PRECISION FARMING RESEARCH: EVALUATION OF YIELD MONITOR DATA IMPORTANCE TO NASS, by Paul W. Cook, Spatial Analysis Research Section, Geospatial Information Branch, Research and Development Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. 20250-2000, October 2001, NASS Research Report RDD-01-xx.

ABSTRACT

The National Agricultural Statistics Service (NASS) currently uses farmer-reported (list-frame) based procedures to make county and state-level crop yield estimates. NASS also collects objective yield data on small field sites as a complement to the farmer provided information. Two crops of particular interest for crop yield determination are corn and soybeans. In conjunction with Purdue University, NASS collected yield monitor data and soils' variables associated with objective yield plots chosen within two selected fields of corn and soybeans at the Purdue Davis Research Farm in east-central Indiana.

This study provided an analysis of relationships of corn and soybeans yields with soils' information and compared the objective yield data with the yield monitor data. Crop yields for corn and soybeans did vary within the fields for the yield monitor data. The yield monitor yield levels were not always in agreement with the objective yield estimated crop yields, particularly for corn. Yield monitor data was not found to be of importance for direct use by NASS. The Economic Research Service (ERS) of USDA has recommended questions about yield monitors be incorporated into future surveys. Precision farming procedures that aid in greater uniformity in crop yields within fields should help NASS objective yield procedures be more reliable over time. However, there are also many other changes occurring in crop production practices, such as new crop varieties, that will be a challenge to the objective yield program.

KEYWORDS

Geographic Information Systems(GIS); Global Positioning System (GPS); Objective Yield; Precision Farming; Soils' Variables; Yield Monitor.

ACKNOWLEDGMENTS

My special thanks go to George Hanuschak, Branch Chief of NASS's Geospatial Information Branch for his efforts to arrange the agreement with Purdue University, oversee many aspects of the project as needed, travel to help with laying out the NASS samples, and ensure that I obtained needed training. My thanks go as well to Ralph Gann, State Statistician, Indiana for his support of the research with Purdue University and leading the data collection effort of the NASS objective yield data collection for the two study fields at Purdue Davis Farms. For technical assistance, my thanks go to Martin Ozga, Research and Development Division, for FLDHARV, the program that made preliminary corrections to the yield monitor data, Jim Burt, Research, for preliminary studies of the yield monitor data, Keith Morris, Purdue University, for obtaining GPS locations for the NASS samples, Tim Keller, Sampling and Estimation Research Section, for providing information on objective yield procedures and current research efforts, and Gail Wade and Jeff Johnson for suggestions to make better use of ArcView 3.2, an ESRI Geographic Information System (GIS) computer program. The Illinois State Office did the two fields' laboratory work on the NASS objective yield sample data during 1997 and 1998. Bob Hale, Head, Spatial Analysis Research Section, and Charles Day have provided a careful review and editorial recommendations that have helped improve my paper.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This report documents a small pilot study by NASS staff to become familiar with yield monitor data. The USDA Under Secretary of Research, Education, and Economics (REE) authorized NASS to conduct this study. This Summary Section and the Discussion Section provide a complete executive summary. The remaining sections of the report provide an analysis of one year of corn and soybeans data for two fields within the Purdue University Davis Research Farm. The limited quantity of data prevents any statistical inferences regarding yield monitor data characteristics outside the two fields studied.

Estimating crop yields is an important part of the NASS mission. Changes in farming practices have increased crop yields, but crop yield estimation relies on farmer-reported (list frame based) estimates for both county and state-level crop yield estimates. Objective yield procedures use a sample of small hand harvested plots of crop data to provide an important indication of Regional and State level yield estimates.

The intention of precision farming research is to develop new technologies to monitor and reduce within-field crop yield variability. Their goal is to reduce the costs of farming while improving the crop yields. Developers of precision farming procedures use the latest in available technology in four areas:

1) Development of yield monitors, often with positional capabilities (Global Positioning System (GPS)) to locate their observations and include crop moisture levels, protein, and water stress (AgLeader Technology, 1997),
2) Evaluation of improved variable rate technology to deliver water, seeds, nutrients or chemical sprays only where needed within the field thereby reducing waste and improving efficiency (Rawlins, 1996),
3) Production of Geographic Information Systems (GIS) maps displaying the field's crop yield variability to develop a plan for variable rate technology devices (Blackmore, 1999), and
4) Production of more detailed soils' maps to provide soil type and nutrient differences within a field.

The focus of this paper is to provide an understanding of the basics of precision farming research and its importance to NASS in setting crop yields. This goal requires examining an example of precision farming using yield monitor data (category one above). This evaluation provides a better understanding of the strengths and weaknesses of yield monitor data for evaluating crop yields. This report analyzes the yield monitor data collected by Purdue University in 1998 for two fields at the Purdue Davis Research Farm. An analysis of NASS objective yield sample data and soils' data for the two fields are an integral part of the analysis as well.

Three general conclusions regarding the yield monitor data and its importance to NASS are evident from these analyses. First, the large quantity of data required for analysis and low rate of adoption of yield monitors would argue that NASS will not find yield monitor data to be an

important data set for U.S. or State level crop yields at this time. Second, lack of data collection uniformity among various yield monitors, potential analysis errors, and complexity of yield monitor data prevent its use in crop yield surveys. Finally, yield monitor data cannot substitute for weigh wagon, crop marketing (sales tickets), or on-farm storage volume information because its development is not sufficiently advanced.

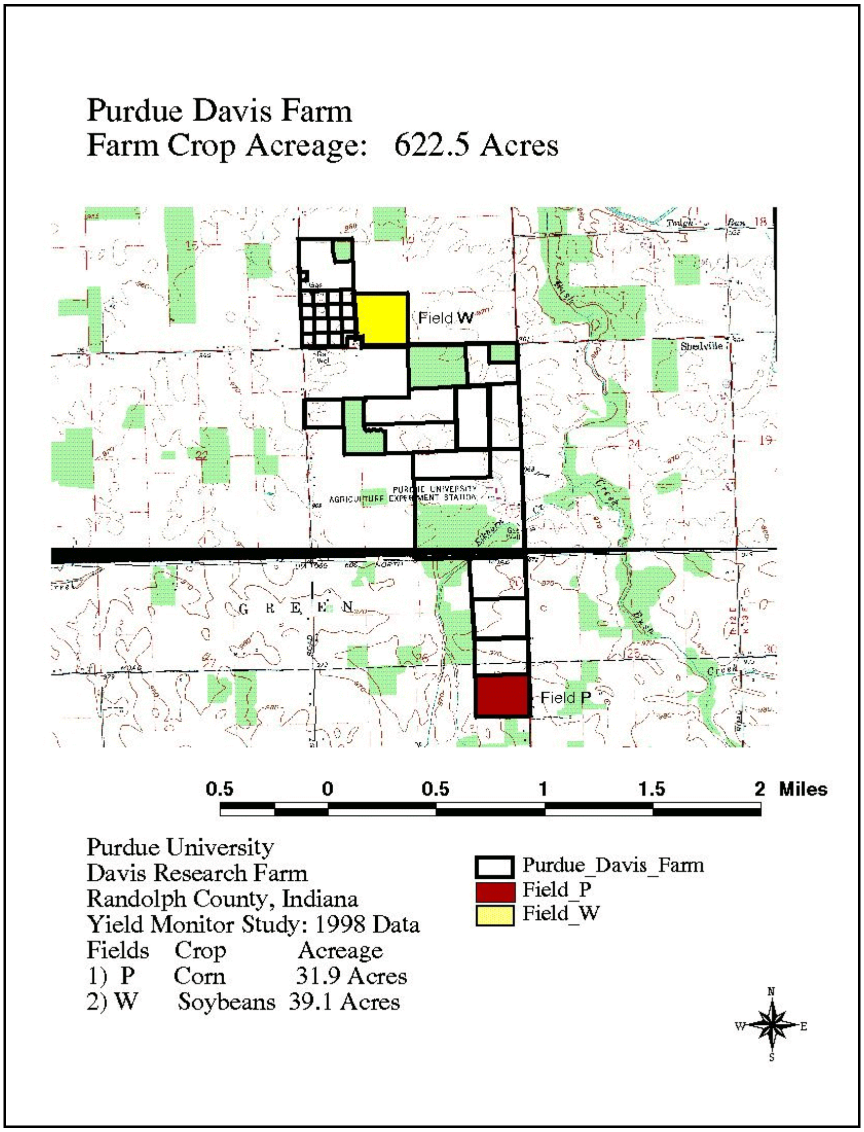
1. BACKGROUND

This project combined analysis of Purdue University yield monitor and soils' data with NASS objective yield data. NASS uses the objective yield procedures with list frame data to estimate State level crop yields. Specifically, NASS will harvest small portions of grain from two plots, typically for 300 fields, to provide early indications of State-level crop yields. Also, NASS mails questionnaires to a list frame sample of cooperating farmers to confirm a State's crop yields. Use of these surveys provides NASS a method of making both forecasts and estimates of crop yields close to the time of harvest.

Measuring the soils' nutrient availability through soils sampling is an important component of precision farming methodology. When the soils' Organic Matter (OM), Phosphorous (P), Potassium (K), Magnesium (Mg), Calcium (C), pH, or Cation Exchange Capacity (CEC) are outside accepted levels for a particular crop, the farmer will take corrective action to maintain future crop yields at optimal levels. Evaluation of the two fields soils' characteristics is a part of this study. However, one year of data does not permit examination of the soils' effects on crop yields.

Precision farming is a recently developed field of study; however, university researchers and commercial interests have already published many journal articles. Many authors praise precision farming procedures for improving crop yields with reduced farming costs. Others discuss potential problems in using the data from precision farming. This report focuses on the latter.

Yield monitors are electronic devices measuring crop yields from harvesters at intervals of one second to five seconds. GPS receivers can provide associated latitude and longitude coordinates from satellites. Additional instruments can provide other crop measurements such as moisture percentage. Since the yield monitor requires calibration of its sensors for accurate readings, the yield monitor data can exhibit irregularities. Typically, yield monitors require calibration with weigh wagon records for sub-field areas (Ag Leader Technology, 1997). The literature documents many problems with using yield monitor data. For example, *Precision Agriculture, Proceedings of the 3rd International Conference* (1996) has articles about corrections necessary to make yield monitor data of value. Of course, the yield monitor data can help the farmer focus on sub-field areas where crop yields need improvement. One particular Midwestern benefit has been the identification of soil drainage problems due to broken tiles that require replacement.



NASS initiated an agreement in 1996 with Purdue University to examine the capabilities and potential value of precision farming data in our work. The primary emphasis of the studies would be yield monitor data. Five professors with the Agriculture Engineering Department of Purdue University were the primary researchers cooperating in the research. Dr. Dan Ess was the lead researcher during the agreement's early years. Dr. Sam Parsons assumed that role in early 1998. The remaining three professors were Dr. John Trott, Director, Purdue Agricultural Centers, Dr. Chris Johannsen, Director, Laboratory for Applications of Remote Sensing, and Dr. Gaines Miles, Assistant Professor of Agricultural Engineering. A group of Purdue graduate students worked with the project as well. Two students, Patrick Willis and Montie O'Neal, wrote papers (Willis, 1999 and O'Neal, 2000a and 2000b).

Purdue University chose the Davis-Purdue Research Farm (DPAC) in Randolph County, Indiana (near Muncie, Indiana) as the study's site. The farm contains 622.5 acres of land with fields of corn and soybeans (Figure 1). Although many fields were the focus of research, Purdue University shared two years of yield monitor data for fields' P and W (Figure 1). NASS collected objective yield samples from both fields during 1997 and 1998. Purdue University obtained detailed soils' data in both 1994 and 1998 and collected yield monitor data during 1997 and 1998.

Purdue University collected data from other sources including weigh wagon, remote sensing, and weather data. Patrick Willis (Willis, 1999) analyzed the available remote sensing data and yield monitor data for both fields. Montie O'Neal (O'Neal, 2000a) evaluated weather effects on crop yields for many fields within the Purdue Davis Research Farm. His second paper analyzed possible yield monitor errors (O'Neal, 2000b). This paper studies relationships among the 1998 yield monitor data, the NASS objective yield samples, and the two soils' samples (1994 and 1998) for fields P and W.

2. DATA COLLECTION

The Purdue Davis Farm obtained an Ag Leader yield monitor in 1996 to collect crop yield data. NASS had suggested obtaining traditional objective yield (Huddelston, 1978) point sample data for fields P and W at Davis during 1997 and 1998 to compare against the yield monitor data. Purdue shared data from 1997 and 1998 from fields P and W with NASS. There were a number of problems that led NASS not to use the 1997 data. First, a GPS antenna positioning error led to many points with clearly incorrect latitude and longitude values. Indeed, many points had two or more yield values. Second, increased natural variability caused by poor weather conditions (O'Neal, 2000) made identification and correction of data collection errors with the desired degree of confidence difficult. These data problems with GPS errors and the atypical growing conditions made the 1997 data of little value to NASS in evaluating the yield monitor data. Fortunately, both fewer GPS errors and more typical growing conditions made the 1998 data of value to NASS. Consequently, this report focuses on an analysis of the Purdue 1998 yield monitor (Figure 1) and associated data sets.

2.1 PURDUE UNIVERSITY DATA COLLECTION EFFORTS

Purdue University collected three sets of data associated with this project: yield monitor data from the Ag Leader yield monitor, soils' variables data for 1994 and 1998, and weather data from on farm weather stations. This report does not analyze the weather data. University researchers also assisted with collection of the NASS objective yield site data by providing GPS locations of the selected locations within the fields and with laying out the samples.

2.1.1 YIELD MONITOR DATA: FIELDS P AND W

The yield monitor data provided for each geographic point is not only from that area. The harvester must fill the hopper before a reading (lag), and then the readings remain low as the hopper fills (ramp-up). As a pass ends, the yield monitor readings exhibit decreasing readings (ramp-down). Unequal grain movement can prevent associating areas from two consecutive previous GPS readings. Each individual reading can contain portions of grain from many areas within the harvesting path. High accuracy differential GPS using a stationary GPS unit can provide locations accurate to within five meters (Blackmore, 1999). Movement of the tractor at varying speeds as the operator adjusts for changes in the field conditions can decrease accuracy. Improved procedures continue to increase the positional accuracy (Rawlins, 1996). Further automation of the harvester and other agricultural equipment will aid in these efforts.

Other errors with the collected data can cause inaccuracies. These errors include the following: flyers (stray points from incorrect GPS readings), an overlay of consecutive points (from incorrect GPS antenna placement) and the equipment's inability to measure blank spots. Montie

O'Neal (2000a) has recently catalogued errors inherent in the yield monitor data. Each data set can exhibit all possible errors, but both the frequency and severity of errors can vary. The header not remaining full can cause fluctuations in the measured weight of grain during harvest. This error is difficult either to detect or correct.

This reduced accuracy of the yield monitor data from errors does not invalidate the estimates of yields obtained from the yield monitor data. However, effective use of the yield monitor data requires careful checks against weigh wagon data. Of particular interest is the scale at which the yield monitor data is of value. Some researchers have evaluated the data at the one-meter level (Willis, 1999). However, characteristics of the data that suggest aggregating to a larger area for evaluation are the following:

1) Inaccuracies from GPS errors (Nolan, 1996) and analysis of the Purdue data,
2) Possible miscalibrations of the yield monitor sensors preventing accurate measurement of grain weight and moisture (Missotten, 1996),
3) A partially full header with no objective data to verify it to be full, as with Purdue, and
4) Considerable variations both in location and grain weight data (as with fields P and W).

The complexity of the yield monitor data prevents concluding from two fields what would be an ideal sized area to consider. Not all authors agree as to an ideal sized area to aggregate the yield monitor data. For example, O'Neal (2000b) chose 9.13 meter grid cells corresponding to two passes of the harvester with a fifteen-foot header.

One author (Dunn, 1998) established how rapidly the yield monitor would need to adjust in measuring field areas containing zero yield, normal yield, and 1.5 times normal yield for corn. He concluded that a minimum size area would be a 100 foot grid cell that is equivalent to 30 meter grid cells. Another author (Nolan, 1996) stated larger areas provide greater accuracy. His estimate was that calculated yields are within 5% for a 400 square meter area (that is, a 20 meter grid cell). A 30 meter grid cell should provide an accuracy within that range,

Data collection accuracy limitations make selection of a larger area than that of the collected yield monitor data to aggregate the yield monitor data essential. One goal was to avoid correcting data locations beyond the possibility of evaluating the accuracy of corrections. A second goal was to make the evaluation of the aggregated data more meaningful. Since LANDSAT TM data has a 30 meter pixel size and has been successful in crop acreage estimation, a 30 meter aggregation area for yield monitor data would seem to be reasonable. Many yield monitor datasets would be needed to evaluate the strengths and weaknesses of other grid cell sizes. The discussion above indicates that a 30 meter grid cell will at least provide sufficient accuracy to make evalaution at that size meaningful. Therefore, this study uses only a 30 meter grid cell (900 square meters).

2.1.2 COLLECTION OF SOILS' INFORMATION: FIELDS P AND W

The first soils' mappings for fields P and W were in 1994. Top Soil Testing Service (Top Soil, 1994) provided a mapping of the basic soils types for the farm. Also, they collected information on pH, Phosphorous (P), Potassium (K), Organic Matter (OM), and Cation Exchange Capacity (CEC). Field P had 63 one-half acre samples while field W had 16 2.5 acre grids that provided 16 sample points.

To examine variability of yields within the fields, Purdue University obtained a second set of soils' data for various nutrients during the 1998 growing year (A. and L. Laboratories, 1998). This soil's sampling was only for fields P and W. A. and L. Laboratories (1998) collected information on pH, P, K, OM, and CEC. Their approach included GPS latitude and longitude locations. This sample point data made possible creation of gridded soils' maps for the nutrients.

Top Soil Testing Service included maps of the following soils' nutrients: pH, P, K, and OM. The maps gave 0.5 acre grids for field P and 2.5 acre grids for field W. They provided tables of needed limestone, P, and K to improve yields. Their analysis gave targeted optimums for P and K levels as well.

Purdue's 1998 data were gathered before the harvest of fields P and W and so provided information about the soils' conditions during the crop's growth. The data set for field P had 63 data points in a uniform grid across the field (Figure 2). Many points were near the NASS objective yield sample points. The 1994 data collection provided 0.5 acre grid cells to correspond with the 1998 data collection points. Table 1 (below) provides soils' descriptive statistics for field P for both years. S-Plus (MathSoft, Splus 2000, Seattle, Washington) software was used for all tables and calculation of statistical comparisons in this report.

Table 1. Descriptive Statistics for Purdue Davis Farm Field P, 1994 and 1998 Soils Variables:

Variables[1]	OM_94	P_94	K_94	pH_94	CEC_94	OM	P	K	Mg	C	pH	CEC
Min	2.1	8	70	5.6	13.9	2.2	7	70	245	950	5.4	9.4
Max	5.4	179	282	7.9	39.0	4.9	97	296	775	4750	8.2	28.1
Mean	3.5	39.6	163.8	7.0	22.4	3.2	42.2	142	566	2148	7.1	16.5
S. Dev.	0.75	26.1	50.9	0.56	5.7	0.64	26.5	44.9	110	745	0.72	3.97

[1] OM_94 = Organic Matter_1994, P_94 = Phosphorous_1994, K_94 = Potassium (K)_ 1994, pH_94 = pH_1994, CEC_94 = Cation Exchange Capacity (CEC)1994. The 1998 soils' variables were OM = Organic Matter (OM), P = Phosphorous (P), K = Potassium (K), Mg = Magnesium (Mg), C = Calcium (C), pH = pH Reading, CEC = Cation Exchange Capacity (CEC). The soils' variables are from a 63 location soil's sample from the 1994 and 1998 seasons.

