Impacts of Observation Deleting Standards

on Profitability Analysis of Precision Agriculture

Baohui Song, Ph.D. (Contact author)
Assistant Professor
California State University, Chico/College of Agriculture
229 Plumas Hall
Chico, CA 95929-0310
Phone: (530) 898-3056
E-mail: bsong@csuchico.edu

Carl Dillon, Ph.D.
Associate Professor
University of Kentucky/Department of Agricultural Economics
417 C. E. Barnhart Bldg.
Lexington, KY 40546-0276
Phone: (859) 257-7283
E-mail: cdillon@uky.edu

Tom Mueller, Ph.D.
Associate Professor
University of Kentucky/Plant and Soil Sciences
N-122G, Ag. Science Bldg. North
1100 Nicholasville Road
Lexington, KY 40546-0091
Phone: (859) 257-8887
E-mail: mueller@uky.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Dallas, TX, February 2-6, 2008

Impacts of Observation Deleting Standards
on Profitability Analysis of Precision Agriculture

Abstract: This research explores a possible reason for the inconsistent results from previous study on the profitability analysis in precision agriculture—different standards of identifying possible erroneous observations for PA datasets. By comparing the results from the different standards of identifying possible erroneous observations, this research raises concerns about the negative impacts of different standards of identifying possible erroneous observations on the profitability analysis of PA, and provides some suggestions for the standard which could be used in the future profitability analysis of PA.

Introduction

Precision agriculture (PA) refers to the practice of modifying management techniques to fit specific areas of the farm instead of treating the whole farm with same techniques. The main goal of PA is to increase farming profitability and to conserve environmental resources. Precision agriculture technology has evolved to include a vast number of practices that provide information for producers in making production decisions. Some main techniques used in PA include but not limited to yield monitor, variable rate input application, infrared mapping, light bar, GPS unit, and grid soil sampling. The adoption rates of PA technologies in the U.S. for main crops are shown in table 1 (USDA, 2004). Data in table 1 show that, in general, PA technology adoption rates for corn and soybeans are higher than other crops. In 2001, over one-third of corn produced in the U.S. adopted yield monitor, one-fourth adopted soil map, and almost 14% adopted yield map. For soybeans, the adoption rate for yield monitor was 28.7%, for soil map was 11.2%, and for yield monitor was 10.7%.

Table 1: USDA Survey Results of PA Adoption Rate in the U.S.

Crop and Year	Yield Monitor	Yield Map	Geo-referenced Soil Map	Remotely Sensed Map	Guidance System
Corn (2001)	36.5	13.7	25	3.4	6.9
Soybeans (2002)	28.7	10.7	11.2	1.7	6.8
Wheat (2000)	9.1	<1	12.2	3.9	NA*
Cotton (2003)	1.7	<1	4.8	4.6	5.9

*NA = not available.

The variable rate technology (VRT) adoption rates for main crops in the U.S. are shown in table 2 (USDA, 2004). The survey results indicate that the adoption rates of

VRT in crop production were much lower than other PA technologies. Only less than 10% of corn, 5% soybeans adopted VRT of fertilizers in the United States.

Table 2: USDA Survey Results of VRT Adoption Rate in the U.S.

Crop and Year	Fertilizer	Seed	Pesticides
Corn (2001)	9.8	2.4	3.8
Soybeans (2002)	5	<1	1.3
Wheat (2000)	3.1	<1	<1
Cotton (2003)	3.7	1	1.6

One possible reason for the low adoption rate for VRT in the main crop production in the U.S. may related the past research results. Currently, no consensus conclusion has been drawn that whether or not the adoption of VRT technology in production practice is profitable. After an updated review of recent literatures based on the previous work done by Peone, et al. in 2004, Bullock and Lowenburg-DeBoer (2006) pointed out that "The empirical results on the profitability of generating information and using variable rate technology are mixed. Some studies find information seeking and VRT to be profitable, and others find just the opposite." They also mentioned that the three possible reasons for the inconsistence in the research results include "insufficient use of spatial analysis", "the need for longer-term data, and the need for ex ante analysis". In this research, another possible reason for this inconsistence will be explored—different observation deleting standards for possible erroneous observations could have significant impacts on the results of profitability analysis of VRT.

In the research of precision agriculture (PA), the datasets are normally very large. One dataset may contain many thousands of observations. However, not all these observations can be used in the empirical estimation directly due to some possible erroneous observations. Before using the dataset in empirical estimation, the possible erroneous observations need to be identified and eliminated. However, which standard to use in identifying and eliminating the possible erroneous observations could have significant impacts on the final coefficient estimation. In the PA literature, not many papers convey what standards were used to identify and eliminate those possible erroneous observations. Therefore, it is difficult to compare research findings from different projects pertaining to the profitability analysis of PA technology adoption. This research raises concerns about this problem and provides some suggestions for reference in future research in PA profitability analysis.

