

10 Collocating Multiple Self-Generated Data Layers

Viacheslav I. Adamchuk and Chenguang Wang

CONTENTS

10.1 Executive Summary.....	185
10.2 Introduction	186
10.3 Methods	187
10.4 Results	191
10.5 Conclusions	195
10.6 Supplemental Files	195
References.....	197

10.1 EXECUTIVE SUMMARY

Through original attempts to adopt precision agriculture management strategies, it has been observed that a single self-generated data layer does not provide sufficient information with regard to a production field. Therefore, an analysis of multiple data layers is important.¹ Based on geographic coordinates, collocated point measurements that belong to different data layers can be grouped to investigate relationships between alternative point estimates (e.g., crop yield versus a particular soil property). Unfortunately, this task cannot be accomplished using every geographic information system (GIS) software package, and frequently is inaccessible to low-level users.

As described in this chapter, DM_Comp software was developed as a stand-alone application suitable for performing primitive comparisons of several self-generated point data layers. Input data must be available in the form of delimited text files with geographic coordinates represented in decimal degrees. The output files will follow the same format and will be placed in the same directory as the input files. Further analysis of the output data can be performed using either a spreadsheet or an actual GIS software package.

The main function of the DM_Comp includes collocating multiple self-generated data layers in (1) points that belong to one of these layers (using user-defined fixed radius or nearest neighbor averaging) and (2) centers of rectangular grid cells (using grid cell averaging). In addition, a simple statistical filter has been included to remove

outlier points from yield data files available in AgLeader® (AgLeader Technology, Inc., Ames, Iowa) advanced text export format. A primitive graphic interface was added to observe the integrity of different data layers in terms of field coverage.

The supplemental example includes three years of soybean yield, laboratory analysis of soil sample results, a soil electrical conductivity map, and a map produced using a soil mechanical resistance on-the-go sensor. The software presented is an example of primitive stand-alone applications that can aid spatial data analysis when specific straight-forward processing is required and/or the access to high-level GIS software packages is not available.

10.2 INTRODUCTION

The majority of data layers generated through precision agriculture practices represents a one-dimensional array of qualitative values such as yield parameters or soil properties georeferenced to a common system of geographic coordinates (most likely WGS-84). In many instances, useful information pertaining to relationships between various data layers can be observed only if these data sets are related based on common coordinates. However, the process of defining these relationships is not trivial.

In the modern GIS environment, the most frequent comparison between different field data layers is done through generating, and then analyzing, surfaces (raster maps) representing spatial structure in a given sparse data set. For example, each data set (yield, on-the-go sensor, or soil laboratory analysis records) can be filtered and interpolated to the extent of the field boundaries using a common interpolation procedure (e.g., inverse distance weighting, kriging, etc.). Then, collocated cells that belong to several interpolated surfaces (with values predicted based on each data layer) can be included in a consistent numerical and/or logical computation. Therefore, quantitative determination of the relationship (correlation) between different data layers can be accomplished.

Unfortunately, many of the procedures involved are not available in the low-level GIS software commonly available in production agriculture. Also, the inappropriate use of many existing surface interpolation methods can lead to significant errors due to low mapping density in combination with weak spatial structure. Therefore, there is a need for a stand-alone program (or module) to help combine various precision agriculture data sets into a single file with a simple data structure, which could be further analyzed using a commonly available spreadsheet and/or low-level GIS software packages.

The objective of this chapter is to report on an example of such a software application. The required functions were to (1) read and recognize various text-delimited data files with WGS-84 longitude and latitude columns (e.g., soil laboratory reports, on-the-go soil sensor, or yield data), (2) filter unprocessed yield files (AgLeader® advanced text export format) based on statistical limits for individual sensor outputs and overall yield estimates, (3) compare multiple data sets to a single sparse data set (typically with the smallest density) using the simplest procedures (averaging of nearest neighbor values and/or averaging of data points within a fixed radius area), (4) graphically display the data extent and generate rectangular grids

with full coverage of data domain (adjust size to avoid split cells), (5) determine average data values within each grid cell, and (6) output the combined data sets as a text-delimited file.

