

# The Pitfalls of Constructing Yield Maps

R. Anthony Valvo  
Ernest S. Humphrey  
John G. Lee

Presented at Western Agricultural Economics  
Association 1997 Annual Meeting  
July 13-16, 1997  
Reno/Sparks, Nevada

July 1997

## The Pitfalls of Constructing Yield Maps

R. Anthony Valvo, Ernest S. Humphrey,  
and John G. Lee\*

\* R. Anthony Valvo is a graduate student in the Department of Agronomy, Ernest Humphrey is a graduate research assistant in the Department of Agricultural Economics, John G. Lee is an associate professor in the Department of Agricultural Economics, all at Purdue University, West Lafayette, IN, 47907.

**Abstract**

A yield map can be an important tool in guiding the variable rate application of inputs and in assessing the affects that variable rate input application can have on crop yields for a farm field. This paper identifies and examines several factors which determine the quality of a yield map.

**Introduction**

Precision farming (variable rate input application) is based on the premise that a uniform application of inputs ( including fertilizers, herbicides, and seeds) in a given field can lead to an over and/or under application of inputs to certain areas of the field due to the heterogeneity of various factors within the field. The consequences of the misapplication of inputs can include unnecessary fertilizer expenses, water quality degradation, and depressed crop yields.

An accurate assessment of the possible benefits to precision farming for a given farm field requires the ability to determine the 'most efficient' application rates for inputs across the field. Yield maps can help guide the determination of input management zones for a field, and subsequently, lead to the prescription of more efficient input application rates by allowing the identification of relatively high and low yielding areas of the field. In addition, yield maps can help quantify the affects (or lack there of) that variable rate input application have had on crop yields for a given farm field. The possible benefits of having

a yield map for a given farm field are dependent on the accuracy of the map. An inaccurate yield map may lead to the prescription of inefficient input application rates and/or lead to an inaccurate assessment of the benefits of variable rate input application for a farm field. Therefore, it is important to understand the factors which affect the quality of a yield map. The bottom line is that farmers choosing to commit resources to collect and process monitored yield data should know how to ensure they are going to produce or receive quality information. This paper will describe the issues involved in the construction of a yield map. The paper will suggest how to deal with the systematic errors inherent in the collection of yield monitored data by describing the procedure used to acquire monitored yield data for one thousand acres in Southwestern Indiana and Southeastern Illinois. Finally, the paper will identify and illustrate issues associated with the processing of this data into yield maps.

### **Yield Mapping Issues**

There are many issues involved in the construction of a yield map. There are issues associated with data acquisition, data interpolation, and map representation. Through the collection of yield monitored data the farmer will encounter several sources of systematic error. These errors should be minimized as it is not possible to construct an “accurate” yield map using data which is of “poor” quality. If the farmer has collected data of “adequate” quality the farmer will then need to use a yield software package to create a yield map. The quality of yield map produced by a software package depends on the data interpolation algorithm used to transform the “raw” yield monitored data into a yield map.

In addition, the color scheme and legend used on a “good” yield map for a field should allow anyone viewing the map to easily identify the significant areas of yield variation within the field.

### **Data Acquisition**

Prior to the description of the procedure of how the yield monitored data was collected it is important to identify some of the equipment that will be used in the collection of this data. A tangential flow combine was used to harvest the grain. As the name implies, it does so by combining the functions of cutting and threshing into one machine. A mass-flow sensor was mounted on the discharge end of the clean grain auger of the combine. A moisture sensor was mounted in grain tank of the combine. A 486 CPU was mounted in the cab of the combine. A Differential Global Positioning System (DPGS) and AM band receiver was fixed to the top of the cab. The DGPS signal is corrected to a relative accuracy of approximately one meter. It is important to point out that the use of a DGPS signal minimizes a potentially significant source of error that can occur by the use of an ‘uncorrected’ GPS signal which has only a relative accuracy of approximately three hundred meters. A type one PCMCIA port, which is part of the CPU, was used to store to the files containing the “raw” yield monitored data.

The farm selected for this study is a one thousand acre corn and soybean farm located in west central Indiana, the name and exact location are withheld from this paper at the owners request. The field from which yield monitored data was collected that will subsequently be used to illustrate the issues associated with yield monitored data

processing is rectangular and approximately 180 acres in total area. The standing crop in this field was “Number Two Yellow Dent Corn”.

