Addis Ababa University
Office of Graduate Program

Faculty of Science
Department of Statistics

APPLICATION OF SPATIAL MODELING TO THE STUDY OF SOIL FERTILITY PATTERN

BY
DECHASSA OBSI

A Thesis submitted to the Office of Graduate Programs of
Addis Ababa University in Partial fulfillment
of the requirement for the Degree
of Master of Science in Statistics

September, 2008

Addis Ababa University
School of Graduate Studies

Title of Research: **Application of Spatial Modeling to the Study of Soil Fertility Pattern**.

Name Candidate: Dechassa Obsi

Department: Statistics

Faculty: Natural Science

Approved by

	Advisor	Examiner	Examiner
Name:	Dr. Girma Taye	Dr. Butte Gotu	Dr. Fantahun Abegaz
Signature	_____	_____	_____

Date September, 2008

ACKNOWLEDGEMENT

First of all, I would like to thank my advisor Dr. Girma Taye for his continuous follow up excellent advice and offering comments and suggestions that led to significant improvements during planning and execution of my research work.

I would like to thank the Ethiopian Agricultural Research Organization (EARO), Biometrics and Data Processing Department allow me to use their data.

I am thankful to my Father Ato Obsi Gudeta, mother Birki Tadesse my brothers and sisters Bayisa Obsi (BSc), Ayansa Obsi (BA), Yeshi Obsi (BA), Yeshi Nagassa, Ebise Tolera and Kenate Mamo(BA) for moral and financial support throughout my study of this program.

My special thanks goes to my lively wife, Dose Abdissa, my children (Talile and Kumale) for support assistances and encouragement without which achievement of the current status of education would have been difficult.

ABSTRACT

Spatial statistical analysis was undertaken to study the variability of potato yield data due to soil fertility pattern. Seed potato yield were measured from field uniformity trial conducted at Hollota and Kulumsa agricultural research centers in year 2001 on an area of 0.15 hectare at each site. The harvested area was divided in to basic units of 1.2x1.5 m and a total of 1658 and 931 plots were considered from Hollota and Kulumsa respectively. The basic units were combined in different plot sizes to acquire the required plot dimension.

In this study, the Monte Carlo test for completely spatial randomness is applied and the result shows no complete spatial randomness detected in the series of potato yield data for both sites.

Thus, to set a model adjusted for spatial pattern, Moran's index and Geary's coefficient were applied to test for global and local spatial autocorrelation respectively. The result shows positive spatial autocorrelation detected among potato yield data using Rook's weighted neighboring plot relations. The result also shows, increasing plot size will not generally make the observed spatial autocorrelation insignificant.

An autoregressive model, that is adjusted for presence of spatial autocorrelation in simulated plot size of $12m^2$ is fitted for row and column effect for each site. The result shows significant positive association between neighboring plots row effect and the adjusted potato yield. In addition, the result from comparison of model adjusted for presence of spatial autocorrelation and conventional OLS based analysis of variance shows the autocorrelation parameter accounts significant percent of variation among potato yield in both sites.

The classical and robust variogram models were used to produce a map of predicted potato yield data, taking in to account plot variation from the model prediction for Kulumsa agricultural research center. The result shows the predicted values at each of the grid locations do not differ greatly for the two variogram models. However, in the comparison of Gaussian and Spherical models the standard error of kriged prediction for the spherical model is subsequently larger than the Gaussian model.

TABLE OF CONTENTS

Appendices:

LIST OF TABLES

LIST OF FIGURES

List of Figures **Page**

CHAPTER I
INTRODUCTION

1.1. Background of the Study

Data that are tied to position on the earth's surface, that are spatial or geo-referenced data, often serves as the empirical backbone of much of the research that is presently done. The statistical analysis of spatial data forms the subject matter of "spatial statistics." Spatial statistics can be employed for description, inference, and modeling/predictions. Descriptive measures of spatial pattern include Moran's coefficient, point pattern, semivariograms and spatial connectivity or weight matrices.

Spatially arranged measurements and spatial patterns occur in surprisingly wide variety of scientific disciplines. The origin of human life link studies of the evolution of galaxies, the structure of Biological cells, and settlement patterns in archeology. Foresters and agriculturalists need to investigate plant competition and account for soil variations in their experiments. The abundance of research in many fields come together in to the analysis of spatial data and to make particularly available the methods made possible by computer software and computer graphics, which are on the essential tool in spatial Statistics. (Brian D. Ripley, 1981)

Generally speaking, spatial statistics is a form of specialized applied multivariate statistics, concerned with the statistical analysis of geo-referenced data.

Spatial statistics differ from classical statistics in that the observations analyzed are not independent; this single assumption violation is the crux of the difference. The other most important future that distinguishes spatial statistics from classical statistics is that, spatial statistics uses the spatial coordinates to model statistical dependence among data, rather than assuming the data are independent. The spatial dependence is often termed as autocorrelation, if the data are of one type and cross-correlation if the data is of another type, (Jay, Cressie and Paul, 2004).

Cressie (1991, p-3) characterize this subject as:

> '*independence is a very convenient assumption that makes much of mathematical-statistical theory tractable. However, models that involve statistical dependence are more realistic.'*

Observations are correlated strictly due to their relative locational position (referred to as spatial autocorrelation), resulting in redundant information to be present in data values. The redundancy increase as the degree of locational dependence increases. This duplication of information produces complications in the statistical analysis of geo-referenced data that remains dormant in the statistical analysis of traditional data, composed of independent observations. The net result is that classical statistics applied to geo-referenced data fail to capture locational information.

Types of Spatial Data

Spatial data may come in many various ways, and so that it is not easy to arrive at a system of classification that is simultaneously exclusive, exhaustive, imaginative, and satisfactory. Through advanced computer technology in spatial statistical theory, practitioners have gained powerful new ability to obtain, organize, and study complex relationships between farming and its natural resource base, like soil fertility pattern and topography of the agricultural field trial. Application of these capabilities is broad, ranging from investigating the farm level benefits of precision farming to assessing the impact of conservation practices over wide areal locations. Opportunities that arise from these developments also arise numerous questions relating to the interpretation, manipulation, and analysis of spatial data.

There are three broad categories statistics for spatial data: Point pattern, Lattice data, and Geo-statistics.

Point pattern

Treat locations as points to be analyzing the distance separating points, the density of points in areas and sub-areas, and the degree of correspondence between point patterns. The discussion of point pattern analysis in general will point up the fact that this approach to location analysis has been most fruit full in terms of analytical results.

The locations of points relative to one another, and the distance between nearest neighbors, are important expressions of arrangement in a point pattern. Plant ecologists, such as Clark and Evans (1954), Morisita (1959), and Pilou (1959), first developed these methods for analysis of plant communities.

Lattice Data

A lattice of location evokes an idea of regularity spaced points in $\Re^d$, linked to nearest neighbors, second-nearest neighbors, and soon. Here "Lattice" refers to a countable collection of (spatial) sites, distinguished as being either spatially regular or irregular (Birkhoff, 1967; Hammerslay and Mazzarino, 1983) with slit differences of usage Basag (1974,1975) and Cressie, (1993, p-383) reversed the word lattice for spatially regular sites.

Geo-statistics

Hart (1954), coined the term "Geo-statistics" in geographical context to denote statistical techniques that emphasis location within areal distributions, Matheron (1963b), used the term in geological context to denote theory and methods for inferring ore reserves from data spatially distributed throughout an ore body.

Agriculture and Statistics go back a long way. The origins of statistical science can be found in the design and analysis of agricultural experiments (Gower, 1988). Classical experimental design of agricultural field trials ignores the spatial position of the treatment in the design; however, it has been realized that more efficient treatment-contrast estimator can be obtained by exploiting spatial variation.

A uniformity trial, is an experiment where all the plots are treated with the control treatment, the data consists of plot yield for one specific crop, like potato. By aggregating adjacent plots into new plots and computing mean squared error as a function of a new plot size, an optimal plot size can be determined for future crop yield trials. Selected colons' and different varieties of crop were evaluated for adaptability and better yield in different regions for local check.

Soil variability introduces an extra source of variation in the field experiment and inflates the experimental error. The precision of an experiment is increased when experimental error is estimated precisely. In agricultural experiments, comparative studies on plot size (experimental units), has been carried out and found that determination of plot size increased the precision of results of an experiment.

When spatial data arise from controlled (or even particularly controlled) experimental situations of agricultural uniformity trial, an important area of interest is to design experiments to assess treatment effects. This includes designing the spatial layout of treatment blocks in order to counter the inferences of spatial correlation between the blocks (Kiefer and Wynn, 1983). This route is clearly not open to areas that depend on observational data but sampling schemes can be devised to minimize correlation effects.

Spatial modeling refers to a particular form of disaggregating, in which an area is divided into a number (often-large number) of similar units: typically grid square or polygons. Spatial data analysis requires models that describe spatial variations. Data that are close together in space (and time) are often more alike than those that are far apart, (Tobler, 1997). A spatial model incorporates this spatial variation in to the generating mechanism, in contrast to a non-spatial model.

In writing about his Cornell Theory Centers Super computer project, Durrett notes (1994,p-4) that:

> *'for a half century, the literature … has been dominated by models in which spatial location is ignored and each individual (site) is assumed to interact equally with all others. Such models provide an acceptable approximation in many contexts, but there is growing list of examples of phenomena that must be treated by models that are spatially explicit ….'*

Girma (2005), presented about spatial modeling selection as:

> *'Despite the current level of development in spatial modeling and design used for field trials, there are still some problems that require lasting solution. Past developments in spatial models were of significant practical value. There are no sound spatial model selection methods in place and Biometricians and researchers are not in position to select variable spatial models. The modeling of global and local variations was not successful due to confounding problem and further work is required in that line.'*

Spatial prediction is a simple way to smooth or interpolate the value of a surface by a weighted average value at the data points and incorporates spatial dependence. A simple and popular spatial prediction method is *ordinary kriging*. The word "*kriging*" is synonymous with "*optimal prediction.*" Kriging is a method for interpolation named after South African mining engineer named D. G. Krige (1951), who developed the technique in an attempt to more accurately predict ore reserves. Kriging provides optimal estimates given the

hypothesized relationship, and error estimates can be mapped to determine if spatial pattern exists.

1.2. General Objectives of the Study

The overall objective of this thesis is to study the fertility pattern of the experimental field trial, to compare the method of conventional statistical analysis with yield data adjusted for spatial autocorrelation based analysis and offer better modeling approach for improvement of precision and accuracy of farm trials. In general the aim is to show that the spatial analysis techniques can considerably increase our capacity of understanding the spatial patterns associated to agronomic yield data from field trial.

To achieve these objectives, spatial statistical approach is applied based on potato yield data collected from potato breading programs under Ethiopian Institute of Agricultural Research (EIAR), Hollota and Kulumsa agricultural research centers in year 2001.

The specific objectives of the Study are:

- To diagnosis for complete spatial randomness and detect spatial pattern among plot in agronomic field trial on potato yield.
- To fit spatial autoregressive model for potato yield adjusted to auto-correlated error.
- To identify important predictor variables of unknown values using random process, model variogram and “kriging” or “optimum predictor”.

1.3. Organization of the study

The study is organized in to five chapters as outlined in the table of contents. Chapter II deals with review of related literatures and summarizes works cited in the past either in or outside of Ethiopia on method of spatial data analysis and prediction of values for a given dataset. In Chapter III, the general spatial process, test for complete spatial randomness (CSR), methods of detecting spatial pattern in spatial analysis and systematic spatial model selection, and prediction methods are discussed. In Chapter IV, results and discussion based on the sample data are presented with SAS computer Package software. Chapter V, presents conclusion of the study, some methodological recommendations and limitations in the study, appendices for sources of reference, the original data and SAS computer software codes applied in the analysis are included.

CHAPTER II

LITERATURE REVIEW

2.1. Introduction

Potato (Solanum Tuberosum L) is one of the important food crops of the world. The potato, which has always been the 'poor man's frind' is originated from South America in the Central Andean region. The important potato growing countries in the world are U.S.S.R., Poland, U.S.A., China, India, Germany and Spain (Chhidda Singh, 1986). Also it is one of the most important vegetable crops in Ethiopia. It is very important food and cash crop specifically in the high lands of Ethiopia. According to a report of international Potato center (1982); it is estimated that in Ethiopia potato is produced on a total of 30,000 hectare of land annually with a national average yield of 5.3 tones/ha, and stated that, this yield is extremely low as compared to the national average yield obtained from many African countries that grow potato. To upgrade the productivity of potato perunit area and in order to partially cobra hunger in the country, it becomes very important to initiate national potato improvement projects in the country.

The variety trial for potato crop were conducted at several locations distributed across different agro-ecological zones of Ethiopia, like Hollota, Bako, Addet, Jimma, Hawasa and Sinana all of which were practiced for adaptability and better yield, and according to the repot of international potato center (1982), the cultivator gave significantly higher mean tuber yield than the local check. Potato

breeding programs generate many advanced new clones that need to be selected for yield
potential and adaptation every year. This breeding step requires high experimental precision to identify the best clones. The experimental precision involves experimental design, optimum plot size, and adequate replication number adjusted to the availability of experimental area.

The principal purpose of this thesis is to illustrate how to implement and evaluate spatial data analysis techniques in general and in particular, to increase experimental precision of yield performance trials for data aggregated by area in the study of plot size variation among potato clones for selected sites. Here, analysis of spatial autocorrelation (exploratory analysis intended to identify the structure spatial correlation that describes the yield data) is used to estimate the magnitude of the spatial autocorrelation between the plots. In this case, the objective is to develop and investigate the usefulness of spatial correlation analysis for characterizing the nature of the soil fertility pattern and determining optimum plot size. A descriptive statistic based on well defined areas, rectangular in shape, of a random location approach is taken to quantify spatial point pattern with specified length and width.

The Monte Carlo test method is used to diagnosis for complete spatial randomness. Moran (1950), and Geary (1954), introduced the coefficient of autocorrelation test using normality and randomization of the observations amongst the auto-correlated regions. Additionally, to illustrate the different sorts of spatial autocorrelation, global and local indicators of spatial autocorrelation are used to describe the performance of each model in planning and designing optimum plot size.

Typically, a model that takes in to account the spatial effects, referred as spatial autoregressive model, with short description are presented. The method of prediction known as “ordinary kriging” lies in determining the value of unobserved from the neighborhood of the observed value of spatial data also discussed. It is used to assess the information content of existing station by determining to what extent value can be predicted from other (neighboring) plots.

2.2. Review of Literature

Many authors in various disciplines discuss the dependence among observations of varieties in space and several different approaches will be taken to quantify types of spatial data analysis and prediction. Cliff and Ord (1973), gave formal definition of spatial autocorrelation, and conclude that, the analysis of spatially located data is one of the basic concerns of the geographer and so it becoming increasingly important in other fields of study. R. A. Fisher (1935) was clearly aware of spatial dependence in agricultural field experiments because he went to such great lengths to remove it. In the 1920’s and 1930’s, at Rothamsted experimental station in England, Yates (1938), established the principle of randomization, blocking, and replication and as well as controlling for unwanted bias, randomization also neutralizes (but does not remove the effect of spatial correlation). Mecrcer and Hall (1911), Fair Field Smith (1938), reported that experiments on agricultural field trials are carried out to compare varieties of crops and to investigate soil fertility variation.

Cliff & Ord (1981, P-8), provide formal definition for autocorrelation discussed in Haning (1980), stated that spatial autocorrelation as a term used to describe the tendency for similarly valued observations to cluster in space, distorts

inferential test and intensive estimates; and Upton and Fingleton (1985), with main emphasis testing the hypothesis the observations are auto-correlated against the observations are not auto-correlated. Petersen (1985), stated randomization is used to satisfy the condition of independence of the observations. It is supposed to ensure that no treatment is consistently favored by being placed under the best field conditions. Mulla *et al*. (1990); van ES and van Es (1993), stated that autocorrelation can be imparted to crop data because non-treatment variables such as soil and topography are auto-correlated and they influence crop response. Olson *et al.*, (1985) conclude as randomization may not completely neutralize this autocorrelation when it is too non-random to promote equal likelihood of treatments being placed under all possible field conditions.

Bartlett (1978); Besag and Kempton (1986); Green *et al.*, (1984); Zimmerman and Harville (1991) and Brownie *et al.*, (1993) used methods of analysis of field experiments that take in to account spatial autocorrelation citing evidence of up to 30 percent greater efficiency. The spatial process of Wiener-Kolmogorov prediction theory has been developed and used by Matheron (1965); Delfiner (1976); David (1977); and Journal and Huijbergts (1978), exposed different points of view for stochastic process prediction. Crissie (1993), stated kriging prediction is a weighted linear combination of all output values already observed and conclude that, the closer the inputs are, the more positively correlated the output are. Gabrosek and Cressie (2002), showed that when location error is large, kriging after adjusting for location error performs much better than that without adjustment, in terms of both prediction bias and the mean squared prediction error. Lewise and Hutchinson (200) developed an expression for estimating the importance of location error for a given dataset. Atkinson (1977), shows that the location error is result in an increase variogram only at small lags. The behavior of robust kriging estimation for dependent data has been

investigated by Gastwirth and Rubin (1975), and stated location estimation in the presence of serial correlation shows that robust estimators obtain higher efficiency relative to mean for Gaussian observations. Huber (1972) and Bartels (1977) stated that linear predictions are very sensitive to outlying observations and effect of outliers on inference is substantial. Hawkins and Cressie (1984) propose a way of using neighboring values to determine how much the outlier should be down weighted.

Kiefer and Wynn (1938), designed the spatial layout of treatment blocks in order to counter the inferences of spatial correlation between blocks. Papadakis (1937); Bartelett (1938, 1978); Wilkinson et al. (1983); Besag and Kempton (1986), used the nearest-neighbor method for analyzing agricultural field trials attempt to take spatial dependence in to account, independently, by using residual from neighboring plots as covariates, or by differencing.

Narayara and Ramanta (1985), determined optimum plot sizes by considering the relation between the coefficient of variation (CV) and plot size graphically: optimum plot size is represented by the point at which the maximum rate of change in the estimated CV per unit of increment in plot size takes place.

Smith (1938), developed a method of estimation of soil heterogeneity index. Later on Kock and Rigney (1951); Hatheway and Williams (1951); Federer (1955) & Narayana and Ramanta (1983) contributed to the development of the technique by improving Smith's method. Girma *et.al.*, (2000) developed Optimum plot size and shape for wheat, Mohamed (1987) developed for Durra Sorghum, Naraya and Ramanata (1985) for pigeonpea and Girma *et.al.*(2002), for potato yield under considerable soil heterogeneity using Smith's empirical model for two seasons to estimate the coefficients of heterogeneity and

optimum plot size. The result concluded that the main (rainy) season trials show less heterogeneity index than off-season trials and using two replications has increased precision in examination of the plot shape and size.

Kinfemikael A.(2000), applied spatial statistical methods to test spatial autocorrelation among 'neighboring' plots over the experimental farm and compared row-sowing method with scattered sowing methods for wheat yield and conclude that there is an over all high adjacent plot correlation throughout the experimental farm, arising from soil fertility pattern. Girma T.(2005), discussed various methods of spatial data analysis, methods of fitting variogram models (prediction) and methods of modeling spatial variation. Results on a potato uniformity trial was demonstrated for relationship between small (basic) and large plot sizes and conclude that as plot get larger the strength of correlation between neighbors decreases.

There is great interest in the method of analysis of field experiments that take in to account spatial autocorrelation, for example, Grondona and Cressie (1991), found that spatial analysis that approximate the error variance-covariance structure are more precise than are conventional analysis that assume this structure to be independent (Arlinghaus, 1996). Spatial statistics is devoted to measuring neighboring influence and involved in different area of research. Awareness of complications attributable to locational information latent in spatial data, especially in terms of their impact on the validity of traditional statistical analysis, has emerged recently among scientists, catapulting spatial statistics into forefront of much data analysis discussion (Arlinghaus, 1996 p-3). In fact, the analysis of spatial data has become a major preoccupation of numerous statisticians only rather recently. Accordingly, increasing attention has begun to focus on the general field of spatial and geo- statistics. For

instance, the announced goals in the solicitation of proposals for the NCGIA (NSF, 1987) included the objectives of promoting advances in spatial statistics with in context of GIS. And, the Board of Mathemathical Sciences of the National Research Council (1999b) has targeted spatial statistics as one of twenty-seven topics of national concern in mathematics (its rank is 17). And characterized this subject area as: '*one of the most rapidly growing areas of statistics, rife with fascinating research opportunities.*' Arlinghaus (1996, p-2)

2.3. Statement of the Problem

In agricultural field trials, it is likely that plots located close to one another share the same micro-environmental soil condition than those farther apart. Soil fertility gradients often occur in patches in the field and it is the distance and direction of the neighboring plots that is important for the presence of spatial autocorrelation. This leads to spatial model fitting and spatial prediction of neighboring plot variations. However, unassumed amount of autocorrelation can inflate the level of significance in data and can lead to incorrect inference.

In spite of increasing attention has begun to focus in the current development in spatial modeling and designs used in experimental field trial, there is a shortage of work in spatial statistical techniques in case of Ethiopia. Thus, based on the generated data of potato yield, the spatial modeling approaches that take into account spatial effect and spatial techniques of analysis is applied that may minimize the problem stated.