In this research, focusing on the impacts of different standards used to identify those possible erroneous observations on the profitability analysis of PA technology, specific objectives include:

1. To apply different standards of identifying possible erroneous observations to three different datasets, then estimate the production function and compare the estimation results from different standards;

2. To conduct the comparative analysis of the impacts of the estimated coefficients on the production functions resulting from the different standards;

3. To raise concerns about the impacts of different standards of identifying possible erroneous observations on the profitability analysis of PA; and

4. To provide some suggestions about the standards that could be used to identify those possible erroneous observations.

Experiment and Data

Site Description

The study was conducted during the 2002, 2003, and 2004 growing season at three sites, located at Worth and Dee Ellis Farms in Eminence, KY (**coordinates and elevation**). The soils at the Ellis sites were comprised of a mixture of generally well drained Lowell and moderately well drained Nicholson silt loam soils (fine, mixed, active, mesic Typic Hapludalfs and fine-silty, mixed, active, mesic Oxyaquic Fragui). This location had been in a corn (*Zea mays* L.)-soybean (*Glycine max* L.)-wheat (*Triticum aestivum* L.) rotation for more than 20 years. The soils at the Preston sites consisted of a mixture of a Cumberland silty clay loam (fine, mixed, semiactive, thermic Rhodic Paleudalf) and a well drained Crider silty clay loam (fine-silty, mixed, active mesic Typic Paleudalf). This location had been cropped in a corn (*Zea mays* L.)-soybean (*Glycine max* L.) rotation. Corn at these sites was grown utilizing no-tillage practices.

Experimental treatments and field design

Rows were spaced 30 inches apart and seeding rates were on the order of 26,000 to 28,000 seeds per acre. The producer applied 10-15 lbs of a 28% UAN nitrogen solution concurrently with corn plantings to ensure adequate nutrients for seed development. Lime, phosphorous (P), potassium (K), herbicide, and insecticide were applied by the producer according to recommended rates.

Experimental plots were identified across the entire field using ArcGIS software (ESRI, Redlands, CA). An experimental unit was classified as an application rate applied to a 100 by 100 foot block occupied by mid season growth corn. A randomized complete block design was utilized. Location was identified as the extraneous source of variation, and plots were blocked accordingly. N application rates were randomly applied to plots within each block. The number of replications varied according to field size.

Four N application levels (120, 140, 160, 180 lb N/acre) were applied to the 2002 and 2003 sites and 6 application levels (50, 90, 120, 160, 180, 210 lb N/acre) were applied to the 2004 sites. Sidedress application of the different N rates occurred in late May/early June using a variable rate applicator. This applicator utilizes a Rawson controller driving a positive displacement piston pump, allowing for map based N application with a 28% UAN nitrogen solution.

Comparison of Different Observation Deleting Standards and Their Impacts

In this part, two basic observation deleting measures will be reviewed and compared: absolute value measure (Shears Method) and standard deviation measure. The absolute measure refers to setting an interval for possible values of the observations based on experts' opinions or designed traits of the machine. For example, by expert opinion, the highest possible corn yield in Kentucky for 2002 is 180 bushels/acre, and any observations with yield value higher than 180 bushels/acre will be deleted from the dataset. For moisture data, possible range is between 10-35%, and any observation with a value outside this range will be deleted. By the designed trait of the harvester, the travel speed should between 25 and 140 inches per second. If an observation value of harvester travel distance is less than 25 inches or greater than 140 inches will be deleted. The attractiveness of this measure is that the result from this measure is more reasonable and logical from the technical or production practice perspectives. However, this measure also has its weakness: these absolute values could vary a lot from region to region, from year to year, or from machine to machine due to different soil types, different weather conditions, or different machine designs. Another weakness of this measure is that the standard set based on expert opinions is subjective decision. Different experts may have different standards. It may increase the chances of human errors. In addition, it is difficult to compare the research results from different research due to the different standards set by different researchers/experts.

The other measure is the standard deviation measure: first calculate the standard deviation of the variable; then, if an observation is outside a certain standard deviations away from its mean it could be erroneous and should be deleted. One advantage of this measure is that if all the researchers choose to use this measure in the PA analysis, it is easy to compare the research results from different researchers, different projects, or different locations. Another advantage is that this measure is easy to use in data management. However, this measure has its weakness too. Observations deleted by this measure are purely decided from the statistical perspective. It is difficult to interpret this measure from the technical or practical perspectives. Another problem is to decide how many standard deviations should be adopted to delete those possible erroneous observations.