The described program, DM_Comp.exe, was developed using a C++ code to accommodate all of these requirements. The current version does not contain help and user guide options, and it may malfunction since not every combination of input parameters or corrupt input files has been tested. The main purpose of this software was to illustrate the functionality that could be pursued by future developers.

10.3 METHODS

DM_Comp (Data Management – Comparison) is a stand-alone program that can be used to combine various common sources of spatial data such as yield, soil laboratory analysis, topography, electrical conductivity or other on-the-go soil sensor measurements. This software includes two main utilities: (1) process yield and other spatial data files to determine values corresponding to a specific sparse data set (such as laboratory soil analysis) and (2) develop a rectangular grid pattern and determine corresponding average attributes from every input data file corresponding to each grid cell. DM_Comp should not be viewed as an alternative to GIS software in any way. Rather, it serves as an effective complementary program used to conduct multilayer analyses of spatial data. Such analyses can be performed in a spreadsheet or statistical software environment using delimited output text files, and may include, for example, regression analysis between different data layers (i.e., yield versus soil nutrients level or soil electrical conductivity versus selected soil properties) or development of crop response equations.

Every input data layer must be available as a delimited text file with two columns corresponding to geographic position (latitude and longitude) in decimal degrees with respect to WGS-84 ~~data~~ (commonly used by Global Positioning System). The existence of a header row is optional. The type of data delimiter (comma or tab) is detected automatically.

There is no interpolation process involved. All data manipulations are accomplished solely by matching geographic coordinates from various data layers. Simple arithmetic averaging is done when more than one data point satisfies the requirements when searching across different layers. Unavailable matching data points result in incomplete rows within the output file that may be excluded from further statistical analysis.

Flexibility and simplicity are the biggest advantages of this program, which takes only 2.16 MB of storage space and requires no installation. To start the program, a user should simply double click on the program icon that appears in Windows Explorer or through the Run option of the Windows Start menu.

The main window of DM_Comp (Figure 10.1) appears right after the program is started. This window allows a user to (1) select a sparse data file and define the corresponding yield and other data files to be matched, (2) set the power for yield data filtering and a method for the multilayer data search, and (3) execute the data collocating routine and switch to the Graph View window mode. The most common Windows control options (e.g., closing the program, selecting multiple files, select-

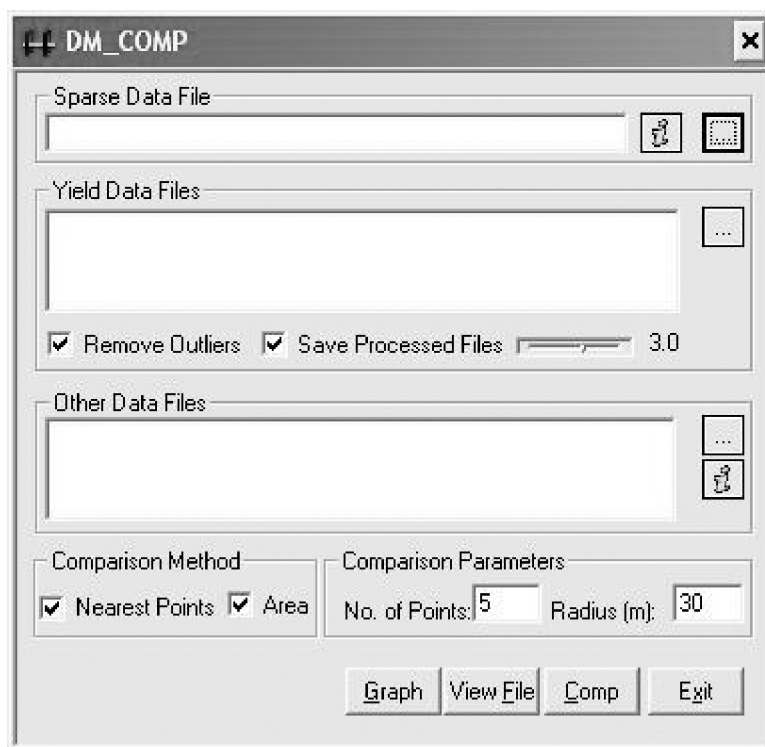


FIGURE 10.1 DM_Comp main window.

ing/deselecting data columns, etc.) have been included to provide an intuitive user interface. Information about DM_Comp can be viewed when the program icon in the upper left corner is pressed and scrolled down to the About Comp menu option.

