Site-Specific Management Guidelines

J. Kleinjan, D.E. Clay, C.G. Carlson, and S.A. Clay

SSMG-45

(10/06)

Developing Productivity Zones from Multiple Years of Yield Monitor Data

Summary

The ability to instantaneously record yield data at harvest has been widely available to agricultural producers since the mid-1990s. Yield is the ultimate integrator of landscape and climatic variability and therefore should provide useful information for identifying management zones. However, due to year-to-year climatic variation, identifying useful management zones based on a single year's yield map is difficult. Increasing the number of years used to define zones may be a solution to this problem. A technique to define a type of management zone known as a productivity zone based on multiple years of yield monitor data involves creating common grid-cells across years and then calculating mean yield and standard deviation maps. 'Mean yield' maps created from multiple years of data may be used to determine yield goals and fertilizer recommendations, while standard deviation maps may be used to identify areas requiring corrective management. The preferred method for explaining yield variability used a combination of average yields and standard deviation to delineate productivity zones.

Introduction

Producers who have been collecting yield monitor data for multiple years are asking: "How can this yield data be used to improve management?" A potential use for these data sets is incorporating them to define a type of management zone known as a productivity zone. There are at least two different approaches proposed for identifying productivity zones. The first approach is to calculate the impact of zone boundaries on fertilizer recommendations. Chang et al. (2004) reported that landscape specific yield goals, combined with grid-cell sampling, can be used to improve nitrogen (N) and phosphorus (P) fertilizer recommendations by 35 and 59%, respectively. This approach requires that extensive soil sampling be conducted to define initial soil conditions and then a model be used to calculate fertilizer recommendations for each zone. A second approach is to determine the impact of productivity zones on yield variability (Bakhsh et al., 2000; Fridgen et al., 2000; Diker et al., 2002; and Kitchen et al., 2002). This approach assumes that the best method of zone delineation minimizes yield variability. Due to the widespread availability of multiple-year yield monitor data sets and relatively simple method of collection (versus grid soil sampling), many producers would opt for the second method of productivity zone delineation.

Removing Erroneous Data Points

Yield monitor data may be exported as text files (.txt extension) from the yield monitor software and viewed in common spreadsheet software programs in either basic or advanced format. Basic format contains longitude/latitude coordinates and yield in bushels per acre (bu/A). Advanced format contains many aspects of machine operation including flow rate, header up/down, distance, speed, etc. These factors may be used to remove erroneous data. Yield monitor data sets normally contain values that are incorrect. For example, erroneous data can be collected when harvest swaths are not the full header width, the combine's speed is rapidly changing, or in areas near the end rows. Yield monitor data accuracy is improved by removing erroneous points…cleaning the data (Blackmore and Moore, 1999; Thylen and Algerbo, 2000). From a visual perspective, removing these erroneous data points may have little impact upon yield map appearance. From an analytical perspective, removing erroneous data may significantly impact the ability to compare yield data with other decision support system information. There are several precision agricultural software packages ranging in price from $500 to $700 that automatically clean or filter yield data when it is loaded into the program for analysis. Some no-cost programs available for cleaning yield monitor data are available.

For example, find "Field Analyst" at: >http://plantsci.sdstate.edu/precisionfarm/paper/publicationSoftware.aspx< "Yield Editor 1.00" is available at: >www.fse.missouri.edu/ars/decision_aids.htm<. Both programs clean either AgLeader CSMS or Greenstar advanced text file formats.

Assigning Yield Data to a Common Grid

Once yield data are cleaned, the data should be converted to a common coordinate system in order to compare

the same field areas from multiple years. One technique to accomplish this task is to identify common grid-cells. "Field Analyst" converts multiple years of yield data to a common grid-cell format. This is accomplished by: 1) calculating the maximum and minimum latitude and longitude for a field, 2) selecting a desired grid-cell size, and 3) determining the average yields for each grid-cell. This study used grid-cell size of three times the combine header width or 45 by 45 ft. A 45 by 45 ft. grid-cell size results in over 3,400 grid-cells in a 160 acre field. The assigned grid-cell values are the average of all yield data that fall within the boundaries of that particular grid-cell. For each grid cell, "Field Analyst" calculates a coordinate (Row #, Column #), the latitude and longitude coordinates in decimal degrees, feet, and meters (to the center of the grid-cell), and the number of yield measurements in the grid-cell and average yield value within each.

