IMPROVING FARM MANAGEMENT DECISION MAKING: EXPERIENCES FROM SPATIAL ANALYSIS OF YIELD MONITOR DATA FROM FIELD-SCALE ON-FARM TRIALS

T.W. Griffin, J.M. Lowenberg-DeBoer and R.J.G.M. Florax

Site-Specific Management Center
Spatial Analysis @ Purdue
Dept. of Agricultural Economics
Purdue University
West Lafayette, Indiana

ABSTRACT

Typical yield monitor data analysis may use data subjected to a default correction process of standard farm level software. Recent advances to farm-level software include the integration of professional data filtering capabilities. This paper presents results from both traditional and spatial analyses of differing levels of yield data quality including: raw harvester measurements, accepting default settings from a range of popular standard farm software packages, and a conscious removal of erroneous points and flow delay adjustments. The decisions that the rational farm manager would make for each of these scenarios are presented, answering the question of the importance of quality yield data. Human capital costs necessary to filter yield monitor data and whether the benefits of better decisions outweigh these costs are considered. Results suggest that farmers using yield monitors can make reliable farm management decisions based on their limited replication field-scale experiments if the data are of a sufficient quality and appropriate spatial statistical inference are conducted.

Keywords: yield monitor data, spatial analysis, data filtering

INTRODUCTION

The introduction of Global Positioning Systems (GPS) into commercial agricultural production systems have allowed farmers to test inputs and rates before making farm management decisions across large areas of the farm. Yield monitor technology has driven the resurgence of on-farm testing because farmers can collect yield measurements without interfering with other harvest-time operations. Farmers conducting on-farm trials base their production decisions on this yield data. The overall goal of our work is to help them make better decisions from the data that they are collecting.

Over the last several years, researchers have analyzed numerous farm-level datasets while developing appropriate spatial analysis techniques for field-scale on-farm comparisons using precision agriculture technologies. Because suitable spatial analysis methods were not widely available or understood at the farmer, consultancy, and university outreach levels, rudimentary analyses continue to be conducted and farm management decisions based on their results. Whether advanced spatial econometric or rudimentary analyses are conducted, yield monitor data at some level of data quality is used. Typically, yield monitor data is filtered with the default settings of the farm level software used to import raw data as recorded by the harvester. Recently, advanced yield monitor data filtering software has become available and integrated into existing farm-level software packages. Questions have arisen on the value of increased effort and human capital needed to subject yield data to filtering protocols.

The general objective of this study was to document commonalities among proper spatial analyses of field-scale experiments so that farmers and those that advise them may have better tools on which to base decisions. Specific objectives of the paper include: 1) to report on commonalities among the most appropriate analysis techniques across crops, treatments, and experimental designs, and 2) document differences in harvester yield monitor data quality for farm management decision making. The overall hypothesis was that the same farm management recommendations were made regardless of yield data quality level.

This study demonstrates the implications of using differing levels of yield monitor data quality for spatial data analysis of empirical field-scale on-farm comparisons. The field scale experiments were conducted by farmers in collaboration with the authors. The research questions, i.e. treatments, were pertinent to and chosen by the farmer. The choice of treatments and experimental design was ultimately the farmer's with guidance from the authors.

BACKGROUND

With the increased ease of site-specific data collection, an increased lack of ease emerged from data analysis. Problems with analyzing field-scale precision agriculture datasets include 1) lack of important spatially autocorrelated variables, 2) disparate spatial data layers, and 3) choosing an estimator. Excluding an important spatially autocorrelated variable from model specification is an omitted variable problem. One ramification of an omitted variable problem is spatial autocorrelation in the residuals under ordinary least squares (OLS) estimation leading to inefficient parameter estimation. Spatial statistics explicitly model spatial autocorrelation providing efficient treatment effect estimates. Disparate spatial data layers give rise to the modifiable areal unit problem (MAUP) potentially causing manipulated datasets to have differing moments than the true data (Anselin, 1988; Bockstael, 1996; Cressie, 1993; Gotway and Young, 2002; Openshaw and Taylor, 1979, 1981; Schabenberger and Gotway, 2005; Yule and Kendall, 1950). Disparate spatial data layers may be aligned to minimize adverse affects of MAUP. Appropriate estimators are chosen based on properties of the estimator and characteristics of the data. The OLS estimator is known to be inefficient with the spatially autocorrelated precision agriculture data. Maximum likelihood (ML) estimators are limited to moderate sample size datasets (<500 observations) due to computational difficulty in calculating eigenvalues and evaluating the Jacobian of the spatial weights matrix (Kelejian and Prucha, 1999; Bell and Bockstael, 2000). In addition, it is uncertain if not unlikely that a precision agriculture dataset will have normally distributed errors. Therefore, the general moments (GM) estimator is a likely candidate for spatial analysis of precision agriculture data.

