Improving Yield Data Accuracy: Challenges and Solutions

Ali Mirzakhani Nafchi, Assistant Professor and SDSU Extension Precision Agriculture Specialist
Yashar Askarzadeh, Graduate Research Assistant, South Dakota State University

January 2024

1. Background

A yield monitor is a tool designed to measure and record crop yield in real-time as agricultural machineries harvest the crops. Initially, crop yields were monitored using volume or weight measurement systems per harvested area. Still, later, yield monitors were developed to determine the real-time volume or weight of the crop through harvesters, stamped with GPS coordinates and the crop moisture content to compute on-site yield. This technology evolved from manual readings to automated sensor-based systems, improving data collection precision. Yield monitoring generates files filled with significant amount of crop production, crop moisture, and spatial information. After harvest, yield data is extracted from the yield monitor and imported into geospatial agricultural mapping software (such as SMS). It's important to note that each yield monitor brand may generate data in different formats. For example, a John Deere system would provide information like point coordinates, field name, track, swath, width, distance, elevation, and other factors directly affecting crop yield. Despite improvements in the accuracy of commercially available yield monitors and the introduction of various types, challenges in data quality persist. There are two main types of issues: sensor system-related problems affecting data accuracy and operational issues leading to data collection problems. Without data quality control, the resulting yield will be inaccurately represented, impacting the final prescription maps developed based on the spatial yield information.

2. Components influencing yield measurements

Yield measurement errors often result from improper sensor calibration and operational errors. Key components in yield data computation include travel distance, header swath, crop moisture, travel time, flow, and reading cycles. To ensure sensor accuracy, routine checks and calibrations before harvest involve visual inspections to detect visible equipment damage or obstructions. Sensor tests follow, verifying moisture sensors' correct identification of moisture levels, which is crucial. Wet grains sticking together can disrupt flow and skew yield data, while excessively dry seeds can produce dust, interfering with sensors. Calibration is another step to ensure synchronization between the yield monitor's data, accurate detected moisture content, recorded header swath correctly when different changes occur during the harvest. Proper inspection and calibration maintain yield data consistency and accuracy.

3. Pre-harvest considerations

The condition of harvesting equipment and machinery, including wear and tear, cleanliness, and functionality, can impact the accuracy of the data collected. A dirty or worn-out sensor may not detect grain flow accurately, causing errors in data collection. Therefore, proper inspection and regular maintenance checks are necessary to ensure all parts are functioning optimally, and to prevent blockage or malfunction.

Weather conditions also influence data collection accuracy. Sensors might not be able to provide reliable readings when the fields are wet; high humidity can cause seeds to stick together, which can affect the grain flow rate and yield estimation. Extreme temperature fluctuations might also interfere with the machine's optimal functioning. Crop moisture content can be reduced significantly in hot weather, leading sensors to overestimate yield, in contrast, cold weather also may affect readings. Fields that become muddy from rain, can make machinery skid, or get stuck, impacting both harvest and data collection. It is advisable to cross-reference yield data with weather condition reports from the operation date to determine if extreme conditions might have affected the readings.

Seeds' moisture content is perhaps the most important aspect to consider before harvesting. Excessively wet seeds may clump together, causing blockages in the harvester and leading to unreliable data. Conversely, seeds that are too dry could flow too quickly, making it difficult for sensors to capture accurate readings and increasing the risk of grain damage during harvest. Farmers should ensure that the crop's moisture content is within the optimal range for harvesting to facilitate the best data collection.

4. Factors causing errors

4.1. Header position and swath width

The header's position controls the start and end of data collection in many yield monitoring systems. A specialized switch, installed on the harvester's feeder housing or directly on the header, manages this process. Some machines also have a control switch by the driver's seat to monitor the header's position. Improper switch setup can lead to inaccurate yield data (Figure 1), such as zero values (Figure 2) in productive areas or exaggerated readings from empty or previously harvested sections. The system must stop collecting data when the header is lifted to avoid recording irrelevant information.