Five paired t-tests evaluated the same soils' variables relationships between the two years of data collection. The t-tests compared Field P's soils pH, OM, P, K, and CEC variables between years.

Since multiple t-tests were necessary, this report employs the use of Bonferroni correction to reduce the risk of a significant test occurring at random. Based on five t-tests with the correlations among the variables given in Table 3, a nominal p-value of 0.0178 provided a true 0.05 probability of a Type I error (SISA) for each test.

Table 2 presents the results of the paired t-tests. The tests indicate that the soils' pH (pH_94 and pH), OM (OM_94 and OM), and soils' P (P_94 and P) are similar for 1994 and 1998. However, K and CEC do differ significantly, allowing the conclusion that the soil properties have changed . Therefore, one component of the improved yields obtained in 1998 might come from improvements in overall soils' properties. However, as pointed out by Mallarino (Mallarino, 1999), the soils' variables explaining yield variability will likely differ across fields. Therefore, making conclusions about what soils' variables to sample will remain a difficult choice when making relationships to crop yields. Of course, weather for 1998 (O'Neal, 2000a) provided remarkably better rainfall than in 1997. However, providing a more detailed evaluation of weather data is outside the scope of this report.

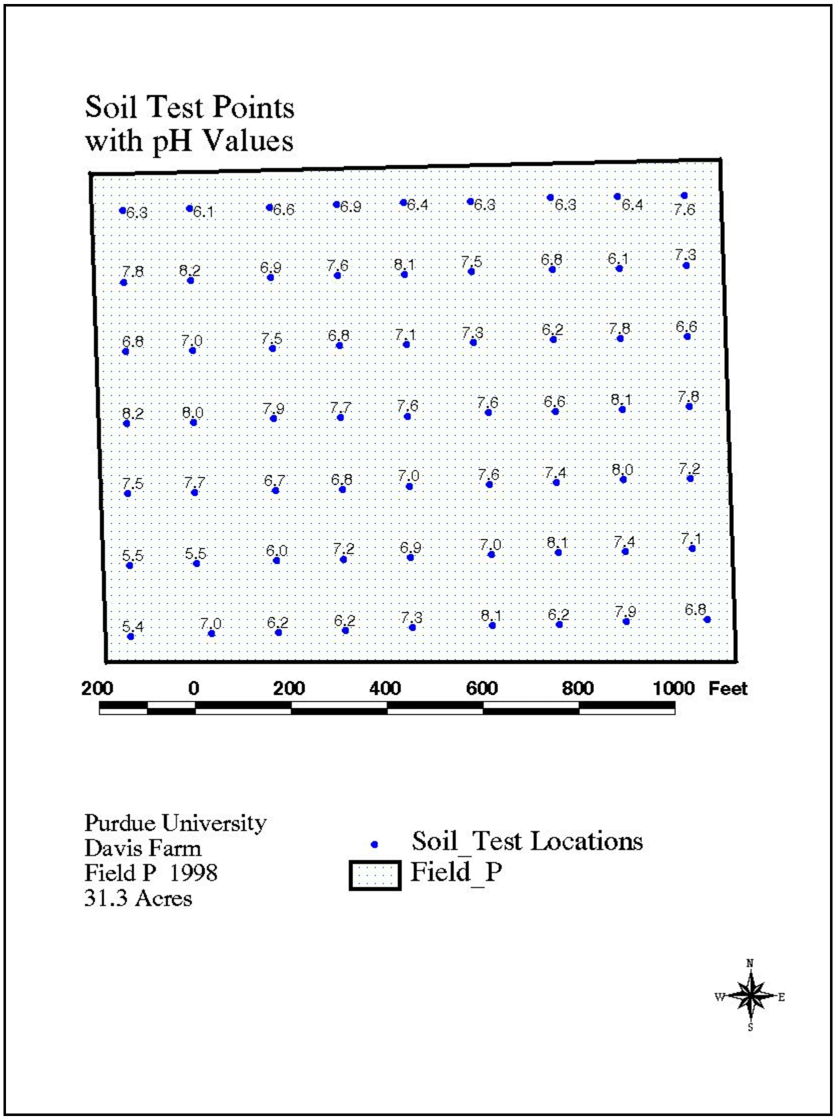
Table 2. Paired t-test results for Purdue Davis Farm Field P, 1994 and 1998 Soils Variables

Paired t-Test Results: 1994 v.s. 1998 Field P Soils Data[1]					
	pH	OM	Phosphorous(P)	K	CEC
t-Value	1.1024	-2.4142	0.6591	-2.9232	-7.2487
df	62	62	62	62	62
p-value	0.2745	0.0187	0.5122	0.0048	0
Min	-0.09	-0.53	-5.2	-37.6	-7.6
Max	0.31	-0.05	10.2	-7.1	-4.3
Mean	0.11	-0.29	2.5	-22.3	-6.0

[1] pH = pH_1994 minus 1998 pH, OM = OM_94 minus 1998 OM, Phosphorous(P) = P_94 minus 1998 Phosphorous(P), K = Potassium (K)_ 1994 minus 1998 Potassium(K), and CEC = CEC_94 minus 1998 Cation Exchange Capacity (CEC). The soils' variables are from a 63 location soil's sample from the 1994 and 1998 seasons.

Interpretation of these results requires some care. First, the t-tests in Table 2 effectively refer to the 0.5 Acre gridded areas for field P, since the sampled locations are not exactly the same for each of the two years. Also, the years studied did not use the same procedures for measuring the soils' properties. Nevertheless, the differences for K and CEC are highly significant. The two years do differ. Indeed, the differences in spatial relationships are more pronounced (Appendix D). Spatial distributions of grid contours are sensitive to the relationships between the surrounding points and can display differences for samples with no statistical differences in their means (Cressie, 1993).

Mallarino, and others (1999) provide one way of evaluating the soils' variables by using descriptive statistics (as in Table 1 above) and through correlations. Their study focused on five corn fields in the Corn Belt. Similarly, Table 3 (next page) provides an accounting of the correlations among the soils' variables using the two years of data collection. Their conclusion from a factor analysis of the soils' variables in relationship to crop yields were that soils' variables that might explain yield variability will likely differ across fields. They did not address relationships between weather and soils' variables on yields. One year's data for these two fields is not sufficient to analyze these more complex relationships.



Mallarino (1999) also stated that the factor analysis showed that groupings of soils' variables were possible in explaining crop yields. For example, Mallarino considered three factors to be growth, soil fertility, and weed control, but found that the interpretation of the specific factors would vary among the fields. From this article, there is evidence that the soils' effects on crop yields can vary considerably from one field to another. Quite understandably, having only one field for each of corn and soybeans must limit the conclusions that are possible to be drawn from this study.

Management of the field has resulted in substantial changes in the nutrients available since correlations of variables across years are no more than 0.32 (P) and typically much less. Indeed, crop yields from both fields were much better in 1998 than in 1997 (Willis, 1999). Part of that improvement could be due to changes in the soils' properties, although weather was likely a significant contributor due to improved rainfall in 1998 (O'Neal, 2000b). Without soils testing from 1997, this report cannot make a more definitive statement of the relationship between the soils changes and crop yield.

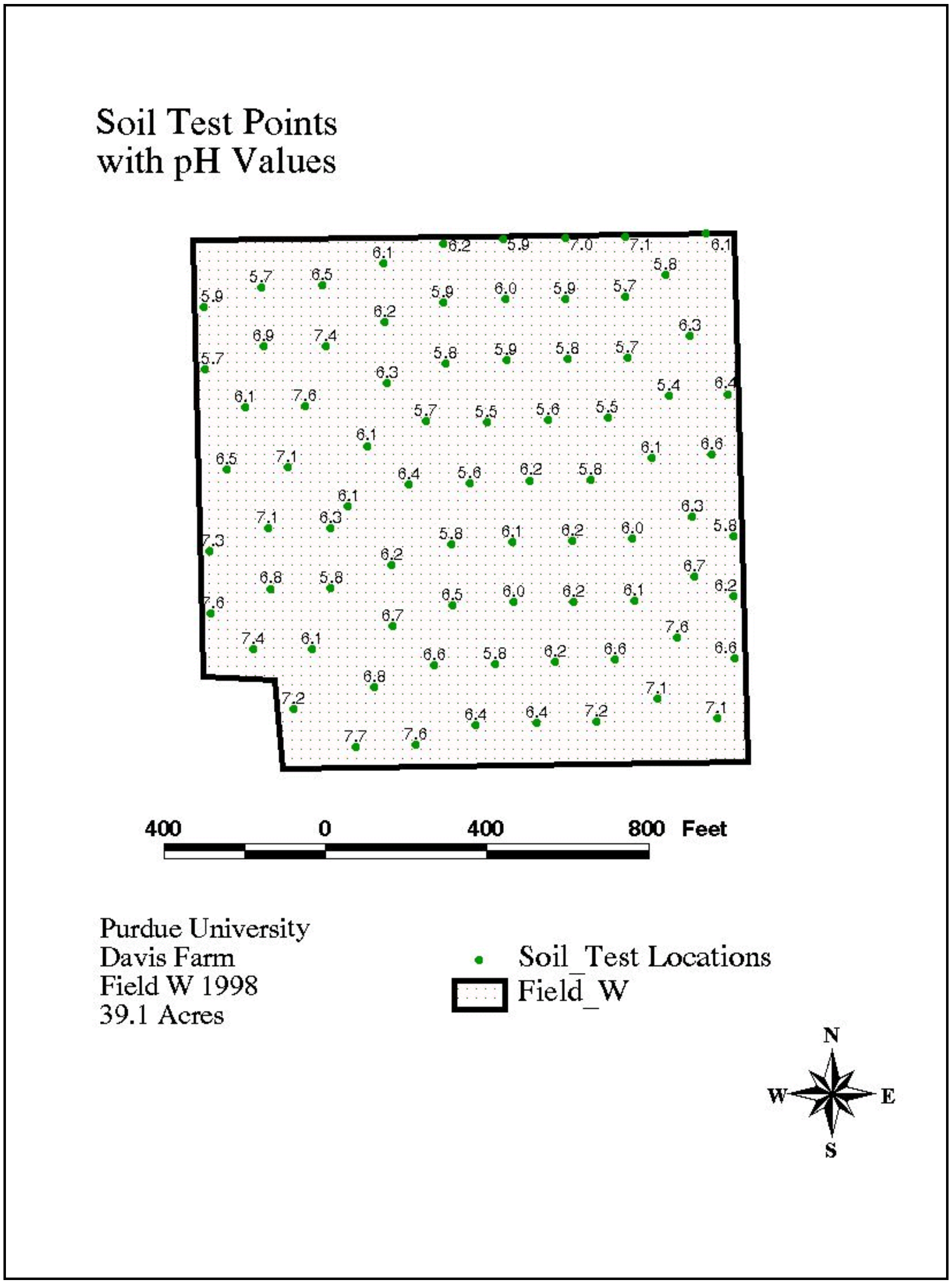
Table 3. Correlations for Purdue Davis Farm Field P, 1994 and 1998 Soils Variables

Variables[1]											
	OM_94	P_94	K_94	pH_94	CEC _94	OM	P	K	Mg	C	pH
P_94	0.45										
K_94	0.59	0.68									
pH_94	-0.44	-0.17	-0.23								
CEC_94	0.44	0.57	0.69	0.08							
OM	0.05	-0.05	0.02	0.07	0.03						
P	0.11	0.32	0.31	-0.02	0.29	0.44					
K	0.07	0.22	0.20	0.07	0.19	0.51	0.76				
Mg	-0.02	0.00	0.01	0.27	0.07	0.15	0.07	0.37			
C	0.09	0.19	0.04	0.17	0.10	-0.26	0.02	0.19	0.39		
pH	-0.02	0.04	0.02	0.27	0.07	-0.64	-0.37	-0.25	0.40	0.61	
CEC	0.08	0.19	0.06	0.13	0.11	-0.07	0.17	0.38	0.52	0.95	0.18

[1] OM_94 = Organic Matter_1994, P_94 = Phosphorous_1994, K_94 = Potassium (K)_ 1994, pH_94 = pH_1994, CEC_94 = Cation Exchange Capacity (CEC) 1994. The 1998 soils' variables were OM = Organic Matter (OM), P = Phosphorous(P), K =Potassium(K), Mg = Magnesium (Mg), C = Calcium (C), pH = pH Reading, CEC = Cation Exchange Capacity (CEC). The soils' variables are from a 63 location soil's sample from the 1994 and 1998 seasons.

The field W 1998 soils' data set (Figure 3, next page) contained 79 data points on an angled pattern from east to west that also had many points near the NASS samples. Table 4 provides those soils' data means while Table 5 presents t-tests between the soils' variables for 1994 and 1998. The substantial difference in the sample pattern for the 1994 soils data prevented doing an analysis of correlations as for field P data. Instead, Table 6 presents the correlations only for the 1998 soils' variables without including the 1994 soils' variables since the locations were so different both in location and number for the two years. Comparisons were made for soils pH, OM, P, K, and CEC. Table 5 of the t-tests (paired t-tests were not possible here) shows that the soils' K (K_94 and K) and P (P_94 and P) have remained close enough in value to be considered the same for 1994 and 1998. Of course, the spatial distribution of the soils' variables is often different although the means might be considered the same. The remaining soils' pH, OM, and CEC differ after adjustment for Bonferroni correct- ions to a p-value of 0.023 as described previously. These measurements most probably differed due to field management practices.

Table 6 (page after next) shows that most soils' variables for field W exhibit positive correlations. The one exception is the soils' pH that has a small negative correlation of -0.07 with OM and 0.18 with CEC. Field P relationships were more mixed, with both large negative and positive correlations among the variables. Although found within the same farm and separated only by the other fields within the farm (Figure 1), the soils relationships differ for the two fields. Of course, some of these differences might be due to soil types. Field P had Glynwood silt loam (12.9 Acres), Pewamo silty clay loam (8.8 Acres), Morley clay loam (5.0 Acres), and Blount silt loam (1.8 Acres). However, field W had Pewamo silty clay loam (22.8 Acres), Blount silt loam (13.6 Acres), and Glynwood silt loam (1.6 Acres). The different mix of soil types might be one factor in the differing relationships among the soil properties as observed. Of course, no easy procedure is available for making a more quantitative statement regarding these descriptive differences in soils properties. Another possible factor in contributing to the soils differences is that field W is a gently rolling field from east to west whereas field P is a nearly level field that might improve the ability to



evenly spread soils' supplements. Again, a quantitative comparison would not possible to make.

Table 4. Descriptive Statistics for Purdue Davis Farm Field W, 1994 and 1998 Soils Variables

Variables[1]	OM_94	P_94	K_94	pH_94	CEC_94	OM	P	K	Mg	C	pH	CEC
Min	2.5	7	76	5.1	16.5	2.0	7	72	255	1000	5.4	9.0
Max	6.2	50	240	7.2	32.7	5.0	151	501	402	3000	7.7	17.9
Mean	3.9	19.5	142.8	5.9	22.4	3.3	31.6	148.2	484	1937	6.4	15.8
S. Dev.	1.24	12.4	48.5	0.54	5.1	0.75	24.9	61.2	104	419	0.59	2.69

[1] OM_94 = Organic Matter_1994, P_94 = Phosphorous_1994, K_94 = Potassium (K)_ 1994, pH_94 = pH_1994, CEC_94 = Cation Exchange Capacity (CEC)1994. The 1998 soils' variables were OM = Organic Matter (OM), P = Phosphorous (P), K = Potassium (K), Mg = Magnesium (Mg), C = Calcium (C), pH = pH Reading, CEC = Cation Exchange Capacity (CEC). Soils data from 1994 were from 16 samples within 2.5 acre grids of field W. However, in 1998, there were 79 soils' sample points for field W.

Table 5. t-test Results for Purdue Davis Farm the Field W, 1994 and 1998 Soils Variables

t-Test Results: 1994 v.s. 1998 Field W Soils Data[1]	pH	OM	Phosphorous(P)	K	CEC
t-Value	-3.0125	2.4956	-1.8882	-0.3327	8.4256
df	93	93	93	93	93
p-value	0.0053	0.0143	0.0621	0.7401	0
	Reject	Reject	Reject	Accept	Reject

[1] pH = pH_1994 minus 1998 pH, OM = OM_94 minus 1998 OM, Phosphorous(P) = P_94 minus 1998 Phosphorous(P), K = Potassium (K)_ 1994 minus 1998 Potassium (K), and CEC = CEC_94 minus 1998 Cation Exchange Capacity (CEC). Soils data from 1994 were from 16 samples within 2.5 acre grids of field W. However, in 1998, there were 79 soils' sample points for field W.

Table 6 (below) shows that most soils' variables for field W exhibit positive correlations. The one exception is the soils' pH that has a negative correlation of -0.07 with OM and 0.18 with CEC. Field P relationships were more mixed (Table 3), with both large negative and positive correlations among the variables. Although found within the same farm and separated only by the other fields within the farm (Figure 1), the soils relationships differ for the two fields. As mentioned earlier, explanations for these difference are difficult to make quantitatively.

Table 6. Correlations for theField W, 1998 Soils Variables

Variables[1]							
		OM	P	K	Mg	C	pH
P		0.54					
K		0.64	0.83				
Mg		0.23	0.40	0.44			
C		0.78	0.59	0.65	0.69		
pH		-0.07	0.30	0.22	0.79	0.42	
CEC		0.83	0.58	0.68	0.60	0.94	0.18

[1] OM = Organic Matter (OM), P = Phosphorous (P), K = Potassium, Mg = Magnesium (Mg), C = Calcium (C), pH = pH Reading, CEC = Cation Exchange Capacity (CEC). The K through CEC variables come from the 79 sample soils' samples for the 1998 season.

2.2 NASS OBJECTIVE YIELD DATA COLLECTION: FIELDS P AND W

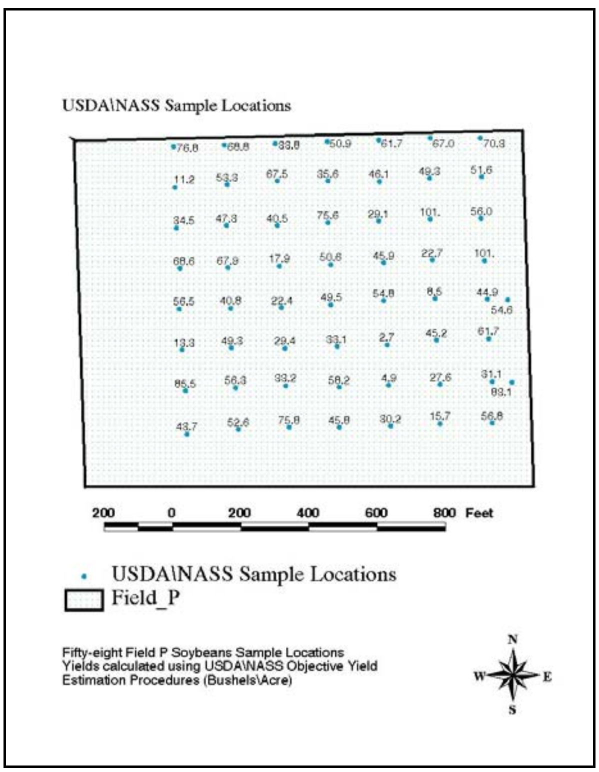
During an onsite visit in the spring of 1997, sample points within fields' P and W for the NASS objective yield sites were selected. Distances between sample locations in field P (Figure 4) were not uniform throughout the field as planned; however, the distances between sample locations in field W (Figure 5) were uniform. NASS does not usually use objective yield data to examine individual fields, but to help in estimating crop yields at the State level (Huddelston, 1976). Purdue University helped to obtain GPS coordinates of the samples within each field and provided access to the sample areas. Each sample area was quite small (Statistical Methods Branch, 1998). The usual NASS procedure requires the collection of data from two paired samples within each sampled field in a State-wide sample.