The techniques applied includes:

- Measure soil fertility pattern using method of spatial analysis.
- Test for observing significant spatial dependencies in statistical analysis.
- Generating stochastic model that take in to account the spatial autocorrelation effect and compare the result with that do not include spatial effect in the field trials.
- Method of spatial prediction called ordinary kirging is applied to predict unobserved values.

CHAPTER III

MATERIALS AND METHODS

As outlined in the introductory chapter and in literature review, the principal objective of this chapter is to illustrate and discuss how to implement spatial statistical techniques, and to show basic differences between spatial and other statistical methods of analysis based on effect of dependent neighboring plots on agronomic yield data in field trials. In section 3.1, data source and method of collection are discussed. The methodology part includes general spatial process; test for complete spatial randomness and spatial autocorrelation are discussed in section 3.2. In section 3.3, spatial autoregressive models fitting method are discussed with its advantage over the analysis of variance method that does not include spatial effect. Lastly, spatial prediction methods including the variogram and ordinary kriging methods are discussed in section 3.4.

3.1. Data Source and Method of Collection

The data that has been utilized in this thesis are collected from the agricultural research centers Hollota and Kulumsa by well-trained enumerators and under careful supervision of EIAR, department of Biometrics and Data Processing, such as verification, data entry, and tabulation was done at the head office, so as to ensure the quality of the data.

The data is collected from the uniformity trial based on the study of the fertility pattern of the experimental field trial on Potato yield, Hollota and Kulumsa Agricultural research centers on an area of 0.15 hectare on each site, under Ethiopian Institute of Agricultural Research (EIAR) in year 2001. Hollota (located in West Shoa zone at 2495 meter above sea level, with 15^0C average annual temperatures and 1160 mm annual average rainfall) and Kulumsa (located in Arsi zone at 1600 meters above sea level , 14^0C average annual temperature and 823 mm average annual rainfall). In each experimental area seventy five cm spacing between rows and 60 cm spacing between plants were applied. The harvested area was divided into subplots (basic units) of 1.2x1.5 m^2, with plot size ranging from 1.8 to 120 m^2 with various shapes.

A uniformity trial is a field experiment all the plots are treated with control treatment and the data consists of plot yields for one specific crop.

3.2. Methodology

3.2.1. The General Spatial Process

Let $S \in \Re^d$ be a generic data location in d-dimensional Euclidean space and suppose that the potential datum $Z(S)$ at spatial location s is a random quantity. Now let s vary over index set $D \in \Re^d$ so as to generate the multivariate random process:

(3.1) $\{Z(S) : S \in D\},$

with a realization of a random process denoted as $\{z(s) : s \in D\}$, where D is a fixed subset $\Re^d$ with positive d-dimensional volume. To emphasize the source of randomness ***(3.1)*** is some times written as $\{Z(s,w) : s \in D, w \in \Omega\}$, where

$\{\Omega, £, \rho\}$, is a probability space; where Ω is the realizations sample space $\{z(s): s \in D\}$, $£$ is field of fixed set $D \in \Re^d$ and $\rho: Z(s) \to \Re^d$ is a probability measure correspond realizations to a particular value of w, say $w = w_0$.

In geo-statistical data, D is a fixed subset of $\Re^d$ that contains a d-dimensional rectangle of positive volume; $z(s)$ is a random vector at location $s \in D$.

The flexibility of ***(3.1)*** is now apparent, and although it is clear that D and $z(s)$ could be even more general random quantities, the fact that D is a subset of $\Re^d$ allows ***(3.1)*** to be called a spatial process.

A Point Process Theory

A point process is a stochastic model governing the location of an event $\{s_i\}$ in some set $X \subset \Re^d$ (applications are usually for $d = 2$ *or* $d = 3$). The process then consists of $\{Z(s_i): s_i \in D\}$, $D \equiv$ a spatial point process and $Z(s_i) \equiv$ a random set located at $s_i \in D$. Data from this process are often observed at a realization of $X = U\{Z(s_i): s_i \in D\}$, which referred to as generalized Boolean (Serra, 1980).

3.2.3. Diagnosis for Spatial Data

Spatial dependence

Spatially dependence is a common characteristic of spatial data that is of intrinsic interest as well as creating problems for the application of certain statistical procedures. Methods of diagnosis for spatial structure are needed for detecting and representing spatial dependency. It is important to assess the sensitivity of results to modeling assumptions. Certain modeling assumptions (such as how spatial relation ships between observations are represented and

what conditions are assumed at the boundary of the study region) could be of considerable importance in affecting model fit (Robert Haning, 1990, p-5).

Spatial dependency leads to the spatial autocorrelation problem in statistics since it violates standard statistical techniques that assume independence among observations. For example, regression analysis that do not compensate for spatial dependence can have unstable parameter estimates and yield unreliable significance tests. It is also appropriate to view spatial dependency as source of information rather than some thing to be corrected.

3.2.3.1. Complete Spatial Randomness

Randomization strives to neutralize the effect of spatial correlation and renders valued test for the hypothesis of equal treatment effect. Grodona and Cressie (1991), have shown that resorting to a spatial statistical model can yield more efficient estimators of the treatment contrasts than classical statistical approaches. In doing, such a spatial statistical analysis gives a more complete understanding of the phenomena influencing observed values.

Complete spatial randomness (CSR) is the “white noise process” of spatial point process. It characterizes the absence of structure in the data. As such, it is often the null hypothesis in the statistical test to determine whether there is spatial structure in a given point pattern. Because all statistical process have component of randomness in them, the complete spatial randomness is used and is synonymous with homogeneous Poisson process. The method of Monte Carlo test for complete spatial randomness is discussed in the next section.

3.2.3.2. Monte Carlo Test for Completely Spatial Randomness

There are various statistics for testing complete spatial randomness (CSR). The Monte Carlo test is implemented by simulating values of the test statistic under CSR and comparing them to the corresponding statistics calculated from the observed pattern (Cressie, 1993). Let U denote a test statistic and let $\{U_i : i=1, 2, \ldots, k\}$ denote $(k-1)$ values of the statistic generated by independently simulated data (of the same size) $(k-1)$ times under the null hypothesis of CSR. Call U_1 the observed value of U and ask whether U_1 is different from, or typical of, the other $U_{(i)}$'s. This is ascertained by ordering $U_{(1)} \leq U_{(2)} \leq \ldots \leq U_{(k)}$ and rejecting the null hypothesis if U_1 is one of the smallest or one of the largest order statistic (depending on the direction of the alternative hypothesis with regard to U).

Consider a bounded region $A \subset \Re^2$, Pielou's (1959), used statistic

(3.2) $$U_{(j)} = \pi \hat{\lambda} \sum_{i=1}^{n} X_i^2 \Big/ n$$

for testing CSR versus clustering , where $\hat{\lambda} = N(A) \Big/ |A|$ is an estimator of the global intensity and $X_1, X_2, \ldots, X_n$ is a random sample of $n \leq N(A)$ point- to-nearest – event distances. The null hypothesis of CSR involves the intensity parameter λ, the conditioned sufficient statistics $N(A)$, and the number of events in A. The Monte Carlo test can be carried out, because large values of U indicate clustering-type departure from CSR, for size-$\alpha (0 < \alpha < 1)$ Monte Carlo test is performed as follows:

Suppose $U_1 = U_{(j)}$ for some $j \in \{1, \ldots k\}$, then reject the null hypothesis if $(k+1-j)/k \leq \alpha$. Based on the 99 simulations (i.e., k=100), rejection at 5% level occurs if U_1 is $U_{(96)}, \ldots, U_{(100)}$. If a CSR hypothesis is rejected, then we use often fit a model that departure from CSR and have a tendency of regular spatial pattern; that is, statistical modeling that is adjusted for spatial autocorrelation may considered.

3.2.4. Global Indicators of Spatial Autocorrelation

What is spatial autocorrelation?

Spatial autocorrelation exists when the value at any one point in space is dependent on value at neighboring points. Spatial autocorrelation statistics measures and analyzes the degree of dependency among observations in neighboring areal units. If there is any systematic pattern in the spatial distribution of a variable, it is said to be spatially autocorrelated. If nearby or neighboring areas are more alike, this is positive spatial autocorrelation. Negative autocorrelation describes patterns in which neighboring areas are unlike. Zero result indicate no spatial autocorrelation. This is, random patterns exhibit no spatial autocorrelation. Spatial autocorrelation that is more positive than expected from random indicate the clustering of similar values across specific areal-location, while significantly spatial autocorrelation for a data set.

The possibility of spatial homogeneity suggests that the estimated degree of autocorrelation may vary significantly across geo-space. Local spatial autocorrelation statistics provides estimates disaggregated to the level of the spatial analysis unit, allowing assessments of the dependency relationship across space.

Cliff and Ord (1973), defined formally as:

> *'Let there be a study lattice which is exhaustively partitioned in to* $n = rc$ *non-overlapping plots, where* r *and* c *represent the total number of rows and columns respectively. Further, let* X *be a measurable variate, such that* x_i *is the observed value of* X *at* i^{th} *plot. Now, if for every pair of plots* i *and* j *in the study area the drawings which yield* x_i *and* x_j *are uncorrelated, then there is no spatial autocorrelation in the plot system on* X*. Conversely, spatial autocorrelation exists if drawings are not all pair wise uncorrelated (or drawings are influenced by neighboring plots).'*

Why spatial autocorrelation is important?

Most statistics are based on the assumption that the values of observations in each sample are independent of one another. Positive spatial autocorrelation may violet this, if the sample were taken from near by areas. Questions: How similar are near values? Are values grouped? Or are they randomly distributed? May be answered by calculating spatial autocorrelation.

Tobler (1979), claims that "everything is related to everything else, but near things are more related than distant things." This 'First law' of Geography captures a spatial regularity known as distance decay relationship. It implies that objects close to each other may be very similar in their attributes, pinpointing the source of many persistent problems that spatial scientists have to deal with in analyzing spatial data. Because objects are spatially related, each observation is no longer independent. Thus, spatial autocorrelation has to be taken into account in analyzing spatial data.

Goals of spatial autocorrelation

1. Measure the strength of spatial autocorrelation among neighboring areal units.
2. Test the assumption of spatial independence or randomness.

Methods of measuring spatial autocorrelation

Spatial autocorrelation is conceptually as well as empirically, the two dimensional equivalent of redundancy. It measures the extent to which the occurrence of an event in an areal unit contains, or makes more probable the occurrence of an event in neighboring areal unit.

3.2.4.1. Defining Spatial Weights Matrix

Stetzer (1982), was one of the first researcher to systematically addressed the issue of how a spatial weights matrix should be specified. According to Arlingaus (1996, p-8), often a spatial data analyzer needs to know the answer to questions asking:

(1). Which spatial autoregressive model and accompanying spatial weight matrix performs best, irrespective of sample size and miss specification;

(2). about the consequence over-specifying the weight matrix (i.e., including false spatial dependence adjacencies); and

(3). about the consequence under-specifying the weight matrix (i.e., omitting or including true spatial dependence adjacencies).

In spatial autocorrelation analysis, some measure of contiguity is required. Contiguity has a rather broad definition depending on the research question; however, most analyses in spatial autocorrelation adhere to a common definition

of neighborhood relations. Namely, neighborhood relations are defined as either rook's case, bishop's case or queen's (king's) case. These are rather simple and intuitive as their names suggest see Figure 3.1 below.

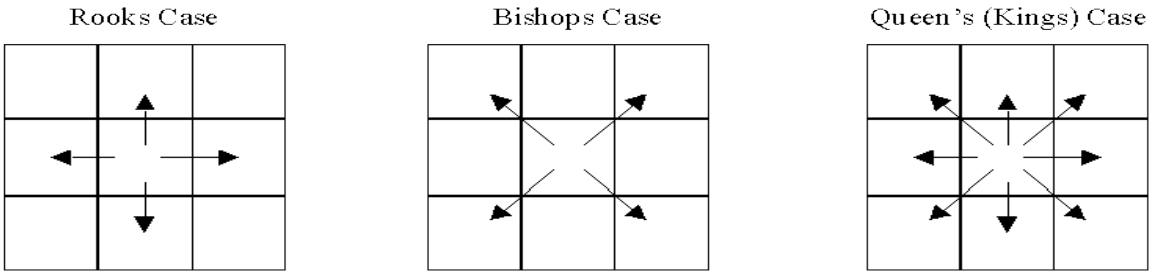


Figure 3.1. Different definitions of contiguity matrix

Rook's case contiguity is by a neighborhood of four locations adjacent to each cell, Bishop's only considers the diagonals of the relation and queen's or king's case considers a neighborhood of eight cells. These are the most common forms of contiguity used in spatial autocorrelation when considering continuous data in a raster format. Of these three, the rooks' case is the most commonly used and most programs only will compute this particular case. Hereafter when talk about the contiguity matrix or weight matrix below, only those cells that have a rook's case contiguous relation are given a value of 1.

Suppose an nxn matrix spatial weighted matrix $W = \{W_{ij}\}$ indicating elements of computations, which are to be included or excluded

(3.3)
$$W_{ij} = \begin{cases} 1, & \textit{if two plot are adjacent} \\ 0, & \textit{other wise} \end{cases}$$

and general cross product statistic: $\Gamma = \sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})$

This matrix is symmetric and binary, but these futures are not pre-requisites. More generally, if the source data polygon, for example, well defined

agricultural tracts, the adjacency weights rook's case might be chosen to reflect the relative length of shared boundaries. There are a number of general measures of spatial autocorrelation that are derived from the general cross product statistic: Moran's I, Geary's C, are the two common indices of spatial autocorrelation.

3.2.4.2. Test Statistics for Spatial Autocorrelation

I). Moran's Index, I

The most commonly used specification test for spatial autocorrelation is derived from statistics developed by Moran (1984, 1950a, 1950b) as the two dimensional analog of a test for univariate time series correlation (see also Cliff and Ord (1972, 1973). It is one of the oldest indicators of spatial autocorrelation (Moran, 1950), still defects standard for determining spatial autocorrelations, campers the value of the variable at any one location with the value of all other location and applied to point with continuous variable associated with them.

Now having a set of observed values $\{x_i\}$ and a set of weights $\{W_{ij}\}$, it is possible to look for some way of combining this information, by multiplying the location value together, optionally adjusting for over all mean value for all locations and including the adjustment and weight included, which gives equation ***(3.3)*** above.

As noted above, assuming the weights W_{ij} are binary, they simply identify which elements of the computation are to be included or excluded in the calculation. The expression looks quite like the covariance of the selected data.

To adjust for the number of items included in this sum and produce a covariance value, we need to divide through by the sum of the weights used,

(3.4) i.e. $$\frac{\sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j W_{ij}}$$

To standardize the covariance expression, divide ***(3.4)*** by the variance of the data, this is given as:

(3.5) $$\frac{1}{n}\sum_i (x_i - \bar{x})^2$$

The ratio of these two expressions ***(3.4)*** and ***(3.5)*** gives an index, known as *Moran's* I *statistics:*

(3.6) $$I = \frac{n\sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_i \sum_j W_{ij})\sum_i (x_i - \bar{x})^2}$$

where, n is number of cases

x_i is the variable value at particular location

x_j is the variable value at another location

$\bar{x}$ is the mean of the variable

W_{ij} is the weight applied to the comparison W between location i and location j defined in equation ***(3.3).***

In general, Moran's index serves as a test where the null hypothesis is the spatial independence; in case its value would be zero. Depending on the configuration of the weights used, the upper and lower limit of this index

typically range from approximately +1, representing completely positive spatial autocorrelation, to approximately -1, representing complete negative spatial autocorrelation, and with zero result representing no spatial autocorrelation, indicates that the pattern of values is random across space.

Here, from equation ***(3.6)*** we notice that the term in the denominator $\sum_i \sum_j \left(x_i - \bar{x}\right)^2$ cannot be zero because the equation would be undefined in such a case. In other words, if this term is zero then there is no variance in the original data and the I measure is inappropriate. This seems intuitively problematic since a field with no variance is perfectly auto-correlated since all values are similar, that is, exactly the same. Furthermore, $\sum_i \sum_j W_{ij}$ cannot be zero if there are adjacency relations considered. The only case where Moran's I equals zero is when $(x_i - \bar{x})(x_j - \bar{x})$ equals zero.

Once calculated, it is important to establish its statistical validity. In other words, would the measured values represent a significant spatial autocorrelation? To estimate the significance of the index it will be necessary to associate it to a statistical distribution, usually, the normal distribution.

Testing Statistical Significance of Moran's I

The statistical significance of Moran’s index I can be assessed through a number of formulae that have been derived either by the normal approximation or by randomization experiments. For the general cross product, randomization is used. However, through randomization Huber (1981) show that the general cross-product statistic can add additional flexibility to task specific needs and does not require a theoretical distribution in order to test significance of the result. For Moran's I there are formulae for the normal approximations and for

randomization but the normal formulae are far simpler than those for the randomization tests are. First, under the normality assumption the variance of I, $Var_N(I)$, is given as:

$$Var_N(I) = \left(\frac{1}{S_0^2 (n2-1)} (n^2 S_1 - nS_2 + 3S_0^2) \right) - E_N(I)^2 \quad \textbf{(3.7)}$$

whereas, the variance of Moran's I under Randomization - $Var_R(I)$ - is given as:

$$Var_R(I) = \frac{\left(\left[n \left| (n^2 - 3n + 3) S_1 - nS_2 + 3S_0^2 \right| \right) - \left(k \left| (n^2 - n) S_1 - 2nS_2 + 6S_0^2 \right| \right)}{(n-1)(n-2)(n-3) S_0^2} - ER(I)2 \quad \textbf{(3.8)}$$

The following variables in the variance equations are defined as:

n = Number of Observations

$$E_N(I) = \frac{-1}{(n-1)}$$

$$E_R(I) = E_N(I)$$

$S_0 = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} W_{ij}$ The sum of the Spatial Weight Matrix

$S_1 = \frac{\sum_{i=1}^{i=n} \sum_{j=i}^{j=n} (W_{ij} + W_{ji})^2}{2}$ If weight matrix symmetric then $S_1 = 2 \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} W_{ij}$

$S_2 = \sum_{i=1}^{i=n} (W_{i\bullet} + W_{\bullet i})^2$ The sum of the (i^{th}column + i^{th}Row)2 of weight matrix. If symmetrix $S_2 = 4 \sum_{i=1}^{i=n} W_{i\bullet}^2$

$k = \frac{\left[\sum_{i=1}^{i=n} (x_i - \bar{x})^4 \Big/ n \right]}{\left[\sum_{i=1}^{i=n} (x_i - \bar{x})^2 \Big/ n \right]^2}$ Involves the sum of each value in the data matrix minus the mean

$$k = \frac{m_4}{m_2^2}, \qquad m_r = \frac{1}{n_i} \sum_{i=1}^{n} (x_i - \bar{x})^r$$

The Standard deviation and standard z-scores of I are given by:

$$SD_{NorR}(I) = \sqrt{Var_{NorR}(I)} \quad \textbf{(3.9)}$$

and the z - score is given as

$$z = \frac{(I - E_{NorR}(I))}{\sqrt{Var_{NorR}(I)}}$$

From the expected value and Variance Z score and significance level to I can be computed under the alternative distributional assumptions. If the computed values of I are all different from its expected value, which is $\left(\frac{-1}{n-1}\right)$, thus the data of the series of observations are truly dependent. That is, if the values of Moran's I is equal to $\left(\frac{-1}{n-1}\right)$, then the spatial distribution is no different than if the values are assigned randomly to each location and there would be no spatial autocorrelation evident.

The null hypothesis is that there is no spatial autocorrelation; if Z(I) > Zcrit (Zcrit=1.96 for p=0.05, 2-tailed), then reject the null hypothesis at the p=0.05 confidence level.

II). Geary's Coefficient, C

One alternative to Moran's I form of computation involves sum of squared differences of attributes, $\sum_i \sum_j (x_i - x_j)^2$. These expressions once again standardized by dividing the variance and adjusted for the ratio of location or point pairs sampled. If this expression is used the resulting measure is known as the Contiguity ratio or Geary's ratio, C:

(3.10)
$$C = \frac{(n-1)\left(\sum_i \sum_j W_{ij}(x_i - x_j)^2\right)}{2\left(\sum_i \sum_j W_{ij}(x_i - \bar{x})\right)^2}$$

Computation of Geary's C (1954), results in a value with in the range of 0 to +2, with zero being a strong positive spatial autocorrelation, +1 through +2

which represents a strong negative spatial autocorrelation. If values of any one location are spatially unrelated to any other location, the expected value of C will be 1.

Testing Statistical Significance of Geary's C

In the formulae below all variables are the same as defined for the Moran's I significance tests.

$$Var_N(C) = \left(\frac{1}{2(n+1)S_0^2} ((2S_1 + S_2)(n-1) - 4S_0^2) \right)$$

$$Var_R(C) = \frac{\left[(n-1)S_1(n^2 - 3n + 3 - (n-1)k)\right] - \left[\frac{1}{4}\left((n-1)S_2\left\langle n^2 + 3n - 6 - (n^2 - n + 2)k\right\rangle\right)\right] + \left[S_0^2\left(n^2 - 3 - (n-1)^2 k\right)\right]}{n(n-2)(n-3)S_0^2}$$

Sokal and Oden (1978), note that empirical results computed by both Moran's I and Geary's C are similar they are not identical. The numerator of Geary's C will tend to be sensitive to absolute differences between neighboring variates due to the squared term in the numerator making up the W_{ij} component. Geary's C is computed identically to Moran's I with the exception of the numerator W_{ij} term which does not include the difference of each variate from the mean but rather each variate from all other variates. That is, Geary's C, is inversely related to Moran's I, does not provide identical inference because it emphasize the difference in value between pairs of observations, rather than the co-variation between the pairs. Moran's I uses the inference between each point and global average give more global indicator, where as the Geary coefficient is more

sensitive to differences in small neighborhoods. The Z-score of Geary's C is given by:

(3.11) $$Z(C) = \frac{C - E(C)}{S_E(C)},$$

where $S_E(C) = SQRT[Var_{N\,or\,R}(C)]$

Most often spatial autocorrelation is considered for raw data at various locations. Another application of spatial autocorrelation is in model specification. Specifically, it is applicable in the testing of regression residuals for spatial autocorrelation. The use of spatial autocorrelation in the residuals of a regression is similar to that of regular temporal autocorrelation in the residual of regression as discussed in the next section.