Statistical Calculations

To calculate the standard deviation for a given grid-cell, it may be necessary to convert or normalize yields from different crops to a common scale. Normalization allows yields of corn, soybeans, wheat, or any other crop to be directly compared. The normalization process is not necessary for all types of analysis. When comparing multiple years of the same crop, normalization is usually not required. However, when comparing the yields of two crops with different yield levels, such as corn and soybean, normalization is necessary. An in-depth discussion of normalization methods is beyond the scope of this paper.

Basic statistics, such as multiple year average yields, standard deviations, and the coefficients of variation (CV) can be calculated on the grid-cell output file using spread-sheet software. Average yields, or means, are used to describe the central tendency of the data set. The mean is determined by summing all the values within the data set and then dividing this value by the number of observations. One of the most widely used operators in statistics is the summation symbol Σ. The summations term is best described by an example, as follows.

If four measurements…1, 5, 6, and 3…are obtained from a population, then the sum of these values (1+5+6+3) is 15. Each individual value can also be written in termi-nology where X_1 is 1, X_2 is 5, X_3 is 6 and X_4 is 3. Using this terminology, it is possible to write universal mathematical expression that can be used to calculate a variety of factors. For example, the sum of 1+ 5+6+3 can be written as $\sum_{i=1}^{4} Xi$. The $\sum_{i=1}^{n} Xi$ symbol means to sum all the Xi values where i goes from 1 to n. In this example n is 4 and the values for X_1, X_2, X_3, and X_4 were defined above.

The mean, or average, of this sequence of numbers is the sum divided by the number of values contained in the data set. A notation for the mean or average that is often used is $\bar{X}$. The mean for the example above is $\bar{X} = \frac{15}{4} = 3.75$. Mathematically, this is expressed by the equation,

$$\bar{X} = \frac{\sum_{i=1}^{4} Xi}{n} = \frac{15}{4} = 3.75.$$

The *variance* provides a measure of variation or

differences among the individual sample values in the data set. The variance is a measure of precision, not accuracy. Precision and accuracy are different values. Precision represents the degree of scatter in a data set while accuracy is the degree of variation around a specific value. Accuracy requires that the true value be obtained. For example, when I play darts, if I consistently shoot 'tens' when aiming at the bull's-eye, I am very precise but not very accurate. The variance (s^2) is calculated using the equation:

$$s^2 = \frac{\sum_{i=1}^{4}(Xi - \bar{X})^2}{n-1}$$ [1]

where $\bar{X}$ as described above is the data set mean, X_i are each of the individual measurements, and n is the number of random observation in the data set. Using a data set provided above, where $X_1=1$, $X_2=5$, $X_3=6$, and $X_4=3$, the variance is calculated as follows:

$$s^2 = \frac{\sum_{i=1}^{4}(Xi - \bar{X})^2}{n-1} = \frac{(1-3.75)^2 + (5-3.75)^2 + (6-3.75)^2 + (3-3.75)^2}{4-1} = 4.922$$

The *standard deviation* is the square root of the variance (s). In this example the standard deviation is 2.22 ($4.922^{0.5}$).

The *coefficient of variation* (CV) provides an indication on the degree of precision relative to the mean. The CV is calculated using the equation, $CV = \frac{s}{\bar{X}} * 100$ where s is the standard deviation (square root of the variance, s^2) and $\bar{X}$ is the mean. For the example provided above, the CV is calculated as follows: $CV = \frac{s}{\bar{X}} \frac{2.22}{3.75} * 100$. The larger the CV data set, the less precision there is relative to the mean. An example of a calculated statistics file is shown in **Table 1**.