The corn sample size can vary somewhat because corn is planted in rows. Sample size for soybeans is more uniform, though the laboratory counts for soybeans are usually taken on a small sub-sample and expanded to represent the full sample. To avoid evaluating all these considerations, the following sizes were chosen to represent the samples: 18 square feet for soybeans, field P, and 15 square feet for corn, field W. Eighteen square feet is typical of the soybeans sample size used in standard NASS procedures, but the corn sample is often as much as seventy-five square feet. Both sending samples for laboratory evaluation and the physical demands from collecting the data limited the number of possible data locations to 58 in field P and 64 objective yield sites in field W.

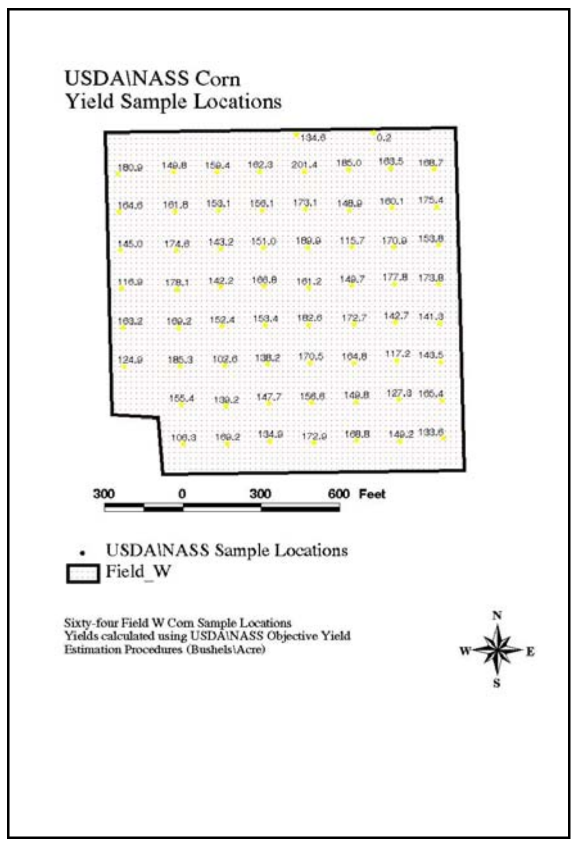
Ralph Gann, State Statistician, Indiana SSO, led 12 NASS enumerators in obtaining the USDA/NASS objective yield data. These objective yield data sets and soils sample data sets for the two fields provided a source of comparisons to the 1998 yield monitor data. An overlay with USGS digital maps at 1:24,000 scale to ensure accurate location of the data sets provided confirmation of the correct location and overlay of the data sets (Appendix C).

3. PREPARATION OF YIELD MONITOR DATA FOR ANALYSIS

Before examining the available data sets, each data set was overlaid on a 7.5 minute map Digital Orthophoto Quadrangle (DOQ) obtained from USGS. This overlay process was necessary to ensure the following: the data sets were in the same projection; each data set properly overlaid all others and the map; and basic distances and areas were correct for the field. Creation of the gridded data sets required all the original data sets to be in decimal degrees, with the ArcView (ESRI, GIS software) projection being the same as the USGS map. Specifically, the map projection chosen was the following: Universal Transverse Mercator (UTM) 1927, Clarke 1866 spheroid, and NAD27 Datum.



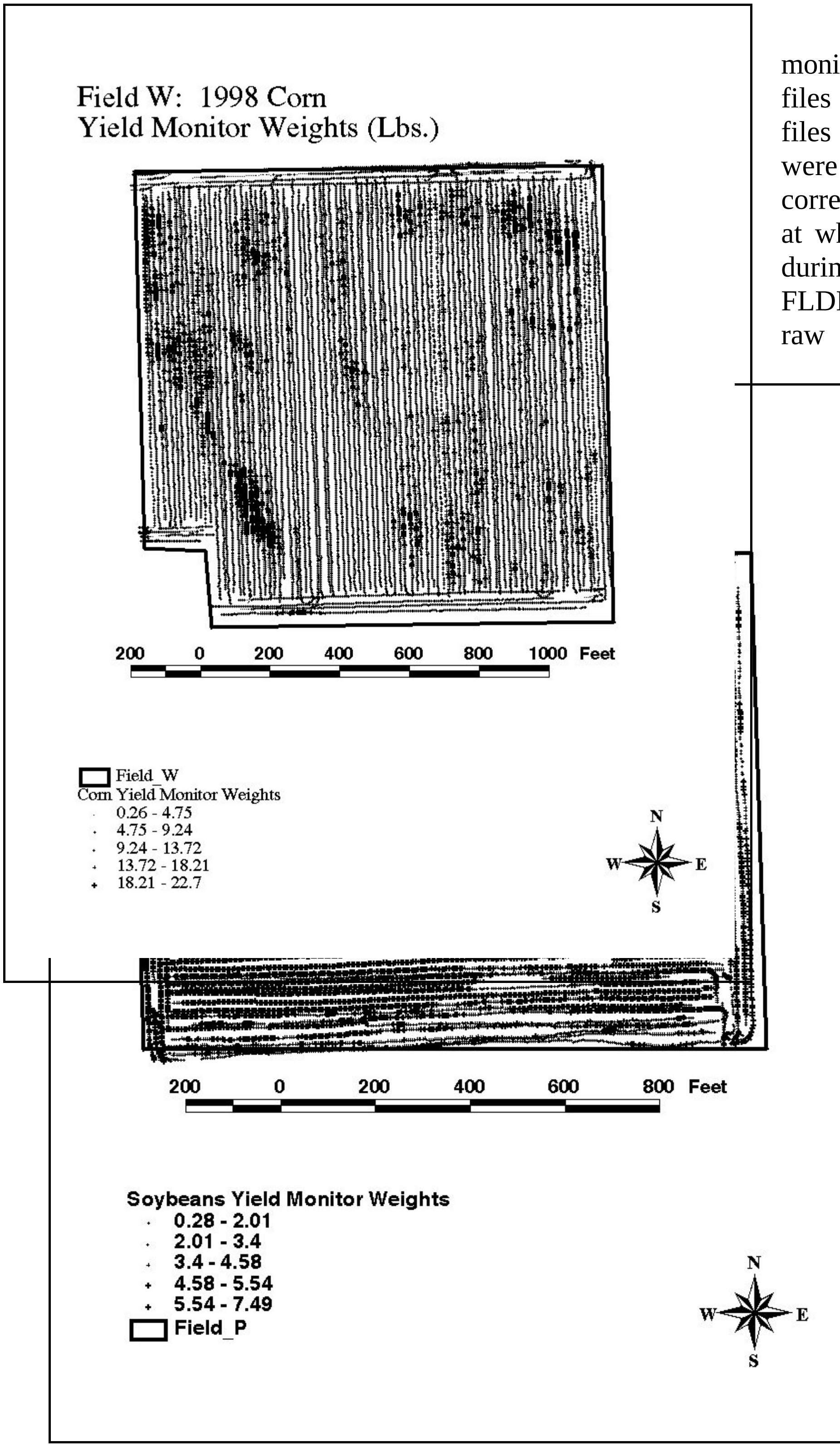
Fifty-eight Field P Soybeans Sample Locations
Yields calculated using USDA\NASS Objective Yield
Estimation Procedures (Bushels\Acre)



Sixty-four Field W Corn Sample Locations
Yields calculated using USDA\NASS Objective Yield
Estimation Procedures (Bushels\Acre)

3.1 CORRECTION OF MAJOR POSITIONAL AND RECORDING ERRORS IN THE YIELD MONITOR DATA

After delays in developing software to correct the locations of the raw yield monitor data, Martin Ozga, Research and Development Division, programmed FLDHARV (Appendix B) in Delphi 3. He translated algorithms into code intended to correct for lag, ramp-up, ramp-down, and zero data points. Preparing the data for use in ArcView 3.2 with the Spatial Analyst add-in (ESRI,1996) remains a laborious process. However, FLDHARV did make possible consistent corrections of many data errors for fields P and W (Appendix B). Correcting the errors in a spreadsheet would have been possible, but would have required much more labor and been more error-prone. Purdue did not ensure that harvest paths were straight nor prevent overlap of passes. Recent developments in associated precision agriculture (such as a light bar tracking system to align the harvester passes) may aid in improving the data positioning over what was observed in these datasets. The reasoning for avoiding corrections of this kind was that no additional information was available to make accurate determination of the correctness of the resulting locations.



Preparing the yield monitor data required using ten files for field P and thirteen files for field W. These files were of varying sizes corresponding to the intervals at which the data was saved during harvest. Using the FLDHARV program with the raw ASCII files allowed an examination of each point. The program summed small data points corresponding to the ramp-up and ramp-down, eliminated zero data points corresponding to the lag, and adjusted the data values to find the starting values for each row more accurately. Elimination of many zero values and properly placing the observed values made the data appear much more regular. However, additional errors remained, since rows varied in their closeness to each other along the field's length.

After correcting these major positional and recording errors in the yield monitor data, our first step was to provide a mapping of the initial values. Actual readings obtained were in pounds of crop. The usual way of printing out these preliminary values is to convert them to Bushels/Acre in a GIS. Since FLDHARV did not provide accurate calculations of the new distances between recorded points, accurate yield estimates were not possible on a location by location basis. However, providing a plot of the weights collected in pounds will give a simple idea of patterns of lowest and greatest yield within the field, along with errors in the yield monitor data locations. Figure 6 (previous page) provides this mapping for field P while Figure 7 provides the same kind of mapping for field W. These maps help to show the variations in measured crop weights within the field. They also show some of the uneven movements by the harvester down the rows. However, the maps are not intended as a direct evaluation of harvested yields within the fields.

Examining these Figures reveals many cases of overlapping data, breaks in the data, open areas with no data values, and values outside the mapped field boundaries. Another component of the yield monitor data that is evident is the changes in recorded weights that vary throughout the fields and within each pass. Few adjacent recorded weights are the same so that calculated yields will vary in a similar manner.

The correction of all errors in the data would require making many assumptions about the data that would be difficult to justify. Potential mapping corrected many of these errors and permitted analysis of the resulting gridded data.

3.2 EVALUATION OF THE FLDHARV SUPPLEMENTARY OUTPUTS

Evaluation of the corrections made by FLDHARV showed considerable variability within each data file. Running each file individually within FLDHARV helped by permitting evaluation of data in the order recorded. The program output provided the following information:
1) Calculation of the number of points within a pass,
2) A count of zero weight values before a positive weight value (lag),
3) The number of points with positive values before more stable values start (ramp- up),
4) Evaluation of the reduced weights at the end of a pass (ramp-down), and
5) A count of the corresponding zero values with header-up at the end of a pass (Appendix B).

3.3 POTENTIAL MAPPING OF THE YIELD MONITOR DATA

Although the data exhibited additional errors not corrected from this first step, the process of potential mapping described by Blackmore and Marshal (1996) helped to make the data useable for analysis. Blackmore and Marshal explained that potential mapping uses the summation of all available yield monitor weights within grid cells of a certain size. A 30 meter square (900 square meters) was the starting point for this analysis. Other researchers have considered areas as small as one-meter (Willis, 1999). However, this report attempted to reduce overall errors in the yield monitor data by making only small position corrections and crop weight corrections using FLDHARV. Smaller areas than 30 meter squares would require much more work with the data to develop methods of dealing with the many problems present. Potential mapping does especially well in correcting for a partially full header, zero weights, and variations in positional accuracy.

However, adjustments for lag, ramp-up and ramp-down are necessary to make potential mapping procedures work correctly. FLDHARV made these corrections these corrections possible with greater reliability than possible if using spreadsheets to examine the groupings of data.

Potential mapping required adding up the total yield within each 30 meter grid cell. Selection of the 30 meter grid size was somewhat arbitrary, since no general agreement exists among researchers on an optimal grid cell size. Smaller grid cells require greater confidence in the accuracy of the crop weights measured, the GPS readings, the various corrections necessary to position the readings accurately, and a full header. The width of the header (15 feet) means that six passes of the combine will be included in a 30 meter grid width. Since passes were not always straight line due to various possible errors, use of this size grid will make the effect of various uncorrected data errors of less impact on the final analyses. Indeed, this procedure helps to correct for many errors inherent in the yield monitor data by simply aggregating the many yield monitor readings within each square. The evaluation of the data can now focus on larger blocks of data rather than on the more variable individual readings. Researchers have found this technique to be particularly helpful in comparing the yield monitor data with other data sources such as soils' data.

4. INTERPOLATION OF SOILS' VARIABLES

Creation of surface from an initial data set associated with point data is known as interpolation. The purpose of this surface creation is to permit the comparison of different data sources that cover the same area. Estimated values for locations were data is not available permit these comparisons for even those cases in which data exists at the different locations for the various data sources.

4.1 BACKGROUND ON INTERPOLATION METHODS

IDW (inverse distance weighting), splining (a minimum curvature technique), and other spatial statistics procedures can create a grid of interpolated soil data. Comparing these grids with the potentially mapped yield monitor data is one place to start in evaluating the data. However, unlike the IDW procedure, splining can extrapolate the data outside the contained region and can provide a closer fit to the original data values. Doucette and Beard (2000) compared four techniques of interpolation and found splining to be acceptable for gap fills. Using the authors' experience as a guide, the default parameters for the splining procedure in ArcView were chosen to create the corresponding gridded surfaces. These interpolated NASS estimated crop yields were compared with the potential mapped Purdue yield monitor data using the same grid cell size (30 meter).

The only change made to the splining procedure was the use of the tension splining option to force the interpolated data to fit the NASS objective yield values more closely. This choice made the surface more closely conform to the NASS values to aid in evaluating only our the objective yield procedures. The splining procedure was particularly important for field P, since the objective yield locations missed the western edge of the field. Estimating the values in the western part of the field required extrapolation, which was not possible using IDW procedures.

As mentioned earlier, splining is one way provided by the Spatial Analyst software add-on for ArcView to create an interpolated surface from a given data source. Other procedures such as kriging are available, but require more analyst interaction with the data and more time to produce a gridded surface. Kriging requires an evaluation of the variogram (Cressie (1993), a spatial statistic showing the degree of correlation of data with changing distances. Evaluation of the variogram would require focusing on the data in much greater detail than would be appropriate for this report.

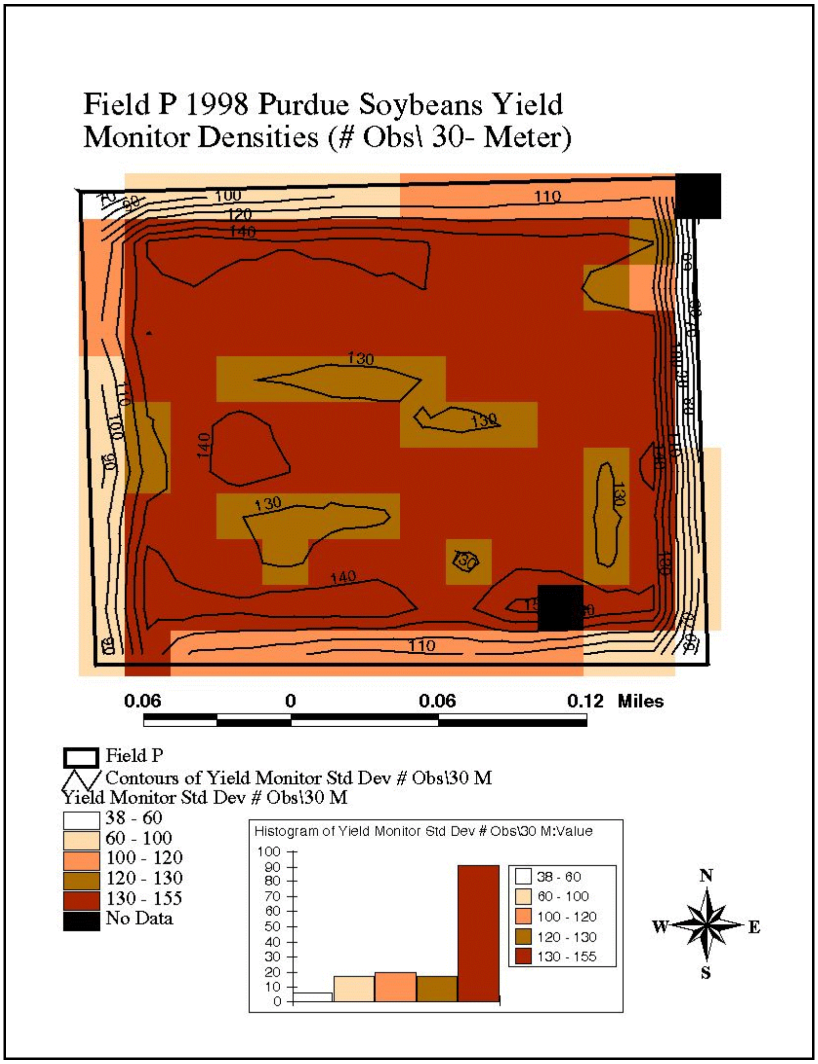
The splining procedure used the standard setting of a second order equation with 12 adjacent points and the tension option in creating the grid. Use of splining permitted creation of a surface for each soils' variable along with both the yield and moisture values from the NASS objective yield data. Unlike kriging, splining the data does not generate estimates of the accuracy of the predictions. Although interpolation of the yield monitor data would be possible, this study focused only on the use of potential mapping for the yield monitor data. Splined surfaces with 30 meter grid cells for soils' data and the NASS objective yield data provided the capability of comparing similar locations with the sampled soils' data.

4.2 MAPPING SPLINED SOILS' VARIABLES: FIELDS P AND W

Use of ArcView's Spatial Analyst to create the maps of the splined soils' variables made possible creation of a map for each soils' variable examined earlier: pH, Organic Matter (OM), Phosphorous (P), Potassium (K), and Cation Exchange Capacity (CEC) for 1994 and 1998. Magnesium (Mg) and Calcium (C) were available only for the 1998 soils' testing for both fields. Using a 30 meter grid cell for mapping soils' variables meant that assumptions about the locations of the 1994 soils' sample location were not as critical to creation of the maps. For field P, the 1998 data locations were accepted for both years of data since these points were within the original Top Soils 0.5 acre grids. Field W, however, had such a difference that arbitrary points were chosen within the 2.5 acre grids to represent the 16 sample locations. Consequently, the soils' maps for field W were much less accurate for the1994 data.