At least three characteristics of farm-level precision agriculture data influence statistical inference, e.g., spatial effects such as spatial heterogeneity and spatial dependence, disparate spatial data layers, and inaccuracies in spatial data measurements. Spatial effects such as spatial heterogeneity and spatial autocorrelation have been addressed in the precision agriculture literature by using spatial econometric techniques (Griffin et al., 2006; Griffin et al., 2005a; Anselin et al., 2004; Lambert et al., 2003, 2004, 2006; Hurley et al., 2001, 2004 ; Liu et al., 2006). Farm-level precision agriculture datasets often have many yield observations and very few observations of continuous covariates, resulting in disparate spatial data layers. Aligning spatially disparate data layers is an issue for geography in general and precision agriculture in particular. With automated measurements and on-the-go sensors commonly used, both dependent and explanatory variables may potentially be measured incorrectly. Raw yield monitor data has many erroneously measured observations in terms of value and location, inducing errors in variables (Anselin, 1989). Erroneous yield monitor measurements can be adjusted with specialized software (Drummond, 2006; Griffin et al., 2005b). This paper focuses on disparate spatial data layers and spatial data measurement errors of the dependent variable.

Since the advent of precision agriculture, geostatistical procedures such as spatial interpolation have been used for visualization, input application, and data analysis. Empirical studies have used spatial interpolation during data analysis to 1) predict values at unmeasured areas to align spatially disparate data layers to a common sample size, 2) convert elevation into a surface in order to calculate slope and other elevation derivatives, and 3) smooth over errors in measurements such as erroneous yield monitor measurements. These motivations for interpolation have been addressed in this paper from an alternative spatial econometric perspective. The issues of interpolating elevation into a surface to calculate slope have been addressed by using spatial econometric cross-regressive variables (Lowenberg-DeBoer et al., 2007).

This study addressed the latter two points. Point 2 is addressed by describing how this study chose to manage disparate data layers and discusses the issues of the MAUP as it applies to precision agriculture. Point 3 is addressed by discussing the ramification of differing yield data quality levels and the availability of advanced software capable of filtering erroneous yield measurements and correcting location.

Disparate Spatial Data Layers

The modifiable areal unit problem (MAUP) has to do with the size, shape, and orientation of a polygon superimposed over an area. Aggregating data points within cells result in different means and variance than the original data inducing spatial effects (Schabenberger and Gotway, 2005). Differences in response function estimation from differing econometric estimations lead to potentially erroneous economic results for both aspatial and spatial models.

The modifiable areal unit problem arises from a host of issues in precision agriculture including yield monitor measurements and assimilating disparate spatial data layers. Using grids or other areal polygon units for interpolating or averaging measurements amplifies the problem since yield, soils and other biological factors are measured in a discrete manner over a continuous surface (Openshaw and Taylor, 1979). In any of these cases, a method of aggregating the data points within the resulting cell is the new value, albeit with differing properties than the original data in terms of mean, variance, and spatial autocorrelation. This phenomenon occurs at the relatively small areal unit from which yield monitor data is collected.

Aligning Disparate Spatial Data Layers

Site-specific agriculture data layers tend to be disparate, meaning that some point data layers have higher resolution or that a data layer is more dense. To assimilate the data layers into meaningful units for spatial analysis, some sort of spatial aggregation is required to align dependent and independent observations. Rather than predict missing values of a surface for a sparse data layer via interpolation it may be appropriate to aggregate a dense data layer to the same resolution as the least dense data layer (Griffin et al., 2005 a,b).

Measurement Errors in Yield Data

At least two sources of measurement error exist in yield monitor data. The first source is measurement error in the location attribute and pertains only to spatial analysis. The second source is the yield measurement itself, leading to larger error variances and is pertinent to both traditional aspatial and spatial analysis. Although both sources of measurement error result in errors in variables and spatial autocorrelation (Anselin, 1989), adjustments potentially correct these errors.

