SOIL AND CROP MANAGEMENT

Spatial–Temporal Analysis of Yield and Soil Factors in Two Four-Crop–Rotation Fields in the Sacramento Valley, California

J. F. Perez-Quezada, G. S. Pettygrove, and R. E. Plant*

ABSTRACT

Data were analyzed from two 30-ha laser-leveled commercial fields in the Sacramento Valley, California, for crops grown between 1995 and 1999. Crops were wheat (*Triticum aestivum* L.), tomato (*Lycopersicon esculentum* Mill.), bean (*Phaseolus vulgaris* L.), sunflower (*Helianthus annuus* L.), and corn (*Zea mays* L.). One of the fields had a wheat–tomato–bean–sunflower rotation while the other had a wheat–tomato–sunflower–corn rotation during the same period. Yield data were collected with a commercial yield monitor, except for tomato yield data, which were collected with an experimental yield monitor. After georeferencing, interpolation, and correction of yield-monitored data, analysis was performed on grid cells representing an area of 20-m square. Soil cores were extracted from the fields on a 60-m square grid. Yield and sand content were both separated into large- and small-scale components by median polish. Persistent large-scale trends in yield, which were consistent with large-scale trends in sand content, were observed. Six 2-yr comparisons via linear regression were performed in each field. Yield of one crop was a poor predictor of yield of another crop grown in another year, but after standardizing and averaging the yields, areas with the same average performance tended to be clustered together spatially. The standardized yields were also analyzed using *K*-means clustering. This provided a different spatial configuration of clusters from that of the standardized average but also a high level of spatial autocorrelation, which shows that both methods may be helpful in delineating management zones at the scale normally used by growers.

Precision agriculture, or site-specific management (SSM), involves the management of the crop at a spatial scale smaller than that of the field. It depends on understanding the processes and factors that regulate crop responses to within-field variability and on being able to predict the spatial pattern of yield response to these factors. The practical implementation of SSM has been greatly facilitated by the introduction of commercial yield monitors that permit the measurement and analysis of yield distribution at a spatial distribution on the scale of meters. Data collected from these yield monitors, and supplemented by other agronomic data of varying types, provide the opportunity to analyze the spatial distribution of yield and to relate it to the spatial distribution of agronomic factors at a previously unattainable level of precision at the commercial field scale.

A number of researchers have examined interannual variability in yield spatial distribution. Colvin et al. (1995) analyzed yield transects in a 4-yr corn–soybean [*Glycine max* (L.) Merr.] rotation in Iowa. They found a high level of variability in most areas but did not specifically quantify this variability statistically. One means of characterizing interannual variability is through correlation and regression analysis. Huggins and Alderfer (1995) used multiple regression to study the effects of various factors on yield variability in a 34-yr small-plot–based study of corn fertility. They reported that 67% of the variation was explained by climatic and other factors while only 8% was explained by site variability. Sadler et al. (1995, 1998) analyzed yield sequences in corn, wheat, and soybean plots. Interannual yield correlations were statistically significant although the coefficients of determination were fairly low. Lamb et al. (1996, 1997) observed similar results in a 6-yr sequence of corn and corn–soybean systems. Jaynes and Colvin (1997) studied spatial patterns in a 4-yr corn–soybean rotation using median polish and variogram analysis. Lark and Stafford (1997) and Lark et al. (1997) used fuzzy clustering to divide an experimental field into regions characterized by similar yield trends.

Crop response to environment at the field scale is a dynamic process involving the interaction of spatial and temporal effects. One conceptual model that has been put forth for such processes as a means of organizing their complexity is to consider the data as consisting of the sum of two additive components: a large-scale *deterministic* process that is considered primarily reactive in nature and a smaller-scale *stochastic* process that is considered primarily interactive in nature (Cliff and Ord, 1981, p. 222; Cressie, 1991, p. 25; Isaaks and Srivastava, 1989, p. 531). The terms “reactive” and “interactive” are used here in the sense of Cliff and Ord (1981, p. 141). Briefly, reactive processes are those in which yield variability is primarily due to reaction to an external, spatially varying agent while interactive processes are those in which the primary cause of spatial variability is spatial interaction of the process itself. Jaynes and Colvin (1997) used the technique of median polish (Emerson and Hoaglin, 1983; Cressie, 1991) to separate the spatial data into a large-scale trend and a small-scale stochastic structure. Although no biological or physical interpretation of the results of the median polish process is necessary for the statistical analysis, the reactive and/or interactive model given above is consistent with a subdivision by scale. The question of separating spatio-temporal data into large- and small-scale components

J.F. Perez-Quezada, Graduate Group in Hortic. and Agron., G.S. Pettygrove, Dep. of Land, Air, and Water Resour., and R.E. Plant, Dep. of Agron. and Range Sci. and Dep. of Biol. and Agric. Eng., Univ. of California, Davis, CA 95616. Received 31 May 2002. *Corresponding author (replant@ucdavis.edu).

Published in Agron. J. 95:676–687 (2003).

Abbreviations: CV, coefficient of variation; SOM, soil organic matter; SSM, site-specific management.

is therefore of interest for two reasons: empirically, to determine the extent to which it facilitates data analysis, and fundamentally, to determine the extent to which it can be used to model the process of crop interaction with environment.

This paper examines different methods for analyzing spatial and temporal patterns of soil characteristics and yield at the field scale, using data from a 4-yr study of two commercial fields in California. The fields in this study are typical of those found in the lower Sacramento and upper San Joaquin Valleys where crop rotations include grains, vegetables, legumes, and oilseeds. We examine the use of trend analysis to study both crop and soil variability. We also compare various methods for characterizing the interannual relationship of crop yield values. Our objective is to determine whether spatiotemporal yield variability can be statistically characterized. We analyze interannual relationships of yield data and the relationship of yield to soil physical characteristics of the field.

MATERIALS AND METHODS

Both fields subject to study are commercially managed in the Sacramento Valley, California (38°32′ N, 121°58′ W). The fields are identified in this study following the grower's numbering system as Fields 5 and 58. Field 5 is made up of three soil types, Capay silty clay (fine, smectitic, thermic Typic Haploxerepts), Brentwood silty clay loam (fine, smectitic, thermic Typic Haploxerepts), and Yolo silty loam (fine-silty, mixed, superactive, nonacid, thermic Mollic Xerofluvents). There is a textural gradient of increasing clay content from the south end (Yolo) to the north end of the field (Capay). Field 58 has two soil types, Brentwood silty clay loam and Rincon silty clay loam (fine, smectitic, thermic Mollic Haploxeralfs). Both fields are 30 ha in area, and their linear dimensions are approximately 400 by 800 m, with the long axis of Field 5 running in the north–south direction and that of Field 58 in the east–west direction. Both fields have been laser-leveled. The climate is Mediterranean, with an average annual rainfall of 550 mm, almost all of which occurs during the winter. Precipitation during the winter of 1995–1996 was 822 mm, which represents 150% of normal. This occurred during the first year of the study and affected the winter wheat crop grown that season (Fig. 1). During the remaining years, the farmer grew summer crops so that annual precipitation had less effect in this fully irrigated production system. Soil samples to a depth of 30 cm were taken in the first year using a 60-m grid, resulting in 86 samples in Field 5 and 78 samples in Field 58. A detailed description of the measurement procedures is given by Plant et al. (1999).

The experiment was conducted from 1995 through 1999. Crops grown in the fields during this period were wheat, tomato, bean, sunflower, and corn. Field 5 had a wheat–tomato–bean–sunflower rotation, whereas Field 58 had a wheat–tomato–sunflower–corn rotation. The southern portion of Field 5 was removed from the study in the 1998 and 1999 seasons when the grower used it for different crops. Grain and oilseed crops were harvested with a combine equipped with an Ag Leader GPS yield-mapping system (Ag Leader Technol., Ames, IA), and tomato was harvested with an experimental tomato yield monitor. Yield data were collected once per second. In some years, it was not possible to get full GPS coverage, so data were missing from some parts of the fields in some years.