Birrell, and others (1996) present an excellent account of how interpolation method, selection of the number of nearest neighbors used in the interpolation process, and even the selected grid size may influence the final interpolated soils' map. Others suggest that even the method of sampling (for example, simple random sampling, grid sample, or a directed sampling approach) the locations of the soils' samples can affect the final map produced (Pocknee, 1996). The individual field's history, cropping practices, and overall field management can have important contributions to the complexity exhibited by the interpolated soils' variables maps as well. Therefore, the maps presented in this paper provide only one possibility of many for interpretation of the sampled soils' variables. The possibilities for interpolating the soils' data include use of kriging, IDW (inverse distance weighting), and splining, which was chosen here. Each method chosen would produce a somewhat different final map. However, emphasis here will be more on evaluating broad categories in the mappings. Soils' data will be compared in greater detail against interpolated yield monitor data and NASS objective yield data.

Choice of contours at regular intervals for each variable made possible creation of more



easily interpreted black and white maps using ArcView's Spatial Analyst for this paper. The contours are in effect a summarization of the grids into zones within each field corresponding to groups of 30 meter grid cells. As much as possible, these contour intervals were kept consistent between the two years, but the natural variation in the soils' sample data between years meant that ranges at the low and high end of values were different. Increasing the number of ranges for the data and decreasing the size of the contouring intervals would have allowed more accurate evaluation of the contours. However, more contours would increase the complexity of the maps and require more careful interpretation to evaluate them. Also, the number and accuracy of the soils' data samples would need further improvement to justify the use of greater precision. Please refer to Appendix D to examine examples of the soils' maps in detail for field P.

5. EVALUATION OF THE POTENTIAL MAPPED YIELD MONITOR DATA

The objective of this phase of the analysis was to evaluate one possible way of preparing the different data surfaces. Since many other algorithms are available, additional studies to compare the accuracy obtained using the different procedures would be possible. For example, Purdue University has used extensive data preparation procedures for the yield monitor data, but has still seen the need to combine data using potential mapping procedures (O'Neal, 2000). Evaluation of these other methods would be the subject of another paper.

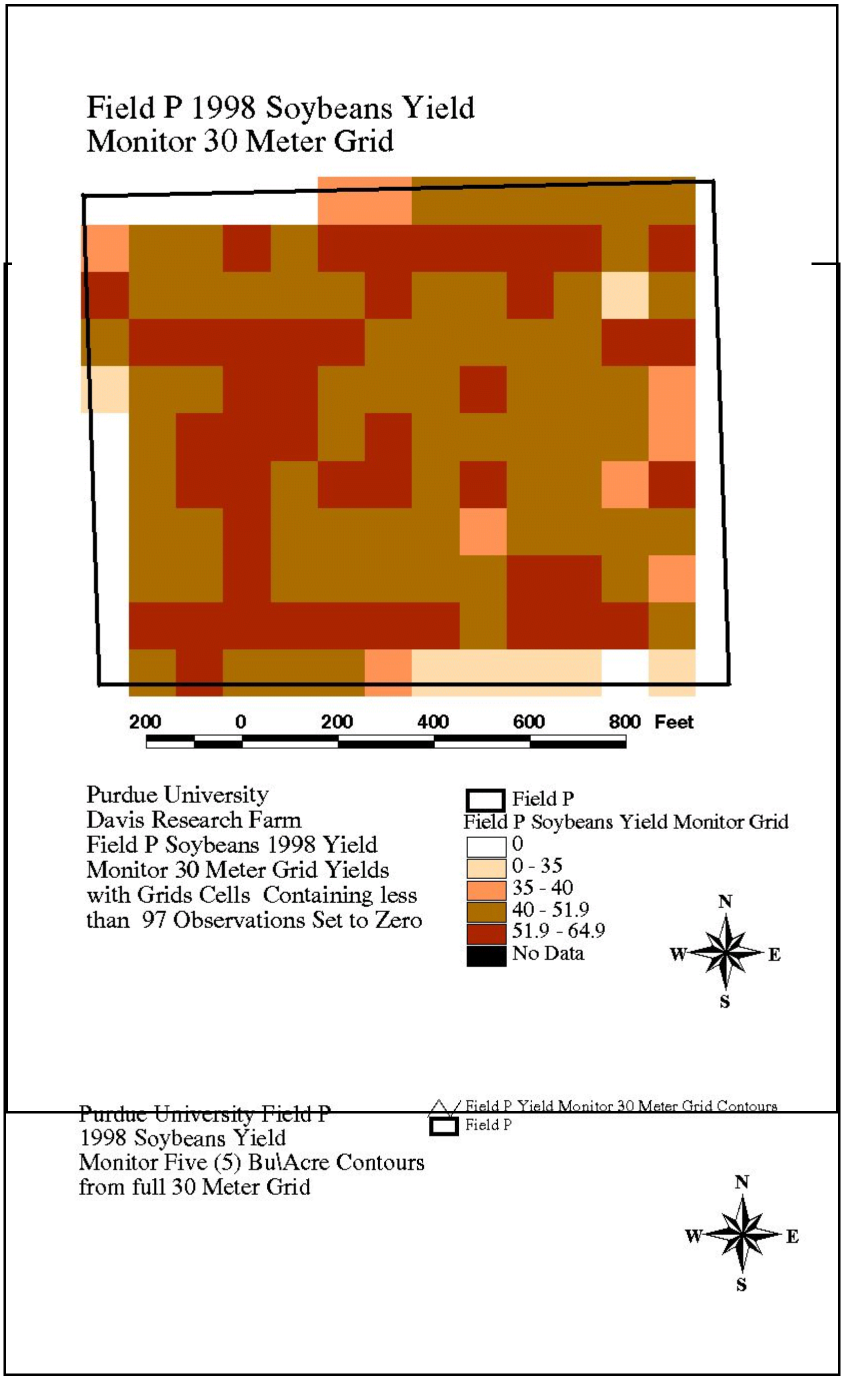
The process of potential mapping as done in this study was to create 30 meter square gridded cells from the yield monitor data by aggregating the observed yield monitor data within the gridded area. A digitized outline of the field provided a mask for any data outside the field. After study of various possibilities to maintain use of the data, an approximate density count of yield monitor data values within each grid cell (Figure 8) was used to determine which grid cells to retain in the analysis. Only those grid cells containing more than one standard deviation below the mean number of observations within the grid cells were retained. Grid cells were of particular concern in the eastern and western edges of both fields, but the northern and southern edges had some difficulties as well, due to the reduced number of included samples.

One concern with potential mapping for the yield monitor data in both fields was the presence of lower data values and many zero weight values within the end row areas. Since this data was part of the field, inclusion of the end row areas was necessary. The lower yield values in the end-rows should create an edge-effect for the eastern and western portions of field P. Similarly, the end-rows should give lower yields for the northern and southern portions of field W. Strong evidence that the yields were lower in the end-rows of field W comes from two NASS data points in the northern edge of the fields that were 134.7 and 0.2 Bu/Acre, respectively: both values were below the field average yield for the NASS data points. Potential mapping at 30 meter grid cell sizes reduced the number of observations within the edge grid cells for both fields since these grid cells that often went outside the field accuracy of the field boundaries could be in error.

5.1 EXAMINATION OF FIELD P POTENTIAL MAPPED YIELD MONITOR DATA

After correction of the original data set using FLDHARV, the remaining 19,419 weights show a total of 92,743 pounds of soybeans was obtained from field P. Without moisture adjustment, the yield would be 48.5 Bushels/Acre without regard to the planned potential mapping of the data. As a comparison to this data, weigh wagon data provided a value of 45.5 Bushels/Acre. An agreement within three Bushels/Acre is certainly acceptable at this scale since this difference would mean only 100 bushels difference at the field level.

Creating the grid cells for field P required summing the yield monitor weights (in pounds) on a 30 meter grid to obtain total weights. Field P (planted with soybeans for 1998) with 31.9 acres produced an initial grid with 11 rows and 14 columns for a total of 154 grid cells. After using the average moisture values and adjusting to a standard moisture of 12.5%, the total weights for each grid were converted to acres (a factor of 4.49636 Grid Cells/Acre) and divided by 60 pounds per bushel within ArcView's Spatial Analyst. These calculations produced estimated soybean bushels per acre yields for each grid cell.



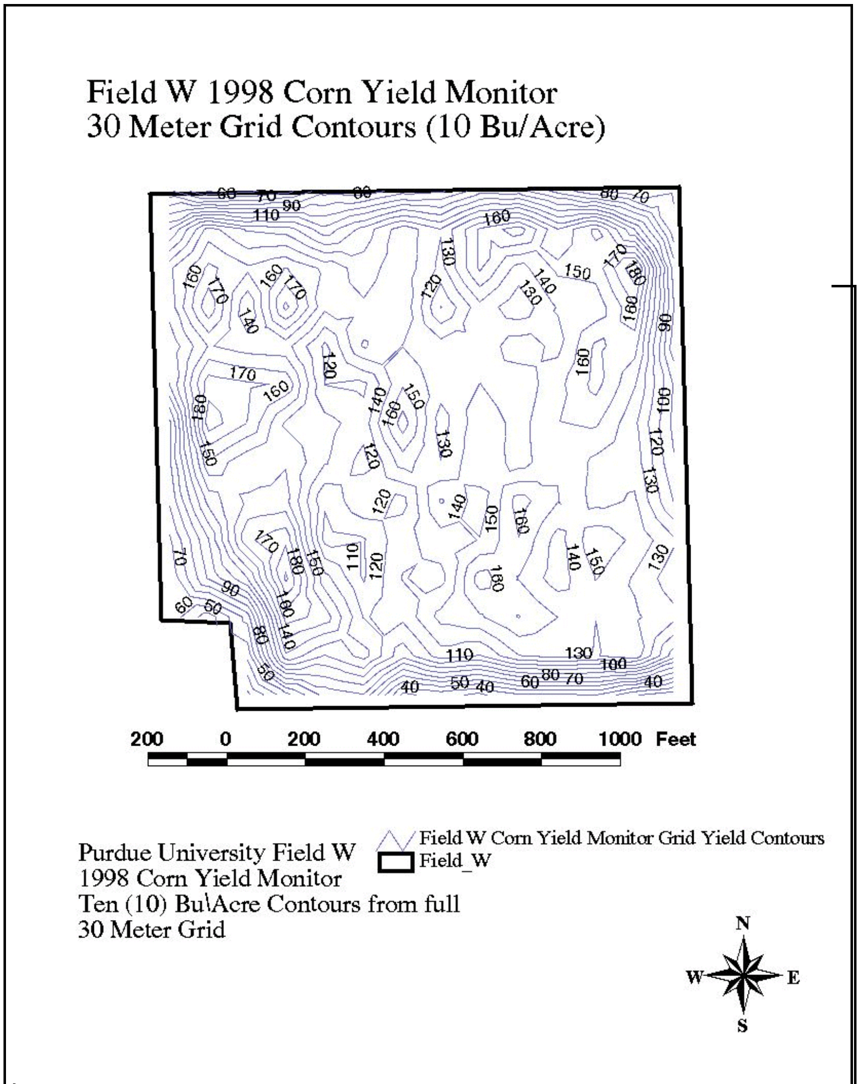
After creating the gridded data set, ArcView's Spatial Analyst was used to create contours using the full set of grid cells at a five (5) Bushel/Acre resolution. These contour maps would help in evaluating the accuracy of this first attempt. An edge effect on the eastern edge of field P is especially apparent in the closely spaced contours that ramp-up quickly from a minimum of five (5) Bushels/Acre to 35 Bushels/Acre (Figure 9). The remainder of the field shows less variation except one section in the northeastern part of the field that had a much lower yielding grid surrounded by larger yielding grid cells (Figure 10).

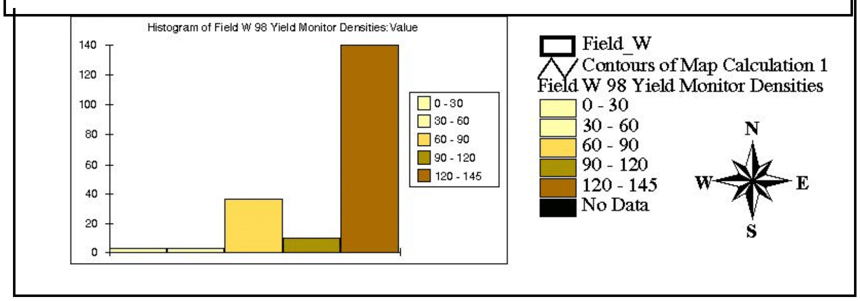
Two causes of the reduced yields in the eastern edge of the field are readily apparent. The first is that this edge of the field contains end-rows that are typically lower yielding than the remainder of the field. The second cause of reduced yields for these grid cells is that the number of observations in these grid cells is much less than for the remainder of the field. Choosing to include only those grid cells with greater than minus one standard deviation resulted in accepting only those grid cells with less than 97 observations. This choice required deleting the entire eastern column of grid cells along with additional grid cells in the western and northern edges where the field boundaries did not match well with the yield monitor locations. Eliminating the eastern column of grid cells, many along the western border of the field, along with modifications to the northern and southern edges of the field reduced the number of grid cells to 131 (Figure 10).

The reduced grid of data will be used in later analysis of the yield monitor data with the NASS interpolated data and the recorded soils data. Only these selected grid cells will be used in the calculation of correlations and other comparisons. This final grid surface portrays the surface of yields throughout the field since they exclude extreme values due to reduced numbers of observations from the calculations.

5.2 EXAMINATION OF FIELD W POTENTIAL MAPPED YIELD MONITOR DATA

After correction of the original data set using FLDHARV, the remaining 22,569 recorded weights totaled 299,275 pounds of corn for field W. Without moisture correction, the yield would be 134.1 Bushels/Acre without considering the planned potential mapping of the data.





Creating the grid cells for field W required summing the yield monitor weights (in pounds) on a 30 meter grid to obtain total weights by grid cell as for field P. Field W (planted to corn for 1998) had 39.1 acres that produced an initial grid containing 14 rows and 14 columns for a total of 196 grid cells. After using the average moisture values and adjusting to a standard moisture of 15.5%, we expanded the total weights for each grid to acres (a factor of 4.49636 Grid Cells/Acre) and divided by 56 pounds per bushel. These calculations produced estimated corn Bushels per acre yields for each grid cell (Figure 11).

After creation of the gridded data set, contours were generated from the full set of grid cells at ten (10) Bushels/Acre resolution to help in evaluating the accuracy of this first attempt. The edge effects for all contours near the field W boundaries are evident from the closely spaced contours that ramp-up quickly from minimums of 40 Bushels/Acre in southern part of the field, 60 Bushels/Acre in the western and northern sections, and 90 to the east. The central sections of the field show contours that generally range from 120 to 170 Bushels/Acre (Figure 12).

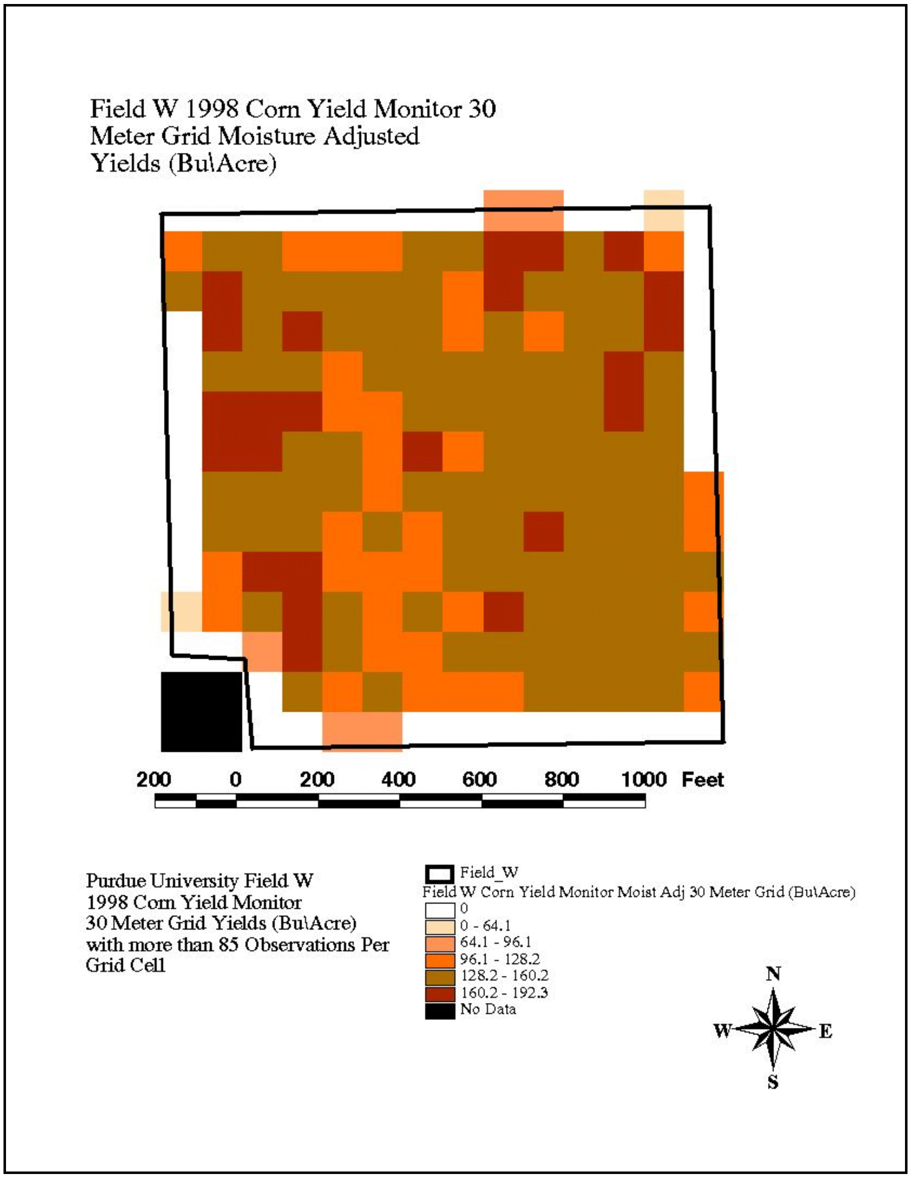
Two causes of the reduced yields in the southern and northern edges of the field are apparent. The first is that these edges of the field contain end-rows which are typically lower yielding than the remainder of the field. The second cause of reduced yields for these grid cells is that the number of observations in these grid cells is much less than for the remainder of the field. Choosing to include only those grid cells with greater than minus one standard deviation resulted in accepting only those grid cells with 85 or more observations (Figure 13, next page). This choice required deletion of most boundary grid cells. Using this rule resulted in deleting 41 grid cells on the boundary (including the four grid cells that included a farmstead in the southwestern corner of the field). These deletions reduced the number of grid cells to 155.

6. INTERPOLATED NASS YIELD SURFACES

Splined surfaces for the NASS yield estimates provided estimates for the same grid cells as the yield monitor data described above for each field. To test the results of different methods of obtaining the gridded data, the following two methods were used to generate NASS grid surfaces for the two fields:

1) Create a grid using only the calculated sample location yield estimates using the traditional objective yield formulas, and
2) Generate small grid cells, corresponding to the NASS sample area, for each component of the two component parts of the yield; plants per acre and weight of the crop per plant, then multiply those two grids to calculate the 30 meter grid surface.