Raw yield monitor measurements (*raw*) are typically subjected to automated filtering processes (*default*) when imported into standard farm-level software to correct for location, adjust yields according to moisture, and remove erroneous observations. Default software parameter settings can be adjusted to reflect specific characteristics or be removed to revert back to the raw yield monitor data; however no diagnostics exist to determine if the settings are appropriate nor are the iterations user-friendly. A conscious removal of erroneous data and relocation to appropriate positions (*filtered*) are possible with specialized software and training (Drummond, 2006). It is unclear how the additional human capital and other costs of filtering data compares to the benefits.

Yield monitor data may be subjected to a yield data filtering protocol (Drummond, 2006; Griffin et al., 2005b; Ping and Doberman, 2005) to consciously filter raw yield monitor data. Under a certain set of known harvester characteristics, the yield monitor is unable to make accurate yield and location measurements. It is under these conditions yield observations are either adjusted for location or deleted from the data. ‘Flow delays’ result from mechanical lags within the harvester due to differences from the time crop is harvested to recording. Flow delays are corrected by repositioning yield measurements to the most appropriate location. Yield measurement values are not adjusted except for moisture corrections, although erroneously measured observations are deleted once detected. It is expected that consciously *filtered* yield monitor data is superior to *default* data which is superior to *raw* data, however only the differences in inference and farm management recommendations based on analysis using these levels of data quality can be measured.

METHODS

In order to perform spatial analysis on precision agriculture on-farm trial data, spatial data layers not only must be assembled into a single dataset but locations must coincide. The disparate spatial data layers were assembled into a single GIS data layer using techniques presented in Griffin et al. (2005b). The following sections describe procedures relevant to yield monitor data quality. Depending upon the condition of the data when acquired, differing steps may be required.

Yield Data Quality

To test if differences exist among levels of yield data quality with respect to inference and especially decision making, each data quality level is compared by conducting separate analyses. Yield quality levels include 1) raw harvester measurements (*raw*), 2) accepting default settings from popular standard farm-level software packages (*default*), and 3) a conscious locational adjustment and removal of erroneously measured yield measurements (*filtered*).

In many cases the analyst acquires yield monitor data after it has been subjected to the default yield filtering procedures of the farm-level software, thus only *default* and *filtered* are possible. In this case, the *filtered* data originating from *default* data may differ from *filtered* data derived from *raw* data. Assuming that individual data layers are acquired in the raw form, i.e. yield monitor data as exported from harvester without conversion or filtering (*.yld, *.gsy, *.gsd), this paper proceeds by describing steps necessary to appropriately assimilate raw data into a final dataset suitable for analysis. To perform comparisons of yield data quality, the raw yield data is imported and exported with the farm-level software twice, once to obtain raw yield data (*raw*) and once to obtain yield data subjected to the software's default parameter settings (*default*). The *raw* data was used as the base data quality as well as the beginning data in the yield filtering procedure.

Once filtering parameters of the farm-level software (e.g. AgLeader SMS, JDOffice, MapShots, FarmWorks, etc.) were removed or set to zero, yield monitor data (*.yld, *.gsd, *.gsy, or other yield monitor file formats) were imported into one of the farm-level software packages. The default filtering parameters usually include flow delay, start pass delay, end pass delay, minimum yield, and maximum yield (Table 1). In the event that yield data is acquired after the default filtering procedure has been applied, it is important to note the default parameters used. The yield data file was exported as an AgLeader Advanced export format (Drummond, 2006; Griffin et al., 2005b) which can be performed with several farm-level software packages.

Table 1.Typical default yield data processing settings

	Corn	Soybean
Flow Delay (s)	12	12
Start Pass Delay (s)	4	4
End Pass Delay (s)	4	4
Minimum Yield (bu ac-1)	10	10
Maximum Yield (bu ac-1)	400	200

The *raw* yield monitor data, or alternatively *default* yield data if this is the data originally acquired by the analyst, were imported into Yield Editor (Drummond, 2006). Eleven filtering parameters can be set with a combination of trial and error, *a priori* information, and analyst intuition. In addition to flow delay, end pass delay, start pass delay, minimum yield, and maximum yield available in many of the farm-level field mapping software packages, Yield Editor supplements these parameters with maximum velocity, minimum velocity, smooth velocity, minimum swath, and standard deviation filters (see Table 2 for examples of analyst chosen parameters). When appropriately set, these filters remove erroneously measured yield observations and adjust measurements to the appropriate location. As with other data analysis procedures, a learning curve is involved in gaining human capital sufficient to appropriately filter yield data. Once all levels of yield data quality were complete, they were imported into a GIS to combine with explanatory variables for each dataset.