Identifying Productivity Zones

In this example, 4 years of corn yield data were used to develop the productivity zones. Yields were highly variable from year to year due to climatic differences (**Figure 1**). Productivity zones based on the average yield and standard deviation values can be identified using several different approaches including cluster analysis, producer preferences, and natural boundaries (Chang et al., 2004; Fleming et al., 1999; Kitchen et al., 2002; Walker et al., 2004). This paper selected four approaches that can be easily applied using a standard spreadsheet software program, by using mean and standard deviation values. In method 1, the average yield approach, arbitrary yield values based on producer preference were chosen as the boundaries between each of three zones (low, medium, high). In method 2, productivity zones were sorted in a similar manner based on standard deviation values. Again, producer preferences were used to sort values into three zones (low, medium, high). In method 3, productivity zones based on the coefficient of variation (CV) were also separated into three zones.

In method 4, both yield and standard deviation data are used to identify four different productivity zones. Both yield and standard deviation were split into two categories, above and below the average value. Combining these categories resulted in four productivity zones with the following characteristics: 1) high yield, high deviation,

Table 1. A portion of a summary file with elevation, seasonal and average yields, standard deviations, and coefficients of variation for 4 years of corn yield data.

	Row Number	Column Number	Long Feet	Lat Feet	Elevation (ft)	Corn96	Corn98	Corn00	Corn02	Average Yield (bu/ac)	Standard deviation (bu/ac)	Coefficient of Variation (%)	Yield zone	De
1														
2	1	1	22.50	22.50	1660.48	111.35	176.87	151.40	110.72	137.59	32.38	23.53	1.00	
3	1	2	22.50	67.50	1660.68	112.98	176.25	156.60	95.43	135.32	37.50	27.71	1.00	
4	1	3	22.50	112.50	1660.28	107.71	217.89	155.27	89.88	142.69	57.23	40.11	1.00	
5	1	4	22.50	157.50	1659.72	114.23	212.60	153.29	106.48	146.65	48.51	33.08	1.00	
6	1	5	22.50	202.50	1659.09	115.57	204.69	165.56	112.79	149.65	43.98	29.39	1.00	
7	1	6	22.50	247.50	1658.46	109.43	216.24	186.51	110.30	155.62	54.21	34.84	1.00	
8	1	7	22.50	292.50	1657.81	140.38	207.91	183.71	105.00	159.25	45.70	28.70	1.00	
9	1	8	22.50	337.50	1657.15	97.93	226.14	184.03	103.47	152.89	62.71	41.02	1.00	
10	1	9	22.50	382.50	1657.23	125.86	222.65	186.52	103.47	159.62	54.74	34.29	1.00	
11	1	10	22.50	427.50	1657.64	136.68	216.81	170.89	110.49	158.72	45.95	28.95	1.00	
12	1	11	22.50	472.50	1658.07	126.84	215.48	164.53	143.61	162.62	38.47	23.66	1.00	
13	1	12	22.50	517.50	1658.42	121.93	213.91	169.01	143.04	161.97	39.62	24.46	1.00	
14	1	13	22.50	562.50	1658.82	124.45	219.43	170.19	145.65	164.93	40.86	24.77	1.00	
15	1	14	22.50	607.50	1659.24	124.60	212.76	172.22	144.38	163.49	38.22	23.38	1.00	
16	1	15	22.50	652.50	1659.63	121.23	193.17	163.22	135.32	153.23	31.83	20.77	1.00	
17	1	16	22.50	697.50	1660.06	110.04	183.88	145.64	124.54	141.02	32.09	22.76	1.00	
18	1	17	22.50	742.50	1660.55	112.05	174.86	148.57	123.39	139.72	27.96	20.01	1.00	
19	1	18	22.50	787.50	1661.10	119.06	165.14	142.10	120.78	136.77	21.62	15.81	1.00	