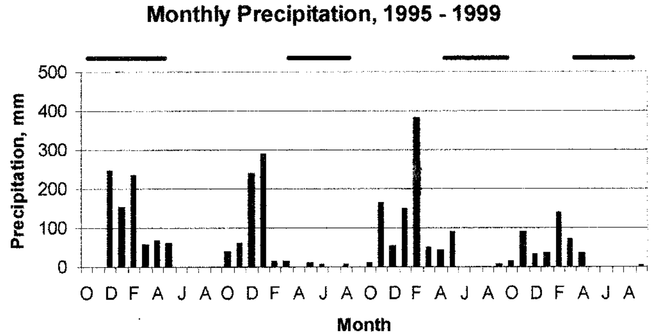

Fig. 1. Monthly precipitation recorded at the California Irrigation Management Information System weather station at Winters, CA, October 1995 through September 1999. Precipitation is measured by season from 1 October through 30 September. Dark bars indicate months in which a crop was grown in the two fields. The first crop was grown in the winter (November– May), and the rest were grown in the summer (April– October). Total precipitation in each season was as follows: 1995–1996, 822 mm; 1996–1997, 683 mm; 1997–1998, 967 mm; and 1998–1999, 423 mm.

Before analysis, yield data were aggregated to a grid of 20- by 20-m cells, where each cell value represented the average of all points contained within that 20- by 20-m square. This process was performed using the same mask coverage each year so that the grids were coincident, which permitted the comparison of yields for different years in the same locations. Data manipulation was performed using ArcView and ArcInfo (ESRI, Redlands, CA). Sample statistics were computed using standard procedures (Steel and Torrie, 1980). Median polish (Cressie, 1991) was performed as described by Jaynes and Colvin (1997). Analysis was performed using Excel (Microsoft, Redmond, WA), SAS (SAS Inst., Cary, NC), and Minitab (Minitab, State College, PA). Experimental variograms and autocorrelation statistics were computed using Variowin (Pannatier, 1996), GS+ (Gamma Design Software, Plainwell, MI), and the statistical package of Lee and Wong (2001).

Cluster analysis was performed using Statistica (StatSoft, Tulsa, OK). Before this analysis, yields were standardized using the formula:

$$Y_s = \left(\frac{Y_i - \overline{Y}}{\mathrm{SD}}\right) \times 100$$

where Y_i is the yield for the ith cell, $\overline{Y}$ is the average yield, and SD is the standard deviation of that particular crop-year. Analysis was performed using K-means clustering, which is an algorithm based on the square-error clustering method (Jain and Dubes, 1984). In this method, each sample is assigned to one of K clusters so that the variance within clusters is minimized and the variance between clusters is maximized. The K-means algorithm achieves this by recalculating the mean of each cluster every time a sample is assigned to it. The K-means clustering purposefully excluded spatial information so that spatial relationships of the cluster sets could be analyzed statistically.

To obtain a temporal trend analysis, data were standardized using a formula based on a similar analysis of Blackmore (2000):

$$Y_{s,i} - \left(\frac{Yi}{\overline{Y}}\right) \times 100$$

where $Y_{s,j}$ is the standardized yield value of the cell. By adding the standardized yields, a statistic called the *sumscore* was calculated. This represents the average standardized yield over the 4-yr period. To characterize the temporal stability, working also with the standardized yields, the coefficient of variation (CV) over years was calculated for each cell as in Blackmore (2000), using the 4 yr of data.

Multiple-regression analysis was performed in which yield was the response variable and relatively persistent soil quantities measured at the start of the experiment were predictors. We attempted a regression tree analysis similar to that which Plant et al. (1999) performed on the first year's data. However, possibly because of the reduced number of data values available for the full 4-yr sequence of crops, the analysis was not successful. The regression tree grew only one node pair for the data of Field 5 and was unable to grow any nodes for the data of Field 58.

RESULTS

Figures 2 and 3 show yield maps for the two fields, with data aggregated to a 20-m grid as described in the Methods section. Table 1 shows correspondence between basic statistical values of original yield data sets and aggregated data sets used in our analyses. As usual with a change of support (Isaaks and Srivastava, 1989, p. 190), the mean values were little affected while the standard deviation decreased considerably. The decrease in CV was least for wheat, indicating that yield data for that crop had the least short-range variability. The values obtained for the reclassified data sets demonstrate the smoothing effect of aggregating and the re-

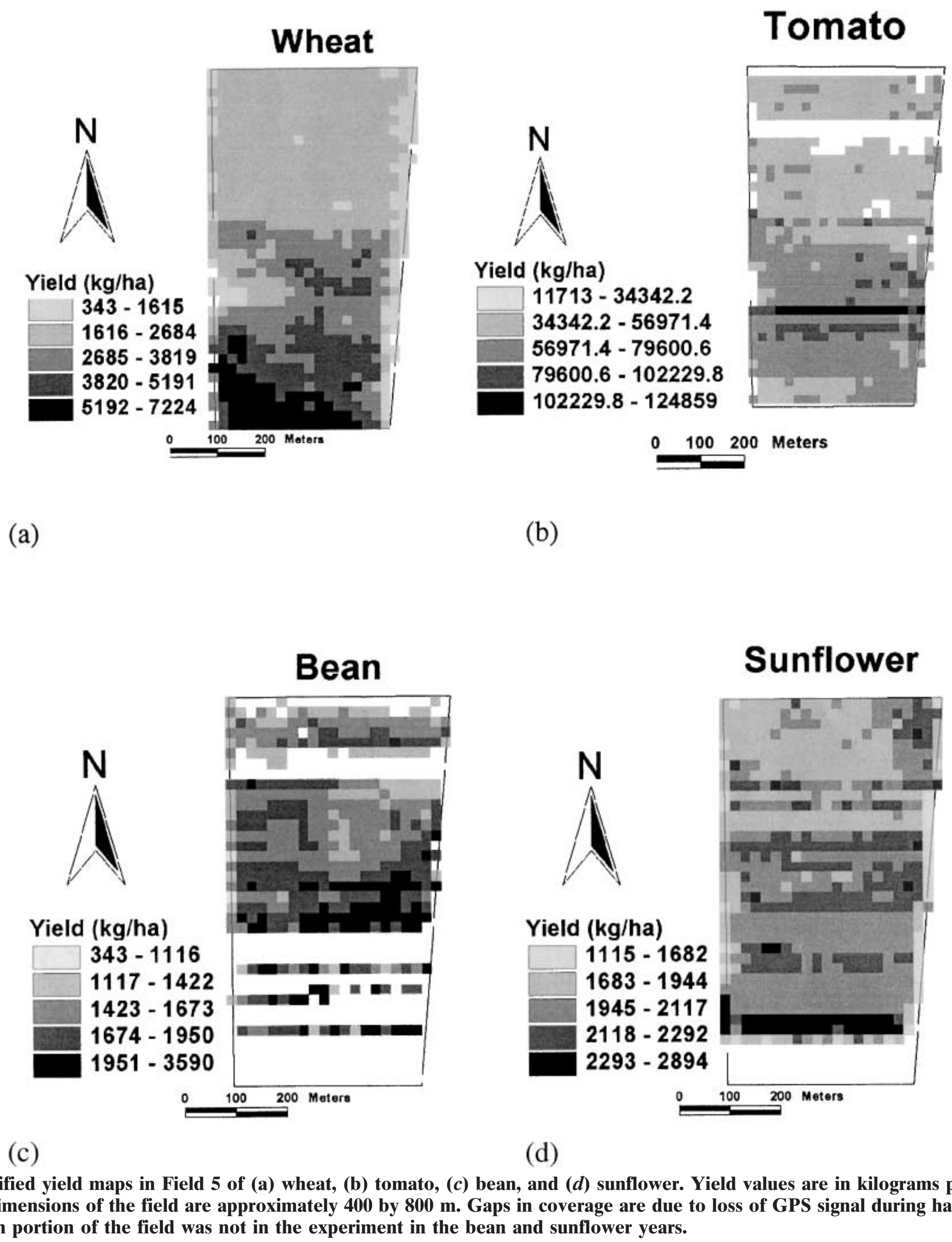

Fig. 2. Reclassified yield maps in Field 5 of (a) wheat, (b) tomato, (*c*) bean, and (*d*) sunflower. Yield values are in kilograms per hectare, and the linear dimensions of the field are approximately 400 by 800 m. Gaps in coverage are due to loss of GPS signal during harvest. Note that the southern portion of the field was not in the experiment in the bean and sunflower years.

moval of outliers, which relates to the sensitivity of the standard deviation to extreme values. When the reclassified mean yield values were compared with those the grower registered from net truck weights, the differences ranged from 0.6 to 81% (Table 2). The monitored yields could in principle be corrected directly by assuming that the truck weights were correct and that the yield monitors worked with the same calibration throughout the field. However, this would be possible only for the crop-years in which the data set was complete (see Fig. 2 and 3). Because the primary interest of this work was spatial yield distribution rather than absolute yield values, we worked with uncorrected data, which assumes that the spatial patterns delineated by the yield monitor were correct.