Before the actual grain harvest, the combine’s data acquisition equipment was tested to insure a continuous DGPS signal was being received. This was accomplished by driving the combine around the farmstead with the “corn-head” down. The reason the corn-head is in the down position, is that the electronics are setup in such a way that when the header is raised the sensors and the DGPS receiver shut off. The farmer who was to drive the combine used to collect the data was made aware of the two sources of systematic error that he would encounter when collecting the yield monitored data. He was told the first source of error he would face would be termed “lag time” which was the period between the time the head is lowered and the time it takes to fully load the combine with grain. He was made aware of the fact that evidence of this error could be seen in areas where the combine enters a standing crop, and by the alternate low spots at the end of the field. The farmer was instructed to try and minimize this error by isolating the direction of combine travel and harvest patterns. The second source of error the farmer was warned about could be introduced from unknown crop width. This problem arises when the header does not have a full width of standing crop entering it. The problem occurs when the combine overlaps a previously cut area. The equation used to calculate the dry yield weight, assumes the header is always full. Therefore, the farmer was instructed to be very diligent in navigating the combine on each field.

Once we were satisfied the farmer understood why we were concerned with these types of errors, how we were trying to deal with them, and that the equipment was

operating properly, we did our last minute calibrations and adjustments. These calibrations included, the selecting of a time interval in which each DGPS fix or position report, would be recorded. For this project one second fixes were used. The “field name” and “crop type” were entered into the on-board computer prior to beginning harvesting corn.

At one second time intervals, and an average combine speed of five kilometers/hour, the computer recorded a “dry bushel per acre yield” every seven and half feet as the combine traverses the field. This is an average lag time and will differ depending on combine speed, crop type, and crop moisture. The yield attribute was calculated by subtracting the moisture value from the mass flow reading. These values were being fed continuously to the on-board computer, but are only time stamped with a DGPS fix at intervals of one second. In summary, the computer is recording a latitude and longitude every second and is attaching an attribute of dry grain weight to that geographic location. This information is being stored on type one PCMCIA storage media located in the on-board computer.

Once the field was cut, the combine was brought back to the farmstead and the PCMCIA card was removed from it. The PCMCIA card was inserted into a card reader on the MS-DOS computer detailed earlier. The raw data was then downloaded to a software program which allowed viewing of the raw data and exporting it out in different display options and various formats. For this study, ‘comma delimited, ASCII text’ format was used. The output display was manipulated so that the data stream read latitudes in the first column, longitudes second, yield attributes third and any headers were

removed before further processing . This data structure and format established an absolute or Euclidean coordinate geometry system for X, Y and Z or latitude, longitude and Attribute respectively.

As mentioned earlier, ASCII text is the format we used and is the most widely accepted by the majority of yield mapping software, and other operating systems. The authors wish to point out that prior to the importation of the ‘raw’ yield data into a yield mapping software package the data would have needed to be inspected for ‘spurious’ GPS readings (if the yield mapping software the authors chose did not provide this function). The process of ‘throwing out’ spurious points is called ‘clipping’ the data. A complete discussion of the issues involved in the ‘clipping’ of yield monitored data is in contained in Rands (1995) and Saunders, et. al (1996)<sup>1</sup>.

### **Data Processing Issues**

It is important for the farmer (or crop consultant) to know that all types of yield mapping software do not use the same algorithm to transform the ‘raw’ yield data from a field into a yield map. Unfortunately, there is no consensus as to which algorithm is the “best” algorithm to use when creating yield maps from yield monitored data. However, it is important to know the types of algorithms which can be used to create yield maps. The farmer (or a crop consultant or a colleague whom the farmer trusts) should know and be comfortable with the type of algorithm used by a mapping software before committing his resources to the purchase of the software.

---

<sup>1</sup> It is important to note that even if a DGPS signal is used there is still an inherent ‘wandering’ error in the signal which can cause locational readings which are outside the boundaries of the field.



