DETERMINING YIELD MONITORING SYSTEM DELAY TIME WITH GEOSTATISTICAL AND DATA SEGMENTATION APPROACHES

S. O. Chung, K. A. Sudduth, S. T. Drummond

ABSTRACT. *In combine harvesting, knowledge of the delay time from cutting the crop to sensing the grain flow is required for accurate spatial location of grain yield data. Currently, either an assumed, fixed delay time is used or the delay time is determined by visual inspection of yield maps. Geostatistical and data segmentation methods were developed to estimate yield monitoring system delay time using objective criteria. The methods were validated with an ideal dataset and with elevation and soil electrical conductivity datasets having known delay times. When applied to yield and moisture content measurements collected with a commercial yield monitoring system, the methods successfully estimated delay time. In most cases, the results agreed (±1 s) with results achieved using a visual method. Grain yield and grain moisture content exhibited different delay times at different locations within test fields. Thus, it may be appropriate to apply delay time corrections to homogeneous sub–field areas, instead of on a whole–field basis. Use of these new estimation methods could allow for more accurate and efficient processing of yield monitor data.*

Keywords. Precision agriculture, Yield monitor, Yield map, Delay time, Geostatistics, Data segmentation.

Yield measurement and mapping have been key in the development of precision agriculture. On–the–go sensor–based acquisition of crop yield and position data resulting in yield map creation was accomplished in the late 1980s by Searcy et al. (1989). Since that time, yield monitoring systems have been improved significantly, with some commercial systems providing average load accuracies of approximately ±1% (Murphy et al., 1995). The creation of yield maps is critical to precision agriculture because yield maps provide data important for the evaluation phase of the precision agriculture cycle. For example, a yield map can provide local information on nutrient absorption, soil variability, and effects of treatment strategies (Reitz and Kutzbach, 1996). Furthermore, crop yield may be used as a basis for many agricultural input recommendations.

Previous research on yield monitors and mapping algorithms has been well reported (Colvin and Arslan, 2000; Pierce et al., 1997; Birrell et al., 1996; Stafford et al., 1996;

Missotten et al., 1996). Yield monitoring and map creation is conceptually easy to understand, but obtaining an accurate and reliable yield map is challenging due to six major factors (Blackmore and Marshall, 1996):
- Time lag (transportation delay time) of grain through the threshing mechanism.
- Unknown crop width entering the header during harvest.
- The inherent "wandering" error from the GPS.
- Surging grain through the combine transport system.
- Grain losses from the combine.
- Sensor accuracy and calibration.

Numerous researchers have proposed methods to address these problems and improve the accuracy of yield mapping. Post–harvest data filtering and correction techniques have commonly been used (O'Neal et al., 2000; Blackmore and Moore, 1999; Beck et al., 2001; Rands, 1995). Nolan et al. (1996) indicated that the coefficient of variation in yield data was reduced from 32% to 10% by the correction of errors from GPS wandering, overlapped harvest area, and delay time. Searcy et al. (1989) estimated delay time and crop redistribution by an averaging technique and smoothed yield and position data. Measurement of actual swath width (Sudduth et al., 1998) and correction of overlapped harvest areas (Han et al., 1997; Drummond et al., 1999) have been implemented to improve point yield estimations.

Blackmore and Marshall (1996) emphasized the importance of ascertaining the delay time between cutting and sensing, because if the delay time is not assessed correctly, then all yield positions will be offset by an incorrect amount. They further noted that delay time might not be constant across a field. Lamb et al. (1995) also regarded delay time as a major source of yield map error. Moore (1998) reported that incorrect delay time could significantly over– or under–estimate crop yield, and would cause a position offset of yield measurements within a field.

The authors are **S**un–Ok Chung, **ASAE Member**, Graduate Research Assistant, Biological Engineering Department, University of Missouri, Columbia, Missouri; and **Kenneth A. Sudduth**, **ASAE Member Engineer**, Agricultural Engineer, and **Scott T. Drummond**, Computer Specialist; USDA–ARS Cropping Systems and Water Quality Research Unit, Columbia, Missouri. **Corresponding author**: Kenneth A. Sudduth, 269 Agricultural Engineering. Bldg., University of Missouri, Columbia, MO 65211; phone: 573–882–4090; fax: 573–882–1115; e–mail sudduthk@ missouri.edu.

In general, previous researchers have chosen to use constant delay times when creating yield maps of a field. Several methods have been used to determine these delay times. Whelan and McBratney (2002) attempted to better understand the dynamics of grain flow through the combine by painting individual rows of grain sorghum different colors, and investigating color variations in the combine grain flow stream over time. Birrell et al. (1996) and Nolan et al. (1996) estimated transportation delay time by comparing the time at which the combine head entered or exited the crop to the time at which the measured grain flow began or stopped. Chosa et al. (2000) used the elapsed time between an operator switch closure and the initial mass flow signal from a grain flow sensor to estimate delay time.

Stott et al. (1993) used visual inspection of the time–varying yields obtained when harvesting in alternating directions to estimate delay time. They determined delay time by matching grain yield for two different directions across a known zero–yielding portion of the field. This method does not need additional hardware or sensors and does not require fieldwork to measure the delay time. However, visual inspection of each yield map after harvest is time–consuming, and the criteria used to estimate delay time are subjective. Nevertheless, we have generally relied on this method to determine delay time in our research data, because we believe it gives the most reliable estimate.

It is clear that delay time cannot be universal, as it can vary with make and model of harvest equipment, speed, ground slope, load, and whether the harvester is moving into or out of the crop (Nolan et al., 1996). In fact, it is clear that delay time may vary significantly within a field or even within a single pass of the harvester. Measuring and calibrating delay time for every operating time/condition would be tedious and time–consuming, if not impossible. Application of the same value to different sensors on a single combine (i.e., grain flow and moisture) is not reasonable either, since the sensors are often mounted at different positions in the grain flow stream. Yield and grain moisture content may exhibit different delay times, and these delay times may vary across locations within a field. If this is the case, then it may be appropriate to apply variable delay time corrections to yield and grain moisture on homogeneous sub–field areas. Clearly, a systematic and objective means of determining delay time that could be applied to multiple sensors and sub–field areas would be extremely useful.

The objectives of this study were to: (1) develop systematic and objective methods to determine delay time for combine yield and moisture sensor data, and (2) apply these methods to whole field and sub–field areas to investigate their performance.

ANALYSIS APPROACH

The visual, map–based method of estimating delay time can be described as follows. Assuming that the direction of travel is alternating, as shown in figure 1, then the delay time that causes the greatest visual similarity between adjacent transects over several instances is selected, and homogenous areas become clear (14 s in this case). If an inappropriate delay time is used, then the edges of those areas will become blurred, with something of a "saw tooth" pattern (i.e., 12 s or 16 s). In other words, the visual method assumes that the features in a yield map should appear in natural, continuous, spatial patterns.

In order to successfully employ the visual method of delay time estimation on a dataset, several requirements exist. It is critical that the dataset have significant, widespread spatial variation in the measured variable. Secondly, variation must be on a spatial scale compatible with the sampling intensity of the dataset. Finally, it is important that the measurement error be relatively small in relation to the variation of the measured variable. When all three of these requirements are met, the datasets have significant "clarity," and we can be quite confident in the selection of delay time. As these conditions are less well met, the datasets are more ambiguous, and the estimation of delay time becomes more problematic.