Sheet1 / Sheet2 / Sheet3
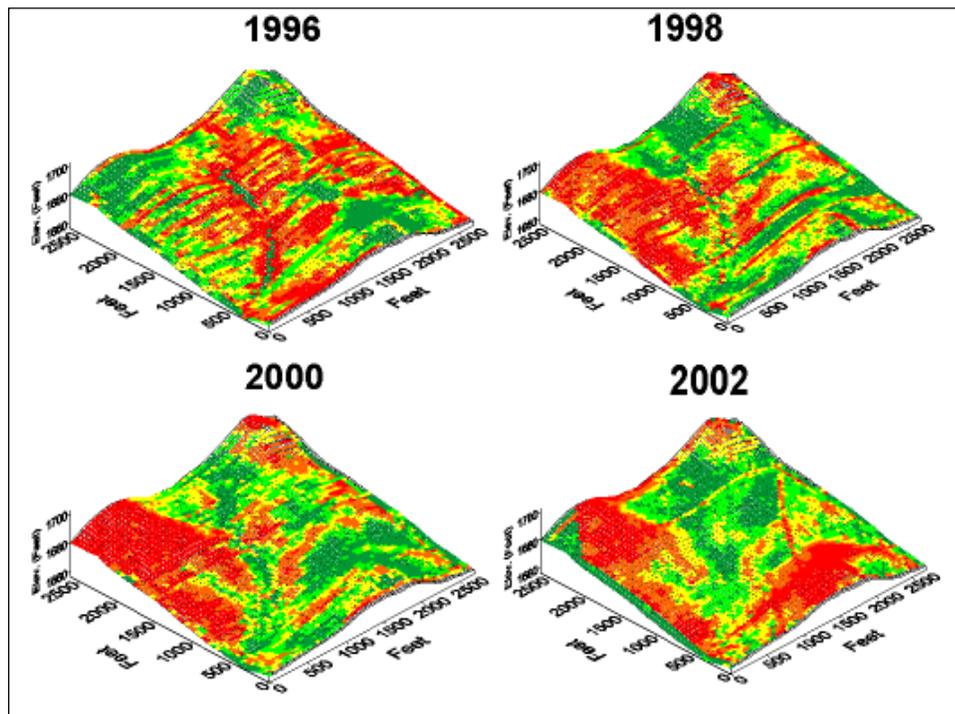
Figure 1.   Four years of corn yield data showing temporal differences in yield patterns.

2) low yield, high deviation, 3) high yield, low deviation, and 4) low yield and low deviation are created.

Results

Method 1: Production zones created from average yields separated the field into areas with low, medium, and high yields. Low yields were observed in the summit or shoulder landscape positions, medium yields were generally observed in the backslope positions, and high yields often were found in footslope or toeslope positions (**Figure 2**). However, this approach does not take into consideration variability extremes that occur on a temporal basis. Areas of high production in one year may have limited production in a different year if different weather conditions prevail. For example, in a wet year, yield in low landscape positions may be poor due to flooded soils. Average yield zones can help producers to set production goals for different field areas and then vary inputs to account for the expected differences in production.

Method 2: Examination of maps produced by the grid-cell standard deviations should improve management. Low standard deviation areas were typically located in summit, shoulder, or backslope areas. Average or medium standard deviation areas are typically located in transition areas such as backslopes. Toeslope and footslope areas typically have a high standard deviation because either very high or very low yields may occur, depending on climatic conditions. Areas with high standard deviation may require additional management. For example, in toeslope areas, tile drainage may be installed to remove excess water. By understanding within-field variability, producers should be able to manage variability and improve production.

Method 3: The coefficient of variation (CV) is another method to delineate zones based on variability. The CV uses the field average yields and standard deviations to create zones. The zones created with CV values are similar to those created by standard deviations (**Figure 2**).