The algorithms most commonly used can be categorized as nearest neighbor, inverse distance, exponential decay, or polynomial trend surface. The premise of the nearest neighbor, inverse distance, and exponential decay algorithms is to estimate the value of yield at a point (location) for which the raw yield data provides no estimate of yield as a function of all of the raw yield data points (for which there exists an estimate of crop yield) which are located within a certain radius of the point (j). A polynomial trend surface uses least squares to fit a polynomial to the area of a given field. The polynomial is a function of the location of the point at which yield is being estimated expressed in Euclidean coordinates. Equations (1)-(4) represent different equations used by each of the aforementioned methods to estimate crop yield at a point (j) at which the raw yield data provides no yield estimate. The first equation is a nearest neighbor method where yield at each location is calculated as:

$$(1) \quad \bar{Y}_j = \sum_{i=1}^n \lambda * Y_i$$

where  $Y_i$  for  $i= 1, \dots, n$  denotes all the yield data points within a radius of influence ( $rd$ ) of point  $j$ ,  $\lambda$  is a weighting factor equal to the inverse of the number of points ( $n$ ) within the radius of influence.

An inverse distance specification is calculated as follows:

$$(2) \quad \bar{Y}_j = \sum_{i=1}^n \left( \frac{Y_i}{D} \right)$$

where  $Y_i$  for  $i= 1, \dots, n$  denotes all the yield data points within a radius of influence ( $rd$ ) of point  $j$ ,  $D$  is the distance between point  $i$  and point  $j$ .

The exponential decay formulation is given by:

$$(3) \quad \bar{Y}_j = \sum_{i=1}^n \left( \frac{Y_i}{e^D} \right)$$

where  $Y_i$  for  $i=1, \dots, n$  denotes all the yield data points within a radius of influence ( $rd$ ) of point  $j$ ,  $D$  is the distance between point  $i$  and point  $j$ ,  $e$  is a constant.

The final method considered in this study is a polynomial trend surface represented by:

$$(4) \quad \bar{Y}_j = \sum_{i=1}^n aX_i + bZ_i + cX_iZ_i$$

where  $(X_i, Z_i)$  represent the location of point  $j$  in Euclidean coordinates, and  $a$ ,  $b$ , and  $c$  represent parameters estimated from an ordinary least squares procedure.

Figures (2)-(5) are yield maps constructed using the Vision Professional software by Rockwell of 1996 corn yield for the 180 acre field in Southwestern Indiana using equations (1)-(4), respectively, using a radius of influence of twelve meters.

Figure 2. Nearest Neighbor  
1996 Corn Yield

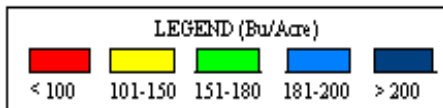
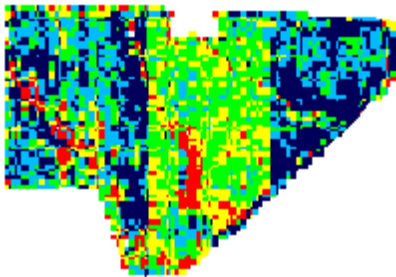


Figure 4. Exponential Decay  
1996 Corn Yield

Figure 3. Inverse Distance  
1996 Corn Yield

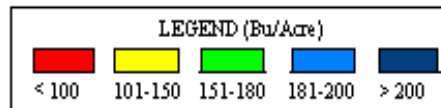
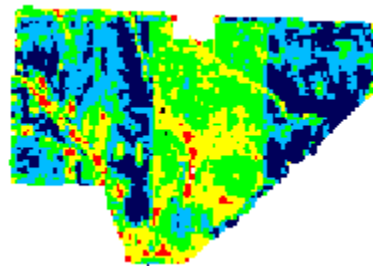
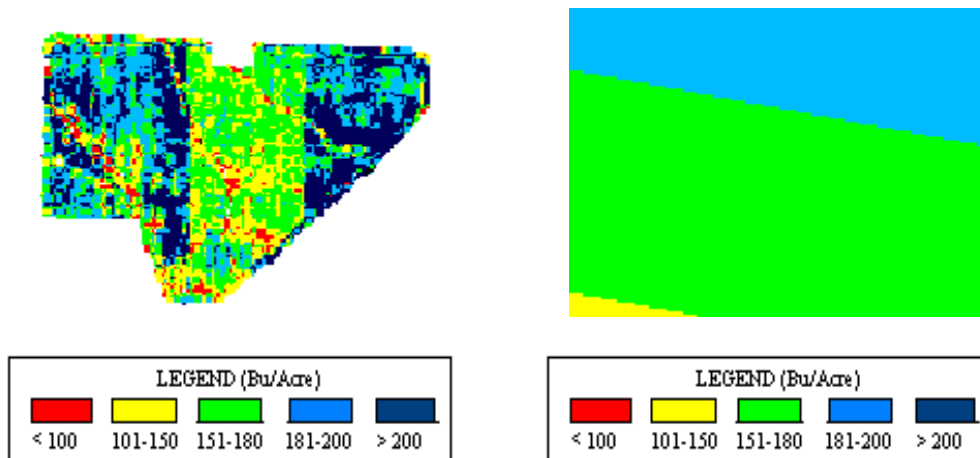


