THE SPATIAL VARIATION OF SEED POPULATION OF WHITE CORN BASED ON SOIL MOISTURE HOLDING CAPACITY

By

STEPHEN R. JOHNSON, JR.

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1995

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 1997

THE SPATIAL VARIATION OF SEED POPULATION OF

WHITE CORN BASED ON SOIL MOISTURE

HOLDING CAPACITY

Thesis Approved:

David A. Waits

Thesis Adviser

Stephen W. Tweedie

Jonathan C. Comer

Wayne B. Powell

Dean of the Graduate College

ACKNOWLEDGEMENTS

I wish to express gratitude to my thesis adviser, Dr. David Waits, for the encouragement, support, and patience that he has provided during the completion of this study. Also, I am indebted to Dr. Stephen Tweedie and my co-workers Mr. Daniel Umbach and Mr. Shawn Fitzgerald for their insights and ideas which helped edify the project. I must also express appreciation to the Department of Geography and NASA Oklahoma Space Grant funds for supporting the research.

A special thanks must be given to Mr. Wayne Gray, and all of Gray Farms, Incorporated, for their willingness to undertake this project; without a progressive and visionary agriculturalist, this study would not have been possible. His support, hospitality, and enthusiasm for the study is greatly appreciated.

I would finally like to extend my gratitude to my parents, for their support through all the necessary stages that preceed an accomplishment such as this. The seeds of curiosity must be planted in a person and supported for years in order for intellectual curiosity to finally bear fruit.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

LIST OF FIGURES

Figure Page

NOMENCLATURE

K	Soil potassium content, in parts per million
P	Soil phosphorus content, in parts per million
OM	Percentage soil organic matter composition
CEC	Cation exchange capacity index
Ca	Soil calcium content, in parts per million
Mg	Soil magnesium content, in parts per million
pH	Soil pH
SPI	Soil productivity index
GIS	Geographic Information System
GPS	Global Positioning System
$Z(x)$	Value of regionalized variable at location x
$Z(x + h)$	Value of regionalized variable at distance from x h
$VAR(h)$	Variance between two points separated by distance h
$\gamma(h)$	Semi-variance between two points separated by distance h
RMS	Root mean squared error
ANOVA	Analysis of variance
$\mu(x)$	Mean value of variable x
$\sigma(x)$	Variance of variable x

STD	Standard deviation
R^2	Coefficient of determination
LSD	Least squares difference
ΔD	Change in distance

CHAPTER I
INTRODUCTION

Spatial Variability and Agriculture

Since the advent of mechanized agriculture, the possibility of spatially variable crop management has intrigued both farmers and the academic agronomic community. The presence of within-field differences in crop growth parameters has long been suspected, but no means have traditionally been available to quantify that variability and alter management practices accordingly. Such a variable management paradigm would emphasize within-field differences in crop production capabilities and seek to capitalize on those areas of the field which are the most productive and apply crop inputs only where they are needed.

However, the ability to implement such systems has only recently become a reality, due to the rapid growth in several complementary technologies. These include the launching of the NavStar constellation of Global Positioning System (GPS) satellites which enable practitioners to accurately record positions on the earth's surface, the ability to acquire and process multi-spectral satellite imagery for the assessment of crop conditions, and the movement of Geographic Information Systems (GIS) software from the domain of large organizational computer systems to desktop microcomputers. All of these factors working in tandem have made it possible for agricultural producers to implement sub-field level management schemes.

This development opens the door to incorporate the results of geographic (site-specific) and agronomic (context-specific) research to optimize soil performance at every location in a farm field. The distinction between the two aforementioned approaches arises because of the differences between the holistic approach of geography, which seeks to incorporate various types of data to create a narrative describing the uniqueness of specific locations, and the highly controlled approach of agronomy, which seeks to understand how crops behave given specific parameters in a particular location.

The discipline of geography provides a good backdrop for the study of the impact of human land management on the physical surroundings; indeed, the geographer Richard Chorley noted that "Ecosystem management is therefore primarily associated with the maximization of existing productivity, the minimization of wastage by the adoption of suitable harvesting strategy, pest control, or the scientific cropping of native flora and fauna" (Chorley, 1973). Peter Gould, a quantitative geographer whose influence on the development of spatial theory cannot be overstated, pointed to the need for applied geography in the management of data complexity, reflecting that "With an increased willingness to tackle complicated systems and structures came the necessity to distance oneself a bit from all the complexity, and try to simplify (to model), abstract, and theorize" (Gould, 1985). Thus, geographic inquiry within the context of agronomy, which also deals intimately with man/land relationships, can provide valuable insight with the multitude of sub-field level data that are now available.

Understanding how to manage crops given certain conditions and ascertaining where in a field those conditions exist is precisely why geographic and agronomic techniques are being married in a burgeoning new sub-discipline which has been referred to as "farming by the foot" or "precision farming" (Reichenberger and Russnogle, 1989; Sudduth et al., 1991). This integration of two traditional academic studies seeks to enable agricultural producers to make environmentally responsible decisions while simultaneously accruing the economic benefits associated with increased crop productivity and the reduced application of crop inputs.

One problem that innovative agricultural producers are faced with is that of assessing the success of a spatially variable management technique. The nature of precision farming is such that its implementation alters the conditions of crop production to various extents within the field; thus, any yield which results from such a practice must be compared to crop performance on the same plot of land managed homogeneously, i.e., with no intended spatial variability. Also, if two different methods of implementing spatial variability are applied, they must be compared not only with traditional methods, but also with each other if an objective assessment of the results is to be made.

The former comparison is flawed because although it considers a common geography, the field contains different management practices; the second comparison presents an even further difficulty in that it compares both different management practices and different geographies. Thus, any assessment of the performance of a spatially variable management practice must include sufficient data to compare yield differences with such discrepancies removed from the data. Normalization for the differences between the two scenarios under scrutiny must be included in the study if an accurate assessment of the success of one practice over another is to be made.

The Statement of Purpose

The specific problem this study addresses is the difficulty in comparing different spatially variable farm management practices with each other. The ultimate goal is to assess the performance of two management practices based on their respective abilities to improve crop yield.

The Requirements of the Study

The specific objective of this study is to construct a database which catalogues two agricultural fields supporting the same crop, to which different spatially-variable management techniques are applied. These data are augmented by control zones within each field to which traditional aspatial methods have been applied. A geographic approach is then taken to quantitatively assess the differences both between the two spatially variable management schemes and between the application of aspatial and spatially variable management practices. The field of spatial statistics is heavily employed in an effort to examine the impact of spatially variable management. The following null hypotheses are investigated:

1) Ho: There is no difference between crop yield in portions of a field managed spatially and aspatially
2) Ho: There is no difference between crop yield between two different fields to which unique spatial management regimes were applied

CHAPTER II

LITERATURE REVIEW

Spatial Variability and Agronomy

Concurrent with technological advances and the emergence of commercially available variable-rate crop input application equipment, the field of agronomy has produced studies which have analyzed the amount of within-field variability in crop growth parameters under a variety of conditions (Forcella, 1993) (Lark and Stafford, 1996). These studies have served to address the gap between the intuitive knowledge of within-field differences and the actual quantitative application of management strategies. Many of these studies address how to assess the within-field variability of some specific criterion for the application of a particular management technique; this has been demonstrated repeatedly in soil fertilizer applications and soil-specific tillage operations (Voorhees et al., 1993; Ferguson et al., 1996).

Older studies in agronomy are not devoid of acknowledgment of within-field differences; rather, variability is simply cast in the light of available management strategies. For example, in a 1947 study of fertilizer applications for various crops in Oklahoma, Harper discusses at length the relationship between soil acidity and phosphorus availability on various soil types. While it is noted in the study that different soil structures will behave differently under conditions of acidity, fertilizer application recommendations are given on a crop-by-crop basis (Harper, 1947a). The underlying assumption is that while a field may encompass several soil types, it is not feasible to address them individually, and thus the field should be treated homogeneously and managed according to a specific crop's requirements.

Soil sampling, which seeks to understand soil conditions within a fieldhas typically not addressed within-field differences, but has been conducted on a field-by-field basis for fertilizer recommendations. In a 1978 field guide for the sampling and description of soils, the author notes that "Soil samples for laboratory analysis often weigh a few grams, a small portion of the whole profile, and we use such samples to represent many hectares..." (Hodgson, 1978). The book does not elaborate on the spatial variability of soils; although many descriptions are given of various ways to determine the properties of individual soil profiles, no proposal is given of the spatial resolution at which soil properties may vary.

More recent studies, however, have appeared which highlight within-field differences in crop growth potential. Brukler et al. present an analysis of nitrate variability in both space and time for Mediterranean salad crops (Brukler, 1996). The study noted that the range of spatial dependence of nitrate leaching potential was very small, and demonstrated quantitatively how the percent uncertainty around an estimate could be significantly decreased by increasing the spatial resolution of sampling. Blackmer et al. outline a strategy for incorporating aerial photography to assess within-field differences in nitrogen availability based on known relationships between crop response to nitrogen and crop canopy reflectance of light (Blackmer et al., 1996).

Crop Population Density

Corn planting population prior to the development of hybrids and chemical fertilizers was restricted to usually about one kernel every 40 inches in a row, which translated into about 10,000 plants per acre (Aldrich and Leng, 1965). However, densities in excess of 20,000 plants per acre are common today. The influence of population density on yield depends largely on planting data and available moisture. Earlier planting dates typically favor higher population densities because of increased mortality and greater moisture availability (Aldrich and Leng, 1965).

The effect of population density during periods of drought has been investigated by agronomists under several scenarios. Norwood and Currie found over a four year study of dryland corn in Kansas that yields in drier years decreased with higher plant populations (Norwood and Currie, 1996). Another study at experiment stations in Georgia noted that maximum yields were experienced for higher population densities during optimal management, but when moisture stress occurred, lower population densities yielded the highest. In the current study, moisture stress is expected at the end of the growing season, so areas of the field which are less susceptible to drought (those with higher moisture-holding capacity) should experience higher yields (Cummins, 1976).

Relationship of Soil Nutrients and Soil Chemistry to Crop Growth

The relationship between crop yield and soil nutrients has been well documented in the agronomy and soil science literature. The database for this research includes several of these soil parameters, each of which has been established as a determinant of yield. Typically, there is a positive correlation between the values of each and yield, but this is not always the case, as will be demonstrated for phosphorus and soil pH. The degree to which crop development is influenced differs for each soil parameter, so brief descriptions of each will be given below. Soil parameters which are typically analyzed include K, P, OM, CEC, Ca, Mg, and soil pH. Soil nitrogen levels, widely recognized as a major limiting factor in corn development, is excluded from the database because of the difficulty inherent in describing the spatial distribution of such a highly mobile nutrient (De Willegen, 1991).

Potassium is a nutrient occurring in the soil medium which does not comprise any organic compound within plant tissues, but nonetheless plays an important role in crop development (Sauchelli, 1965). Potassium is essential for the energy conversion process of photosynthesis; it also helps to decrease the adverse effects of inadequate moisture levels within the plant. A study in Aberdeen, Scotland, found that plants with adequate levels of K were not as susceptible to scorching and wilting as were plants with insufficient K (Cowie, 1951). In the current study, if precipitation events are sparse at the end of the growing season, as is expected in southern Illinois, then this could enhance the positive relationship between yield and K.

Phosphorus is vital to a number of plant physiological processes, including seed germination, cell division, and the formation of fat. Although phosphorus occurs in the soil in the form of phosphate, elemental phosphorus is extracted inside plant cells and is highly mobile (Sauchelli, 1965). The elemental form moves from older tissues to younger, developing tissues, so one would expect, in general, a positive correlation between P and yield. However, if it is applied continually, phosphorus reservoirs can build in the subsoil, which may result in detrimental conditions to crop health if application is continued

(Swain, 1983). Thus, a positive or negative correlation may be detected between P and yield.

Organic matter and cation exchange capacity (CEC) are soil parameters which describe the structure of the soil medium. Organic matter is the amount of biodegradable material in the soil, which is rich in carbon, and CEC is a measure of a soil's ion transfer gradient. Both parameters are highly correlated with each other; higher organic matter levels in the soil represent availability of ion exchange sites (USDA, 1993). Since both are measures of the soil's ability to utilize available nutrients, a positive correlation between CEC and yield is expected.

Soil pH is measured on a logarithmic scale in which 7.0 is considered neutral, values < 7 are acidic, and values > 7 are basic, or alkaline. Typically, lime is applied to ameliorate acidic soils, which can be toxic to plant health, due to the oxidation of metals such as aluminum and manganese (Ritchie, 1989). However, the optimal growth pH for a plant is usually some value less than 7; thus, the relationship between pH and yield will be dependent on the variability of pH within the field and the specific crop being cultivated. Also, since lime application neutralizes acidic soils, the relationship between soil pH before and after harvest can be an indicator of the success of such management practices which seek to remove soil acidity.

Much like cation exchange capacity, soil pH exerts an influence on a soil's potential for chemical interactions. Thus, the spatial coincidence of certain levels of acidity and soil nutrients can influence the degree to which a particular nutrient can be utilized by the crop. For example, Harper notes that increasing levels of acidity in soils are responsible for the transformation of phosphates into insoluble forms which cannot be absorbed by plants (Harper, 1947b). Also, the decomposition of organic materials in acidic soils is slowed, thus resulting in lower levels of available sulfur and nitrogen. These relationships illustrate the interaction of nutrient levels and a soil's chemical potential.

Spatial Distribution of Soil Properties

Typically, the levels of soil nutrients such as those listed above are measured within a field in an attempt to determine the amount of fertilizer to spread. However, the joint spatial distribution of several of these variables may give clues to the structure of a soil, assuming that there has been no site-specific management of nutrients within a field. Zones of homogeneous soil characteristics are typically delineated with respect to soil variables which describe the structure and physical properties of the soil. Khakural et al. computed a soil productivity index (SPI) which was intended to divide a field in central Minnesota into zones of common production capabilities; this index was a function of available water-holding capacity, soil pH, soil permeability, and bulk density (Khakural et al., 1996). The index was then used in a regression model which attempted to predict yield; the resulting model explained 90 percent of the variation in yield for that particular field.

Soil structure is an important determinant of crop yield because a soil's structure governs the interface between available moisture and a plant's root system. In a study of the relationship between soil structure and corn yield over a seven-year period in Iowa, Cambardella et al. found a strong relationship between aggregate size distribution of soil particles for each of the seven years, as well as notable consistent relationships between such related factors as bulk density and volumetric water content (Cambardella et al., 1994). Thus, spatial variability in within-field crop production can be a function of variable moisture-holding capacities throughout the field.

Geostatistics and Soil Fertility Data

Regionalized Variable Theory

Spatial data analysis, especially within the context of geographic information systems, has experienced a great deal of attention in recent literature. Bailey and Gatrell offer a synopsis of the role of GIS in analyzing spatial data, and Berry discusses the differences between classical and spatial statistics within the context of GIS (Bailey and Gatrell, 1995; Berry, 1995). The treatment of data observations as non-independent entities marks the major departure of spatial statistics from traditional statistics; data which interact in space are assumed to be, to some extent, non-randomly distributed. However, it is expected that the distribution of spatial data will assume a pattern which is neither completely regular nor completely random.

This assumption of "ordered chaos" pervades the study of regionalized variable theory, from which the technique of kriging was spawned. Journel and Huijbregts illustrate how functions which describe the disribution of a regionalized variable must include both a random and a stochastic component (Journel and Huijbregts, 1968). Oliver and Webster note that geomorphic features, such as soil fertility parameters, are to some extent spatially dependent; in other words, there is some degree of spatial autocorrelation in the data (Oliver and Webster, 1986). The method of kriging, which creates interpolated surfaces based on the structure of point-level data, will be used in the creation of the database in this research.

A regionalized variable can be defined as, at a simple level, any variable which is distributed in space. This variable can be described by some function [f(x)] which takes place in either two or three dimensions (since the current study does not attempt to ascertain levels of soil parameters at various depths, only two dimensions will be considered here). The function [f(x)] assumes some value at every point x within some region. The peculiar characteristic of this function is that it is both locally erratic, sometimes even almost random at a large scale, but generally structured, with predictable patterns emerging (Journel and Huijbregts, 1968). Thus, at some scale, structure will not appear in sampled data, but at some other larger scale, relationships between points within the region of interest can be described by a mathematical function.

A random function describing the data could be computed by taking into account all realizations of the regionalized variable, e.g., the values of the variable at all discrete locations within the region of interest. With this in mind, the expression $Z(x)$, which describes the value of regionalized variable Z at location x, is locally random, but two points separated by some distance h are in some way correlated; thus $Z(x)$ and $Z(x + h)$ are not independent of each other. The nature of this correlation between two points separated in space is described by the average relationship between other points in the sampled distribution which are separated by the same distance h. Thus, the key assumption in regionalized variable theory, which has been termed the hypothesis of stationarity, states that the difference between any two points $Z(x)$ and $Z(x + h)$ separated by a lag distance h is dependent not on the locations of the two points, but on the distance that separates them in space. The function which describes the variance of a single point's value about the mean is given in the following equation (Journel and Huijbregts, 1968):

$$\text{Spatial Variance} = \mathrm{Var}\{Z(x)\} = \Sigma\{[Z(x) - m(x)]^2\}$$

where Z is the regionalized variable, x is some point, Z(x) is the value of the regionalized variable Z at point x, and m(x) is the mean of all points within the study area. The semi-variance quantifies the relationship between two points separated by some distance, and is given by the following equation (Journel and Huijbregts, 1968):

$$\text{Semi-Variance} = \gamma(h) = (\mathrm{Var}\{Z(x1)\} - \mathrm{Var}\{Z(x2)\}) / 2$$

where h is the distance between points x1 and x2, and Z(x) is the value of variable Z at location x.

The method of spatial interpolation by kriging seeks to estimate the value that some variable will assume at various locations within a study area based on the values at certain sampled locations. Kriging is based on regionalized variable theory, and thus it needs to quantify the relationships between points separated by various distances in the study area in order to build a matrix of weights which can be used to estimate values. The idea here is that when estimating the value of some variable Z at some discrete location x, the values at all known sampled locations within the study area are to be used. The amount of influence that a point has on the location to be estimated will be given by the semi-variogram computed for two points separated by that distance.

In order to produce a matrix of weights which describe the amount of spatial dependence between points in a study area separated by some distance, the semi-variance must be calculated for every separation distance between points in the sampling scheme. However, when interpolation with kriging is actually performed, there may be a distance between the location to be estimated and the available sample points for which no semi-variogram has been computed, due to the few distances that separated the sample points.

Thus, to obtain a continuous function which provides a semi-variance value for any distance between points in the study area, a graph is created in which the semi-variances at known separation distances are plotted on the y-axis and the distances are plotted on the x-axis. The resulting graph is known as a semi-variogram (Journel and Huijbregts, 1968).

A function can be fitted through the points displayed in the resulting graph to obtain an estimate of semi-variance at every possible separation distance of discrete points within the study area. Several models are available which can compute this function; each model has its own particular assumptions about the shape of the distribution of semi-variances. In order to objectively assess which model is the most appropriate, however, one can run an interpolation on a set of points for several semi-variogram models and then perform a cross-validation analysis to determine which model is the closest to reality (Hosseini et al., 1994). Cross- validation looks at the sum of squared deviations between the values of the regionalized variable at actual sampled locations and the values estimated at those same locations. The formula for performing cross-validation is as follows (Hosseini et al., 1994):

$$RMS = ([\Sigma\{Z(x1) - Z(x2)\}^2] / N)^{1/2}$$

where N = number of sampled points, Z(x1) = estimated value at point x, Z(x2) = actual value at point x, and the summation is iterated from 1 to N.

Applications of Kriging to Soil Fertility Data

Although the method of spatial interpolation by kriging was developed in the field of Geology for application in mineral exploration, it has been successfully applied to other spatial variables which meet the assumptions of a regionalized variable. The current study utilizes kriging specifically for the interpolation of soil parameters. Oliver and Webster first applied kriging to geomorphological data, introducing the concepts of regionalized variable theory to the geographic literature, using small-scale soil properties such as stone content, sandiness, and percentage mottling in color of soil (Oliver and Webster, 1986a). Oliver and Webster followed this study with another which investigated model-fitting procedures for the same types of soil variables (Oliver and Webster, 1986b). These two studies were focused primarily on methodological considerations when applying regionalized variable theory to soils data, thus paving the way for more detailed analyses in the literature.

Han et al. established the primacy of kriging as a spatial interpolator in the soil sciences (Han et al., 1992). Noting that data observations in soil parameters were spatially dependent, this study illustrated the flaws in using interpolation methods based on regression, also known as trend-surface analysis, because of the parametric nature of regression. Kriging assumes nothing about the distribution of the regionalized variable of interest, but rather seeks to understand the nature of the spatial dependence of observations within a study area. Han et al. also point out that this is superior to interpolation methods which use a simple pre-defined distance decay effect, such as in inverse distance-weighted interpolation. This method acknowledges the spatial dependence of data observations, but applies a single constant which describes the spatial dependence (some root power of the distance between points), as opposed to kriging, which is sensitive to how the scale of observation influences spatial dependence within the area of interest.