What to do if there is significant spatial autocorrelation in the residuals? First thing is to fined an additional predictor (s) variable that explains the autocorrelation, and that is to the $z = (x_i - \bar{x})$ variable. In other words, the model is actually incomplete because there is left out an important explanatory variable. One way to think of it, to quote Odland (1988), is that:

> *'The 'problem' is not the presence of autocorrelation in the model. The 'solution' is to develop a model that does account for spatial autocorrelation.'*

3.3. Fitting the Autoregressive Response Model For Agronomic Field Trial

The method discussed in section 3.2.4 is based on measuring and testing for spatial autocorrelation that could reveal spatial pattern in the data. Modeling is facilitated with spatial autocorrelation specifications. Some of the major models namely, the simulated autocorrelation model (SAR), the moving average model (MA), and the autoregressive response model (AR). Since these models share similar computational procedures, in this section describes only the AR model, because of as stated in the review literature AR model accommodates plots that are located on the nodes of regular lattice. Thus in case of spatial data, where spatial dependence is detected, it is very unlikely that the standard hypothesis of uncorrelated observation is true. The usual graphical analysis tools and the residual mapping can provide the first indications that the observed values are more correlated than would be expected under a condition of independence. In this case, Moran and Geary- spatial autocorrelation tests on the regression residuals warn of the presence. If the autocorrelation exists, we must specify a model that takes in to account the effect of it.

In this section, methods of spatial autoregressive model that take in to account the spatial effect are discussed. The explicit inclusion of spatial effects in regression model can be done in one of the simplest spatial regression model called autoregressive model (AR), assuming that it is possible to capture the structure of spatial autocorrelation in one parameter only, that is added to traditional regression model. First the ignored spatial autocorrelation is attributed to dependent variable Y. Since spatial dependence is taken in to account by the addition to the regression model of a new term in the form of a

spatial relation for the dependent variable. The AR- model arises when explanatory variable are required. It states that the realization of response variable Y at areal unit i is a function of its expected realization at i ($X\beta$), plus realization of Y locations connected to i (ρWY), plus an error term ε. AR model specification describes the spatial variation in the (nx1) crop response vector Y is given by:

(3.13) $$Y = \rho W^{**} Y + X_p \beta_p + \varepsilon$$

where, ρ is a constant spatial autocorrelation parameter, W^* retains its original definition as a row –standardized version of connectivity matrix W, with

(3.14) $$\{ w_{ij}^{***} \} = \frac{w_{ij}}{\sum_j w_{ij}}$$

where, $$w_{ij} = \begin{cases} 1 \; if \;\; j \; is \;\; connected \;\; to \;\; i \\ 0 \;\; otherwise \end{cases}$$

The term $\rho W^* Y$ is a crop response vector whose value in adjacent locations are specified by the (n x n) configuration matrix W^*, and are auto-correlated by an amount ρ, β is a vector of p slope parameters, X is an (n x p) matrix of classification indicator variables, and ε is the error vector. The null hypothesis of the nonexistence of autocorrelation is that $\rho = 0$. The basic idea with this model is to incorporate the spatial autocorrelation as a component of the model. In terms of individual components, the model in equation ***(3.13)*** can be expressed as:

(3.15) $$y_i = \rho \left(\sum_j w_{ij} ** y_j \right) + \sum_i x_i \beta_i + \varepsilon_i$$

Spatial autoregression model Assumptions

- The independence assumption centers on the errors rather than predictor variable X and response variable Y.

- All diagonal elements of W^{**} are zero.
- ρ and β are unknown parameters to be estimated.
- $(I - \rho W^{**})$ is nxn non-singular matrix.

Therefore, equation ***(3.13)*** is rewritten with respect to the error term ε as:

(3.16) $$\varepsilon = (I - \rho W^{**})Y - X\beta$$

where I is the identity matrix. Equation ***(3.16)*** specifies the basic form of the error variance-covariance matrix. Accordingly, the actual error sum of squares, D, for AR model with non-constant mean is

(3.17) $$D = \left[(I - \rho W^{**})Y - X\beta\right]^T \left[(I - \rho W^{**})Y - X\beta\right]$$

Suppose that the form of the autocorrelation model is specified as:

$B(Y - \mu) = D\varepsilon$, where $B = \{b_{ij}\}$ and $b_{ii} = 1$ for all i.

With the general form of the function defined, the covariance matrix of the observations on the areal unit becomes:

$$\begin{aligned} V &= E\left[(Y - \mu)(Y - \mu)^T\right] \\ &= E\left[\left(B^{-1}D\varepsilon\right)\left(B^{-1}D\varepsilon\right)^T\right] \\ &= E\left[B^{-1}D\varepsilon\varepsilon^T D^T\left(B^{-1}\right)^T\right] \\ &= E\left[B^{-1}DV_\varepsilon D^T\left(B^{-1}\right)^T\right] \\ &= E\left[B^{-1}DV_\varepsilon D^T\left(B^{-1}\right)^T\right] \\ &= \sigma^2\left[B^{-1}DD^T\left(B^{-1}\right)^T\right] \end{aligned}$$

Substituting $D = I$ and $B = (I - S)$, we have

$$V = \sigma^2 \left[(I - S)^T (I - S) \right]^{-1}$$

One particular form of S that is widely adopted among spatial statisticians is $S = \rho W^{**}$, where ρ is a constant spatial autocorrelation parameter and W^{**} is a row-standardized variation of connectivity matrix defined in equation (3.14).

Substituting $S = \rho W^{**}$, where S is the correlation matrix among regions i and j whose diagonal elements are all zero and assuming that $(I - \rho W^{**})$ is being non-singular, it is invertible, we have

(3.18) $$V = \sigma^2 \left[(I - \rho W^{**})^T (I - \rho W^{**}) \right]^{-1}$$

The covariance matrix defined by equation ***(3.18)*** is one of the most important future distinguishing AR- model from others. This way of defining the model also permits the establishment of a joint density function for Y, which is the basis for parameter estimation and interface with the model.

Estimation of the AR model requires the maximum likelihood (ML) technique. The ML is based on the principle of selecting value for the parameters that maximize the probability of the observed data. ML estimates of ρ consistently converge to the true parameter with increasing n [Cliff and Ord, 1975; Haning, 1978], are restricted to a feasible parameter space, and are associated with a probability density function that integrates to unity [Griffith, 1988a], this solve the complicated mathematical structure of the autocorrelation parameter, ρ.

The ML estimate of ρ is found by maximizing with respect to ρ the following nonlinear equation:

(3.19) $$L = (2\pi)^{\frac{-\pi}{2}} \det\left[(I - \rho W^{**})^T (I - \rho W^{**}) \right]^{\frac{1}{2}} (\sigma^2)^{\frac{-\pi}{2}} \exp\left[\frac{-D}{2\sigma^2} \right]$$

where $\det\left[\left(I - \rho W^{**}\right)^T \left(I - \rho W^{**}\right)\right]$ is the Jacobian (J) that is written in terms of the eigenvalues of matrix W^{**}. Equation ***(3.17)*** expresses the probability density function of the normally distributed errors of the response variable Y. For matrix W^{**}, n eigenvalues need to be calculated.

Equation ***(3.16)*** is iteratively estimated until the mean sum of squared error D, weighted by $\exp(J)^{\frac{1}{2}}$, riches a minimum, and the value of ρ is selected that is associated with the minimum. Conceptually, this procedure is similar to computing orthogonal component, in principal component to an uncorrelated mathematical space. The ML function accomplishes normalizing factor ensuring that this transformation yields probability density function where complete integration equals unity (Griffith, 1990).

3.3.1. Eigen Values of Neighboring Plots of Weighted Matrix

One of the steps in AR modeling is computing the eigenvalues from the neighboring plots configuration matrix. The purpose of this matrix is to describe the spatial connectivity of the sample locations of the crop response variable represented as arrow-by-column grid plots. Eigenvalues computed from this matrix reflect the systematic behavior in the arrangement, interconnectivity, and interdependence of the yield observations. Furthermore, the eigenvalues facilitate computation of the Jacobian of the transformation from an auto-correlated to an unautocorrelated mathematical space. In practice, AR-models frequently are specified in terms of matrix W^*, the stochastic form of binary matrix W (given in equation ***(3.14***)). Matrix W^* is derived, as shown above, by dividing each entry of the binary matrix of the corresponding row sum resulting

in each row summing to a unity. The principal eigenvalue should be equal to unity for matrix W^{*}. To simplify the computation of eigenvalues the square root row sum is used in the denominator of equation ***(3.14)*** above. That is,

$$\{ w_{ij}^{**} \} = \frac{w_{ij}}{\left(\sum_{j} w_{ij} \right)^{\frac{1}{2}}}$$

3.3.2. Estimating The Jacobian Term

Having computed eigenvalues from the stochastic configuration matrix W^{**}, the next step in AR modeling is computation of the Jacobian term. As stated above, this term provides a means of transforming the data from an auto-correlated to an unauto-correlated mathematical space.

Griffith(1990), specified computation of the Jacobian by identifying the following functions that accurately approximate this term based on the matrix W^{**} for regular lattice data.

(3.20)
$$J^{\frac{1}{2}} = 2\alpha_n \ln(\delta_n + \rho) - \alpha_n \ln(\delta_n - \rho)$$

where, α and δ are synthetic parameters for a rectangular girded region, and ρ is the autocorrelation parameter. The synthetic parameters are obtained by doing another least square regression of generated value of J on the sampled value of ρ. Values of J in equation ***(3.20)*** may be generated for values of ρ systematically sampled from the interval between the smallest and largest eigenvalues using:

(3.21)
$$J^{\frac{1}{2}} = \sum_{i=1}^{n} (-\ln(1 - \rho\lambda_i))$$

where the $\lambda_i^{'}$s are the eigenvalues extracted from matrix W^{**}, n is the number of observations, and ln is the natural logarithm [Griffith, 1988b). The value J is a logarithmic sum of n eigenvalues for the locations of n observations that are uniquely for depicted by a matrix W^{**}. The eigenvalues constrain the absolute value of ρ to be less than unity. The upper limit of this space, ρ =1, is determined by the principal eigenvalue of matrix W^{**}, which equals unity.

For computational purpose, multiplying both sides of equation ***(3.13)*** above by $\exp\left(J^{\frac{1}{2}}\right)$ the statistical form of AR-model is represented by:

(3.22)
$$Y\exp\left(J^{\frac{1}{2}}\right)=\left(X_k\beta_k+\rho W^{*}Y-\rho\beta_0\right)\exp\left(J^{\frac{1}{2}}\right)+\varepsilon\exp\left(J^{\frac{1}{2}}\right)$$

where, ρ is the autocorrelation parameter, X_k specifies binary indication variables for k different classis, rows and columns, β_0 expresses mean grain yield in units for the k^{th} class and X_k takes on a values 1 for observations belonging to the k^{th} classes and 0 otherwise.

Now, adjusting the yield observations for autocorrelation according to the equation:

(3.23)
$$Y_{adj}=Y-\left(\rho W^{*}Y-\rho\beta_0\right),$$

where, $Y-\left(\rho W^{*}Y-\rho\beta_0\right)$ equals the adjusted values of grain yield and ρ is the autocorrelation parameter obtained from its non linear estimation mean square error, given by:

(3.24)
$$MSE=\frac{SSE}{(n-p-1)},$$

where, the error sum of square (SSE) is retuned from the generalized linear model (GLM), p regression terms (including the intercept) and one degree of

freedom for the autocorrelation parameter estimate, $\hat{\rho}$. And compare the results of AR-based and Conventional ANOVA tables to draw conclusion.

For more basic information, potential user refer Griffith (1993) and Arlinghaus (1996, p-251-277), discussed the necessary steps for approximating the Jacobian, eigenvalues and Autoregressive response model for agronomic field trial. In the next section the method of spatial prediction of unobserved values form observed ones is discussed.

3.4. Optimal Prediction in Random Field

Consider the random field $\{z(s): s \in D \in \Re^2\}$ observed at locations $s_1, \ldots, s_n$,and corresponding data vector $Z(S) = [z(s_1), \ldots, z(s_n)]^T$. Following the notation in Cressie (1993), the domain $D \varepsilon \Re^2$ is fixed and continuous; hence, it is dealing with geo-statistical data. The sample is an incomplete observation of the surface $z(s,w)$ that is, out come of random experiment with realization W. One of the previous problems in spatial statistics is the prediction of Z at some specified location $S_0 \in D$, this can be a location that is part of the set of location, where, $Z(s,w)$ has been observed, or a new (or unobserved) location. In principle, this estimator, call it $Z(S_0)$, is calculated as a weighted linear combination of the measured values in the neighborhood.

3.4.1. Spatial Prediction and Kriging

Spatial prediction is a method that incorporates spatial dependence, which requires a model of the spatial continuity, or dependence in the form of covariance or semi varigram. Thus, it involves two steps, first, modeling the

covariance or semi variogram of the spatial process, this involves choosing both a mathematical form and the values of the associated parameters. The second step is using the dependence model in solving the kriging system at specified set of spatial points, which result in predicted values and associated standard errors.

3.4.2. Use of the Variogram for Lattice Data

The standard tool used in analysis of geo-statistical data is the variogram. When the variogram is applied to a lattice data, most commonly, the data associated with each region are assumed to have been observed and arbitrarily assigned at the center of the region. Distance between centroids is then used to develop the spatial covariance structure through the variogram function directly. This arbitrariness of assigning at the center of centroied causes concern because the spatial structure estimated by the variogram depends heavily on the distance between observations. The investigation here is that what happens to the estimation of the variogram when each lattice value is placed randomly with its associated region; and examine that this randomly placed location has on the empirical variogram, the fitted theoretical variogram, and testing for existence of spatial correlation.

Lattice data analysis was developed to model data collected as summaries of some process over a regions, which most often from a partition of the whole space of interest. Usually lattice data involves method of adjacencies or neighborhood status.

Given observations $Z(S) = Z(s_1), \ldots, Z(s_n)$ from spatial lattice process $Z(S)$, where each observation represents the value associated with each region S_i in the lattice, and suppose

$Var(Z(s_1) - Z(s_2)) = 2\gamma(s_1 - s_2)$ for all $s_1, s_2 \varepsilon D$. The quantity $2\gamma(s_1 - s_2)$ which is a function of only the increment $s_1 - s_2$ is called the variogram (and $\gamma(s_1 - s_2)$ is called a semi variogram)

or

(3.24) $$\gamma(h) = \frac{1}{2}\text{var}[Z(s+h) - Z(s)]$$

where h is lag distance between each region said s_i, s_j and estimation of the variogram is given as:

(3.25) $$\hat{\gamma}(h) = \frac{1}{2}\frac{1}{|N(h)|}\sum_{N(h)}\{Z(s_i) - Z(s_j)\}^2$$

where, $N(h) = \{(s_i, s_j) : s_i - s_j = h, \quad i, j = 1, \ldots, n\}$

The aim is to investigate the use of variogram for lattice data, in particular examine how robust its estimation is to employ geographic simulation techniques in which we let the spatial point location associated with each; and to investigate the effect of this random location (regular square lattice, like the field plot) has on the estimate of spatial variogram. Gastwirth and Rubin (1975) have investigated the behavior of robust estimator for dependent data. They consider the spatial case of location estimation in the presence of serial correlation and show that robust estimators obtain higher efficiency relative to the mean for Gaussian observations. Deleting outliers when estimating a variogram may be sensible but when predicting observations, an alternative way of dealing with them is needed. One of the methods of treating is applying random spatial location effect for regular lattice.

3.4.2.1. Random Spatial Location Effect for Regular Lattice

A regular lattice is a region for which each of its sub regions is of equal size and of the same shape. The data referred to as spatial lattice data, are summaries of each sub region of the lattice. Given, a realization of a spatial process from a kxk regular lattice, considering how much the estimation of the variogram changes if spatial location of each data point is randomly chosen with in their respective region. In general, given a lattice process, there is no correct point location since each observation represents a summary of the entire cell. Thus, observations are treated as if they were from some unknown process on the regular lattice.

3.4.2.2. Variability in the Empirical Variogram

We now consider how the random placement of the spatial location with in each region affects the empirical variogram estimation. We assume for simplicity throughout that the variogram is isotropic and use h to represent the length of $s_i - s_j$. Because it can happen that only one pair of locations are exactly h apart and thus very little information would be used for each point estimator of, $\gamma(h)$ we carry out a grouping over nearby values of h. That is, we make a set of chosen lag of interest $h(l), l = 1, \ldots, k$ and let $2\hat{\gamma}(h) = ave\{(Z(s_i) - Z(s_j)): (i, j) \in N(h);\ h \in T(h(.))\}$, where the "tolerance region" $T(h(.))$ distance around , $h(l), l = 1, \ldots, k$, and chosen to be disjoint.

This is similarly done for the robust variogram estimator $\bar{\gamma}(h)$, given by Cressie and Hawkins (1980),

(3.26)

$$\bar{\gamma}(h) = \frac{1}{2\left(0.457 + \frac{0.494}{|N(h)|}\right)} \left\{ \frac{1}{|N(h)|} \sum_{N(h)} |Z(s_i) - Z(s_j)|^{\frac{1}{2}} \right\}^4$$

Empirical variograms are formed separately for the two different increments of tolerances. Thus $\hat{\gamma}(h)$ is the average of the pair wise differences, $\frac{[(Z(s_i) - Z(s_j))]^2}{2}$, over a number of pair of observation that are with in a distance $h(l)$ of each other. The separation distance $h(l)$ is taken as the average of the pairwise distance that fall into the tolerance interval. By plotting the $\hat{\gamma}(h(l))$ versus $h(l)$, we set the empirical classical variogram. The procedure is similar for the empirical robust variogram. The true exponential variogaram that the data were generated from is also overlapped. The empirical variograms obtained at centroid are close to the true variogram, while random location pattern flattens quite a bit around the true variogram especially for small lag. At larger lags each of the empirical variogram deviates from the theoretical are based on just one realization of spatial process and so we should not expect them to follow the theoretical variogram exactly, specially at large lags where the sampling variability is expected to be larger.

The locations are generated from uncorrelated bi-variate uniform distribution placed on each square in the grid. The first thing is fit the classical and robust empirical variograms and control the tolerance for each of the random spatial patterns. For each of the two cases (classical and robust) we overlap the resulting kxk empirical variograms in one plot. This variability is calculated using the approximation formula given in Cressie (1993, p-96),

(3.27)

$$Var\left(2\hat{\gamma}(h(l))\right) \cong \frac{2\left\{\gamma(h(l));\ \hat{\theta}\right\}^2}{|\ N(h(l))\ |},$$

where $\hat{\theta}$ are the fitted parameter values from the theoretical exponential model fitting the empirical classical variogram at central location. The reason for including the sampling variability is that when a variogram is used for lattice data, it contains two types of variability: the usual sampling variability and an additional variability due to uncertain location. By comparing the empirical variability due to the random or uncertain location with the sampling variability of the variogram estimators, we capture how much influence the lattice structure has on the variability of the variogram.

3.4.2.3. Variability in Estimated Theoretical Variogram

Now, the interest is examining how the fit of a theoretical variogram model to each of the empirical variogram is affected. Using nonlinear ordinary least squares to fit an exponential variogram model to each of the empirical variogram estimates in exponential variogarm is:

(3.28)

$$\gamma(h\ ;\theta) = \begin{cases} 0\ , & if\ \ h = 0 \\ c_0\ + c_e\left\{1 - \exp\left(-\|\ h\ \| \big/ a_e\right)\right\}, & if\ \ h \neq 0 \end{cases}$$

$$\theta = \left(c_{0\ ,}\ c_e,\ a_e\right)^T, where\ \ c_0 \geq 0,\ \ c_e \geq 0,\ and\ \ a_e \geq 0$$

The parameters the nugget, the partial sill, and the range are represented by $c_{0\ ,}\ c_e, and\ a_e$, respectively. And lastly, fitting the theoretical exponential variograms generated at kxk different random locations.

Once an appropriate varigram model is chosen the next step is producing the kriging surface prediction model. After an appropriate variogram model is chosen, there are a number of choices involved in producing the kriging surface.

Models for spatial autocorrelation depend on the distance and direction separating the two locations, and are constrained so that for all possible set of locations, the covariance matrices implied from the models remain negative definite. Based on spatial correlation, optimal linear prediction can be constructed that can yield complete maps of spatial fields from incomplete and noisy spatial data. This methodology is called kriging if the data are of only one type and cokriging if it is of two or more variable type (Jay, Cressie and Paul, 2004).