A **Sparse Data File** is any text-delimited spatial data file containing geographic longitude and latitude columns, in decimal degrees, followed by other attributes described in the header line, which ~~are~~ recommended but not necessary. This file is the basis for the output, and defines the number of rows (data points) for which corresponding values from other data layers are found.

The button symbolized by three dots opens the file selection dialog box (Figure 10.2). Names of files with .txt extension will appear first. If an input file has a different extension, the “all files” extension type should be used. Each button symbolized with an “i” opens the file information dialog box (Figure 10.3), which will also appear after making a sparse data file selection. In this box, the user can set the coordinate columns (if different than longitude and latitude) and select or unselect columns to be included or excluded from the output file. The output file will be written to the same directory as the original sparse data file. The original name will also be kept with “_comp” added before the extension in the output file name.

A **Yield Data File** is a raw file obtained from an AgLeader yield monitor using the **advanced text export** option. If processed, ~~text-delimited~~ yield data files should be considered as “other data files” (described below). Similarly, the button with three

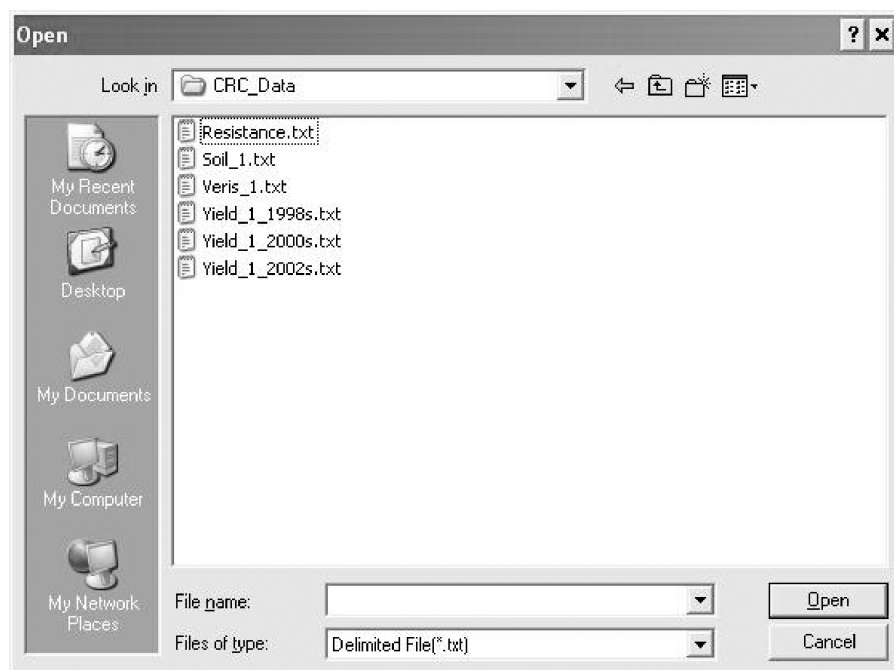


FIGURE 10.2 Open file dialog box.

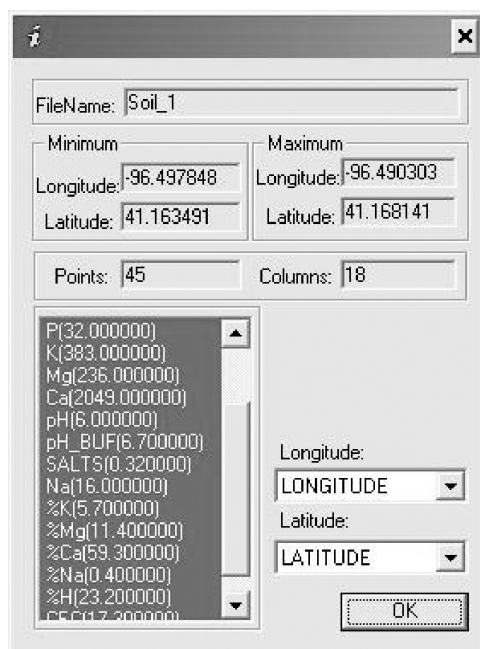


FIGURE 10.3 Data file information window.