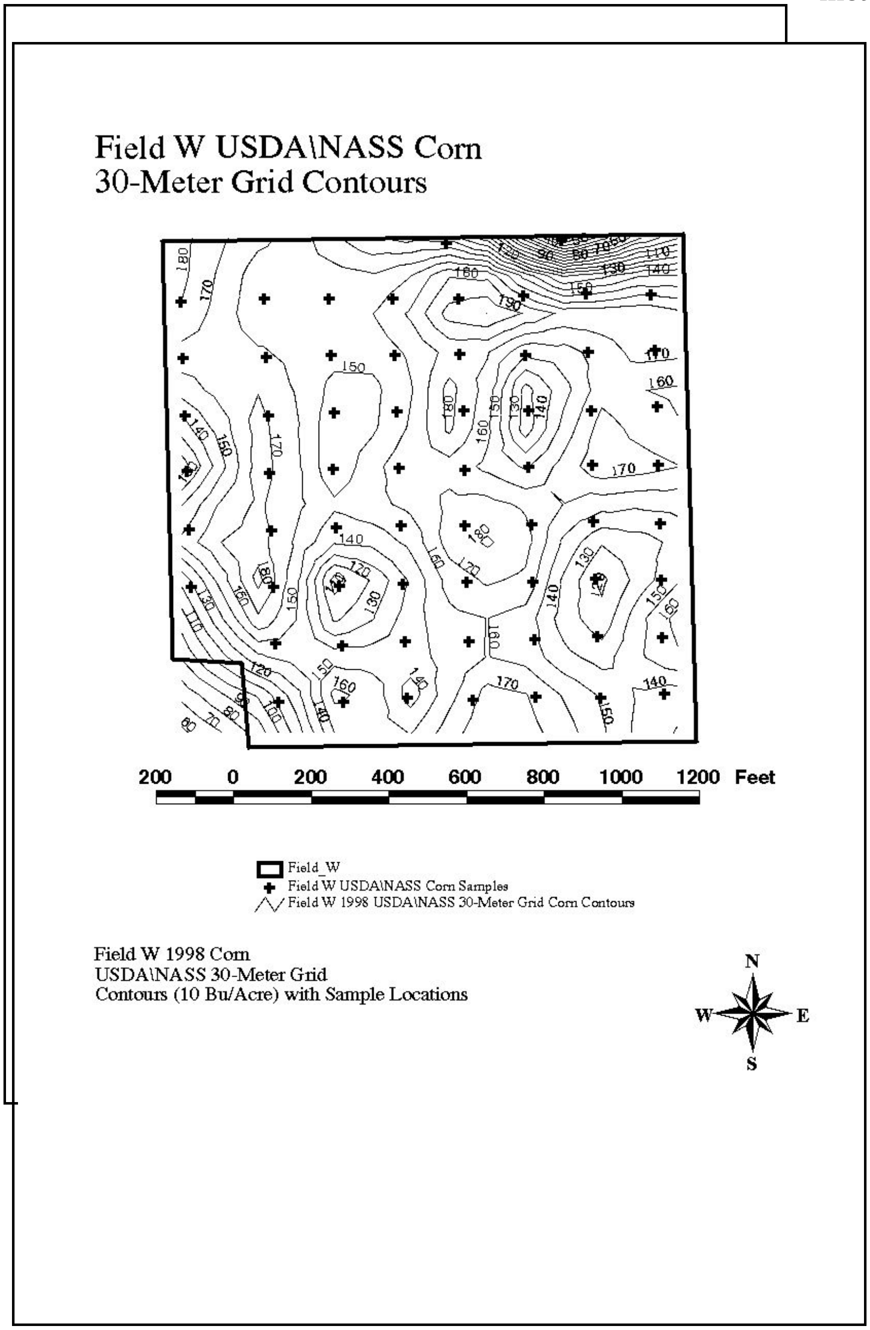
Some differences were evident between the resulting grid-cell values, but these differences were usually less than two bushels per acre. Method two was conceptually more accurate and was used to obtain the values for comparison with the yield monitor and soils' values. Method one included here seemed sufficiently accurate for creation of the figures.



The mean of the original sampled points for field P was 47.8 Bushels/Acre after subtracting two Bushels/Acre, the objective yield harvest loss factor for 1998 in Indiana (Personal Communication). The western edge of the field required extrapolation of the data values and was thus less accurate than the remainder of the field. Using all the grid cells permitted creation of a contour surface (Figure 14) to compare against the contour surface (Figure 9) generated from the yield monitor data. These areas of the field are not accurate, but are included to aid in understanding the lack of accuracy in data at the edges of the field.

Evaluation of these NASS interpolated yield contours shows a much more complex surface of soybeans yields than was evident for the Purdue yield monitor derived contours. A possible explanation of the eleven (11) circular areas with either increasing or decreasing soybeans yields is that the use of the tension splining process forces the interpolated surface through the selected sample locations. If one sample location has a substantially different value from those surrounding it, then these circular areas of rapidly changing yields might occur.

The southeastern corner of the field has an area with unrealistically large soybeans yields due to one sample point with a large yield value of 89.1 Bushels/Acre. This Figure includes all the 30 meter grid cells as in Figure 8. Therefore, the eastern edge of the field with the northern and western areas corresponding to excluded grid cells from Figure 9 has contours of much less accuracy. The contours do exhibit considerable complexity with many changes throughout the field. Even at the 30 meter grid cell level, there is little evidence of uniform yields within the field.



Field W's interpolated 30 Meter grid of corn yields had a mean of 149.1 Bushels/Acre after adjustment for harvest loss of 2.7 Bushels/Acre (personal communication). This yield compares well with the 153.3 Bushels/Acre calculated from the original 64 NASS-collected data points (after adjustment). However, neither of these estimates for field W's corn yield agree with the weigh wagon estimate of 139.9 Bushels/Acre (Willis, 1999).

The upper northeast corner of field W displays closely spaced increasing grid contours due to the very low NASS sample value of 0.2 Bushes/Acre (Figure 15). Since this sample is in the end-rows and near the edge of the field, highly variable corn yields are quite possible.

More centrally located regions of closely spaced circular groupings of increasing or decreasing corn yields are likely due to the use of a splining with tension option procedure, as happens for field P. Note that these places in the field are fewer (six in number) than for field P and more similar to the yield monitor yield contours. However, the field W NASS corn yield contours are generally larger than the contour yields for the Purdue yield monitor data.

7. CROP YIELDS RELATIONSHIPS TO 1998 SOILS' VARIABLES

This phase of the analysis will rely on the sample locations from the 1998 soils' surveys of fields P and W along with the yield monitor potential mapped grids at 30 meter and the NASS splined 30 meter grids. Additionally, the requirements were that any sample location that occurred in portions of the overlaying grids that contained fewer than minus one standard deviation of the mean number of observed points for that field's grid cells were discarded. For field P, this requirement reduced the original 63 sample soils' locations to 58 locations. Field W had ten locations excluded for this analysis so that the original 79 locations were reduced to 69 locations. Tables 7 and 8 exhibit the same variables included in earlier tables to show the effect of the deletions on the descriptive statistics and correlations for the soils' variables.

To make comparisons among the yield monitor, the NASS objective yield estimates and the soils' data, a plan was necessary to make the data sets as equivalent as possible. Use of a sample to permit statistical testing was one criterion used. The next criterion was to minimize the quantity of interpolated data when making the comparisons. The use of the soils' sample location along with the soils' measured values would provide the least amount of interpolated data. Selection of the splined NASS 30 meter grid cell estimates and the potential mapped 30 meter yield monitor grid cell estimates located at the soils' sample locations would provide one possibility that would meet the two above criteria. To ensure choosing the most accurate data, all grid cells not meeting the chosen minimum number of observations for that field would be excluded.

7.1 ANALYSIS OF FIELD P 1998 SOYBEAN YIELDS RELATIONSHIPS TO 1998 SOILS' VARIABLES

To summarize the creation of yield monitor 30 meter grids of yield estimates, the procedure was to sum data weights for field P within each 30 meter grid cell to create a total weight of grain. Multiplication of the weights within each grid cell by 4.4936 converted the values to an acreage basis. Next, was the correction of the grid cell weights to a standard 12.5% moisture. Finally, dividing by 60 pounds/Bushel converted the values to bushels per acre. These calculations produced a grid with a minimum cell value of 2.9 Bushels/Acre and a maximum of 67.5 Bushels/Acre. Its mean was 45.3 Bushels/Acre with a standard deviation of 13.0 Bushels/Acre. This calculated grid agrees well with both the weigh wagon value of 45.5 Bushels/Acre and the full data set of yield monitor data at 48.5 Bushels/Acre. After adjustment for harvest loss, the NASS sample provides a mean yield of 47.8 Bushels/Acre

Analysis of the data sets started with an examination of the moisture readings determined for each grid cell from both the NASS sample points and the yield monitor averaged values. The moisture readings were an average of many moisture readings obtained within the grid cell for the yield monitor. However, the NASS moisture readings were the results of a splined data surface. Without deletion of any created grid cells, yield monitor moisture averages for field P ranged from 10.5% to 13.4%. Using the same full grid of data, the mean of the moisture readings was 11.4% with a standard deviation of 0.56. However, NASS moisture readings for the same grid cells ranged from 7.7% to 8.4%. The moisture mean was 8.0% with a standard deviation of 0.13.

Tables 7 and 8 provide the statistics of the NASS original sample data, the interpolated NASS soybean objective yield 30 meter grid cells obtained at the reduced 58 locations, the Purdue yield monitor data potential mapped at 30 meter grids at the same 58 locations, and the soils' variables at these same locations (as originally recorded). Although every effort was made to make the compared datasets as comparable as possible, caution must be exercised in interpretation of the results for both fields P and W. Correlations, in particular, have shown much variability in other studies from one field to another. Looking again at Table 1 and Table 3, the deletion of five sample locations has not significantly altered the means or correlations of the soils' variables. Of course, the two methods of obtaining the 30 meter grid-cell values for the NASS objective yield data and the Purdue yield monitor data are different processes. Not interpolating the soils' data makes comparisons at the 30 meter grid cell level of less comparability. However, minimizing alterations to the data was emphasized.

Table 7. Descriptive Statistics for the Field P, 1998 Variables: NASS Soybean Yield, Yield Monitor Soybean Yield, and Soils Variables

Variables[1]	SY	INY	YMY	INM	MYM	OM	P	K	Mg	C	pH	CEC
Min	2.7	2.1	22.4	7.8	10.6	2.2	10.0	70	385	1350	5.5	11.1
Max	101.1	90.4	60.8	8.2	12.5	4.9	97.0	296	775	4750	8.2	28.1
Mean	47.8	46.1	48.7	8.0	11.3	3.3	43.9	145	576	2170	7.1	16.6
S. Dev.	22.0	18.5	6.9	0.1	0.4	0.63	26.8	44.8	100	711	0.67	3.9

[1] SY = Sample NASS Yields, INY = Interpolated NASS Yields, YMY = Yield Monitor Yields, INM = Interpolated NASS Sample Moisture, MYM = Yield Monitor Moisture Average, OM = Organic Matter (OM), P = Phosphorous (P), K = Potassium (K), Mg = Magnesium (Mg), C = Calcium (C), pH = pH Reading, CEC = Cation Exchange Capacity (CEC). NASS sample yields come from the original 58 sample points for field P. The Interpolated NASS yields and moisture readings with the yield monitor values correspond to the interpolated values at the remaining 58 soils sample locations for field P after deletion of the grid-cells containing less than 97 observations. The OM through CEC variables are based on this reduced 58 sample soils readings from the 1998 season.

Although the NASS and Purdue yield monitor soybeans' sample of 30 meter grid-cells mean yields are quite close, the moisture values do differ (paired t-test results as follow):

Yields	Moisture
t = -0.9778	t = -50.5898
df = 57	df = 57
p- value = 0.3323	p-value < 0.0001
Infers to accept no difference in yields	Infers to reject no difference in moisture.
	Note: A p-value of near 0 does not require Bonferroni correction.

===

Neither the NASS soybeans yield nor the yield monitor soybeans yield for the 30 meter grid cells located at the 58 locations of the soils sample show very large correlations with any other variable measured. Indeed, the two yields show a small negative correlation of -0.07, although there was no significant difference between their means according to the paired t-test (above). Examination of Figure 9 and Figure 14 show contours that generally do not show a close degree of spatial relationship. Of course, the ranges for the interpolated soybeans yields in these figures have similar ranges for the two data sets. Some areas do show general agreement, however, the overall relationships appear to differ.

Table 8. Correlations for the Field P, 1998 Variables: NASS Soybean Yields, Yield Monitor Soybean Yields, and Soils Variables

Variables[1]	INY	YMY	INM	MYM	OM	P	K	Mg	C	pH
INY										
YMY	-0.07									
INM	-0.20	0.13								
MYM	0.17	-0.28	-0.51							
OM	0.15	-0.22	0.05	0.12						
P	0.26	-0.24	-0.38	0.50	0.43					
K	0.31	-0.33	-0.29	0.42	0.49	0.75				
Mg	0.08	0.02	-0.09	-0.08	0.12	0.05	0.34			
C	-0.10	0.09	0.21	-0.26	-0.28	-0.01	0.16	0.37		
pH	-0.19	-0.23	0.01	-0.24	-0.75	-0.46	-0.37	0.29	0.55	
CEC	0.00	0.01	0.16	-0.21	-0.08	0.15	0.36	0.52	0.95	0.36

[1] INY = Interpolated NASS Yields, YMY = Yield Monitor Yields, INM = Interpolated NASS

Sample Moisture, MYM = Yield Monitor Moisture Average, OM = Organic Matter (OM), P = Phosphorous(P), K = Potassium (K), Mg = Magnesium (Mg), C = Calcium (C), pH = pH Reading, CEC = Cation Exchange Capacity (CEC). NASS sample yields come from the original 58 sample points for field P. The Interpolated NASS yields and moisture readings with the yield monitor values correspond to the interpolated values at the remaining 58 soils sample locations for field P after deletion of the grid-cells containing less than 97 observations. The OM through CEC variables are based on this reduced 58 sample soils readings from the 1998 season.

One possible explanation of this result is that the two procedures, the yield monitor and the NASS objective yield data collection, are not measuring the same characteristics of the field. A 95% confidence interval for the yield monitor grid cell sample of 58 locations is 46.9 to 50.6 Bushels/Acre whereas the NASS objective yield is 41.3 to 51.0 or a range of more than double. To use a sample of only 18 square feet, the objective yield has required the assumption of much greater uniformity within each 900 square meter grid-cell than was found using the yield monitor. That is, choosing a small sample within the 900 square meter area will not necessarily represent the larger area. A larger range of grid-cell values (88.3 Bushels/Acre for the objective yield versus 38.4 Bushels/Acre for the yield monitor) for the objective yield sample shows that creation of the splined surface using the objective yield samples still maintains much of this greater variability created by expanding a small 18 square foot area of the objective yield sample to represent a much larger area. This variability of the surface remained although groups of 12 objective yield samples were used in creating the grid-cell value.

7.2 ANALYSIS OF FIELD W 1998 CORN YIELDS RELATIONSHIPS TO 1998 SOILS' VARIABLES

Using the same procedures as for field P, summing the yield monitor field W corn data weights within 30 meter grid cells permitted creating a total weight of grain for each grid cell. Correcting the field W corn yield monitor 30 meter grid cells to standard 15.5% moisture and dividing by 56 pounds/Bushel converted the data to corn Bushels/Acre. The resultant grid containing all the grid cells generated by the potential mapping procedure had a minimum value of 24.3 corn Bushels/Acre with a maximum of 192.3 corn Bushels/Acre for field W. Using the same process, the mean value for the corn yield monitor grid was 124.2 Bushels/Acre with a standard deviation of 37.04 Bushels/Acre. Unlike the field P soybeans yield monitor data, the corn 30 meter grid cell mean did not agree well with the weigh wagon of 139.9 Bushels/Acre (Willis, 1999). The mean of full set of gridded corn yield monitor data agreed neither with the calculated yield monitor mean using all the data of 134.7 Bushels/Acre nor with the NASS sample plots at 153.3 Bushels/Acre after adjustment for harvest loss. Differences in moisture readings are not sufficient to explain why the NASS estimates and the yield monitor data differ.

Field W moisture reading averages ranged from 13.8% to 18.5% from the yield monitor. The corn moisture mean was 16.1% with a standard deviation of 0.82. NASS moisture values for the same grid cells ranged from 11.7% to 15.9%. The NASS moisture mean was 13.3% with a standard deviation of 0.74 (Table 9). Indeed, Tables 9 and 10 (next two pages) provide the statistics of the NASS original sample data, the interpolated NASS objective yield grid-cells for the reduced set of 69 locations, the Purdue yield monitor potential mapped at 30 meter grids for the same 69 locations, and the soils' variables at the same locations (as originally recorded).

Mallarino, Oyarzaral, and Hinz (1999) present similar mean values as in Table 9 and correlations as in Table 10 for five corn fields in the Corn Belt. Field three in that article had the highest yield of 12.5 Mg/ha (121.6 Bu/Acre) for corn. That field had means for P, K, and pH similar to those for Purdue's Field W (Table 9), but the OM means differed. Correlations for the variables in field three of the article were also very similar with the Purdue field W. The remaining variables measured were not the same so no direct comparisons could be made for them. At least, this article helps in understanding that the data is within the realm of possibility. These other studies show the effects of soils' variables being within the ranges observed here. Soils' variables must be within certain ranges to permit corn yields of certain ranges, but weather variables will determine whether expected yields actually occur.

Table 9. Descriptive Statistics for the Field W, 1998 Variables: NASS Corn Yields, Yield Monitor Yields, and Soils Variables

Variables[1]												
	SY	INY	YMY	INM	MYM	OM	P	K	Mg	C	pH	CEC
Min	0.2	30.0	51.0	12.0	14.6	2.0	7.0	72	255	1000	5.4	9.0
Max	201.4	195.6	192.3	15.7	18.5	5.0	111	501	730	2650	7.7	21.0
Mean	153.3	154.8	137.4	13.4	16.2	3.3	28.3	143	476	1896	6.3	15.6
S. Dev.	28.1	23.4	25.1	0.8	0.8	0.76	18.2	58.4	100	400	0.58	2.7

[1] SY = Sample NASS Yields, INY = Interpolated NASS Yields, YMY = Yield Monitor Yields, INM = Interpolated NASS Sample Moisture, MYM = Yield Monitor Moisture Average, OM = Organic Matter (OM), P = Phosphorous (P), K = Potassium (K), Mg = Magnesium (Mg), C = Calcium (C), pH = pH Reading, CEC = Cation Exchange Capacity (CEC). NASS sample yields come from the original 64 sample points for field W. The Interpolated NASS yields and moisture readings with the yield monitor values correspond to the interpolated values at the remaining 69 soils sample locations for field W after deletion of grid cells containing less than 85 observations. The OM through CEC variables is based on the same 69 sample soils readings from the 1998 season.

Unlike field P, the NASS corn objective yield and Purdue corn yield monitor sample of 30 meter grid-cells mean corn yields differ and as well as the moisture values(paired t-test results follow):

Yields	Moisture
t = 5.5384	t = - 24.2836
df = 68	df = 68
p- value < 0.0001	p-value < 0.0001
Infers to reject no difference in yields	infers to reject no difference in moisture.

Note: p-values of near zero (0) do not require Bonferroni corrections.