Table 2. Parameter settings for yield monitor data filtering procedure

Experiment	POPSEED		FUNG		SOYSEED		SEEDTRT	
Crop	popcorn		soybean		soybean		popcorn	
Yield data form	*default*		*default*		*raw*		*default*	
Experimental design	strip trial		split field		strip trial		pseudo-replicated strip trials	
Parameter	**Value**	**Deleted obs**	**Value**	**Deleted obs**	**Value**	**Deleted obs**	**Value**	**Deleted obs**
Maximum velocity mph	3.75	44	3.5	380	5.25	8	7	0
Minimum velocity Mph	2.5	2946	1.5	1042	4	1802	2	564
Smooth velocity %	0.05	874	0.1	911	0.2	408	0.15	429
Maximum yield bu ac^{-1}	130	1221	85	406	80	49	250	35
Minimum yield bu ac^{-1}	30	160	0	0	0	0	10	0
Standard deviation	3	1257	4	853	4	1524	3	393
Flow delay s	-1	92	-3	511	3	561	NA	0
Start pass delay s	4	357	4	679	4	745	3	214
End pass delay s	5	357	4	172	0	0	2	214

[a] For *default* beginning data; flow delay, start and end pass delays were conducted during importing raw data into farm level software by the farmer

[b] The number of deleted observations is not the sum of the individual parameters

Each dataset was subjected to the same general yield data filtering protocol; however individual studies received customized parameter settings such that appropriate values were assigned. For all but one study in the selected sample (Table 2), the yield data was acquired in the *default* format impacting the overall control of filtering. In these studies (POPSEED, FUNG, and SEEDTRT), the flow delay, start pass delay, and end pass delay parameters were set to 12, 4, and 4 seconds, respectively, in the farm level software. Observation position was still able to be appropriately relocated, although some observations were already deleted. In addition, care must be taken to compare the final parameter setting across studies. The 12 second default setting for flow delay may be an adequate average, but this value varies across farmers, equipment, and within fields. The yield data layers were appended with explanatory variables such as elevation, soils, and treatment variables as described in Griffin et al. (2005b).

Statistical Estimation and Estimators

Estimators are chosen based on properties of that estimator with respect to econometric model, sample size, distribution, and desired inference properties. Commonly used estimators include OLS, maximum likelihood (ML) and generalized moments (GM). OLS is often used with spatial data to facilitate spatial diagnostics on the residuals. In addition, OLS is used for the cross-regressive model. Spatial autoregressive error (SAR) models can be estimated with ML and GM, among others.

Spatial process models estimated with ML uses an eigenvalue computation from the Jacobian matrix which loses numerical precision beyond 1,000 observations (Anselin, 1988; Kelejian and Prucha, 1999; Bell and Bockstael, 2000). Kelejian and Prucha (1999) go on to say that the eigenvalues for spatial weight matrices under ML estimation were not correctly estimated with more than 400 observations. Due to the computational limitations of ML for large sample-size datasets, GM estimators are considered because estimation can be conducted for very large datasets of several thousand observations.

Although GM is not as efficient in general as ML, this limitation may be overcome with large sample sizes although Bell and Bockstael (2000) state that “the GM estimator makes no claim to being asymptotically efficient” (p. 79). In addition, ML requires a known distribution while GM estimation does not. For these reasons, Kelejian and Prucha (1999, 2004, 2006) suggest a GM estimator for the autoregressive parameter in spatial models. GM models are particularly useful in the large samples of unknown distribution of precision agriculture data. However, the specification of the spatial weights matrix may influence parameter estimation more than the choice of estimator (Bell and Bockstael, 2000).