Recently, studies have appeared which apply kriging and other spatial interpolation techniques to a larger scale of soil data, such as that encountered in individual farm-fields.

Delcourt et al. examined the spatial variability of soil fertility parameters in Belgium using the semi-variogram (Delcourt et al., 1996). In this study, kriging was performed on carbon, phosphorus, pH, calcium, magnesium, and sodium. Logarithmic transformations were applied to some of the data to exaggerate their spatial variability at the scale of measurement, or sampling. A weighted least squares method was used to determine which model fit the semi-variograms best; this form of quantitative cross-validation was in contrast to many previous studies, which used visual interpretation to assess the goodness-of-fit of a particular model.

Another interesting approach taken by this study was the elimination of what was termed zonal drift, or local trends in the data which could not be applied to the field as a whole. To overcome this, historical management records of the field were used to delineate zones of homogeneous management within the fields; each sampled value was subtracted from the mean value of the regionalized variable within its management zone. In this way, the authors were able to remove the local perturbations in the data. This method of data de-trending by dividing the sampled locations into blocks was also used successfully by Tevis et al. in the estimation of soil pH (Tevis et al., 1991). It is interesting to note that since this method successfully eliminated the local error in the estimation, it demonstrated that the spatial distribution of soil parameters can be as much a function of human land use and management as it can be a function of geomorphological or pedo-genetic (soil-forming) processes.

Analysis of Spatial Data

Spatial Application of Analysis of Variance (ANOVA)

One tool which has been used in geographic data analysis is the spatial analysis of variance (ANOVA). The objective in a spatial ANOVA is to determine whether or not a significant difference exists between the means of three or more independent samples of variables distributed in space. The distinction between sample groups could be based on some spatial blocking pattern, for conditions which exert an influence on the means are assumed to be homogeneous (McGrew and Monroe, 1993). Wessman et al. used ANOVA in a study of landuse impact on vegetation of the Konza Prairie to determine whether or not burning or grazing significantly produced significantly different vegetation distributions (Wessman et al., 1995). Using remotely-sensed multi-spectral imagery, this study attempted to isolate land management as the culprit in explaining differences in vegetation given otherwise similar conditions.

In an agricultural study, blocking strata based on homogeneous soil conditions which are known to influence yield (from a regression model) can be created, and ANOVA can be used to determine whether or not significant yield differences are present between the strata (Waits and Johnson, 1996). Since there are several soil properties which can exert an influence on yield, the analyst is faced with two possible solutions; either perform a multiple ANOVA, which seeks to analyze the effect of several blocking strata, or produce an unmeasured surrogate variable which describes the spatial variability of a number of variables within an agricultural field. Multivariate methods for quantifying the joint variability of *n* variables create new regions of homogeneity which can be used as blocking strata in an ANOVA.

If a significant difference is detected from an ANOVA, then multiple comparisons can be used to determine which blocks are different, which are similar, and rank them from high yielding to low yielding blocks (Hicks, 1964). ANOVA simply tests the null hypothesis Ho: $\mu 1 = \mu 2 = \ldots = \mu x$; however, if this hypothesis is rejected, then the question arises as to which means are significantly different from others. By comparing differences in observations between blocks with some threshold which represents statistical significance, a multiple comparison can group the means of blocking strata in such a way that illustrates which blocks are similar (according to the set threshold) and which blocks are different.

Regression Analysis Using Spatial Data

Regression analysis attempts to describe the relationship between some number of covariates and a response variable (Clark and Hosking, 1986). A model is constructed in which the independent variables, which are used to predict the response (dependent) variable, are included and their relationship with respect to the dependent variable is described. These variables can interact with each other to vary the dependent variable, and the regression model can be altered to include these interactions. The goodness-of-fit of the resulting model is given by the coefficient of determination, which is represented by R^2. This value is bounded from 0 to 1, and represents the percentage of variation in the dependent variable which is accounted for by the model. Further, each component of the model can be ranked in a stepwise regression such that only those components which exert a significant influence on the dependent variable are included, and these components are ranked according to their ability to influence the dependent variable.

The regression of spatial data can be used to describe the degree to which soil characteristics exert an influence on crop yield. The problem with applying regression to spatial data is the presence of spatial dependence, mentioned above in the discussion of regionalized variable theory, which gives rise to autocorrelation of residuals from the regression model (Clark and Hosking, 1986). Elston et al. note this problem in their study of regression of spatial data in a GIS, proposing that independent variables be aggregated into surrogate variables using principle components in an attempt to reduce the mean squared error of the regression model (Elston et al., 1997). However, although the presence of autocorrelation violates the assumption of independent error terms, unbiased estimates of beta coefficients can still be obtained (Clark and Hosking, 1986). Since the purpose of the regression model in this project is to rank soil parameters based on the degree to which they exert an influence on yield, and not necessarily to obtain a high coefficient of determination, the presence of autocorrelation should not present a problem.

Cluster Analysis and the Delineation of Homogeneous Regions

The spatial variability of any onevariable can be used to analyze the spatial distribution of a response variable; however, there are often many variables within the area of interest which exert an influence on the response variable, and can be measured. Once the spatial variability of each variable has been quantified using the aforementioned process of kriging, these resulting patterns can be analyzed to produce zones of homogeneity using cluster analysis. Cluster analysis in many applications seeks to taxonomically classify a number of observations in a tree diagram, given a certain number of variables which describe each observation (Hartigan, 1975). Cluster analysis using a k-means algorithm, however, plots a number of variables in n-dimensional space for a number of objects, or observations, and computes a matrix of euclidean distances from each object to all other objects (Hartigan, 1975). In this way, it is possible to delineate a number of clusters which seek to maximize between-cluster distances and minimize within-cluster distances.

To use the k-means algorithm, the analyst must choose the number of clusters (k) into which the data will be partitioned. This preliminary step arises because of the iterative nature of cluster analysis; each iteration attempts to create clusters in which the within-cluster variance is minimized and the between-cluster variance is maximized. As a result, at each iteration, there must be partitions defined into which outlying observations can be moved. The iteration stops when movement of observations into neighboring clusters no longer increases the between-cluster variance.

The application of cluster analysis to produce geographic regions based on a number of variables has been demonstrated in a study of Alpine treelines by Allen and Walsh (Allen and Walsh, 1996). This study used a k-means cluster analysis to determine the spatial organization within alpine treeline ecotones, or transition zones. Variables derived from satellite imagery were used as inputs to the cluster analysis; these included edge density, fractal dimension, contagion, and other metrics which describe the pattern of adjacent image pixels. The interesting aspect of this application was that in order to determine the number of clusters (k) to partition the data into, the authors ran the analysis for several possible k partitions, and recorded the variable means (and distances between means) for each number of clusters. In this way, the authors were able to determine the number of clusters at which between-cluster distances begin to diminish (Allen and Walsh, 1996).

CHAPTER III

METHODOLOGY

Data Collection - 1996 Growing Season

The research was conducted over the summer 1996 growing season on a farmstead near Claremont, Illinois, managed by Gray Farms, Inc. Two fields were under examination (see Figure 1); GF76 and GF120. Both fields were planted to white food-grade corn. The seed population (planting density) of these crops was varied such that the crop was more dense in areas with a higher soil moisture holding capacity. The logic behind this particular variable rate management technique was that rainfall events in southern Illinois are such that the distribution of precipitation is not constant throughout the growing season; thus, areas which can retain moisture better than others are more likely to promote crop health and vigor during longer periods of little or no precipitation which can occur during the growing season. Concentrating crops in those areas with a higher production capacity would then make economic sense to a farmer who wishes to mitigate the effects of stochastic rainfall events.

However, since a soil's moisture holding capacity within a field cannot be measured directly, it becomes necessary to develop surrogate variables which characterize the spatial distribution of this phenomenon. To this end, a knowledge of known biophysical parameters within the field can be combined with a producer's

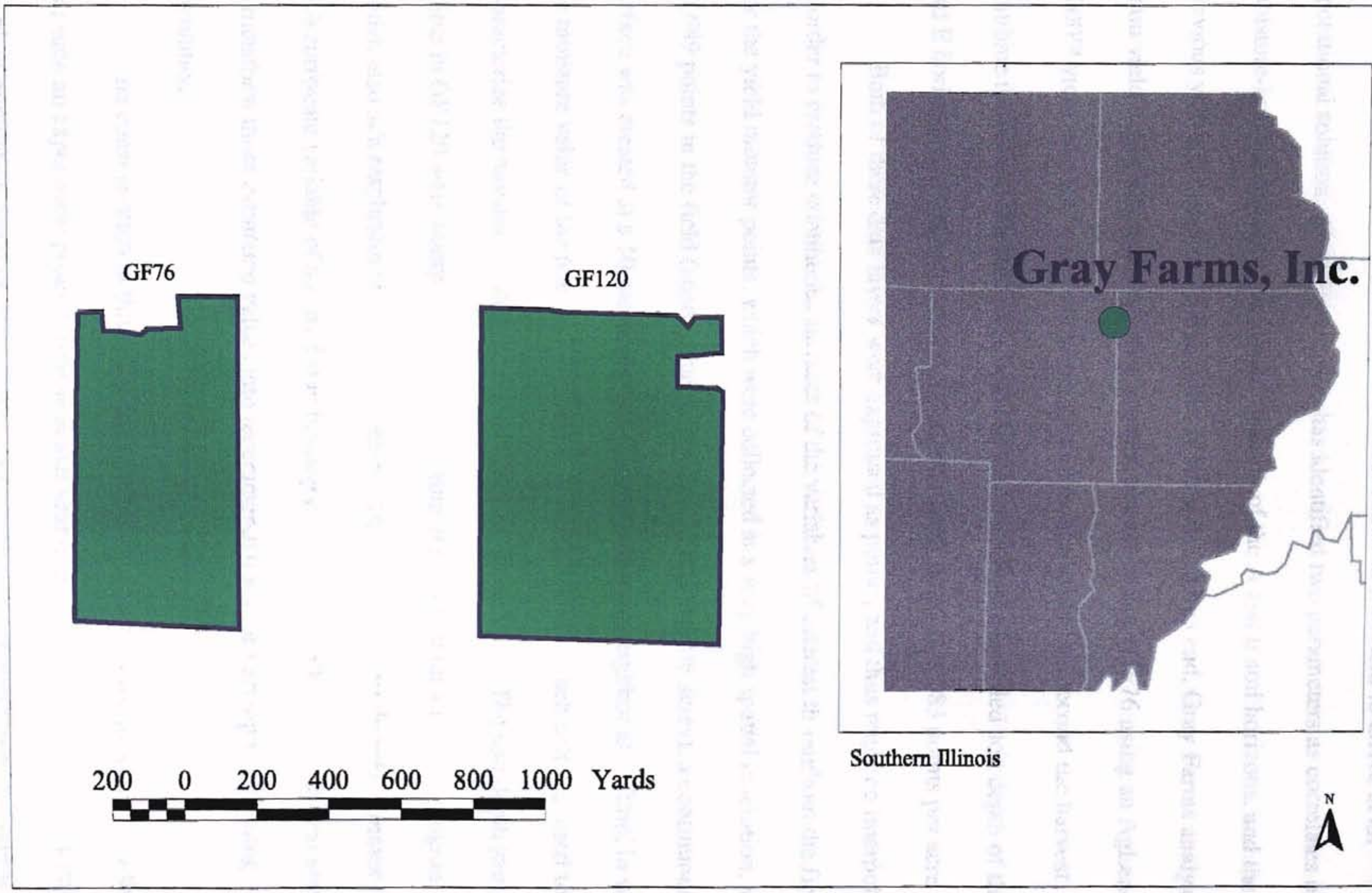


Figure 1. Fields Under Investigation

intuitive knowledge and experience in the cultivation of that field to arrive at an operational solution. Gray Farms, Inc., has identified two parameters as correlates of soil moisture-holding capacity; the combined depth of the A and E soil horizons, and the previous year's crop moisture at the time of harvest. To this end, Gray Farms analyzed grain yield monitor data for the 1995 growing season on field GF76 using an AgLeader 2000™ yield monitor, which measured grain moisture every one second the harvest combine traveled in the field. For field GF120, Gray Farms sampled soil depth of the A and E horizons at 117 discrete points within the field, or about 1.083 points per acre.

Both of these data layers were expressed as points, and thus required interpolation in order to produce continuous surfaces of the variables of interest throughout the fields. For the yield monitor points, which were collected at a very high spatial resolution, with 20,689 points in the field (approximately one point every 0.0036 acres), a continuous surface was created at a 10-meter resolution using a nearest neighbor algorithm, in which the moisture value of the point falling closest to the centroid of each cell was used to characterize the moisture within the entire 10 square meter area. The soil depth sample points in GF120 were interpolated using a kriging algorithm with a linear variogram model, also at a resolution of 10 square meters. These surfaces produced estimates of each surrogate variable of soil moisture holding capacity for both fields; the next step was to transform those estimated values into recommendations for varying the planting population.

One consideration in this study from the point of view of the producer was the risk that such an experiment poses during an actual year of production. Too much variation in the plant population could exaggerate the resulting variation in yield, which could be beneficial for analysis, but as a result would produce unacceptably low yields in certain portions of the fields, both of which were under food-grade production contracts (Gray, 1996). Since this study relies on actual production fields in agriculture, which are large enough to encompass spatial variability and are managed according to real-world constraints, some care had to be taken to both produce quantifiable results and satisfy the producer's contractual and financial obligations. Thus, the producer decided to vary the planting population between 18,000 and 25,000 plants per acre in field GF76, and between 18,000 and 26,000 plants per acre in field GF120. Polygonal maps of the variability of the two surrogate variables were created and used as recommendations for varying the seed population rate for the summer 1996 corn crop (Figures 2 and 3). The relationship between the measured criterion for varying population and the plant population in seeds per acre is outline in Table I. Each measured variable was classified into discrete categories based on a quantile classification scheme, which seeks to place an equal number of observations in each category (McGrew and Monroe, 1993). Note that although the ranges of the measured variables are very different, the ranges within which the plant populations were varied is comparable. This reflects the fact that there was no quantifiable model linking the seed population to either of the surrogate variables; rather, an attempt was made to spread the population variability out over the surrogate variable's variability distribution.

Figure 2. Seed Planting Population Zones In Field GF76

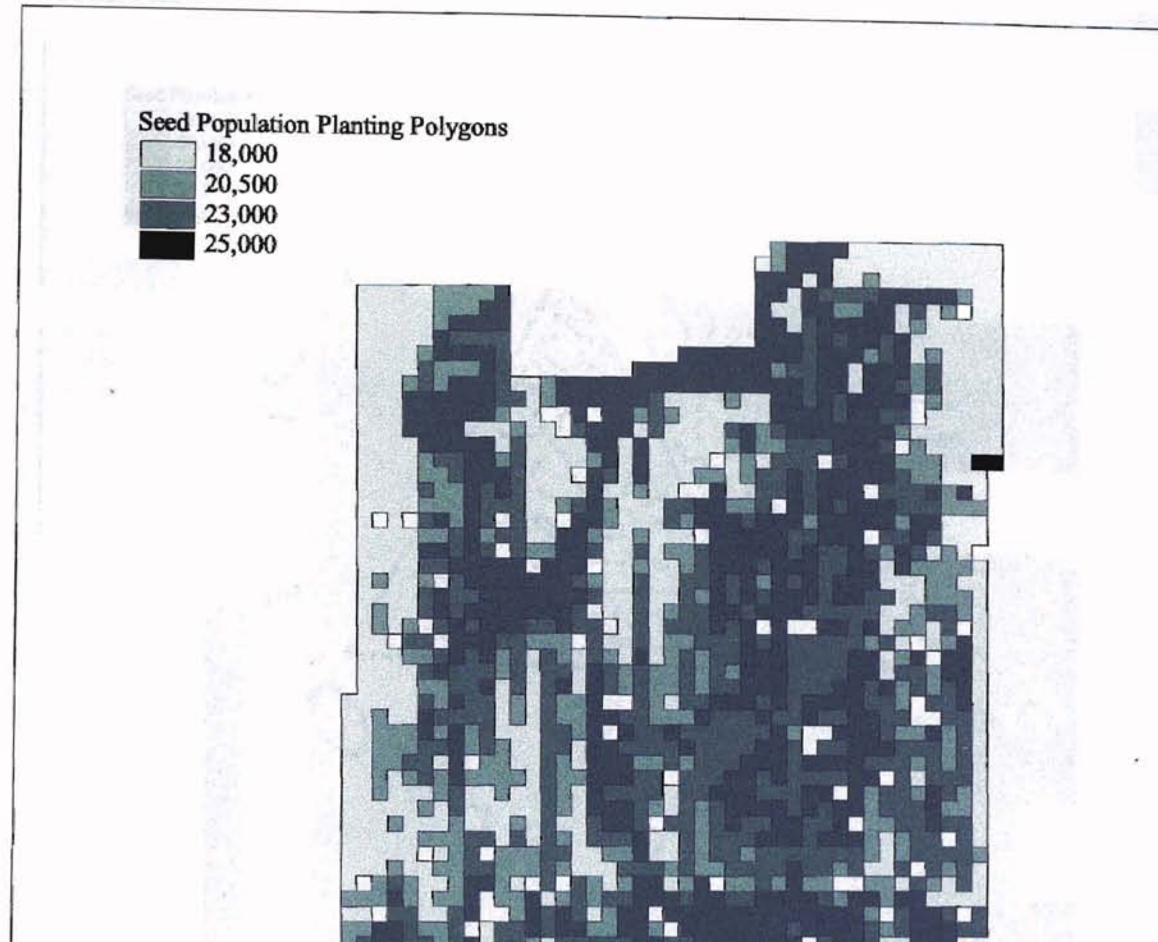


Figure 3. Seed Planting Population Zones In Field GF120

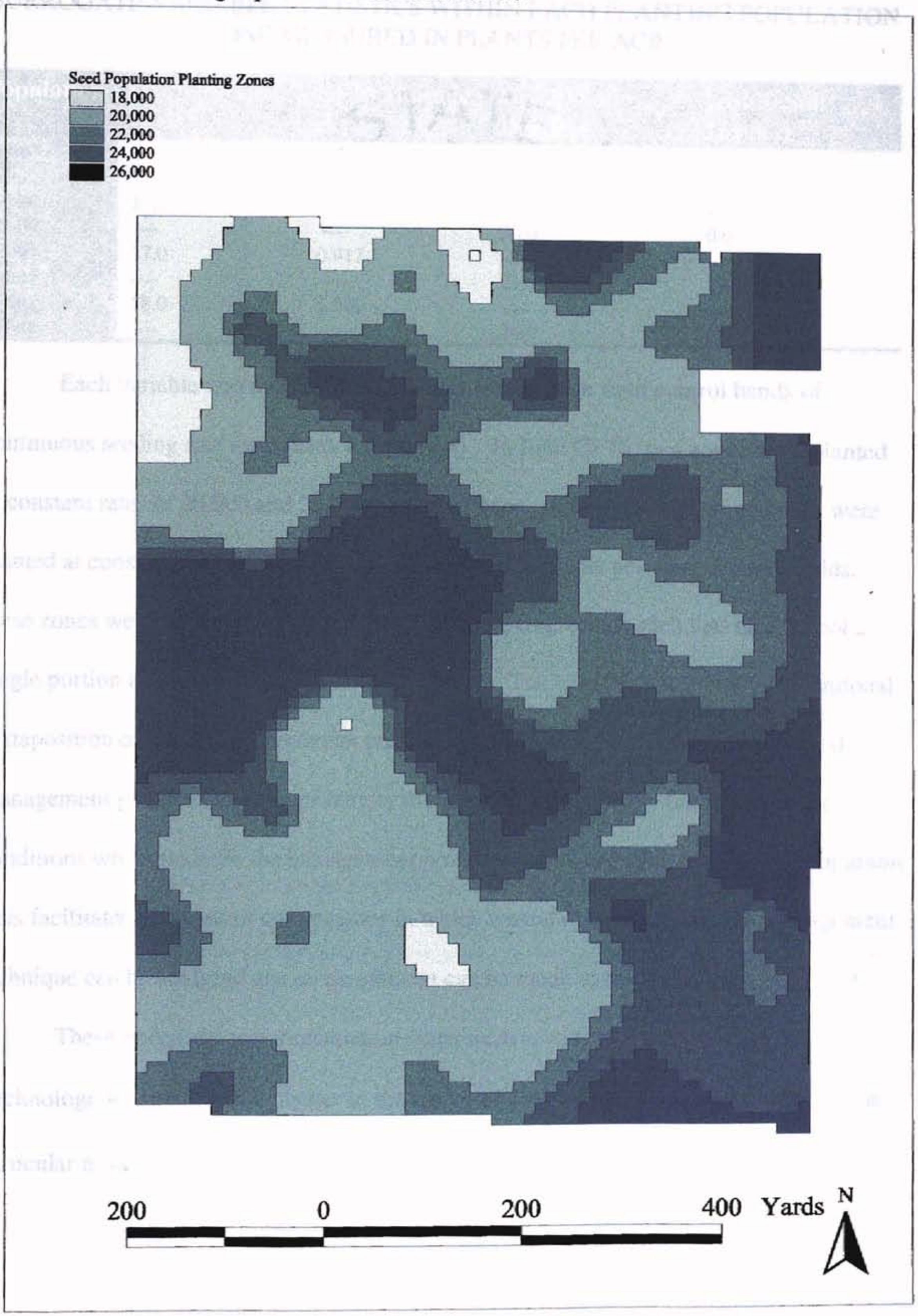


TABLE I

SURROGATE VARIABLE STATISTICS WITHIN EACH PLANTING POPULATION ZONE MEASURED IN PLANTS PER ACRE

Population	GF76 Moisture μ	GF76 Moisture σ	GF120 Soil Depth μ	GF120 Soil Depth σ
18000	15.0	0.778	9.0	2.0
20000	----	----	16.0	2.0
20500	17.0	1.199	----	----
22000	----	----	20.0	0.0
23000	17.0	0.917	----	----
24000	----	----	21.0	1.0
25000	18.0	1.541	----	----
26000	----	----	26.0	3.0

Each variable was used on a single field, and in each field control bands of continuous seeding rate were planted (Figure 4) . In field GF76, two zones were planted at constant rates of 20,000 and 25,000 plants per acre. In field GF120, three zones were planted at constant rates of 20,000, 25,000, and 30,000 plants per acre. In both fields, these zones were interspersed with variable-rate planting zones, such that there is not a single portion of the field designated for "control". The benefit of this type of intentional juxtaposition of the two management practices is that placing the spatial and aspatial management practices in close proximity increases the probability that the growing conditions which underlie the management practices will be similar, or at least comparable. This facilitates the types of comparisons in which spatial differences in each management technique can be analyzed and an assessment can be made as to their relative successes.