3
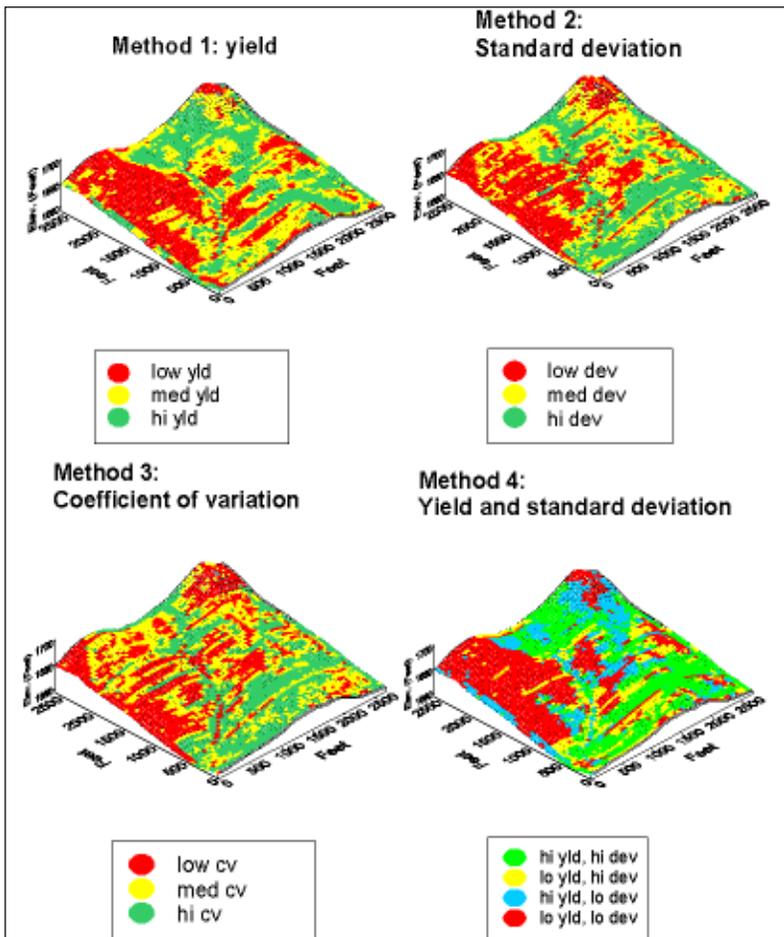
Figure 2. Productivity zone maps created using four different delineation methods.

Method 4: A better understanding of how the average yields and standard deviations in a field interact may be attained if both factors are used to develop zones. In the example field, areas with both high yields and high standard deviations were found in footslope areas that have occasional problems with excess moisture. These areas produce high yields in dry years and low to average yields in wet years. These areas are good candidates for installing tile drainage…provided drainage is allowed by the NRCS or other authorities. Selecting yield goals for these areas is difficult due to the large variability. Footslopes areas in dry years produce corn yields greater than 200 bu/A in this field, whereas yields can be as low as zero/A in wet years. Areas with low yield and high deviation are typically toeslopes that are extremely wet or in compacted regions surrounding wet field areas. High variability observed in these areas may be controlled by limiting compaction, installing drainage, or planting extremely low areas to grass waterways. Field areas with high yields and low deviations were typically backslope areas that had adequate soil moisture and do not flood. Nutrient inputs may need to be increased in these areas due to crop removal from high yields that occur on a regular basis. Field areas that have low yields and low deviations are typically summit or shoulder areas with limited soil moisture holding capacity. These areas may be eroded with low soil organic matter content and limited production capabilities. Inputs may be decreased in the summit or shoulder areas due to low stable yields and corresponding low nutrient removal rates.

Conclusion

There are several ways to use yield data to analyze yield trends both spatially and temporally. Production zones will become more accurate with time as more and more growing seasons are added to the database. Different maps can be used for different purposes. Average yield maps can be used to define yield goals. However, field areas that have high standard deviation in yield will have yield goals that are year-dependent. Standard deviation maps are very useful in identifying areas requiring corrective treatments. By comparing yield and standard deviation maps, the potential yield losses associated with not implementing a corrective treatment can be determined. For example, a 10-acre footslope area with excess water in 1 out of 10 years that results in a 90% yield loss for that season. The monetary loss from that year would be $4,500 (10 acres @ 180 bu/A loss @ $2.50/bu). This situation occurred in this field in 1996.