GEOSTATISTICS

Geostatistics provides a theoretical basis for the use of this visual method. In essence, the spatial dependency of near locations is greater than that of far locations (Webster and Oliver, 1990). Geostatistics, based on the theory of regionalized variables, deals with the variance structure of spatial variables and is the primary tool of spatial variability analysis. Bakhsh et al. (2000), Borgelt et al. (1997), and Chung et al. (2000), among others, have described applications of geostatistics to precision agriculture. Semivariance, the index of spatial dependency, is expressed in equation 1:
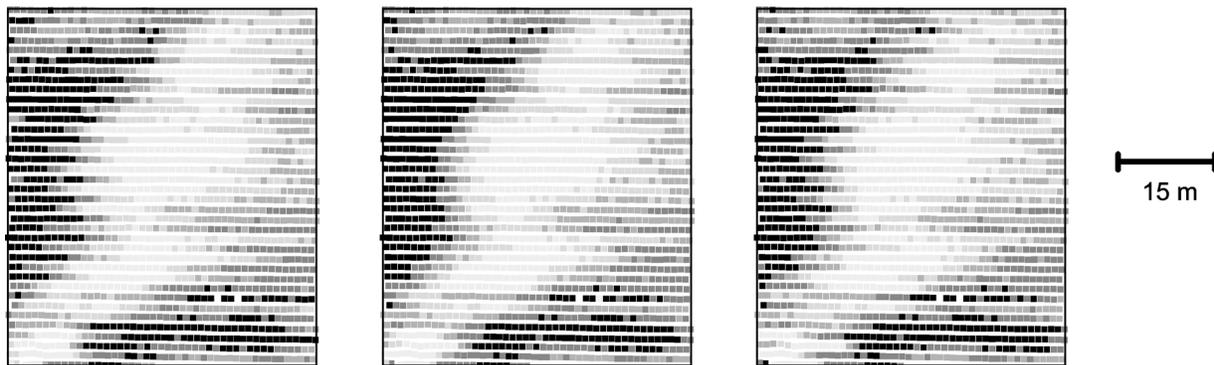

Figure 1. Effect of different delay times on homogeneity of mapped data: 12 s (left), 14 s (center), and 16 s (right).
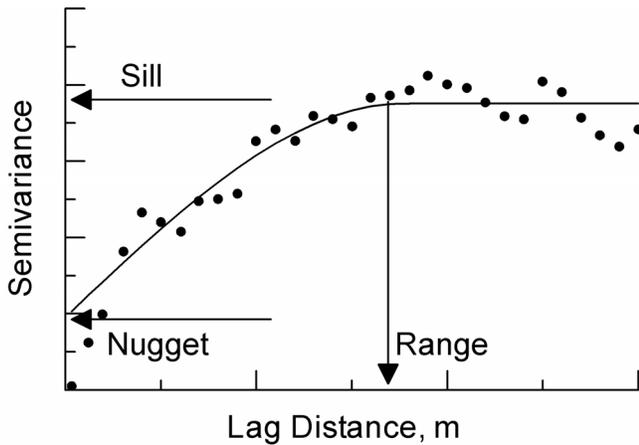
Figure 2. Parameters of a typical semivariogram.

Figure 3. Expected pattern of nugget variance, with minimum value at the true delay time.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{z_i - z_{i+h}\}^2 \qquad (1)$$

where

h = separation distance or lag
$\gamma(h)$ = semivariance for interval distance class h
z_i = measured sample value at point i
z_{i+h} = measured sample value at point $i+h$
$N(h)$ = total number of pairs for the lag interval h.

Figure 2 shows a typical semivariogram and its parameters. As the separation, or lag, distance increases, semivariance increases and then reaches a maximum at the level known as the sill. Range, the limit of spatial dependence, is defined as the separation distance at which the variogram reaches its sill (Webster and Oliver, 1990).

Another characteristic of the variogram is the nugget, which appears on the variogram as a discontinuity at the origin. A nugget variance is due to measurement errors and/or micro–variability over distances smaller than the minimum distance between observations. In practice, the nugget and other semivariogram parameters are calculated by regression of semivariances over a certain lag distance with an appropriate model.

Figure 1 shows that localized areas of a yield map are less variable if the correct delay time is used. The "saw tooth" pattern imparted by incorrect delay times increases the variability between transects, assuming that most actual yield variations are relatively smooth and well–behaved. Since nugget variance measures the amount of small–scale variation in a spatial dataset, we would expect the relationship of nugget and delay time to be similar to figure 3. Nugget variance should be minimized at the true delay time.

Standard semivariograms are fit over a wide range of lag distances and are theoretically restricted to a few functional shapes. Because of this, they may not fit the semivariance pattern at small lag distances particularly well, as shown in the example of figure 2. A linear relationship is most frequently seen in semivariance patterns near the origin (Journel and Huijbregts, 1978). Thus, in this research, we decided to use a linear regression through the first few points of the semivariogram to determine a more reliable nugget–like parameter, rather than using the nugget calculated from the overall theoretical semivariogram.
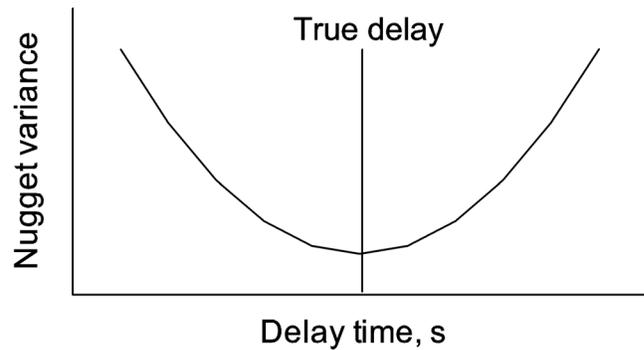
DATA SEGMENTATION

Another approach we used to determine optimum delay time was a data segmentation procedure based on a quadtree decomposition. The quadtree approach was originally used as a two–dimensional representation of binary images (Hanan, 1990) and has been used in many areas related to computer graphics (Vassilakopoulos and Manolopoulos, 1993) and database query. The basic principle of a quadtree is to cover a planar region of interest with a square, and then to recursively partition the square into smaller squares until each square contains a suitably uniform subset of the input. This approach can be used, for instance, to compress bitmap images by subdividing them until each square has the same color value (Leitao et al., 1996).

The quadtree decomposition procedure was explained well by Hanan (1990). As an example, consider the spatial region in figure 4 (left) and the corresponding 2^3 by 2^3 binary array in figure 4 (center). Observe that the ones correspond to pixels within region A, and the zeroes correspond to pixels outside region A. The quadtree decomposition of this data is shown in figure 4 (right). The quadtree decomposition has the property that at each subdivision stage, the image is subdivided into four equal–sized parts. Note that for this dataset, attribute data will be stored for 19 quadtree blocks, as compared to the 64 blocks in the original classified image.

Similar to the geostatistical approach, if a yield dataset is corrected with the true delay time, then the discrepancies between adjacent homogeneous areas are clearer and simpler than in case of an incorrect delay time. Distorted and ambiguous features in a yield map will increase the number of decompositions, or the data segmentation number. The performance of the data segmentation approach will be influenced by crispness of the input dataset, the number of classes used, and the classification scheme selected.

DATA AND ANALYSIS PROCEDURES

DATASETS USED

The feasibility and performance of the geostatistical and data segmentation methods were investigated using a variety of test datasets for which delay times were well established. The methods were then applied to grain yield and moisture content measurements from three research fields harvested with a combine equipped with a commercial yield monitoring system.
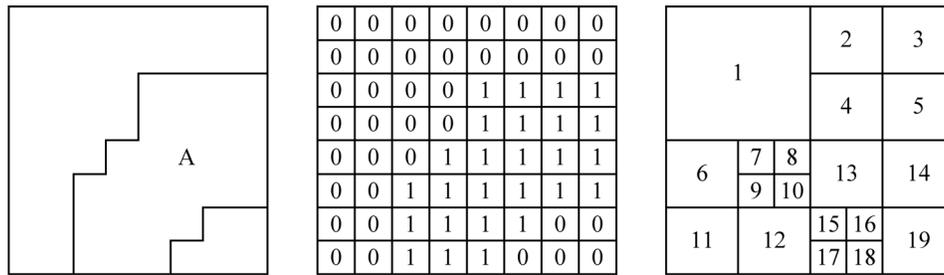
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	1	1	1	1	1
0	0	1	1	1	1	1	1
0	0	1	1	1	1	0	0
0	0	1	1	1	0	0	0