===

Neither the NASS corn yields nor the yield monitor corn yields for the 30 meter grid cells located at the 69 locations of the field W soils sample show very large correlations with any other variable measured. The corn yields from the NASS objective yield grid-cells and the yield monitor 30 meter grid cells do show a positive correlation of 0.42, although their means differ. Although not a high correlation by any means, the positive correlation and different mean corn yield values for theses two grid cells exhibit exactly the opposite behavior as did the soybeans data for field P, since their means did differ according to the paired t-test (above). Examinations of Figure 11 and Figure 15, however, show that the contours do have less spatial relationship with each other than did those contours for field P (Figures 9 and 14). The greatest difference here is that the ranges of the contours exhibit areas of greater yields for the NASS objective yield contours than for the yield monitor contours.

Table 10. Correlations for the Field W, 1998 Variables: NASS Corn Yields, Yield Monitor Corn Yields and Soils Variables

Variables[1]										
	INY	YMY	INM	MYM	OM	P	K	Mg	C	pH
INY										
YMY	0.42									
INM	0.25	-0.04								
MYM	0.08	-0.07	0.27							
OM	0.14	-0.04	0.36	0.56						
P	-0.04	-0.37	0.24	0.31	0.55					
K	-0.10	-0.28	0.17	0.53	0.61	0.84				
Mg	-0.23	-0.21	-0.04	0.06	0.17	0.30	0.37			
C	0.00	-0.13	0.24	0.46	0.79	0.51	0.59	0.94		
pH	-0.23	-0.18	-0.08	-0.14	-0.17	0.13	0.10	0.76	0.33	
CEC	0.04	-0.12	0.26	0.49	0.84	0.57	0.65	0.55	0.94	0.08

[1] INY = Interpolated NASS Yields, YMY = Yield Monitor Yields, INM = Interpolated NASS Sample Moisture, MYM = Yield Monitor Moisture Average, OM = Organic Matter (OM), P = Phosphorous (P), K = Potassium (K), Mg = Magnesium (Mg), C = Calcium (C), pH = pH Reading, CEC = Cation Exchange Capacity (CEC). NASS sample yields come from the original 64 sample points for field W. The Interpolated NASS yields and moisture readings with the yield monitor values correspond to the interpolated values at the remaining 69 soils sample locations for field W after deletion of grid cells containing less than 85 observations. The OM through CEC variables is based on the same 69 sample soils readings from the 1998 season.

Unlike field P, this result of positive correlation with differing means for the two procedures, the yield monitor and the NASS objective yield data collection, shows that although the two procedures might be measuring similar characteristics of the corn field W, the NASS sample estimates have consistently higher corn yields for the 69 location sample of 30 meter grid cells than do the yield monitor sample grid cells for the same 69 locations. A 95% confidence interval for the corn yield monitor grid cell sample of 69 locations is 131.3 to 143.4 Bushels/Acre whereas the NASS objective yield 95% confidence interval is 149.2 to 160.5. The two confidence intervals have nearly the same range. However, the mean of the NASS corn objective yield 30 meter grid-cell sample values has the larger 154.8 Bushels/Acre value (that is, 16.6 Bushels/Acre larger).

To use a sample of only 15 square feet, the objective yield has required the assumption of much greater uniformity within each 900 square meter grid-cell than was found using the yield monitor. That is, choosing a small sample within the 900 square meter area will not necessarily represent the larger area. However, unlike the NASS objective yield soybeans' sample for field P, the NASS objective yield corn sample for field W provides a consistently higher yield value, but with a positive correlation of 0.42. One possible explanation is that the spacing of corn plants grown in field W should create lower yields even in a time of plentiful rainfall than the NASS model is assuming based on usual relationships between ear weight and number of ears per acre. If the small 15 square foot sample should have included only one plant more than is usually available within other 15 square foot areas within the field, then the expanded plants per acre estimate would be too large and thereby result in increased corn yield estimates.

8. DISCUSSION

This report has examined relationships among three sets of data: yield monitor with weigh wagon data, NASS objective yield sample locations, and two sets of soils data from 1994 and 1998. These sources of data were available for two fields, P and W, on Purdue University's Davis Research Farm for 1998. The purpose of this study was to provide an analysis of relationships among the three sets of data for each field and assist in showing how yield monitor data might aid in providing a better understanding of soybeans and corn yield variability within a field.

8.1 OPINIONS DERIVED FROM THIS STUDY

At this time, NASS cannot conclude that yield monitor data will be an important data set in determination of U.S. or State yields. Lack of uniformity, potential errors in analysis, and overall complexity of the yield monitor data would argue against its use in any surveys obtaining crop yields. Yield monitor data cannot at this time be a substitute for weigh wagon information, crop marketing (sales tickets) data, or on-farm storage volumes. Nor can it be a useful input into remote sensing studies without overcoming its multiple shortcomings. While these multiple limitations of the yield monitor yields continue to be significant, NASS data collection to use farmers' reported data from yield monitors is not recommended. Monitoring the degree of adoption of yield monitor data and its importance to farmers should be, at this time, the only potential interests that NASS should maintain.

The results of this study show quite different relationships between the NASS objective yield 30 meter interpolated grid cells and Purdue yield monitor 30 meter potential mapped grid cells for the two fields in the study. The objective yield grid cells for the NASS soybeans objective yield locations have a similar mean as the Purdue soybeans yield monitor 30 meter grid cells based on the t-test analysis. However, the NASS objective yield 30 meter interpolated grid cells values have a greater range and low correlation with the Purdue yield monitor 30 meter potential mapped soybeans' grid cells at the soils' sample locations. These facts would argue that the two procedures are estimating different processes for field P. Conversely, field W NASS estimated objective yield values at the soils' sample locations have a larger mean than do the Purdue corn yield monitor 30 meter potential mapped grid cells. They also have a positive correlation of 0.42 and a 95% confidence interval for the mean yield that is nearly the same as for the Purdue yield monitor corn 30 meter potential mapped grid cells at the soils' sample locations. These results seem to suggest that the two processes are similar, but that the NASS objective yield method of estimating the number of plants or ears per acre might be in question for this field.

The differences in soils variable information obtained for field W and that in the cited Mallarino article show how an accepted set of soils' properties can vary from one study to another. Also, the many soils' characteristics that are possible to examine can make selection of an affordable sample difficult. Of course, Organic Matter (OM), Phosphorous (P), Potassium (K), and pH are four variables often studied. Yield monitor data provided a different corn yield estimate for field W and had a different relationship to the soils variables (Tables 9 and 10). The yield monitor data does not appear to represent field W in the same way as does the NASS sample. Although this degree of difference may not always occur, this field provides a contrary example as follows: The yield monitor will not measure all fields in a way that will generate estimates close to those obtained from objective yield measurement sites. Complexity of the yield monitor data and the need to make frequent calibrations would argue that this might be a frequent occurrence.

Factors that increase the uniformity of crop yields across the field will help the effectiveness of the NASS objective yield methodology. Any detrimental soils' or drainage conditions within the field can increase the crop yield variability and thereby reduce the effectiveness of the NASS objective yield methodology. How well farmers follow good farming practices in the management of their fields by proper field preparation and the provision of required crop nutrients will also play a role in the uniformity of crop yields.

One remark regarding the uniformity of crop yields seems evident from the NASS objective yield sample data and the yield monitor data. Specifically, the possibility exists for rather large degrees of crop yield variability within a field. End rows, in particular, tend to have particularly large variations in crop yields, due to stress from weather conditions and poor soils' conditions. Since the NASS objective yield assumes uniformity of crop yields both at the sample level and the field level, there is evidence from these data sets that insufficient uniformity can be present in the fields, as in the Purdue field W corn data, for these assumptions to work well at least at the field level. Further studies would be necessary to learn the degree to which nonuniformity would be present at the State level. At the very least, the variability will increase the confidence interval range of the estimate determined by the NASS objective yield sample.

The author can catalog just a few of the possible difficulties that might cause the NASS objective yield methodology troubles, over time, as follows:

1) Widespread Planting of Numerous Crop Varieties within a State:
New crop varieties may change germination rates as well as how much uniformity of the crop plants throughout the field in response to soils' variables, weather, and planting rate,

2) Modern Farming Practices that Emphasize Denser Planting Rates:
The planting rate uniformity within the field can have profound effects on the resulting density of plant stands and thereby the resulting crop yields throughout the field,

3) Proper Field Management Practices to Maintain the Field's Soils' Properties:
Greater uniformity (not always observed, as is evident from fields P and W) in the soils' variables observed within a field should result in greater uniformity of resulting yields, and

4) Maintaining Field Drainage by Repair of Field Drainage Tiles: Improved drainage throughout the field with no broken drainage tiles should be a factor in more uniform yields throughout the field as, for example, broken tiles were related to decreased yields in Purdue field P (Willis, 1999).

Any of the above-enumerated possibilities can affect the accuracy of the NASS objective yield. Additional studies would be necessary to evaluate how consistent the results of this study would be under different cropping conditions. Since even small changes in conditions in other studies have resulted in quite different yields and correlations between yields and soils' variables, the possibility that relationships between the yield monitor data and the NASS objective yield estimates might differ for corn and soybeans is very likely. The small area and sample size that are necessary due to costs involved in making data collections, will always be a limitation for NASS in making accurate estimates for the State level. However, improved farming practices should aid in all four areas listed above and thereby improve the uniformity of the field yields.

The most important component of yield monitor information for NASS to examine further at this time would be a possible follow-up survey, every three years or so, to obtain the following information (as follows):

1) Survey farmers to determine those who have adopted use of yield monitors,

2) Question those farmers using yield monitors as to how valuable they have found the use of yield monitor data to be for them,

3) Question these same farmers regarding any difficulties that they might be having in using information obtained from yield monitors, and

4) Finally, question the degree of agreement that the farmers have found that their yield monitor yields have had to weigh wagon data, crop sales tickets, or on-farm storage volume.

9. REFERENCES

1. Ag Leader Technology, "Yield Monitor 2000, Operators Manual," Ames, Iowa, June 1997.

2. A and L Laboratories, "Soils Report on Field P and W," Fort Wayne, Indiana, August 1998.

3. Birrell, S. J. and K. A. Sudduth, "Nutrient Mapping Implications of Short-Range Variability," Precision Agriculture, Proceedings of the 3rd International Conference, American Society of Agronomy, Crop Science of America, and Soils Science Society of America, Minneapolis, Minnesota, p. 207, June 23-26, 1996.

4.4. Blackmore, S. and M. Moore, "Remedial Correction of Yield Map Data," in Precision Agriculture, Pierre Robert (Editor-In-Chief), Volume 1, Number 1, p.53, January 1999.

5. Brown, J. R., Editor, "Recommended Chemical Soil Test Procedures for the North-Central Region," Missouri Agricultural Experiment Station, January 1998.

6. Cressie, N. A. C., "Statistics for Spatial Data," Revised Edition, New York, John Wiley and Sons, 1993.

7. Doucette, P. and Beard, K., "Exploring the Capability of Some GIS Surface Interpolators for DEM Gap Fill," Photogrametric Engineering & Remote Sensing, ASPRS, vol. 66, no. 7, p. 881, July 2000.

8. Dunn, R. F., "Where Oh Where Does the Grain Go?", in Precision Ag Illustrated, Clear Window Multimedia, St. Louis, MO, December 1998.

9. Environmental Systems Research Institute, Inc., ArcView Spatial Analyst, Applied Spatial Analysis Using Raster and Vector Data, Redlands, California, 1996.

10. Huddleston, H. "Sampling Techniques for Measuring and Forecasting Crop Yields," Economics, Statistics, and Cooperative Service, USDA, 193 pages, 1978.

11. Mallarino, A., E. Oyarzabal, and Hinz, P., "Interpreting Within-Field Relationships Between Crop Yields and Soils and Plant Variables Using Factor Analysis," in Precision Agriculture, Pierre Robert (Editor-In-Chief), Volume1, Number 1, p. 15, 1999.

12. Missotten, B., Strubbe, G., de Baedemaeker, J., "Accuracy of grain and straw yield mapping," Precision Agriculture, Proceedings of the 3rd International Conference, American

Society of Agronomy, Crop Science of America, and Soils Science Society of America, Minneapolis, Minnesota, p. 713, June 23-26, 1996.

13. Nolan, S., Haverland, G., Goddard, T., Green, M., Penney, D., Henriksen, J., Lachapelle, G., "Building a Yield Map from Geo-referenced Harvest Measurements," Precision Agriculture, Proceedings of the 3rd International Conference, American Society of Agronomy, Crop Science of America, and Soils Science Society of America, Minneapolis, Minnesota, p. 885, June 23-26, 1996.

14. O'Neal, M., Frankenberger, J., Parson, S., Ess, D., Crisler, M., Strickland, R., "Correcting yield monitor data for improved yield mapping," Presented at the 2000 ASAE Annual International Meeting, Paper No. 001088, ASAE, St. Joseph, MI USA.

15. O'Neal, M., Frankenberger, J., Ess, D., Grant, R., "Spatial Variability of Precipitation as a Factor of Yield Variability," Presented at the 2000 ASAE Annual International Meeting, Paper No. 991144, ASAE, St. Joseph, MI USA.

16. Personal Communication, Mark Schleusener, NASS.

17. Pocknee, S., and others, "Directed Soil Sampling," Precision Agriculture, Proceedings of the 3rd International Conference, American Society of Agronomy, Crop Science of America, and Soils Science Society of America, Minneapolis, Minnesota, p. 159, June 23-26, 1996.

18. Precision Agriculture, Proceedings of the 3rd International Conference, American Society of Agronomy, Crop Science of America, and Soils Science Society of America, Minneapolis, Minnesota, June 23-26, 1996.

19. Rawlins, S., "Moving from Precision to Prescription Farming: The Next Plateau," Precision Agriculture, Proceedings of the 3rd International Conference, American Society of Agronomy, Crop Science of America, and Soils Science Society of America, Minneapolis, Minnesota, p. 283, June 23-26, 1996.

20. SISA, Simple Interactive Statistical Analysis, http:www.home.clara.net/sisa/bonhlp.htm.

21. Statistical Methods Branch, "The Yield Forecasting and Estimating Program of NASS," Washington, D.C., June 1998. NASS Staff Report Number SMB 98-01.

22. Top-Soil Testing Service, "Soil Test Report," Frankfort, Illinois, Spring, 1994.

23. Willis, P., "Evaluation of Crop Yield Estimation Using Yield Monitor Data, Remotely Sensed Imagery, and Hand-Harvest Data," M.S. Thesis, Purdue University, 1999.

APPENDIX A - YIELD MONITOR DATA CHARACTERISTICS AND PREPARATION

Although yield monitor data has many potential errors that make its use challenging, careful preparation of the data can reduce some errors. The simplest errors to correct are those of the lag, ramp-up, and ramp-down that are clearest from evaluating the data. The emphasis of the program FLDHARV, written in Delphi 3 by Martin Ozga, was the correction of only these most troublesome characteristics of the yield monitor data. Correction of other errors such as not keeping the header full, stray points because of GPS loss of signal, and incorrect distances between the recorded points were not addressed. The hope was that the use of potential mapping, that is, aggregation of the adjacent data points, would reduce the overall measurement errors. Use of a 30-meter grid size appeared to increase this averaging effect, so we did not evaluate smaller grid cell sizes. This grid cell size is one that would permit comparison with Landsat Thematic Mapper data that has a 30-meter pixel size. However, funding to purchase Landsat TM data for this study was not available.

Yield monitor data collected by Purdue University was from an AgLeader 2000 yield monitor (Willis, 1999). The latitude and longitude data came from a Vision System Omnistar 4000 DGPS with sub-meter accuracy. A Case 1460 harvester with a 15-foot wide (4.572 meters), that is, a six-row header, harvested the crop. An example of the collected data obtained is in Table A1. Although other data is in the file, the analysis of the data concentrated on the use of the following: latitude, longitude, flow, moisture, and header. Of course, GPS-provided latitude and longitude coordinates gave the location of each recorded point. The flow is the reading of the harvested crop's weight measured by a sensor for the corresponding area. Moisture readings provided the percentage of crop moisture. The header records provide information on the position of the harvester's header for being up (not harvesting) or down (harvesting).

Field Id provides the location of the harvester for being in the endrows (End) or within the primary part of the field (Bulk). The time shows that each location taken was one-second after the preceding location. Cycles confirm that each location recording was at one-second intervals. The distance provides a calculated distance between recorded locations while the pass is always one and so did not describe the travel of the harvester well. A swath was always 15 to represent the intended width in feet of the harvester header that was 15 feet wide. The serial number was zero (0) and referred to the serial number of the yield monitor. Load ID was always one and the grain was always beans for field P and corn for field W. DGPS gave a quality value for the GPS signal while the PDOP gave a measure of the GPS signal strength. The values for DGPS and PDOP were generally good for both fields P and W so that location errors were generally not due to inaccuracies in the GPS signal. Finally, the elevation values provided the altitude of each location. Field W was very level, so the values were quite uniform while field P had more variability. However, we know altitude readings from GPS devices to be somewhat inaccurate, so we attempted no analysis of these measurements.

Table A1. Uncorrected Yield Monitor Data Example: Purdue Davis Farm Field P, 1998

Longitude	Latitude	Flow	Time	Cycles	Distance	Swath	Moisture	Header	Pass	Serial	Field	Load	Grain	DGPS	PDOP	Elevation
										Num	ID	ID				
-85.152095	40.241835	0	148885	1	0	15	12.7	Up	1	0	Bulk	1	Beans	2	2.3	298
		Forty Observations with zero Flow Readings														
-85.151743	40.241840	2.86	148931	1	61	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	298
-85.151707	40.241838	2.84	148933	1	61	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	298
-85.151688	40.241838	2.19	148934	1	60	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	298
-85.151688	40.241838	2.67	148934	1	59	15	12.6	Down	1	0	Bulk	1	Beans	2	2.3	298
-85.151653	40.241838	4.41	148936	1	60	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	298
-85.151637	40.241838	3.44	148937	1	60	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151618	40.241838	3.12	148938	1	59	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	298
-85.151600	40.241840	3.45	148939	1	60	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151600	40.241840	2.87	148939	1	60	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151565	40.241840	2.59	148941	1	60	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151550	40.241840	4.54	148942	1	60	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151550	40.241840	4.11	148942	1	61	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151513	40.241840	2.87	148944	1	61	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151495	40.241840	4.41	148945	1	61	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151495	40.241840	4.83	148945	1	61	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151460	40.241842	4.29	148947	1	60	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151442	40.241842	3.11	148948	1	61	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151423	40.241843	2.89	148949	1	61	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151405	40.241843	2.42	148950	1	61	15	13.0	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151388	40.241843	2.36	148951	1	61	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151372	40.241843	1.96	148952	1	61	15	13.0	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151353	40.241845	2.36	148953	1	60	15	13.1	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151353	40.241845	2.52	148953	1	61	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151317	40.241845	2.54	148955	1	60	15	13.0	Down	1	0	Bulk	1	Beans	2	2.3	298
-85.151300	40.241845	2.27	148956	1	60	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151300	40.241845	2.49	148956	1	60	15	12.9	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151265	40.241847	2.55	148958	1	58	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	298
-85.151250	40.241845	2.66	148959	1	57	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151250	40.241845	2.56	148959	1	57	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151215	40.241847	2.60	148961	1	55	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151200	40.241845	2.63	148962	1	56	15	12.6	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151182	40.241845	3.15	148963	1	57	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151165	40.241845	2.80	148964	1	57	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151165	40.241845	4.16	148964	1	57	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151133	40.241845	4.71	148966	1	56	15	12.5	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151117	40.241847	5.36	148967	1	56	15	12.6	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151117	40.241847	5.79	148967	1	55	15	12.6	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151083	40.241848	5.64	148969	1	56	15	12.6	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151068	40.241847	4.74	148970	1	56	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151052	40.241847	4.01	148971	1	56	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151035	40.241845	4.64	148972	1	56	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	299
-85.151020	40.241845	5.61	148973	1	56	15	12.8	Down	1	0	Bulk	1	Beans	2	2.3	300
-85.151020	40.241845	5.71	148973	1	56	15	12.7	Down	1	0	Bulk	1	Beans	2	2.3	300

* AgLeader 2000 Yield Monitor soybeans' data for Purdue University field P for the 1998 crop (AgLeader, 1997).