3.4.3. Basics of Kriging

Kirging is a method of interpolation named after a South African mining engineer named D. G. Krige (1951), who developed the technique in an attempt to more accurately predict ore reserves. Kriging is based on the assumption that the parameter being interpolated can be treated as a regionalized variable. A regionalized variable is intermediate between a truly random variable and a completely deterministic variable in that, it varies in continuous manner from one location to the next and therefore points that are near each other have certain degree of spatial correlation, but points that are widely separated are statistically independent Davis (1986). Cressie (1993), more precisely stated as: a kriging prediction is a weighted linear combination of all output values already observed. These weights depend on the distance between the inputs already simulated. Kriging assumes that the closer the inputs are, the more positively correlated the output are.

In deterministic simulation, kriging has an important advantage over regression analysis: kriging is an exact interpolator; that is, predicted values at observed input value are exactly equal to observed (simulated) output values. In random simulation, however, the observed out put values are only estimates of the true values, so exact interpolation loses intuitive appeal. Therefore, regression uses OLS, which minimizes the residual-squared and summed over all observations. Since estimation and prediction for spatial data result in different interpolations, the differences are explained also graphically by example.

3.4.3.1. Ordinary Kriging

Ordinary kriging is the most commonly used type of kriging, often associated with the acronym **B.L.U.P.** for "best linear unbiased predictor." It is *linear*, because its estimates are weighted linear combinations of the sample values. "*Unbiased*" means that the estimated value are neither overestimated nor underestimated systematically, i.e., local-over and underestimation neutralize each other, and the average estimation error is zero. Ordinary kriging is "best", because it aims at minimizing the variance of the error for each interpolated point; i.e., it assigns optimal weight to each sample point considering not only the distance between the sample and the unknown point, but also the distance between the samples themselves.

The first step in ordinary kriging is to construct a variogram from the scattered plot set to be interpolated. A variogram consists of two parts: an experimental variogram and a model variogram. The experimental variogram is calculating the variance of each point in the set with respect to each of the other points and plotting the variances versus distance (lag) between points. The regular two-dimensional arrangement of the geo- referenced data allows standard time series

lag functions to be used in order to identify pixels laying immediately to the east and south of a given areal unit. Invoking this function necessitates the addition of missing values column to the map, since the extreme eastern pixels have no values immediately to their east. By including a sequential numbering scheme, the data can be recorded by storing on this numbering variable. The consequence is that a lag function can be used to identify values within the boundary that are adjacent to each pixel. Regular square tessellation data requires for similar computer code for undertaking spatial statistical analysis.

Additional problems with kriging, and with spatial estimation methods in general, are related to the necessary assumption of ergodicity of the spatial process. This assumption is required to estimate the covariance or semivariogram from sample data. Details are provided in Cressie (1993, pp. 52–58). Despite these difficulties, ordinary kriging remains a popular and widely used tool in modeling spatial data, especially in generating surface plots and contour maps. An abbreviated derivation of the ordinary kriging estimator for point estimation and the associated standard error is discussed in the following section. Full details are given in Journel and Huijbregts (1978), Christakos (1992), and Cressie (1993).

3.4.3.1.1. Ordinary Kriging Assumptions

The typical; assumption for the practical application of ordinary kriging are:

- Intrinsic stationarity or wide sense stationarity of the field.
- Enough observations to estimate the variogram.

The mathematical conditions for applicability of ordinary kriging are:

- The mean $E\big(Z(s)\big) = \mu$ is unknown but constant.
- The variogram $\gamma(s_1, s_2) = E\big[Z(s_1) - Z(s_2)\big]^2$ of $Z(s)$ is known.

Following the notation in Cressie (1993), the model assumption for $Z(s)$ is :

(3.29) $$Z(s) = \mu + \varepsilon(s)$$

Here μ is the fixed unknown mean of process and $\varepsilon(s)$ is zero mean of spatial random field representing the variation around the mean.

In most practical applications, an additional assumption is required in order to estimate the covariance C_z of the $Z(s)$ process. This assumption is second order stationarity.

$$C_z(s_1, s_2) = E\left[\varepsilon(s_1)\varepsilon(s_2)\right] = C_z(s_1 - s_2)$$

This requirement can be relaxed slightly when Variogram is used instead of the covariance. In this case, second-order statinarity is required of the differences $\varepsilon(s_1) - \varepsilon(s_2)$ rather than $\varepsilon(s)$:

(3.30) $$2\gamma_z(s_1, s_2) = E\left[\varepsilon(s_1) - \varepsilon(s_2)\right]^2 = 2\,\gamma_z(s_1 - s_2)$$

By applying local kriging, the spatial process represented by the equation ***(3.30)*** for $Z(s)$ are more general than they appear. In local kriging at an unsampled location s_0, a separate model is fit using only data in a neighborhood of s_0. This has the effect of fitting a separate mean at each point.

Given the n known measurements at locations $s_1, \ldots, s_n$,and corresponding data vector $Z(S) = \left[z(s_1), \ldots, z(s_n)\right]^T$, to obtain an estimate $\hat{\boldsymbol{Z}}$ of $\boldsymbol{Z}$ at an unsampled location s_0. When the following three requirements are imposed on the estimator $\hat{\boldsymbol{Z}}$ the ordinary kriging estimation is obtained:

(i). $\hat{Z}$ is linear in $z(s_1), \ldots, z(s_n)$.

(ii). $\hat{Z}$ is unbiased

(iii). $\hat{Z}$ minimizes the mean-square prediction error $E\left[z(s_0) - \hat{z}(s_n) \right]^2$

Linearity requires the prediction assumption of the following form for $\hat{Z}(s_0)$

(3.31) $$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i z(s_i)$$

Applying the unbiasedness condition to equation ***(3.31)*** yield

$$E\left[\hat{Z}(s_0) \right] = \mu = \sum_{i=1}^{n} \lambda_i E[z(s_i)] \Rightarrow \sum_{i=1}^{n} \lambda_i \mu = \mu \Rightarrow \sum_{i=1}^{n} \lambda_i = 1$$

Finally, the third condition requires a constrained linear optimization involving $\lambda_1, \lambda_2, \ldots, \lambda_n$ and Langrage parameter 2m. This constrained linear optimization can be expressed in terms of the function $L(\lambda_1, \lambda_2, \ldots, \lambda_n, m)$ given by

$$L = E\left[\hat{Z}(s_0) - \sum_{i=1}^{n} \lambda_i z(s_i) \right] - 2m\left(\sum_{i=1}^{n} \lambda_i - 1 \right)$$

Define nx1 column vector λ by

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$$

And the (n+1)x1 column vector λ_0 by

$$\lambda_0 = (\lambda_1, \lambda_2, \ldots, \lambda_n, m)^T = \begin{pmatrix} \lambda \\ m \end{pmatrix}$$

The optimum is solved by solving:

(3.32) $\frac{\partial L}{\partial \lambda_0} = 0$ in terms of $\lambda_1, \lambda_2, \ldots, \lambda_n$ and m.

The resulting matrix equation of ***(3.32)*** can be expressed in terms of ordinary kriging equation.

(i). Ordinary kriging Equation

Once the experimental variogram is computed, the next step is to define a model variogram. A model variogram is a simple mathematical function that models the trend in the experimental variogram. From the requirements of ordinary kriging stated above the kriging weight fulfils the unbiasedness condition $\sum_{i=1}^{n} \lambda_i = 1$, and result from equation ***(3.32)*** are given by ordinary kriging equation system:

$$\begin{pmatrix} \lambda_1 \\ . \\ . \\ . \\ \lambda_n \\ m \end{pmatrix} = \begin{pmatrix} \gamma(s_1,s_1) & . & . & . & \gamma(s_1,s_n) & 1 \\ . & . & & & . & \\ . & & . & & . & \\ . & & & . & . & \\ \gamma(s_n,s_1) & . & . & . & \gamma(s_n,s_n) & 1 \\ 1 & . & & . & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(s_1,s_0) \\ . \\ . \\ . \\ \gamma(s_n,s_0) \\ 1 \end{pmatrix}$$

The additional parameter ***m*** is a Langrage multiplier used in minimization of the kriging error $\sigma^2{}_k(s)$ to honor the unbiasedness condition.

Using this solution for λ and m, the ordinary kriging estimate S_0 is:

(3.33) $\hat{Z}(s_0) = \lambda_1 z(s_1) + \lambda_2 z(s_2) + \ldots + \lambda_n z(s_n)$

(ii). Ordinary kriging interpolation

The interpolation by ordinary kriging is given by:

$$\hat{Z}(S_0) = \begin{pmatrix} z_1 \\ . \\ . \\ . \\ z_n \\ 0 \end{pmatrix}^T \begin{pmatrix} \gamma(s_1,s_1) & . & . & . \gamma(s_1,s_n) & 1 \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ \gamma(s_n,s_1) & & & \gamma(s_n,s_n) & 1 \\ 1 & & . \ . \ . & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(s_1,s_0) \\ . \\ . \\ . \\ \gamma(s_n,s_0) \\ 1 \end{pmatrix}$$

These formulae are used in the best linear unbiased prediction (BLUP) of random variables, Cressie (1993, pp.119-123).

(iii). Ordinary Prediction error

The associated prediction error is given by:

$$Var\left(\hat{Z}(S) - Z(S)\right) = \begin{pmatrix} \gamma(s_1,s) \\ . \\ . \\ . \\ \gamma(s_n,s) \\ 1 \end{pmatrix}^T \begin{pmatrix} \gamma(s_1,s_1) & . & . & . \gamma(s_1,s_n) & 1 \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ \gamma(s_n,s_1) & & & \gamma(s_n,s_n) & 1 \\ 1 & & . \ . \ . & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(s_1,s) \\ . \\ . \\ . \\ \gamma(s_n,s) \\ 1 \end{pmatrix}$$

or $$\sigma^2{}_k(s_0) = \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_j) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(s_i - s_j)$$

3.4.3.1.2. General Properties of Kriging

- The kriging estimation is unbiased: $E\left[\hat{Z}(s_i)\right] = E[Z(s_i)]$
- The krging estimation honors the actually observed value: $\hat{Z}(s_i) = Z(s_i)$
- The kriging estimation $\hat{Z}(S_0)$ is the best linear unbiased predictor of $Z(S_0)$ if assumptions hold. However;
    - As with any method: if the assumptions do not hold, kriging may bad
    - There might be better non-linear and/or biased methods,
    - No properties are guaranteed, when the wrong variogram is used. However, typically still a 'good' interpolation is archived.
    - Best is not necessary good: for example, in case of no spatial dependence, the kriging interpolation is only as good as the arithmetic mean.
- Kriging provides σ^2_k as a measure of precision. However, this measure relies on the correctness of the variogram.

Kriging is more sophisticated method that interpolates across space according to spatial lag relationship that has both systematic and random components. This can be accommodating a wide range of spatial relation ships for the hidden value between observed locations. Kriging provides optimal estimates that can be mapped to determine if spatial pattern exists, and assumes that the output are closer.

CHAPTER IV

RESULT AND DISCUSSION

In this chapter results and discussion based on the potato yield data using out put from SAS software and Gplot results are presented in the order they appeared in the methodology part of the study.

A potato uniformity trial of 20x12 unit lattice layout, with basic units of 1.5x1.2 m^2 was used to demonstrate the relationship between small (basic) plots by aggregating cells for larger plot size. Uniformity trial is where the same crop and the same soil management practices are applied to whole experiments, to study variability caused by the soil and related environmental conditions. Plot sizes and potato yield values of all units, regardless of how similar or dissimilar they are, will be presented by summary measure, such as arithmetic averaging and weights. For this thesis, small plots are merged to basic units in the analysis of different plot sizes. Because (dis)similarity among areal units is not uniform over large plots and in all directions, merging smaller areal units using different spatial partitioning schemes is the same as smoothing different combinations of spatial neighbors.

Using 1638 basic units from Hollota agricultural research center and 931 basic units from Kulumsa agricultural research center (Appendix I and II for the original data), the average of potato yield were 2.55 and 5.36 Kg/m^2 respectively. The data consist of potato yield data measurements (in kg/m^2)

taken over an approximately a square area. The coordinates are offsets from a point in the southwest corner of the measurement area, with the north and east distances in units of meters. It is also instructive to see the location of measured points in the area where we want to perform spatial prediction. If this is not the case, the prediction error might be unacceptably large where measurements are sparse.

4.1. Diagnosis for Spatial Randomness and Spatial Autocorrelation

4.1.1. Monte Carlo Test for Completely Spatial Randomness

The Monte Carlo test simulates values of the test statistics under completely spatial randomness and comparing them to the corresponding statistics calculated from the observed pattern. To apply the method discussed in sections 3.2.3.1 and 3.2.3.2, the simulated values of Moran's I and Geary's C for small plots combined for the data given in Table 4.1 and Table 4.3 for potato yield data from Hollota and Kulumsa research centers respectively. The data considered is for a plot size of 12 m^2 for each site to test for completely spatial randomness and SAS code for simulation is given in Appendix III, page 122.

Table 4.1. The data for total potato yield from Hollota (plot size of 12m^2)

		Column					
		1	2	3	4	5	6
Row	1	27.10	30.23	29.04	30.86	26.10	29.80
	2	32.40	34.76	29.94	28.73	27.70	32.25
	3	30.14	32.42	27.70	31.66	28.70	26.74
	4	32.20	29.92	34.50	30.97	30.44	31.55
	5	34.21	33.68	29.72	29.62	28.26	28.32
	6	38.32	34.66	34.00	34.00	32.62	31.66
	7	24.34	23.26	22.74	21.26	22.72	21.54
	8	25.40	26.12	22.00	24.48	24.48	26.98
	9	19.66	21.96	21.64	21.14	24.80	23.02
	10	28.84	30.53	31.32	29.14	32.64	34.37
	11	32.82	28.73	31.62	31.08	33.66	32.50
	12	33.04	32.24	29.20	31.62	34.36	32.56
	13	37.60	35.44	34.44	34.74	36.08	32.42
	14	20.58	22.19	28.72	28.99	33.09	35.84
	15	29.32	31.26	29.76	29.62	31.08	29.91
	16	25.48	27.12	29.16	29.66	25.40	26.16
	17	34.84	34.86	31.99	32.05	28.92	32.20
	18	39.62	34.56	36.04	33.10	27.58	28.86
	19	37.12	32.56	32.34	33.62	33.38	31.44
	20	39.56	31.28	34.98	33.48	30.74	30.04

The Monte Carlo test is implemented by simulating values of the test statistics I and C, under the null hypothesis of CSR and comparing these values to the corresponding statistics calculated using Moran's I and Geary's C from the potato yield observed values. The result of Monte Carlo test is shown in Table 4.2 for Hollota and Table 4.4 for Kulumsa.

Table 4.2. Ranked Monte Carlo Test For Completely Spatial Randomness Ranks For Moran's index I Geary's coefficient C- observed Value is Last value:

3x4 Plot Size (Hollota)

I	RANK	C	RANK
-0.084265	4	0.83538653	30
-0.0416473	21	1.04356044	54
-0.0519689	17	1.03201748	53
0.02000641	68	1.05624368	59
-0.0199806	31	1.04805046	56
0.02168668	70	0.51246861	11
0.04491852	82	0.96367826	41
0.01689919	65	0.94205987	39
0.05169187	86	0.48307744	4
0.03225022	75	1.24586693	78
0.00240347	51	0.98431666	43
0.14162361	98	1.46377282	84
-0.0199526	32	1.01947918	49
-0.0524504	15	1.12200078	69
0.01217576	57	1.15823206	75
0.01507152	62	0.53163311	16
-0.0058758	43	0.51423338	12
-0.0100218	42	0.55268512	18
-0.0684436	8	1.02837873	51
-0.0649951	9	1.60672541	96
0.00377165	53	1.28057405	79
0.04143953	79	0.45353448	2
0.12491758	95	0.98646945	44
0.04437988	81	0.62078733	23
-0.0391059	24	0.66433079	26
-0.0104933	40	1.15260423	74
0.01416217	59	1.55265703	92
-0.0556054	13	1.01988838	50
-0.0523065	16	1.1281487	70
0.01460917	61	0.52368841	15
-0.0004581	48	0.6199503	22
-0.029773	29	0.88427355	34

0.00146674	50	1.36944708	81
-0.0506539	19	1.48561039	85
-0.0012094	46	1.00806419	47
-0.015477	37	1.0924862	67
-0.0366257	25	1.63268934	97
0.03492971	77	0.92327276	38
0.03291643	76	0.88550359	35
0.04626623	83	0.86757468	32
-0.0047989	44	0.50915496	10
-0.0169195	35	0.50804869	9
-0.0341168	28	1.60088045	95
0.0597664	89	0.51944052	13
0.04133179	78	1.04527755	55
0.0524375	87	1.54969852	91
-0.0554019	14	1.0120222	48
0.13379391	97	1.33286907	80
0.04171219	80	1.49581651	87
0.09199504	93	1.50871883	88
0.07563303	91	0.9014325	37
0.00972542	55	0.89517236	36
0.00258768	52	0.39352678	1
0.04626743	84	1.0316706	52
0.0260496	73	0.85280624	31
-0.0234529	30	0.65685144	25
0.01442424	60	1.57633884	94
0.01686693	64	0.81393306	29
0.13044161	96	1.51659685	90
0.00040223	49	0.47532043	3
0.00587118	54	1.1626944	76
-0.0341477	27	1.51385989	89
0.051614	85	1.08449778	66
-0.0011392	47	1.7871902	99
-0.292341	1	1.0545973	58
-0.0508103	18	1.49040716	86
-0.0109349	39	0.96338466	40
0.02636884	74	0.54493571	17
-0.0181428	34	0.4979903	7
0.01322818	58	1.07115133	63
0.08550492	92	1.07217435	64

0.1107413	94	1.0684561	62
-0.0724368	6	1.0056421	63
-0.1588424	2	0.485626	5
0.02265958	71	0.56016949	19
0.06608787	90	0.99784043	46
-0.0907916	3	1.14011345	73
-0.0040406	45	0.87542473	33
-0.061893	10	1.08392614	65
-0.0104007	41	1.13940443	72
-0.0410472	22	1.37991948	82
0.14531628	99	0.72134488	27
-0.0616676	11	1.56725419	93
-0.0358303	26	1.05827493	60
0.01060249	56	0.96599198	42
-0.0440682	20	0.4926157	6
0.0241294	72	1.20539454	77
0.01988143	67	0.60517316	21
-0.0586674	12	1.66750177	98
-0.069276	7	1.13731375	71
-0.0780606	5	1.0669618	61
-0.0196682	33	1.10193318	68
0.0160576	63	0.98840402	45
0.05920951	88	0.52301252	14
-0.0139275	38	0.56203322	20
-0.0164698	36	1.04941257	57
0.01752859	66	0.78051067	28
0.2420311	100***	1.89400643	100***

***** Significance at $\alpha = 0.05$ level**

The Monte Carlo test is applied on potato yield data from Tables 4.1 I_1 or C_1 (the last values in Tables 4.2) was compared for each site separately to ranks $I_{(i)}$'s and /or $C_{(i)}$'s calculated from 99 simulations of 120 observations on field trial of 0.15 ha from Hollota agricultural research centers. At $\alpha = 0.05$ level of significance, from Tables 4.2, the values of the relative variance (I) greater than one indicates that the potato yield data are clustered, and if values

of I are less than zero, this indicates a tendency for regular spacing Crissie (1993).

The results in Table 4.2 also shows $I_1 = I_{(100)}$ and $C_1 = C_{(100)}$ for Hollota. The Monte Carlo test for complete spatial randomness discussed in section 3.2.3.2 states that, reject the null hypothesis for complete spatial randomness if I_1 and/or C_1 is one of the smallest or one of the largest order statistic (depending on the direction of the alternative hypothesis with regard to I). Therefore, under the null hypothesis of CSR, we reject the null hypothesis since I_1 and/or C_1 is the largest order statistics as seen in Table 4.2 $I_1 = I_{(100)}$ and $C_1 = C_{(100)}$ in the rank based on I and C . Thus, it can be concluded that there is no complete spatial randomness among plot yields in Hollota agricultural research center.