Figure 4. Binary data segmentation using a region quadtree: original image (left), classified image (center), and segmented image (right).
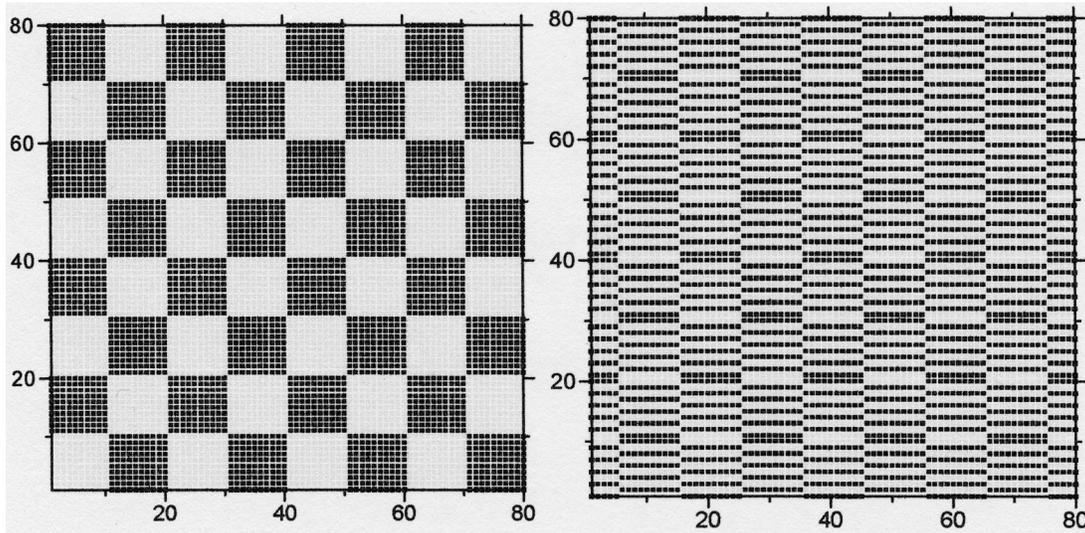

Figure 5. Ideal test dataset with clear regions at zero (left) and 5 s (right) delay times.

The first "ideal" test dataset simulated a block of spatial data with homogeneous sub–field regions clearly separated by sharp boundaries. The left part of figure 5 shows the pattern of this generated dataset. Two types of homogeneous regions were placed next to each other in a checkerboard pattern. Each homogeneous region consisted of 10 rows × 10 columns of grid cells. Data were assumed to have been "harvested" line by line from the bottom left corner, with adjacent rows collected in opposite directions. To simulate the effects of changing delay time, cells were assumed to have been collected on 1s intervals, and 10 additional datasets were created by shifting the values of the cells through the range of ±5 s. The ideal dataset was considered to be a subset of a larger field consisting of the same repeating pattern. As a result, each transect was shifted left or right independently of the rows above or below it (fig. 5, right).

The performance of the developed methods was then examined with elevation and soil electrical conductivity (EC) datasets. The elevation and EC data were collected simultaneously on a 4 ha field using an Ashtech Z–Surveyor real–time kinematic differential global positioning system (RTK–DGPS; 3 cm horizontal and 5 cm vertical rms accuracy) and a Geonics EM38 soil electrical conductivity sensor, as described by Sudduth et al. (2001). These data were collected in the back–and–forth pattern previously de-scribed, with adjacent rows collected in opposite directions. Elevation data were investigated first, because this dataset had clear spatial structure and a delay time of 0 s, disregarding GPS latency issues, which were assumed to be

minor. Based on travel speed and the distance between the EM38 and the RTK–DGPS antenna, the delay time of the EC data was between –1 s and –2 s. The elevation data (fig. 6) exhibited larger homogeneous regions than did the EC data (fig. 7). The field was divided into four approximately equal–size regions to examine possible variations in delay time.

The developed methods were also applied to yield and crop moisture measurements taken with an AgLeader yield monitoring system on three research fields. The grain flow sensor was located at the top of the clean grain elevator, while the moisture sensor was located at the upper end of the inclined auger inside the grain tank. The data were collected in Field 1 (21 ha, corn, 1997), Field 2 (21 ha, corn, 1996), and Field 3 (17 ha, soybean, 1997) near Centralia, Missouri. Figure 8 shows the grain yield and moisture maps of Field 1
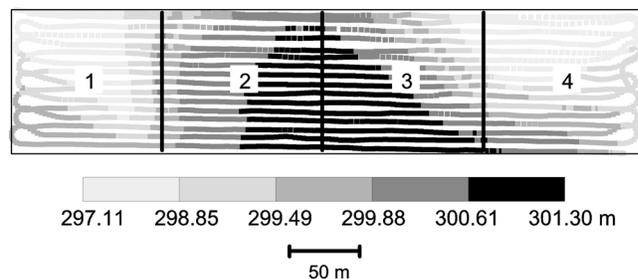

297.11   298.85   299.49   299.88   300.61   301.30 m

50 m

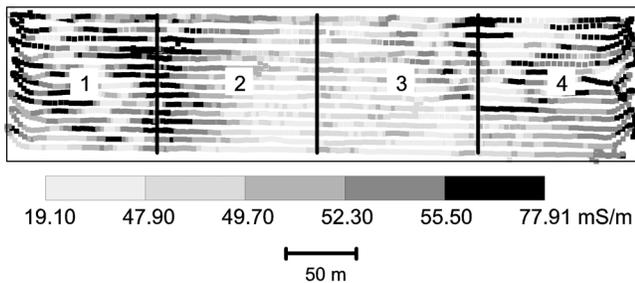**Figure 6. Elevation data used to test methods.**
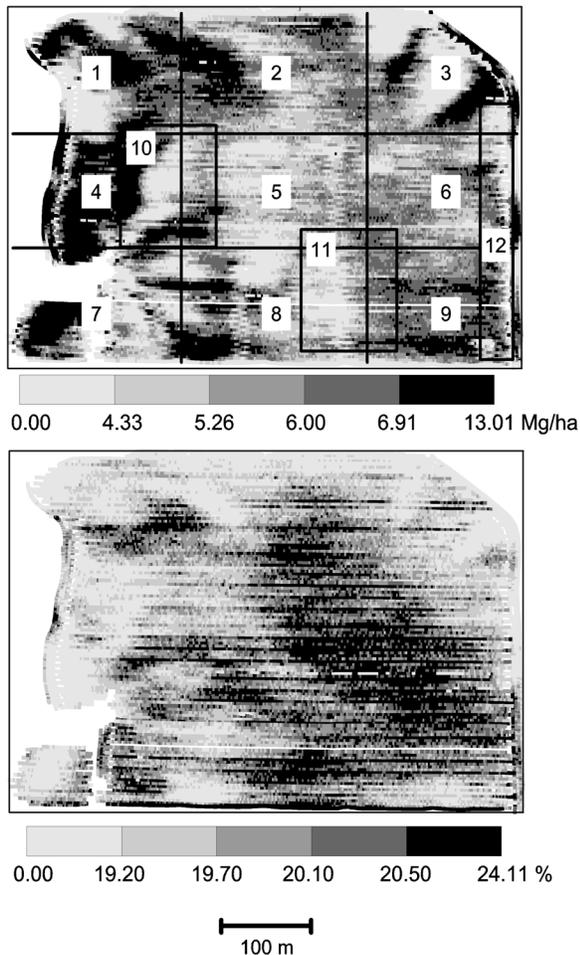
Figure 7. Soil EC data used to test methods.

Figure 8. Field 1 grain yield (top) showing regions defined to investigate variations in delay time. The same division was also applied to the moisture map (bottom).

with delay times corrected by the visual method. Preliminary visual examination of each map revealed the specific spatial structure of the corresponding dataset. For example, the homogeneous areas were smaller in the yield maps compared to the moisture maps. Boundaries between homogeneous areas were clearer in the yield map than in the moisture map for Field 1, while boundaries were clearer in the moisture map for Field 2.

Variations in spatial structure could also be seen within a single map. Because of this, the fields were divided into sub–field regions to investigate possible variations in delay time. For example, Field 1 was divided into 12 regions (fig. 8). First, the field was divided into 9 regions of approximately equal size, with six containing areas where the combine was entering and leaving the crop and three where essentially no starting or stopping of the combine occurred. Regions 10 and 11 were chosen where clear, large homogeneous areas existed. Region 12 was selected to examine delay time within an area where the combine was entering and leaving the crop.

Since the grain flow sensor and crop moisture sensor were located at different positions on the combine, separate delay times were determined. Descriptive statistics of the yield and moisture datasets are summarized in table 1.

ANALYTICAL PROCEDURES

As previously stated, since standard models may not provide a good fit to the semivariance pattern at short lag distances, we used the *y*–intercept of a linear regression fit to the first few semivariances as a more reliable nugget–like parameter. The lag interval used in calculating the semivariance was 1 m for the ideal dataset and 2 m for the elevation, EC, yield, and moisture datasets, approximating the spacing between data points along the measurement transects. These values provided good resolution in the semivariogram while maintaining adequate numbers of data pairs in each lag interval. For yield datasets, the lag interval could be shortened to 1 m; however, we found no advantage to this increased data density.