In this research, the two measures are compared by applying them to three different datasets: the Ellis Farm 2002, 2003 and 2004 datasets. For convenience, we define different standards to be compared as follows:

M0: Original dataset, raw dataset (after deleting the observations with missing values);

MA: Absolute value measure, by expert opinion, using the absolute value to set a possible range for the observations. Any observations outside the range will be deleted;

MSD4-MSD1: Standard deviation measure, if an observation deviates a certain standard deviations away from the mean, it will be deleted. In this research, four standards will be compared: 4 to 1 standard deviations separately (*MSD4*, *MSD3*, *MSD2*, and *MSD1*).

Table 3, table 4, and table 5 show the changes of observation numbers and mean values for yield, moisture and distance for Ellis Farm 2002-2004 datasets when applying different observation deleting standards on these three datasets. Taking table 3 as an example, using different observation deleting standards, observation number reduces from 32074 for *M0* to 15548 for *MSD4* for 2002 dataset. However, from *MA* to *MSD2*, observation numbers do not change much. The big changes occur at *MSD2* and *MSD1*. Compared to the original observation number, the percentage changes of the observation numbers for *MA, MSD4-MSD1* are: 0.86%, 0.32%, 0.90%, 6.28% and 51.52%, respectively, meaning that using 2 and 1 standard deviation standards will reduce the observation number significantly. Similar results can be found for 2003 and 2004 datasets. For the 2003 dataset, the percentage changes of the observation numbers for *MA, MSD4-MSD1* are: 2.33%, 0.16%, 3.46%, 10.13%, and 38.24%, respectively. For the 2004 dataset, the percentage changes of the observation numbers for *MA, MSD4-MSD1* are: 2.00%, 1.25%, 3.01%, 8.13%, and 37.19%, respectively. These tables indicate that under *MSD1*, too many observations will be eliminated from the datasets, causing possible bias when drawing conclusion based on this standard. Therefore we will not consider this standard in the following discussion.

Next, we look at the impacts of different observation deleting standards on the mean values of the selected variables. Setting the original mean value (*M0*) as the base, the means of these three variables (yield, moisture, and distance) do not change much as shown in tables 3, 4, and 5. Taking the 2002 dataset as an example (table 3), the changes of means of yield, moisture, and distance are less than 1% for the measures *MA*, *MSD4*, *MSD3*, and *MSD2*. Therefore, we can conclude that different observation deleting standards do not have much impact on the mean values of the selected variables.

Although different observation deleting standards do not impose much impact on mean values of the selected variables, they may have impacts on the estimated coefficients of production function. We assume that corn yield (*YD*) will response to nitrogen application (*N*), different soil types (*S*), deep electronic conductivity (*DPEC*), and shallow electronic conductivity (*SHWEC*). In addition, we also assume the square terms of N, DPEC, and *SHWEC* as well as the interaction terms between *DPEC* and *N* (*DPEC_N*), and *SHWEC* and N (*SHWEC_N*) also affect corn yield. Based on the above assumptions, a spatial error production function is defined as equation (1) (Discussion of the selection of spatial model will be discussed in another paper):

$$\begin{aligned} YD = \alpha_0 + \alpha_1 S1 + \alpha_2 S2 + \alpha_3 S3 + \alpha_4 N + \alpha_5 DPEC + \alpha_6 SHWEC + \alpha_7 N^2 \\ \alpha_8 DPEC^2 + \alpha_9 SHWEC^2 + \alpha_{10} DPEC\_N + \alpha_{11} SHWEC\_N + \varepsilon \end{aligned} \quad (1)$$

where $\varepsilon = \lambda\varepsilon + u$, λ is a spatial Euclidian distance weight matrix (for detailed information of weight matrix, see Luc Auslin (2003)), and u is i.i.d. error term.

The production function was estimated by GeoDa™ 0.9, developed by Luc Anselin (2003). Table 7, table 8 and table 9 report the estimation results of these spatial error models from 5 different observation deleting standards (*M0, MA, MSD4, MSD3*, and *MSD2*) for the Ellis Farm 2002, 2003, and 2004 datasets respectively. From these tables we find that:

(a) The magnitude of the estimated coefficients changes from standard to standard. For 2003 dataset, the estimated coefficient of *N* decreases from 1.53 for *M0* to 0.61 for *MSD2*, a 60% decrease. The estimated coefficient for *S3* increases from -31.38 for *M0* to -17.97 for *MSD2*, a 43% increase. For 2004 dataset, the estimated coefficient of N decreases from 0.42 for *M0* to 0.32 for *MSD2*, a 24% decrease.

