curve that runs through the set of points that is derived from the maximum yield in 11 image darkness categories.

These data were calculated for entire fields by snapping both yield and image data to a 10 m interval. In these examples, only very weak statistical relationships between the two factors were seen with conventional statistics (Table 1). For the wheat example (Fig. 8a) a clear boundary condition is apparent. Although many factors can limit yield, the maximum yield obtainable for a given darkness level is defined by the response curve. The maximum yield obtainable at low darkness values is 20% less than what is obtainable at high darkness levels.

Similar results were obtained when evaluating the relationship between yield and image darkness for meadowfoam (Fig. 8b). Although the boundary line for meadowfoam has a steeper slope than for wheat, the data should be interpreted with caution since the high yielding points that correspond to high image darkness are relatively rare. Nonetheless, these boundary lines may describe potential spatial limitations on yield potential. Image darkness likely relates to early vegetative growth (Gerard and Buerkert, 1999) in the wheat and moisture levels (Senay et. al., 1998) in the bare ground image. Moisture holding capacity is likely dependent on soil texture and organic matter. Imagery may be helpful in identifying areas where maximum yield is least likely to occur.

Spatial differences in yield potential may be important factors to consider when designing precision agriculture prescriptions (Frazier et. al., 1997). Although some large trees yield poorly, CSA can be used to spatially define yield potential for the hazelnut orchard. Cross sectional area (CSA) is both easy to measure and gives an integrated estimate of tree response to soil and environmental conditions over time. Although the data are less convincing for imagery than CSA, image darkness can likely be used to spatially define yield potential for the wheat and meadowfoam sites.

Although not discussed in this paper, it may be possible to use a spreadsheet-based boundary condition approach to assess if soil factors are associated with yield or yield potential.

Conclusion

A spreadsheet and a shape file viewer can be effectively used to manipulate and view spatial data. Maps can be generated from automated yield monitoring, aerial images, and soil grid sample data. These data can be further manipulated and exported to other software and applications. It is also possible to easily evaluate relationships among yield monitor, imagery, and soil data. Spreadsheets are amenable to boundary condition evaluations, and when applied to large data sets with a spatial component, boundary conditions may define yield potential on a field-by-field basis.

As data collection becomes more automated, information is becoming more readily available. The range of data that is being collected is becoming broader, encompassing more areas, is being collected more frequently, and often the data sets are larger than ever before. With automated data being generated, we can now automatically collect yield measurements, imagery, gridded soils information, water and irrigation practices, and the list is growing with new forms of information on the horizon and more automation likely to emerge. However, these technical advances can be a double-edged sword. As farm managers are required to deal with more, new, and larger amounts of information, the demands on management continue to increase, requiring greater technical expertise and skills.

Currently the average grower has access to more information than he or she has the skill to utilize. The value of spatial data is obvious but the lack of skilled individuals and the available software to perform analysis tasks may be the Achilles' heel of precision agriculture. The ability to perform complicated spatial analysis using a spreadsheet instead of relying on personnel trained in the use of GIS may be attractive to farm and production managers. With the mastery of spreadsheet tools required for spatial analysis, many other data analysis tasks become possible. This dual benefit may not be apparent if one just develops proficiency with GIS software.