In this case study, the preferred method for explaining yield variability used a combination of average yields and standard deviation to delineate productivity zones. Due to the yield variability present in most agricultural production systems, producers need to analyze the different zone creation methods for each field and determine for themselves how management decisions may be altered to increase production efficiency. ◼

References cited

Bakhsh, A., T.S. Colvin, D.B. Jaynes, R.S. Kanwar, and U.S. Tim. 2000. Spatio-temporal analysis of yield variability for a corn-soybean field in Iowa. Trans. ASAE 43:31-38.

Blackmore, S. and M. Moore. 1999. Remedial correction of yield map data. Precision Agriculture. 1:53-66.

Chang, J., D.E. Clay, C.G. Carlson, C.L. Reese, S.A. Clay, and M.M. Ellsbury. 2004. Defining yield goals and management zones to minimize yield and N and P fertilizer recommendation errors. Agron. J. 96:825-831.

Diker, K., G.W. Buchleiter, H.J. Farahani, D.F. Heermann, and M.K. Brodahl. 2002. Frequency analysis of yield for delineating management zones. In P.C. Robert, et al. (eds.) [CD-ROM] Proceedings of the 6th International Conference on Precision Agriculture. July 14-17, 2002. Bloomington, MN. ASA-CSSA-SSSA, Madison, WI.

Fleming, K.L., D.G. Westfall, and D.W. Wiens. 1999. Field testing management zones for VRT. In Site-Specific Management Guidelines. SSMG-21. PPI-FAR, Norcross, GA. Available online at >www.ppi-ppic.org<.

Fridgen, J.J., N.R. Kitchen, and K.A Sudduth. 2000. Variability of soil and landscape attributes within sub-field management zones. In P.C. Robert, et al. (eds.) [CD-ROM] Proceedings of the 5th International Conference on Precision Agriculture. July 16-19, 2000. Bloomington, MN. ASA-CSSA-SSSA, Madison, WI.

Kitchen, N., K.A. Sudduth, S.T. Drummond, J.J. Fridgen, W.J. Wiebold, and C.W. Fraisse. 2002. Procedures for evaluation of unsupervised classification to derive management zones. In P.C. Robert, et al. (eds.) [CD-ROM] Proceedings of the 6th International Conference on Precision Agriculture. July 14-17, 2002. Bloomington, MN. ASA-CSSA-SSSA, Madison, WI.

Thylen, L. and P.A. Algerbo. 2000. An expert filter removing erroneous yield data. In P.C. Robert, et al. (eds.) Proceedings of the 5th International Conference on Precision Agriculture, Bloomington, MN. ASA-CSSA-SSSA, Madison, WI.

Walker, T., M. Cox, W. Kingery, S. Martin, L. Oldham, and J. Street. 2004. Defining and managing yield zones for rice and soybeans – a case study. Better Crops with Plant Food. 88:10-13.

**This Site-Specific Management Guideline
was prepared by:**

J.L. Kleinjan
Research Associate
South Dakota State University
Brookings, SD 57007
Phone: 605-688-5105
E-mail: jonathan.kleinjan@sdstate.edu

Dr. David E. Clay
Professor, Soil Science
214 Ag Hall
South Dakota State University
Brookings, SD 57007
Phone: 605-688-5081
Fax: 605-688-4602
E-mail: david.clay@sdstate.edu

Dr. C. Gregg Carlson
Professor, Soil Science
208 Ag Hall
South Dakota State University
Brookings, SD 57007
Phone: 605-688-4761
Fax: 605-688-4602
E-mail: gregg.carlson@ces.sdstate.edu

Dr. Sharon A. Clay
Professor, Weed Science
Plant Science Department
South Dakota State University
Brookings, SD 57007
Phone: 605-688-4757
E-mail: sharon.clay@sdstate.edu