Field 58 exhibited a generally higher yield and a

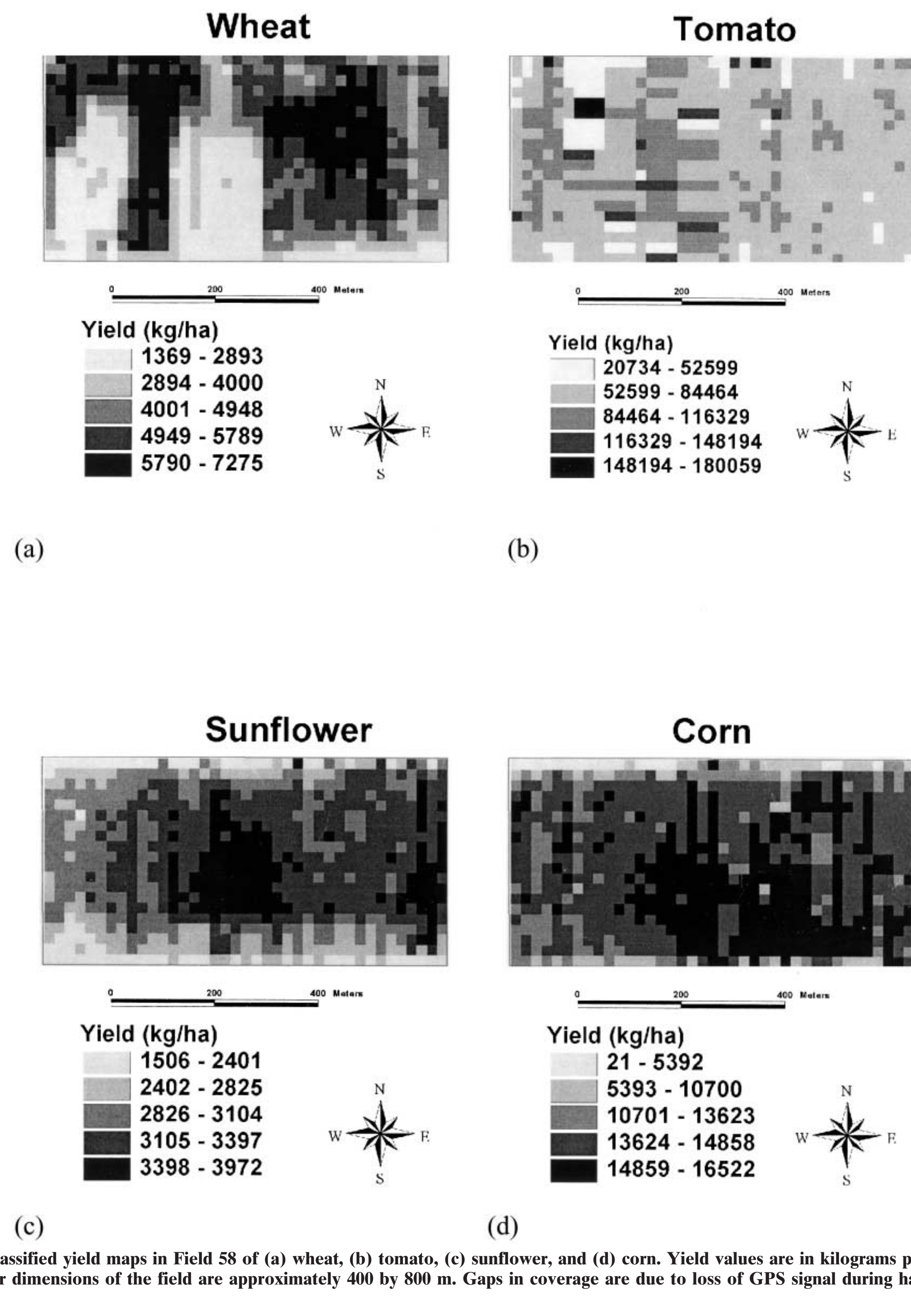

Fig. 3. Reclassified yield maps in Field 58 of (a) wheat, (b) tomato, (c) sunflower, and (d) corn. Yield values are in kilograms per hectare, and the linear dimensions of the field are approximately 400 by 800 m. Gaps in coverage are due to loss of GPS signal during harvest.

Table 1. Basic statistics of monitored yield data before and after aggregation on a 20- by 20-m grid for both fields.

Field	Crop (year)	Before aggregation				After aggregation			
		Data points	Mean	SD†	CV‡	Cells	Mean	SD	CV
			kg/ha		%		kg/ha		%
5	Wheat (1996)	35 816	2 931	1 483	50.6	776	2 927	1 375	50.0
5	Tomato (1997)	17 048	71 041	27 472	38.7	714	69 525	18 990	27.3
5	Bean (1998)	16 716	1 282	512	39.9	545	1 304	374	28.7
5	Sunflower (1999)	22 928	2 034	414	20.4	707	2 012	202	10.0
58	Wheat (1996)	38 159	4 445	1 615	36.3	780	4 428	1 447	32.7
58	Tomato (1997)	12 137	81 549	37 153	45.6	566	82 583	19 448	23.5
58	Sunflower (1998)	25 694	2 800	543	19.4	780	3 048	399	13.1
58	Corn (1999)	25 020	13 211	4 347	32.9	780	13 912	2 392	17.2

† SD, standard deviation.
‡ CV, coefficient of variation.

higher degree of spatial homogeneity (Table 1). The causes of wheat yield variability in Year 1 were examined earlier by Plant et al. (1999), who attributed it primarily to aeration stress in Field 5 and weed competition in Field 58. These results corresponded with the observations of the cooperating grower, who was familiar with the trend toward heavier soil in the north end of Field 5.

Table 3 shows summary statistics of soil properties in both fields. Because silt content was roughly constant, sand and clay content more or less summed to a constant. The most variable component in each field was sand content, which showed a slightly greater range of variation in Field 5 than Field 58. A preliminary partial Mantel correlation test (Smouse et al., 1986) indicated no difference in the level of significance for the relation between sand and yield vs. the relation between clay and yield. Therefore, we elected to use sand content as an index of texture because of its greater variability. Trend analysis of soil sand content was performed using both the median polish (Cressie, 1991) and polynomial curve fit (Cliff and Ord, 1981) methods.

Fitting the percentage sand data with second-degree polynomial trend surfaces resulted in the equation percentage sand $= 43.3 - 0.0239x - 0.0849y - 0.000007x^2 + 0.000056xy + 0.000076y^2$ for Field 5 and the equation percentage sand $= 35.0 - 0.0443x + 0.0025y + 0.000036x^2 - 0.000003xy + 0.000050y^2$ for Field 58. The variables x and y measure position in the field in meters. Table 4 shows the sums of squares and p values for each of the spatial components for each field. Due to the spatial autocorrelation of the data, these p values are unreliable and generally underestimated. The results of the trend surface analysis, however, support the interpretation that the dominant median trend of variation in median soil sand content in Field 5 is in the north–south direction while in Field 58, it is in the east–west direction (i.e., the long axis of each field). Further support is provided by computing the extra sums of squares for sand content for the first-order regression model (not shown). For this reason, after computing the two-dimensional trend using median polish, we computed the median of these trend values over the short axis of each field. Figure 4 shows the medians across the short axes of the median polish values for each of the fields. The figure indicates that Field 5 has a slightly larger overall trend than Field 58 and that the trend in Field 5 tends to be concentrated at the south half of the field while that of Field 58 is spread across the entire field.

Figures 5 and 6 show the results of trend separation through median polish of the yield data. For ease of visualization, the figures show the median across the short axis of the trend surface, similar to Fig. 4. The close correspondence of the wheat yield trend in Field 5 to sand content and the parabolic nature of the tomato yield in this field are visually apparent. Because of the lack of data, trend surfaces for bean and sunflower in Field 5 were only computed in the north end of the field and show a slightly decreasing northward trend, as does wheat yield in this field. In Field 58, the pronounced effect of weeds on wheat yield variability is manifested in two large areas of low yield in the western half of the field (Fig. 2a), and there is a small increasing trend in sunflower and corn yields in the easterly direction.