These polygonal recommendation maps were used to drive a Midwest Technologies® variable rate planter in the spring of 1996, when planting occurred. This particular model of planter communicates back to the recommendation computer the

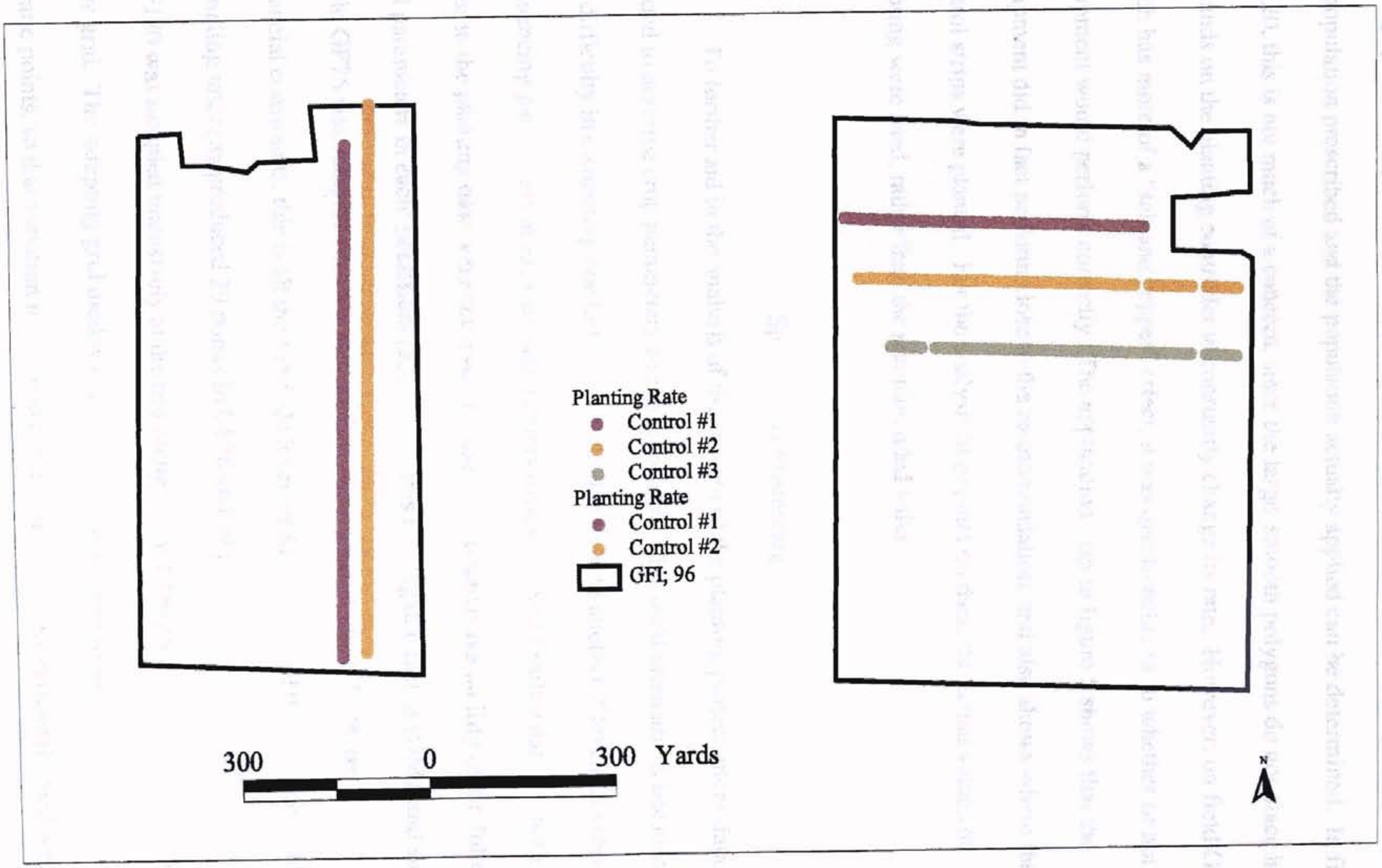


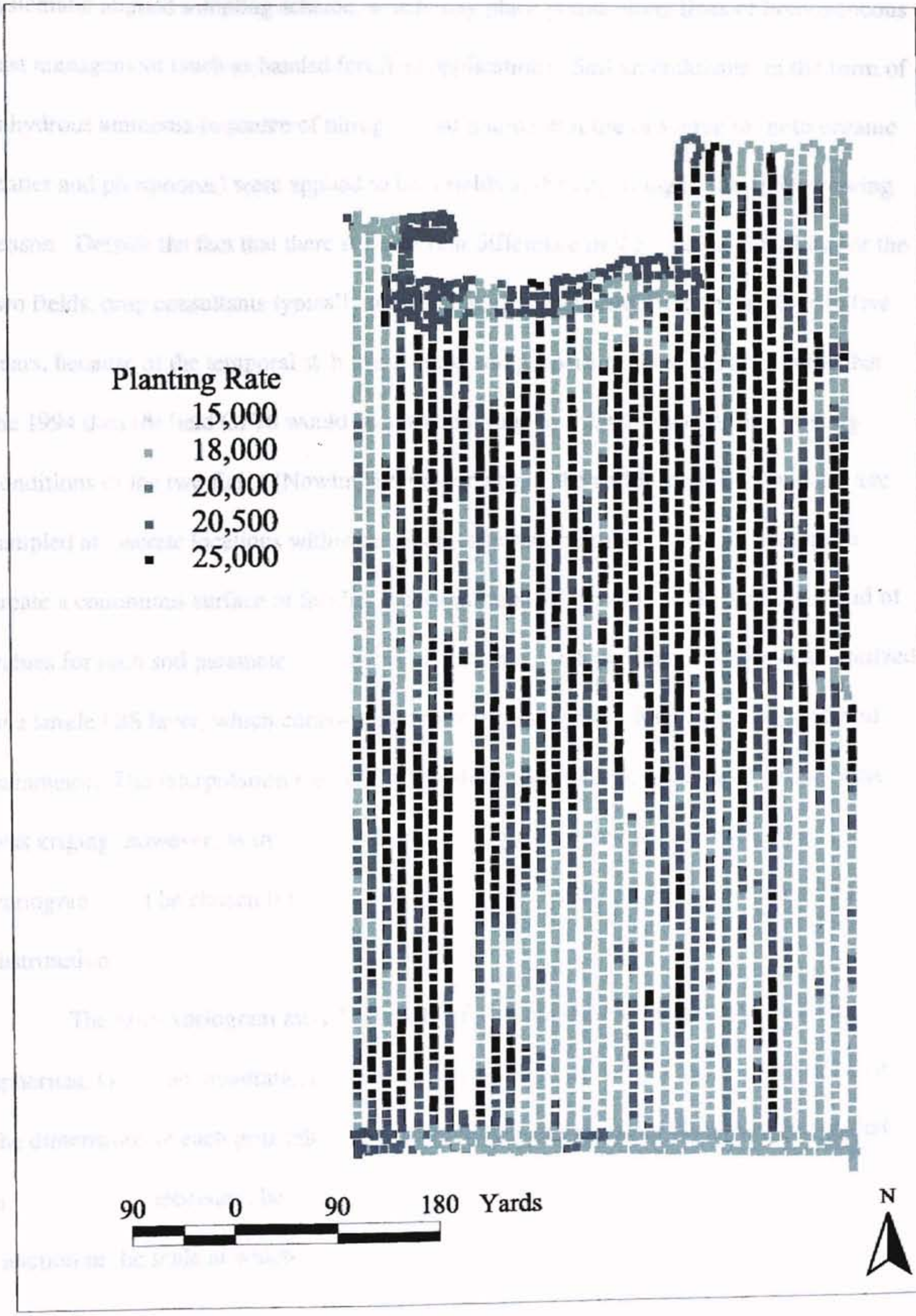
Figure 4. Seed Planting Control Strips in Fields GF76 and GF120

actual population applied to the field at discrete locations, so that the difference between the population prescribed and the population actually applied can be determined. In field GF120, this is not much of a concern, since the large, smooth polygons do not place heavy demands on the planting controller to constantly change its rate. However, on field GF76, which has more of a "salt and pepper" effect, it was questionable as to whether or not the equipment would perform correctly. The application map in figure 5 shows that the equipment did in fact perform close to the recommendation, and also shows where the control strips were planted. For the analysis of population data, the actual values of planting were used, rather than the recommended values.

Spatial Data Processing

To further aid in the analysis of the success of the planting project, more data were needed to describe crop parameters within each field. This need often arises because of the difficulty in comparing one field (or sub-field region) to another if levels of crop-influencing parameters in each are significantly different. Soil fertility data for both fields prior to the planting date were necessary to assess the relative availability of the following soil parameters in each: potassium (K), phosphorus (P), organic matter (OM), and soil pH. Field GF76 was sampled on a 2.5 acre grid for these parameters in 1994; because of financial constraints, this is all the sub-field level fertility data available for that field (the sampling intensity produced 29 points in GF76 and 49 points in GF120). However, field GF120 was sampled intensively at the beginning of the 1996 growing season, also on a 2.5 acre grid. The sampling grid used was a systematic unaligned grid, which produces locally erratic points, so that variation in the field can be captured more efficiently than with a

Figure 5. Planting Rate Performance In Field GF76



systematic aligned sampling scheme, which may place points along lines of homogeneous past management (such as banded fertilizer application). Soil amendments, in the form of anhydrous ammonia (a source of nitrogen) and poultry manure (a source of both organic matter and phosphorus) were applied to both fields at the beginning of the 1996 growing season. Despite the fact that there is a two year difference in the dates of sampling for the two fields, crop consultants typically recommend only sampling fields every four to five years, because of the temporal stability of most soil properties; thus, it was decided that the 1994 data for field GF76 would be sufficient to give a comparison of the growing conditions of the two fields (Nowlin, 1996). However, the soil fertility data which were sampled at discrete locations within the fields required interpolation, in an attempt to create a continuous surface of fertility values throughout the fields. By creating a grid of values for each soil parameter, the spatial variation of each parameter can be characterized in a single GIS layer, which contains attributes for each grid cell representing each soil parameter. The interpolation method used in the creation of this database for each field was kriging; however, as mentioned above, the appropriate model for computing a semi-variogram must be chosen if the estimates are to be reliable, given a particular spatial distribution.

The semi-variogram models available for use in the interpolation were linear, spherical, Gaussian, quadratic, and exponential. The resolution of the resulting grid, or the dimensions of each grid cell, could be set by the analyst; prior determination of that resolution was necessary, because the fit of the semi-variogram models is in part a function of the scale at which interpolation occurs (Journel and Huijbregts, 1968).

The eventual goal of the multi-attribute surface was to produce soil production zones via a k-means cluster analysis algorithm, which produces zones which are homogeneous with respect to n-dimensional data, in this case K, P, OM, and pH. In the Statistica™ statistical analysis package, clusters can be produced from less than or equal to 800 observations; thus, for the largest field, GF120, this would constrain the grid cell size to about 33 meters a side, to prevent a surplus of observations. As a result, this was the grid cell resolution chosen for the study; the same resolution was applied to both fields.

All semi-variogram models increase monotonically near the origin of the semi-variogram; thus, it was initially decided to use a linear model for interpolation. Since the distances to be interpolated in a farm field are much less than distances in a geological application, for which kriging was developed, the commonality of all models near the origin pointed to the linear model as the most logical choice (Surfer® User's Guide, 1994). This suspicion was confirmed for field GF76 in the low error of estimation for kriging with the linear model. The procedure used to estimate the error, known as cross-validation, uses the following equation (Tevis et al., 1991):

$$\text{mean absolute error} = (\Sigma \, |X_{kriged} - X_{actual}|) \, / \, N$$

where X_{kriged} is the estimated value at a sampled location, X_{actual} is the sampled value at that location, N is the total number of sampled locations, and the summation is iterated from 1 to N (Tevis et al., 1991).

Field GF120, however, was not so well behaved in its error of estimation. The linear model produced a noticeable amount of error in its estimations of soil properties. Thus, several semi-variogram models were run in an attempt to produce the best possible surface; the results of those interpolations are summarized in Table II. The amount of variation in the soil parameters in field GF120, compared to the relatively homogeneous GF76 may help explain why GF76 produced less estimation error (see Table III).

TABLE II

MEAN ESTIMATION ERROR OF SOIL PARAMETERS FOR DIFFERENT SEMI-VARIOGRAM MODELS IN FIELD GF120

MODEL	OM ERROR	P ERROR	pH ERROR	K ERROR
Linear	0.02214	3.669	0.0237	10.045
Gaussian	0.0231	3.564	0.0382	25.840
Quadratic	0.02169	3.625	0.0246	10.038
Exponential	0.02222	3.682	0.0241	10.115
Spherical	0.02125	3.610	0.0244	9.970

TABLE III

DESCRIPTIVE STATISTICS FOR SOIL PARAMETERS IN FIELDS GF76 AND GF120

FIELD	P μ	P σ	PH μ	PH σ	K μ	K σ	OM μ	OM σ
GF76	72.345	16.554	6.397	0.258	199.103	48.413	3.145	0.766
GF120	127.776	59.441	6.173	0.397	457.449	133.380	2.594	0.307

In order to choose the appropriate semi-variogram model to use for interpolation of soil parameters in GF120, the estimation error for each soil parameter was divided by that parameter's mean, in an attempt to normalize the error in each parameter such that they have no units and can be compared to each other. The resulting standardized error for each parameter was then summed for each soil parameter and divided by four, which is the total number of parameters, to obtain the overall goodness-of-fit of each model; these results are given in Table IV. According to these results, the spherical model minimizes the average estimation error for soil type parameters, and thus it was decided to use the spherical model in the interpolation of GF120 soil data.

TABLE IV

AVERAGE STANDARDIZED ESTIMATION ERROR
FOR SOIL PROPERTIES IN GF120

Linear	*Gaussian*	*Quadratic*	*Exponential*	*Spherical*
0.0158	0.0249	0.0157	0.01799	0.0156

Yield data were collected at the end of the growing season with an AgLeader™ 2000 grain yield monitor. This device is equipped with a real-time differentially-corrected GPS receiver, so that yield data are available for discrete points within the field at known locations. In the case of both fields, one yield point was collected every second. Each yield point represents the amount of grain mass passing through a sensor on board the combine at a specific time; this is aggregated for the width of the header on the combine, which covered 16 planted rows in this case. The resulting yield maps for each field are shown in Figures 6 and 7.

Since each pass through the field may fill a different width of the combine header with grain, the pass itself can be considered a source of spatial variation in the yield data; thus, while points along the direction of harvesting will be spatially dependent, this may not be the case across harvest passes. Also, there is a lag between the point at which the corn stalk is cut and the point at which the grain is measured by the sensor; this lag will be manifested in different directions for adjacent combine passes, since the vehicle turns around for every pass, which introduces further discrepancy between adjacent yield values. Thus, the yield data were not interpolated, because of the lack of spatial dependence of observations across harvest passes; instead, to facilitate data analysis, point-in-polygon analyses assigned yield averages to polygonal zones in each field, which may represent

Figure 6. Yield Data for Field GF76

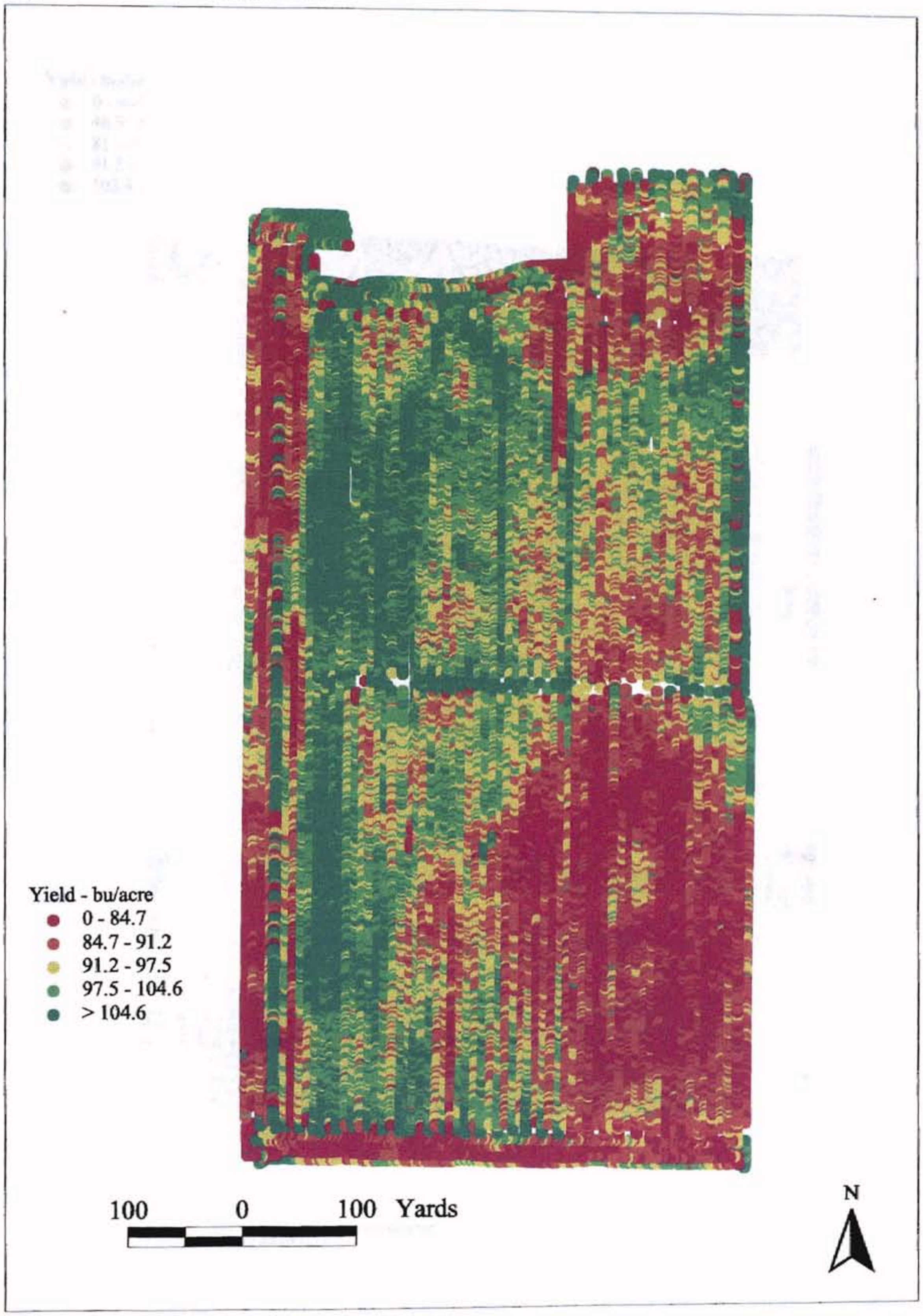
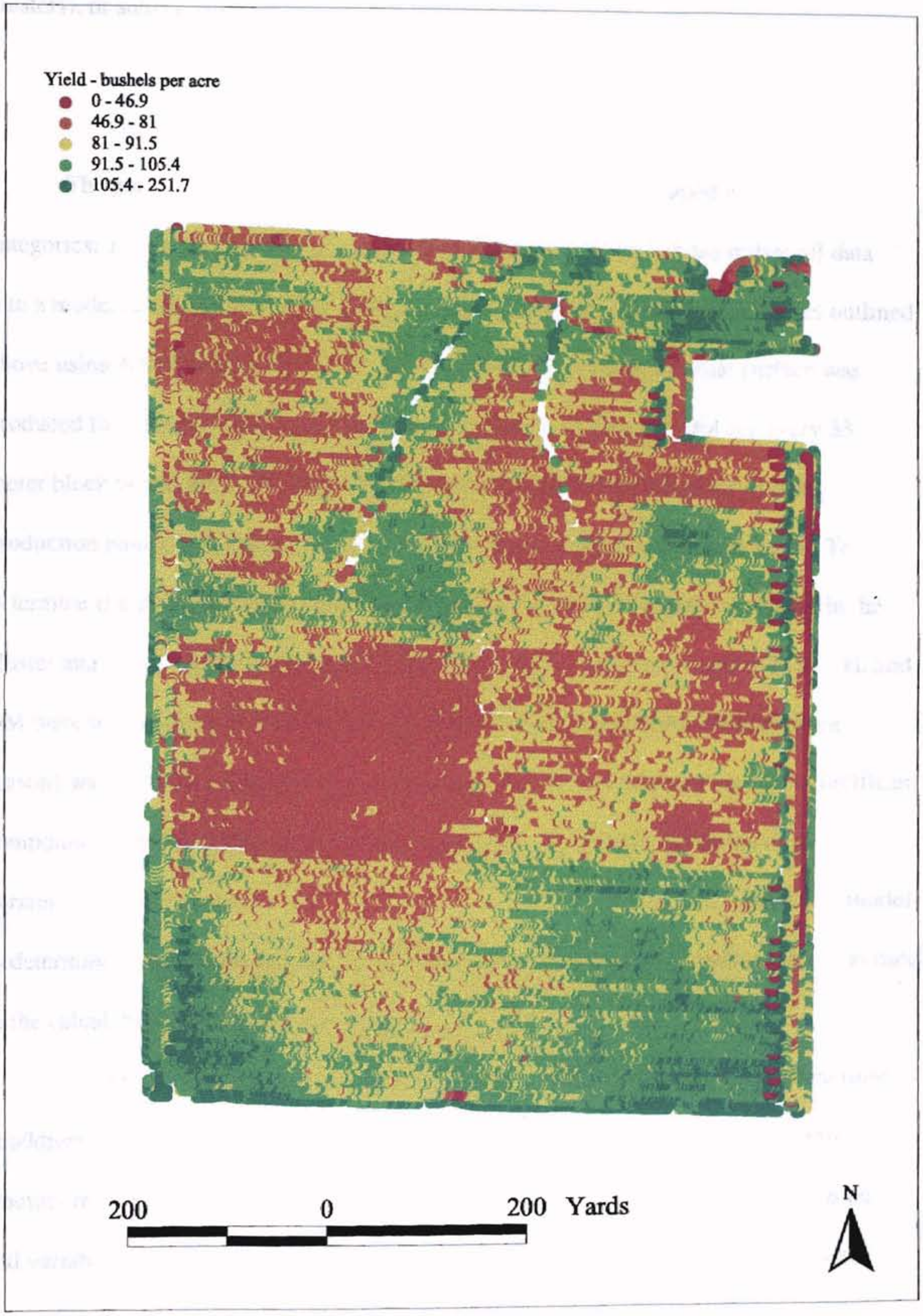