Specifying the Spatial Weights Matrix

Although the inverse distance spatial weights matrix was chosen *a priori* for precision agriculture datasets, the distance band was empirically determined. Spatial weights matrices influence parameter estimation for spatial models, especially when the spatial autoregressive parameter is large. As opposed to several other precision agriculture studies, we use a spatial weights matrix incorporating a distance decay function over a considerable distance. In our experience with field scale precision agriculture data, spatial weights matrices constructed with short proximity, e.g. queen order one and minimum Euclidean distance, produce spatial diagnostics indicating the spatial lag rather than the spatial error model and result in erratic parameter estimates.

In order to determine the appropriate specification of the weights matrix, a series of row-standardized inverse distance weights matrices were created, each with a differing cutoff distance or distance band. The shortest cutoff distance for each weights matrix was below a reasonable level such as 25 meters and the longest distance is a distance such that no spatial correlation is expected to exist in the data, typically 85 meters. Cutoff distances were created on 10-meter intervals such that 25, 35, 45, 55, 65, 75, and 85 meters result.

OLS estimation was conducted on the full econometric model and spatial diagnostics performed on the residuals with each of the spatial weights matrices. Spatial correlograms of LM tests were created to gather information with respect to the appropriate cutoff distance (Cressie, 1993) and spatial process model (Anselin, 1988). The highest chi-squared coefficients of the LM tests indicated the proper distance band for the spatial weights matrix and which model was most appropriate, e.g. OLS or spatial process model. The most appropriate cutoff distance was used to create the spatial weights matrix for spatial analysis.

Spatial process models explicitly model spatial autocorrelation. For precision agriculture datasets, spatial autoregressive error models were expected rather than other spatial models. Spatial diagnostics performed on OLS residuals support this assumption. The spatial error model is given as $y = X\beta + \varepsilon,\ \varepsilon = \lambda W\varepsilon + \mu$ or in reduced form as $y = X\beta + (I - \lambda W)^{-1}\mu$ where y is a n x 1 vector of dependent variables, X a n x k matrix of explanatory variables, β a k x 1 vector of regression coefficients, ε an n x 1 vector of residuals, λ a spatial autoregressive parameter, **W** is an n x n spatial weights matrix, and μ a well behaved, non-heteroskedastic uncorrelated error term (Anselin, 1988). The $(I - \lambda W)^{-1}$ term is the so-called spatial multiplier. When the spatial autoregressive term, λ, is 0, the spatial error model reverts to the familiar aspatial model, $y = X\beta + \mu$. The spatial error process can be characterized by the autoregressive error process resulting in global spillovers due to the spatial multiplier. The spatial error model has no substantive economic interpretation, i.e. estimated parameter coefficients are asymptotically unaffected. When the spatial error model is appropriate, OLS estimates remain unbiased but are inefficient.

RESULTS

Foliar Soybean Fungicide Treatments

Two foliar applied soybean fungicide treatments were applied in Tazewell County, Illinois. Fungicide treatments were applied on July 20, 2005 in strips perpendicular to planter passes. Elevation data was acquired from the automated-guidance GPS on the planter tractor. The inverse distance spatial weights matrix was assigned a distance cutoff of 35 meters.

Table 3. Fungicide agronomic and economic differences

	Agronomic yield difference (kg ha^{-1})		Economic difference ($ ha^{-1})	
	OLS	SAR-GM	OLS	SAR-GM
filtered	-262	-81	-67	-28
default	2073	-222	447	-59

For the fungicide experiment, economic and agronomic decisions were identical, and that fungicide treatment 2 (FUNG2) was superior to fungicide treatment 1 (FUNG) (Table 3). Using default yield data with OLS estimation, the farm manager would have chosen FUNG1 rather than FUNG2. Negative values indicated FUNG2 was superior to FUNG1 and positive values suggest the reverse.

Economic partial budget analysis was conducted by using a soybean price of $0.221 per kg, FUNG1 price of $39.85 per ha, and FUNG2 price of $29.94 per hectare. Assuming GM estimation of the full econometric model best modeled reality, a difference between treatment of $28 per hectare results.