Similar to the observation of Jaynes and Colvin (1997), we found that the residuals from the median polish of the yield data had a distribution significantly different from normal although unlike those reported by Jaynes and Colvin, they were not highly skewed. We employed two iterations of a winsorizing process on

Table 2. Comparison of average yields reported by the grower and those obtained through yield-monitoring (after aggregation procedure).

Field	Crop (year)	Humidity at harvest	Yield measured by the grower	Yield from monitored yield data	Difference
		%	kg/ha		%
5	Wheat (1996)	11	3 298	2 927	−11.2
5	Tomato (1997)	94	58 586	69 525	18.7
5	Bean (1998)	9	1 400	1 304	−6.9
5	Sunflower (1999)	9	1 572	2 012	28.0
58	Wheat (1996)	11	4 498	4 428	−1.6
58	Tomato (1997)	94	75 121	82 583	9.9
58	Sunflower (1998)	9	1 683	3 048	81.0
58	Corn (1999)	15	13 823	13 912	0.6

Table 3. Mean, maximum, minimum, standard deviation (SD), and coefficient of variation (CV) of soil variables.

Field	Variable	No. of points	Avg.	Max.	Min.	SD	CV
							%
5	pH	86	5.8	6.0	5.6	0.1	2.1
5	Sand	86	25.3	44.6	17.3	6.6	26.0
5	Clay	86	37.3	41.5	23.2	6.3	16.8
5	Silt	86	37.5	42.8	30.8	2.4	6.3
5	SOM†	86	2.0	2.5	0.8	0.3	17.3
58	pH	78	5.6	5.9	5.4	0.1	1.6
58	Sand	78	27.4	45.5	17.3	5.9	21.7
58	Clay	78	33.9	46.1	23.2	4.5	13.3
58	Silt	78	38.7	48.1	25.1	4.0	10.3
58	SOM	78	1.5	1.7	1.1	0.1	8.0

† SOM, soil organic matter.

those residuals more than 1.5 standard deviations from the mean. This brought the data closer to a normal distribution as indicated by probability plots (data not shown). In general, a Kolmogorov–Smirnov test still indicated a distribution significantly different from normal ($p < 0.05$), which is not surprising given the large number of data values. Directional variogram analysis indicated no substantial anisotropy (data not shown).

One of the methods we used to determine the temporal structure of the data was correlation analysis. Again reflecting the observations of Jaynes and Colvin (1997), we found that the yield data themselves were highly skewed and that no suitable transformation could be found to normalize them. Therefore, we used the Spearman rank correlation coefficient to measure pairwise correlation between crops within each field. In several cases in Field 5, the regression of yield of one crop against that of another indicated a significant quadratic term. For example, as reported by Pettygrove et al. (1999), the regression of tomato yield against wheat yield is parabolic (cf. Fig. 5). In these cases, we transformed the data by estimating the peak b of the parabola obtained from a quadratic regression of crop y against crop x and then computing the rank correlation coefficient between y and $(x - b)^2$ (Neter et al., 1996, p. 126). Table 5 shows the rank correlation coefficients for each pair of crops in each field. We applied the correction procedure of Clifford et al. (1989) to estimate the effective sample size in the presence of spatial autocorrelation of the data. This procedure was developed for use with the Pearson product moment coefficient rather than the Spearman coefficient. However, the asymptotic normality of the distribution of the Spearman coefficient (Kendall and Gibbons, 1990) lends support to the idea of using this adjustment to estimate the effective sample size for the Spearman coefficient as well. Haining (1990) also used the correction of Clifford et al. (1989) to study the effect of spatial autocorrelation on the significance of the Spearman coefficient. The p values computed in Table 5 are based on the W statistic of Clifford et al. (1989). No significance levels are attached to these values because they should be viewed only as rough approximations of the true p value. The correlation coefficient between sunflower and corn may be spurious because a scatter plot indicated that its large value was due to a relatively small number of influence points (data not shown).

Figure 7 shows the sumscore and CV of each field. There is a tendency of spatial autocorrelation between those areas with similar yield performance in Field 5. The visually apparent spatial clustering was supported by the result of Moran's I and Geary's c tests (Lee

Table 4. Sums of squares and p values for components of the ordinary least-squares curve fit of a second-degree polynomial to the percentage sand data for Fields 5 and 58.†

	Field 5		Field 58	
Component	Sum of squares	p	Sum of squares	p
x	85.1	0.02	1326.0	<0.01
y	1644.9	<0.01	238.8	0.87
x^2	0.5	0.80	218.8	<0.01
xy	250.8	<0.01	0.3	0.88
y^2	1044.8	<0.01	16.7	0.26

† The equation has the form percentage sand = $a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2$, where the values of the a_i are given in the text. The variables x and y measure position in the field in meters.

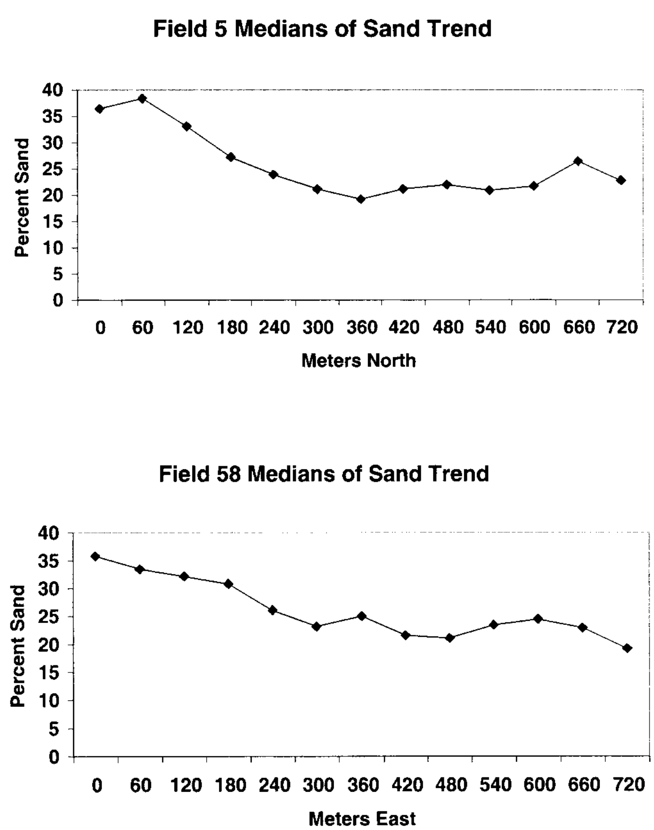

Fig. 4. Medians over the short axis of the field of trend surfaces obtained from median polish for each field.

and Wong, 2001) of autocorrelation. The z value of the Moran's I was 61.9 ($p < 0.001$). The estimated Geary's c index (0.636) also supports the conclusion that the sumscore values are positively autocorrelated. In Field 58, there is also a high visual spatial autocorrelation with the middle parts having higher sumscores (Fig. 7). The visual estimation was again supported by the result of a Moran's I (z value = 46.1, $p < 0.001$) and Geary's c (0.778) tests of autocorrelation, which showed that the distribution of the values of points was clustered. The spatial and temporal patterns are much more homogeneous in this case, except where they are altered by a weed infestation during the first year in the southwestern part of the field (Plant et al., 1999). This situation