Figure 7. Yield Data for Field GF120



either a treatment (population), a particular growth potential area (as given by soil clusters), or some combination of the two.

Methodology for Data Analysis

The goals of data analysis for this study can be broadly grouped into two categories; 1) data preparation, to create soil production regions and assimilate all data into a model conducive to further analysis, and 2) actual testing of the hypotheses outlined above using ANOVA procedures. As mentioned above, a multi-attribute surface was produced for each field which contained estimates of P, K, pH, and OM for every 33 meter block within the field. These grid cells were aggregated into zones of crop production potential based on the spatial variability of all significant parameters. To determine if a given parameter was significant, and therefore warranted inclusion in the cluster analysis procedure, a multiple nonlinear regression was run in which P, K, pH, and OM were independent, or explanatory, variables, and yield (from the 1996 growing season) was the dependent variable. Any of the soil parameters which were not significant components were not included in the cluster analysis. In addition to the four raw variables, transformations of those variables were fitted with a nonlinear regression model to determine if they were significant; if so, then those transformed variables were also used in the calculation of soil production clusters.

Once these clusters were computed, a number of ANOVA procedures were used to address the hypotheses posed above. For each field, it was necessary to determine whether there was a significant difference in the performance of the constant rate zones and variable rate zones; this was accomplished by comparing each constant rate to all variable rate areas in the field which are of the same or similar population as that particular control zone (in other words, comparing a 20,000 plants per acre constant zone with all 20,000 plants per acre variable rate zones within the rest of the field). If a significant difference was found, it was then appropriate to run a multiple comparisons procedure to determine which area (the control or the experimental) had the higher yield.

The next step, assuming a significant difference was found, was to investigate whether any differences found were attributable to the management technique of varying the rate of planting based on a surrogate of moisture-holding capacity, or to different conditions within the fields. This was accomplished by using the crop production areas, as determined by the k-means cluster analysis described above, in other analyses in which the control zones were broken down by crop-production polygons. Each zone could then be analyzed with its variable-rate counterpart (e.g., all 20,000 density control points in production region 3 will be compared with all 20,000 density experimental points which are also in production region 3). Finally, the performance of the variable-rate planting was compared for the two fields, in an attempt to see which criterion for varying the rate of planting (grain moisture or soil depth) was the most effective. This involved comparing each management technique's ability to explain variation in yield, as well as taking into consideration the growing conditions in each field.

CHAPTER IV

DATA PREPARATION AND PROCESSING

Modeling Yield Using Regression

Overview

In an attempt to better understand the relationship between the measured soil parameters in each field and crop yield, regression was used to model the soil parameters and transformed soil parameters against yield. The objective of this modeling exercise is to determine which variables or derived variables are to be included in the cluster analysis which will produce zones of crop-growth potential. Thus, a model which includes only variables which are known determinants of yield would logically produce zones within which yield is homogeneous, and between which yield is variable. The variables available for each field (P, K, OM, and pH) can be transformed in any number of ways; e.g., X^2, X^3, $1/X$, e^X, etc.. The raw variables will be regressed against yield initially, and the least significant of the available variables will be removed from the model before transformation. This step is necessary because although there may be more than one non-significant parameter, often non-significant parameters can make contributions to the model's explanatory power.

In order to produce data for the regression model, the original soil test points were subjected to a spatial join with 1996 yield points using SSToolbox™ GIS. This procedure finds the yield point closest to each soil test point, and assign the attributes of that location to the soil test point data layer's attribute table. Because of the dense nature of yield data within the fields, there was not a problem in finding yield points which were extremely close to each of the original soil test points.

Field GF76

The results of the initial regression, in which dry yield was regressed against P, K, OM, and pH is summarized in Table V. The coefficient of determination (R^2) for this model was 0.36048; the model's parameters are given in the following equation:

$$\text{Dry Yield} = 157.776 + 0.3259(P) + 0.0211(K) - 12.6389(pH) - 2.293(OM) + \varepsilon$$

TABLE V

REGRESSION OF RAW SOIL PARAMETERS AGAINST YIELD FOR FIELD GF76

Variable name	Standard BETA	Error of Std BETA	BETA	Error of BETA	T(24)	P Level
Intercept			157.7766	44.74063	3.52647	0.001725
P	0.5278	0.192715	0.3259	0.11901	2.73876	0.011441
K	0.099776	0.201043	0.0211	0.04245	0.49629	0.624206
PH	-0.319479	0.180359	-12.6389	7.13513	-1.77135	0.089201
OM	-0.171726	0.17938	-2.293	2.39521	-0.95733	0.347947

According to these results, the variable K had the least probability of significance, at 0.624. The standard beta measure, which indicates the beta if all variables were standardized to a mean of 0 and a standard deviation of 1, demonstrates that K contributes less to the model than any of the other parameters. Thus, K was removed from the model, and the following transformations were applied to the remaining independent variables: X^2, X^3, and $1/X$. A stepwise regression was performed on these variables, and the results of this are summarized in Table VI.

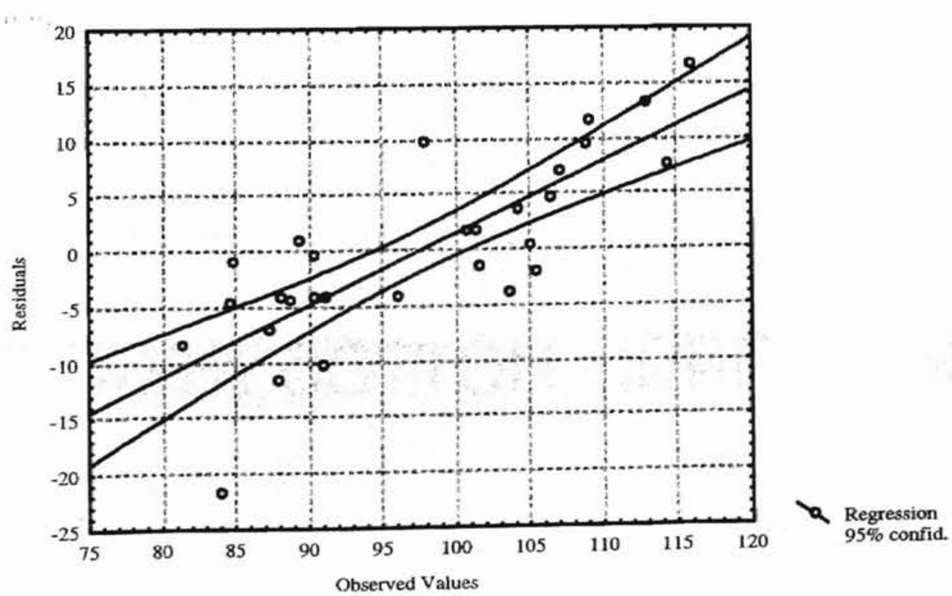
TABLE VI

REGRESSION OF RAW AND DERIVED VARIABLES AGAINST YIELD FOR FIELD GF76

Variable	Standard Beta	Error of Std Beta	Beta	Error of Beta	T(26)	P LEVEL
Intercept			104.4641	14.31591	7.29706	0.0
P	0.55169	0.16529	0.3407	0.10208	3.33772	0.002555
pH^3	-0.368217	0.16529	-0.1202	0.05397	-2.22771	0.034766

The stepwise regression, as shown above, only included two parameters in the model; P and pH^3. The R^2 for this model was 0.33608, which is less than the previous model (which included P, pH, OM, and K, regardless of significance). Thus, a final model was run, in which P, pH, OM, and pH^3 were included; this resulted in an R^2 of 0.35523, which is still less explanatory of yield than the original model with only raw variables; thus, the original model was adhered to. The distribution of residuals resulting from this first model is given in Figure 8.

Figure 8. Observed Yield vs. Residuals From Regression in GF76



Field GF120

The initial model for field GF120, containing only the raw soil parameters, produced an R^2 of only 0.17854. This model , which is summarized in Table VII, yielded the following equation:

$$\text{Dry Yield} = -20.9781 + 17.2909(OM) + 10.8868(pH) - 0.0362(P) + 0.0066(K) + \varepsilon$$

TABLE VII

REGRESSION OF RAW SOIL PARAMETERS AGAINST YIELD FOR FIELD GF120

Variable	Standard Beta	Error of STD Beta	Beta	Error of Beta	T(44)	P LEVEL
Intercept			-20.9781	38.23345	-0.548684	0.585996
OM	0.394228	0.145481	17.2909	6.3808	2.709825	0.009559
PH	0.321053	0.147773	10.8868	5.01094	2.17261	0.035238
P	-0.159589	0.183017	-0.0362	0.04148	-0.871993	0.387946
K	0.06564	0.176699	0.0066	0.01785	0.371481	0.712062

Interestingly, this model also placed the lowest contribution on the variable K, with both the highest p-value and the lowest standard beta. Again, K was removed from the model, and the following transformations were applied to the remaining variables: X^2, X^3, and 1/X. The results of this regression are summarized in Table VIII.

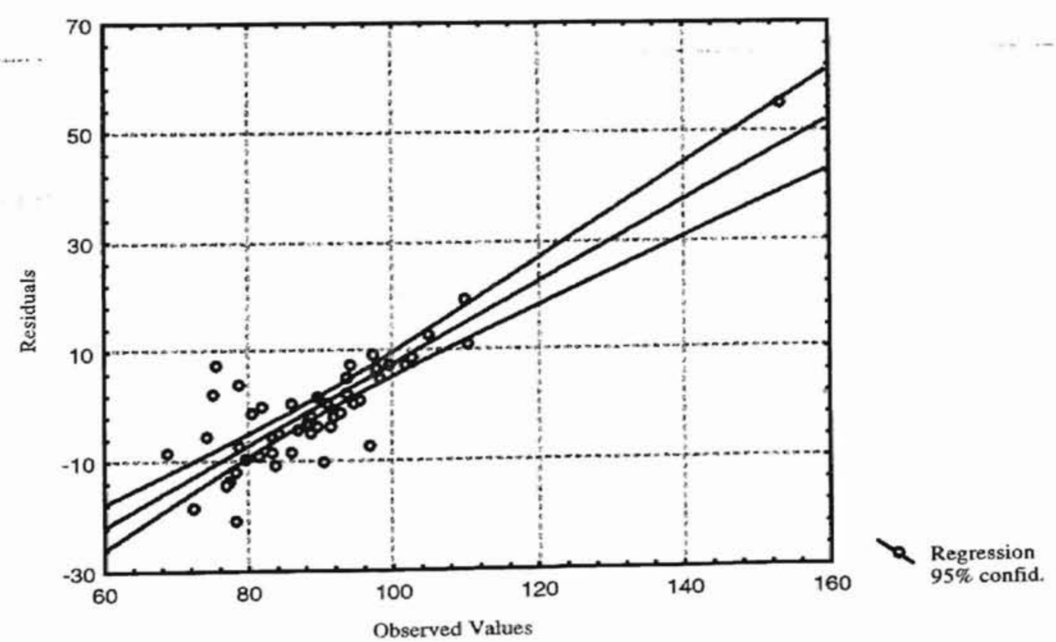
TABLE VIII

REGRESSION OF RAW AND DERIVED VARIABLES AGAINST YIELD FOR FIELD GF120

Variable	Standard Beta	Error of STD Beta	Beta	Error of Beta	T(45)	P LEVEL
Intercept			-113.174	73.26244	-1.54477	0.129405
OM	2.25592	0.957316	98.945	41.98795	2.3565	0.022862
pH^3	0.31225	0.134578	0.09	0.03859	2.3202	0.024922
OM^3	-1.90328	0.954976	-4.145	2.08001	-1.99302	0.052345

The three parameters which remained in the stepwise regression produced an R^2 of 0.23674. This slightly improved the previous model, although the exclusion of P and pH from the stepwise regression may have compromised the predictive potential of the model. Using a standard regression equation (which includes all selected parameters in the model) with the specified transformations produced an ill-fitting matrix for which a solution could not be found. However, by removing the 1/X and X^3 parameters from the model and using a standard regression model, the R^2 was improved slightly to 0.26057. This increase was interesting, since the OM^3 parameter was significant in the previous stepwise model, yet its exclusion from the standard model did not decrease its performance. Figure 9 shows the observed yield values against the residuals for this model. Based on the performance of this final model, the following parameters were used for field GF120: OM, P, pH, OM^2, P^2, and pH^2.

Figure 9. Observed Yield vs. Residuals from Regression in GF120.



Delineation of Crop Production Zones Using Cluster Analysis

Overview

In an attempt to partition each field into zones in which the crop production capacity is relatively constant (or zones in which crop-influencing parameters are homogeneous), multi-dimensional cluster analysis was used. Each dimension in the analysis was one of the parameters in the regression model computed for each field, as shown above. The k-means clustering algorithm requires that the analyst specify prior to the analysis the number of clusters into which to partition the data; thus, exploratory analyses were necessary to determine exactly how many clusters into which the fields were to be partitioned. Ideally, it would be possible to partition each field into enough clusters to separate the inherent variability in each field, while producing spatially continuous regions within the fields.

However, it became necessary to examine both objective mathematical indices of the algorithm's performance as well as a more subjective visualization of the results, in an attempt to assess the appropriate number of clusters to use. The quantitative indicator of performance was the average euclidean distance between cluster means. In order to compute a euclidean distance in k-dimensional space, the following equation was used (Statistica User's Guide, 1995):

$$\text{distance}(x,y) = [\Sigma_i(x_i - y_i)^2]^{1/2}$$

where x and y are the means of two clusters. By computing this distance for all possible cluster mean pairs, it is possible to compute the average distance between any two clusters produced by the analysis.

It must be noted that as the number of clusters increases, the average distance between cluster means will also tend to increase. This phenomenon arises because the means of clusters with just a few partitions represent wider variation in data than cluster means with many partitions. As a result, with more partitions, each cluster mean becomes more representative of its cluster, and thus less influenced by extraneous observations (i.e., observations with unique combinations of variable levels). Thus, included in the quantitative analysis of algorithm performance is a measure of the change in distance, or ΔD, of each successive number of clusters over the previous number of clusters. This measure indicates the relative amount of information gained by adding additional clusters; in other words, it helps to reveal the point of diminishing marginal returns of computing more clusters.

Qualitative assessment of algorithm fitness was determined by joining a variable representing cluster membership into each original observation from each field's parent surface in a GIS. The resulting map of clusters for each field reveals the patterns that these clusters assume. A cluster map which produces large, contiguous zones was deemed preferable to a map which produces a "salt-and-pepper" effect, since crop production potential should involve some degree of spatial dependency. A lack of cohesiveness in the clusters would imply random distribution rather than a spatially predictable pattern, and would thus undermine the logic of producing zones which represent homogeneity. The quality of the analysis of yield data, which will follow the cluster analysis, depends upon the quality of blocking strata, so spatial pattern was an issue in the selection of a number of clusters.

Cluster Analysis in Field GF76

The regression model, computed above for field GF76, included only the original data variables, and none of the transformed variables: thus, only P, pH, OM, and K were used in the cluster analysis for field GF76. It was necessary to standardize these variables, as they are measured on different scales and with different units (pH is logarithmic, OM is a percent, P and K and in parts per million). Standardization ensures that each variable will be weighted equally in the analysis. The results of running the k-means algorithm for 2,3,4,5,6, and 7 clusters are summarized in Table IX.

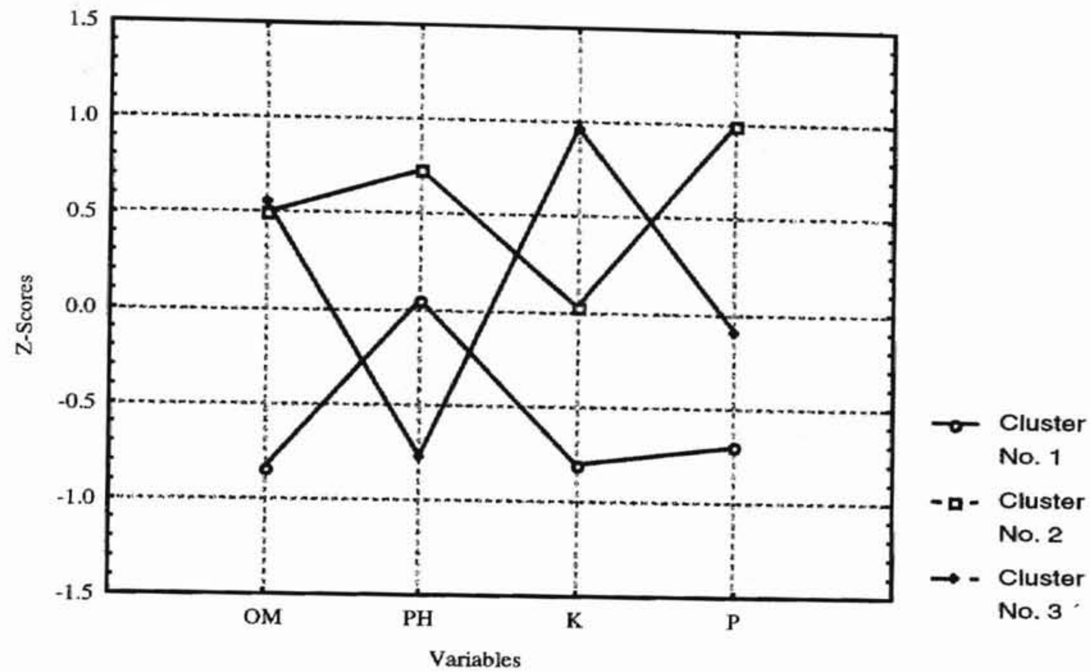
TABLE IX

DISTANCES BETWEEN CLUSTER MEANS FOR DIFFERENT NUMBERS OF CLUSTERS IN GF76

Number of Clusters	Average Distance Between Cluster Means	Change in Distance from Previous Number of Clusters
2	1.11122	N/A
3	1.16808	0.05685
4	1.21579	0.04771
5	1.25429	0.03851
6	1.2687	0.01441
7	1.3111	0.0423

The average distance between clusters increases with each additional cluster; however, after adding the third cluster, the additional information gained by adding clusters drops. When mapped, 2 and 3 clusters produce well-defined regions, while each additional increase produced isolated blocks and fragmented regions within the field. Thus, although it appears that the change in distance is rebounding at 7 clusters, with a ΔD of 0.0423, compared with the 3 cluster addition of 0.05685, the spatial behavior of this many clusters becomes almost unmanageable in an analysis context. As a result, it was decided to use 3 clusters for field GF76. Figure 10 shows the means for each soil parameter in each cluster.