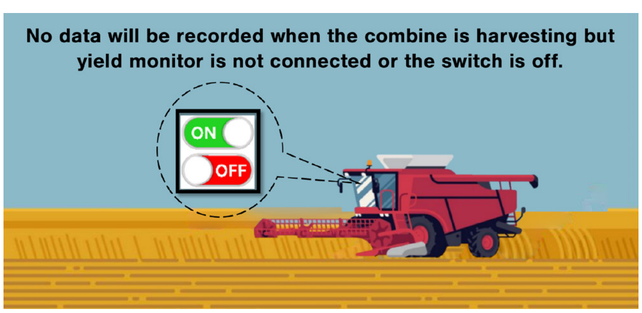

Figure 1. Harvesting with inactive header switch

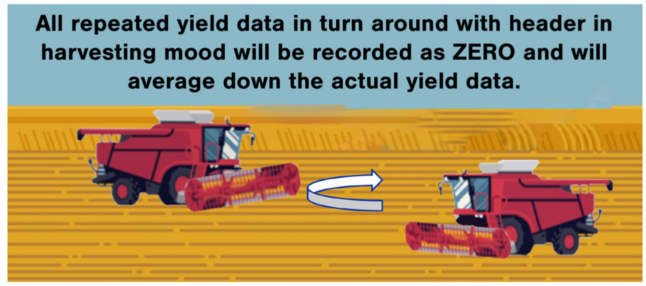

Figure 2. Header switch misalignment impact

The swath width—the area harvested by the header—is equally important. If the system overestimates this width, inaccurate readings can result. Some harvesters may not use a dedicated swath sensor, instead using the machine's full width, which can create errors (Figure 3). Modern systems, such as those from Ag Leader, address these issues by documenting the header's position in data files, allowing for easier error correction afterward.

Figure 3. Swath width miscalibration effects

4.2. Speed changes

Maintaining an optimal speed range is important when operating a combine harvester, as any deviation can lead to inaccurate yield estimations. Yield data tend to be underestimated during rapid acceleration and overestimated during deceleration (Figure 4). This happens because speed affects how quickly kernels pass through the mass flow sensor, causing variations in the logged data points. Data points that show a speed change greater than 15% from one reading to the next are usually considered unreliable and excluded from analysis. Zero values for grain or silage moisture are another type of error that can be easily spotted. These are typically recorded when the combine starts in the field or stops while the material has not yet reached the moisture sensor.

4.3. Low or high speed

Low speeds can be problematic for yield monitors, which are calibrated for typical harvest speeds. When the harvester is moving too slowly or is stationary, the collected data may become unreliable. In some instances, the system might record an ‘infinite yield’ if it detects grain flow but no associated harvested area. Modern systems often avoid logging data at these low speeds to prevent such issues. Similarly, high speeds can compromise yield monitoring precision. Operating a combine faster than recommended can lead to GPS delays, resulting in discrepancies between location data and actual yield and subsequently distorted yield maps. Moreover, at high speeds, crops might not be processed cleanly, leading to blockages and inaccurate yield calculations of unprocessed crops. While modern systems adjust for high-speed errors, the most effective strategy remains to operate within proper speed limits to prevent these problems.

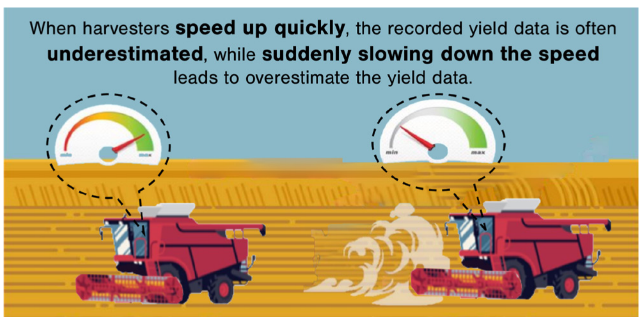

Figure 4. Harvest speed variability impact

4.4. Grain flow delay

The time lag between the actual harvesting of crops until the time the flow rate readings take place by yield monitors is a significant factor affecting the accuracy of yield data. Influenced by the harvester’s design, operating speed, ground slope, and the crop’s load, this delay can lead to data inaccuracies (Figure 5). Setting a standardized delay parameter for each field may compensate for this problem. Typically, grains follow diverse routes from the header to the grain hopper, creating an unpredictable time lag. Yield monitors measure the grain mass at a specific point along this route, making the precise calculation of delay time a complex task. Some monitors use a fixed delay time to approximate the average transit time, while others rely on operators’ estimates to adjust for variations.

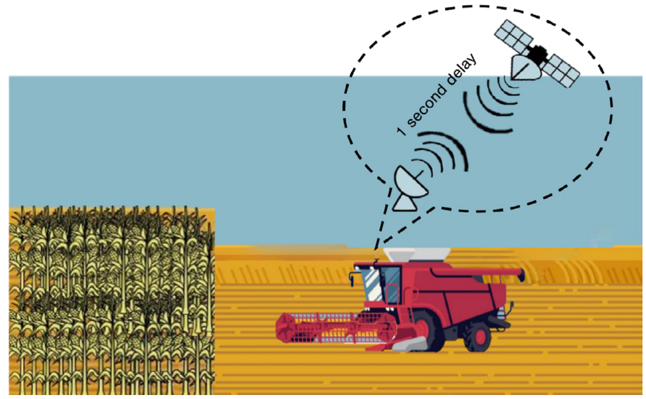

Figure 5. Due to the path a grain passes till reach the grain flow there is time delay happens between the actual location and collected data.

4.5. Sensor errors

Yield estimation errors often originate from improperly calibrated moisture, flow rate, and temperature sensors. Additionally, the use of outdated calibration data by some farmers contributes to inaccuracies. It’s essential to calibrate mass flow sensors across the full range of expected values; uncalibrated ranges can significantly increase error likelihood. Yield monitor calibration should focus on the flow rate rather than the yield itself. For example, harvesting 200 bushels per acre at 1 mph might result in a moderate flow rate, while 100 bushels at 8 mph could produce a very high flow rate. If flow rate data exceed calibrated limits, reliability decreases. Such data should be considered less credible and may need to be excluded from further analysis.