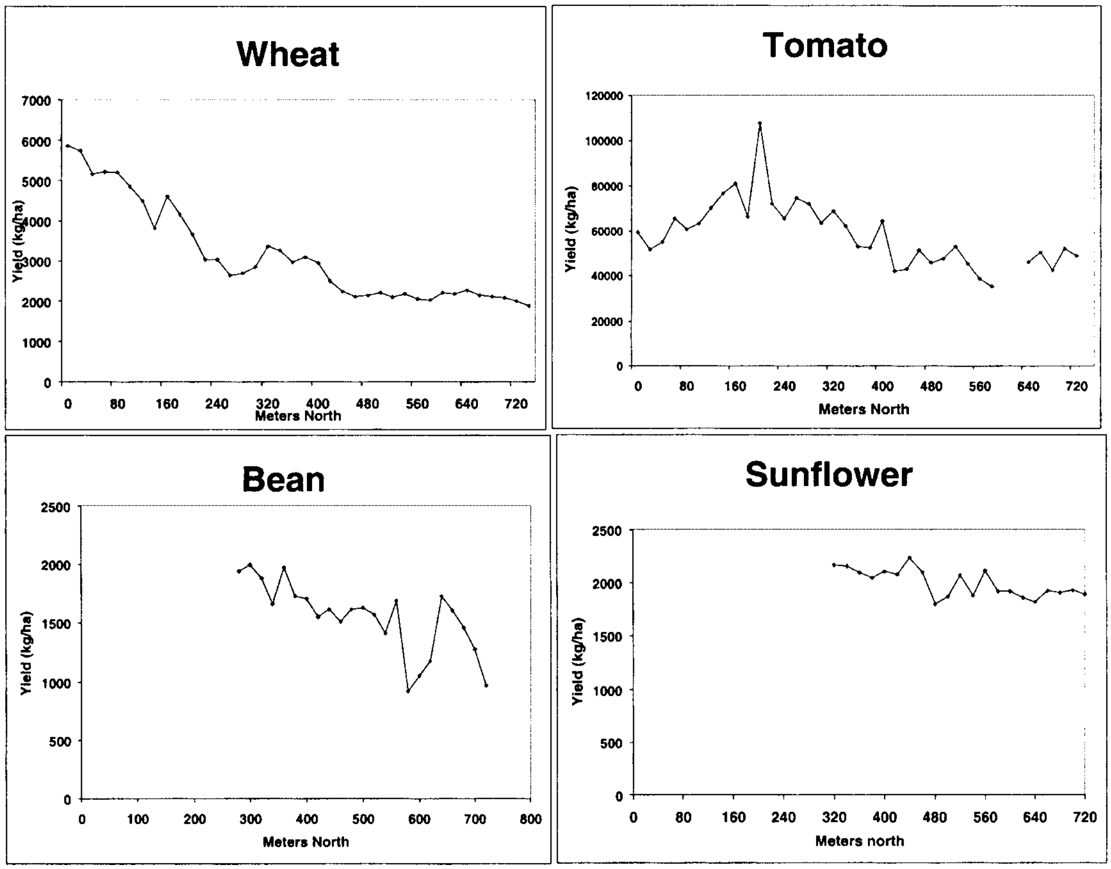

Fig. 5. Medians of the Field 5 yield trends obtained through median polish.

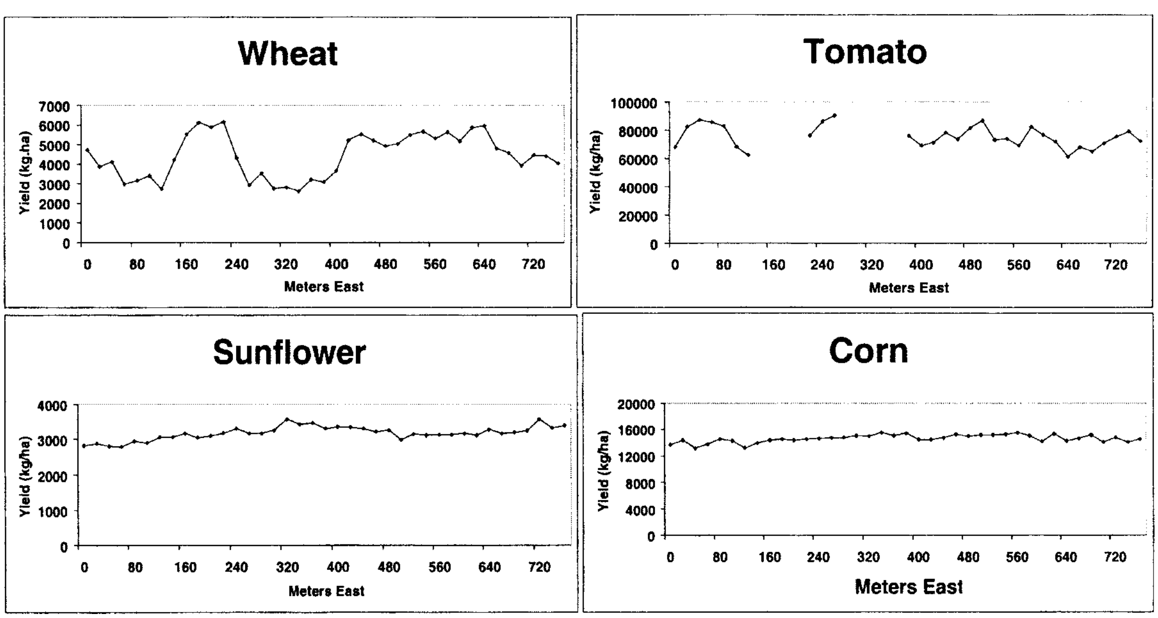

Fig. 6. Medians of the Field 58 yield trends obtained through median polish.

is also reflected in the CV map (Fig. 7) where besides a clear border effect, high variability values are found in the southwest side of the field. The effect was that only 70% of the area in this field had a temporal variability lower than the average spatial variability of the four crop-years (CV = 21.6%, see Table 1). In Field 5, most of the higher CV values (over 34%) are located close to the boundaries due apparently to border effects. In the interior of the field, 65% of the area where the CV was calculated had a value lower than 17%, which can be compared with the average of the spatial variability of the four crops (CV = 29.0%). In other words, 88% of the area in the field had a temporal variability lower than the average spatial variability of each crop-year (CV = 29.0%). Comparable analysis of the data in Field 58 is complicated by the existence of a dominant weed infestation in the wheat crop. In this field, 70% of the area had a temporal variability less than the average spatial variability of the four crop-years (CV = 21.6%).

Table 5. Spearman rank correlation coefficients, r_s, of crop yields. N = number of data values, $\hat{M}$ = effective sample size as computed by the method of Clifford et al. (1989), W = Clifford et al. (1989) W statistic, and p = estimated p value of W statistic based on standard normal distribution. Crop abbreviations: W = wheat, T = tomato, B = bean, S = sunflower, and C = corn. Relationships between other crops and wheat in Field 5 were parabolic so that rank correlation coefficients were computed between that crop and the transformed variable $(W - b)^2$, where b is the value of W at which the parabola attains its peak.

Field	Crop pair	r_s	N	$\hat{M}$	W	p
5	$(W - b)^2$, T	−0.47	702	15.8	−1.80	0.07
5	$(W - b)^2$, B	−0.51	542	26.5	−2.57	0.01
5	$(W - b)^2$, S	−0.44	563	52.7	−3.16	<0.01
5	T, B	0.38	473	34.3	2.19	<0.01
5	T, S	0.12	495	34.2	0.71	0.5
5	B, S	0.32	492	55.9	2.37	0.02
58	W, T	−0.06	566	113.9	−0.7	0.5
58	W, S	0.07	780	50.3	0.53	0.6
58	W, C	−0.03	780	71.4	−0.27	0.8
58	T, S	0.07	566	118.7	0.72	0.5
58	T, C	−0.10	566	80.7	−0.89	0.4
58	S, C	0.42	780	68.8	3.46	<0.01