Figure 10. Soil Parameter Means Within Each Cluster for GF76



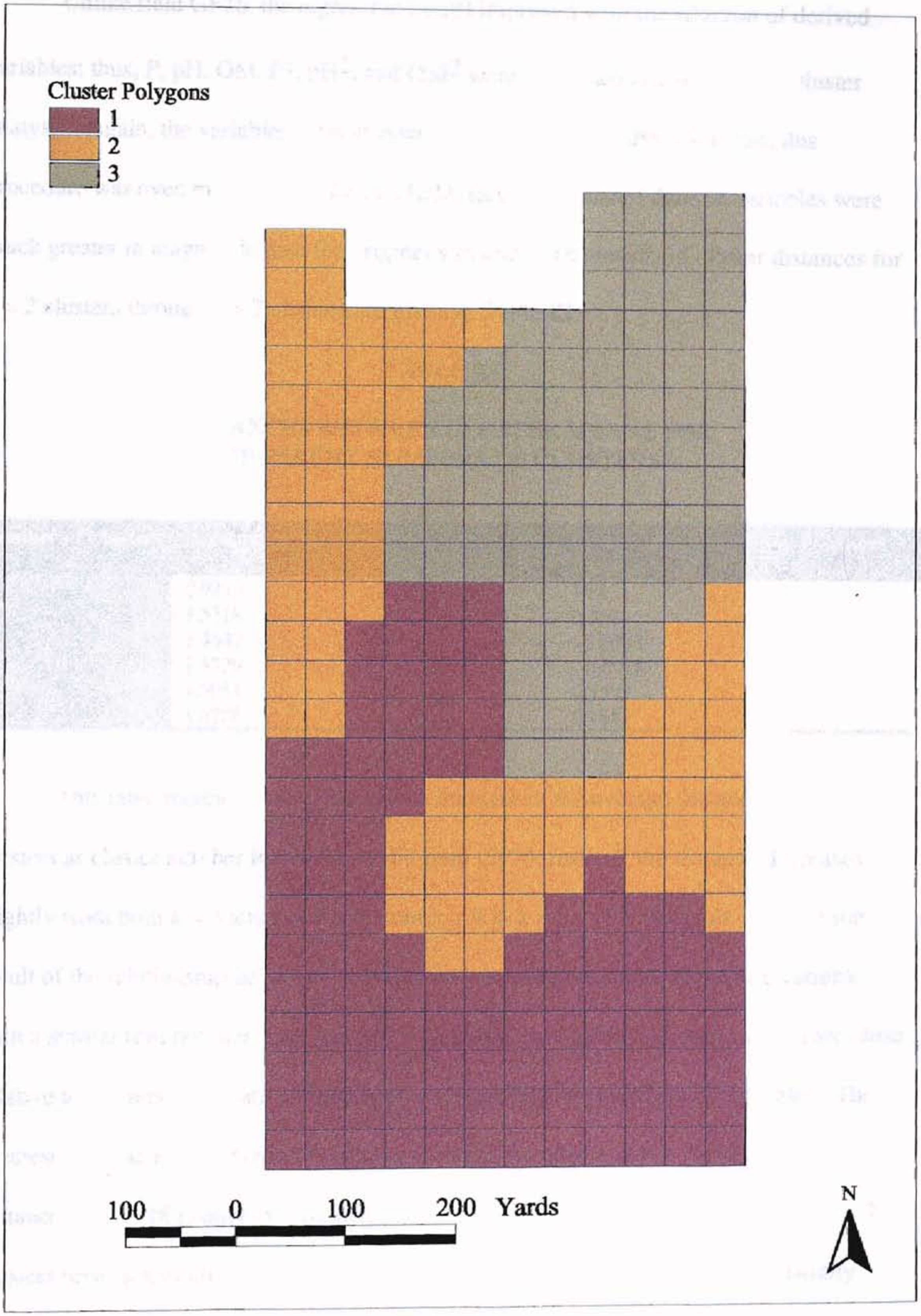
The euclidean distances and squared euclidean distances, which were used in the calculation of the average distance between clusters, are given in Table X, which shows those distances for each pair of clusters in field GF76, with squared distances above the diagonal and distances below the diagonal. Figure 11 shows the actual resulting soil clusters map for field GF76.

TABLE X

CLUSTER DISTANCES FOR FIELD GF76

Cluster Number	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0.0	1.475449	1.54512
Cluster 2	1.214681	0.0	1.095209
Cluster 3	1.243029	1.046522	0.0

Figure 11. Soil Clusters for Field GF76



Cluster Analysis in Field GF120

Unlike field GF76, the regression model improved with the addition of derived variables; thus, P, pH, OM, P^2, pH^2, and OM^2 were computed and used in the cluster analysis. Again, the variables were standardized before the analysis was run; this procedure was even more critical for this field, since the squared derived variables were much greater in magnitude than the original variables. The results of cluster distances for k = 2 clusters through k = 7 clusters are given in Table XI.

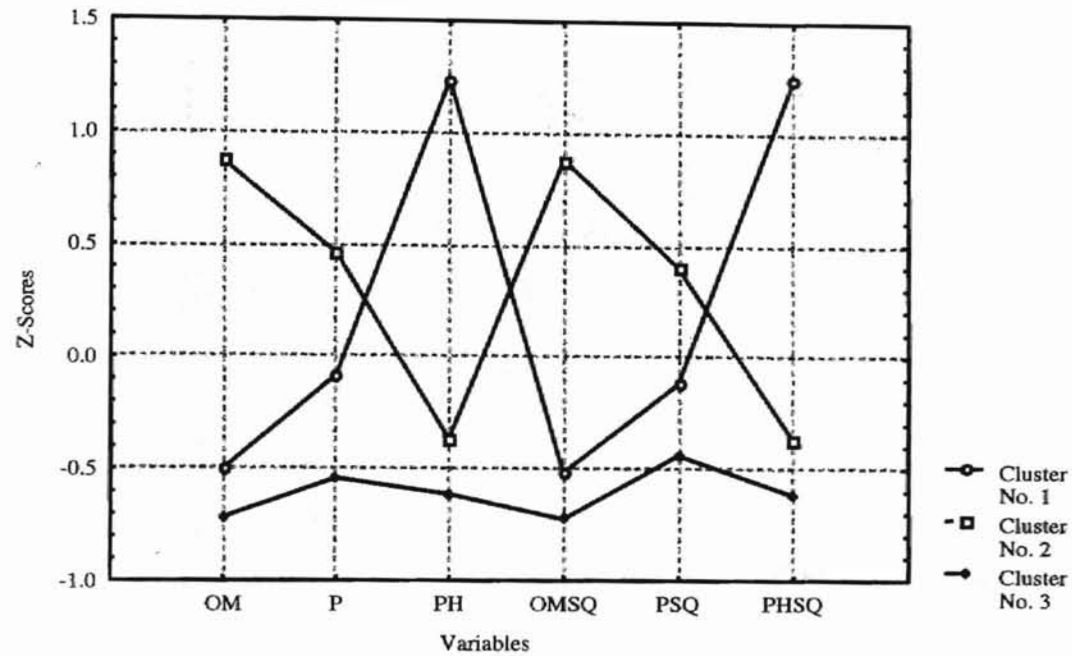
TABLE XI

DISTANCES BETWEEN CLUSTER MEANS FOR DIFFERENT NUMBERS OF CLUSTERS

Number of Clusters	Average Distance Between Cluster Means	Change in Distance From Previous Number of Clusters
2	0.9873	N/A
3	1.5218	0.5343
4	1.4447	- 0.0771
5	1.3729	- 0.0718
6	1.5051	0.1322
7	1.6208	0.1157

This table indicates not a continuous increase in the average distance between clusters as cluster number increases, as did field GF76; instead, the distance decreases slightly from both k = 3 and k = 4 and from k = 4 to k = 5. This behavior could be the result of the relationship between a variable and its transformation; clustering variables with a similar (but not identical) pattern of variation may produce clusters which are close relative to clusters which are derived from a completely unrelated set of variables. The greatest increase in information by adding a cluster number is at k = 3; the average distance of 1.5218 is only exceeded at k = 7; however, as in field GF76, a map of k = 7 clusters reveals spatially fragmented areas within the field, with much more spatially separate cluster distances; thus, it was decided to use a cluster number of 3 in field GF120 as well. Figure 12 shows the means for each soil parameter (and derived parameter) for each cluster; Figure 13 shows the resulting soil clusters map for field GF120.

Figure 12. Soil Parameter Means Within Each Cluster for GF120



The cluster lines for cluster 1 and 2 cross at variable pH and pH^2; this indicates that locations in the field with high nutrient levels tend to have lower pH values. Also, the points along any line for a variable and its square are, when transformed, identical; thus, it appears that although the squares of the variables improved the fit of the regression model, they did little to help in the delineation of crop production zones. The distances and squared distances between each cluster are given in Table XII; again, squared distances are above the diagonal, and distances are below.

Figure 13. Soil Clusters for Field GF120

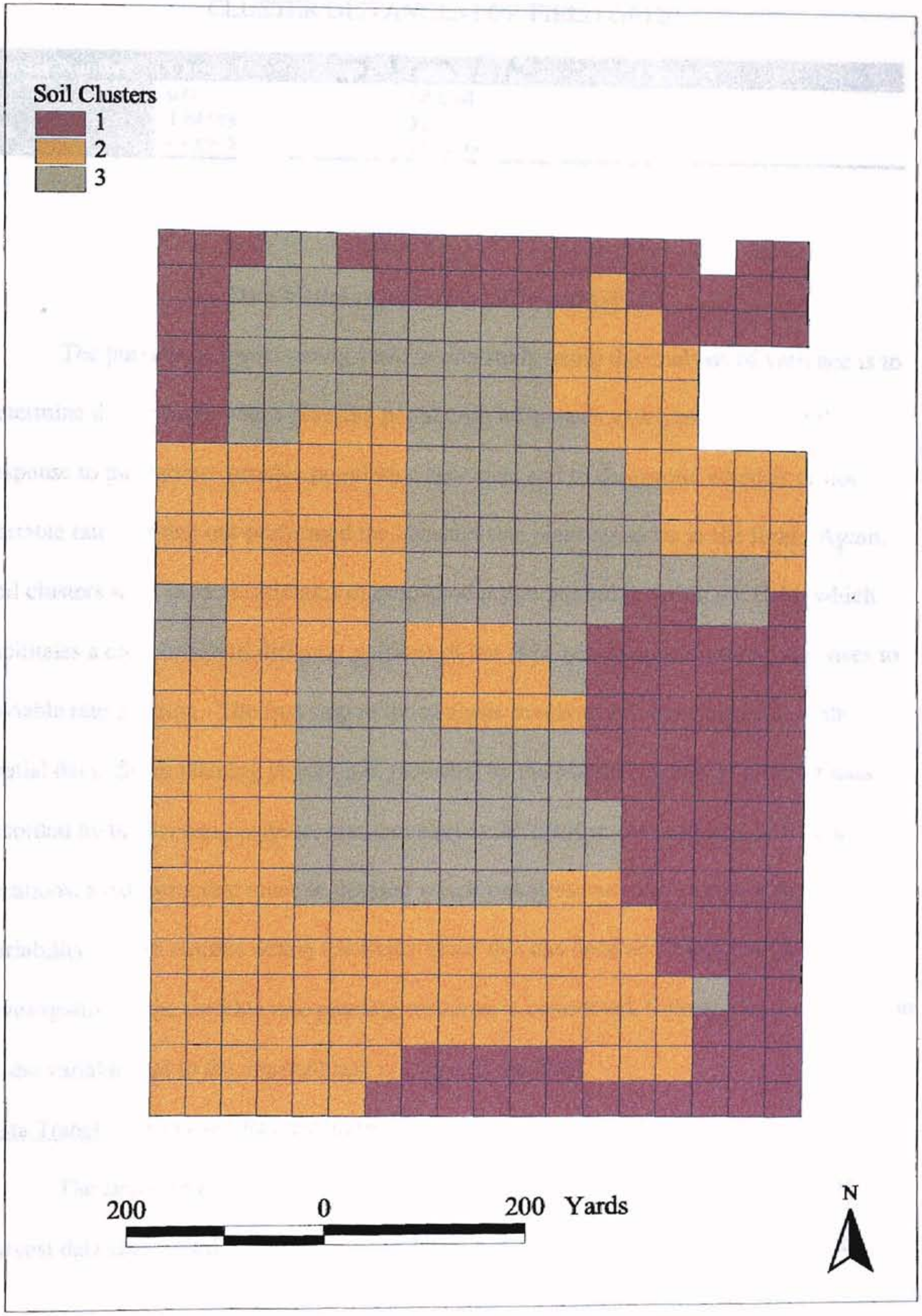


TABLE XII

CLUSTER DISTANCES FOR FIELD GF120

Cluster Number	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0.0	3.405647	0.965658
Cluster 2	1.845439	0.0	3.017745
Cluster 3	0.982679	1.737166	0.0

Data Modeling in Fields GF76 and GF120

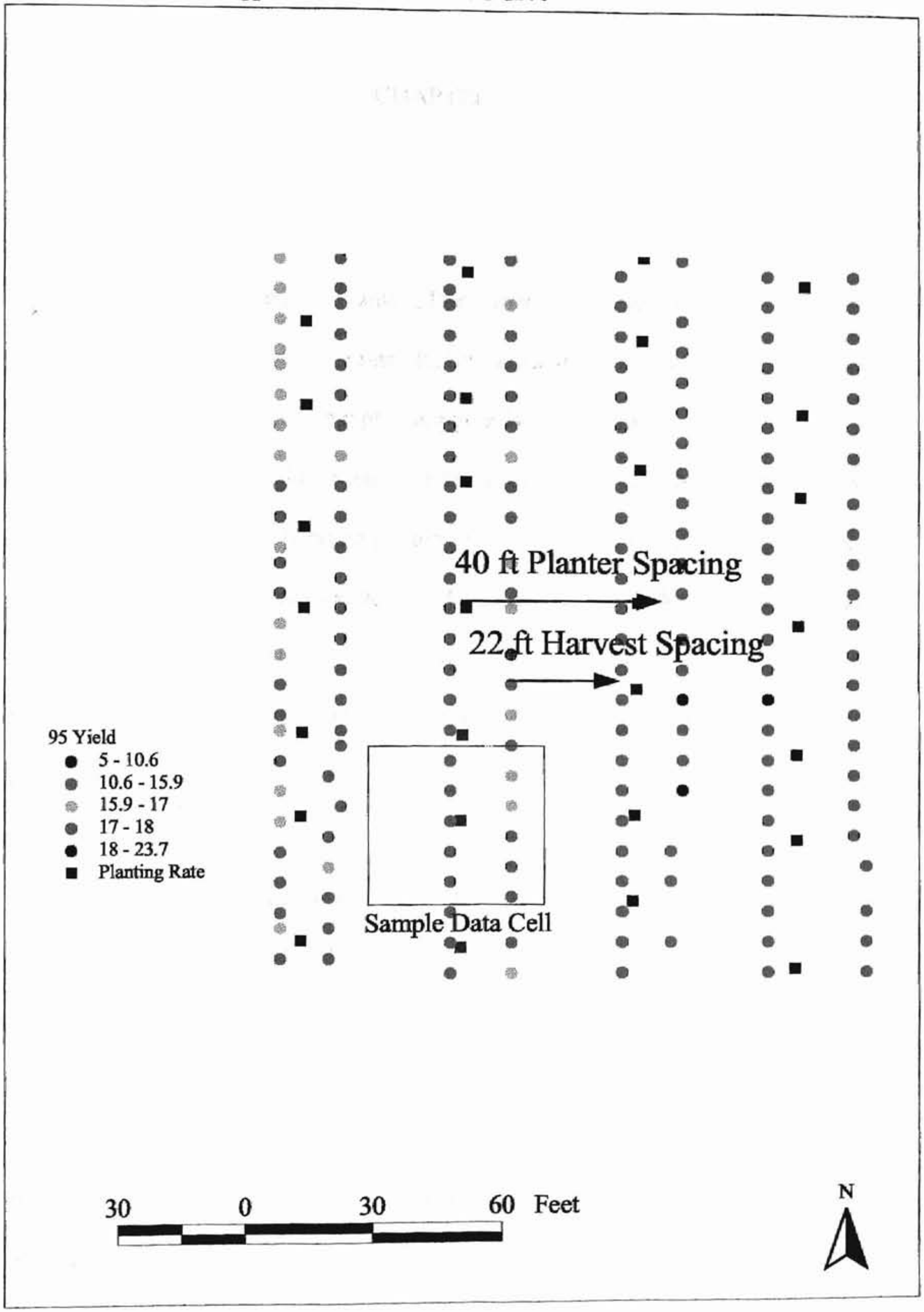
The purpose of investigating yield in this study using the analysis of variance is to determine the extent to which planting population influences yield, based on yield's response to the various variable population schemes, and to determine whether or not variable rate planting out-performed the constant rate-planting strips in the field. Again, soil clusters were used as estimates of crop-production potential within the field, which facilitates a comparison of different portions of the field with respect to their responses to variable rate planting. The first step in the analysis involves the representation of the spatial data. Since planting population, recorded by the planter, and yield monitor data, recorded by the harvest combine, was recorded at different resolutions and at different locations, a data structure must be devised which can simultaneously capture the variability of both entities within the field. Once this has been accomplished, an investigation of the variable rate planting response is conducted, followed by a comparison of the variable rate to the constant rate portions of the field.

Data Transformation and Representation

The first step in the ANOVA, as noted above, is to transform both planter and harvest data such that their variability can be characterized spatially. The discrepancy between the two data sources is illustrated in Figure 14; notice that the planter width, averaging about 40 feet, is roughly twice that of the harvest swath width, which averaged about 22 feet. This is actually quite advantageous to the analysis; since the harvest width is half of the planter width (32 rows per planter swath vs. 16 rows per harvest swath), it is ensured that there will be "pure" harvest data for each of the constant rate planting strips. In other words, the swath width of the combine will, on at least one pass, fall completely within each of the control strips, so that there will be no mixing in that row of constant and variable rate yield data.

To accommodate both data sources, square cells were created for each of the planting population points, with the point being located at the center. The size of the cells was chosen to be 40 feet, since yield passes would be 20 feet on either side of a given planter pass. Once these cells were created, the yield data were averaged within each cell, so that each planter point's polygonal cell also carried a dry yield estimate, measured in bushels per acre (bu/Acre). The cells were then joined spatially to the clusters, in order to determine which soil cluster each cell fell within. Having accomplished this, the resulting data structure was comprised of 40ft wide cells, each of which carried planting population, average yield, and soil cluster as attributes. These cells were used in all subsequent analyses.

Figure 14. Planting Rate Application vs. Yield In Field GF76



CHAPTER V

DATA ANALYSIS

Variable Rate Planting Performance in Field GF76

The performance of variable rate planting in field GF76 was investigated using analysis of variance to describe how planting population and soil cluster influenced yield, as well as to discover if soil cluster and population interacted to influence yield. Next the control strips and their corresponding population in the rest of the field were analyzed using ANOVA and multiple comparisons to determine if their yields were significantly different. Finally, the control strips were analyzed together to determine if they were significantly different in yield than the rest of the field. Recall that the planting population in field GF76 was varied based upon the previous year's grain moisture, as recorded by a yield monitor.

The yield data by population was characterized spatially using the process outlined above. A forty-foot neighborhood was defined around each of the planting population points, as recorded by the planter, and yield points were averaged within each neighborhood and then assigned to the appropriate population point. Next, the soil clusters were spatially joined to the planting population points, giving a soil cluster ID number for each population point, which represented the soil conditions at that location. These data were then exported to Statistica™ for data analysis. An initial analysis of variance was run for planting population in the field as a whole, excluding control strips; the null hypothesis in this analysis is that there is no significant difference in crop yield between planting rates. (NOTE: For all of the ANOVA procedures that follow, a significance level of 95% will be used). The results of this run are given in Table XIII.