(b) The significance of a variable changes. For 2002 dataset, the variable $DPEC^2$ changes from significant for *M0*, *MA*, *MSD4*, and *MSD3* to insignificant for *MSD2* (at a 5% significance level, same through out this paper). For 2003 dataset, the intercept changes from insignificant for *M0*, *MA*, *MSD4*, and *MSD3* to significant for *MSD2*. For 2004 dataset, the variable $SHWEC^2$ changes from insignificant for *M0*, *MA*, and *MSD4* to significant for *MSD3* and *MSD2*.

The above discussion indicates that different observation deleting standards will lead to different estimated coefficients. As a result, the different observation deleting standards will affect the results of profitability analysis of VRT. We use equation (2) to derive the different impacts of different observation deleting standards on estimated production function:

(2) $\Delta YD_{M0,j} = f(\hat{\beta}(DM_{M0}), N, \& Z) - f(\hat{\beta}(DM_j), N, \& Z)$, where j=*MA*, *MSD4*, *MSD3*, and *MSD2*.

Where $\Delta YD_{M0,j}$ is the difference between the estimated yield from raw data *(M0*) and the estimated yields from observation deleting standard DM_j. The yield is a function of estimated parameters, $\hat{\beta}$, nitrogen application, N, and other exogenous variables, Z. The estimated parameter $\hat{\beta}$ is a function of different observation deleting standards,

DM_j. The exogenous variables include different soil types, deep and shallow electronic conductivities, and the interaction terms between nitrogen and electronic conductivities.

The results of the changes of the estimated yields between *M0* to other observation deleting standards (*MA, MSD4-MSD2*) are reported in table 9. These results show that for the Ellis Farm 2002 dataset, the estimated yield decreases from *M0* to *MA*, *MSD4* and *MSD3* about 5-6% and for *MSD2* less than 1%. For the 2003 dataset, the estimated yield decreases 45% from *M0* to *MSD3*, and 16% from *M0* to *MSD2*. For the 2004 dataset, the estimated yield decreases 5-6% from *M0* to *MSD4*, *MSD3*, and *MSD2*. In conclusion, different observation deleting standards do have significant impacts on the estimated yield. As a result, this will definitely affect the profitability analysis of PA.

Caution on the Selection of Observation Deleting Standards on PA Analysis

The above analysis shows that different observation deleting standards could impose significant impacts on the profitability analysis results of PA. Therefore, when we conduct profitability analysis on PA, we need to be very cautious to select an observation deleting standard. In the first part, we compared the two different data deleting measures. From the technical and production practice perspectives, the absolute value measure (Shears measure) is preferred to standard deviation measure, which is purely derived from statistics perspective. However, as discussed in the previous session that Shears measure has its disadvantages. From the convenience perspective as well as for the easy comparison among different researches, the standard deviation measure is recommended. The question is how many standard deviations to use. Table 10 reports yield correlations among different observation deleting standards for 2002, 2003, and 2004 datasets. Results show that the highest correlation occurs between MA and MSD3 for all three datasets, meaning that using 3-standard deviation measure gives the closest result to the absolute value measure for this research (2002, 2003, and 2004 datasets from Kentucky farm). Therefore, we suggest that when choosing an standards to delete the possible erroneous observations 3 standard deviation standard could be used.

Table 10: Yield Correlation among Different Data Management Measures

	Correlation	MA	MSD4	MSD3	MSD2	MSD1
2002 Data	MA	1				
	MSD4	0.56	1			
	MSD3	**0.73**	0.59	1		
	MSD2	0.31	0.22	0.37	1	
	MSD1	0.09	0.05	0.09	0.24	1
	Correlation	MA	MSD4	MSD3	MSD2	MSD1
2003 Data	MA	1				
	MSD4	0.22	1			
	MSD3	**0.53**	0.26	1		
	MSD2	0.43	0.14	0.54	1	
	MSD1	0.15	0.05	0.19	0.36	1
	Correlation	MA	MSD4	MSD3	MSD2	MSD1
2004 Data	MA	1				
	MSD4	0.34	1			
	MSD3	**0.65**	0.63	1		
	MSD2	0.39	0.31	0.50	1	
	MSD1	0.12	0.10	0.15	0.31	1

Conclusions

Conflict results exist in the profitability analysis of PA technologies. One possible reason—different research used different data deleting standards for those possible erroneous observations—is explored in this paper. Three year datasets (2002, 2003, and 2004) from Kentucky Ellis Farm were used in this research. Results show that different data deleting standards do have significant impacts on the profitability analysis of PA technology.