dots should be used to open the dialog box for selecting the appropriate yield files (multiple selections can be made by holding the shift key down). The **Remove Outlier** box should be checked to enable a built-in filter that removes outlying data points from each yield data file. When the **Save Processed File** box is checked, yield files with outliers removed will be saved. The original yield file names will be used with “_out” added before the extension. Otherwise, the yield files with outliers removed will be erased when DM_Comp is closed. The horizontal sliding bar can be used to set the power of the built-in filter. The number indicated represents the multiplier for standard deviations to consider the values of individual sensor measurements or the calculated yield being erroneous. It ranges from 0 (almost all data points removed) to 5 (the most conservative filter). The default filter power is set at 3. Yield, moisture, and elevation values calculated from each input yield file are written to the output file. Blanks are used to substitute unavailable or corrupted values.

An **Other Data File** is any ~~text-delimited~~ file with longitude and latitude columns. Those usually include processed yield files, soil electrical conductivity, and other on-the-go sensor recordings. The function of the file selection and the information buttons is the same as described above, except that the files must be selected one at a time. Multiple columns in each file are allowed, and their order will be preserved when preparing the output file.

Combining Parameters contains two main settings of this program. The user has the ability to define a rule to distinguish collocated values. The values can be defined either as a certain number of nearest neighbors (5 by default), as all points located within a predefined area around individual sparse data points (within 30-m radius by default), or both combined. If neither box is checked, DM_Comp will search for the single nearest point. All distance calculations were conducted according to Adamchuk.²

The **View File** button can be used to open a text view window to review all input and generated output files (if available) in the text mode. The X should be pressed to return to the main window. The **Comp** button starts combining (matching) all specified data files. After the work is completed, the output file is saved and a log message box appears (Figure 10.4). It provides information about program performance such as the number of points left without a qualified match. The **Exit** button or **Alt+F4** should be pressed to quit the program.

The **Graph** button opens a **Graphic View** window (Figure 10.5). Points from each file (sparse, yield, and other data files) can be viewed on the screen if selected. When ID is selected, a small square will appear in the middle of each grid cell (default 4 × 4). The user is able to specify both easting (X) and northing (Y) grid dimensions, or indicate the number of grid cells in each direction. The **Square Grid** option can also be used to ensure similar grid cell dimensions in both directions. Enabling the **Show Grid** option will display the lines of grid cells calculated for a given field using an optimized boundary rule (adjust grid size to avoid partial cells within data domain). If the **Save** button is pressed, the request to specify an output file name will appear. The output file will contain an ID column with the corresponding longitude and latitude for the center of each grid cell followed by the corresponding average values from all highlighted data files (all columns previously

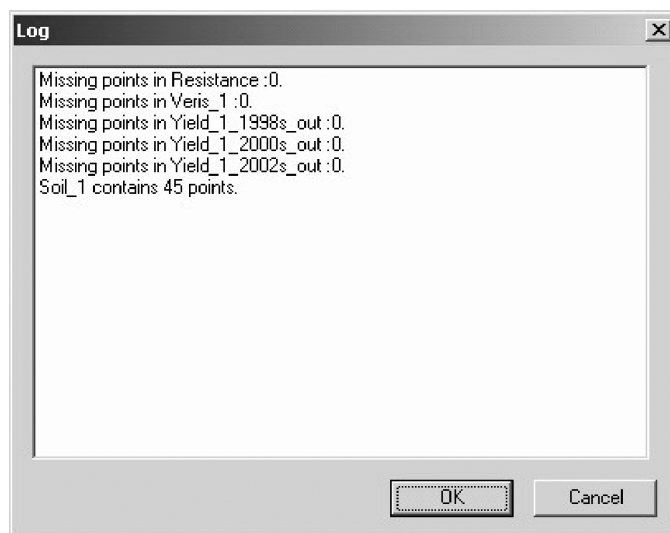


FIGURE 10.4 Spatial data matching log report.

selected). If at least one selected data layer does not contain points within the specific grid cells, these cells are excluded from the output file. Therefore, data layers with different density and/or coverage should be saved separately. This option can be used to generate a center-point grid soil sampling map. Either OK or the X button can be used to return to the main window of the DM_Comp program.