Examination of semivariograms revealed that, for the datasets used in this study, the first 10 or more semivariance points were contained within the initial, relatively linear portion of the curve. A preliminary investigation showed that using a smaller number of points resulted in increased estimation errors for the datasets with known delay times. Therefore, we chose to use 10 points in the linear regression to minimize the effects of noise in any single variogram point while representing the shape of the initial portion of the curve. In the non–ideal datasets, these 10 points spanned a 20–m distance. In our previous experience (e.g., Birrell et al.,

Table 1. Descriptive statistics and ranges of delay times examined.

Statistic	Ideal	Elevation (m)	EC (mS/m)	Field 1 Yield (Mg/ha)	Field 1 Moisture (%)	Field 2 Yield (Mg/ha)	Field 2 Moisture (%)	Field 3 Yield (Mg/ha)
Mean	1.5	299.65	51.42	5.61	19.74	9.11	18.35	2.26
C.V. (%)	33.33	0.32	10.11	31.24	5.72	20.55	4.17	18.95
Range of test data	1	4.16	58.73	12.93	19.1	14.89	17.1	3.40
Number of observations	6400	2624	2624	29914	29906	32799	32795	8089
Ranges of delay times examined (s)	(−5, +5)	(−5, +5)	(−5, +5)	(10, 20)	(13, 23)	(10, 22)	(13, 25)	(8, 20)

1996), a 20–m lag distance has generally been well within the initial, linear portion of the grain yield semivariogram. However, users of this procedure are advised to check that this relationship is also true for their data.

S+SpatialStats version 1.5 (MathSoft, Inc., Seattle, Wash.) was used to calculate semivariances and to perform the linear regressions. A range of delay times was applied to each dataset (table 1), and semivariances were calculated. The delay time at which the y–intercept of the regressed line was a minimum was selected as the optimum delay time.

Implementation of the data segmentation procedure required that the data be in a square grid of side 2^n. Since all datasets, with the exception of the ideal datasets, initially consisted of point data, a method was required to convert these data into a grid form. A nearest–neighbor gridding algorithm implemented in Surfer version 7.0 (Golden Software. Inc., Golden, Colo.) was applied to the point data, with the grid size selected to be approximately equal to the harvest transect spacing (3 to 6 m for these datasets). This method provided a close approximation of the point data, but in the grid form required. The data grid was inserted into a square grid of side 2^n, such that 2^n was greater than or equal to both dimensions of the data grid. Grid cells outside the field boundary were assigned a class value of zero to effectively remove them from analysis. Finally, the non–zero cells in the grid were classified into a user–determined number of classes, based either on equal size intervals or on placing an equal number of observations in each class. The segmentation algorithm was applied to the resulting grid, and the total number of segments with a non–zero classification value was calculated.

In the data segmentation procedure, the number of classes selected could have significant impact on the results of the analysis. Too few classes could result in over–simplification of the yield distribution and the resulting grid. Too many classes could create too much complexity, making the pattern extremely noisy. In addition, the classification scheme used could have significant impact on the results. Selecting intervals of equal size could produce a more intuitive, more natural–looking classification to represent the yield distribution, but this method could be severely compromised when extreme values were present. The equal number of observations per class method could over–exaggerate small variations but would be more stable where extreme values were present. For this analysis, both 5–class and 10–class intervals were investigated, for both equal interval and equal number classification schemes.

With an incorrect delay time, the number of segments obtained from a yield map would be increased due to the "saw tooth" pattern along the boundaries between adjacent homogeneous areas. However, at the correct delay time, this segmentation number should be minimized. Further, this segmentation number would, in some sense, give an indication of the "complexity" of the image. The ratio of the segmentation number at optimum delay time to the maximum possible segmentation number, equal to the total number of grid cells, was calculated for each dataset to assess the complexity of the map and to compare the degree of complexity between datasets. This "complexity ratio" is a minimum when the entire grid contains the same value, and it increases as the number of heterogeneous features next to each other increases, up to maximum when each cell is different from any adjacent cells. It should be clear that as this

ratio nears a maximum, the ability of the data segmentation method to accurately estimate delay time becomes severely reduced.

A subjective, visual method commonly used in determining delay time was also applied to the grain yield and moisture datasets. An individual experienced in processing yield data examined each yield and moisture dataset visually and selected delay times for each region in the field such that discrepancies in visual features were minimized. In addition, the individual rated the selection of each delay time with a confidence factor (1 to 10) that subjectively indicated the level of clarity of the map and the resulting certainty in the selection of that delay time. The results of the geostatistical and data segmentation methods were then compared to the delay times estimated by the visual method.

RESULTS AND DISCUSSION

GEOSTATISTICAL METHOD

Figure 9 shows variograms for the ideal dataset with delay times of 0 and 5 s. The semivariance of the dataset at a delay time of 0 s increases with a linear pattern until lag distance reaches 10 m (10 cells) and then oscillates. This variogram pattern indicates that there is one variation component within 10 m of lag distance and another periodic component beyond that range. The variogram for the dataset at 5 s delay time shows a pure nugget effect, representing the dominance of a small–scale variation without other large–range components. The sills of the two variograms are approximately identical. Figure 10 shows how the intercept of the regression line responded to changes in delay time for the ideal dataset. These results supported our hypothesis that the y–intercept should decrease as the homogenous regions of the map increased in size.

Figure 11 shows results of the geostatistical analysis of elevation data (left) and EC data (right) for both the whole–field and region 1 datasets. The results for the elevation data were more stable and symmetrical across zero delay time, compared with the slightly more asymmetrical case of the EC data. As expected, the minimum value consistently occurred at a delay time of 0 s for the elevation data. The results were not quite as well behaved for the EC data, but the delay times were clearly in the range of –1 s to
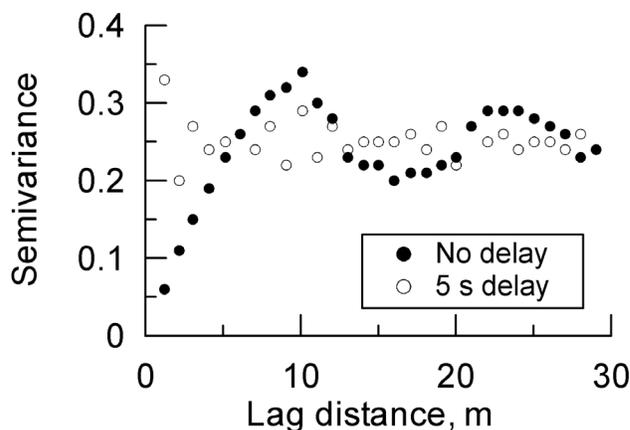
Figure 9. Semivariogram of the ideal test data at zero and 5 s delay times.
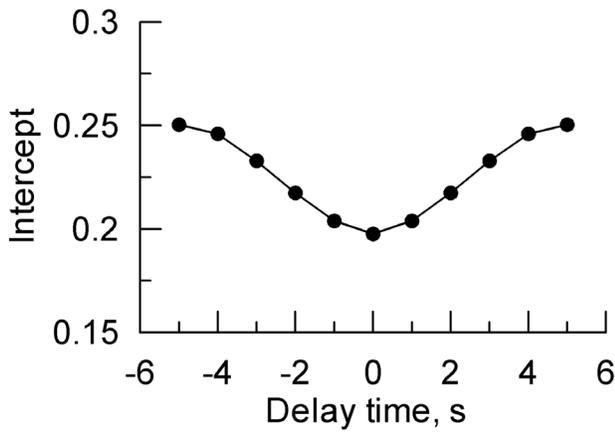
Figure 10. Change of variogram intercept with different delay times for the ideal dataset.

showed a constantly decreasing intercept over the range of examined delay times, not clearly indicating an optimal delay time. The moisture data from region 10 showed slightly better results, locating a marginally optimal delay time of 17 s. Visual analysis of these maps provided similar results (table 3). The visual method gave a delay time of 14 s for grain yield in region 10, with a high confidence rating (9) for this estimate. A delay time of 12 s was chosen for yield in region 11, with a slightly lower confidence rating (8). The visual method was not able to indicate an appropriate delay time for the moisture data in region 11, and in region 10 a delay time of 17 s was selected with the lowest of confidence ratings (1).