5. Data normalization

After applying the previously discussed errors and inaccuracies to yield data, some anomalies may remain. The following step in data cleaning is statistical normalization, which calculates the average and standard deviation of the yield data within a specific pre-defined area. For example, it is common to assume that a block measuring X1 by X1 feet will have relatively consistent yield data. If a recorded yield within this block exceeds X2 standard deviations from the block’s average yield, it is flagged as suspect and might be removed from subsequent analysis. Visual representations, such as graphs or plots, can effectively highlight the impact of idle normalization (Figure 6) by showcasing the data before and after the process (Figure 7).

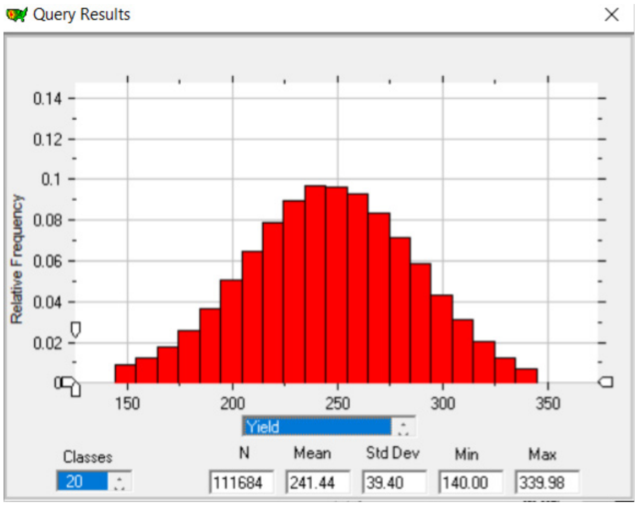

Figure 6. Yield data normalization for mapping

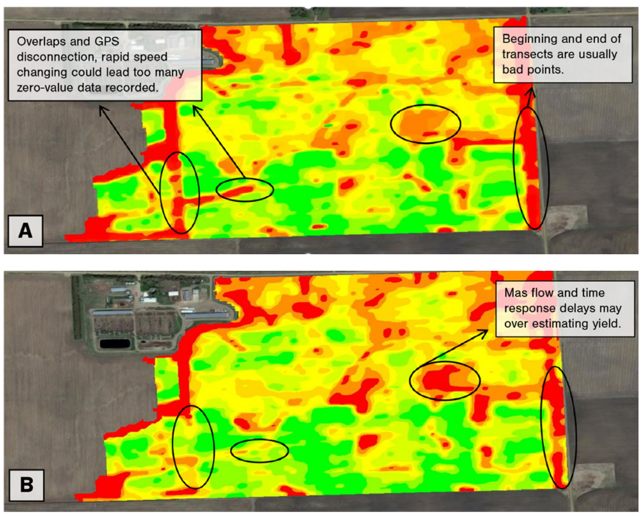

Figure 7. Yield data; before (a) vs. after normalization (b).

6. Summary

There is a need to implement strategies to ensure that yield data is both accurate and precise. Farmers and agricultural professionals who apply these methods can expect to see an improvement in their understanding of crop performance. By refining the data collected, they can make informed decisions that could lead to enhanced productivity and sustainability of their farming operations. Accurate yield maps can help identify low-yielding areas requiring more accurate management of crop inputs to obtain higher yields and profits.

References

Arslan, S., & Colvin, T. S. (2002). Grain yield mapping: Yield sensing, yield reconstruction, and errors. Precision Agriculture, 3, 135-154.

Kharel, T., Swink, S. N., Youngerman, C., Maresma, A., Czymmek, K. J., Ketterings, Q. M., & Hubbard, V. (2018). Processing/cleaning corn silage and grain yield monitor data for standardized yield maps across farms, fields, and years. Ithaca, NY, USA: Cornell University, Nutrient Management Spear Program, Department of Animal Science. Retrieved August 19, 2020.

Kharel, T. P., Swink, S. N., Maresma, A., Youngerman, C., Kharel, D., Czymmek, K. J., & Ketterings, Q. M. (2019). Yield monitor data cleaning is essential for accurate corn grain and silage yield determination. Agronomy Journal, 111(2), 509-516.

Kleinjan, J., Chang, J., Wilson, J., Humburg, D., Carlson, G., Clay, D., & Long, D. (2002). Cleaning yield data. SDSU Publication.

Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield maps. Agronomy Journal, 99(6), 1471-1482.

USDA. 2023. Yield Editor 2.0.7. U.S. Department of Agriculture, Ag Data Commons https://data.nal.usda.gov/dataset/yield-editor-207

South Dakota State University Extension

SOUTH DAKOTA STATE UNIVERSITY®
AGRONOMY, HORTICULTURE & PLANT SCIENCE DEPARTMENT

P-00284