Table 4.3. The data for total potato yield from Kulumsa plot (size of 12m^2)

Table 4.4. Ranked Monte Carlo Test For Completely Spatial Randomness - observed value is the last value: 3x4 plot size (Kulumsa)

I	RANK	C	RANK
-0.037896	**26**	**1.04955379**	**76**
-0.0831681	**12**	**1.05530358**	**81**
-0.0251382	**34**	**1.02730189**	**69**
-0.109863	**5**	**1.08799995**	**93**
-0.0792679	**15**	**1.07475328**	**90**
0.10181445	**95**	**0.84720087**	**4**
-0.0228346	**35**	**1.03131773**	**70**
0.03709801	**75**	**0.96245206**	**31**
0.00198987	**54**	**1.00031636**	**57**
0.04231452	**78**	**0.93691395**	**20**
-0.0941257	**9**	**1.13919609**	**97**
0.03444867	**73**	**0.96414228**	**33**
-0.0663299	**20**	**1.00823141**	**60**
-0.0258988	**33**	**0.98977378**	**51**
0.0015498	**53**	**1.01948784**	**64**
0.0286265	**67**	**0.9818979**	**45**
-0.106483	**6**	**1.09845194**	**94**
-0.019046	**38**	**0.9983187**	**56**
0.1986118	**99**	**0.82163833**	**2**
-0.0151836	**43**	**1.03334651**	**71**
0.13585115	**98**	**0.86565402**	**5**
-0.0811957	**13**	**1.07300621**	**89**
0.03451207	**74**	**0.9626963**	**32**
-0.1556123	**2**	**1.17322566**	**99**
-0.1849021	**1**	**1.1943508**	**100**
0.01944568	**61**	**0.95410381**	**29**
0.03271891	70	0.96428806	35
-0.0166065	40	1.02638213	68
0.08907198	92	0.92486629	12
-0.0283339	31	1.05096197	78
-0.0083978	47	0.99499097	54
-0.0517914	22	1.05349912	80
-0.0068422	50	1.02221532	65
-0.0114605	45	0.99040853	52

5	6
7	65
1	66.7
2	65.2
8	66
1	61.3
5	56.7
6	59.5
5	57.1
9	62.4
3	65.4
5	61.5
1	61.1

0.05426746	82	0.94833664	28
-0.0777257	16	1.06690625	86
0.13533415	97	0.83605906	3
0.0387151	76	0.99527694	55
0.00869749	56	0.98793013	50
-0.0071184	48	0.98633716	49
-0.0748012	18	1.05980565	83
-0.015546	42	1.00772529	59
0.08054021	91	0.92082574	11
0.07578282	90	0.93012779	17
-0.0282975	32	1.01717723	62
0.05008304	80	0.94748938	27
-0.0968634	8	1.06827207	88
-0.0761217	17	1.08391837	92
-0.0709974	19	1.05591416	82
0.02198538	64	0.96925784	37
-0.0120862	44	0.98545793	48
0.03419517	72	0.93534416	19
0.06633772	87	0.94673098	25
-0.0071135	49	1.02348331	67
0.02964555	69	0.97803521	42
0.07499638	89	0.92611526	13
-0.0034283	51	0.94700242	26
0.0416903	77	0.94251127	23
0.05689458	84	0.93385101	18
-0.0369486	27	1.03815794	72
-0.1265933	3	1.14340631	98
0.00604113	55	0.96670474	36
-0.1041104	7	1.12818047	96
0.12219043	96	0.87574568	6
-0.1122674	4	1.10325323	95
0.05342277	81	0.90435005	9
0.01373033	59	0.97285841	38
-0.0161862	41	0.98426775	46
-0.0589954	21	1.04535149	75
0.01995727	63	0.97595339	41
-0.079319	14	1.06072182	84
-0.0293474	30	1.05289399	79
0.02565656	66	0.92955034	16

-0.0215858	36	1.06811855	87
0.01342537	58	0.94394334	24
-0.0859353	11	1.06372546	85
0.09112073	93	0.89812609	7
-0.0407369	25	1.01748691	63
-0.0350556	28	1.04265269	74
0.02923292	68	0.99097163	53
-0.086891	10	1.07876981	91
-0.0188627	39	0.98128443	43
0.00059772	52	0.94153766	22
0.10179084	94	0.90213222	8
0.02263779	65	0.92627679	14
0.01396585	60	0.97518847	40
0.03383861	71	0.98492024	47
-0.0515941	23	1.0506783	77
0.01960086	62	0.9818073	44
0.06272543	86	0.90831747	10
-0.04171	24	1.03857372	73
0.07333778	88	0.92942526	15
-0.0113717	46	1.01474135	61
-0.0202152	37	1.02263294	66
0.05526329	83	0.96416305	34
0.04505544	79	0.95529353	30
-0.0342148	29	1.00738536	58
0.49066753	100***	0.48698223	1***

***

Significance at $\alpha = 0.05$ level

Similar to the above analysis the Monte Carlo test is applied on potato yield data from Tables 4.3 and I_1(the last values in Table 4.4) was compared for each site separately to ranks $I_{(i)}'s$ calculated from 99 simulations of 72 observations on field trial of 0.15 ha from Kulumsa agricultural research centers . At $\alpha = 0.05$ level of significance, from Table 4.4, the values of the relative variance (I) greater than one indicates that the potato yield data are clustered, and if values of I are less than zero, this indicates a tendency for regular spacing Crissie (1993).

The results in Tables 4.4 also shows $I_1 = I_{(100)}$ in case of Kulumsa $I_1 = I_{(100)}$ for rank based on I and $I_1 = I_{(1)}$ for rank based on C. The Monte Carlo test for complete spatial randomness discussed in section 3.2.3.2 states that, reject the null hypothesis for complete spatial randomness if I_1 is one of the smallest or one of the largest order statistic (depending on the direction of the alternative hypothesis with regard to I). Therefore, under the null hypothesis of CSR, we reject the null hypothesis since I_1 is the largest in terms of I and the smallest order statistics in case of C shown in Table 4.4 $I_1 = I_{(100)}$ for rank based on I values and $C_1 = C_{(1)}$ for rank based on C. Thus, it can be concluded that there is no complete spatial randomness among plot yields in Kulumsa agricultural research center. So, once complete spatial randomness is rejected, the next step in a spatial analysis may be to measure the departure from CSR.

4.1.2. Testing for Spatial Autocorrelation

From observed data given in Appendix I and II, the basic unit of potato yield data were combined to simulate different sizes and shape for analysis. The Moran's index I and Geary's coefficient C, are among the most widely implemented measures of spatial autocorrelation between 'neighboring' plots was briefly discussed in the methodology part. In this section, focus is on their application to particular data analysis, the essential task being to seek for spatial pattern. The contiguity matrix or weighted matrix, discussed in section 3.2.3.4, is plot yields that have rook's case contiguous relations for plot size 24 m^2 (6-by-4m) potato yield data from Holota agricultural research center is considered in Table 4.5 to describe neighboring plot relations.

Table 4.5. Weighted matrix W for Neighboring Relation between Plots of 6-by-4 plot variety trial (with portion shown for plots 1-14 and 33-36)

column

row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	33	34	35	36
1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
5	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
6	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0
8	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0
10	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0
12	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0

33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

4.1.2.1. The Moran's Index, I, and Geary's Coefficient, C.

A fundamental aspect of the spatial analysis is characterization of spatial dependence, showing how the observed values are correlated in space. With this context, the spatial autocorrelation is used for estimating how much the observed value of a potato yield is dependent on the values of the same variable in neighboring plots. The significance of a particular autocorrelation statistics is derived in two ways. First, by randomization procedure whose null hypothesis

proposes that the sample is randomly chosen. The second, significance can be tested using statistics based on the normal approximation. The values were then tested for their Moran's I and Geary's C simultaneously.

The main objective of estimating spatial autocorrelation coefficients (Moran's I and Geary's C) is to measure the strength of spatial autocorrelation among neighboring plot yields, to seek for spatial pattern or to diagnosis spatial dependence in regration model and test the assumption of spatial independence or randomness for potato yield data collected from Hollota and Kulumsa agricultural research centers. The estimated result of Moran's index I is also applicable model specification, specially, in testing of regression residuals.

These coefficients are computed to test the null hypothesis: no spatial autocorrelation between yields of neighboring plots. Moran's I and Geary's C summary statistics for autocorrelation test under both normality and randomization assumptions. Moran's index serves as a test where the null hypothesis is the spatial independence; in case its value would be zero. Depending on the configuration of the weights used, the upper and lower limit of this index typically range from approximately +1, representing completely positive spatial autocorrelation, to approximately -1, representing complete negative spatial autocorrelation, and with zero result representing no spatial autocorrelation, that indicates the pattern of values is random across space. Geary's C is computed identically to Moran's I with the exception of the numerator W_{ij} term which does not include the difference of each variate from the mean but rather each variate from all other variates.

Table 4. 6. Summary statistics of spatial autocorrelation, Moran's I and Geary's C, for Hollota Site with different size. Hollota Site (Normal Approximation)

Plot size (m^2)	Width (m^2)	Length (m^2)	N	X_{BAR}	Moran's I					Geary's C				
					I	E(I)	Variance	z value	P-value	C	E(C)	Variance	z value	P-value
4	2	2	360	10.1166111	0.21741567	-0.0027855	0.00262279	3.157337**	0.00079609	1.15525111	1	0.00283504	1.16982137	0.12103643
12	3	4	120	30.1829167	0.49066753	-0.0084034	0.00463128	7.333507**	1.213E-13	0.48698223	1	0.00535551	-7.01023**	1.1897E-12
24	4	6	60	60.6998333	0.47546972	-0.0169492	0.00957008	5.033579**	2.40702E-7	0.49528805	1	0.01132383	-4.74293**	1.05324E-6
36	6	6	40	91.04975	0.41681446	-0.025641	0.01418141	3.715437**	0.00010143	0.59748098	1	0.01670875	-3.11397**	0.00092294
48	12	4	30	116.229333	0.32802265	-0.0344828	0.0193953	2.602952**	0.00462124	0.69246243	1	0.02602947	-1.906187	0.02831297
60	5	12	24	151.749583	0.2995812	-0.0434783	0.03155302	3.05722**	0.001117	0.6011587	1	0.05124829	-1.7618156	0.03905023
120	15	8	12	303.499167	0.04081053	-0.0909091	0.0525103	0.60989562	0.27096548	1.00822431	1	0.07268142	0.03153088	0.48742308

**Sig. at 5% level

Hollota Site (Random Approximation)

Plot size (m^2)	Width (m^2)	Length (m^2)	N	X_{BAR}	Moran's I					Geary's C				
					I	E(I)	Variance	z value	P-value	C	E(C)	Variance	z value	P-value
4	2	2	360	10.1166111	0.21741567	-0.0027855	0.00263105	3.157337**	0.00079609	1.15525111	1	0.00283504	1.16982137	0.12103643
12	3	4	120	30.1829167	0.49066753	-0.0084034	0.00463601	7.329767**	1.1524E-13	0.48698223	1	0.00535551	-7.01023**	1.1897E-12
24	4	6	60	60.6998333	0.47546972	-0.0169492	0.00941333	5.075317**	1.93426E-7	0.49528805	1	0.01132383	-4.74293**	1.05324E-6
36	6	6	40	91.04975	0.41681446	-0.025641	0.01402917	3.735541**	0.00009366	0.59748098	1	0.01670875	-3.11397**	0.00092294
48	12	4	30	116.229333	0.32802265	-0.0344828	0.01927541	2.611035**	0.00451343	0.69246243	1	0.02602947	-1.906187	0.02831297
60	5	12	24	151.749583	0.2995812	-0.0434783	0.03085468	3.091625**	0.00099532	0.6011587	1	0.05124829	-1.7618156	0.03905023
120	15	8	12	303.499167	0.04081053	-0.0909091	0.05235026	0.60989562	0.27096548	1.00822431	1	0.07268142	0.03153088	0.48742308

**Sig. at 5% level

The Result in Tables 4.6 and 4.7 for different plot size in m^2 (the merged unit area), W in m is the plot width, L in m is the length, N is the number of basic units in the area, I is the calculated Moran coefficient, C is calculated Geary ratio, Z is the approximate Z (standard normal) score for I and C, P-value is the probability to reject the null hypothesis that there is no spatial autocorrelation based on Z-score, $E(I)$ and $E(C)$ are the expected value signifying zero spatial autocorrelation.

Results of Moran's index in Table 4.6 show that there is significant positive spatial autocorrelation between neighboring plot yield for plot size $4m^2$ through $60m^2$ considered. While the Geary's C coefficient show significant spatial autocorrelation between neighboring plot yield for a plot of size of $12m^2$ (3-by-4m) through $36m^2$ (6-by-6m). The variances among plot yield increases as plot size and number of plots aggregated in the area increase. This show that there is almost an over all high adjacent plot correlation and indicates the degree of spatial dependence in an observed potato yield throughout the experimental farm, arising from soil fertility pattern in the field trial.

Both the Moran's index and Geary's coefficient show, high significant positive autocorrelation between yields of plots for basic units. Thus, for potato yield in Hollota plot size of $12m^2$, (3-by-4m) through 60 m^2 (5-by-12m) there is a continuous soil fertility pattern in the farm and taking plot size of more than $60m^2$ will minimize or eliminate spatial autocorrelation between adjacent plots.

Table 4.7. Summary statistics of spatial autocorrelation, Moran's I and Geary's C, for Kulumsa Site with different size. Kulumsa Site (Normal Approximation)

Plot size (m^2)	Width (m^2)	Length (m^2)	N	X_{BAR}	Moran's I					Geary's C				
					I	E(I)	Variance	z value	P-value	C	E(C)	Variance	z value	P-value
4	2	2	191	25.7235602	0.01459711	-0.0052632	0.00284772	0.3602549	0.35932826	0.2850027	1	0.00321419	-12.6116**	0.35488459
9	3	3	144	27.2958333	0.36981611	-0.006993	0.00523571	4.926220**	4.19178E-7	0.58126448	1	0.00566291	3.55394**	0.00018976
12	4	3	72	63.5541667	0.4604842	-0.0140845	0.00775962	2.97525**	0.00016646	0.570202	1	0.00919623	-1.4312248	0.07617573
24	4	6	36	127.106667	0.658224	-0.0285714	0.00925608	5.367142**	3.99971E-8	0.40700989	1	0.01650192	-4.25384**	0.00001051
36	6	6	24	190.664167	0.72731499	-0.0434783	0.01261367	4.947658**	3.75557E-7	0.36224306	1	0.03210012	-3.65779**	0.0001272
48	8	6	18	254.213333	0.53574883	-0.0588235	0.03610215	0.77201813	0.022005184	0.54069092	1	0.05741627	0.48360939	0.31433155
108	12	9	8	577.2675	0.02420122	-0.1428571	0.05126527	0.64736479	0.25869794	0.81180342	1	0.07926125	-0.6453267	0.25935775

Kulumsa Site (Random Approximation)

Plot size (m^2)	Width (m^2)	Length (m^2)	N	X_{BAR}	Moran's I					Geary's C				
					I	E(I)	Variance	z value	P-value	C	E(C)	Variance	z value	P-value
4	2	2	191	25.7235602	0.01459711	-0.0052632	0.00235734	0.39595635	0.34606862	0.2850027	1	0.00321419	-12.6116**	0.35488459
9	3	3	144	27.2958333	0.36981611	-0.006993	0.00523135	4.926220**	4.19178E-7	0.58126448	1	0.00566019	3.55394**	0.00018976
12	4	3	72	63.5541667	0.4604842	-0.0140845	0.00533489	3.5882382**	0.00016646	0.570202	1	0.00919623	-1.4312248	0.07617573
24	4	6	36	127.106667	0.658224	-0.0285714	0.00926613	5.367142**	3.99971E-8	0.40700989	1	0.01650192	-4.25384**	0.00001051
36	6	6	24	190.664167	0.72731499	-0.0434783	0.01259836	4.947659**	3.75557E-7	0.36224306	1	0.03210012	-3.65779**	0.0001272
48	8	6	18	254.213333	0.53574883	-0.0588235	0.03386311	0.79713259	0.21268701	0.54069092	1	0.05741627	0.48360939	0.31433155
108	12	9	8	577.2675	0.02420122	-0.1428571	0.05150027	0.64736479	0.25869794	0.81180342	1	0.07922054	-0.6453267	0.25935775

**Sig. at 5% level

Results of Moran's index, I, and Geary's coefficient, C, in Table 4.7 for observed data from Kulumsa show there is significant positive spatial autocorrelation between neighboring plot yields for plot size of $4m^2$, $9m^2$, $24m^2$ and $36m^2$ considered. This indicates that there is high adjacent plot yield correlation throughout the experimental farm, due to among neighboring plots. Moran's I index show a high significant spatial autocorrelation for plots of size 9 m^2 (3-by-3m) basic units through 36 m^2 (6-by-6m) basic units. Thus, for potato yield trial in Kulumsa, plot size of $9m^2$ through $36m^2$ there is high potato yield variation in the field due variation in soil fertility pattern and taking plot size of more than $36m^2$ will minimize or eliminate spatial autocorrelation between adjacent plots.

From above discussion, at α = 0.05 level of significance, spatial autocorrelation measured by I and C statistics for all plot sizes using both agricultural research centers, clearly shows that there is high degree of spatial autocorrelation between yield of adjacent plots and there is a continuous fertility pattern over experimental farm. The result for both agricultural field trials on plot size of approximately 12 m^2 (4-by-3m) or (3-by-4m) through 36 m^2 (6-by-6m) , produce very less variability, and the yield gained from plots are spatially dependent. Hence, even increasing plot size will not generally make the observed spatial autocorrelation insignificant.

4.1.2.2. Comparison of Moran's I and Geary's C

The Geary's coefficient, C, differs from the Moran's index, I, because it uses the difference between the pairs, while Moran uses the difference between each point and the global average. The Moran's I is also used to detect deviation from a random spatial distribution, however Geary's C incorporates the variation of the population with in the area (plots). Thus Geary's coefficient is sensitive to the occurrence of data clusters with in the areas (plots).

Griffith (1987), notes that simulation experiments suggest that the inverse relationship between Moran's I and Geary's C is basically linear in nature. Departure from linearity are ascribed to differences in what each of the two induces measure, that is, Geary's C deals with paired comparisons and Moran's I with co-variations.

Now to observe the relationship between the two, consider the normality approximation for potato yield in table 4.6(a) and table 4.7(a) are summarized in Table 4.8 (a) and (b) below based on the number of plots and area in m^2.

Table 4.8. Relation between Moran's I and Geary's C

(a) Hollota

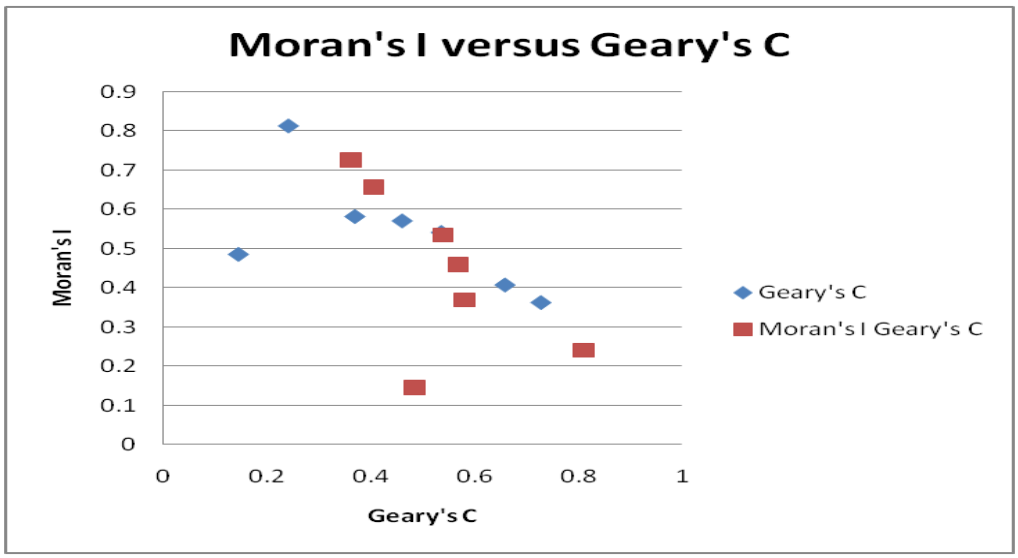
Number of plots	Area (m^2)	Moran's I	Geary's C
360	4	0.217416	1.155251
120	12	0.490668	0.486982
48	30	0.41687	0.495288
40	36	0.416814	0.597481
30	48	0.328023	0.692462
24	60	0.299581	0.601159
12	120	0.408105	1.008224

(b) Kulumsa

Number of plots	Area (m^2)	Moran's I	Geary's C
191	4	0.145971	0.485003
144	9	0.369826	0.581245
72	12	0.460484	0.570202
36	24	0.658224	0.40701
24	36	0.727315	0.362243
18	48	0.535749	0.540691
8	108	0.242012	0.811803

Figure 4.1. Relation between Moran's index I and Geary's C for statistic generated using rook's case neighboring relations

(a)



(b)

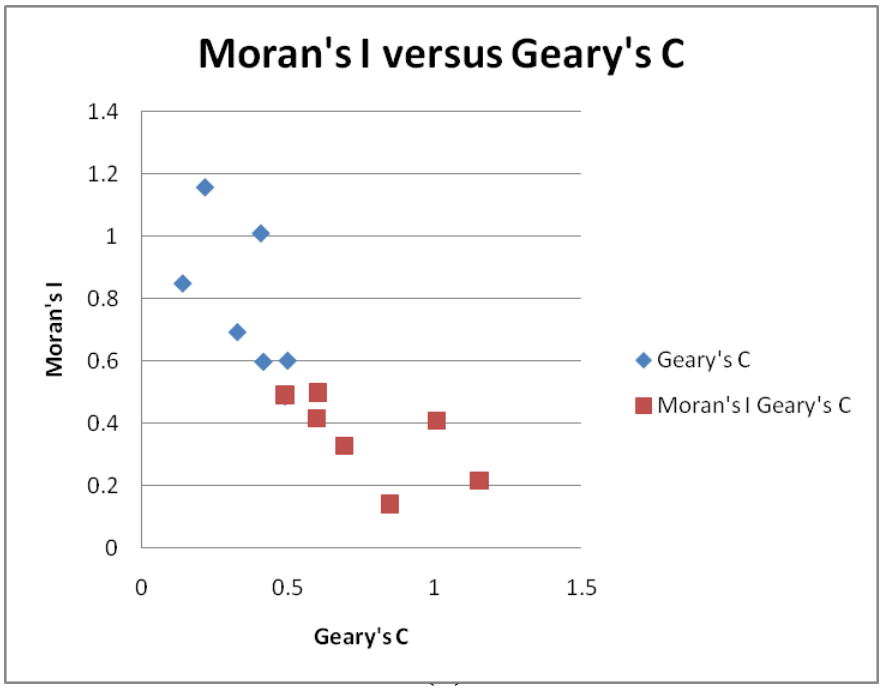


Figure 4.1 compares the relation between Moran's I and Geary's C and shows that the relation between Moran's I and Geary's C is linear and negative for the fact that either statistics will essentially computed for the same aspects, detecting existence of spatial autocorrelation among neighboring plot yield.