Unlike rate trials where a continuum of decisions is possible, the farm manager must select a single treatment with categorical choices such as fungicide treatments. In this example, the least difference in economic returns were estimated to be $28 per hectare using the SAR model estimated with GM. It is suspected that this difference is closest to true difference since the SAR-GM estimation using *filtered* was expected to be the best model and data available. The largest economic difference when the same decision would have been made was $67 per hectare with OLS estimation of a model with *filtered* data. When OLS was used with models using *default* data, the opposite decision would have been made at an even greater economic difference of $447 per hectare. When using *default* data, the optimal decision differed by estimators but not when *filtered* data was used. Nevertheless, the GM estimator indicated the same optimal treatment regardless of yield data quality level. However, the input recommendation derived from aspatial OLS estimation differed between yield data quality levels. This indicates that the choice of estimator impacts the optimal decision the farm manager would have made. In addition, the alternative decision based only on data quality level within the same estimator indicated that data quality is important with categorical experiments especially with aspatial models.

In each case, the AIC is lower for *filtered* than for *default* data and lower for SAR-GM than for OLS estimation (Table 4). Although this example does not answer the question of which data quality or estimator leads to the best economic decisions, it does show that farm management decisions may be affected by data quality and by choice of estimators.

Table 4. Fungicide AIC statistics

	OLS	SAR-GM	# obs
filtered	8,137	7,904	1,124
default	9,962	9,766	1,272

Table 5. POPSEED seeding rate trial results

Estimated agronomic optimal seeding rates 000's ha^{-1}	OLS	SAR-GM
default	58.3	65.0
filtered	68.2	68.7
Predicted yield at each estimated seeding rate Mg ha^{-1}*		
default	4.7	5.7
filtered	5.8	5.8
Estimated profit maximizing seeding rate[a,b] \$ ha^{-1}		
default	52.9	62.8
filtered	67.7	68.2
Economic returns from GM-SAR \$ ha^{-1} *		
default	249	457
filtered	485	486

* Predicted yields from each estimator/data quality combination with the SAR-GM coefficients
[a]Corn price = \$2.50 bu-1; seed cost = \$0.30 per 1,000
[b] Profit maximizing level estimated for each estimator and yield data quality level

Popcorn Seeding Rate

Four popcorn seeding rates were applied in a strip trial design in Tazewell County, Illinois. Popcorn was produced on contract under center pivot irrigation systems. Seeding rates included two blocks of 65,500 and single blocks of 59,800, 68,500, and 73,900 seeds per hectare. Elevation was acquired from the automated guidance GPS signal from the planter tractor.

Aspatial OLS and SAR estimated with GM showed almost the same yield maximizing and profit maximizing popcorn seeding rates with *filtered* data (Table 5). All were within 700 seeds of 68,000 seeds per hectare. With *default* data a much larger difference occurred. The difference in yield maximizing seeding rates by estimator was almost 7,000 seeds per hectare and for profit maximizing seeding rates almost 10,000 seeds per hectare.

Assuming that the popcorn planter can be set to be within 2,500 seeds per hectare, no difference in decision would be made between the OLS and SAR-GM estimators with *filtered* data. However, when *default* data was used, differing decisions, i.e. different planter settings, were made between estimators. Within the same estimator, different decisions would have been made between the two levels of yield monitor data quality. Thus there were differences in farm management decisions depending upon yield data quality.

Table 6. Popcorn seeding rate AIC statistics

	OLS	SAR-GM	# obs
filtered	26,593	25,687	3,446
default	31,540	31,711	3,906

AIC goodness-of-fit diagnostics indicate that the SAR-GM model was superior to the OLS model when *filtered* data was used (Table 6). When *default* data were used, the OLS estimation dominated SAR-GM although spatial diagnostics on OLS residuals indicated the spatial error model was appropriate rather than the aspatial model. Although AIC for the *filtered* dataset was lower than the *default* dataset, direct comparisons can not be made between datasets on the basis of AIC. These examples have not answered the question of which estimator or data quality is best, but differences between them have been demonstrated.

Soybean Seeding Rate Experiment

Five soybean seeding rates were replicated four times in an 19-hectare strip-trial design with two harvester passes wide per treatment strip in Montgomery County, Indiana. Seeding rates were selected by a very low rate (197,600) to a relatively high rate (395,200) in increments of 49,400 seeds per hectare. RTK-GPS was used for the position signal for the automated-guidance on the planter tractor as well on the combine harvester since it was already available.