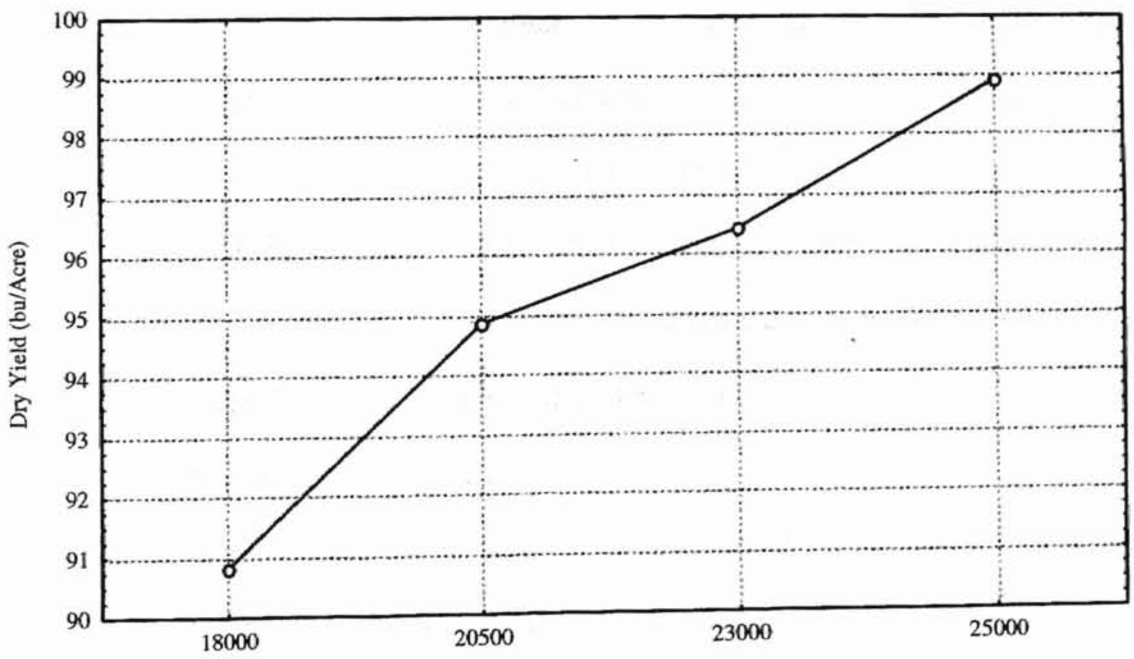
TABLE XIII

POPULATION RATE ANOVA RESULTS IN FIELD GF76

Population DF	*Error DF*	*F Statistic*	*Population MS*	*Error MS*	*P Level*
3	1520.0	31.32393	5174.044	165.1786	0.0000

As this table shows, at the 95% level of confidence, we do in fact detect a significant difference in the yield between the population rates. Figure 15 shows the population rates in field GF76 and their average.

Figure 15. Average Yield by Planting Rate in Field GF76



A post-hoc comparison of seed population rates was then run, using the multiple comparison technique of least-squares differences, in an attempt to determine which populations were significantly different from each other. The results of the LSD are summarized in Table XIV.

TABLE XIV

MULTIPLE COMPARISON OF SEEDING RATES IN FIELD GF76

Population Rate	{1} 92.2 bu/Acre	{2} 94.8 bu/Acre	{3} 96.6 bu/Acre	{4} 99.9 bu/Acre
18000 {1}	----	0.005794	0.000001	0.0
20500 {2}	0.005794	----	0.088175	0.000001
23000 {3}	0.000001	0.088175	----	0.000862
25000 {4}	0.0	0.000001	0.000862	----

This table lists the p-values for a comparison between each population rate and all other rates. This table shows that there were not significant differences between rates 2 (20,500) and 3 (23,000); however, rate 1 was different from 2 and 3, and rate 4 was different from rate 1 and rates 2 and 3. Interestingly, the distance in seeding rates between rates 2 and 3 was 2,500 seeds per acre, and no difference was found between the two, while the distance between rates 3 and 4 was only 2,000 seeds per acre, and a difference was detected in yield between those rates. This implies that perhaps the variable seeding rates chosen were in the critical range in which slight changes can produce noticeable changes in yield; this finding lends credence to the assumption that managing seeding rate variably in a field can produce different crop responses in a field.

The three soil clusters in field GF76 were analyzed next to determine if there was a significant yield response by the derived soil clusters; these yields are given in Figure 16 below. The results of the ANOVA run are given in Table XV.

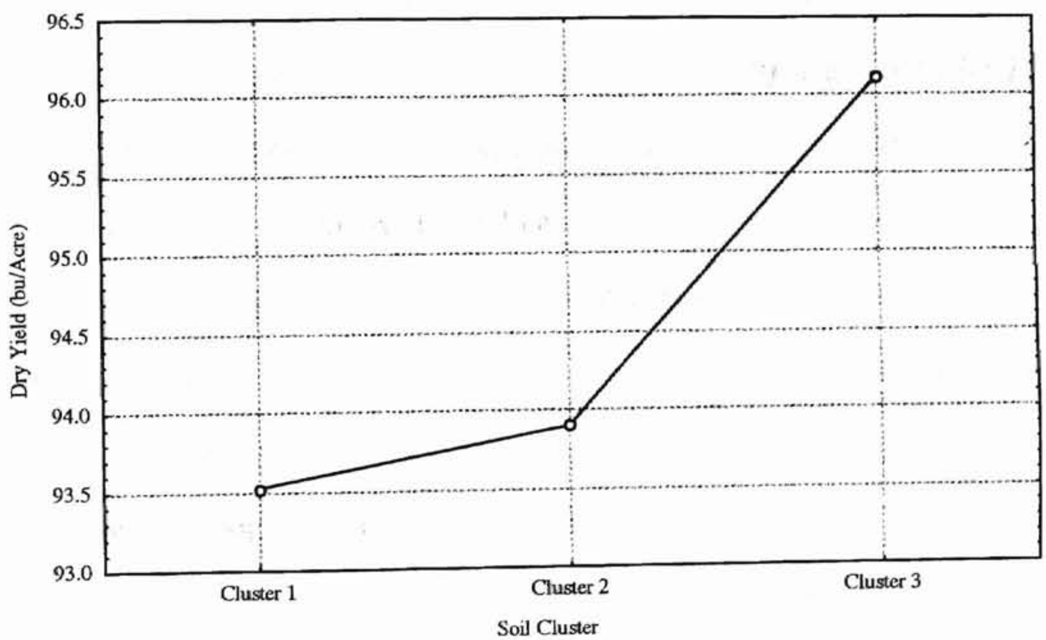
TABLE XV

ANOVA RESULTS FOR SOIL CLUSTERS IN FIELD GF76

Cluster DF	*Error DF*	*F Statistic*	*Cluster MS*	*Error MS*	*P Level*
2	1521	5.413699	942.1804	174.0363	0.004541

These results show that there is in fact a significant difference in yield by soil cluster at the .05 level of significance. This means that, since neither the soil clusters nor their constituent soil properties were used as criteria for varying the seeding rates in GF76, there are areas in the field with inherently different crop production potentials.

Figure 16. Average Yield by Soil Cluster in Field GF76



Again, a multiple comparison was run to explore the soil clusters and determine which clusters were different from each other. The LSD procedure was used, and the results are listed in Table XVI.

TABLE XVI

MULTIPLE COMPARISON OF SOIL CLUSTERS IN FIELD GF76

Soil Cluster	Cluster 1 93.6 bu/Acre	Cluster 2 93.8 bu/Acre	Cluster 3 96.2 bu/Acre
Cluster 1	----	0.641756	0.001824
Cluster 2	0.641756	----	0.010162
Cluster 3	0.001824	0.010162	----

These results corroborate what Figure 16 shows; clusters 1 and 2 are not significantly different from each other, but cluster 3 is different from 1 and 2. Cluster 3 has an average yield that is more than 2 bu/Acre higher than that of cluster 2; this difference is not quite as stark as the difference between seeding rates, which was as much as 3.1 bu/Acre (see Figure 15). However, since a significant difference was found, clusters 1 and 2 will be treated as a single entity in the remaining analyses for field GF76, and cluster 3 will be treated as a separate entity. This gives us two clusters to analyze, which will be referred to as simply cluster 1 and cluster 2.

Before moving on to the comparison of control strips to the rest of the field, it would be illustrative to investigate if there was a significant interaction between seed population and soil cluster in the field. A multi-factor ANOVA was run, with clusters (using all three) and population rates as main effects; this ANOVA also analyzed the interaction between the main effects. If a significant interaction is found in such a design, then the level of yield within a soil cluster is dependent upon which population was planted in it, and vice versa. The results of the multi-factor ANOVA are given in Table XVII.

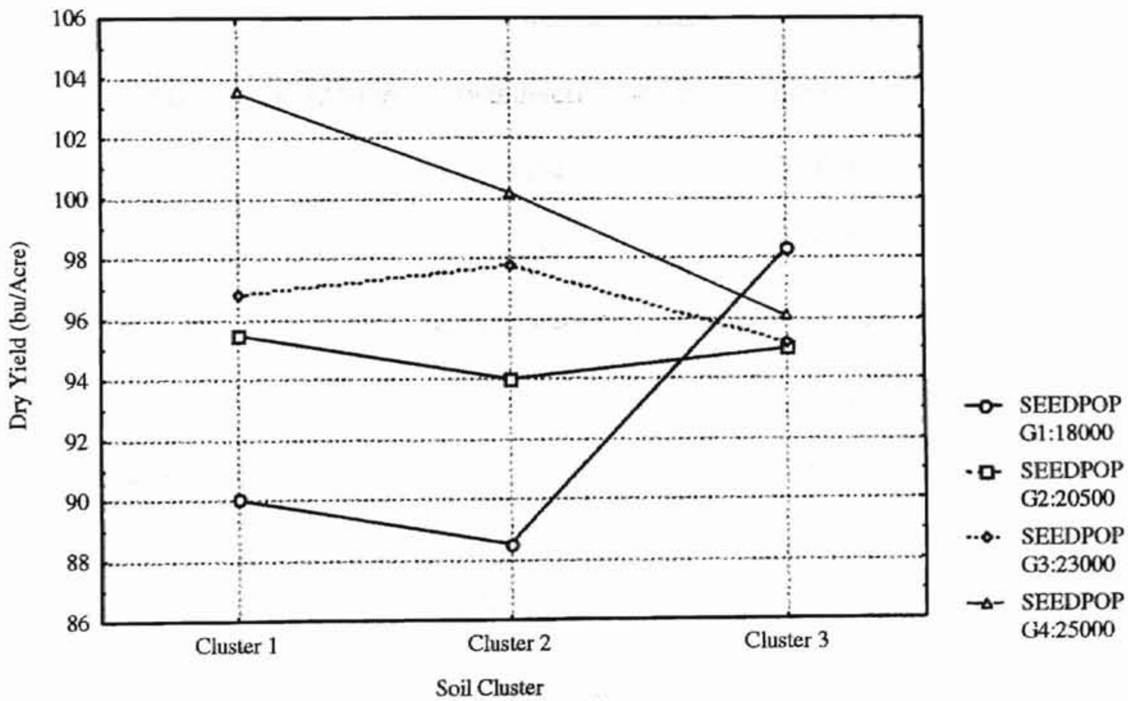
TABLE XVII

MULTI-FACTOR ANOVA IN FIELD GF76

Effect	Effect DF	Error DF	F Statistic	Effect MS	Error MS	P Level
Population	3.0	1512.0	22.67709	3620.023	159.6335	0.0000
Cluster	2.0	1512.0	1.31424	209.797	159.6335	0.268984
Interaction	6.0	1512.0	9.92285	1584.019	159.6335	0.0000

According to these results, there is a significant interaction between soil cluster and planting rate. This interaction is shown graphically in Figure 17.

Figure 17. Interaction Between Soil Clusters and Population Rates



This graph shows that, while populations 20,500, 23,000, and 25,000 do not have lines that cross by soil cluster, seed population 18,000 does cross the other lines in cluster 3. This means that the performance of seeding rate 18,000 plants per acre (which is significantly different from the other rates) is dependent upon the soil cluster that it is planted in.

Constant Rate vs. Variable Rate Planting in Field GF76

The two constant rate strips planted in field GF76 were at populations of 20,000 and 25,000 seeds per acre; the 25,000 rate was also planted variably in the field as a whole. Although the 20,000 constant rate has no counterpart in the field as a whole, there do exist rates of 18,000 and 20,500 in the variable portions of the field, thus facilitating a comparison of the constant strip with variable rates both higher and lower in the field. In addition, the soil clusters in which different rates were planted can be investigated in order to determine if spatial positioning of the management practices was a factor in explaining yield differences. Each of these comparisons attempts to address the issue of whether or not varying the seed population is advantageous within a single farm-field.

An initial summary of yield by management practice in the field is given in Table XVIII. Here, each constant rate and each variable rate applied in the field is listed, with its average yield. Notice that here, each constant rate strip performed lower than its analagous variable rate portion of the field. This helps illustrate how the management practices performed overall, before proceeding with the analysis of variance and comparison of seeding rate practice versus soil production potential.

TABLE XVIII

AVERAGE YIELD BY MANAGEMENT PRACTICE IN FIELD GF76

Seed Population (seeds/Acre)	Management Type	Average Yield (bu/Acre)
18000	Variable	91.40396
20000	Constant #1	89.90929
20500	Variable	95.21201
23000	Variable	96.76443
25000	Constant #2	87.58258
25000	Variable	98.57137

To determine if significant differences exist between each constant zone and its variable counterpart, analysis of variance was used initially with a code representing the combination of seed population and management type (e.g., variable 25,000, constant 25,000) used as the main effect. The results of this analysis are summarized in Table XIX.

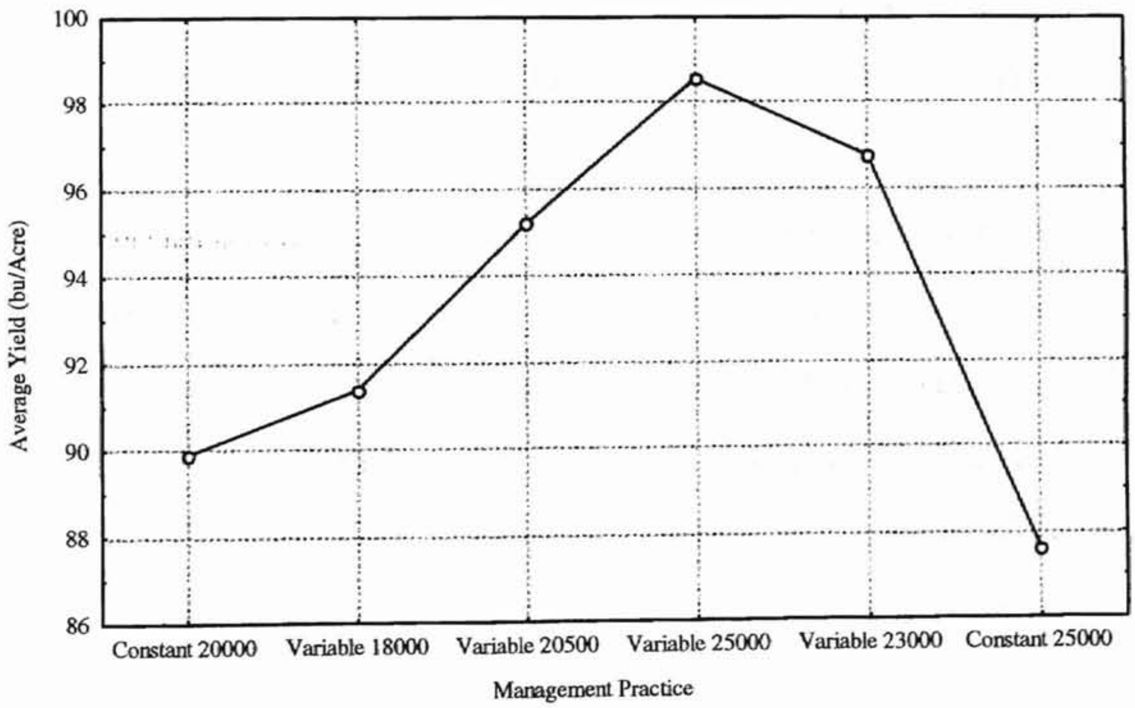
TABLE XIX

ANOVA OF MANAGEMENT PRACTICES IN FIELD GF76

Management DF	*Managment MS*	*Error DF*	*Error MS*	*F Statistic*	*P Level*
5.0	5414.079	2641.0	110.2815	49.09327	0.0000

As this table shows, we are 95% confident that there is a significant difference in the yield within GF76's management types; the average yields are shown in Figure 18.

Figure 18. Average Yield by Management Practice in Field GF76



To determine which management practices were significantly different, the LSD multiple comparison was used. The results of the multiple comparison are given in Table XX.

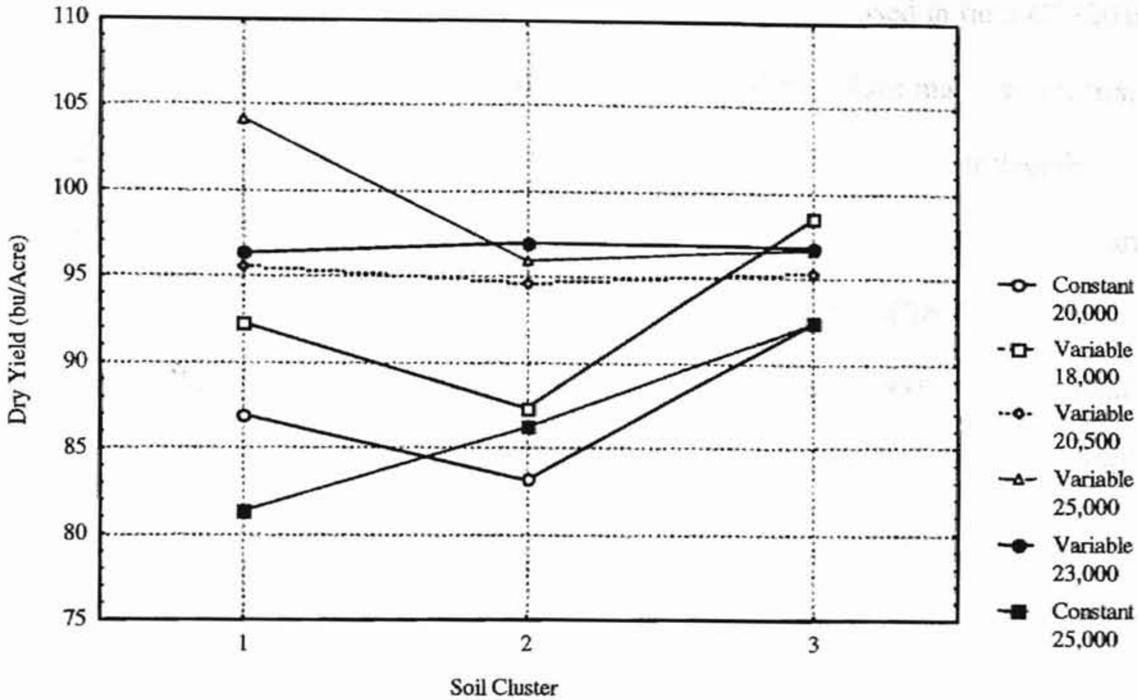
TABLE XX

MULTIPLE COMPARISON OF MANAGEMENT PRACTICE YIELDS IN FIELD GF76

Management Practice	1 (89.9)	2 (91.4)	3 (95.2)	4 (98.5)	5 (96.7)	6 (87.5)
Constant 20000 (1)	----	0.183143	0.000005	0.0	0.0	0.121999
Variable 18000 (2)	0.183143	----	0.0	0.0	0.0093	0.000713
Variable 20500 (3)	0.000005	0.0	----	0.0	0.016005	0.0
Variable 25000 (4)	0.0	0.0	0.0	----	0.003623	0.003865
Variable 23000 (5)	0.0	0.0093	0.016005	0.003623	----	0.0
Constant 25000 (6)	0.121999	0.000713	0.0	0.003865	0.0	----

Notice that at the .05 level of significance, constant rate 20,000 is not different from constant rate 25,000 or variable rate 18,000. Other than those three management practices, which produce similar yields, all other practices have significantly different results. The variable rates in field GF76 clearly out-performed the constant rate strips of the field, with the exception of the lowest variable rate, 18,000 seeds per acre. The next logical question to ask of the data involves the inherent yield potential of the locations of the management practices. Recall that soil cluster 3 had a significantly higher yield than clusters 1 and 2. If any management practice resides predominantly in soil cluster 3, then perhaps the increased yield performance was explained by the soil conditions rather than the management practice applied. To investigate this possibility, the average yield for each unique population/cluster combination was computed and sorted by cluster. These results are shown in Figure 19.