–2 s. This variability was reasonable, given the non–constant speed of the ATV during data collection. Estimated delay times for the elevation and EC data are summarized in table 2. Delay times were within the expected ranges for all analyses of these datasets.

Figure 12 displays results for yield and moisture data from regions 10 and 11 of Field 1. The yield data showed clear patterns, with an optimum delay time of 14 s in region 10 and 13 s in region 11. The results for grain moisture were less clear, possibly due in part to the relatively small range of moisture variation within this field. The results for region 11

Estimated delay times for the grain yield and moisture datasets are summarized in tables 3 and 4. For yield data, the delay times estimated by the geostatistical method were the same as those determined by the visual method in 5 of the 12 regions of Field 1 (table 3). In the seven other regions of that field, the geostatistical method differed from the visual observer by only 1 s. Averaged over all regions of Field 1, the delay time from the geostatistical approach was about 1 s greater than from visual observation. For moisture measurements, estimated delay times were the same as those determined by the visual method in 3 of the 12 regions of Field 1, and 3 other regions showed a 1 s difference. The geostatistical method could not determine optimum delay times in 2 regions where the visual observer also could not find map features clear enough to estimate delay time.
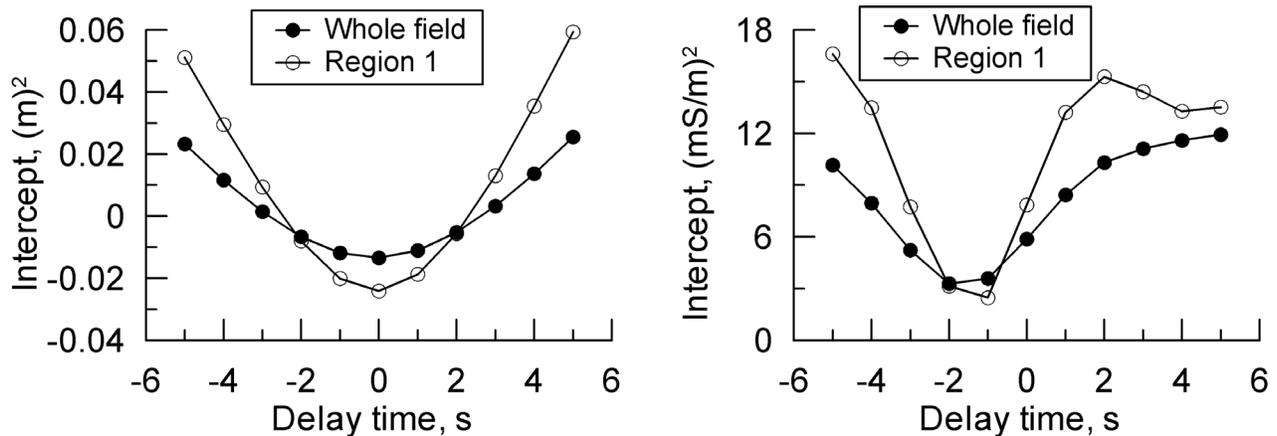

Figure 11. Change of variogram intercept with different delay times for elevation data (left) and soil EC data (right).

Table 2. Delay times determined for elevation and EC data using variogram intercept and segmentation methods.

	Variogram Intercept[a]		Segmentation, 5 Classes[b]				Segmentation, 10 Classes[b]			
			Equal Number		Equal Interval		Equal Number		Equal Interval	
	Elevation	EC	Elevation	EC	Elevation	EC	Elevation	EC	Elevation	EC
Whole field	0 (–0.013)	–2 (3.293)	0 (0.348)	–2 (0.717)	0 (0.327)	–3 (0.332)	0 (0.584)	–2 (0.883)	0 (0.526)	–2 (0.606)
Region 1	0 (–0.024)	–1 (2.470)	–1 (0.427)	–2 (0.660)	0 (0.468)	–3 (0.537)	–2 (0.711)	–1 (0.843)	0 (0.706)	–2 (0.734)
Region 2	0 (–0.001)	–2 (3.555)	0 (0.351)	–2 (0.559)	0 (0.453)	–1 (0.365)	0 (0.671)	–3 (0.807)	0 (0.706)	–1 (0.537)
Region 3	0 (–0.010)	–1 (1.657)	1 (0.256)	–1 (0.763)	1 (0.294)	[c]	1 (0.542)	–1 (0.917)	1 (0.515)	–1 (0.862)
Region 4	0 (–0.021)	–2 (1.733)	0 (0.383)	–2 (0.625)	–1 (0.453)	–2 (0.542)	–2 (0.621)	–2 (0.834)	–1 (0.659)	–2 (0.708)
Range	0	(–2, –1)	(–1, 0)	(–2, –1)	(–1, 1)	(–3, –1)	(–2, 1)	(–3, –1)	(–1, 1)	(–2, –1)
Average	0.00	–1.50	0.00	–1.75	0.00	–2.00	–0.75	–1.75	0.00	–1.50

[a] Value in parentheses is y–intercept: m^2 for elevation and $(mS/m)^2$ for EC.

[b] Value in parentheses is the ratio of segmentation number to the number of grid cells.

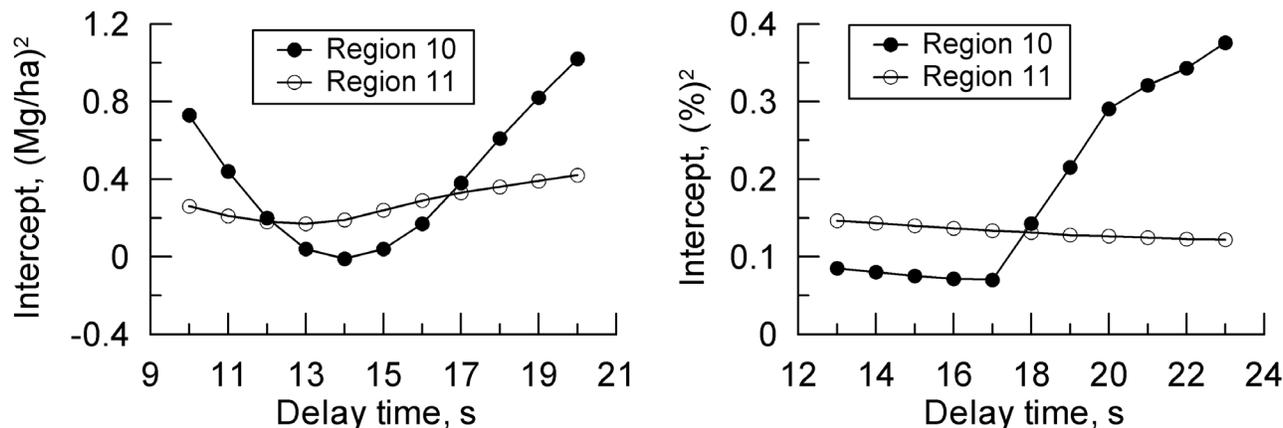[c] Delay time estimation was not obtained within the range examined.

Figure 12. Change of variogram intercept with different delay times for grain yield (left) and moisture data (right) in regions 10 and 11 of Field 1.

Table 3. Summary of delay times determined for Field 1 using variogram intercept, data segmentation, and visual methods.

	Yield			Moisture		
	Intercept[a]	Segmentation[b]	Visual[c]	Intercept[a]	Segmentation[b]	Visual[c]
Whole field	14 (0.342)	14 (0.707)	14 (7)	21 (0.553)	19 (0.790)	19 (3)
Region 1	14 (0.093)	14 (0.584)	14 (7)	19 (0.059)	20 (0.708)	18 (3)
Region 2	14 (0.218)	13 (0.743)	13 (7)	20 (0.074)	20 (0.804)	19 (2)
Region 3	13 (0.734)	14 (0.587)	14 (8)	14 (0.730)	18 (0.693)	18 (1)
Region 4	14 (0.405)	15 (0.774)	14 (9)	18 (0.063)	21 (0.774)	18 (2)
Region 5	13 (0.211)	14 (0.820)	13 (4)	16 (1.322)	19 (0.846)	[e]
Region 6	15 (0.214)	15 (0.820)	14 (5)	[d]	21 (0.819)	[e]
Region 7	14 (0.494)	16 (0.609)	14 (7)	18 (0.287)	17 (0.676)	18 (1)
Region 8	13 (0.237)	13 (0.738)	12 (8)	17 (0.225)	19 (0.836)	18 (1)
Region 9	14 (0.362)	14 (0.856)	13 (4)	21 (0.206)	[d]	[e]
Region 10	14 (–0.011)	14 (0.702)	14 (9)	17 (0.070)	17 (0.833)	17 (1)
Region 11	13 (0.170)	13 (0.801)	12 (8)	[d]	20 (0.911)	[e]
Region 12	14 (0.550)	16 (0.818)	13 (6)	15 (0.605)	[d]	[e]
Range	13 – 15	13 – 16	12 – 14	14 – 21	17 – 21	17 – 19
Average	14.25	13.75	13.33	19.20	17.5	18.0

[a] Value in parentheses is y–intercept: $(Mg/ha)^2$ for yield and $\%^2$ for moisture.
[b] Using 5 classes and equal number classification; value in parentheses is the ratio of segmentation number to the number of grid cells.
[c] Value in parentheses is qualitative confidence factor (1–10 scale).
[d] Optimum delay time was not estimated within the range of times examined.
[e] Map features not clear enough for visual estimate of delay time.

