Step-by-Step Yield Monitor Data Analysis

Bruce Erickson

Introduction

Farmers and agribusinesses have accumulated mountains of yield data since the commercial introduction of combine yield monitors in the U.S. in 1992. One common complaint is that there is no good way to analyze that data. In response to this need, Purdue Site-Specific Management Center researchers have posted step-by-step instructions for their approach to analyzing yield monitor data (see www.purdue.edu/ssmc, select Publications, and look for "Yield Monitor Data Analysis: Data Acquisition, Management and Analysis Protocol").

"We hope farmers and agribusinesses will try our approach and give us feedback," said Terry Griffin, a Purdue Agricultural Economics PhD student working on the use of precision farming data in crop management decision making. Griffin is studying under Jess Lowenberg-DeBoer, Professor and Director of the Site-Specific Management Center (SSMC). "This isn't the only approach to data analysis, and it may not be the perfect approach, but it does provide clear answers, where visual analysis would only give you impressions or suggestions," Griffin said.

The Purdue SSMC researchers focus on what is referred to as "inferential statistics" when using spatial data. In particular, they focus on the reliability of the choice. They are continually asking whether the differences in the data reflect reality or whether it is just the result of random variation. These choices might be "either/or" decisions about inputs such as herbicides, hybrids, insecticides or tillage types, or they might be rate decisions about the quantity of fertilizer or plant population. The focus on inferential statistics at Purdue is in contrast to the more common use of Geographical Information Systems (GIS) to communicate spatial data and facilitate visual comparisons.

The report is based on a decade of work by Lowenberg-DeBoer and his students. They adapted techniques from geology, regional economics, geography and other disciplines with a longer history of spatial data analysis and applied them to the specific problems of agricultural data. These problems include the fact that yield monitor, electrical conductivity (EC) and most other ground-based agricultural sensors collect data that is denser within the pass than among passes. The Purdue group has worked on corn, soybean, wheat, rice and cotton data from Indiana, Illinois, Kentucky, Minnesota, Arkansas, and Arizona. They have also worked with yield monitor data from Canada, Argentina and South Africa.

The step-by-step instructions include the use of free software that is available on-line. For cleaning yield data to remove outliers and bad data, the Purdue group suggests using Yield Editor developed by USDA Agricultural Research Service (ARS) (Drummond, 2005). For exploratory analysis they suggest using GeoDa developed by University of Illinois professor Luc Anselin (Anselin, 2003).

Analysis with Interpolated Data

The Purdue SSMC group suggests that analysis should stick as close to the raw data as possible. In particular they caution that the common practice of using interpolated data to deal with differences in spatial resolution can lead to the wrong decisions, because it introduces spatial patterns that do not exist in the original data.

It is common for data to be collected on different spatial scales. For example, yield monitor data might be available at one, two, or three second intervals (every 7 to 22 feet) on every combine pass, but soil tests may only be on a 2.5 acre grid. One approach would be to interpolate the soil tests to obtain a continuous map and use that interpolated value in the analysis. The problem is that while the interpolated map may

be the best available estimate, it is not perfect and the interpolation methods (e.g. kriging, inverse distance) may bias the statistical analysis.

The Purdue group suggests using the least dense data set as the resolution for the remaining layers, but caution that typical grid or management zone soil test information may be too sparse to be useful. When both yield and EC data is available they suggest using simple averages of yield and EC data within polygons that make sense for the equipment size and field layout. For instance, one option is the use of grids with each side being some multiple of equipment width. Another possibility is to average yield data points within a given radius of less dense soil characteristic data.

Vision of the Future

The SSMC researchers acknowledge that working with yield monitor data is not easy, but they think that if the value of the analysis can be shown, then software engineers will find ways to make it easier to use.

"Twenty years ago if you had told me that thousands of farmers would be using GIS, I would have said you were crazy. Twenty years ago, GIS was for specialists. It was too complicated for the average person. But today many Corn Belt farmers use GIS to organize and visualize data," Lowenberg-DeBoer said. "Similarly, I think that software engineers could make spatial data software easier to use. I don't know if many farmers will use spatial data analysis, but I am sure that their agronomists and crop consultants will."

Value of Analysis

One unanswered question is the value of spatial crop data as part of a farm management system. There are many examples of effective use of spatial data in crop research and in answering specific farm management questions, but no one has been able to link spatial data and overall farm profitability. Kent Olson, University of Minnesota, was unable to find a statistical relationship between yield monitor use and farm profits in data from farm business associations in his state (Olson, 2003)

The problem may be related to how the spatial data was used. The farm profitability data used by Olson only noted whether or not the farm had a combine yield monitor. At best those farms used the yield monitor data to make yield maps and do visual assessments. The type of analysis developed by the Purdue group was not available to farmers at the time when Olson did his study. Some of the farms with yield monitors in Olson's dataset may not have even used their data.

In other parts of the economy, data analysis seems to be most commonly used by very large businesses. Data analysis is a standard business practice among large retailers such as Wal-Mart and Target. It is also a well accepted practice among credit card companies, such as American Express. In livestock agriculture, the success of Murphy Family Farms (now part of Smithfield) was often attributed to analysis of production data from their many farrowing and finishing units. But business data analysis is almost unknown among "Mom-and-Pop" retailers and unusual for regional chains.

"I am convinced that analysis of crop data will be profitable, but the question is for what size farms?" said Lowenberg-DeBoer. "Do we have to have Wal-Mart sized farms before we see that profitability? Or can family size farms make use of these techniques?"

To test potential usefulness for family sized farms the Top Farmer Crop Workshop (www.agecon.purdue.edu/topfarmer) has been offering yield data analysis for workshop participants for the last two years. Most Top Farmer participants have yield monitors in their combines and about half say that they will bring yield data for analysis in the future, but in 2004 only four brought data and in 2005 only three farmers participated in yield data analysis.

Conclusions

Lowenberg-DeBoer, Griffin and the other Purdue SSMC researchers look forward to your comments on the yield data cleaning and exploratory analysis methods that they are suggesting. Please send comments to: lowenbej@purdue.edu or twgriffi@purdue.edu.

For More Information

Anselin, Luc, 2003. GeoDa 0.9 User's Guide. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign, IL. Available at: https://geoda.uiuc.edu/default.php

Drummond, Scott, 2005. Yield Editor 1.00 Beta Version User's Manual November 9, 2004. Available at: http://www.fse.missouri.edu/ars/ye/yield_editor_manual.pdf

Olson, Kent D. and Pascal Elisabeth. An economic assessment of the whole-farm impact of precision agriculture. Presented at American Agricultural Economics Association Annual Meeting, July 27-30, 2003, Montreal, Canada, 2003. 12p., 28cm. Available at: http://agecon.lib.umn.edu/aaea03/sp03ol01.pdf

Griffin, T.W., J.P. Brown, and J. Lowenberg-DeBoer. 2005. Yield Monitor Analysis: Data Acquisition, Management, Analysis, and Interpretation Protocol. Available on-line at: http://www.purdue.edu/ssmc