In summary, the spatial autocorrelation pattern observed is evident under both normality and random approximation(Tables 4.6 and 4.7). Therefore, the normal and random approximations are consistent and do not greatly differ for Moran's I or Geary's C test statistics.

The computation of Moran's index, I, and Geary's coefficient, C, to detect spatial autocorrelation among plot using Rook's case neighboring relation among plots (Figure 3.1). However, when spatial autocorrelation is used with a complex framework of data analysis, such as Queen's case or Bishop's case can be used in spatial autoregressive model fitting. In our case, for all plot sizes considered, Moran's I exceeds its corresponding expected value, E(I), under a null hypothesis of zero autocorrelation in both agricultural research locations. Hence, significant positive spatial autocorrelation is being detected among series of potato yield observations for each plot size between 12 m^2 and 36 m^2 under study. Its effect may inflate level of explained variance in the yield leading to incorrect inference. Thus, it is recommended to apply spatial statistics where such detection of spatial autocorrelation effect is observed in agronomic field trial. In the next section a model adjusted for spatial autocorrelation effect called autoregressive (AR) model fitting method and the comparison of the model with that do not include spatial effect statistical analysis method.

4.2. Autoregressive (AR) Model Fitting

Measuring and testing for spatial autocorrelation discussed in section 4.1.2.1 could reveal spatial pattern in the data. To compare the means of treatment groups in the field trials agronomists use the analysis of variance (ANOVA). To perform an ANOVA, the treatment groups are assumed to be randomly sampled, normally distributed and have equal variances. In practice, the method works well even with non-normal data and having unequal variance provided the groups are of equal size. However, the ANOVA is seriously affected by data that violate the randomization assumption, or are spatially autocorrelated. Agronomists are interested in methods of analysis of field experiments that take in to account spatial autocorrelation. Grodona and Cressie (1991) found that spatial analysis that approximates the error variance-covariance structure are more precise than the conventional analysis that assumes the structure to be independent.

When a spatial pattern is detected, we need to analyze pattern and provide an explanation based on the model that account for spatial autocorrelation. Here, spatial autocorrelation will be incorporated into a model of spatial variation for agronomic field trial, as well as the error term from a classical regression model that has the potato yield as its response variable.

The main focus in this section involves the method of spatial statistical technique known as autoregressive response (AR)-model proposed for agronomic field trial through its specification of the geographic configuration of areal units. The AR-model considered here accommodates plots that are located on the model of a regular lattice in agronomic field trial of potato yield. The method is computed using SAS (Appendix IV) on the dataset collected from Hollota and Kulumsa agricultural research centers in year 2001.

The goals of the method are: Fitting an autoregressive response of (AR) model and compare the basic difference of using AR-based ANOVA and Conventional ANOVA, with regard to results from 120-plots and 72-plots Varity trial from Hollota and Kulumsa agricultural research centers respectively.

The dataset considered in this section is adjusted for total yield in kg/12m^2 for both sites, where significant spatial autocorrelation detected. As it can be observed from section 4.3.1 of Table 4.6 and Table 4.7, the Moran's index for plot size of 12m^2 is 0.5 for Hollota and 0.46, for Kulumsa; exceeds their respective expected values 0.0084 and -0.014 under the hypothesis of zero autocorrelation. Hence positive autocorrelation is being detected among the observations, AR- based ANOVA is advocated here for such field trials, because through specification of aggregated plots of areal units it accommodates plots that are located on the nodes of a regular lattice. In addition, as discussed in methodology part of section 3.2.4, by partialing out of the error term the auto-correlated component of variance, AR-based ANOVA increases the precision of detecting real treatment differences.

4.2.1. Estimating the Jacobian term

The Moran's index and the Eigen values are computed using SAS in the AR-model for the geographic configuration matrix. The purpose of this matrix is to describe the spatial connectivity of the sample locations of the crop response variable represented as row-by-column grid plots. Eigenvalues computed from this matrix reflect the systematic behavior in the arrangement, interconnectivity, and interdependence of the potato yield observations. Furthermore, the eigenvalues facilitate computation of the Jacobian transformation from an auto-correlated to an Unautocorrelated mathematical space estimated using:

$$J^{\frac{1}{2}} = \sum_{i=1}^{n} \left(-\ln(1 - \rho\lambda_i) \right)$$

where, the λ_i's are the eigenvalues extracted from matrix W^{**}, n is number of observation, and $\ln$ is the natural logarithm. (section 3.3.1 and 3.3.2 for further detailed information of the method used).

As discussed in the methodology part of section 3.3, suppose the form of the autocorrelation model is specified as:

$$B(Y - \mu) = D\varepsilon,$$

where, Y is 20-by-1 vector that is spatially correlated potato yield under study. B is parameter matrices with $B = \{b_{ij}\}$ and $b_{ii} = 1$ for all i.

The data in Table 4.9 below contains potato yield data in kg/12m^2 collected from a total of 120 plots arranged in rectangular grid of 20 rows by 6 columns from Hollota.

Table 4.9. six columns by 20 rows rectangular grid of plots with potato yield observations in kg/12m^2 (Hollota)

row \ column	1	2	3	4	5	6
1	27.10	30.23	29.04	30.86	26.10	29.80
2	32.40	34.76	29.94	28.73	27.70	32.25
3	30.14	32.42	27.70	31.66	28.70	26.74
4	32.20	29.92	34.50	30.97	30.44	31.55
5	34.21	33.68	29.72	29.62	28.26	28.32
6	38.32	34.66	34.00	34.00	32.62	31.66
7	24.34	23.26	22.74	21.26	22.72	21.54
8	25.40	26.12	22.00	24.48	24.48	26.98
9	19.66	21.96	21.64	21.14	24.80	23.02
10	28.84	30.53	31.32	29.14	32.64	34.37
11	32.82	28.73	31.62	31.08	33.66	32.50
12	33.04	32.24	29.20	31.62	34.36	32.56
13	37.60	35.44	34.44	34.74	36.08	32.42
14	20.58	22.19	28.72	28.99	33.09	35.84
15	29.32	31.26	29.76	29.62	31.08	29.91
16	25.48	27.12	29.16	29.66	25.40	26.16
17	34.84	34.86	31.99	32.05	28.92	32.20
18	39.62	34.56	36.04	33.10	27.58	28.86
19	37.12	32.56	32.34	33.62	33.38	31.44
20	39.56	31.28	34.98	33.48	30.74	30.04

Table 4.10. Oder of observations in data field including column of 7 missing plots on the right side of lattice (Hollota).

column

row \ column	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	8	9	10	11	12	13	14
3	15	16	17	18	19	20	21
4	22	23	24	25	26	27	28
5	29	30	31	32	33	34	35
6	36	37	38	39	40	41	42
7	43	44	45	46	47	48	49
8	50	51	52	53	54	55	56
9	57	58	59	60	61	62	63
10	64	65	66	67	68	69	70
11	71	72	73	74	75	76	77
12	78	79	80	81	82	83	84
13	85	86	87	88	89	90	91
14	92	93	94	95	96	97	98
15	99	100	101	102	103	104	105
16	106	107	108	109	110	111	112
17	113	114	115	116	117	118	119
18	120	121	122	123	124	125	126
19	127	128	129	130	131	132	133
20	134	135	136	137	138	139	140

Table 4.10. represents how the plots are ordered in the data field. Accordingly, the lattice entries are read from the columns of data field in order of columns 1-6 by row 1-20. Each row of six plots is separated by (decimal) to denote the presence of missing values of variable (Appendix V). This spatial value is encountered in the data file once after each six records and have 20 of the 140 entries are missing value entries.

Figure 4.2. Relationship between geographic arrangement of 6-by-20 plot variety trial and 120-by-120 entry configuration matrix (with portion shown for plots 1-18 and 115-120)

120 Plot Varity trial

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72
73	74	75	76	77	78
79	80	81	82	83	84
85	86	87	88	89	90
91	92	93	94	95	96
97	98	99	100	101	102
103	104	105	106	107	108
109	110	111	112	113	114
115	116	117	118	119	120

120-by-120 configuration matrix W

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	___	115	116	117	118	119	120
1		1					1																		
2	1		1					1																	
3		1		1					1																
4			1		1					1															
5				1		1					1														
6					1							1													
7	1							1					1												
8		1					1		1					1											
9			1					1		1					1										
10				1					1		1					1									
11					1					1		1					1								
12						1					1							1							
13							1							1											
14								1					1		1										
15									1					1		1									
16										1					1		1								
17											1					1		1							
18												1					1								
115																					1				
116																				1		1			
117																					1		1		
118																						1		1	
119																							1		1
120																								1	

The plot numbering is in order of lattice (columns 1-6 by rows 1-20) contains the binary entries of matrix W. Entries equal to 1 for plots that share common boundaries (rook's case neighboring relation) and adjacent and 0 otherwise. Figure 4.2 show 120 plots of, for example, dataset arranged in 20 rows and 6 columns, and the associated (120)-by-(120) binary configuration (weigheted) matrix.

Here it is important to notice that log transformation of the data in Table 4.9. is used to overcome the outlying values observed.

After some iterations results show $D = I$ and $B = (I - S)$, where $S = \rho W^{**}$. Particularly, the model after certain operations on $(Y - \mu)$ is filter out its spatial autocorrelation. The resulting estimate is used to adjust the yield observations for the effect of autocorrelation. This transformation is done through matrix B, as estimated by $V = \sigma^2 \left[\left(I - \rho W^{**} \right)^T \left(I - \rho W^{**} \right) \right]^{-1}$, to compute a value for the mean squared error based squared difference between observed potato yield and predicted yield, Y_{adj}. Estimation of parameter β and the correlation matrix B are shown in Table 4.11 below for Hollota site.

Table 4.11. Non-Linear estimated parameters for autoregressive model of potato yield data from (Hollota site).

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
RHO	0.0374	0.00475	0.0280	0.0468
B0	2.3847	0.1238	2.1390	2.6305
B1	-0.00075	0.00306	-0.00682	0.00532
B2	-0.00546	0.00706	-0.0195	0.00855
B3	-0.1449	0.1918	-0.5255	0.2357
B4	0.1864	0.2025	-0.2155	0.5884
B5	-0.1020	0.2046	-0.5079	0.3039
B6	0.0633	0.2083	-0.3501	0.4766
B7	-0.0861	0.2210	-0.5247	0.3525
B8	-0.0263	0.1827	-0.3889	0.3363
B9	0.0421	0.1917	-0.3383	0.4224
B10	0.1064	0.2030	-0.2964	0.5091
B11	-0.1042	0.1942	-0.4896	0.2811
B12	0.0121	0.2061	-0.3969	0.4212
B13	0.0174	0.2110	-0.4013	0.4361
B14	0.0331	0.1896	-0.3430	0.4093
B15	-0.0180	0.1800	-0.3752	0.3392
B16	0.0397	0.1696	-0.2970	0.3763
B17	-0.0741	0.1645	-0.4005	0.2524
B18	0.0396	0.1658	-0.2894	0.3686
B19	0.0901	0.1673	-0.2420	0.4221
B20	-0.1747	0.1608	-0.4938	0.1444

The results in Table 4.11 are used for estimate the auto-regression model, which allows us to derive the autocorrelation parameter, ρ, and slope parameter, β, estimated simultaneously, in solving the problem of adjusting the data values to correct for the effect of autocorrelation on conventional ANOVA. Each slope term expresses a difference in mean potato yield between the row represented by this slope term and the row and column represented by the intercept. Table 4.11 indicates the autocorrelation parameter ρ is 0.0374 and slope parameter β_0 is 2.3847 which are different from zero (at $\alpha =$ 0.05). This indicates the potato yield data are spatially autocorrelated, is consistent with result of spatial autocorrelation discussed in section 4.2.

Approximate Correlation Matrix for B (Hollota)

	RHO	B0	B1	B2	B3	B4	B5	B6	B7	B8
RHO	1.0000000	-0.9337382	-0.5215995	-0.0294966	-0.0684534	0.0165486	-0.0782071	-0.0756898	-0.1596113	-0.1414246
B0	-0.9337382	1.0000000	0.2539858	-0.1818547	0.0059536	-0.0593684	0.0531487	0.0258071	0.0984267	0.1014153
B1	-0.5215995	0.2539858	1.0000000	0.0214432	0.1013490	0.0960707	0.1517459	0.1642503	0.1965301	0.1962707
B2	-0.0294966	-0.1818547	0.0214432	1.0000000	0.1681932	0.0287313	-0.0977439	-0.0073059	0.0398990	-0.0688868
B3	-0.0684534	0.0059536	0.1013490	0.1681932	1.0000000	0.3829102	-0.1541671	-0.0975645	-0.3501127	-0.6240266
B4	0.0165486	-0.0593684	0.0960707	0.0287313	0.3829102	1.0000000	0.2378393	-0.2502405	-0.1946154	-0.3922356
B5	-0.0782071	0.0531487	0.1517459	-0.0977439	-0.1541671	0.2378393	1.0000000	0.2575383	-0.1449168	0.0668576
B6	-0.0756898	0.0258071	0.1642503	-0.0073059	-0.0975645	-0.2502405	0.2575383	1.0000000	0.2924247	0.0749595
B7	-0.1596113	0.0984267	0.1965301	0.0398990	-0.3501127	-0.1946154	-0.1449168	0.2924247	1.0000000	0.4192201
B8	-0.1414246	0.1014153	0.1962707	-0.0688868	-0.6240266	-0.3922356	0.0668576	0.0749595	0.4192201	1.0000000
B9	-0.0757900	0.0287066	0.1126559	0.0719567	-0.3985269	-0.5772481	-0.2481721	0.1343979	0.1882533	0.4440338
B10	-0.0061745	-0.0042392	0.0333432	-0.0075266	-0.3446791	-0.4660295	-0.4782323	-0.1974483	0.2356725	0.2863271
B11	0.0077602	0.0138098	-0.0181192	-0.0729895	0.0241079	-0.2717399	-0.4450063	-0.4602783	-0.2090892	0.0426581
B12	0.0958523	-0.0665539	-0.0864358	-0.0403110	0.4277164	0.1876013	-0.3397409	-0.4811976	-0.5310236	-0.4393880
B13	0.0926634	-0.0685943	-0.0728778	-0.0366491	0.4200530	0.4668376	0.0657399	-0.3889198	-0.5343910	-0.5294607
B14	0.0131431	-0.0037625	-0.0033300	-0.0431606	0.1678845	0.3642439	0.4013920	-0.0236081	-0.4116350	-0.3494878
B15	-0.0712441	0.0078739	0.0814939	0.2020285	0.2771022	0.1226012	0.2548213	0.3201515	-0.0669056	-0.3076438
B16	-0.1360793	0.0459789	0.1938245	0.1675742	0.2063790	0.2571339	0.0626283	0.2291437	0.2509988	-0.0684353
B17	-0.1823840	0.1061551	0.2681719	-0.0038915	-0.1381530	0.1606213	0.3097925	0.0823286	0.2404323	0.2938454
B18	-0.0961961	0.0408191	0.2206972	-0.0753840	-0.2119936	-0.1686634	0.2379176	0.3431460	0.1282953	0.2897153
B19	-0.1231333	0.0599530	0.2377254	-0.0542400	-0.1555390	-0.2365217	-0.1288975	0.2155808	0.3651417	0.2171416
B20	-0.1913908	0.1432768	0.2429332	-0.0896463	-0.1055394	-0.1098786	-0.1470285	-0.0712044	0.2498665	0.2638057

Correlation matrix (Hollota continued)

	B9	B10	B11	B12	B13	B14	B15	B16	B17	B18	B19	B20
RHO	-0.0757900	-0.0061745	0.0077602	0.0958523	0.0926634	0.0131431	-0.0712441	-0.1360793	-0.1823840	-0.0961961	-0.1231333	-0.1913908
B0	0.0287066	-0.0042392	0.0138098	-0.0665539	-0.0685943	-0.0037625	0.0078739	0.0459789	0.1061551	0.0408191	0.0599530	0.1432768
B1	0.1126559	0.0333432	-0.0181192	-0.0864358	-0.0728778	-0.0033300	0.0814939	0.1938245	0.2681719	0.2206972	0.2377254	0.2429332
B2	0.0719567	-0.0075266	-0.0729895	-0.0403110	-0.0366491	-0.0431606	0.2020285	0.1675742	-0.0038915	-0.0753840	-0.0542400	-0.0896463
B3	-0.3985269	-0.3446791	0.0241079	0.4277164	0.4200530	0.1678845	0.2771022	0.2063790	-0.1381530	-0.2119936	-0.1555390	-0.1055394
B4	-0.5772481	-0.4660295	-0.2717399	0.1876013	0.4668376	0.3642439	0.1226012	0.2571339	0.1606213	-0.1686634	-0.2365217	-0.1098786
B5	-0.2481721	-0.4782323	-0.4450063	-0.3397409	0.0657399	0.4013920	0.2548213	0.0626283	0.3097925	0.2379176	-0.1288975	-0.1470285
B6	0.1343979	-0.1974483	-0.4602783	-0.4811976	-0.3889198	-0.0236081	0.3201515	0.2291437	0.0823286	0.3431460	0.2155808	-0.0712044
B7	0.1882533	0.2356725	-0.2090892	-0.5310236	-0.5343910	-0.4116350	-0.0669056	0.2509988	0.2404323	0.1282953	0.3651417	0.2498665
B8	0.4440338	0.2863271	0.0426581	-0.4393880	-0.5294607	-0.3494878	-0.3076438	-0.0684353	0.2938454	0.2897153	0.2171416	0.2638057
B9	1.0000000	0.4932471	0.1422269	-0.1632676	-0.4946981	-0.4673380	-0.3043156	-0.2705727	0.0167872	0.3124004	0.3337067	0.1274309
B10	0.4932471	1.0000000	0.4459086	-0.0208204	-0.2720177	-0.5041752	-0.4854303	-0.3500712	-0.2271290	0.0552354	0.3576088	0.3197084
B11	0.1422269	0.4459086	1.0000000	0.4132202	0.0307753	-0.1522055	-0.3972453	-0.4196176	-0.3457243	-0.2786760	0.0827968	0.2786010
B12	-0.1632676	-0.0208204	0.4132202	1.0000000	0.4908820	0.1027913	-0.0288048	-0.2346206	-0.3857403	-0.4077974	-0.2999135	-0.0094125
B13	-0.4946981	-0.2720177	0.0307753	0.4908820	1.0000000	0.4465044	0.0909921	0.0482022	-0.2189136	-0.3973970	-0.4446139	-0.2842125
B14	-0.4673380	-0.5041752	-0.1522055	0.1027913	0.4465044	1.0000000	0.3266292	0.0203214	0.0591096	-0.1957820	-0.4619398	-0.3506028
B15	0.3043156	-0.4854303	-0.3972453	-0.0288048	0.0909921	0.3266292	1.0000000	0.3102705	-0.0292041	0.0472814	-0.2718565	-0.3754705
B16	-0.2705727	-0.3500712	-0.4196176	-0.2346206	0.0482022	0.0203214	0.3102705	1.0000000	0.2106171	-0.0591499	-0.0007175	-0.1910465
B17	0.0167872	-0.2271290	-0.3457243	-0.3857403	-0.2189136	0.0591096	-0.0292041	0.2106171	1.0000000	0.2311200	-0.0431184	0.0650454
B18	0.3124004	0.0552354	-0.2786760	-0.4077974	-0.3973970	-0.1957820	0.0472814	-0.0591499	0.2311200	1.0000000	0.2510374	-0.0596233
B19	0.3337067	0.3576088	0.0827968	-0.2999135	-0.4446139	0.4619398	-0.2718565	-0.0007175	-0.0431184	0.2510374	1.0000000	0.3036757
B20	0.1274309	0.3197084	0.2786010	-0.0094125	-0.2842125	-0.3506028	-0.3754705	-0.1910465	0.0650454	-0.0596233	0.3036757	1.0000000

The dataset, described in Table 4.12 below contains values of average potato yield in kg/12m^2 collected from a total of 72 plots arranged in rectangular grid with 12 rows by 6 columns from Kulumsa.