10.4 RESULTS

Enclosed with the program is a typical data set consisting of three years of soybean yield, georeferenced analytical soil lab results, as well as soil electrical conductivity and mechanical resistance data. The following steps will guide the user through an example of completed data processing using the software.

1. Start the program by clicking the **DM_Comp** icon in Windows Explorer (the program will run better if copied to the hard disk first).
2. Press the **Open File** button in the **Sparse Data File** area and navigate to the *Soil_1.txt* file. This file contains a laboratory report from a soil analysis performed on 45 samples (2.5-acre grid sampling). Assuming that we do not want to include laboratory results on percent base saturation in the output, unselect the %K, %Mg, %Ca, %Na, and %H fields. Click OK to return to the main window.
3. Press the **Open File** button in the **Yield Data File** area and select *Yield_1998.s.txt*, *Yield_2000.s.txt*, and *Yield_2002.s.txt* yield files (press and hold the shift key to select multiple files). These are soybean yield files from the 1998, 2000, and 2002 growing seasons. Click OK to return to the main window.

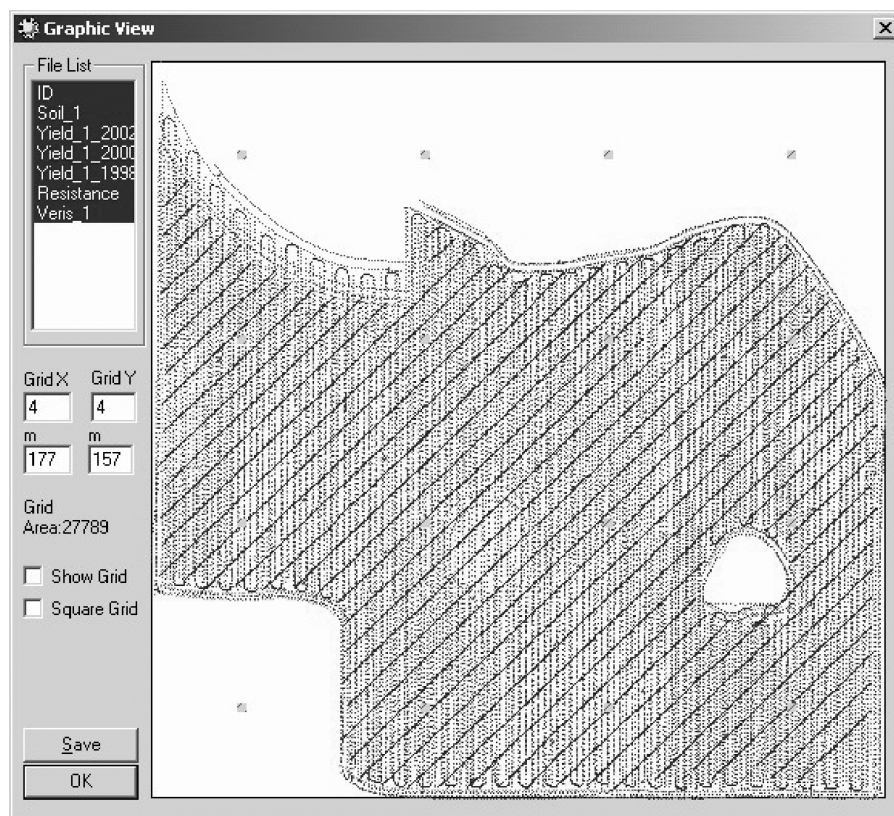


FIGURE 5 DM_Comp graphical view window.