Due to the shape of the yield response curve to soybean seeding rates, small differences in agronomic yield resulted when seeding rates were estimated from differing estimators. Although reasonable physical agronomic optimal rates were estimated with OLS, the optimal economic decision was to produce soybeans at the minimum seeding rate in the trial (197,600 seeds per hectare) due to constrained optimization and no interior solution. For this reason, the economic returns for any level of yield monitor data analyzed with OLS was $321 per hectare. Although expected returns were very similar with *default* or *filtered* yield data under SAR-GM estimation, there was a $104 per hectare difference between *filtered* and *raw*.

Using the same assumption with the popcorn seeding rate dataset (POPSEED) that the planter can be set to plant within 2,500 seeds per hectare, differing decisions would have been made between using *raw* and the other two levels of yield data quality. Differences in the agronomic and economic optimal soybean seeding rates for SAR-GM with *filtered* data were 2,300 and 2,600 respectively (Table 7). The difference in decision between *default* and *filtered* data is less clear and recommendations may be outside the capabilities of the planter. In each level of yield monitor data quality, the SAR-GM estimator dominated the OLS estimator (Table 8). Again, this example has not shown which estimator or yield monitor data quality is best, but a demonstration of the differences between estimators and yield data quality has been presented.

Table 7. SOYSEED seeding rates and returns for yield data quality levels

	Agronomic rate (000's) ha^{-1}		Economic rate (000's) ha^{-1}		Economic returns \$ ha^{-1}	
	OLS	SAR-GM	OLS	SAR-GM	OLS	SAR-GM
filtered	280.3	307.3	197.6	265.7	321	497
default	374.8	305.0	197.6	253.1	321	490
raw	307.8	300.3	197.6	160.7	321	393

Table 8. Soybean seeding rate AIC statistics

	OLS	SAR-GM
filtered	23,954	21,461
default	31,666	29,892
raw	41,328	41,008

CONCLUSIONS

The null hypothesis that the same farm management recommendation would be made with any level of yield data quality was not supported. It has been demonstrated that more variation occurs in recommended farm management decisions from differences in yield data quality than from choice of estimator.

It logically follows that filtered data is most likely superior to default and raw data, however further work is needed to determine the validly of this idea. A simulation experiment that generates yield monitor data along with the associated measurement errors in value and location would be useful in determining if conscious filtering of raw data leads to superior statistical inference and farm management decision making. Another long term test of differing levels of yield monitor data quality would be to continue work with farmers conducting on-farm trials. If the farmer would apply the alternative treatments as indicated as more appropriate from differing levels of yield data quality to at least a small portion of the farm, the cost of the wrong or less appropriate decision could be measured to determine if that decision was less appropriate. The human capital costs of filtering data must also be estimated in order to address the cost benefit analysis.

Many farm-level software providers offer farmer and consultant users the ability to perform advanced yield monitor data filtering above and beyond the default settings of most software packages, potentially without having to import and/or export files between software packages. For future implementation of yield monitor data spatial analysis techniques in farm-level software packages, GM estimation routines are becoming more available.

Elevation data from combine harvesters are typically not conducive to explain variation or noise in data from on-farm experiments, thus elevation data from automated guidance is a more likely candidate. This is one reason for synergy between GPS automated-guidance and yield monitor technology, to facilitate on-farm experimentation. This leads to another reason why ‘raw’ or ‘default’ data probably will not be used. Some sort of data assimilation is necessary to spatially join the elevation and yield observations. With this level of data management and software familiarity, it is expected that care would be taken by the farmer or analyst to properly filter erroneous yield data measurements and properly set the flow delays. It is also anticipated that analysts who use advanced econometric estimators will use high quality data. This assumes that if the analyst goes as far as advanced estimator and spatial econometric models, then filtering yield monitor data would be consistent with their behavior and desired outcome.

ACKNOWLEDGEMENTS

Routines for the GM estimator within the R (R Development Core Team, 2006) statistical language and environment were provided by Luc Anselin and Julia Koschinsky, University of Illinois. Special appreciate is given to USDA-SARE for providing funding to facilitate the research of Griffin's PhD dissertation.

REFERENCES

Anselin. L. 1992. SpaceStat Tutorial: a Workbook for Using SpaceStat in the Analysis of Spatial Data. Available on-line at: http://www.terraseer.com/products/spacestat/docs/spacestat_tutorial.pdf

Anselin, L. 2001. Spatial Effects in Econometric Practice in Environmental and Resource Economics. American Journal of Agricultural Economics 83(3) August 2001: 705-710.