Table 4.12. Six columns by 12 rows rectangular grid of plots with totals potato yield observations in kg/12m^2 (Kulumsa)

row \ column	1	2	3	4	5	6
1	70.82	59.9	65.5	71.9	67.7	65
2	72.16	67.6	70.5	65.8	66.1	66.7
3	65.14	63	67.4	69.1	63.2	65.2
4	63.48	63.6	66.3	64.2	62.8	66
5	65.82	69.4	67.3	65.9	66.1	61.3
6	62.06	51	63.3	55.4	54.5	56.7
7	56.64	58.7	57.8	56.2	57.6	59.5
8	58.48	58.2	56.6	60.1	65	57.1
9	62.38	63.6	63.6	64	59.9	62.4
10	68.35	66.5	63.7	69.7	63.3	65.4
11	63.38	64.3	64.7	69.8	61.5	61.5
12	64.83	64.8	64.2	64	69.1	61.1

Table 4.13. Oder of observations in data field including column of 7 missing plots on the right side of lattice (Kulmsa).

row \ column	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	8	9	10	11	12	13	14
3	15	16	17	18	19	20	21
4	22	23	24	25	26	27	28
5	29	30	31	32	33	34	35
6	36	37	38	39	40	41	42
7	43	44	45	46	47	48	49
8	50	51	52	53	54	55	56
9	57	58	59	60	61	62	63
10	64	65	66	67	68	69	70
11	71	72	73	74	75	76	77
12	78	79	80	81	82	83	84

Table 4.13 represents how the plots are ordered in the potato yield data from Kulumsa. Accordingly, the lattice entries are read from the columns of data field in order of columns 1-6 by row 1-12. Each row of six plots is separated by (decimal) to denote the presence of missing values of variable (Appendix VI). This spatial value is encountered in the data file once after each six records and have 12 of the 72 entries are missing value.

Table 4.14. Non-Linear estimated parameters for autoregressive model for potato yield data from Kulumsa site.

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
RHO	0.0114	0.00283	0.00576	0.0171
B0	3.4536	0.1671	3.1192	3.7880
B1	0.00193	0.00297	-0.00401	0.00788
B2	-0.00330	0.00449	-0.0123	0.00570
B3	0.0322	0.0903	-0.1485	0.2130
B4	0.1265	0.0923	-0.0581	0.3112
B5	-0.0413	0.0983	-0.2381	0.1556
B6	-0.1351	0.1008	-0.3368	0.0667
B7	0.1008	0.0966	-0.0926	0.2942
B8	0.1686	0.0964	-0.0243	0.3615
B9	0.0601	0.0926	-0.1254	0.2455
B10	-0.1136	0.0953	-0.3043	0.0771
B11	0.0712	0.0934	-0.1157	0.2581
B12	0.1680	0.0938	-0.0197	0.3558

In case of Kulumsa, based on result show in Table 4.14, ρ and β_0 are 0.0114 and 3.4536, respectively, which are significantly different from zero at $\alpha =$ 0.05 level of significance. This indicates a moderate strong degree of positive spatial autocorrelation is observed among potato yield data from neighboring plots, which confirms conclusions made on autocorrelation test conducted in section 4.2.1. Therefore, when spatial autocorrelation is considered, on original explanatory variables, plus the sample mean, provides significant influence on the realization of the response variable in an area.

Approximate Correlation Matrix for B (Kulumsa)

	RHO	B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
RHO	1.0000000	-0.9834600	-0.1729417	-0.0595804	-0.0351549	0.0323437	-0.0547983	-0.1000306	-0.2277171	-0.2151587	-0.2014215	-0.3265184	-0.1850357	-0.1963344
B0	-0.9834600	1.0000000	0.0267909	-0.0482114	0.0289850	-0.0478409	0.0445132	0.1132024	0.1903604	0.1481254	0.1336201	0.2687955	0.1353086	0.1417880
B1	-0.1729417	0.0267909	1.0000000	0.0811065	0.0557489	0.0669576	0.0624895	0.0486951	0.3278353	0.3484934	0.3049849	0.3216448	0.2912858	0.2928747
B2	-0.0595804	-0.0482114	0.0811065	1.0000000	-0.0077258	0.0580596	0.0218728	-0.1920014	-0.1268939	0.1460058	0.2146649	0.1038089	0.0436401	0.0934610
B3	-0.0351549	0.0289850	0.0557489	-0.0077258	1.0000000	0.0018358	-0.2568588	0.0078402	0.1566249	0.0206935	-0.6762299	0.0228493	-0.1193712	0.0179604
B4	0.0323437	-0.0478409	0.0669576	0.0580596	0.0018358	1.0000000	0.0030055	-0.2709894	0.0052965	0.1578894	0.0237830	-0.6265808	0.0141621	-0.1227109
B5	-0.0547983	0.0445132	0.0624895	0.0218728	-0.2568588	0.0030055	1.0000000	0.0042198	-0.4795688	0.0307872	0.2107572	0.0336978	-0.5888015	0.0261958
B6	-0.1000306	0.1132024	0.0486951	-0.1920014	0.0078402	-0.2709894	0.0042198	1.0000000	0.0647422	-0.4867548	-0.0080665	0.1953486	0.0241348	-0.5792794
B7	-0.2277171	0.1903604	0.3278353	-0.1268939	0.1566249	0.0052965	-0.4795688	0.0647422	1.0000000	0.1433922	0.0237423	0.1626757	0.4499339	0.1318361
B8	-0.2151587	0.1481254	0.3484934	0.1460058	0.0206935	0.1578894	0.0307872	-0.4867548	0.1433922	1.0000000	0.1705532	0.0954237	0.1454294	0.4638950
B9	-0.2014215	0.1336201	0.3049849	0.2146649	-0.6762299	0.0237830	0.2107572	-0.0080665	0.0237423	0.1705532	1.0000000	0.1719827	0.2287691	0.1439475
B10	-0.3265184	0.2687955	0.3216448	0.1038089	0.0228493	-0.6265808	0.0336978	0.1953486	0.1626757	0.0954237	0.1719827	1.0000000	0.1533223	0.2497146
B11	-0.1850357	0.1353086	0.2912858	0.0436401	-0.1193712	0.0141621	-0.5888015	0.0241348	0.4499339	0.1454294	0.2287691	0.1533223	1.0000000	0.1278668
B12	-0.1963344	0.1417880	0.2928747	0.0934610	0.0179604	-0.1227109	0.0261958	-0.5792794	0.1318361	0.4638950	0.1439475	0.2497146	0.1278668	1.0000000

4.2.2. Comparison of AR-based ANOVA and Conventional ANOVA

The information we get from the above estimation is used to adjust the data values to correct for the effect of autocorrelation on conventional ANOVA. The conventional ANOVA method is based on observed yield that is not adjusted for presence of autocorrelation. The AR method is based on observed yield that is adjusted for autocorrelation. As discussed in methodology part fitting spatial AR-response model using the adjusted yield, Y_{adj}, that is given by

$$Y_{adj} = Y - \left(\rho W^{**} Y - \rho\beta_0\right).$$

To compare AR-based and conventional ANOVA, consider potato yield data in Table 4.12 for Hollota site. with list values for the coefficient of determination (R^2), error sum of squares, mean square error and F values for the error, and total source of variation that is computed from correlated plots for potato yield dataset. The correct conditional sum of squares for the autocorrelation parameter is the difference between the total sum squares gained from OLS method and that gained from AR model.

The classification of variables in row and column have been included for ease of performing ANOVA. These variables will be used once the yield observations have been adjusted for autocorrelations by means of AR-modeling. The binary indicator variables allow the AR- method to cast the ANOVA as a regression of the potato yield. Accordingly, 5 indicator variables define all the columns except one and 19 indicator variables define all rows except one. The resulting intercept term β_0 expresses the combined mean of the excluded row and column. Each statistical form of the AR-model statement used in SAS code is:

$$Y\exp(J^{1/2}) = \left(X_k\beta_k + \rho W^{**} Y - \rho\beta_0\right)\exp(J^{1/2}) + \varepsilon\exp((J^{1/2}),$$

here, ρ is the autocorrelation parameter, and X_k specifies binary indicator variables for k different classes of row and columns, β_0 expresses mean potato yield in kg/m^2 for the kth class and X_k takes a value of 1 for observations belonging to the kth class and 0 otherwise. The potato yield response vector $W^{**}Y$ the weighted average of neighboring plots.

Table 4.15. ANOVA table for comparing the AR-based and conventional based model analysis result for Hollota Site

(a) Conventional based ANOVA (Hollota)

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	R^2
Row	19	1.13391470	0.05967972	3.3487865**	0.412356
Column	5	0.04159051	0.00831810	0.4667505	
Error	95	1.67520194	0.017821297		
Total	119	2.85070715			

(b) AR-based ANOVA (Hollota)

Dependent Variable: Y_{adj}

Source	DF	Sum of Squares	Mean Square	F Value	**R^2**
ρ	1	0.18314438	0.18314438	23.7357866**	0.728104
Row	19	1.92508880	0.02057202	2.6661648**	
Column	5	0.01717340	0.00309257	0.4008017	
Error	94	0.72530057	0.00771596		
Total	119	2.85070715			

** significance at 5% level of probability

In case of Hollota results in Tables 4.15 (a) and (b) shows there is significant difference for rows effect in both conventional ANOVA and AR-based ANOVA; while column effect is non-significant. The AR-based ANOVA shows that the difference for row effect is more insignificant than one would find for unauto-correlated data, while the difference for column effect is less insignificant. The amount of explained variation (R^2) in grain yield increases from about 41% for OLS –based ANOVA to about 71% for AR-based ANOVA. This is accompanied by a reduction in the MSE from 0.017821297 of OLS-based ANOVA to 0.00771596 for AR-based ANOVA. This difference represents a 10% reduction in experimental error, or improvement in precision, for AR-based ANOVA over OLS-based ANOVA.

The autocorrelation parameter, ρ, accounts for 6% of the variation in potato grain yield. Meanwhile, the sum of squares for rows is 0.020572021 rather than 1.13391470. The reduction in sum of squares for rows in AR-versus OLS-based ANOVA corresponds to a decrease in explained yield variation from 67% to 40% relative to the total sum of squares.

Similarly, estimated values of AR-model estimation for Kulumsa site using log transformed total potato yield of Table 4.13 with plot size of 12 m^2 (4-by-3m) arranged in 12 rows by 6 columns with total of 72 plots is presented in Table 4.16.

Table 4.16. ANOVA table for comparing the AR-based and conventional based model analysis result for Kulumsa Site

(a). Conventional ANOVA (Kulumsa)

Dependent variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	$\mathbf{R^2}$
Row	11	1.52144495	0.13831318	8.66**	0.655377
Column	5	0.14822649	0.02964530	1.86	
Error	55	0.87797739	0.01596323		
Total	71	2.54764883			

(b). AR-based ANOVA (Kulumsa)

Dependent Variable: Y_{adj}

Source	DF	Sum of Squares	Mean Square	F Value	$\mathbf{R^2}$
ρ	1	2.1974768	2.1974768	1114.780085**	0.690390
Row	11	0.22629225	0.02057202	10.43618672**	
Column	5	0.01546285	0.00309257	1.56886091	
Error	54	0.10841693	0.00197122		
Total	71	2.54764883			

** significance at 5% level of probability

Similarly, in case of Kulumsa, results in Tables 4.16 (a) and (b) shows there is significant difference for rows effect in both conventional ANOVA and AR-based ANOVA . And column effect is non-significant for both conventional ANOVA and AR-based ANOVA. The AR-based ANOVA shows that the difference for row effect is more insignificant than one would find for unauto-correlated data, while the difference for column effect is less insignificant. The amount of explained variation

(R^2) in grain yield increases from about 66% for OLS –based ANOVA to about 69% for AR-based ANOVA. This is accompanied by a reduction in the MSE from 0.01596323 OLS-based ANOVA to 0.00197122 for AR-based ANOVA. This difference represents a 10.4% reduction in experimental error, or improvement in precision, for AR-based ANOVA over OLS-based ANOVA.
The autocorrelation parameter, ρ, accounts for 86% of the variation in potato grain yield. Meanwhile, the sum of squares for rows is 0.22629225 rather than1.52144495. The reduction in sum of squares for rows in AR-versus OLS-based ANOVA corresponds to a decrease in explained yield variation from 60% to 9% relative to the total sum of squares.

These results from soil variability trial suggest that randomization does not completely provide for independence for observations in data that are spatially auto-correlated. Un assumed amount of autocorrelation can inflate level of explained variance in the potato yield, thereby leading to incorrect inference concerning the significance of the difference in row effect in the potato yield data. The outcome of AR-based ANOVA differs from its conventional counterpart because extreme variance in the yield due to autocorrelation has been affected out by $\rho W^{*}Y$ in equation (3.13). Consequently, the R^2 value is larger and the sum of squares for error and explanatory variables is smaller in AR-based ANOVA versus conventional ANOVA.

Furthermore, the decrease in the mean–square error has increased capacity for detecting real treatment differences. That is, spatially dependent data and error render significance test unreliable because variances are underestimated in such situations. Spatial auto-regression model are appropriate alternatives to traditional regression for spatial data because they eliminate the dependencies that cause

problems, and are sensitive to the specification of the weighted matrix, as it is demonstrated in the above application on potato yield data.

In conventional ANOVA, positive spatial autocorrelation tends to inflate the sum of squares for error and deflates those for row relative to the total sum of squares. The AR method gives improved performance because autocorrelation can be partialed out in a regression equation. The method is advised to be adoptable to agronomic field data because the result of regression are identical to an ANOVA of a continuous response variable on any number of classificatory indicator variables; ANOVA with regression is essentially the same approach used in generalized linear models. Therefore, many statistical procedures can be implemented with the method.

4.3. Variogram Prediction and Kriging

Kriging is a technique used in the analysis of spatial data. Once the spatial correlation structure of a variable has been identified, the data from the measured locations can be used to estimate the variable at locations where it had not been measured. This extrapolation from measured locations to unmeasured locations is called kriging. The assumption we made here is that second-order stationary when we fit a variogram model as discussed in the methodology part of section 3.4.3.2.

A common test for the existence of spatial structure in the lattice data is Moran's I, that essentially calculates a weighted correlation between each observation and all its defined neighbor, or observations. Applying this test to our sample of potato yield data of 12x6 lattice from Kulumsa agricultural research center using the rook's weighted relation matrix definition of neighbors, which treats the regions sharing the same boundary equally weighted neighbors. The result based on the hypothesis no spatial correlation is rejected. This is not surprising since we generated the data

from a dependent spatially correlated process. A two dimensional scatter plot results of measurement locations based on original potato yield data (Appendix III) are presented as shown in Figure 4.3 below for Kulumsa agricultural research centers.

Figure 4.3. Two dimensional scatter plot of measurement locations for potato yield data.

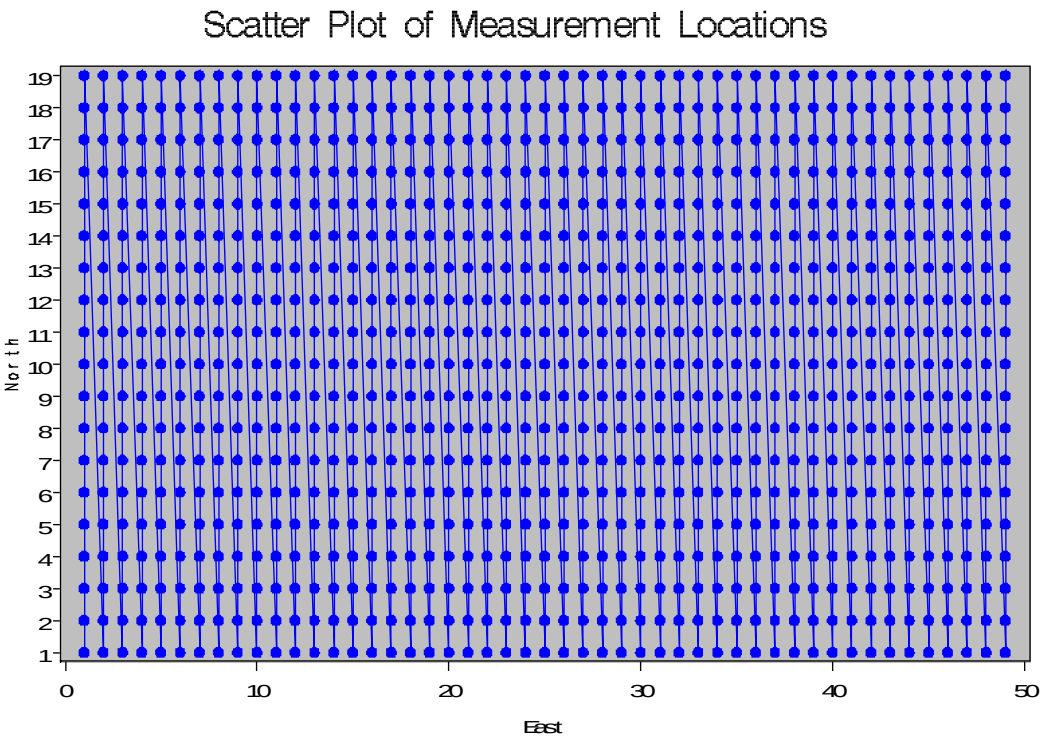


Figure 4.3 indicates, the locations are ideally spread (regular lattice) around the prediction area, i.e., it shows the centroid regular location pattern where the given data points are generated.

Now it is possible look at a surface (three-dimensional) plot of the measured variable, the potato yield data, with their respective location, from result shown in Figure 4.4 below. This is a crucial step for removing any obvious surface trend before computing and estimating the model of spatial dependence (the semivariogram model).

Figure 4.4. Three dimensional scatter plot of measurement locations for potato yield data. The vertical lines are sited at the spatial locations and their heights are equal to the corresponding potato yield (in kg/m^2).

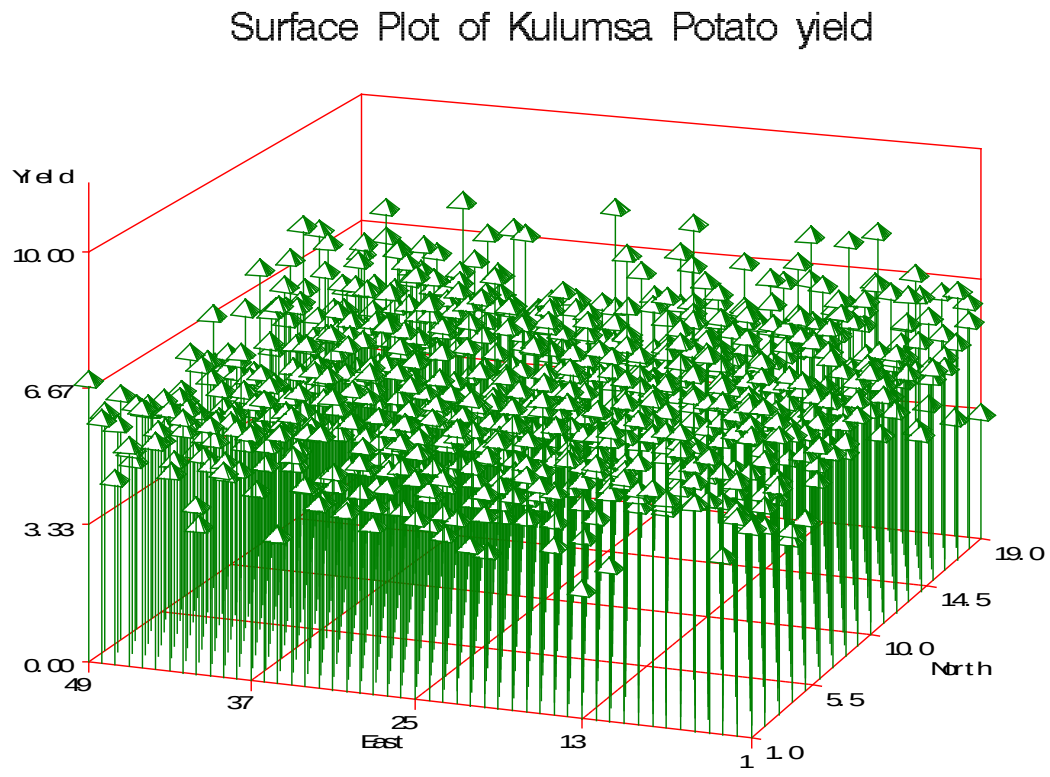


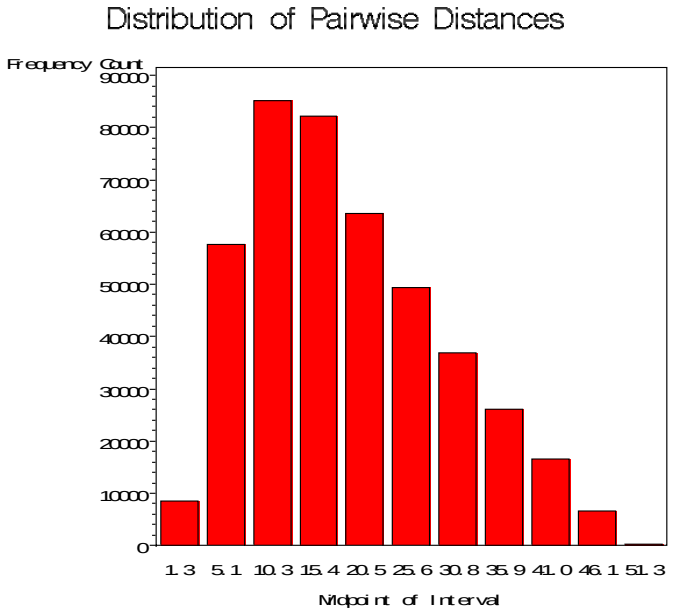
Figure 4.4 shows there are outlying values in the dataset from Kulumsa. But the large-scale variation, typical of spatial data, there appear to be surface trend from north to south and east to west directions as observed from Figures 4.4. Hence, it is better to work with the residuals yield data rather than original form of a trend surface fit.

The variogram is used to produce a modified histogram of the pairwise distances in order to find reasonable values for the lag distance and maximum lags. In the following analysis, a potato field may blocked off on a regular grid, and yield within each block is taken as the spatial lattice data. Potato yield data are considered from 72 plots each $12m^2$ (4m east-by-3m north) from Kulumsa site (Table 4.14) for regular lattice. The main interest is how much the estimation of the variogram changes if spatial location of each data point is randomly chosen with in their respective plots.