Two basic measures—absolute value measure and standard deviation measure were compared regarding to their impacts on the analysis results of PA technology. From the technical and product practice perspective, the absolute value measure is preferred. However, it brings the difficulty to compare the analysis results from different researches. To avoid this problem, standard deviation measure with 3 standard deviation standard is recommended in the future research. In future study, more dataset analyses from other states, or other research projects should be conducted to confirm the conclusion from this research.

Reference:

Anselin, Luc. “GeoDa™ 0.9 User’s Guide”. *Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics University of Illinois, Urbana-Champaign Urbana, and Center for Spatially Integrated Social Science,* June 15, 2003. Website: http://sal.agecon.uiuc.edu/.

Bullock, David S., and James Lowenburg-DeBoer. “Using Spatial Analysis to Study the Value of Variable Rate Technology and Information.” Invited paper prepared for presentation *at the International Association of Agricultural Economists Conference,* Gold Coast, Australia, August 12-18, 2006.

U.S. Department of Agriculture, Economics Research Service (USDA-ERS), 2004. Briefing Rooms: Agricultural Chemicals and Production Technology: Questions and Answers. Website: http://www.ers.usda.gov/Briefing/ARMS/, accessed April, 2007.

Lambert, Dayton M. and James Lowenberg_DeBoer. “A Comparison of Four Spatial Regression Models for Yield Monitor Data: A Case Study from Argentina.” *Precision Agriculture*, 5, 579-600, 2004

Gao, Xiaolu, Yasushi Asami, Chang-Jo, F. Chung. “An Empirical Evaluation of Spatial Regression Models.” *Computers & Geosciences*, 32, 1040-1051, 2006.

Ripley, B. D.. “Modelling Spatial Patterns.” *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 2, 172-212, 1977.

USDA, ERS, “Precision Agriculture Adoption Continues to Grow.” *Agricultural Outlook,* November 2002. Washington, DC.

Haslett, John. “Spatial Data Analysis-Challenges.” *The Statistician*, Vol. 41, No. 3, Special Issue: Conference on Applied Statistics in Ireland, 1991. pp. 271-284, 1992.

Richardson; S., C. Guihenneuc, and V. Lasserre. “Spatial Linear Models with Autocorrelated Error Structure.” *The Statistician*, Vol. 41, No. 5. (1992), pp. 539-557.

.

Table 3: Descriptive Summary of Ellis Farm 2002 Dataset (Selected Variables)

Measure	Original (M0)			Absolute Value Measure (Yield<180) (35>Moisture>10) (140>Distance>25 (MA)			Standard Deviation Measure (> 4 std) (MSD4)			Standard Deviation Measure (> 3 std) (MSD3)			Standard Deviation Measure (> 2 std) (MSD2)			Standard Deviation Measure (> 1 std) (MSD1)		
	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST
N	32074	32074	32074	31798	31798	31798	31970	31970	31970	31784	31784	31784	30059	30059	30059	15548	15548	15548
Min	0.13	3.41	1.00	0.13	10.50	25.00	0.13	3.41	11.00	0.13	3.41	24.00	20.16	3.41	37.00	46.13	4.06	49.00
Max	483.68	30.00	125.00	179.71	30.00	96.00	175.92	30.00	96.00	149.94	30.00	96.00	124.01	30.00	87.00	98.09	30.00	74.00
Mean	72.11	20.22	61.62	71.75	20.22	61.97	71.74	20.21	61.71	71.50	20.21	61.87	71.39	20.21	62.13	70.78	20.55	59.89
Std	25.98	2.62	12.71	24.51	2.58	12.19	24.61	2.62	12.52	24.09	2.62	12.27	21.70	2.63	11.56	13.73	2.55	7.31

Note: YLD: Corn yield, (bushels/acre);
MST: Moisture (%);
DST: Distance (inches/second);
Std: Standard Deviation;
N: Observation Number;

Table 4: Descriptive Summary of Ellis Farm 2003 Dataset (Selected Variables)