4. Use the default filter settings. However, you can also change the power of the filter by dragging (or clicking) the scroll bar to the left or right.
5. Press the **Open File** button in the **Other Data Files** area and select the *Veris_1.txt* file. This is the output from the Veris® 3100 (Veris Technologies, Inc., Salina, Kansas) electrical conductivity mapping system. After pressing OK, you can select columns of interest (SHALLOW and DEEP). These data represent two depths of electrical conductivity measurements. Press OK.
6. Press the **Open File** button in the **Other Data Files** area again and select the *Resistance_1.txt* file. This is the output from an experimental soil mechanical resistance mapping tool.³ After pressing OK, select three depth layers representing mechanical resistance (in units of pressure MPa) at 0–4-in, 4–8-in, and 8–12-in depth intervals. Press OK.
7. Set the comparing method to all the points within a 30-m radius (only the **Area** checkbox should be on). Press **Comp** to perform the calculations. This may take a few minutes.

8. Observe the log message and press OK. The output file named *Soil_1_comp.txt* will be written to the drive. It will include columns from the *Soil_1.txt* file as well as the matched values from all other files.
9. Click the **View File** button and review the input and output text files (use the pull-down menu at the top). Press X in the upper right corner to close the **File View** window.
10. Press the **Graph** button to open the **Graphic View** window. Examine the map by selecting and unselecting different fields in the upper left selection window. When done, select all fields except *Soil_1* (to avoid data with density much smaller than the size grid cells to be defined). Click **Show Grid** to see the boundaries of the default (4 × 4) grid. Enter 40 m to define the easting grid interval (X). To make the grid square, check the **Square Grid** option. In this case, both dimensions will be automatically adjusted and each grid cell will represent an approximate 0.4-acre area.
11. Click **Save** to record the output file with coordinates for the center of each grid cell and averages of all previously selected data for each cell. The user must specify a descriptive file name (for example: *0.4_acre_grid.txt*). Click OK to return to the main window and exit the program.

The two output files can be processed further using common software packages. The resulting *Soil_1_comp.txt* can be used to determine correlations between specific soil properties and each additional data layer (Table 10.1), and to further study specific relationships. For example, one could explore dependency between (1) soil pH and soybean yield, (2) cation exchange capacity and soil electrical conductivity, or (3) soil organic matter content and soil mechanical resistance (Figure 10.6). Yield

TABLE 10.1
Pearson coefficients of correlation summary

Soil property	Soybean Yield			Electrical Conductivity		Soil Mechanical Resistance		
	2002	2000	1998	0–30 cm	0–90 cm	0–10 cm	10–20 cm	20–30 cm
OM	–0.02	0.03	–0.02	0.11	0.16	–0.18	–0.01	0.21
P	–0.48	–0.45	–0.23	–0.43	–0.46	0.25	–0.29	–0.38
K	–0.07	–0.18	0.10	–0.02	–0.01	–0.12	–0.37	–0.25
Mg	0.37	0.27	0.16	0.68	0.59	–0.14	0.42	0.45
Ca	0.15	0.08	0.03	0.39	0.28	0.03	0.24	0.26
Soil pH	0.00	–0.04	0.01	0.11	0.05	0.07	0.07	0.04
Buffer pH	0.15	0.18	–0.06	0.29	0.29	0.09	0.20	0.26
Salts	0.21	0.30	0.29	0.42	0.37	–0.18	0.14	0.28
Na	0.02	0.08	–0.29	–0.01	0.00	–0.21	0.07	0.22
CEC	0.23	0.11	0.05	0.45	0.35	–0.09	0.26	0.26