Figure 19. Interactions Between Soil Clusters and Management Types



This graph reveals an interesting relationship; although overall, soilcluster 3 outperformed clusters 1 and 2, which were not significantly different in yield, at the lower populations, this finding is supported, but at higher density populations (> 20,500) it no longer holds. This implies that there is an optimal population for different soil clusters as well as for different soil moisture holding capacities. This fact becomes especially obvious for population 25,000, in which the average yield for cluster 3 was 96.167, and the average yield for cluster 1 was 100.889, the highest in the field. However, of all the locations in the field that were planted at a rate of 25,000 seeds per acre, 22.41% were in cluster 1, while 44.10% were in cluster 3.

Variable Rate Planting Performance in Field GF120

In a general sense, the same methodology for analysis was used in field GF120 that was applied to GF76. However, this field differed from GF76 in three major ways; first, the criterion for varying planting population was measured soil depth, rather than the previous year's grain moisture level. Secondly, the degree to which population was varied was much greater in field GF120; whereas the seeding rate in field GF76 ranged from 18,000 to 25,000 plants per acre, the rate for GF120 varied from 15,000 to 30,000 plants per acre. Finally, instead of two control strips, three strips were planted in field GF120 at rates of 20,000, 25,000, and 30,000 plants per acre. This facilitates a more rigorous comparison of the performance of single-rate versus variable rate seeding practices.

The methodology for analyzing field GF120 nevertheless follows closely that of GF76; the seeding population was analyzed using ANOVA, as were the soil clusters. The interaction between soil cluster and population was then investigated, followed by a comparison of the three control strips to the variable rate portions of the field. This step was modified somewhat to accomodate the wealth of control strip data available in field GF120. The next and final step in the analysis is a comparison of the performance of variable rate planting between field 76 and 120.

The seeding populations in the variable portions of field GF120 were planted at rates of 15,000, 18,000, 20,000, 22,000, 24,000, 25,000, and 26,000. However, these population rates were not applied to even proportions of the field; some were only applied to very small areas within field GF120. This could potentially skew the results; if only a few observations were logged for a certain population, then comparing its yield to that of another population with many points gives the smaller population's yield data an over-representation. Table XXI shows the number of observations logged from the planter for each planting rate applied variably to field GF120.

TABLE XXI

NUMBER OF OBSERVED PLANTING SITES FOR DIFFERENT RATES IN GF120

Planting Population (seeds/Acre)	*Number of Observations*
15000	14
18000	239
20000	1884
22000	1090
24000	1312
25000	3
26000	1063

Notice that for rates 15,000, 18,000, and 25,000, less than 1,000 observations were logged. As a result of this discrepancy, it was decided to only use rates 20,000, 22,000, 24,000, and 26,000 in the analysis. Another advantage of using these particular rates is that they are all 2,000 seeds apart; thus, the problem of comparing classes of unequal increases in seeding rate is also circumvented. An ANOVA procedure was run on these population rates to determine whether or not populations yielded differently; the results of this analysis are summarized in Table XXII.

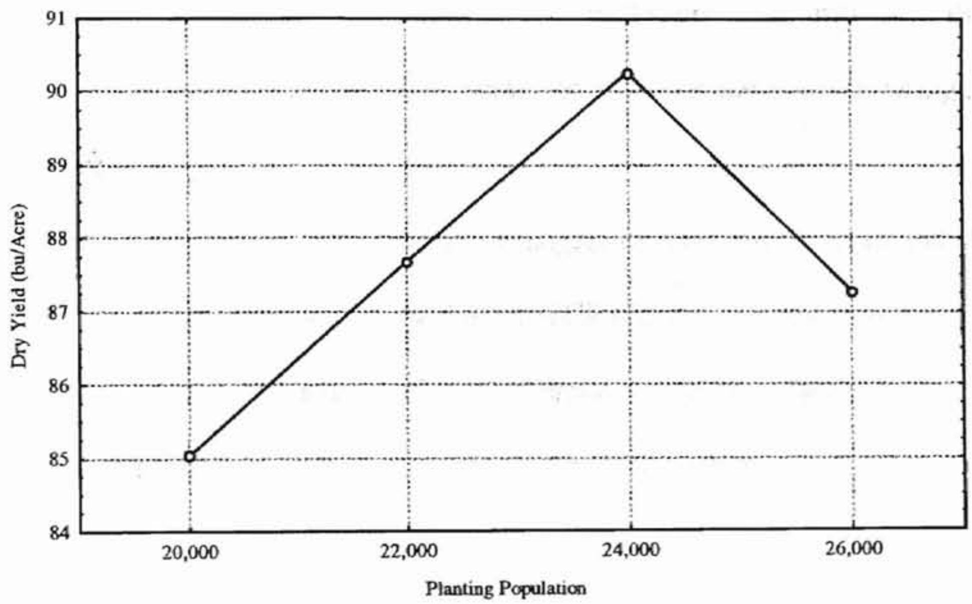
TABLE XXII

POPULATION RATE ANOVA RESULTS IN GF120

Effect DF	*Effect MS*	*Error DF*	*Error MS*	*F-Statistic*	*P Level*
3.0	7145.583	5345.0	139.2804	51.30359	0.0000

At the .05 level of significance, these results indicate that there is a significant difference in yield between some of the population rates. Figure 20 shows average yield within populations. Notice that there is a decrease in yield at a population of 26,000, which was the highest population included in the analysis; this indicates that perhaps within the distance between 24,000 and 26,000 plants per acre is the critical point beyond which increasing the population will exceed the field's growing capacity for corn.

Figure 20. Average Yield by Planting Rate in Field GF120



To delineate which populations differed significantly in yield from each other, a multiple comparison was run subsequent to the ANOVA. Again, LSD was the chosen method for comparing yield differences. Table XXIII summarizes the results of the multiple comparison.

TABLE XXIII

MULTIPLE COMPARISON OF SEEDING RATES IN GF120

Population	{1} 85.0	{2} 87.6	{3} 90.2	{4} 87.2
20000 {1}	----	0.0	0.0	0.000001
22000 {2}	0.0	----	0.0	0.430371
24000 {3}	0.0	0.0	----	0.0
26000 {4}	0.000001	0.430371	0.0	----

According to these results, there was no significant difference between populations 26,000 and 22,000; however, all other populations were different from each other. The highest yielding population, 24,000, produced a yield of 90.2 bu/Acre; increasing the population to 26,000 drops the yield to 87.2 bu/Acre, which is at the same level as 22,000 seeds per acre. Thus, even though rates 22,000 and 26,000 are not significantly different in their ability to produce yield, there are obvious economic benefits to operating at a lower seeding rate.

The next step in the analysis was to investigate the possible effect of soil cluster on yield. Again, three soil clusters were used in field GF120; an ANOVA was run to test the null hypothesis that there was not significant differences in yield between the three clusters. The results of this analysis are presented in Table XXIV.

TABLE XXIV

ANOVA RESULTS FOR SOIL CLUSTERS IN FIELD GF120

Effect DF	*Effect MS*	*Error DF*	*Error MS*	*F Statistic*	*P Level*
2.0	7927.614	5602.0	140.2723	56.5159	0.000

According to these results, there was in fact a significant difference in yield between soil clusters at the .05 level of significance. Again, this suggests that there is yield variability in the field explained by factors other than the variable rate seeding management practice. The average yield in each soil cluster is presented in Figure 21. To investigate which clusters were significantly different, the LSD procedure was run on soil clusters; these results are given in Table XXV.

Figure 21. Average Yield by Soil Cluster in Field GF120.

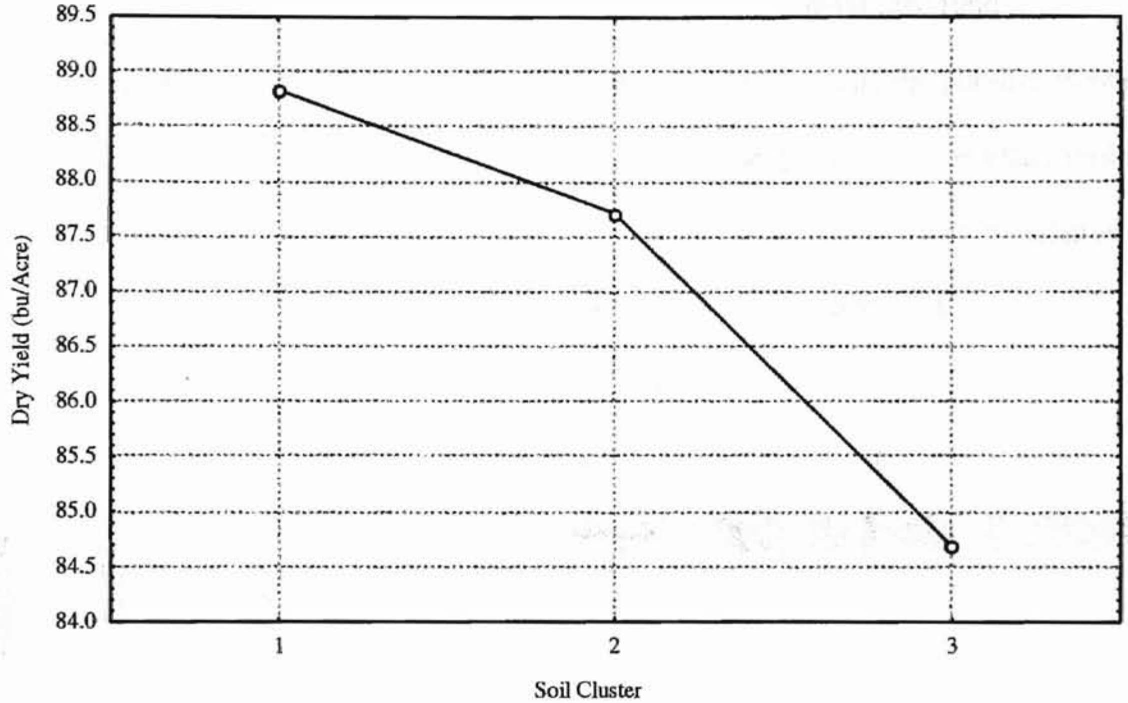


TABLE XXV

MULTIPLE COMPARISON OF SOIL CLUSTERS IN FIELD GF120

Soil Cluster	1 (88.8)	2 (87.7)	3 (84.6)
Cluster 1	----	0.0036	0.0
Cluster 2	0.0036	----	0.0
Cluster 3	0.0	0.0	----

According to this table, there was a significant difference in yield between all three soil clusters. This means that, unlike field GF76, field GF120 encompassed three distinct types of crop-growth potential. Recall that the clusters in field GF120 did not form large contiguous regions, as they did in field GF76, but rather produced smaller, more fragmented zones; this indicates that there was more inherent variability in the soil fertility levels in field GF120 than in field GF76.

The variability in soil parameters for field GF120 may manifest itself in complex interactions between the soil clusters and the population rates within the field. The possibility of such interactions, in which the performance of a particular population rate is dependent upon which soil cluster it fell within, was investigated by performing a multi-factor ANOVA on population rate, soil cluster, and the interaction between population rate and soil cluster. The results of this ANOVA are given in Table XXVI.

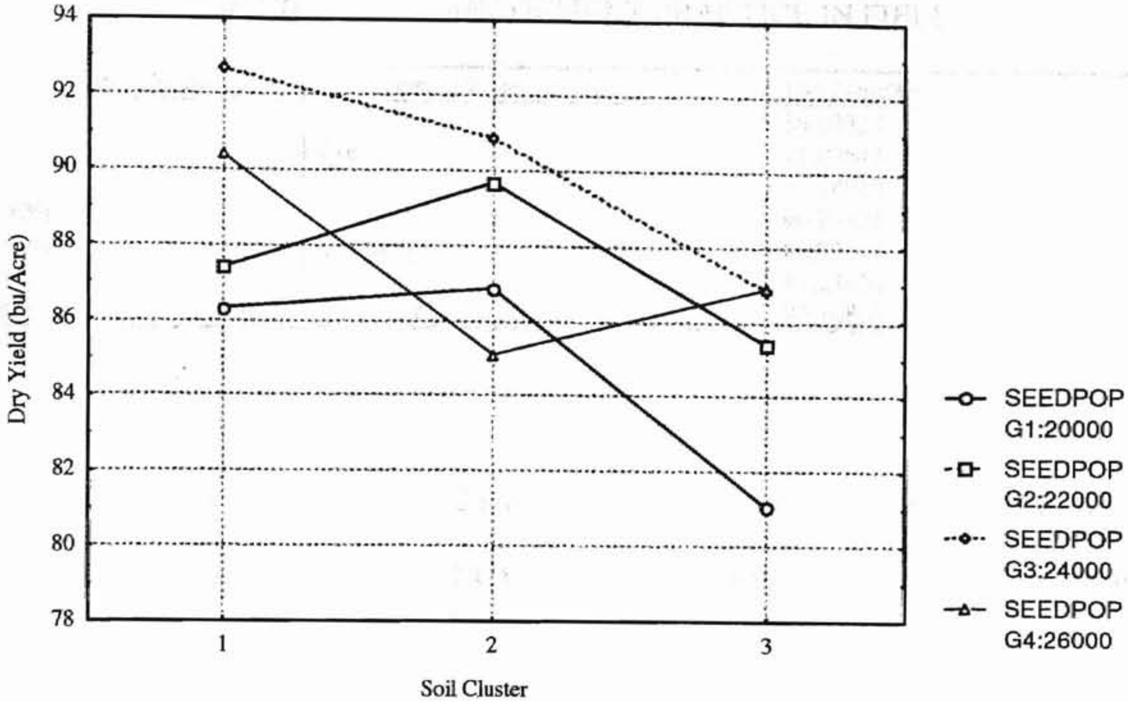
TABLE XXVI

MULTI-FACTOR ANOVA IN FIELD GF120

Effect	Effect DF	Effect MS	Error DF	Error MS	F Statistic	P Level
Population	3.0	6680.449	5337.0	133.9434	49.87516	0.0000
Soil Cluster	2.0	7251.067	5337.0	133.9434	54.13531	0.0000
Interaction	6.0	1733.361	5337.0	133.9434	12.94099	0.0000

The results of the multi-factor ANOVA indicate that there are, in fact, significant interactions occurring between soil cluster and seeding population in field GF120. This implies that, since soil clusters by themselves contained significantly different yields from each other, and since they interact with population density in their ability to produce yield, perhaps soil clusters should be taken into account when determining how to vary the seeding population for highly variable fields. Since the populations planted by soil depth produced significantly different yields, it would even be possible to recompute soil clusters, using soil depth as an additional variable in the cluster analysis. A visualization of the interactions between soil cluster and population is given in Figure 22.

Figure 22. Interaction Between Soil Clusters and Population Rates



Constant Rate vs. Variable Rate Planting in Field GF120

The next step in the analysis of field GF120 is to investigate how the three constant rate strips performed compared with the variable portions of the field. The constant rate strips were planted at seed densities of 20,000, 25,000, and 30,000 seeds per acre; 20,000 was also planted at a number of sites within the variable portion of the field. The 25,000 rate was also planted in both a control strip and in the variable portions of the field, although only a few observations were available in the field as a whole. The under-represented rates of 15,000, 18,000, and 25,000 variable rate were removed from the above analysis, they will also be removed from the comparison of constant rate to variable rate that follows. Table XXVII shows the average yield in all population/management combinations within field GF120.

TABLE XXVII

AVERAGE YIELD BY MANAGEMENT PRACTICE IN FIELD GF120

Planting Population	*Management Zone*	*Yield (bu/Acre)*
20000	Variable	85.03324
20000	Control #3	91.07641
22000	Variable	87.68045
24000	Variable	90.27324
25000	Control #2	80.93014
26000	Variable	87.27926
30000	Control #1	85.10889

The 20,000 control strip seems to out-perform the 20,000 variable observations, while the other two control strips at 25,000 and 30,000 performed at the lower end of field GF120's range of yield. An ANOVA was run to determine if there were significant differences between yield in different population/management combinations. The results of this analysis are presented in Table XXVIII.

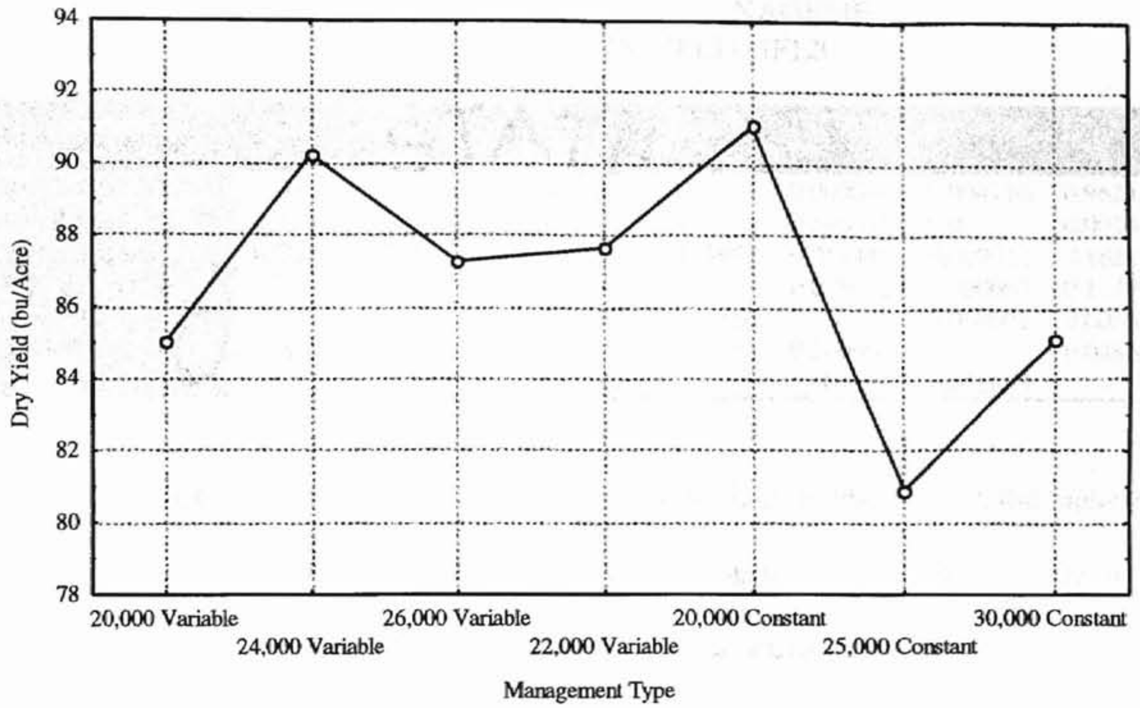
TABLE XXVIII

ANOVA OF MANAGEMENT PRACTICES IN FIELD GF120

Effect DF	*Effect MS*	*Error DF*	*Error MS*	*F Statistic*	*P Level*
6.0	4232.162	5529.0	136.016	31.11517	0.0

The results of the ANOVA indicate that at the .05 level of significance, there is a significant difference in the yield of different management schemes. A graph of the average yield within each population/management combination is shown in Figure 23. Notice that while the 20,000 constant rate performed well, the other two control strips yielded low compared to other population/management combinations.

Figure 23. Average Yield by Management Practice in Field GF76



The fact that the performance of 25,000 plants per acre at a constant rate performed lower than the higher populations of 26,000 and 30,000 indicates that the the yield in that strip was not low because the population carrying capacity of the field was exceeded, but rather because the varying conditions within that strip were not being addressed. Although this differs from the situation presented by GF76, in which there was a direct relationship between the seed population density and yield, it does indicate that for different locations within a field, there is an carying capacity which will support a maximum yield given a certain seeding population. A multiple comparison was run to determine which population/management combinations were significantly different from each other. The results of this analysis are presented in Table XXIX.

TABLE XXIX

MULTIPLE COMPARISON OF MANAGEMENT PRACTICE YIELDS IN FIELD GF120

Management Practice	{1} 85.0	{2} 90.2	{3} 87.2	{4} 87.6	{5} 91.0	{6} 80.9	{7} 85.1
20000 Variable {1}	----	0.0	0.000001	0.0	0.000046	0.004116	0.962519
24000 Variable {2}	0.0	----	0.0	0.0	0.590619	0.0	0.001435
26000 Variable {3}	0.000001	0.0	----	0.424891	0.011446	0.000012	0.182239
22000 Variable {4}	0.0	0.0	0.424891	----	0.023616	0.000003	0.113796
20000 Constant {5}	0.000046	0.590619	0.011446	0.023616	----	0.000001	0.005639
25000 Constant {6}	0.004116	0.0	0.000012	0.000003	0.000001	----	0.048653
30000 Constant {7}	0.962519	0.001435	0.182239	0.113796	0.005639	0.048653	----

In general, this table reveals three categories of yield in field GF120; the highest category included the 24,000 variable rate and 20,000 constant rate; the middle category included the 20,000 variable, the 22,000 variable, 26,000 variable, and 30,000 constant; the lowest category included only the 25,000 constant. Thus, there was no clear distinction between variable/constant in field GF120, but rather a distinction between low versus high population densities, with the lower densities out-performing the high. The only pair of management practices that were constant/variable with the same population was with population 20,000, and here the constant rate out-performed the variable rate; however, the constant rate of 25,000 was the lowest in the field, and the constant rate of 30,000 performed in the median yield category.