Figure 5. Polynomial Trend  
1996 Corn Yield



Using figures (2)-(5) we can illustrate the possible consequences that the choice of data interpolation algorithm can have for a farmer who uses a yield map to guide variable rate input application or to assess the affects that variable rate input application had on the crop yields for a field. Initially, let us assume the farmer wants to use a yield map to guide the application of  $P_2O_5$  fertilizer across the one hundred and eighty acre field, and that he wants to apply a specific rate to all areas of the field on which the crop yield was less than or equal to one hundred bushels per acre. The nearest neighbor method used to construct the map in figure 2. would direct the farmer to apply this rate to 12.08 acres of the field. The exponential decay method (figure 3.) would direct the farmer to apply the rate to 9.20 acres of the field. The inverse distance method (figure 4.) would recommend the farmer apply the rate to 5.02 acres while the polynomial trend would direct the farmer not to apply this application rate on the field. Now assume the farmer applied a higher rate of  $P_2O_5$  fertilizer application to the center of the field and chose to use a yield map to assess the affects that the higher rate of application of  $P_2O_5$  had on the crop yields for that area of the field. The choice of interpolation algorithm would affect his interpretation of the

magnitude of the changes in crop yields. The use of a polynomial trend would lead him to infer the largest gains in yield from increased fertilization while the use of the nearest neighbor method would cause the farmer to infer the lowest changes in crop yields. It is beyond the scope of this paper to suggest which is the ‘best’ data interpolation algorithm. However, it is apparent that the choice of data interpolation algorithm can affect decisions made by the farmer which are based on the information provided by the yield map.

It important to make clear at this point that the use of a nearest neighbor, inverse distance, or an exponential decay yield mapping algorithm entails the choice of a radius of influence ( $rd$ ) . Therefore, even if two yield mapping software packages use the same type of algorithm the maps produced by the two packages using the same data may be different if different radii of influence are used by each package in creating yield maps. This point is illustrated in figures (6)-(7). Each is a yield map of the 180 acre field generated using a nearest neighbor algorithm. However, the radius of influence used to generate the map in figure (6) is five meters while the radius of influence used to generate the map in figure (7) is two hundred meters.

Figure 6. Nearest Neighbor  
five meter radius of influence

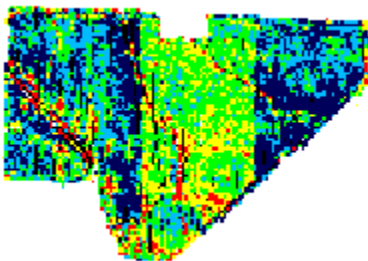
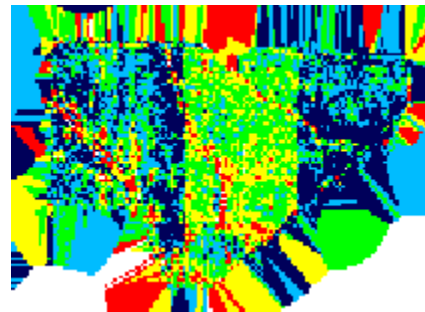
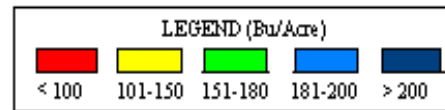
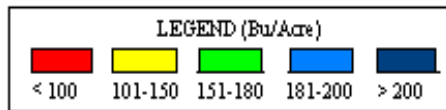


Figure 7. Nearest Neighbor  
two hundred meter radius of influence





The radius of influence used to generate the map in figure (6) is too small as use of this radius fails to interpolate points encompassing the total area of the field (the black areas in figure(6)). The radius of influence used to generate the map in figure (7) is too large as use of this radius interpolates yield values for points whose location is outside the boundaries of the field. The bottom line is that the choice of the radius of influence can affect a yield map. Consequently, it is important for a person choosing yield mapping software to know the radius of influence used in the yield mapping algorithm as well as the type of the algorithm itself. The radius of influence used in creating a yield map should be consistent with the manner in which the yield monitored data was collected by the farmer.