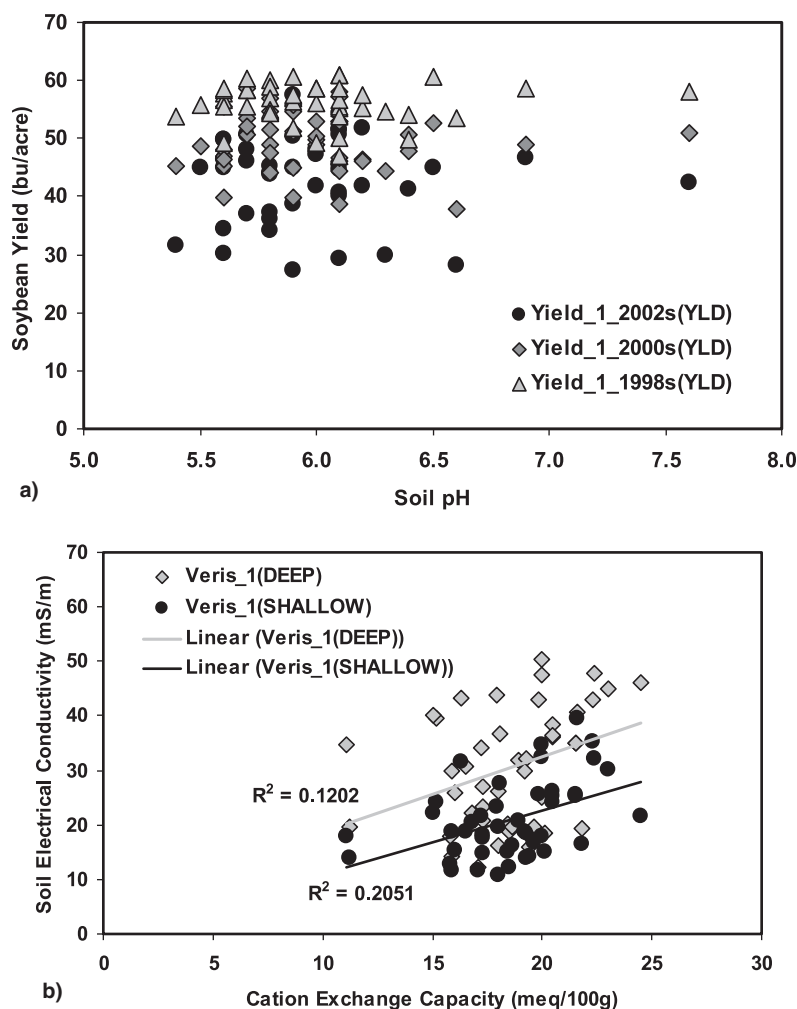


FIGURE 10.6 Example of relationships between (a) soil pH and soybean yield, (b) cation exchange capacity and soil electrical conductivity, and (c) soil organic matter content and soil mechanical resistance using 45 soil sampling points as processed with DM_Comp software.

response and other important relationships can be defined when similar output files from different fields are combined.

The other output file, *0.4_acre_grid.txt*, can be illustrated by creating original and processed deep electrical conductivity maps using a simple GIS software package (Figure 10.7). In the case of 0.4-acre averaging, each data point includes an ID number as well as the grid cell average for selected attributes. Missing points mean that the corresponding grid cells did not contain at least one original data point from at least one selected data layer.

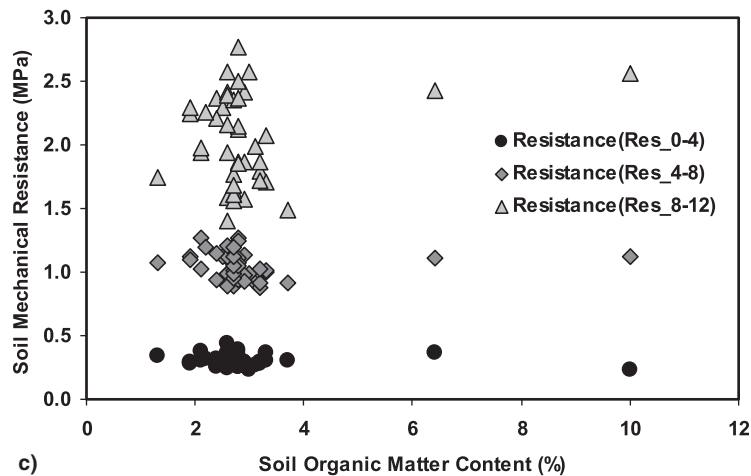


FIGURE 10.6 (continued)

10.5 CONCLUSIONS

The DM_Comp software was developed as a stand-alone application capable of collocating multiple self-generated point data layers. It illustrates the concept of task-oriented spatial data handling without the need for the actual GIS software. A similar approach can be pursued to facilitate other specific data management routines (e.g., filtering, primitive geostatistical analysis or management zone delineation).