References

- Bhatti A.U., and Frazier B.E. 1991. Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and Thematic Mapper images. *Remote sensing of the environment*. 37(3):181-191.
- Evanylo G.K., and M.E. Sumner. 1987. Utilization of the boundary line approach in the development of soil nutrient norms for soybean production. *Commun. in Soil Sci. Plant Anal.*, 18(12):1379-1401.
- Franzen, D.W., Reitmeier, J.F. Giles, and A.C. Cattanach. 1999. Aerial photography and satellite imagery to detect deep soil nitrogen levels in potato and sugarbeet. p. 281-290. In: P.C. Robert, R.H. Rust, and W.E. Larson (ed.), *Proc. Fourth Intl. Conf. Precision Agriculture*. Amer. Soc. of Agronomy; Crop Sci. Soc. of Amer.; Soil Sci. Soc. Amer.; Madison, Wis.
- Frazier B.E., C.S. Walters, and E.M. Perry. 1997. Role of remote sensing in site-specific management. ASA-CSSA-SSSA. *Precision agriculture 1997: papers presented at the first European Conference on Precision Agriculture*, Warwick University Conference Center, UK. vol 1, p. 149-160.
- Gerard, B., and A. Buerkert. 1999. Aerial photography to determine fertilizer effects on pearl millet and *Guiera senegalensis* growth. *Plant and Soil*. 210:167-199.
- Heym, J., and E. Schnug. 1995. A mathematical procedure for the development of boundary lines from XY scattered data. p. 137-142. In: *Field experiment techniques: 11-13 December 1995 Churchill College, Cambridge*. Wellesbourne, Warwick, UK.
- Isaaks, E. H., and R. M. Srivastava. 1989. *Applied geostatistics*. Oxford University Press, Oxford, New York.
- Kitchen, N.R., K.A. Sudduth, and S.T. Drummond. 1999. Soil electrical conductivity as a crop productivity measure for claypan soils. *J. Prod. Agric.* 12:607-617.
- Lark, R. M. 1997. An empirical method for describing the joint effects of environmental and other variables on crop yield. *Ann. Appl. Biol.* 131:141-159.

- Merrill, E.H., M.K. Bramble-Brodahl, R.W. Marrs, and M.S. Boyce. 1993. Estimation of green and herbaceous phytomass from Landsat MSS data in Yellowstone National Park. *Journal of Range Management*. 46, no. 2, p. 151-157.
- Mulla, D. J. 1991. Using geostatistics and GIS to manage spatial patterns in soil fertility. p. 336-345. In: G. A. Kranzler (ed.), *Automated Agriculture for the 21st Century*. Amer. Soc. Agr.; St. Joseph, MI.
- Mulla, D. J., and A. U. Bhatti. 1997. An evaluation of indicator properties affecting spatial patterns in N and P requirements for winter wheat yield. p.145-153. In: J.V. Stafford (ed.), *Precision Agriculture '97*. Vol.1 BIOS Sci. Publ. Oxford, UK OX4 1RE.
- Righetti, T.L., and M.D. Halbleib. 2000. Pursuing precision horticulture with the internet and a spreadsheet. *Hort. Tech.* 10(3):458-467.
- Schnug, E., J. Heym, and F. Achwan. 1996. Establishing critical values for soil and plant analysis by means of the boundary line development system (Bolides). *Commun. Soil Sci. Plant Anal.* 27(13 -14):2739-2748.
- Senay G.B., A.D. Ward, J.G. Lyon, N.R. Fausey, and S.E. Nokes. 1998. Manipulation of high spatial resolution aircraft remote sensing data for use in site-specific farming. *Transactions of the ASAE*. (St. Joseph, Mich., American Society of Agricultural Engineers). 41(2)489-485.
- Thomas, D.L., C.D. Perry, G. Vellidies, J.S. Durrance, L.J. Kutz, C.K. Kvien, B. Boydell, and T.K. Hamrita. 1999. Development and implementation of a load cell yield monitor for peanut. *Appl. Eng. Agric.* 15(3):211-216.
- Walworth, J.L., W.S. Letzsch, and M.E. Sumner. 1986. Use of boundary lines in establishing diagnostic norms. *Soil Sci. Soc. Am. J.* 50:123-128.
- Webb, R.A. 1972. Use of the boundary line in the analysis of biological data. *J. Hort. Sci.* 47:309-319.
- Westwood, M.N. 1978. *Temperate zone pomology*. W.H. Freeman and Co. San Francisco, California.
- Whitney, J.D., W.M. Miller, T.A. Wheaton, M. Salyani, and J.K. Schueller. 1999. Precision farming applications in Florida citrus. *Appl. Eng. Agric.* 15(5):399-403.