A possible explanation of yield differences other than population or constant/variable management may be the soil conditions within the field; recall that each of the three soil clusters for field GF120 was found to have a significantly different yield. One interesting point to note is that the constant rate strip of 25,000 seeds per acre was planted exclusively in clusters 2 and 3, and not in cluster 1, which had the highest yield;

this may explain the poor performance of that particular management regime. This fact precludes the use of a multi-factor ANOVA to attempt the determine which soil cluster/management type practices were higher than others, since all combinations of soil cluster and management type were not accounted for in the dataset. However, Table XXX displays the average yield within each cluster/management type combination.

TABLE XXX

AVERAGE YIELD WITHIN SOIL CLUSTER AND MANAGEMENT COMBINATIONS

Management Type	*Soil Cluster*	*Average Yield (bu/Acre)*
20000 Control #3	1	93.315
20000 Control #3	2	90.38812
20000 Control #3	3	90.63611
20000 Variable	1	86.31525
20000 Variable	2	86.87141
20000 Variable	3	81.07164
22000 Variable	1	87.41022
22000 Variable	2	89.69357
22000 Variable	3	85.42297
24000 Variable	1	92.6773
24000 Variable	2	90.84704
24000 Variable	3	86.85122
25000 Control #2	2	80.57325
25000 Control #2	3	81.42241
26000 Variable	1	90.41235
26000 Variable	2	85.1124
26000 Variable	3	86.89298
30000 Control #1	1	79.717
30000 Control #1	2	83.18889
30000 Control #1	3	87.14314

This table reveals that, in general, the 20,000 constant rate produced better than other management types regardless of soil cluster. The 24,000 variable rate did slightly better than 20,000 constant rate in soil cluster 2, and was nearly as productive in cluster 1. Interestingly, this table also shows that for the constant rates in the field, 30,000 had its highest yield within soil cluster 3, which was the lowest yielding soil cluster in the field, as did the 25,000 constant rate. Most of the variable rate portions of the field yielded higher in soil cluster 1. This tends to strengthen the notion that perhaps soil clusters should be accounted for when creating an application map for variable rate seeding practices.

Comparison of Variable rate Seeding Between GF76 and GF120

The two fields included in this experiment each had a unique criterion by which seeding rate was varied; this raises the question as to which method performed the best. The problem with comparing the yield performance in two different fields such as these is that each field has its own production capacity and expected performance; therefore, it becomes difficult to relate the yield observations from different fields, because it is impossible to distinguish between a yield difference that is attributable to the field's inherent production capabilities or the management practice applied to that field. In an attempt to alleviate this problem, a technique was used on yield data that is traditionally used to compare yield from different crops; a "normalized" yield was computed for each yield observation in each field. Normal in this context refers not to a standard data distribution, but rather to an index of yield computed by dividing each yield observation by its field's maximum yield value. This results in a number between 0 and 1 that describes that observation's position within the entire range of yield values for a particular field.

Thus, by computing this normalized yield, a numeric index was obtained that facilitates cross-field comparison. For each seeding rate and management type (control or variable) in both fields, an average normalized yield was computed which describes that seeding rate/management combination's performance. The results of this calculation are presented in Table XXXI.

TABLE XXXI

AVERAGE NORMALIZED YIELD BY SEEDING RATE, FIELD, AND MANAGEMENT PRACTICE

Management	*Seed Population*	*Field*	*Average Normalized Yield*
Variable	20000	GF120	0.507
Variable	22000	GF120	0.524
Variable	24000	GF120	0.537
Variable	25000	GF120	0.516
Variable	26000	GF120	0.506
Constant #1	30000	GF120	0.511
Constant #2	25000	GF120	0.486
Constant #3	20000	GF120	0.547
Constant #1	20000	GF76	0.596
Constant #2	25000	GF76	0.586
Variable	15000	GF76	0.640
Variable	18000	GF76	0.593
Variable	20000	GF76	0.678
Variable	20500	GF76	0.621
Variable	23000	GF76	0.647
Variable	25000	GF76	0.648

One thing that becomes immediately apparent upon studying this table is that in field GF76, the normalized values are not as low as those in GF120; this reflects the lower amount of yield variability in field GF76. However, for field GF76, it is also striking how, with the exception of variable rate 18,000, all of the variable rates had normalized yields above 0.6, while the constant rates were below 0.6, making them the lowest producing areas in the field. In contrast, field GF120 had several variable rate areas that produced lower yields than constant rate areas; indeed, from the above analysis, the highest yielding areas of the field in GF120 were variable rate 24,000 and constant rate 20,000 seeds per acre. This tends to imply that using soil depth as the variable rate criterion was not as accurate in portraying crop production potential in the field as the grain moisture in GF76.

This finding is supported by looking at how well the criterion layer for varying seed population is associated with yield variation in each field. This was accomplished by computing a correlation in which yield was correlated with the population variability criterion for each field; this was previous year's grain moisture for field GF76, and topsoil depth in field GF120. The result of these correlations was that GF76 had a .22 correlation between yield and grain moisture, and field GF120 had a .09 correlation between yield and soil depth. These results show while neither field displayed a very high correlation between the variable rate criterion and yield, GF76 did have a better response than GF120.

The low correlations presented here also demonstrate that the complexity of yield variability in production fields such as these cannot be modeled with a single variable; a number of variables of different types interact to produce yield variability.
Obviously, we cannot include soil clusters in this comparison of fields to attempt to normalize crop production potential between the two fields. Different variables and variable transformations were used to compute the clusters in each field, and the variability of each was different, such that soil cluster 1 in field GF76 does not have the same characteristics as soil cluster 1 in field GF120.

CHAPTER VI

SUMMARY AND DISCUSSION

Summary

The comparison of yield within management zones between different portions of a field and between different fields is confounded in production agriculture by within-field variability of crop-growth parameters other than the management zones themselves. In the current study, these inherent locational differences within fields were intensified due to the fact that the management practice under investigation was performed differently in each field under consideration. However, by utilizing the information stored in a spatial database, it becomes possible to attempt to normalize crop-growth conditions within a field in an attempt to adequately compare the yield results from different locations.

The calculation of soil clusters was performed to accomplish this task; their validity in explaining yield variation was confirmed by the rejection of the null hypothesis in both fields that there was no yield difference between at least some of the derived clusters. The benefit of these clusters for this study was that they facilitated within-field comparisons of variable rate versus constant rate seeding practices by describing crop-growth conditions in the field; thus, if a difference was found between constant and variable seeding practices within the same soil cluster, then that difference could be attributed to the management practice itself and not external conditions that may have been either beneficial or detrimental to yield. In the case of field GF76, where clusters 1 and 2 did not have significantly different yields from each other, they were combined and treated as a single entity. Since the clusters were derived from different variables in each field, and since they represented different soil conditions, they could not be used in normalizing conditions between the two fields.

The practice of varying seed population based on previous year's grain moisture percentage produced significantly different yields, with higher yields associated with more dense populations. They also significantly out-performed their constant rate counterparts, although there were not constant rate zones available at every variable seeding rate in the field for comparison. Varying seed population based on depth of A and E soil horizons did not prove to be clearly superior to constant rate seeding; however, the lack of a clear direct relationship between seeding population and yield indicates that different areas of a field can support different densities of crop growth. These two findings suggest that if a different variable rate seeding recommendation criterion were used in field GF120, perhaps there would have been a clear advantage over constant rate seeding practices.

The comparison of the performance of variable rate seeding practices between the two fields suggests that in field GF76, there was a much higher response in yield to the population rate than in GF120; however, GF76 also clearly encompassed less yield variability than field GF120. The lack of ability to normalize within-field conditions between the fields compromised the ability to compare their respective yields, but by examining the relationship between yield and the variable rate criterion, it was demonstrated that the relationship between yield and variable criterion in GF76 was much stronger than that of field GF120, suggesting that previous year's grain moisture is a better measure of soil moisture-holding capacity than top soil depth.

Discussion

The current study required not only a very large dataset in order to describe conditions within each field, but also a good deal of data processing in order to convert raw data into data models which are conducive to analysis. Spatial data layers were necessary to record raw data describing soil moisture-holding capacity, derived themes which represent variable rate seeding recommendations, soil fertility points and interpolated soil fertility surfaces to describe soil characteristics, and yield data from a grain yield monitor were the inputs to the study. Other derived spatial layers such as soil clusters and observation polygons which include soil cluster, population, and average yield were also necessary. This point highlights the difference in methodology and experimental design between traditional agronomic studies, which are conducted on small, homogeneous plots with variability and external factors kept under control, and geographical studies which occur in real-world production scenarios, in which external variability exists and must be accounted for, and the producer faces economic constraints and risk factors which may compromise the quality of the dataset. Without a powerful geographic information system and a host of data transformation and representation techniques, the implementation of concept research conducted on an actual production farm would not be practical.

Even armed with the wealth of data available in this study, there were limitations in the dataset. The ability to compare yield between constant rate and variable rate portions of the field could have been improved by a one-to-one correspondance between control rates and population rates; e.g., for each variable rate in a field, one control strip of the same rate should have been included. In addition, crop influencing factors could have been included in the soil clusters; variables such as slope, aspect, and drainage quality of the soil would have been beneficial in accounting for yield variability. Also, with GF76, the soil fertility data was collected two years previous to the study. Further research is needed to determine the degree to which soil fertility variables vary as a function of time, as well as how landscape factors interact with a soil's chemical and physical properties to influence crop yield. The ability to compute a single index of crop-growth potential that would have been applicable to both fields would have also been beneficial to this study, in that it would have allowed a more direct comparison of the performance of the two variable rate criteria.

The methodological progression of this research provides a model for integrating disparate data types into an analytical context. Data stored at different scales representing various entities must be normalized spatially such that their distributions can be accounted for at a similar scale of observation. A comprehensive field-level dataset in production agriculture may include such objects as coarse-resolution points (such as fertility samples), fine resolution points (such as yield monitor data), and polygons (such as soil types or soil clusters). These data must be characterized such that their spatial distributions can be compared. The data model developed for this study provides an example of how to integrate such seemingly incompatible data types by normalizing the scale of observation (through interpolation or aggregation) and by joining different data into a single data structure by querying spatially to create a multi-attribute surface. The research presented here incorporates the work of various projects of a more specific scope (e.g., interpolation of soil parameters, calculation of a fertility index, etc.) and provides an example of how a multitude of data can be integrated for spatial decision support (see Chapter II, Literature Review for specific studies).

This study also presents a methodology for analyzing the performance of a spatial management strategy by attempting to explain variability in the response parameter (in this case, crop yield) caused by the mangement technique itself. This requires that extraneous variability, which could be location-dependent, be accounted for. This research addressed variable crop production capacities of the fields involved through the creation of soil clusters, which defined regions of similar soil characteristics. By using these clusters and analyzing the potential interaction between soil conditions and seeding rate, it was possible to explain differences in yield according to field conditions and according to the management technique of interest.

Despite the limitations of the dataset, insight was gained into how well these variable rate management techniques performed. The ability to store management data spatially and to subsequently analyze the data set for spatial relationships makes it feasible for a large-scale producer to design experimental trials in which a proposed spatial management technique is applied and assess the success of that techinque. As the scale of management of agricultural lands continues to grow, producers and farm managers are faced with the challenge of knowing what conditions exist at different locations, and how to best deal with those conditions. The storage of spatial and temporal data in a geographic information system will increasingly prove to be a useful decision-making tool in production agriculture.

REFERENCES

Aldrich, Samuel R., and Earl Leng. (1965) *Modern Corn Production*. F&W Publishing Corp., Cincinnati, Ohio.

Allen, Thomas R., and Stephen J. Walsh. (1996). *Spatial and Compositional Pattern of Alpine Treeline, Glacier National Park, Montana*. Photo. Eng. & Remote Sensing 11:1261-1268.

Bailey, Trevor C., and Anthony C. Gatrell. (1995). *Interactive Spatial Data Analysis*. Longman Scientific & Technical, Essex, England.

Berry, Joseph K. (1995). *Spatial Reasoning for Effective GIS*. GIS World Books, Fort Collins. Colorado.

Blackmer, Tracy M., James S. Schepers, Gary E. Varvel, and George E. Meyer (1996). *Analysis of Aerial Photography for Nitrogen Stress Within Corn Fields*. Agronomy Journal 88:729-733.

Bruckler, L., A.M. de Cockborne, P. Renault, and B. Claudot. (1996). *Spatial and Temporal Variability of Nitrate in Irrigated Salad Crops*. Irrigation Science 17:53-61.

Cambardella, C.A., T.B. Moorman, J.M. Novek, T.B. Parkin, D.L. Karlen, R.F. Turco, and A. Konopka. (1994). *Field-Scale Variability of Biological, Chemical, and Physical Soil Properties in Central Iowa Soils.* Soil Sci. Soc. Am. J. 24:36-41.

Chorley, Richard J. (1973). *Directions in Geography*. Methuen & Co. Limited, London, England.

Clark, W.A.V., and P.L. Hosking. (1986). *Statistical Methods for Geographers*. John Wiley and Sons, NewYork.

Cowie, G.A.. (1951). *Potash: Its Production and Place in Crop Nutrition*. Edward Arnold & Co., London, England.

Cummins, D.G.. (1976). *The Influence of Plant Population, Nitrogen Rate, and Hybrids on Production and Quality of Corn for Silage*. Research Bulletin #183, Georgia Experiment Station. 10 pp.

De Willegen, P. (1991). *Nitrogen Turnover in the Crop-Soil System; Comparison of Fourteen Simulation Models*. Fertilizer Research 27: 141-149.

Delcourt, Hans, Paul L. Darius, and Josse De Baerdemaeker. (1995). *The Spatial Variability of Some Aspects of Topsoil Fertility in Two Belgian Fields*. Computers and Electronics in Agriculture 14:179-196.

Elston, D.A., G. Jayasinghe, S.T. Buckland, D.C. MacMillan and R.J. Aspinall. (1997). *Adapting Regression Equations to Minimize the Mean Squared Error of Predictions Made Using Covariate Data from a GIS*. Int. J. Geographical Information Science 11:265-280.

Ferguson, R.B., C.A. Gotway, G.W. Hergert, and T.A. Peterson. (1996). *Soil Sampling for Site-Specific Nitrogen Management*. In: Robert, P.C., Rust, R.H., and W.E.Larson. *Proceedings of the Third International Conference on Precision Agriculture*. Minneapolis, Minnesota: ASA, CSSA, SSSA. pp 141-150.

Forcella, Frank. (1993). *Value of Managing Within-Field Variability*. In: American Society of Agronomy. *Soil Specific Crop Management*. Madison, Wisconsin: American Society of Agronomy. pp 125-132.

Gould, Peter. (1985). *The Geographer at Work*. Routledge & Kegan, London, England.

Gray, Wayne. (1996). Personal Communication.

Han, S., C.E. Goering, M.D. Cahn, J.W. Hummel. (1992). *Blocking of Spatial Data For Site-Specific Crop Management*. Presented to the American Society of Agricultural Engineers International Summer Meeting, Charlotte, North Carolina.

Harper, Horace J. (1947a). *Crop Adaptation to Soils of Varying Acidity or Alkalinity*. Stillwater, OK: Oklahoma Agricultural Experiment Station, Bulletin B-315. 12 p.

Harper, Horace J. (1947b). *Soil Reaction and Availability of Plant Nutrients*. Stillwater, OK: Oklahoma Agricultural Experiment Station, Bulletin B-315. 12 p.

Hartigan, John A. (1975). *Clustering Algorithms*. John Wiley & Sons, New York.

Hicks, Charles R. (1964). *Fundamental Concepts in the Design of Experiments*. Holt, Rinehart, and Winston, Inc., New York.

Hodgson, John Micheal. (1978). *Soil Sampling and Soil Description*. Oxford University Press, New York.

Hosseini, E., J. Gallichand, and D. Marcotte. (1994). *Theoretical and Experimental Performance of Spatial Interpolation Methods for Soil Salinity Analysis*. Transactions of the American Society of Agricultural Engineers 37:1799-1807.

Journel, A.G., and Charles Huijbregts. (1968). *Mining Geostatistics*. Academic Press, London, England.

Khakural, B.R., P.C. Robert, and A.M. Starfield. (1996). *Predicting Corn Yield Across a Soil Landscape in West-Central Minnesota Using A Soil Productivity Model*. In: Robert, P.C., Rust, R.H., and W.E.Larson. *Proceedings of the Third International Conference on Precision Agriculture*. Minneapolis, Minnesota: ASA, CSSA, SSSA. pp 197-206.

Lark, R.M., and J.V. Stafford. (1996). *Consistency and Change in Spatial Variability of Crop Yield Over Successive Seasons: Methods of Data Analysis*. In: Robert, P.C., Rust, R.H., and W.E.Larson. *Proceedings of the Third International Conference on Precision Agriculture*. Minneapolis, Minnesota: ASA, CSSA, SSSA. pp 141-150.

McGrew, J. Chapman, and Charles B. Monroe. (1993). *Statistical Problem Solving in Geography*. Wm. C. Brown Publishers, Dubuque, Iowa.

Norwood, Charles A., and Randall S. Currie. (1996). *Tillage, Planting Date, and Plant Population Effects of Dryland Corn*. J. Prod. Agric. 9:119-122.

Nowlin, Bruce. (1996). Personal Communication.

Oliver, Margaret A., and R. Webster. (1986a). *Combining Nested and Linear Sampling for Determining the Scale and Form of Regionalized Variables*. Geographical Analysis 18:227-242.

Oliver, Margaret A., and R. Webster. (1986b). *Semi-Variograms for Modelling the Spatial Pattern of Landform and Soil Properties*. Earth Surface Processes and Landforms, 11:491-504.

Reichenberger, L., and J. Russnogle. (1989). *Farming by the Foot*. Farm J. Mid-March:11-15.

Ritchie, G.S.P.. (1989). *The Chemical Behaviour of Aluminum, Hydrogen, and Manganese in Acid Soils. Soil Acidity and Plant Growth*. 12:1-45.

Sauchelli, Vincent. (1965). *Phosphates in Agriculture*. Reinhold Publishing Co., New York.

STATISTICA User's Guide [Computer Software]. (1995). Tulsa: Statsoft.

Sudduth, K., S. Borgelt, and J. Harrison. (1991). *Prescription Farming Research in Missouri.* Status report on MSEA Water Quality Res. for 1991. Univ. Mo., Columbia.

SURFER User's Guide 6.04 [Computer Software]. (1994). Golden: Golden Software.

Swain, R.W. (1983). *Approaches and Methods for Evaluating and Increasing the Crop Potential of Soils*. Martinus Nijhoff Publishers, The Hague, Netherlands.

Tevis, Joe W., A. Dale Whittaker, and Darrell J. McCauley. (1991). *Efficient Use of Data In the Kriging of Soil pH*. Presented to the American Society of Agricultural Engineers International Summer Meeting, Albuquerque, New Mexico.

Verstegen, Jos, Ruud Huirne, Aalt Djikhuizen, and Jack P.C. Kleijnen. (1995). *Economic Value of Management Information Systems in Agriculture: A Review of Evaluation Approaches*. Computers and Electronics in Agriculture 13:273-288.

Vorhees, W.B., R.R. Allmaras and M.J. Lindstrom. (1993). *Tillage Considerations in Managing Soil Variability*. In: Robert, P.C., R.H. Rust and W.E. Larson, eds. *Soil-Specific Crop Management*. Minneapolis: ASA CSSA SSSA. pp 95-112.

Waits, David A., and Stephen R. Johnson, Jr. (1996). *"A Stratified Soil Sampling and Analysis Strategy for Agricultural Producers"*. Presented to the Association of American Geographers, Charlotte, North Carolina.

Wessman, Carol A., C. Ann Bateson, and Tracy L. Benning. (1995). *Detecting Fire and Grazing Patterns in Tallgrass Prarie Using Spectral Mixture Analysis.* Ecological Applications 7:493-511.

USDA. (1993). *Soil Survey Manual*. United States Department of Agriculture, Washington, D.C..

2

VITA

Stephen R. Johnson, Jr.

Candidate for the Degree of

Master of Science

Thesis: THE SPATIAL VARIATION OF SEED POPULATION OF WHITE CORN BASED ON SOIL MOISTURE HOLDING CAPACITY

Major Field: Geography

Biographical:

Personal Data: Born in Fairborn, Ohio, November 24, 1973, son of Steve and Cheryl Johnson.

Education: Graduated from Oklahoma State University in May, 1995 with Bachelor of Science Degree; completed requirements for the Master of Science Degree at Oklahoma State University in December, 1997.

Professional Experience: Graduate research assistant in Oklahoma State University Department of Geography, May 1995 through December 1995; NASA Oklahoma Space Grant recipient January 1996 through July 1996; GIS Programmer/Analyst at SST Development Group, Stillwater, Oklahoma, August 1997 to present.

Professional Memberships: Association of American Geographers, Gamma Theta Upsilon, Agricultural Retailers Association.