NMSP Yield Monitor Data Cleaning Project Improves Information for Farmers and Researchers

By Lisa Fields

In recent years, advances in yield monitor technology have increased the availability of crop yield data for both farm managers and researchers. An exciting project is improving the usefulness of these data by evaluation through a data cleaning process. Initiated in 2017, "Forage Yield Monitor Data Processing for Accurate Maps," is led by Professor Quirine Ketterings, Director of Cornell Nutrient Management Spear Program (NMSP). She elaborated, "We know that variations in field topography, drainage, and equipment operation can cause data errors that skew yield estimates from harvest equipment. To be able to use yield monitor data for crop and nutrient management on-farm research we needed a standardized protocol for cleaning raw yield monitor data. At the farm level, proper data cleaning is essential, as accuracy of yield data plays a big role in crop management decisions that can affect the farm's bottom line."

Rick Bair of Selden Stokoe and Sons Farm invested time into learning to use Yield Editor to improve yield data quality (photo credits: Kevin Keenan).

The Forage Yield Monitor Data Processing project is funded by the Northern New York Agricultural Development Program (NNYADP), Federal Formula Funds and a multi-state USDA-NIFA grant that aims to use advanced statistical methods to analyze data from on-farm strip trials. The project is in collaboration with counterparts at the University of Missouri and the Iowa Soybean Association. The cleaning protocol utilizes Yield Editor software to identify errors in yield monitor datasets so they can be removed from yield calculations. The software is available free of charge through the website of the USDA-ARS.

Ketterings noted, "In late 2016 we connected with Yield Editor's IT specialist, Scott Drummond of USDA-ARS to become familiar with the software and see if we could use it to standardize cleaning of corn silage data as well as corn grain data. We were successful in applying Yield Editor to corn silage data so we developed the standardized cleaning protocol based on the software."

The Yield Monitor Data processing project is integrated with another NNYADP funded initiative, "Re-Evaluating Yield Potential of Corn Grain and Silage in Northern NY." Its purpose is to update the database for corn grain yield potentials of soils, to develop an independent database for corn silage yield potentials, and to evaluate corn nitrogen management guidelines based on yield levels. It is really important that data that are shared are properly cleaned as trustworthy yield data from farmers' fields are essential for this project. Supporting farmers and farm consultants who wish to use Yield Editor is an important part of our work."

At the 2018 Corn Congress, Selden Stokoe and Sons Farm employee Rick Bair connected with the NMSP project. "We grow about 4,000 acres of commodity grain crops. Applying Yield Editor seemed to be a really good fit for my skills," he said. Bair's former career was Technical Sales and Business Development support in the corporate world of data management. He explained, "I was recruited to the farm because of my background. The speed at which data-generating technology was changing was beyond the time the guys could spend to keep up with it and make good use of the information. The desktop data

management software we use has some filters built in, and Yield Editor adds more capability and is much faster. It takes the noise out of the yield maps."

Bair described the data cleaning process. "The raw data files from the harvesting machinery have data "artifacts." These are inaccuracies caused by things such as the delay at the beginning of a harvest pass, speed changes or field features such as washes. Another cause of errors we see is when a combine has picked 12 rows, and then on the next pass it picks 11, the yield estimate will show a false low because the machine assumes it picked 12 rows both times. With Yield Editor you can check data continuity by comparing a data point to all points within a defined number of machine widths (I use five), from the point. Points more than four standard deviations from the mean are identified and can be eliminated. This allows you to remove false lows and highs caused by differences in number of rows picked in a pass or field topography. Another editing feature is to change the time delays at the ends of field passes. This identifies outliers caused by differences in turn speed and when the combine head was raised and lowered."

Sheryl Swink, NMSP Research Aide, worked with the data from Selden Stokoe and Sons Farm. Together with postdoctoral researcher Tulsi Kharel, and other team members including several undergraduate students, the team has cleaned farm data from about 41,000 acres of corn silage and 21,000 acres of corn grain so far, much of it from multiple crop years. Swink said, "The use of yield monitors for corn silage harvest is fairly recent. We see different monitor data errors than in grain harvest. There are a lot more odd-shaped fields that cause non-uniform travel patterns, and far more acceleration changes. The data cleaning experience enabled us to develop guidelines for farms. It covers critical factors for both silage and grain growers, such as calibration, and consistency in both velocity and pass alignment. The good news is that we see far fewer errors about three years after farms start using yield monitors during harvest as they get used to the technology."

Bair commented, "The reality of a field can be quite different from the raw data map. I enjoy getting the ground level view during harvest season. I apply the picture I have in my head when I'm back in the office, looking at numbers on the screen and cleaning the data. We do everything on this farm by prescription, so the more precise the yield information is, the more precise our management choices such as fertilizer and seeding rates will be."

"It's exciting to have software that helps us gather so much information quickly enough to take on such a large project. We have a whole lot more datasets in the pipeline to process," Swink said. "We need a robust dataset of fields for each soil type across crop years to apply cleaned yield information to the Soils Yield Potential database and identify the factors that affect yield stability. From there we can determine whether or not the N recommendations need to be altered."

Summarizing his view of the two NMSP projects, Bair said, "Ultimately, what I hope to gain from the time investment with Yield Editor is to use the tool to get better prepared yield maps to do more of our analysis at the digital level. If we can do that instead of using strip plots to evaluate performance of crop varieties, population rates and other management practices we can save a lot of time and I think get better answers." He added, "The big picture benefit is that improving the quality of yield monitor data can feed into the database of soil type yield potential. It's good to be part of this project and contribute to something that's useful to all of NY agriculture."

(June 4, 2018)

The **Nutrient Management Spear Program** (NMSP) is an applied research, teaching and extension program for field crop fertilizer and manure management on dairy and livestock farms. It is a collaboration among faculty, staff and students in the Department of Animal Science, Cornell Cooperative Extension, and PRO-DAIRY. Our vision is to assess current knowledge, identify research and educational needs, facilitate new research, technology and knowledge transfer, and aid in the on-farm implementation of strategies for field crop nutrient management including timely application of organic and inorganic nutrient sources to improve farm profitability while protecting the environment. An integrated network approach is used to address research, extension and teaching priorities in nutrient management in New York State. For more information on NMSP projects and extension/teaching activities, visit the program website (http://nmsp.cals.cornell.edu) or contact Quirine Ketterings at qmk2@cornell.edu or (607) 255-3061.