Table 4. Summary of delay times determined for Field 2 and 3 using variogram intercept, data segmentation, and visual methods.

	Yield, Field 2			Moisture, Field 2			Yield, Field 3		
	Intercept[a]	Segmentation[b]	Visual[c]	Intercept[a]	Segmentation[b]	Visual[c]	Intercept[a]	Segmentation[b]	Visual[c]
Whole field	17 (1.181)	21 (0.846)	17 (4)	20 (0.196)	22 (0.798)	22 (5)	12 (0.053)	10 (0.888)	12 (4)
Region 1	19 (2.087)	23 (0.831)	18 (2)	20 (0.279)	18 (0.830)	22 (4)	12 (0.057)	17 (0.923)	12 (3)
Region 2	15 (0.930)	[d]	17 (1)	[d]	22 (0.873)	23 (3)	10 (0.047)	11 (0.939)	11 (1)
Region 3	19 (0.869)	[d]	[e]	20 (0.096)	21 (0.894)	21 (1)	13 (0.074)	12 (0.918)	12 (2)
Region 4	17 (1.491)	19 (0.800)	17 (3)	21 (0.246)	21 (0.821)	22 (3)	12 (0.033)	10 (0.893)	12 (4)
Region 5	16 (0.698)	[d]	[e]	21 (0.118)	22 (0.859)	22 (3)			
Region 6	16 (1.012)	19 (0.955)	16 (1)	18 (0.072)	23 (0.909)	22 (1)			
Region 7	15 (1.483)	23 (0.823)	16 (2)	[d]	[d]	21 (2)			
Region 8	17 (0.655)	23 (0.864)	17 (1)	[d]	23 (0.931)	[e]			
Region 9	18 (0.940)	[d]	[e]	17 (0.085)	20 (0.945)	[e]			
Range	15 – 19	19 – 23	16 – 18	17 – 21	18 – 23	21 – 23	10 – 13	10 – 17	11 – 12
Average	16.89	21.40	16.83	19.50	21.25	21.86	11.75	12.50	11.75

[a] Value in parentheses is y–intercept: $(Mg/ha)^2$ for yield and $\%^2$ for moisture.
[b] Using 5 classes and equal number classification; value in parentheses is the ratio of segmentation number to the number of grid cells.
[c] Value in parentheses is qualitative confidence factor (1–10 scale).
[d] Optimum delay time was not estimated within the range of times examined.
[e] Map features not clear enough for visual estimate of delay time.

The geostatistical method also performed quite well for Fields 2 and 3 (table 4), where map features were generally less clear than for Field 1. Delay times for yield were within 2 s of those determined by the visual method for both fields. The confidence level assigned by the visual observer was low in all cases. For moisture data from Field 2, the geostatistical method determined a delay time in 6 of 9 regions, and in 4 of these 6 regions the delay time differed by less than 2 s from the visual method. When averaged over all regions of Field 2, the geostatistical method under–estimated moisture delay times by about 2 s compared with the results by the visual method. Yield delay times were nearly identical for geostatistical and visual methods when averaged over Field 2 and Field 3.

Figure 13 illustrates an issue associated with the application of the geostatistical method to region 3 of Field 1. The left image in this figure shows the yield map created with a 13 s delay time, as determined by the geostatistical method, while the right image used a 14 s delay time, as selected by the visual observation method. By comparing the circled areas, it is clear that the 14 s delay was superior to the 13 s delay. However, there were a number of data points at the start or end of several transects where measured yield was quite low, but the adjacent yield in the end rows was high. When these low–yield points fall very near high–yield points in the end rows, the semivariance at very short lags (i.e., the nugget) increases rapidly. This indicates that the effect of spurious data may be significant for the geostatistical method, and must be considered before the method is applied.

Another point can be demonstrated by the yield map for region 7 of Field 1 (fig. 14). The delay times estimated by both the geostatistical and visual methods were the same (14 s). However, a more detailed visual investigation of the map revealed different delay times in the smaller regions seen in the left and right portions of the figure. The left portion showed a clearer image at 14 s (fig. 14, left), while the right portion showed better results with a 13 s delay time (fig. 14, right). This illustrates that subdivision of fields into smaller areas may result in better estimation of optimal delay time.

DATA SEGMENTATION METHOD

The results from the segmentation method were similar to those for the geostatistical method. Figure 15 shows the results from the ideal dataset, where the segmentation method produced an estimated delay time of zero, as did the geostatistical method.

Segmentation was performed on the elevation and EC data with 5 and 10 classes based on the equal number and equal interval classification schemes, with the results summarized in table 2. The estimated delay time for the EC data was occasionally outside the range of expected values (–1 to –2 s) with the 5–class/equal interval and the 10–class/equal number classification schemes. The 5–class/equal number and 10–class/equal interval classification schemes provided estimated delay times within the expected ranges in all cases.
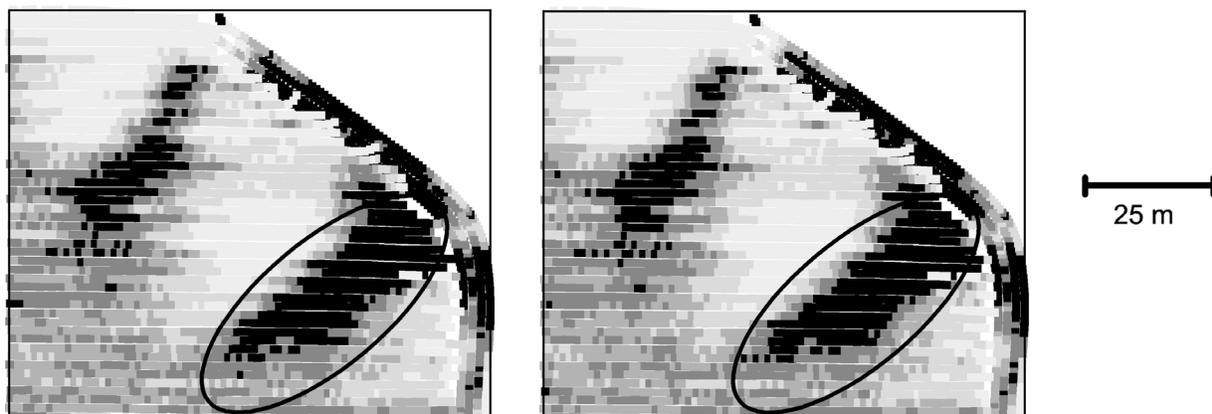

Figure 13. Yield maps for region 3 of Field 1 (fig. 8), after application of the 13 s delay time from the geostatistical method (left) and the 14 s delay time from the visual method (right).
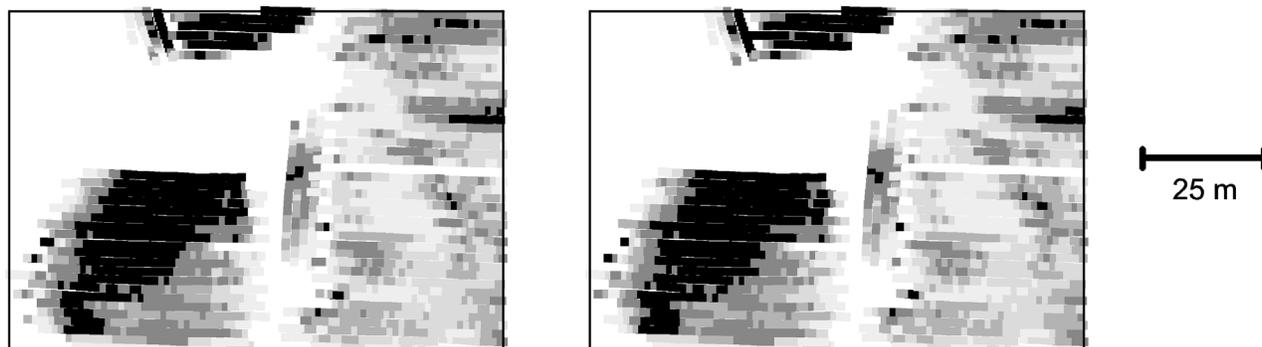

Figure 14. Yield maps for region 7 of Field 1 (fig. 8), after application of a 14 s delay time (left) and a 13 s delay time (right).