Measure	Original (M0)			Absolute Value Measure (Yield<180) (35>Moisture>10) (140>Distance>25 (MA)			Standard Deviation Measure (> 4 std) (MSD4)			Standard Deviation Measure (> 3 std) (MSD3)			Standard Deviation Measure (> 2 std) (MSD2)			Standard Deviation Measure (> 1 std) (MSD1)		
	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST
N	7741	7741	7741	7561	7561	7561	7729	7729	7729	7473	7473	7473	6957	6957	6957	4781	4781	4781
Min	0.10	0.26	6.00	0.10	11.01	25.00	0.10	0.26	8.00	18.36	0.77	20.00	67.99	11.87	31.00	117.63	18.91	42.00
Max	441.68	29.20	80.00	298.72	29.20	80.00	354.43	29.20	80.00	306.28	29.20	79.00	266.35	29.20	74.00	216.81	29.20	63.00
Mean	167.16	23.50	52.61	168.63	23.49	53.41	167.03	23.50	52.67	171.61	23.49	53.37	176.97	23.44	54.49	172.92	23.46	54.25
Std	49.65	1.81	11.18	46.25	1.76	9.97	49.20	1.81	11.08	41.22	1.78	10.19	31.16	1.72	8.82	21.97	1.66	5.49

Note: YLD: Corn yield, (bushels/acre);
MST: Moisture (%);
DST: Distance (inches/second);
Std: Standard Deviation;
N: Observation Number;

Table 5: Descriptive Summary of Ellis Farm 2004 Dataset (Selected Variables)

Measure	Original (M0)			Absolute Value Measure (Yield<180) (35>Moisture>10) (140>Distance>25 (MA)			Standard Deviation Measure (> 4 std) (MSD4)			Standard Deviation Measure (> 3 std) (MSD3)			Standard Deviation Measure (> 2 std) (MSD2)			Standard Deviation Measure (> 1 std) (MSD1)		
	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST	YLD	MST	DST
N	12381	12381	12381	12133	12133	12133	12226	12226	12226	12008	12008	12008	11375	11375	11375	7776	7776	7776
Min	12.03	5.37	5.37	12.03	14.07	25.29	44.67	5.37	11.80	81.86	5.37	23.04	119.45	5.37	34.29	156.93	5.37	44.97
Max	460.62	20.27	123.65	299.22	20.27	123.65	341.84	20.27	100.05	303.35	20.27	87.68	269.56	20.27	77.56	232.04	18.87	66.89
Mean	194.48	16.18	55.90	194.15	16.19	56.62	195.13	16.18	56.02	196.11	16.18	56.47	197.26	16.20	57.22	198.45	16.24	57.67
Std	37.23	0.57	11.05	34.31	0.55	9.86	33.80	0.57	10.35	30.64	0.57	9.60	26.72	0.56	8.38	18.43	0.54	5.63

Note: YLD: Corn yield, (bushels/acre);
MST: Moisture (%);
DST: Distance (inches/second);
Std: Standard Deviation;
N: Observation Number;

Table 6: Estimation Comparison among Different Data Management Measures for Ellis Farm 2002 Dataset

Measure	Original (M0)			Absolute Value Measure (Yield<180) (35>Moisture>10) (140>Distance>25 (MA)			Standard Deviation Measure (> 4 std) (MSD4)			Standard Deviation Measure (> 3 std) (MSD3)			Standard Deviation Measure (> 2 std) (MSD2)		
Variable	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|
Intercept	69.4191	6.3620	<0.001	65.4511	5.8027	<0.001	66.3395	5.8513	<0.001	66.2344	5.6901	<0.001	67.4077	5.4618	<0.001
S1	2.8600	2.3812	0.2297	2.2647	2.2103	0.3055	2.1397	2.2272	0.3367	1.9835	2.1716	0.3610	2.0349	1.9893	0.3063
S2	-0.4428	1.4320	0.7572	-1.0631	1.3586	0.4339	-1.2185	1.3643	0.3718	-1.8467	1.3387	0.1677	-1.1820	1.2374	0.3395
S3	-0.5480	1.2813	0.6689	-0.6579	1.2182	0.5892	-0.5861	1.2230	0.6318	-0.6895	1.1995	0.5654	0.0700	1.1148	0.9500
N	0.1547	0.0484	0.0014	0.1419	0.0437	0.0012	0.1338	0.0442	0.0025	0.1352	0.0429	0.0016	0.1278	0.0418	0.0022
DPEC	-0.4840	0.4490	0.2810	-0.5091	0.4090	0.2132	-0.4073	0.4120	0.3228	-0.4497	0.4012	0.2624	-0.5736	0.3854	0.1367
SHWEC	0.0815	0.5407	0.8802	0.3809	0.4907	0.4376	0.2804	0.4956	0.5715	0.2795	0.4809	0.5612	0.3523	0.4547	0.4384
N^2	-0.0006	0.0001	<0.001	-0.0005	0.0001	0.0001	-0.0005	0.0001	0.0002	-0.0005	0.0001	0.0001	-0.0005	0.0001	0.0001
$DPEC^2$	-0.0104	0.0052	0.0460	-0.0099	0.0048	0.0406	-0.0113	0.0048	0.0191	-0.0108	0.0048	0.0231	-0.0071	0.0046	0.1260
$SHWEC^2$	-0.0032	0.0076	0.6766	-0.0057	0.0070	0.4159	-0.0046	0.0070	0.5097	-0.0039	0.0068	0.5718	-0.0059	0.0064	0.3559
DPEC_N	0.0112	0.0023	<0.001	0.0112	0.0021	<0.001	0.0111	0.0021	<0.001	0.0112	0.0020	<0.001	0.0104	0.0019	<0.001
SHWEC_N	-0.0101	0.0027	0.0002	-0.0104	0.0025	<0.001	-0.0104	0.0025	<0.001	-0.0103	0.0024	<0.001	-0.0098	0.0023	<0.001
λ	0.8736	0.0048	<0.001	0.9000	0.0042	<0.001	0.8956	0.0043	<0.001	0.9012	0.0041	<0.001	0.8967	0.0042	<0.001
Number of Observations			32074			31798			31970			31784			30059
R^2			0.50			0.55			0.54			0.56			0.54