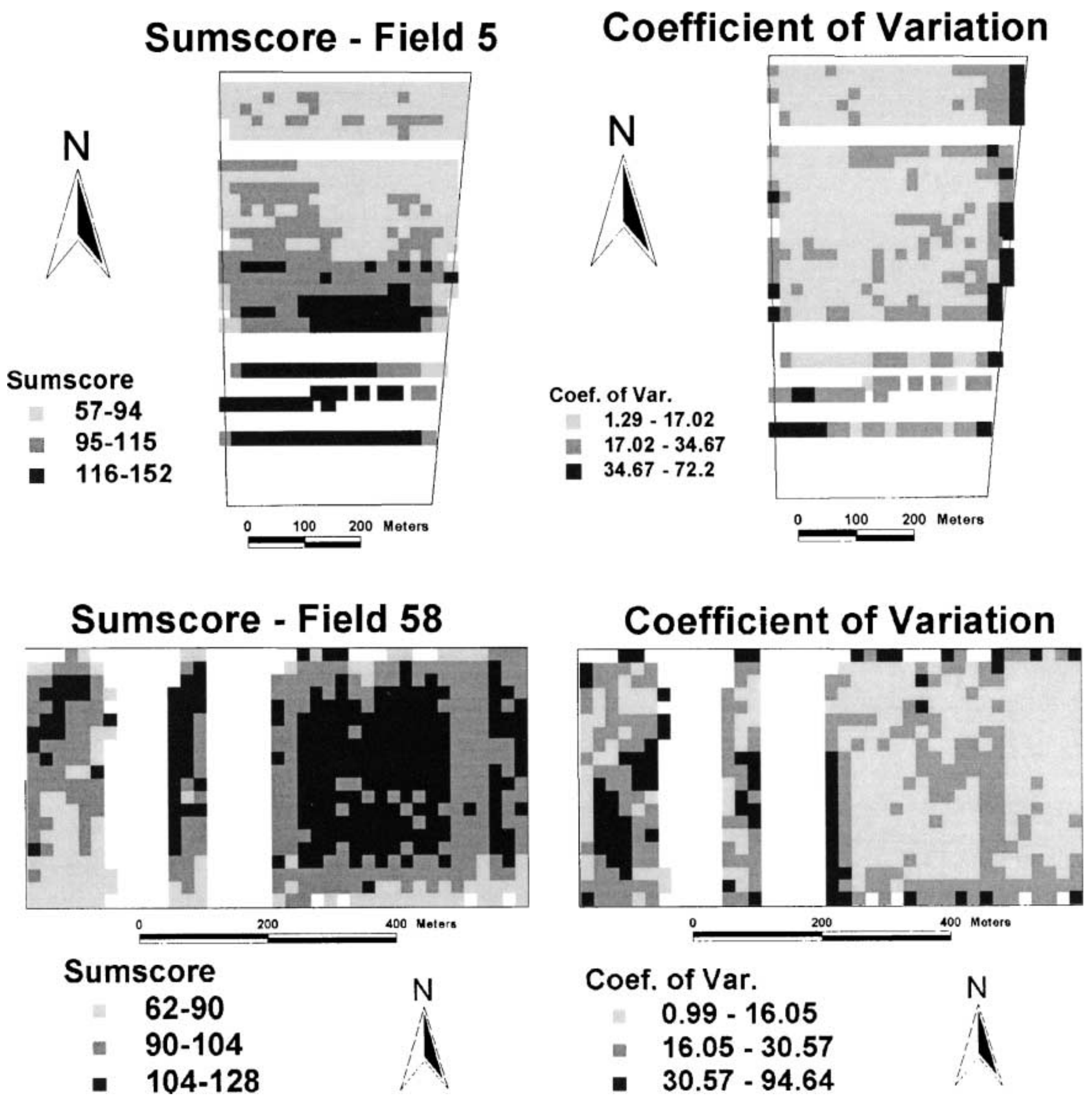

Fig. 7. Sumscore and coefficient of variation of standardized yields in each field. Each yield was standardized to its own mean and variance. The line shows the border of the field, and different shades represent each index where data were available for the 4-yr period. A sumscore of 100 represents the standardized mean.

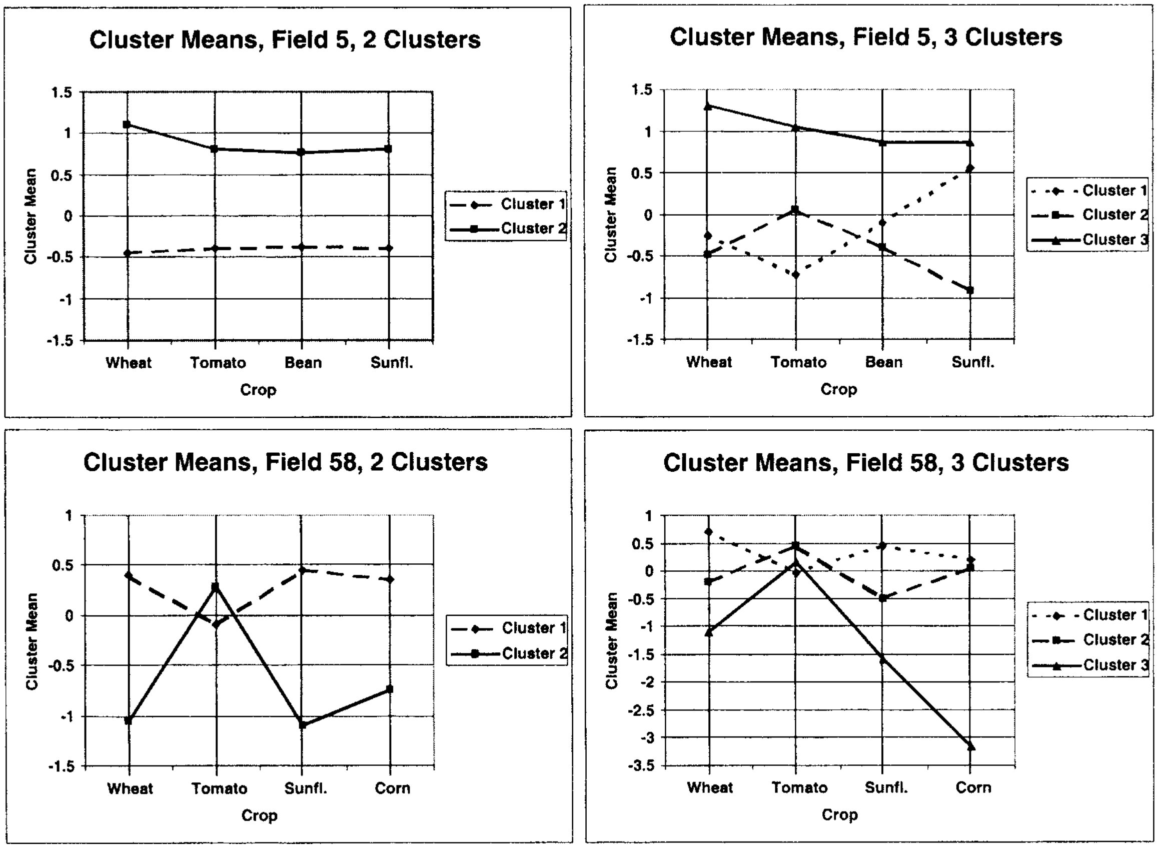

Fig. 8. Plots of the mean values of clusters of standardized yield in Fields 5 and 58 for each crop obtained by *K*-means clustering for $K = 2$ and $K = 3$.

In summary, Field 5 presents more spatial variability within each year (average CV = 29.0%) compared with Field 58 (average CV = 21.6%), but across years, Field 5 is slightly more stable (average CV = 16.7%) compared with Field 58 (average CV = 18.5%).

Figure 8 shows the results of the *K*-means cluster analysis of the 4 yr of standardized yields for Fields 5 and 58 using two and three clusters, respectively. The cluster analysis identified two groups that are respectively consistently high or consistently low in yield, and there is a decreasing dispersion from the first to the last year, graphically shown as the means of each cluster moving closer to the average of that particular year. In the three-cluster division, the lower-yielding cluster splits into two, one of which is consistently low in yield and one of which varies from low to high. Figure 9 shows the spatial arrangement of the clusters in each field. Visual inspection clearly indicates that the clusters are spatially correlated. This is confirmed by the significance of spatial contiguity of the clusters using Moran's *I* statistic, based on nearest neighbors. The *z* value of this statistic for the three clusters was 61.9 ($p < 0.001$). In Field 58, the cluster analysis indicated that there are two natural clusters, one of which consists of consistently high-yielding areas and one of which is characterized by lower yields, except in the tomato year. The three-cluster division separates a border effect from the weed infestation area that occurred during the wheat year. The *z* value of the Moran's *I* statistic for the three-cluster set was 46.1 ($p < 0.001$), again confirming a visually obvious spatial autocorrelation.