For yield and crop moisture measurements, the 5–class/equal number classification scheme was chosen since this classification provided more reliable delay time estimates on elevation and EC data. Results are summarized in tables 3 and 4. Variation in segmentation number with changes in delay time for region 10 of Field 1 for both yield and moisture data is displayed in figure 16. The pattern of variation in segmentation number appeared more complex but gave similar results to those from the geostatistical and visual methods. For 4 of the 12 regions of Field 1, the segmentation method achieved the same delay time for yield as did the visual method. For 6 additional regions, the results differed by only 1 s. Only two regions of the field differed by more than 1 s. As with the geostatistical method, there were two regions of the field (regions 9 and 12) where the segmentation method was unable to estimate delay time for the moisture data.

Delay times determined by the data segmentation method differed by less than 1 s from the visual method in 5 of 9 regions for the Field 2 moisture data, and by less than 2 s in 3 of 4 regions for the Field 3 yield data (table 4). The data segmentation method did not perform well for the Field 2 grain yield measurements. Delay times were not determined in 4 regions, and were over–estimated by about 5 s when compared to the visual method over all regions. Visual estimation was also difficult for this dataset.

COMPARING METHODS

The performance of the visual method was influenced by the clarity of the features in the dataset examined. When features on a map were clear and obvious to the visual observer, the confidence factor selected was relatively high. From clearest to most ambiguous, the datasets were: ideal case, elevation, EC, Field 1 yield, Field 1 moisture, Field 2 moisture, Field 3 yield, and Field 2 yield. Map clarity, as indicated by the confidence rating of the visual observer, will not necessarily have the same meaning as the complexity ratio of a map indicated by the data segmentation method (one can clearly envision a map that is very complex but still contains many clear, small, spatially distributed patterns). However, for the datasets investigated in this study, the complexity ratios showed very similar trends to the confidence ratings of the visual method. The complexity ratios were 0.123 for the ideal dataset, 0.256 to 0.427 for elevation

data, and 0.559 to 0.763 for EC data when the 5–class/equal number classification scheme was applied (table 2). For the yield and moisture maps, Field 1 had smaller complexity ratios than Field 2 and Field 3. The complexity ratios of yield maps were smaller (0.584 to 0.856) than those of moisture maps (0.693 to 0.911) on Field 1 (table 3), and moisture data had smaller values (0.798 to 0.945) than yield (0.800 to 0.955) on Field 2 (table 4). One likely factor causing the high complexity ratio on Field 2 was a systematic noise component (caused by management factors) contained in the yield map. The Field 2 yield map showed a number of alternating adjacent strips of high and low yield running in a north–south direction, which would be detrimental to the performance of the data segmentation method. When compared to the visual method, the geostatistical method showed better performance for most yield datasets, while the data segmentation method generally provided better results on the moisture datasets (tables 3 and 4).

In spite of noise in the datasets, the geostatistical and data segmentation methods performed quite well and showed similar results in the estimation of delay times for data collected in a back–and–forth pattern. The geostatistical method used the y–intercept of a linear regression in the short–range, rapidly increasing portion of a variogram to produce a "nugget–like" result. This method produced a
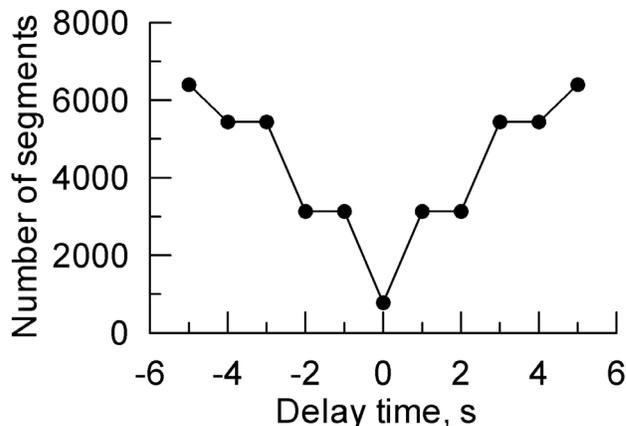

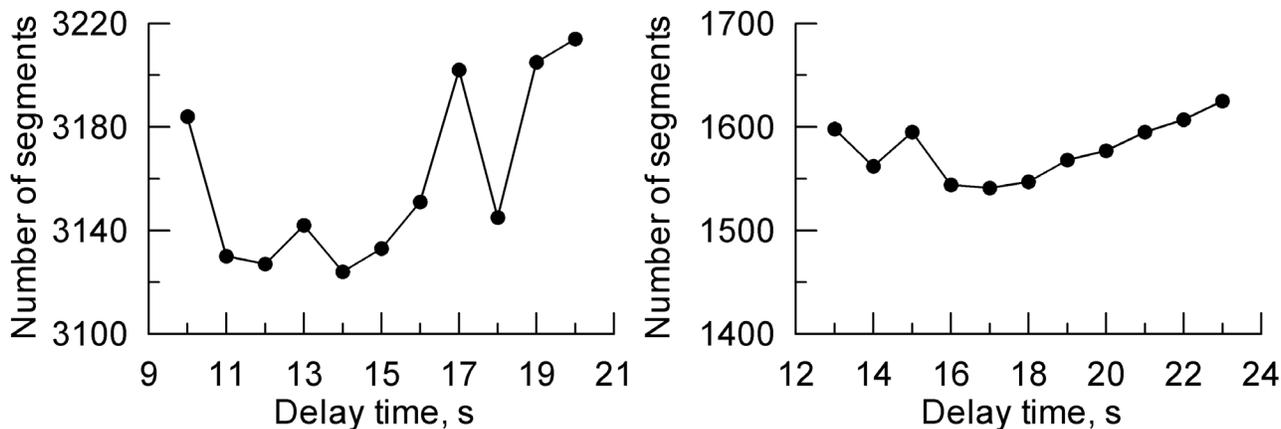Figure 15. Number of segments for the simulated ideal dataset, showing a minimum at zero delay time.

Figure 16. Number of segments for the yield data (left) and moisture data (right) of region 10 of Field 1.

more reliable pattern in the objective function used to estimate delay time than did the data segmentation method. This method, in effect, measured the amount of variability over small lag distances, identifying variations in the borders between adjacent homogeneous areas. For this reason, large differences in measurement values that are very close together, whether due to edge effects or spurious data, would cause significant errors in the estimation of delay time.

The data segmentation approach should be far less susceptible to these problems, since individual spurious data points would, in general, affect at most a very small region of the grid. In fact, in region 3 of Field 1, where the geostatistical method had significant difficulty, as shown by a relatively large intercept (0.734 (Mg/ha)^2), the segmentation method selected the same delay times as the visual method. The visual method experienced difficulty in determining moisture content delay times, having very low confidence ratings (6 regions) or no estimate at all (5 regions) on 11 of the 12 regions in Field 1. In contrast, at least one of the two objective methods estimated a delay time for every region in Field 1 (table 3). Only for one case (moisture data for region 7 of Field 2) could neither method determine a delay time. The visual confidence rating for this data was also very low (table 4). We might expect that an approach combining both methods could increase the accuracy of delay time estimation. Automated procedures to estimate delay times for yield monitoring systems seem to show promise.