APPENDIX B - FLDHARV: A PROGRAM TO CORRECT LAG, RAMP-UP, AND RAMP-DOWN OF YIELD MONITOR DATA

Use of the yield monitor continues to be operator dependent as for the saving of data. Ten files of uncorrected data for field P corresponded to the following: (1) two files for the endrows, (2) six files for the main section of the field, and (3) two files with only zero flow readings. Similarly, field W had 14 files with the following: (1) three files for the endrows, (2) nine files for the main section of the field, and (3) two files with all zero flow readings. Because the data within each file was not always adjacent, the processing of the data went from one file at a time. Of course, the category (3) files with zero flow readings were not used.

The purpose for FLDHARV was to correct the data locations for the lag, ramp-up, and ramp-down errors that are evident in all uncorrected yield monitor data-sets. Although many programs are commercially available to do this preprocessing, such programs usually make assumptions about the lag by using a constant value. Ramp-up and ramp-down variations are not usually considered for correction.

Development of FLDHARV came from some basic considerations about the data from field W. First, the definition of a row was set as starting when a header down occurred (no matter what flow value that observation had - including zero). Next, the program counted the number of downs with zero flow readings, the number of ups, and the number of ups with zero flow readings. Algorithms to adjust for the ramp-up and ramp-down effects had three cases:

1) Zero < = Number of header down observations with zero flows Minus Number of header up observations with positive flow < = 5,

2) -5 < = Number of header down observations with zero flow Minus number of header up observations with positive flows < = -1, and

3) If (1) and (2) are both false, then provide no correction to the recorded data.

For case (1), FLDHARV added together flow values for beginning and ending flow observations to adjust the number of observations with the header down and zero flow to equal the number of the header up with zero flows so that the correction for location of the lag at the beginning would not displace the locations of the points at the end of the row. Case (2) required adding together more values to adjust for the ramp-up and fewer values for the ramp-down.

FLDHARVworked better in field W (Table B.4). Running FLDHARV on field P (Table B.2) data corrected the data, however, greater variability in the number of header downs with rows defined for field P produced more variable results. However, evaluation of the data within ArcView showed that the corrections were satisfactory for input into potential mapping so FLDHARV was not updated for field P data. To handle other datasets successfully, more developmental work would be necessary. However, the basic philosophy of using the counts of the data types at the beginning and ending of the rows would be followed. The great amount of variability within the individual files as regards the number of zero downs (lags) within individual passes shows that at least for this data set, an assumption of a standard lag adjustment would often be incorrect and thereby provide reduced accuracy.

Output from FLDHARV included a file containing counts of the output file of corrected data and a file containing the following: the number of output header downs, the input number of header downs with zero flows, the input number of header ups, and the input number of header ups with zero flows for each row.

Table B.1. Purdue Field P: 1998 Soybeans Weights Recorded by Purdue Yield Monitor: Output from the FLDHARV program

File	Passes	H_Downs	Downs_0	H_Ups	Ups_0
2	1	259	15	17	11
2	2	27	12	15	11
2	3	179	15	14	10
2	4	11	11	12	6
2	5	210	12	11	9
2	6	13	12	12	11
2	7	10	10	11	10
2	8	208	13	13	11
3	1	199	16	1	1
3	2	31	0	5	5
3	3	31	2	11	9
4	1	211	16	11	8
4	2	13	12	12	12
4	3	14	12	13	12
4	4	208	13	14	13
5	1	30	12	13	11
5	2	22	13	12	11
5	3	226	15	14	11
5	4	220	13	12	9
5	5	221	13	12	10
5	6	227	19	14	10
5	7	230	11	13	10
5	8	35	13	18	16
5	9	205	13	14	11
5	10	218	12	13	10
5	11	208	13	13	11
5	12	195	10	12	10
6	1	237	13	15	11
6	2	12	11	18	18
6	3	12	12	17	17

6	4	234	20	13	11
6	5	231	14	16	11
6	6	214	23	17	12
6	7	228	12	13	11
6	8	194	20	14	12
6	9	226	13	17	11
6	10	17	12	15	13
6	11	18	12	14	13
6	12	220	13	15	11
6	13	219	12	14	11
6	14	218	12	12	10
6	15	224	12	14	11
6	16	219	13	12	10
7	1	227	14	15	12
7	2	15	10	13	12
7	3	13	11	18	17
7	4	13	11	16	15
7	5	212	14	14	10
7	6	219	12	14	11
7	7	216	13	15	12
7	8	216	12	14	11
7	9	218	13	13	11
8	1	237	13	14	11
8	2	23	12	14	12
8	3	22	11	16	12
8	4	221	14	14	10
8	5	188	13	3	3
8	6	30	0	12	11
8	7	219	12	13	11
8	8	188	13	3	3
8	9	29	0	14	12
8	10	34	11	5	5
8	11	90	0	1	1
8	12	96	0	13	10
8	13	46	15	4	4
8	14	22	0	14	11
8	15	136	16	12	11
8	16	43	11	14	12
8	17	218	12	13	10
8	18	222	13	15	11
8	19	17	11	14	11
8	20	18	12	14	12
8	21	16	12	14	13
8	22	55	15	1	1
8	23	15	0	4	4
8	24	13	0	7	7
8	25	15	5	1	1
8	26	132	0	15	10

8	27	224	13	14	10
8	28	221	14	12	10
8	29	220	13	15	11
8	30	218	11	13	10
8	31	225	12	13	8
8	32	23	10	13	11
8	33	21	11	14	13
8	34	221	12	13	10
8	35	222	12	13	10
8	36	59	12	3	3
8	37	167	0	13	11
8	38	226	14	13	10
8	39	226	12	14	11
8	40	241	12	13	10
8	41	223	12	12	10
8	42	249	12	13	11
8	43	18	11	13	12
8	44	31	11	12	11
8	45	22	11	12	10
8	46	51	11	15	12
8	47	130	12	15	11
8	48	243	13	13	10
8	49	223	12	13	11
8	50	242	12	13	10
8	51	125	11	2	2
8	52	21	0	2	2
8	53	15	0	4	4
8	54	75	0	13	10
8	55	24	13	3	3
8	56	20	0	6	6
8	57	218	3	15	11
8	58	20	12	1	1
8	59	73	0	5	5
8	60	29	1	1	1
8	61	40	0	3	3
8	62	112	0	12	10
8	63	283	14	10	9
8	64	247	12	15	11
8	65	258	13	13	11
8	66	229	11	1	1
8	67	43	0	1	1
8	68	14	0	37	32
9	1	225	14	4	4
9	2	9	0	8	7
9	3	192	14	13	10
9	4	139	12	12	10
9	5	222	13	13	11
9	6	218	13	11	10

9	7	228	12	10	7
9	8	223	14	17	12
9	9	27	26	5	5
9	10	105	23	13	11
9	11	28	16	7	7
9	12	227	14	14	11
9	13	248	12	12	10
9	14	197	13	8	5
9	15	231	13	12	10
9	16	76	14	0	0
9	17	23	20	2	2
9	18	7	7	3	3
9	19	196	12	16	15
9	20	214	12	14	10
9	21	188	19	15	14
9	22	62	12	3	2
9	23	116	39	3	2
9	24	207	12	14	12
9	25	32	12	1	1
9	26	161	0	19	14
Totals		Note: 2			
1038	3008	19671	1616	1664	1366

[1] Files = The File Number for Field P with Yield Monitor Data, Passes = Pass Defined by FLDHARV as defined above, H.Downs = Output Header Downs for the designated Pass, Downs.0 = Input Weights with Harvester Header in Down Position with Zero Flows for the designated Pass, H.Ups = Input Weights with Harvester Header in Up Position for the designated Pass, and Ups.0 = Input Weights Equal Zero (0) with Harvester Header in Up for the designated Pass.

2. The reduced number of weights (19,419) mentioned in the text was the result of deletion of 252 zero weights missed by FLDHARV.

Development of FLDHARV stressed the capability of handling the large degree of variability in lag, ramp-up, and ramp-down readings from the yield monitor. Often other commercial programs may use a standard lag value to adjust the locations of each pass (one continuous stream of yield monitor data that crosses the width or length of a field). Both Field P and W exhibit how a constant lag, ramp-up or ramp-down would not do well in correcting for these positional errors. Other methods of working with the yield monitor data that rely on correct positioning of the yield monitor data will not work well under these assumptions. Even kriging to interpolate a surface will not work correctly. Potential mapping in particular will be sensitive to incorrectly positioned data, although, the larger grid size of 30-meters used in this study will reduce the effects somewhat.

Some smaller rows (less than 100 observations) are the result of the harvester stopping and\or lifting the header possibly to avoid obstacles within the field. Adjustments to such passes within a file that had less than one-half of the median number for that file were not corrected since there would be the uncertainty of stopping and starting points for the pass and its relationship to other passes. Similarly, many observations exceeding three times the median within that file were not corrected either. Proportionately, these cases affected few of the passes within a field.

Table B2. Purdue Field P 1998 Soybeans Weights Recorded by Purdue Yield Monitor: Statistics of the File Output Information from the FLDHARV program

File	Passes	Statistics	H.Downs	Downs.0	H.Ups	Ups.0
2	8	Min	10	10	11	6
NA	NA	Max	259	15	17	11
NA	NA	Mean	114.6	12.5	13.1	9.9
3	3	Min	31	0	1	1
NA	NA	Max	199	16	11	9
NA	NA	Mean	87	6	5.7	5
4	4	Min	13	12	11	8
NA	NA	Max	211	16	14	13
NA	NA	Mean	111.5	13.3	12.5	11.3
5	12	Min	22	10	12	9
NA	NA	Max	230	19	18	16
NA	NA	Mean	169.8	13.1	13.3	10.8
6	16	Min	12	11	12	10
NA	NA	Max	237	23	18	18
NA	NA	Mean	170.2	14	14.8	12.1
7	9	Min	13	10	13	10
NA	NA	Max	227	14	18	17
NA	NA	Mean	149.9	12.2	14.7	12.3
8	68	Min	13	0	1	1
NA	NA	Max	283	16	37	
NA	NA	Mean	119.7	8.6	10.5	8.6
9	26	Min	7	0	0	0
NA	NA	Max	248	39	19	15
NA	NA	Mean	146.2	14.2	9.6	7.9

[1] Files = The File Number for Field P with Yield Monitor Data, Passes = Pass Defined by FLDHARV as defined above, H.Downs = Output Header Downs for the designated Pass, Downs.0 = Input Weights with Harvester Header in Down Position with Zero Flows for the designated Pass, H.Ups = Input Weights with Harvester Header in Up Position for the designated Pass, and Ups.0 =

Input Weights Equal Zero (0) with Harvester Header in Up for the designated Pass.

Table B3. Field W 1998 Corn FLDHARV Output: Weights Recorded by the Purdue Yield Monitor

File	Passes	H_Downs	Downs_0	H_Ups	Ups_0
1	1	1	1	0	0
1	2	182	24	18	17
1	3	63	15	19	12
1	4	227	19	21	16
1	5	29	21	19	18
1	6	204	16	19	14
1	7	204	19	18	13
1	8	29	15	20	15
2	1	238	17	17	13
2	2	239	14	16	12
2	3	235	14	15	12
2	4	236	13	15	11
2	5	234	14	17	12
2	6	240	14	14	11
2	7	236	13	18	12
2	8	237	14	16	11
2	9	236	14	16	12
2	10	234	13	18	12
2	11	235	13	19	12
2	12	240	13	15	11
2	13	237	13	18	13
2	14	240	14	16	13
2	15	234	13	17	13
2	16	233	12	18	13
2	17	237	13	16	11
2	18	238	15	19	13
3	1	258	19	20	15
3	2	30	14	31	27
4	1	240	15	17	12
5	1	11	11	19	19
5	2	20	15	21	17
5	3	21	15	18	17
5	4	259	14	23	17

5	5	27	19	18	15
5	6	161	19	19	16
5	7	13	13	16	12
5	8	281	25	20	14
5	9	18	17	19	17
5	10	266	15	17	13
5	11	70	17	17	14
6	1	236	15	15	12
7	1	230	11	17	10
7	2	237	14	17	12
7	3	239	15	16	13
7	4	243	16	17	13
7	5	238	14	19	13
7	6	235	14	17	13
7	7	241	14	16	12
7	8	235	14	18	12
7	9	243	14	21	13
7	10	238	14	18	13
7	11	234	12	16	13
7	12	240	16	17	12
7	13	237	13	16	12
7	14	239	13	16	12
7	15	239	14	16	13
7	16	239	14	17	13
7	17	240	13	18	13
7	18	238	14	17	12
7	19	237	13	15	12
7	20	236	14	17	13
7	21	234	13	17	13
7	22	237	14	17	13
7	23	236	14	17	13
7	24	237	14	16	12
7	25	239	14	15	11
7	26	236	14	15	11
7	27	239	15	16	12
7	28	237	15	15	11
7	29	238	15	16	13
8	1	236	14	17	13
8	2	240	13	17	12
8	3	238	14	18	13
8	4	239	13	16	12
8	5	236	14	17	12
8	6	236	13	15	12
8	7	14	13	20	19
8	8	225	13	19	12
8	9	239	13	17	12
8	10	241	15	16	12
8	11	239	13	15	10

8	12	240	16	17	12
8	13	196	13	17	13
10	1	69	17	19	12
10	2	25	14	17	15
10	3	36	14	18	12
10	4	50	12	18	15
10	5	1	1	0	0
10	6	21	9	18	14
10	7	36	16	15	13
11	1	196	14	19	13
11	2	195	12	20	13
11	3	199	17	18	13
11	4	194	13	14	10
11	5	193	14	18	13
11	6	197	13	16	11
11	7	198	14	16	13
11	8	196	15	16	11
11	9	195	14	17	12
13	1	238	16	18	13
13	2	237	15	20	13
13	3	238	15	18	12
13	4	241	15	17	13
13	5	235	15	17	13
13	6	239	13	17	13
13	7	236	12	17	13
13	8	237	14	18	13
13	9	235	15	17	13
13	10	231	15	16	13
13	11	231	14	18	13
13	12	236	13	15	11
13	13	232	13	17	12
14	1	236	14	17	13
14	2	193	15	18	13
14	3	236	14	17	11
14	4	191	15	16	11
Totals		Note 2.			
816	978	22742	1652	1982	1484

[1] Files = The File Number for Field P with Yield Monitor Data, Passes = Pass Defined by FLDHARV as defined above, H.Downs = Output Header Downs for the designated Pass, Downs.0 = Input Weights with Harvester Header in Down Position with Zero Flows for the designated Pass, H.Ups = Input Weights with Harvester Header in Up Position for the designated Pass, and Ups.0 = Input Weights Equal Zero (0) with Harvester Header in Up for the designated Pass.

2. The deletion of 173 zero weights after FLDHARV ran gave the reduced number of weights (22,569) as mentioned in the text.

Table B4. Purdue Field W 1998 Corn Weights Recorded by Purdue Yield Monitor: Statistics of the File Output Information from the FLDHARV program

File	Passes	Statistic	H.Downs	Downs.0	H.Ups	Ups.0
1	8	Min	1	1	0	0
NA	NA	Max	227	24	21	18
NA	NA	Mean	117.4	16.3	16.8	13.1
2	18	Min	233	12	14	11
NA	NA	Max	240	1	19	13
NA	NA	Mean	236.6	13.7	16.7	12.1
3	2	Min	30	14	20	15
NA	NA	Max	258	19	31	27
NA	NA	Mean	144	16.5	25.5	21
4	1	Min	240	15	17	12
NA	NA	Max	NA	NA	NA	NA
NA	NA	Mean	NA	NA	NA	NA
5	11	Min	11	11	16	12
NA	NA	Max	281	25	23	19
NA	NA	Mean	104.3	16.4	18.8	15.5
6	1	Min	236	15	15	12
NA	NA	Max	NA	NA	NA	NA
NA	NA	Mean	NA	NA	NA	NA
7	29	Min	230	11	15	10
NA	NA	Max	243	16	21	13
NA	NA	Mean	237.6	13.9	16.7	12.3
8	13	Min	14	13	15	10
NA	NA	Max	241	16	20	19
NA	NA	Mean	216.8	13.6	17	12.6
10	7	Min	1	1	0	0
NA	NA	Max	69	17	19	15
NA	NA	Mean	34	11.9	15	11.6
11	9	Min	193	12	14	10
NA	NA	Max	195	13	16	11
NA	NA	Mean	195.9	14	17.1	12.1
13	13	Min	231	12	15	11
NA	NA	Max	241	16	20	13
NA	NA	Mean	235.8	14.2	17.3	12.7
14	4	Min	191	14	16	11
NA	NA	Max	236	15	17.3	13
NA	NA	Mean	214	14.5	17	12

[1] Files = The File Number for Field P with Yield Monitor Data, Passes = Pass Defined by FLDHARV as defined above, H.Downs = Output Header Downs for the designated Pass, Downs.0 = Input Weights with Harvester Header in Down Position with Zero Flows for the designated Pass, H.Ups = Input Weights with Harvester Header in Up Position for the designated Pass, and Ups.0 =

Input Weights Equal Zero (0) with Harvester Header in Up for the designated Pass.

APPENDIX C - NASS OBJECTIVE YIELD SAMPLES: FIELDS P AND W

Collection of the NASS objective yield samples for fields P and W required considerable effort by a group of twelve NASS enumerators led by Ralph Gann. They sent the sample data to our Illinois State Statisitical Office (SSO) for evaluation. The laboratory did the grain counts on the soybeans pods from field P, the corn ears from field W, and the moisture measurements for both the pods and corn kernels. Counts of plants for the soybeans and stalks for the corn were all done in the field. Table C1 provides descriptive statistics for field P soybeans' data while Table C2 provided field W data descriptive statistics for corn. Although comparisons to the yield monitor data are available in this paper, further analysis of this data would require comparison with available NASS objective yield data for corn and soybeans from additional fields for Indiana.

Table C1. Field P 1998 Soybeans NASS Variables: Descriptive Statistics

Variables[1]													
	RA	RB	RC	RD.	PA	PB	APd	BPd	PdD	PdU	Abn	Mst	AYld
Min	0	0	1	2	34,848	34,848	0	0	21	0	35.8	7.8	2.7
Max	21	17	17	15	302,016	302,016	155.5	147.5	447	4	206.6	8.2	101.1
Mean	7.3	8.2	8.6	8.2	187,959	187,960	72.1	76.2	184.0	0.74	109.7	8.0	47.8
SD	4.5	3.9	3.4	3.1	54,805	54,806	36.1	33.6	86.9	1.0	37.1	0.12	22.2
LCL	6.2	7.2	7.7	7.4	173,548	173,549	62.6	67.2	161.2	0.48	99.9	8.0	41.9
UCL	6.5	9.3	9.5	9.0	202,369	202,370	81.6	85.1	206.9	1.00	119.4	8.0	53.6

[1] RA = Plants in Row A, RB = Plants in Row B, RC = Plants in Row C, RD. = Plants in Row D, PA = Population Row A, PB = Population Row B, APd = Row A Pods, BPd = Row B Pods, PdD = Pods Rows A and B Summed, PdU = Ratio of Pods with Beans to Total Fruit, Mst = Pod Moisture, Ayld = Calculated Soybeans Yield Bushels\Acre after deletion of 2.02 Bushels\Acre Harvest Loss Average.