Bell, K.P. and Bockstael, N.E. 2000. Applying the Generalized-Moment Estimation Approach to Spatial Problems Involving Micro level Data. The Review of Economics and Statistics, February 2000, 82 (1): 72-82.

Cressie, Noel A.C. 1993. Statistics for Spatial Data. John Wiley & Sons: New York.

Drummond, S. 2006. Yield Editor 1.02 Beta Version. USDA-ARS, Cropping Systems and Water Quality Unit, Columbia, MO. Available on-line at: http://www.ars.usda.gov/services/software/download.htm?softwareid=20

Gotway, C.A. and Young, L.J. 2002. Combining Incompatible Spatial Data. Journal of the American Statistical Association. June 2002, Vol. 97, No. 458.

Griffin, T.W., Fitzgerald, G, Lambert, D.M, Lowenberg-DeBoer, J., Barnes, E.M., and Roth, R. 2005a. “Testing Appropriate On-Farm Trial Designs and Statistical Methods for Cotton Precision Farming,” *Proceedings of the Beltwide Cotton Conference*, January 4 – 7, 2005, New Orleans, LA. Available at: http://www.cotton.org/beltwide

Griffin, T.W., Brown, J.P., and Lowenberg-DeBoer, J. 2005b. Yield Monitor Data Analysis: Data Acquisition, Management, and Analysis Protocol. Available on-line at: http://www.purdue.edu/ssmc

Hurley, T., Kilian, B. and Dikici, H. 2001. “The Value of Information for Variable Rate Nitrogen Applications: A Comparison of Soil Test, Topographical, and Remote Sensing Information.” Selected Paper, *AAEA Annual Meeting, Chicago, IL, August 5-8, 2001*. Available on line at: http://agecon.lib.umn.edu

Hurley, T.M., Malzer, G. L., and Kilian, B. 2004. Estimating Site-Specific Nitrogen Crop Response Functions: A Conceptual Framework and Geostatistical Model. Agronomy Journal 96: 1331-1343.

Kelejian, H.H. and Prucha, I. 2006 "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," Department of Economics, University of Maryland. Working Paper. Available on-line at: http://www.econ.umd.edu/~prucha/Papers/WP_Prucha_3_2006.pdf

Kelejian, H. and Prucha, I. 1998. "A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," Journal of Real Estate Finance and Economics. 17(1): 99-121.

Kelejian, H.H. and Prucha, I.R. 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. International Economic Review, 40, 509-533. Available on-line at: http://www.econ.umd.edu/~prucha/Papers/IER40(1999).pdf

Lambert, D.M., Lowenberg-DeBoer, J. and Bongiovanni, R. 2004. A Comparison of Four Spatial Regression Models for Yield Monitor Data: A Case Study from Argentina. Precision Agriculture 5 pp. 579-600.

Lambert, D.M. 2005. Spatial Regression Models for the Economic Analysis of On-farm Site-Specific Management Trials. PhD Dissertation, Purdue University, West Lafayette, IN.

Lambert, D.M., Lowenberg-DeBoer, J., and Malzer, G. L. 2006. Economic Analysis of Spatial-Temporal Patterns in Corn and Soybean Response to Nitrogen and Phosphorus. Agronomy Journal 98:43–54.

Liu, Y. Swinton, S.M. and Miller, L.R. Forthcoming. Is Site-Specific Yield Response Consistent over Time? Does It Pay? American Journal of Agricultural Economics.

Lowenberg-DeBoer, J., Griffin, T.W., and Florax, R.J.G.M. 2007. Local Spatial Autocorrelation in Precision Agriculture Settings Accounting for Micro-Scale Topography Differences. Forthcoming in Proceedings of the 8th International Conference on Precision Agriculture and Other Resource Management.

Openshaw, S. and Taylor, P. 1979. A Million or so Correlation Coefficients. In Statistical Methods in the Spatial Sciences, ed. N. Wrigley, London, Pion, pp. 127-144.

Openshaw, S. and Taylor, P. 1981. The Modifiable Areal Unit Problem, in Quantitative Geography; A British View, eds N. Wrigley and R. Bennett, London, Routledge and Kegan Paul, pp. 60-69.

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Yule, G.U. and Kendall, M.G. 1950. An Introduction to the Theory of Statistics (Charles Griffin, London).