In this example, DM_Comp was used to compare several common self-generated data layers. Analysis of relationships between different data layers can be pursued through a decision-making process. For example, the user could identify soil properties to be considered as yield limiting factors. An ability to relate laboratory analysis of the limited number of soil samples and high-density soil sensor measurements can help calibrate the sensor to better delineate field areas with potential nutrient deficiency. Grid cell averaging can be pursued when either simply defining a rectangular grid cell pattern for a follow-up soil sampling, or when analyzing spatial data files with measurement spacing different with respect to northing and easting directions. More advanced analytical methods will certainly require a GIS software package with data smoothing, interpolation, or raster computation capabilities.

10.6 SUPPLEMENTAL FILES

The following files should be copied to the same folder to follow the example above:

1. *DM_Comp.exe* is a program itself.
2. *Soil_1.txt* is a georeferenced soil laboratory analysis report.
3. *Yield_1_1998s.txt* is the 1998 soybean yield data saved using the AgLeader advanced text export format.
4. *Yield_1_2000s.txt* is the same for 2000 growing season.

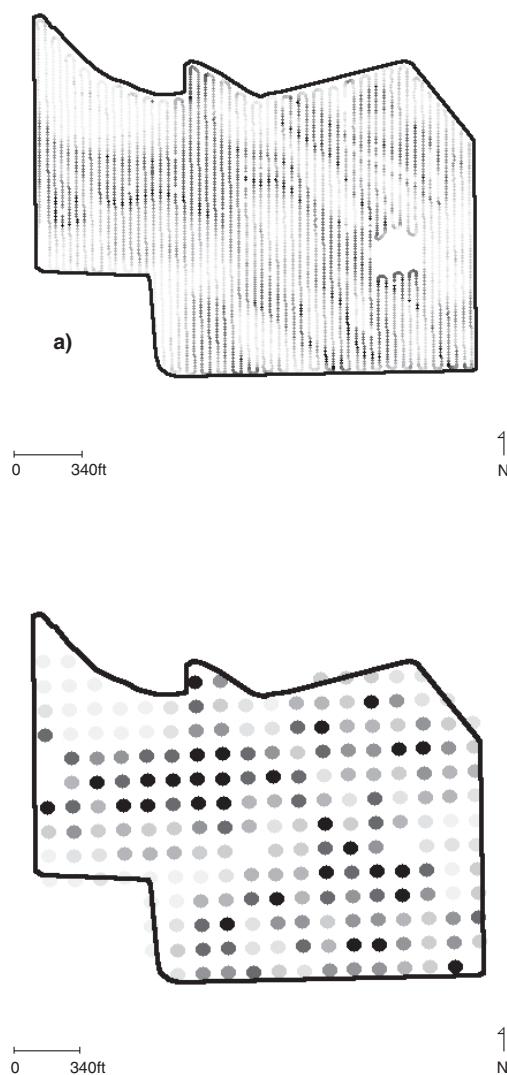
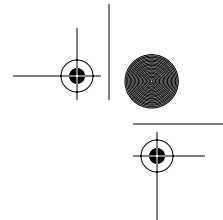


FIGURE 10.7 Deep soil electrical conductivity maps produced using (a) the original data and (b) 0.4-acre square grid averaging option.

5. *Yield_1_2002s.txt* is the same for 2002 growing season.
6. *Veris_1.txt* is a Veris® 3100 EC Surveyor output file.
7. *Resistance_1.txt* is a logging file with on-the-go soil mechanical resistance measurements.



REFERENCES

1. Adamchuk, V.I., Dobermann, A., and Ping, J., Listening to the story told by yield maps, Precision Agriculture extension circular EC 04-704, University of Nebraska, Lincoln, Nebraska, 2004.
2. Adamchuk, V.I., Untangling the GPS data string, Precision Agriculture extension circular EC 01-157, University of Nebraska, Lincoln, Nebraska, 2001.
3. Siefken, R.J., Adamchuk, V.I., Eisenhauer, D.E., and Bashford, L.L., Mapping soil mechanical resistance with a multiple blade system, *Appl. Eng. Agr.*, 21, 15, 2005.