Table C2. Field W 1998 Corn NASS Variables: Descriptive Statistics

Variables[1]													
	Stlk1	Stlk2	Er1	Er2	ErW1	ErWt	GrnWt	Moist	EarsA	WtEar	Fract	lbEar	AYld
Min	16	9	3.0	3	0.2	54	37	11.7	3484.8	0.067	0.69	0.47	0.2
Max	33	27	26.0	27	11.3	549.7	454.9	15.8	30,201.6	0.491	0.85	0.43	201.4
Mean	23.7	23.3	23.2	22.8	9.0	326.7	268.9	13.4	26,698.7	0.385	0.82	0.32	153.3
SD	2.0	2.8	2.9	3.5	1.5	82.6	70.4	0.89	3455.4	0.06	0.029	0.011	28.1
LCL	23.2	22.6	22.4	21.9	8.6	306.1	251.3	13.1	25,835.5	0.37	0.81	0.31	146.3
UCL	24.2	24.0	23.9	23.7	9.3	347.3	286.5	13.6	27,561.8	0.40	0.83	0.34	160.3

[1] Stlk1 = Stalks in Row 1, Stlk2 = Stalks in Row 2, Er1 = Ears in Row 1, ErW1 = Ears Weight in Row 1, GrnWt = Grain Weight from Ears in Row 1 plus Row 1, Moist = Moisture of Grain, EarsA = Number of Estimated Ears\Acre (Calculated), Fract = Shelling Fraction (Calculated), EarsA = Number of Estimated Ears\Acre (Calculated), lbEar = Estimated weight of shelled corn from on ear (pounds), Ayld = Calculated Corn Yield Bushels\Acre after deletion of 2.7 Bushels\Acre Harvest Loss Average.

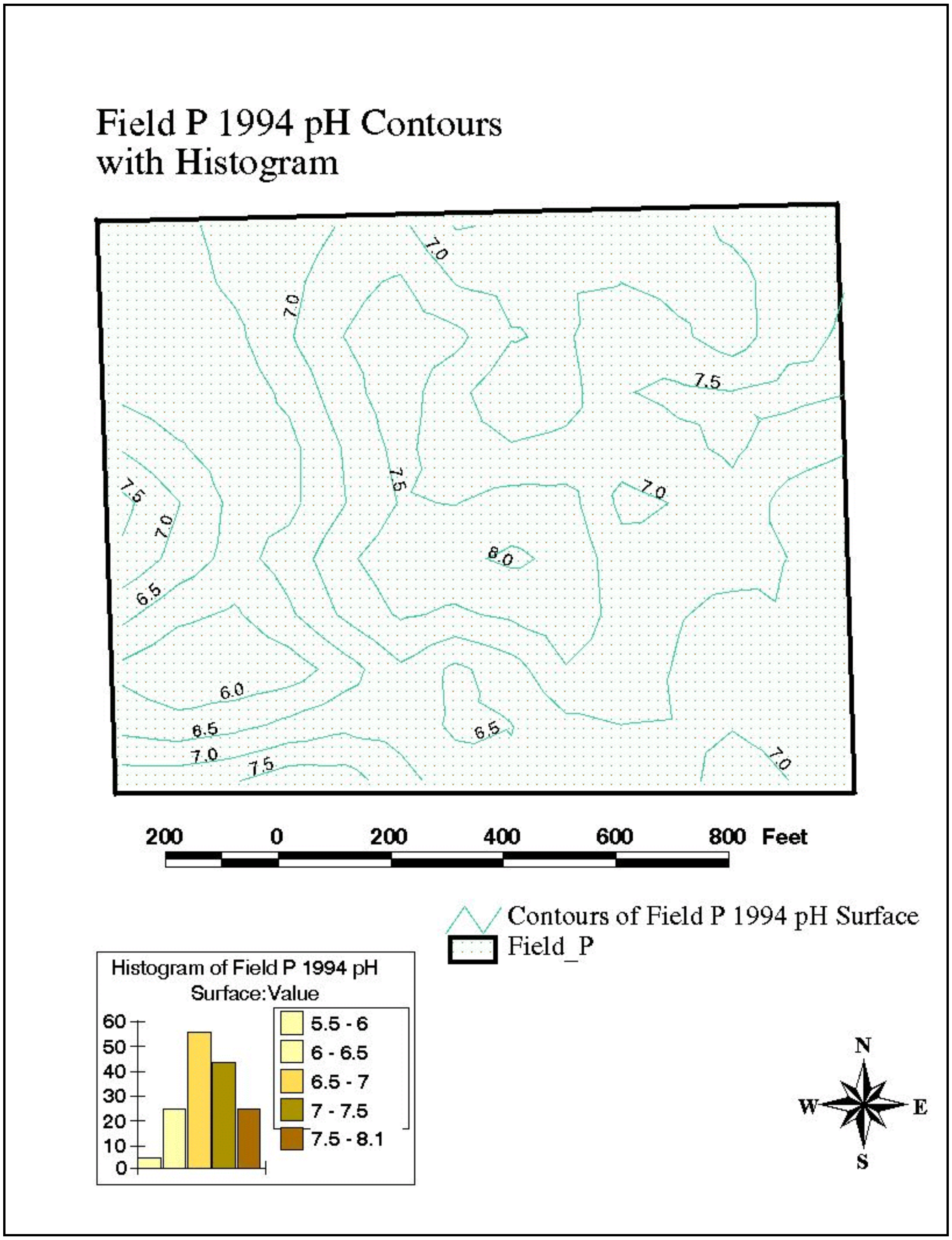
.

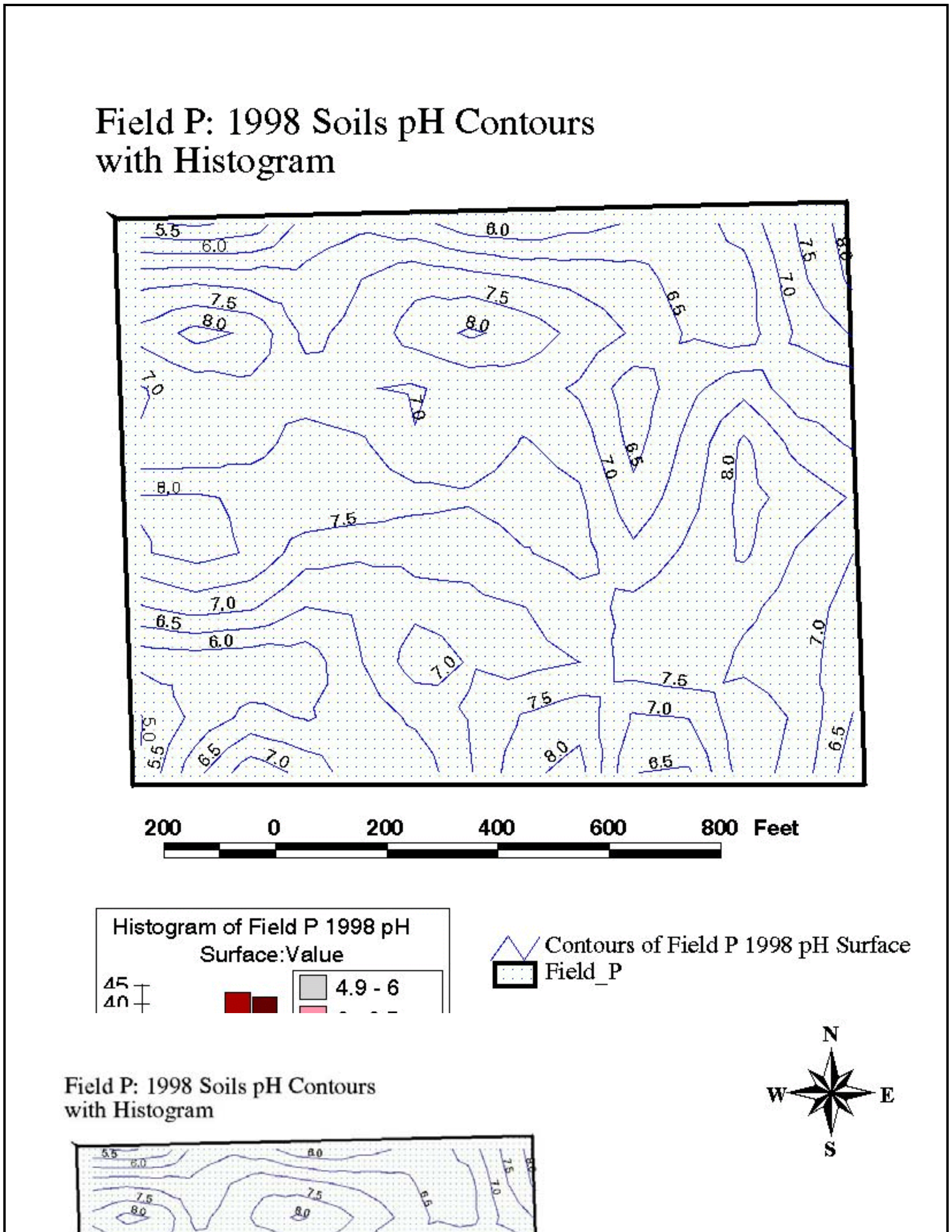
APPENDIX D - FIELD P SOILS' VARIABLE MAPPING FOR 1994 AND 1998: EXAMPLES

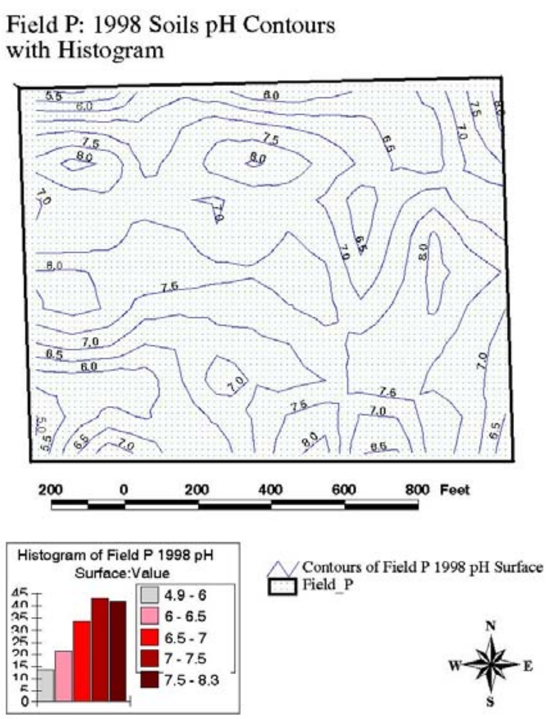
Creation of the soils' variables surfaces for the 1994 and 1998 years followed the previously described procedures. Specifically, splining the soils' sample point information using ArcView's Spatial Analyst to create 30-meter gridded surfaces as a first step. Then, the Spatial Analyst helped to create contours from the gridded surfaces at predefined intervals. Although color maps of the gridded soils' variables are usually easy to interpret, the black and white images used here are easier to interpret as contours. Histograms of the counts of the 30-meter gridded values within each field are provided with each Figure to give an idea of the distribution of interpolated soils' values within each gridded image. Since the gridded surfaces are not highly accurate, these graphs give a representation of the relative ranges of soils' values within the field while the contours provided in the Figures give an idea of the spatial distribution of those values within the field.

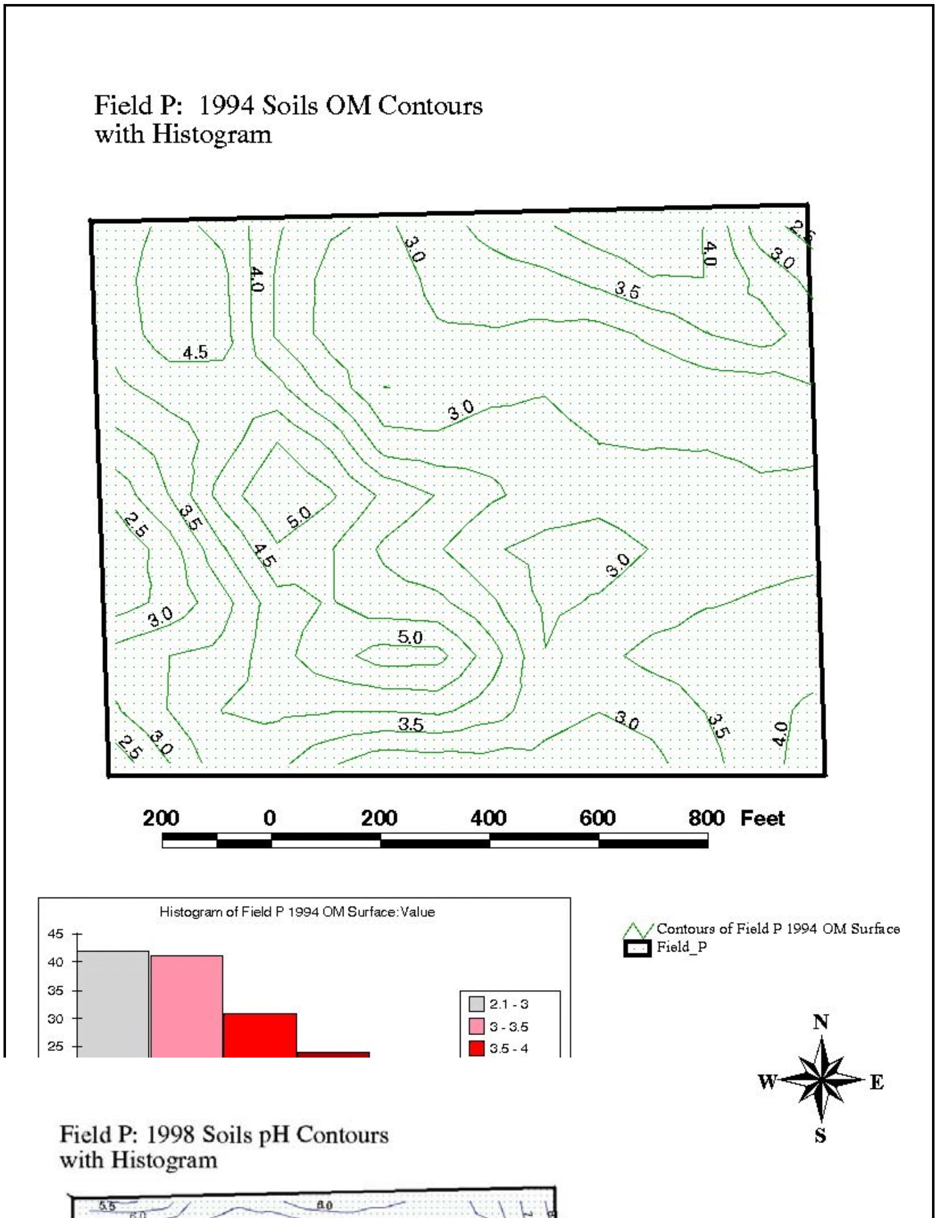
The first two Figures (Figure 16 and Figure 17) provide contours of soil pH values for field P with accompanying graphs of the number of grid cells for each range of the contours provided in the years 1994 and 1998. The typical nomenclature for pH readings is to characterize them as "acid" below seven, "neutral" at seven, and "alkaline" above seven. Although the paired t-test found the distributions of the test results to be similar for the two years of data, the soils' Figures give a different impression of the spatial distribution of the pH contours within the field. Counts in the histogram for 1994 would appear that the field is more acid. However, in 1998 the field might be more alkaline since the number of 30-meter grid cells within the field is larger in number.

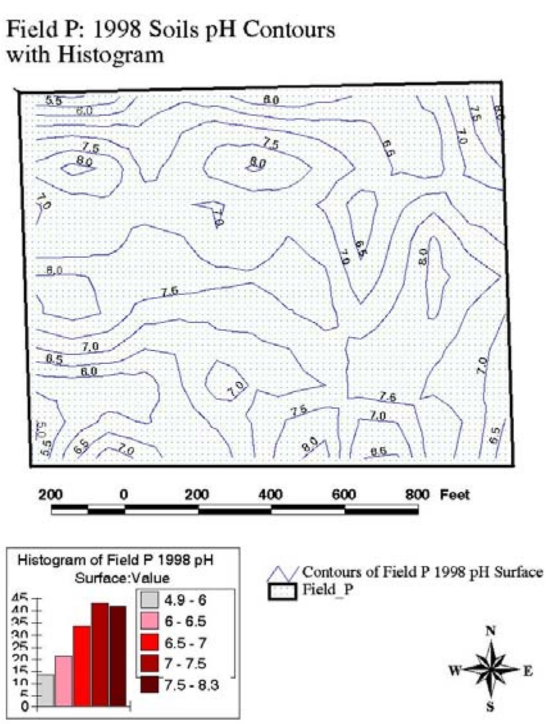
The next two Figures (Figure 18 and Figure 19) are the field P soils Organic Matter (OM) contours for 1994 and 1998, respectively, with associated histograms of the gridded OM polygons. Although the range of 3% to 3.5% occurred most frequently in both years, the spatial distribution for the two years is quite different with more variability in 1994 than in 1998. The western parts of the field in particular exhibits lower Organic Matter (OM) values.


Field P 1994 pH Contours
with Histogram
7.0
7.5
6.5
6.0
8.0
200
0
200
400
600
800 Feet
Contours of Field P 1994 pH Surface
Field_P
Histogram of Field P 1994 pH
Surface:Value
60
50
40
30
20
10
0
5.5 - 6
6 - 6.5
6.5 - 7
7 - 7.5
7.5 - 8.1
N
W
E
S


Field P: 1998 Soils pH Contours
with Histogram
200 0 200 400 600 800 Feet
Histogram of Field P 1998 pH
Surface:Value
4.9 - 6
Contours of Field P 1998 pH Surface
Field_P
N
W
E
S


Field P: 1998 Soils pH Contours
with Histogram
200 0 200 400 600 800 Feet
Histogram of Field P 1998 pH
Surface:Value
4.9 - 6
6 - 6.5
6.5 - 7
7 - 7.5
7.5 - 8.3
Contours of Field P 1998 pH Surface
Field_P
N
W
E
S


Field P: 1994 Soils OM Contours
with Histogram
200 0 200 400 600 800 Feet
Histogram of Field P 1994 OM Surface:Value
2.1 - 3
3 - 3.5
3.5 - 4
Contours of Field P 1994 OM Surface
Field_P
N
W
E
S


Field P: 1998 Soils pH Contours
with Histogram
200 0 200 400 600 800 Feet
Histogram of Field P 1998 pH
Surface:Value
4.9 - 6
6 - 6.5
6.5 - 7
7 - 7.5
7.5 - 8.3
Contours of Field P 1998 pH Surface
Field_P
N
W
E
S

Field P 1998 Soils OM Contours with Histogram