Correlation and forward stepwise multiple regression analyses were performed between each crop-yield data set and sand, silt, and clay content; soil organic matter (SOM); and pH (Table 6). There is an evident high level of multicollinearity among predictor variables. In Field 5, sand is strongly covariant with clay and also with SOM. The highest coefficient of correlation with yield determined which of the three variables (sand, clay, or SOM) was included in the stepwise analysis. In Field 58, silt content also showed high correlation with sand and SOM. Therefore, the same criterion was followed, and silt and clay were included together in the model only when one of them had the highest coefficient of correlation. The regression coefficients and variables of the models obtained using forward stepwise analysis and the result from the standard multivariate analyses are contained in Table 7. In Field 5, the best fit was obtained for wheat where a model including only clay explained 62% of the yield variation. For tomato and bean, the models explained close to half of the variation in yield, with the reduced models obtained through stepwise analysis being very close to the models including all of the variables that had no multicollinearity problems. The low R^2 obtained for sunflower yield model is probably due to the low spatial variability of this oilseed crop in Field 5 (see Table 1). In Field 58, the scenario is quite different: The spatial variability of the crop

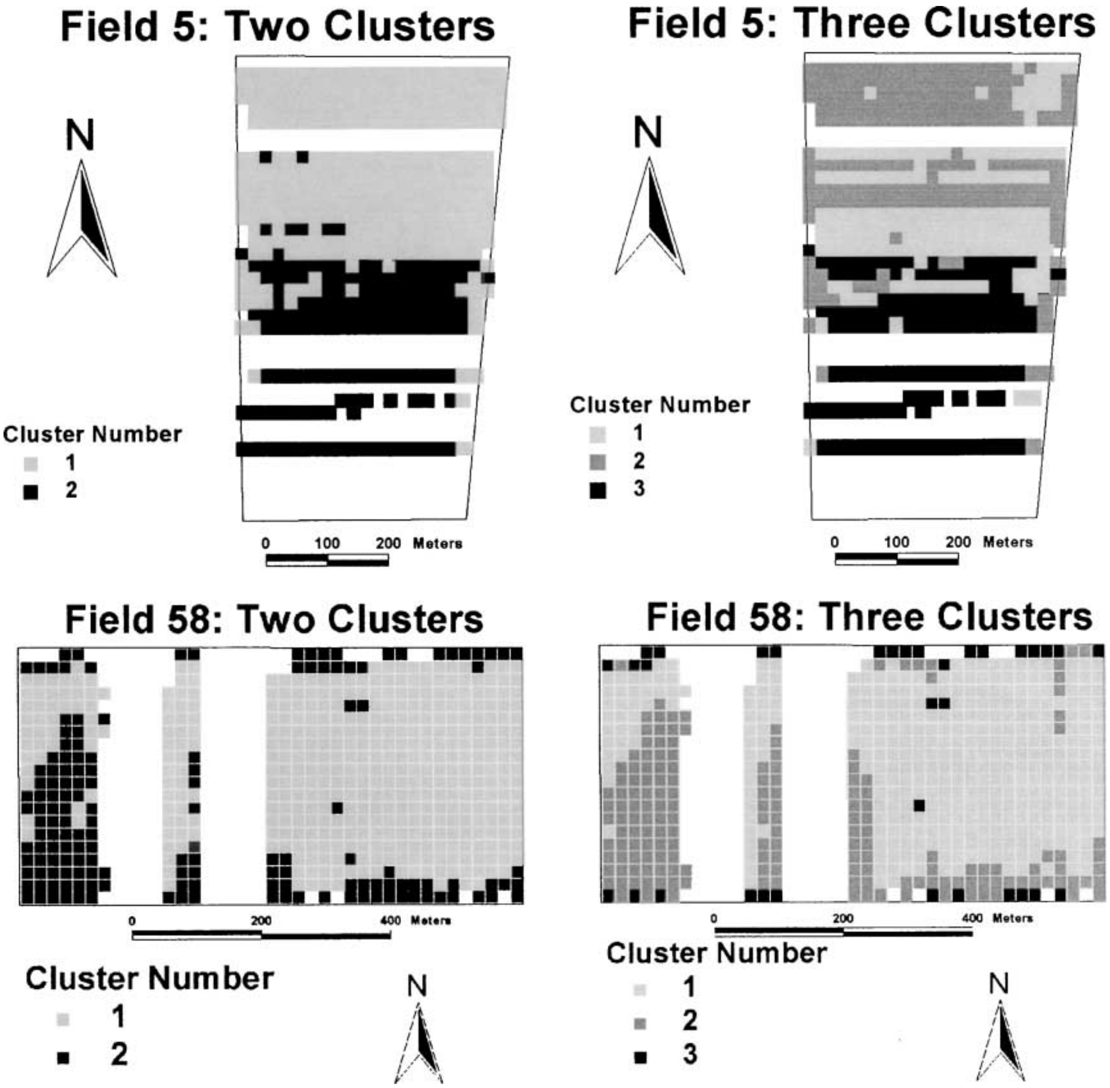

Fig. 9. Distribution of cluster sets of the two- and three-means cluster analysis in Fields 5 and 58. The line shows the border of the field, and different shades represent each location where data were available for the 4-yr period.

yields was not explained very well by the measured soil variables. Again, wheat had the highest coefficient of determination, $R^2 = 0.25$ for the standard regression method, and was the only crop where it was possible to generate a model through forward stepwise regression ($R^2 = 0.22$, including only sand content).

Regression analyses were also performed between the sumscore and soil characteristics (Table 7). When sumscore was used as the response variable in Field 5, the model that includes pH, silt, and SOM explained 56% of the variation. The reduced model included only SOM, and it still explained 44% of the 4-yr average yield (sumscore). Because the sumscore could be calculated only where data were available for the four crops, the number of observations is lower (43 points) and basically represents the bean year, which was the year with the least complete data set (Fig. 2). The regression model for sumscore in Field 58 ($R^2 = 0.16$) was obtained using the 60 soil samples that matched locations where yield data for all 4 yr was available. The coefficient of determination for sumscore was not high but was higher than for tomato, sunflower, and corn yield models.

DISCUSSION

In general, using the net truck weights as the reference, the yield monitor overestimated crop yields, except for bean and wheat, where the latter crop was underestimated in both fields although in Field 58, the

Table 6. Matrix of correlation coefficients between soil properties measured in both fields in 1995–1996 season.

Field	Variable	pH	Sand	Clay	Silt
5	Sand	−0.02			
5	Clay	0.16	−0.93		
5	Silt	−0.36	−0.30	−0.06	
5	SOM†	−0.49	−0.65	0.53	0.42
58	Sand	−0.04			
58	Clay	0.12	−0.74		
58	Silt	−0.08	−0.66	−0.02	
58	SOM	−0.18	−0.69	0.50	0.47

† SOM, soil organic matter.

Table 7. Variables and regression coefficients of models obtained through forward stepwise and standard regression analyses to explain crop yields.

		Forward stepwise			Standard regression		
Field	Crop	Variables	N	R^2	Variables	N	R^2
5	Wheat	Clay	86	0.62	pH, silt, clay	86	0.63
5	Tomato	Clay, silt	86	0.46	pH, silt, clay	72	0.45
5	Bean	SOM†	56	0.48	pH, silt, SOM	56	0.49
5	Sunflower	None	74	–	pH, silt, clay	74	0.15
5	Sumscore	SOM	43	0.44	pH, silt, SOM	43	0.56
58	Wheat	Sand	78	0.22	pH, sand	78	0.25
58	Tomato	None	60	–	pH, silt, clay	60	0.04
58	Sunflower	None	78	–	pH, silt, clay	78	0.00
58	Corn	None	78	–	pH, SOM	78	0.09
58	Sumscore	None	60	–	pH, silt, clay	60	0.16

† SOM, soil organic matter.

difference was negligible (1.6%). As a standard, yield-monitoring should be within 5% of the actual yield (Pierce et al., 1997). In our case, only two out of eight crop-years were within that range, sunflower being the crop that was least accurately monitored (28 and 81% difference in Field 5 and Field 58, respectively). This indicates that yield monitoring, when done commercially, does not necessarily provide an accurate estimation of total yield. Our analysis is based on the assumption (which we cannot prove) that the spatial pattern of yield indicated by the yield map reflects that of the actual field. The yield maps indicate a high level of short-range variability, at least some of which may be due to yield monitor error.

These laser-leveled, fully irrigated fields both showed a relatively high level of temporal structure. This is in contrast to the situation commonly found in the rainfed, unleveled systems of the Midwest, in which temporal variation often dominates spatial variation (Eghball and Varvel, 1997). The two fields studied represent an interesting contrast in that one is characterized by a strong spatial trend in texture over a portion of its extent (and virtually no trend over the other portion) while the other field is characterized by a milder trend over its entire range. Intuitively, one would expect a higher level of temporal structure in the area of greater spatial structure. In other words, one would expect temporal stability to be connected with a high level of spatial variability in soil texture. To some extent, this is born out by the results. The plots of yield and sand content trends obtained by median polish indicate the strongest relationship, albeit still characterized by a high level of variability, in the region of greatest soil texture trend. Visual comparison of the trends in yield and sand content also indicates a considerable deviation, and it is clear that a biological interpretation of the two-component model, as suggested in the introduction, requires further analysis of other cases.