Note: Para.: Parameter Estimate;
S.E.: Standard Error.

Table 7: Estimation Comparison among Different Data Management Measures for Ellis Farm 2003 Dataset

Measure	Original (M0)			Absolute Value Measure (Yield<180) (35>Moisture>10) (140>Distance>25 (MA)			Standard Deviation Measure (> 4 std) (MSD4)			Standard Deviation Measure (> 3 std) (MSD3)			Standard Deviation Measure (> 2 std) (MSD2)		
Variable	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|
Intercept	21.6073	28.2305	0.4440	9.9007	26.8743	0.7126	20.4609	27.8138	0.4619	42.5757	26.4838	0.1079	144.5950	25.4157	<0.001
S1	-16.1481	6.0526	0.0076	-13.5612	5.8475	0.0204	-15.8048	5.9833	0.0083	-10.3158	5.3365	0.0532	-5.8083	4.1759	0.1643
S2	2.1131	5.1329	0.6806	1.5487	4.7491	0.7443	2.4400	5.0613	0.6297	-2.1268	4.4393	0.6319	-1.2766	3.3940	0.7068
S3	-31.3754	3.9124	<0.001	-32.4883	3.7231	<0.001	-30.8441	3.8550	<0.001	-27.1378	3.6697	<0.001	-17.9713	3.1516	<0.001
N	1.5293	0.1838	<0.001	1.6334	0.1796	<0.001	1.5291	0.1803	<0.001	1.4793	0.1866	<0.001	0.6100	0.2128	0.0042
DPEC	-11.6539	6.0308	0.0533	-10.1831	5.6558	0.0718	-10.7513	5.9418	0.0704	-15.7788	5.6307	0.0051	-13.7148	4.9554	0.0056
SHWEC	20.0897	8.8379	0.0230	19.3804	8.2375	0.0186	18.8906	8.6840	0.0296	22.2080	8.1466	0.0064	9.3877	7.3984	0.2045
N^2	-0.0023	0.0006	0.0001	-0.0026	0.0005	<0.001	-0.0022	0.0006	0.0001	-0.0028	0.0006	<0.001	-0.0006	0.0007	0.3736
$DPEC^2$	0.3226	0.2059	0.1171	0.2444	0.1924	0.2041	0.2938	0.2035	0.1488	0.2974	0.1784	0.0956	0.4541	0.1453	0.0018
$SHWEC^2$	-0.8433	0.3900	0.0306	-0.7291	0.3656	0.0461	-0.7794	0.3841	0.0424	-0.7499	0.3426	0.0286	-0.4517	0.2953	0.1261
DPEC_N	0.0288	0.0238	0.2270	0.0251	0.0223	0.2606	0.0281	0.0233	0.2290	0.0546	0.0237	0.0214	0.0182	0.0225	0.4185
SHWEC_N	-0.0920	0.0358	0.0101	-0.0886	0.0331	0.0074	-0.0918	0.0351	0.0088	-0.1001	0.0338	0.0030	-0.0286	0.0324	0.3772
λ	0.8970	0.0083	<0.001	0.9017	0.0081	<0.001	0.9024	0.0080	<0.001	0.8589	0.0102	<0.001	0.8285	0.0116	<0.001
Number of Observation	7741			7561			7729			7473			6957		
R^2	0.60			0.26			0.26			0.22			0.17		

Note: Para.: Parameter Estimate;
S.E.: Standard Error.