There are also issues to consider concerning the features of a hard copy output of a yield map produced by a yield mapping package. There a few basic questions which need to be answered when examining a hard copy output of a yield map. Does the color scheme of the map clearly identify the variation of yield within the field? Does the legend let one easily identify the variation in crop yield which is important to the user of the yield mapping software?

Examining figures (2)-(4) it is clear that the color scheme and legend used by each yield map clearly allow the identification the areas of significant yield variation within the field. The Vision Professional software used to create these images also gives us the ability to change these parameters. In addition, recall that the radius of influence used in

constructing each map was twelve meters which was consistent with the combine settings used to collect the “raw” yield data. Therefore, the maps illustrated in figures (2)-(4) created using the Vision Professional software are what the authors would consider “good” quality yield maps.

It is clear from examining figure (5) that it does not clearly identify any areas of significant variation within the field which the authors knew existed by examining the “raw” yield data for the field. The map fails to correctly identify even the distinction between the highest and lowest yielding area of the field. Therefore, the map illustrated in figure (5) is an example of what the authors would consider a yield map of “poor” quality.

### **Summary**

This paper has identified and illustrated many of the problems associated with the collection and processing of yield monitored data. Through the description of a procedure used to collect data from a 180 acre farm in Southeastern Indiana issues of data quality were identified and methods to minimize these types of errors were suggested by the authors. Finally, issues associated with yield data interpolation and map presentation were identified and illustrated using yield maps constructed using the Vision System software.

### **Conclusions**

Yield maps can be important tools in guiding the variable rate application of inputs and quantifying the affects of variable rate input application on crop yields. Therefore, it is important to consider the factors which determine the quality of a yield map. There are issues associated with data collection, data interpolation, and map output. The farmer needs to understand the importance of minimizing the error propagation inherent in the

collection of yield monitored data. The farmer must realize that without exception bad data equals useless yield map. The farmer need not worry about how a yield map is created if he has collected yield monitored data which is of 'poor' quality. If the farmer has collected 'good' data he must then choose a yield mapping software package to create a yield map from his 'raw' yield monitored data. This software should allow for the use of an algorithm for the interpolation of yield data which is consistent with how the raw yield data was collected by the farmer. In addition, the software should allow the use or choice of a color scheme and legend for an individual yield map which makes the variation in crop yield for the field easily identifiable for anyone who views the map. If the farmer is to realize any benefits from the investment of his resources into the collection and processing of yield monitored data he must take steps to ensure the quality of the yield maps he constructs or receives from a crop consultant.

## References

- Blackmore, B. S., and C. J. Marshall. 1996. "Yield Mapping; Errors and Algorithms". *In Proceedings of the 3<sup>rd</sup> International Conference on Precision Agriculture, June 23-26, 1996, Minneapolis, MN.*
- Frazen, David W., and Ted R. Peck. "Field Soil Sampling Density for Variable Rate Fertilization". *Journal of Production Agriculture*. 8(1995):568-574.
- Gotway, C. A., R. B. Ferguson, G. W. Hergert, and T. A. Peterson. "A comparison of kriging and inverse-distance methods for variable rate nitrogen application". *Soil Sci. Soc. Am. J.* 60(1996):1237-1247.
- Rands, M. 1995. "The Development of an expert filter to improve the quality of yield mapping data". Unpublished MSc. Thesis, Silsoe College, Cranfield University.
- Rockwell Vision System Software. Professional Version 2.9B. 1996. Rockwell and Delta Data

Systems, Inc. Rockwell Agricultural Systems, Cedar Rapids, IA.

Rockwell Vision System Software User's Guide Advanced and Professional.1995.

Rockwell

and Delta Data Systems, Inc. Rockwell Agricultural Systems, Cedar Rapids, IA.

Saunders, S. P., G. Larscheid, B.S. Blackmore, and J. V. Stafford. 1996. "A method for direct

comparison of global positioning systems suitable for precision farming". *In Proceedings of the*

3<sup>rd</sup> International Conference on Precision Agriculture, June 23-26, 1996, Minneapolis, MN.

Wollenhaupt, N. C., R. P. Wolkowski, and M. K. Clayton. "Nutrient Management:Mapping Soil

Test Phosphorus and Potassium for Variable-Rate Fertilizer Application." *Journal of Production Agriculture*. 7(1994): 441-448.