Consistent with the results of other investigators (Jaynes and Colvin, 1997; Lamb et al., 1996, 1997; Sadler et al., 1995, 1998), the correlation coefficient between yields in different years is generally low, even in cases of seemingly obvious dominance of a soil trend. However, after standardizing and then averaging yields across time, the resulting values (which we called the sumscore) indicated both a visual and statistically significant spatial autocorrelation. Cluster analysis also seemed to give a reasonable summary of field trends. Cluster analysis was also used to good effect by Lark and Stafford (1997) and by Stafford et al. (1999) to describe yield trends. These authors used fuzzy clustering, which is advocated for soil studies by Burrough (1989). We used the standard, or *crisp*, version of clustering. While perhaps not providing the same level of detail as fuzzy clustering, ordinary clustering is considerably simpler to implement and appears to have much to recommend it.

We argue that in some sense, cluster analysis may be a more natural indicator of spatiotemporal pattern than correlation and regression analysis. Both methods begin with an aspatial analysis in which data are studied cell by cell with no consideration of the spatial arrangements of the cells. Both methods can be subjected to hypothesis testing but in different ways. Correlation analysis makes no use of spatial structure but rather considers the relation between yields in each individual land area (or grid cell) separately. Spatial autocorrelation between grid cells acts to confound the analysis and reduce the level of significance of the test. Cluster analysis, on the other hand, first sorts the grid cells in an aspatial manner into groups on an ordinal scale. Hypothesis testing can then be used to determine whether these groups have a significant spatial structure (Cliff and Ord, 1981). In some ways, this may be a closer approximation to the way a farmer would think about the field. The farmer might not think of individual small areas and how they relate to each other on an interannual basis but rather might classify areas of the field as good yielding, poor yielding, etc., and then consider the spatial arrangement of these areas. Moreover, if the cluster groups do have substantial spatial structure, they may provide a first step to the delineation of management zones for site-specific farming.

Our results support the idea that data collection at the high level of resolution provided by devices such as yield monitors (as well as technologies such as remote sensing and bulk electrical conductivity measurement) has a scientific utility beyond developing a SSM strategy for a particular field. It permits the testing of scientific hypotheses in commercial fields using a method other than replicated field trials. The data may be studied using observational methods more commonly employed by ecologists and epidemiologists. Such methods may open the way to a greater ability to carry out meaningful agronomic research in commercial fields.

ACKNOWLEDGMENTS

This research was supported in part by the Fertilizer Research and Education Program of the California Department of Food and Agriculture. The authors are grateful to Jiayou Deng and Julie Young for expert technical support.

REFERENCES

Blackmore, S. 2000. The interpretation of trends from yield maps. Comput. Electron. Agric. 26:37–52.

Burrough, P.A. 1989. Fuzzy mathematical methods for soil survey and land evaluation. J. Soil Sci. 40:477–492.

Cliff, A.D., and J.K. Ord. 1981. Spatial processes: Models and applications. Pion, London.

Clifford, P., S. Richardson, and D. Hemon. 1989. Assessing the significance of the correlation between two spatial processes. Biometrics 45:123–134.

Colvin, T.S., D.L. Karlen, J.R. Ambuel, and F. Perez-Muñoz. 1995. Yield monitoring for mapping. p. 3–14. *In* P.C. Robert, R.H. Rust, and W.E. Larson (ed.) Proc. Int. Conf. on Precision Agric., 2nd, Bloomington and Minneapolis, MN. 27–30 Mar. 1994. ASA, CSSA, and SSSA, Madison, WI.

Cressie, N.A.C. 1991. Statistics for spatial data. John Wiley & Sons, New York.

Eghball, B., and G.E. Varvel. 1997. Fractal analysis of temporal yield variability of crop sequences: Implications for site-specific management. Agron. J. 89:851–855.

Emerson, J.D., and D.C. Hoaglin. 1983. Analysis of two-way tables by medians. p. 166–210. *In* D.C. Hoaglin, F. Mosteller, and J.W. Tukey (ed.) Understanding robust and exploratory data analysis. John Wiley & Sons, New York.

Haining, R. 1990. Spatial data analysis in the social and environmental sciences. Cambridge Univ. Press, Cambridge, UK.

Huggins, D.R., and R.D. Alderfer. 1995. Yield variability within a long-term corn management study: Implications for precision farming. p. 397–416. *In* P.C. Robert, R.H. Rust, and W.E. Larson (ed.) Proc. Int. Conf. on Precision Agric., 3rd, Minneapolis, MN. 23–26 June 1996. ASA, CSSA, and SSSA, Madison, WI.

Isaaks, E.H., and R.M. Srivastava. 1989. An introduction to applied geostatistics. Oxford Univ. Press, Oxford, UK.

Jain, A.K., and R.C. Dubes. 1984. Algorithms for clustering data. Prentice-Hall, Upper Saddle River, NJ.

Jaynes, D.B., and T.S. Colvin. 1997. Spatiotemporal variability of corn and soybean yield. Agron. J. 89:30–37.

Kendall, M., and J.D. Gibbons. 1990. Rank correlation methods. Edward Arnold, London.

Lamb, J.A., J.L. Anderson, G.W. Rehm, and R.H. Dowdy. 1996. Grain yield stability in continuous corn and corn–soybean systems on a sandy landscape. p. 419–424. *In* P.C. Robert, R.H. Rust, and W.E. Larson (ed.) Proc. Int. Conf. on Precision Agric., 3rd, Minneapolis, MN. 23–26 June 1996. ASA, CSSA, and SSSA, Madison, WI.

Lamb, J.A., R.H. Dowdy, J.L. Anderson, and G.W. Rehm. 1997. Spatial and temporal stability of corn grain yields. J. Prod. Agric. 10:410–414.

Lark, R.M., and J.V. Stafford. 1997. Classification as a first step in the interpretation of temporal and spatial variability of crop yield. Ann. Appl. Biol. 130:111–121.

Lark, R.M., J.V. Stafford, and M.A. Froment. 1997. Exploratory analysis of yield maps of combinable crops. p. 887–894. *In* J.V. Stafford (ed.) Precision agriculture '97. BIOS Sci. Publ., Oxford, UK.

Lee, J., and D.W.S. Wong. 2001. Statistical analysis with ArcView GIS. John Wiley & Sons, New York.

Neter, J., M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. 1996. Applied linear statistical models. Irwin, Chicago.

Pannatier, Y. 1996. Variowin: Software for spatial data analysis in 2D. Springer-Verlag, New York.

Pettygrove, G.S., S.K. Upadhyaya, M.G. Pelletier, T.K. Hartz, R.E. Plant, and R.F. Denison. 1999. Tomato yield–color infrared photograph relationships. p. 1483–1492. *In* Proc. Int. Conf. on Precision Agric., 4th, St. Paul, MN. 19–22 July 1998. ASA, CSSA, and SSSA, Madison, WI.

Pierce, F.J., N.W. Anderson, T.S. Colvin, J.K. Schueller, D.S. Humburg, and N.B. McLaughlin. 1997. Yield mapping. p. 211–244. *In* F.J. Pierce and E.J. Sadler (ed.) The state of site-specific management for agriculture. ASA, CSSA, and SSSA, Madison, WI.

Plant, R.E., A. Mermer, G.S. Pettygrove, M.P. Vayssieres, J.A. Young, R.O. Miller, L.F. Lackson, R.F. Denison, and K. Phelps. 1999. Factors underlying grain yield spatial variability in three irrigated wheat fields. Trans. ASAE 42:1187–1202.

Sadler, E.J., W.J. Busscher, P.J. Bauer, and D.L. Karlen. 1998. Spatial scale requirements for precision farming: A case study in the southeastern USA. Agron. J. 90:191–197.

Sadler, E.J., W.J. Busscher, and D.L. Karlen. 1995. Site-specific yield histories on a SE coastal plain field. p. 153–166. *In* P.C. Robert, R.H. Rust, and W.E. Larson (ed.) Site-specific management for agricultural systems. Proc. Int. Conf., 2nd, Minneapolis, MN. 27–30 Mar. 1994. ASA, CSSA, and SSSA, Madison, WI.

Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matric correspondence. Syst. Zool. 35:627–632.

Stafford, J.V., R.M. Lark, and H.C. Bolam. 1999. Using yield maps to regionalize fields into potential management units. p. 225–238. *In* Proc. Int. Conf. on Precision Agric., 4th, St. Paul, MN. 19–22 July 1998. ASA, CSSA, and SSSA, Madison, WI.

Steel, R.G.D., and J.H. Torrie. 1980. Principles and procedures of statistics: A biometrical approach. McGraw-Hill, New York.