SUMMARY

Geostatistical and data segmentation methods were developed to estimate delay time in yield monitoring data with objective criteria. These methods were validated with ideal datasets and near–ideal datasets (elevation and EC data) and were then applied to grain yield and moisture data. Major results of the study were:

- A geostatistical method was developed to take advantage of the variability structure over small lag distances as a measure of discrepancy between adjacent homogeneous areas. A "modified nugget" was calculated, and the optimum delay time was determined at its minimum value.
- A data segmentation method was developed by modifying the concept of a region quadtree. This method evaluated the degree of complexity required to accurately represent a given set of data, with the assumption that less complexity would be seen in data with optimal delay times. Data were transformed into a grid, classified, and segmented recursively until all the values were the same within a given sub–field region. The number of classes and the classification scheme used influenced the performance of this method.
- Both methods performed well on test datasets when results were compared to known delay times. Furthermore, they provided reasonable estimates of yield and moisture delay times when compared to a more subjective, visual method. The geostatistical method performed more reliably in general, but the data segmentation method yielded better results near field boundaries and for moisture data in some cases. The performance of the

methods was influenced by the clarity and complexity of datasets.

Optimal delay times for grain yield and moisture varied in different parts of the test fields. For given sub–field regions, yield and moisture content had different delay times.

REFERENCES

Bakhsh, A., D. B. Jaynes, T. S. Colvin, and R. S. Kanwar. 2000. Spatio–temporal analysis of yield variability for a corn–soybean field in Iowa. *Trans. ASAE* 43(1): 31–38.

Beck, A. D., S. W. Searcy, and J. P. Roades. 2001. Yield data filtering techniques for improved map accuracy. *Applied Eng. in Agric.* 17(4): 423–431.

Birrell, S. J., K. A. Sudduth, and S. C. Borgelt. 1996. Comparison of sensors and techniques for crop yield mapping. *Comput. Elect. Agric.* 14: 215–233.

Blackmore, B. S., and C. J. Marshall. 1996. Yield mapping: Errors and algorithms. In *Proc. 3rd Int'l. Conf. on Precision Agric.,* 403–415. P. C. Robert, R. H. Rust, and W. E. Larson, eds. Madison, Wisc.: ASA, CSSA, and SSSA.

Blackmore, B. S., and M. R. Moore. 1999. Remedial correction of yield map data. *Precision Agric.* 1: 53–66.

Borgelt, S. C., R. E. Wieda, and K. A. Sudduth. 1997. Geostatistical analysis of soil chemical properties from nested grids. *Applied Eng. in Agric.* 13(4): 477–483.

Chosa, T., K. Kobayashi, M. Omine, and K. Toriyama. 2000. Yield–mapping algorithm for a head–feeding rice combine. In *Proc. 5th Int'l. Conf. on Precision Agric.,* CD–ROM. Madison, Wisc.: ASA, CSSA, and SSSA.

Chung, S. O., J. H. Sung, K. A. Sudduth, S. T. Drummond, and B. K. Hyun. 2000. Spatial variability of yield, chlorophyll content, and soil properties in a Korean rice paddy field. In *Proc. 5th Int'l. Conf. on Precision Agric.,* CD–ROM. Madison, Wisc.: ASA, CSSA, and SSSA.

Colvin, T. S., and S. Arslan. 2000. A review of yield reconstruction and sources of errors in yield maps. In *Proc. 5th Int'l. Conf. on Precision Agric.,* CD–ROM. Madison, Wisc.: ASA, CSSA, and SSSA.

Drummond, S. T., C. W. Fraisse, and K. A. Sudduth. 1999. Combine harvest area determination by vector processing of GPS position data. *Trans. ASAE* 42(5): 1221–1227.

Han, S., S. M. Schneider, S. L. Rawlins, and R. G. Evans. 1997. A bitmap method for determining effective combine cut width in yield mapping. *Trans. ASAE* 40(2): 485–490.

Hanan, S. 1990. *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS.* Boston, Mass.: Addison Wesley.

Journel, A. G., and C. J. Huijbregts. 1978. *Mining Geostatistics.* London, U.K.: Academic Press.

Lamb, J. A., J. L. Anderson, G. L. Malzer, J. A. Vetch, R. H. Dowdy, D. S. Onken, and K. I. Ault. 1995. Perils of monitoring grain yield on–the–go. In *Site–Specific Management for Agricultural Systems,* 87–90. P. C. Robert, R. H. Rust, and W. E. Larson, eds. Madison, Wisc.: ASA, CSSA, and SSSA.

Leitao, J. M., A. Sousa, and F. Ferreira. 1996. Ray–tracing for stereoscopic images. In *Modeling and Graphics in Science and Technology,* 244–257. J. C. Teixeira and J. Rix, eds. Berlin, Germany: Springer.

Missotten, B., G. Strubbe, and J. De Baerdemaeker. 1996. Accuracy of grain and straw yield mapping. In *Proc. 3rd Int'l. Conf. on Precision Agric.,* 713–722. P. C. Robert, R. H. Rust, and W. E. Larson, eds. Madison, Wisc.: ASA, CSSA, and SSSA.

Moore, M. R. 1998. An investigation into the accuracy of yield maps and their subsequent use in crop management. PhD diss. Silsoe, U.K.: Cranfield University, Silsoe College.

Murphy, D. P., E. Schnug, and S. Haneklaus. 1995. Yield mapping: A guide to improved techniques and strategies. In *Site–Specific Management for Agricultural Systems,* 33–47. P. C. Robert, R. H. Rust, and W. E. Larson, eds. Madison, Wisc.: ASA, CSSA, and SSSA.

Nolan, S. C., G. W. Haverland, T. W. Goddard, M. Green, and D.C. Penney. 1996. Building a yield map from geo–referenced harvest measurements. In *Proc. 3rd Int'l. Conf. on Precision Agric.,* 885–892. P. C. Robert, R. H. Rust, and W. E. Larson, eds. Madison, Wisc.: ASA, CSSA, and SSSA.

O'Neal, M. R., J. R. Frankenberger, S. D. Parsons, D. R. Ess, M. T. Crisler, and R. M. Strickland. 2000. Correcting yield monitor data for improved yield mapping. ASAE Paper No. 001088. St. Joseph, Mich.: ASAE.

Pierce, F. J., N. W. Anderson, T. S. Colvin, J. K. Schueller, D. S. Humburg, and N. B. McLaughlin. 1997. Yield mapping. In *The State of Site–Specific Management for Agriculture,* 211–245. F. J. Pierce and E. J. Sadler, eds. Madison, Wisc.: ASA, CSSA, and SSSA.

Rands, M. 1995. The development of an expert filter to improve the quality of yield mapping data. MS thesis. Silsoe, U.K.: Cranfield University, Silsoe College.

Reitz, P., and H. D. Kutzbach. 1996. Investigations on a particular yield mapping system for combine harvesters. *Comput. Elect. Agric.* 14: 137–150.

Searcy, S. W., J. K. Schueller, Y. H. Bae, S. C. Borgelt, and B. A. Stout. 1989. Mapping of spatially variable yield during grain combining. *Trans. ASAE* 32(3): 826–829.

Stafford, J. V., B. Ambler, R. M. Lark, and J. Catt. 1996. Mapping and interpreting the yield variation in cereal crops. *Comput. Elect. Agric.* 14: 101–119.

Stott, B. L., S. C. Borgelt, and K. A. Sudduth. 1993. Yield determination using an instrumented Claas combine. ASAE Paper No. 931507. St. Joseph, Mich.: ASAE.

Sudduth, K. A., S. T. Drummond, W. Wang, M. J. Krumpelman, and C. W. Fraisse. 1998. Ultrasonic and GPS measurement of combine swath width. ASAE Paper No. 983096. St. Joseph, Mich.: ASAE.

Sudduth, K. A., S. T. Drummond, and N. R. Kitchen. 2001. Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. *Comput. Elect. Agric.* 31: 239–264.

Vassilakopoulos, M., and Y. Manolopoulos. 1993. Analytical results on the quadtree storage–requirements. In *Computer Analysis of Images and Patterns: 5th Int'l. Conf., CAIP '93,* 41–48. D. Chetverikov and W. G. Kropatsch, eds. Budapest, Hungary.

Webster, R., and M. A. Oliver. 1990. *Statistical Methods in Soil and Land Resource Survey.* New York, N.Y.: Oxford University Press.

Whelan, B. M., and A. B. McBratney. 2002. A parametric transfer function for grain–flow within a combine conventional harvester. *Precision Agric.* 3(2): 123–134.