Table 8: Estimation Comparison among Different Data Management Measures for Ellis Farm 2004 Dataset

Measure	Original (M0)			Absolute Value Measure (Yield<180) (35>Moisture>10) (140>Distance>25 (MA)			Standard Deviation Measure (> 4 std) (MSD4)			Standard Deviation Measure (> 3 std) (MSD3)			Standard Deviation Measure (> 2 std) (MSD2)		
Variable	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|	Para.	S.E.	Pr > |t|
Intercept	171.2306	11.1959	<0.001	171.0288	10.3247	<0.001	172.4636	10.1871	<0.001	170.9018	9.1914	<0.001	176.5564	8.0809	<0.001
S1	3.4295	1.7929	0.0558	2.9039	1.6687	0.0818	2.9869	1.6457	0.0695	2.6661	1.4816	0.0720	1.6322	1.2970	0.2082
S2	2.2615	2.3531	0.3365	1.6294	2.2102	0.4610	1.5835	2.1815	0.4679	0.9540	1.9593	0.6263	-0.2902	1.6688	0.8619
S3	-1.2823	2.9225	0.6608	-0.7256	2.7437	0.7914	-0.6309	2.7105	0.8160	-0.6663	2.4408	0.7849	-1.8487	2.1498	0.3898
N	0.4179	0.0980	<0.001	0.4127	0.0888	<0.001	0.4195	0.0875	<0.001	0.3672	0.0789	<0.001	0.3207	0.0703	<0.001
DPEC	1.9654	1.6171	0.2242	1.9564	1.4954	0.1908	0.6297	1.4784	0.6702	0.8200	1.3310	0.5378	0.4471	1.1647	0.7011
SHWEC	-2.7814	2.3038	0.2273	-2.8857	2.1218	0.1738	-1.4964	2.0966	0.4754	-0.8128	1.8865	0.6666	-0.6863	1.6625	0.6797
N^2	-0.0011	0.0003	0.0002	-0.0012	0.0003	<0.001	-0.0012	0.0003	<0.001	-0.0009	0.0002	0.0002	-0.0009	0.0002	<0.001
$DPEC^2$	-0.0054	0.0364	0.8815	-0.0123	0.0339	0.7180	0.0121	0.0336	0.7182	0.0043	0.0302	0.8871	-0.0011	0.0263	0.9669
$SHWEC^2$	-0.0755	0.0608	0.2140	-0.0650	0.0566	0.2507	-0.0878	0.0560	0.1169	-0.1060	0.0503	0.0352	-0.0972	0.0443	0.0283
DPEC_N	-0.0123	0.0074	0.0944	-0.0117	0.0067	0.0808	-0.0066	0.0066	0.3210	-0.0071	0.0060	0.2349	-0.0028	0.0053	0.5966
SHWEC_N	0.0141	0.0111	0.2065	0.0150	0.0101	0.1386	0.0080	0.0100	0.4238	0.0086	0.0090	0.3407	0.0073	0.0080	0.3566
λ	0.7636	0.0103	<0.001	0.7939	0.0095	<0.001	0.7930	0.0095	<0.001	0.7934	0.0095	<0.001	0.7700	0.0101	<0.001
Number of Observations	12381			12133			12226			12008			11375		
R^2	0.34			0.39			0.38			0.40			0.38		

Note: Para.: Parameter Estimate;
S.E.: Standard Error.

Table 9: Changes of Estimated Yields from Different Management Measures

Measure	Original (M0)	Absolute Value Measure (Yield<180) (35>Moisture>10) (140>Distance>25) (MA)			Standard Deviation Measure (> 4 std) (MSD4)			Standard Deviation Measure (> 3 std) (MSD3)			Standard Deviation Measure (> 2 std) (MSD2)		
Year	Estimated Yield	Estimated Yield	Change from M0	Percent	Estimated Yield	Change from M0	Percent	Estimated Yield	Change from M0	Percent	Estimated Yield	Change from M0	Percent
2002	82.48	78.10	-4.38	-5.32%	77.49	-4.99	-6.05%	78.08	-4.40	-5.33%	82.65	0.17	0.21%
2003	168.87	180.21	11.34	6.71%	155.32	-13.55	-8.03%	92.57	-76.30	-45.18%	142.44	-26.43	-15.65%
2004	207.56	206.14	-1.42	-0.68%	196.32	-11.24	-5.42%	194.97	-12.59	-6.07%	194.60	-12.96	-6.24%

Note: By equation (4), we calculated the profit changes from different data management measures. We take the average of 16 different scenarios: 4 different soil types combined with 4 different nitrogen application levels (120, 140, 160, 180 lbs/acre). For the deep electronic conductivity and shallow electronic conductivity, we took the means. If the variable is not statistically significant at a 5% significance level, we set the parameter estimate as zero in the final calculation for production functions.