Table 4.17. Predicted values for large lag distance

Obs	LAG	LB	UB	COUNT	PER
1	0	0.0000	2.5632	8572	0.01980
2	1	2.5632	7.6896	57538	0.13291
3	2	7.6896	12.8160	85102	0.19658
4	3	12.8160	17.9424	82266	0.19003
5	4	17.9424	23.0688	63630	0.14698
6	5	23.0688	28.1952	49451	0.11423
7	6	28.1952	33.3216	36890	0.08521
8	7	33.3216	38.4480	26157	0.06042
9	8	38.4480	43.5744	16532	0.03819
10	9	43.5744	48.7008	6547	0.01512
11	10	48.7008	53.8272	230	0.00053

Figure 4.5. Histogram for pairwise distance among plots with large lag distance



For plotting and estimations purposes of potato yield, it is desirable to have as many points as possible for the plot of the varigram $\gamma_z(h)$ against the lag distance h. This corresponds to having as many distance intervals as possible, that is, having a small value for the lag distance of 7. However, a rule of thumb used in computing sample semivariograms is to use at least 30 point pairs in computing a single value of the empirical or experimental semivariogram.

If the lag distance is too small, there may be too few points in one or more of the intervals. On the other hand, if the lag distance is too large value, the number of point pairs in the distance intervals may be much greater than that needed for estimation precision, thereby "wasting" point pairs at the expense of variogram points. Thus, for lag distance of 1 to 7 the variability of the estimated variogram is

relatively big at small lags and the chance of getting a large average of $\frac{[(Z(s_i) - Z(s_j))]^2}{2}$ increases. Similarly this can occur at lag values ≥ 7.

In order to calculate the variogram, we will need to specify a lag distance and a maximum number of lags. The lag distance is the size into which the pairwise distances are grouped. Then for each distance, the variance of the pairwise differences in the measured variable (in this study, average potato yield) is calculated.

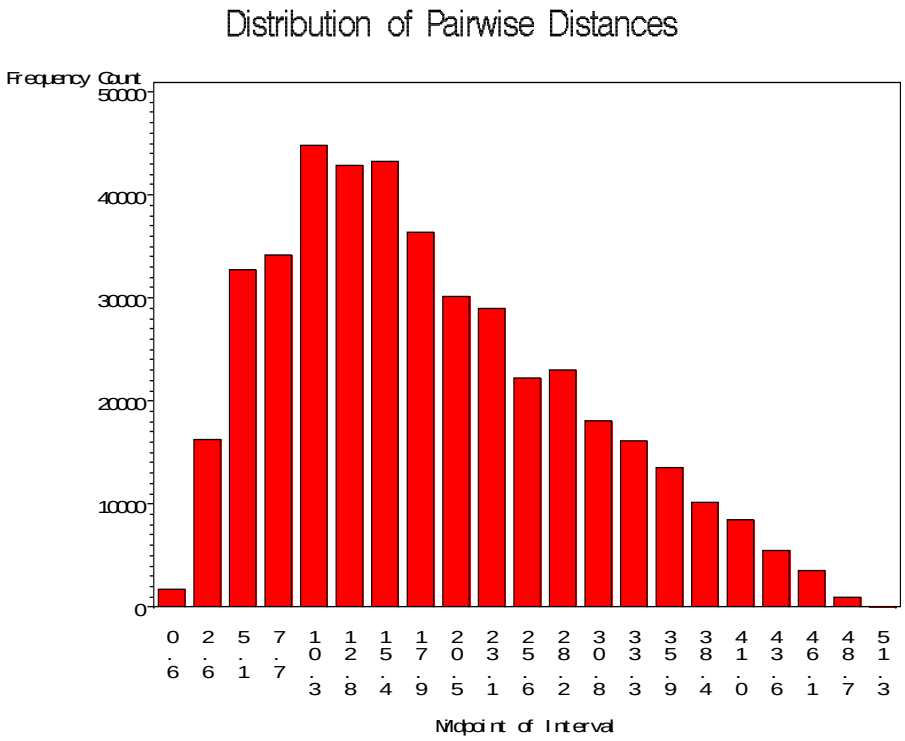
Table 4.17 and Figure 4.5, show the first few distance intervals for maximum lag distance of 10, corresponding to lag 0 and lag 1, are typically the limiting intervals. This is particularly true for lag 0 since it is half the width of the remaining intervals. For the default of the lag 0 class contains 8572 points, which is reasonably, but the lag 1 class contains 57538 points. If maximum lag distance of 20 is used instead, these numbers become 1794 and 16294 for lags 0 and 1, respectively. As the result in Table 4.18 and Figure 4.6 shown below. Because of the asymmetrical nature of lag 0, it is advised to violate the rule of thumb for the 0th lag. However, here we have sufficient numbers in lag 1 and above.

Table 4.18. Predicted values for small lag distance

Obs	LAG	LB	UB	COUNT	PER
1	0	0.0000	1.2816	1794	0.00414
2	1	1.2816	3.8448	16294	0.03764
3	2	3.8448	6.4080	32704	0.07554
4	3	6.4080	8.9712	34178	0.07895
5	4	8.9712	11.5344	44828	0.10355
6	5	11.5344	14.0976	42830	0.09893
7	6	14.0976	16.6608	43250	0.09990
8	7	16.6608	19.2240	36372	0.08402
9	8	19.2240	21.7872	30149	0.06964
10	9	21.7872	24.3504	28934	0.06684
11	10	24.3504	26.9136	22179	0.05123
12	11	26.9136	29.4768	23023	0.05318
13	12	29.4768	32.0400	18142	0.04191
14	13	32.0400	34.6032	16077	0.03714
15	14	34.6032	37.1664	13485	0.03115
16	15	37.1664	39.7296	10177	0.02351
17	16	39.7296	42.2928	8542	0.01973
18	17	42.2928	44.8560	5461	0.01261
19	18	44.8560	47.4192	3565	0.00823
20	19	47.4192	49.9824	897	0.00207
21	20	49.9824	52.5456	34	0.00008

The problem with using the maximum lag value is that it includes pairs of points so far apart that they are likely to be independent. Using pairs of points that are independent adds nothing to the empirical semivariogram plot; they are essentially added noise.

Figure 4.6. Histogram for pairwise distance among plots with small lag distance



4.3.1. Variability in the Emperical Variogram

Given observations $Z(s_1), \ldots, Z(s_n)$ from spatial lattice process $Z(S)$, where each potato yield observation represents the value associated with each plot S_i in the lattice, is the variogram given as $\gamma(h) = \frac{1}{2}\text{var}[Z(s+h) - Z(s)]$ where h is lag distance between each plot s_i, s_j and the variogram estimated as (section 3.4.2)

$$\hat{\gamma}(h) = \frac{1}{2}\frac{1}{|N(h)|}\sum_{N(h)}\{Z(s_i) - Z(s_j)\}^2$$

where, $N(h) = \{(s_i, s_j) : s_i - s_j = h, \quad i, j = 1, \ldots, n\}$. The robust estimation given by the equation

$$\bar{\gamma}(h) = \frac{1}{2\left(0.457 + \frac{0.494}{|N(h)|}\right)} \left\{ \frac{1}{|N(h)|} \sum_{N(h)} |Z(s_i) - Z(s_j)|^{\frac{1}{2}} \right\}^4$$

down-weightes the outliers where as in classical variogram

$$\hat{\gamma}(h) = \frac{1}{2} \frac{1}{|N(h)|} \sum_{N(h)} \{Z(s_i) - Z(s_j)\}^2$$

the squared term aggravate the outliers.

Table 4.19. The Variogram estimation result for both classical and robust estimations

Obs	Lag Class Value (in LAGDIST= units)	Number of Pairs in Class	Average Lag Distance for Class	Variogram	Robust Variogram
1	-1	931	.	.	.
2	0	15020	2.3182	0.82760	0.8033
3	1	97158	7.3518	0.85834	0.82769
4	2	117368	13.9981	0.92400	0.89349
5	3	82649	20.8269	0.98053	0.95735
6	4	59063	27.8086	0.91391	0.89524
7	5	38748	34.7626	0.84806	0.83430
8	6	19705	41.5061	0.87148	0.86308
9	7	3204	48.9557	0.90746	0.88227

Both classical and robust variogram estimator were computed for the potato yield in Table 4.19 and the fitness is shown in Figure 4.7. It gives variogram plot for each lag distance up to 7 where large spatial outlier is observed in the data and spatial dependence changes with increasing distance from a point. That is, points separated by short distances are correlated and the correlation decreases as the distance between points increases. This can be seen in the increase in the variogram up to a lag distance of about 7. If our outcome variable is free of spatial correlation, we will expect the variogram to appear relatively constant across all distances.

Figure 4.7. Comparison of Estimated variogram classical (*) and robust(◇) estimators of potato yield.

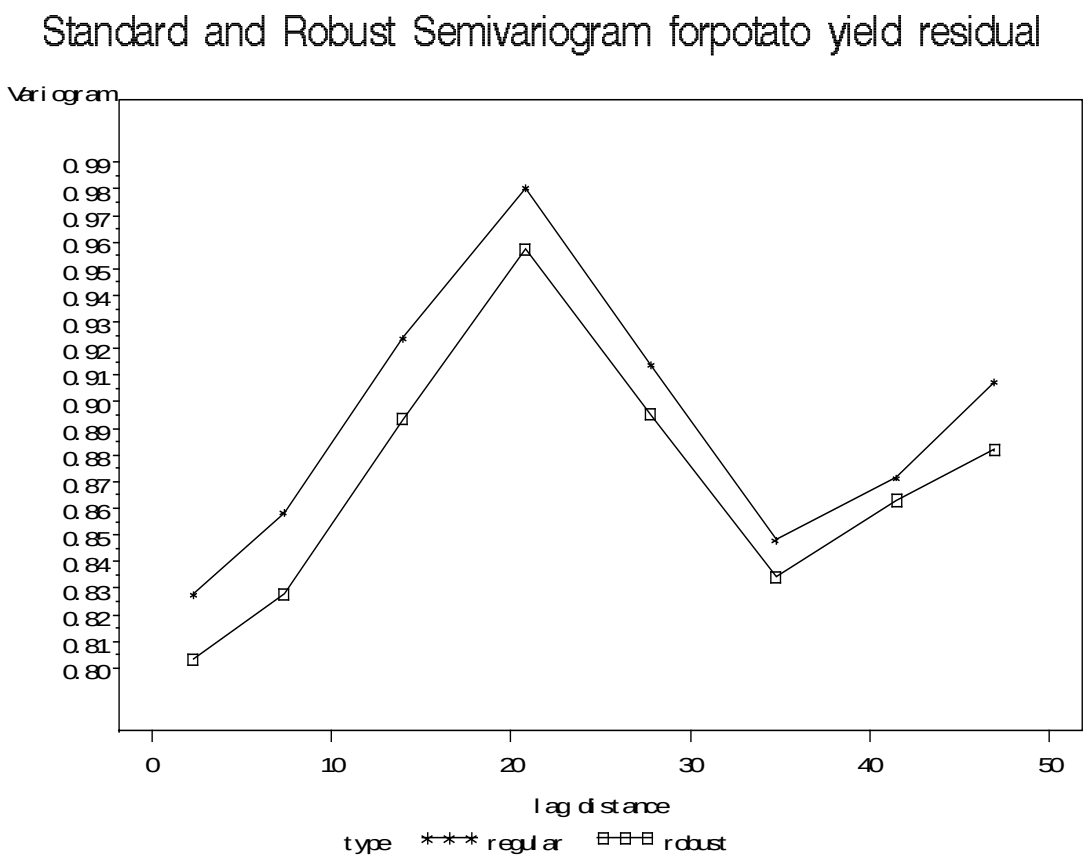


Table 4.19 also shows the sampling variability of the classical empirical variogram at distinct lags. The estimated variogram shapes in Figure 4.7 are very similar. That is, there is no obvious difference between the classical and robust variogram with regard to the empirical variability due to the random location of the potato yield data.

4.3.2. Variability in Estimating Theorotical Variogram

Based on the section of the methodology chapter on the variogram model, a particular variogram is chosen for the potato yield data. The chosen variogram is Gaussian with a scale (sill) of c0 = 7.5, and a range of a0 = 30. This choice of the variogram is based on a visual fit a comparison of the plots of the regular, robust sample variograms and the Gaussian variogram for various scale (sill) and range values.

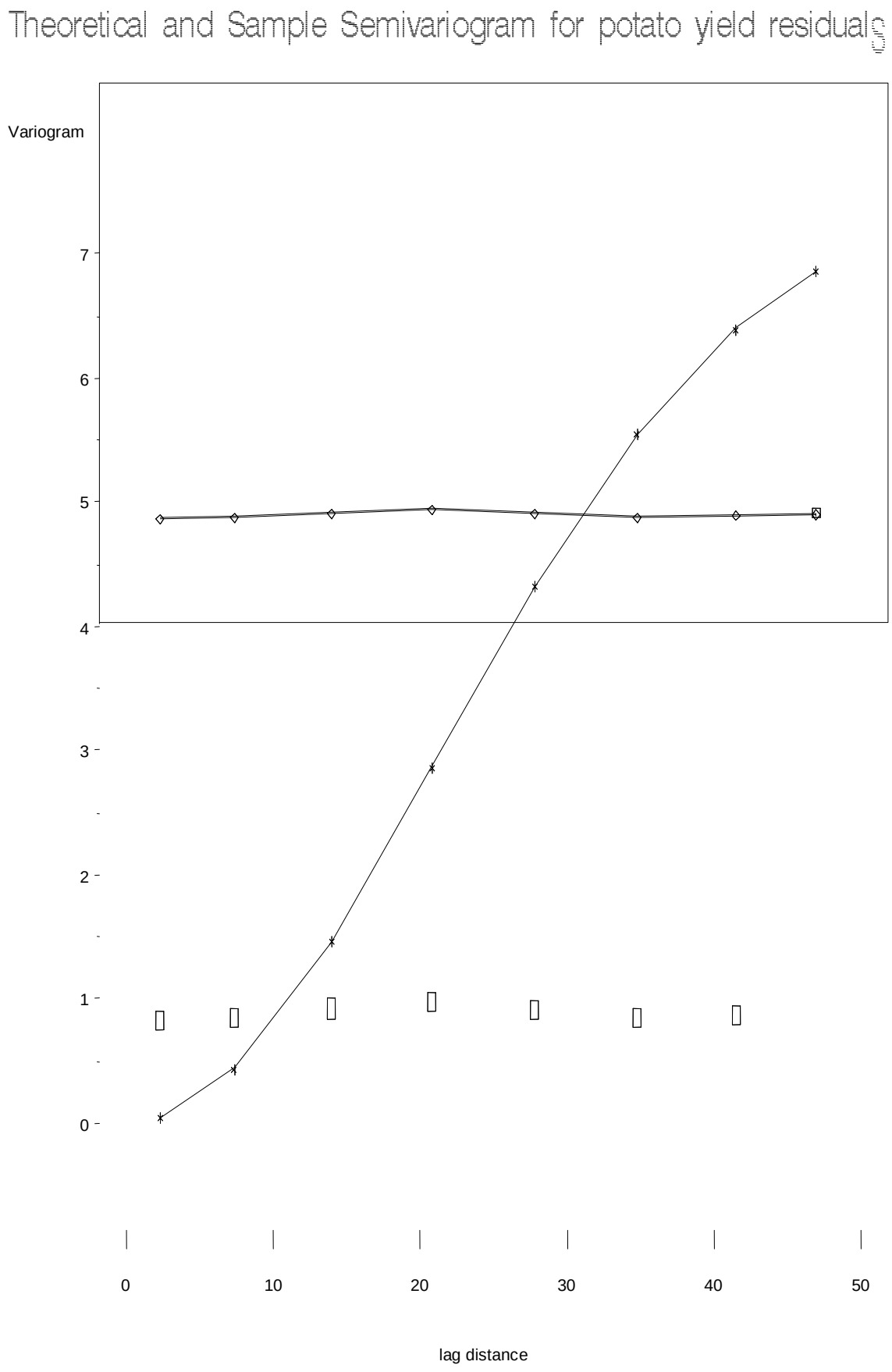
Another possible choice of model is the spherical variogram with the same scale (sill) of c0 = 7.5 but with a range of a0 = 60. This choice of range is again based on a visual fit; while not as good as the Gaussian model, the fit is reasonable. It is generally held that spatial prediction is robust against model specification, while the standard error computation is not so robust.

The theoretical exponential variogram model fitted only for the lag distances less than seven due to increased variability at larger lags. Figure 4.8 shows a plot of the Gaussian, classical & exponential variograms versus lag distance $d(h)$ fitted generated for 12x6 m^2 at different random locations.

$$\gamma(h;\theta)=\begin{cases}0\ , & \text{if } h=0\\ 7.5\left\{1-\exp\left(-\|h\|/30\right)^2\right\}, & \text{if } h\neq 0\end{cases}$$

$$\theta=(7.5,\ 30)$$

Figure 4. 8. The plot result for theoretical and sample variograms.


Theoretical and Sample Semivariogram for potato yield residuals
Variogram
0
1
2
3
4
5
6
7
0
10
20
30
40
50
lag distance
type
Gaussia
regular
robust

This is also shown in Figure 4.8 that the pattern of the 72 overlapped empirical variograms have very similar shape. This is due to the fact that the robust method is meant to be robust to outlying values, while have we are changing the distance from a points, not the observed yield values. That is, in theoretical models considered previously, the lag distance h entered as a scalar value. This implies that the correlation between the spatial process at two point pairs P1; P2 is dependent only on the separation distance h =| P1P2 |, not on the orientation of the two points.

We found that at small lags, estimation of the spatial structure could be seriously affected. What matters here is that the tolerance interval are large and spatial locations are regular. The superimposed star (*) in Figure 4.8 on continuous line shows the weighted least square fit of an exponential variogram model for potato yield and we observe the classical and robust empirical variograms in one plots were overlapped.

In Figure 4.8 a scale of c0=7.5 and a range of a0=30 displayed to fit reasonably well for both the robust and standard semivariogram. The horizontal line at 7.5 variance units corresponding to c0 is called the "sill." In the case of the spherical model, Z(h) actually reaches this value. For the other two model forms, the sill is a horizontal asymptote. In particular, the dimension of c0 is the same as the dimension of the variance of the spatial process $\{Z(S): r^2 \ \varepsilon \ D \subset R^2\}$. The dimension of a0 is length with the same units as lag distance h.

From our findings about the robustness of the variogram for potato yield data, we feel it is wise to perform and present analysis based on several different random spatial locations rather than assigning the lattice values to just one arbitrary chosen point. The method discussed in this section are easily extendable to any shaped

lattice and can be used to examine the robustness of any analysis of lattice data that uses lag distance between plots.

From theoretical consideration discussed the robust variogram estimates are generally smaller than the classical estimates for potato yield data. It is noted from Figure 4.7 and Figure 4.8 that the major distinguishing feature of the Gaussian and exponential forms is the shape in the neighborhood of the origin h = 0. In general, we conclude that small lags are important in determining an appropriate theoretical form based on a sample semivariogram estimated from the potato yield data.

Once an appropriate varigram model is chosen the next step is producing the kriging surface prediction model. After an appropriate variogram model is chosen, there are a number of choices involved in producing the kriging surface.

4.3.3. Spatial Prediction

Spatial prediction is any prediction method that incorporates spatial dependence. In agricultural field trial analysis, data are available at specific spatial locations (such as experimental plot yield from the field), and the goal is to predict unsampled locations. The unsampled locations are often mapped on a regular grid, and the predictions are used to produce surface plots or contour maps. A simple and popular spatial prediction method as discussed in section 3.2.5.3, is ordinary kriging. Ordinary kriging requires a model of the spatial continuity, or dependence. This is typically in the form of a covariance or semivariogram. In ordinary kriging, we allow the variable values from the measured locations to act as covariates for the values at unmeasured locations. For each predicted location, we fit a linear function to these covariates (opting for all measured points or, more often, a subset of points in the neighborhood of the predicted point), applying the variance-covariance

structure implied in the variogram. This generates an unbiased linear prediction and a standard error of the value at the location S_i.

Tables 4.19 shows the model $\hat{\gamma}(h)$ of equation (3.26) chosen to achieve approximate unbiasedness for the estimator $\hat{Z}(s_0) = \sum_{i=1}^{n} Z(s_i)\lambda_i$, *where* $\sum_{i=1}^{n} \lambda_i = 1$ as a predictor.

Thus, for a lag 10 (ten), the predicted value is:

$$\hat{Z}(s_0) = 0.0198Z_1 + 0.13291Z_2 + 0.19658Z_3 + 0.19003Z_4 + 0.14698Z_5 + 0.11423Z_6 + 0.08521Z_7 + 0.06042Z_8 + 0.03819Z_9 + 0.01512Z_{10} + 0.00053Z_{11}$$

The last step in contouring the data is to decide on the grid point locations. A convenient area that encompasses all the data points is a rectangle of length 3m and 4m. The spacing of the grid points depends on the use of the contouring; a spacing of 3 distance units (m) is chosen for plotting purposes. Table 4.20 shows result of kriged standard error estimates:

$$\sigma^2{}_k(s_0) = \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_j) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(s_i - s_j)$$

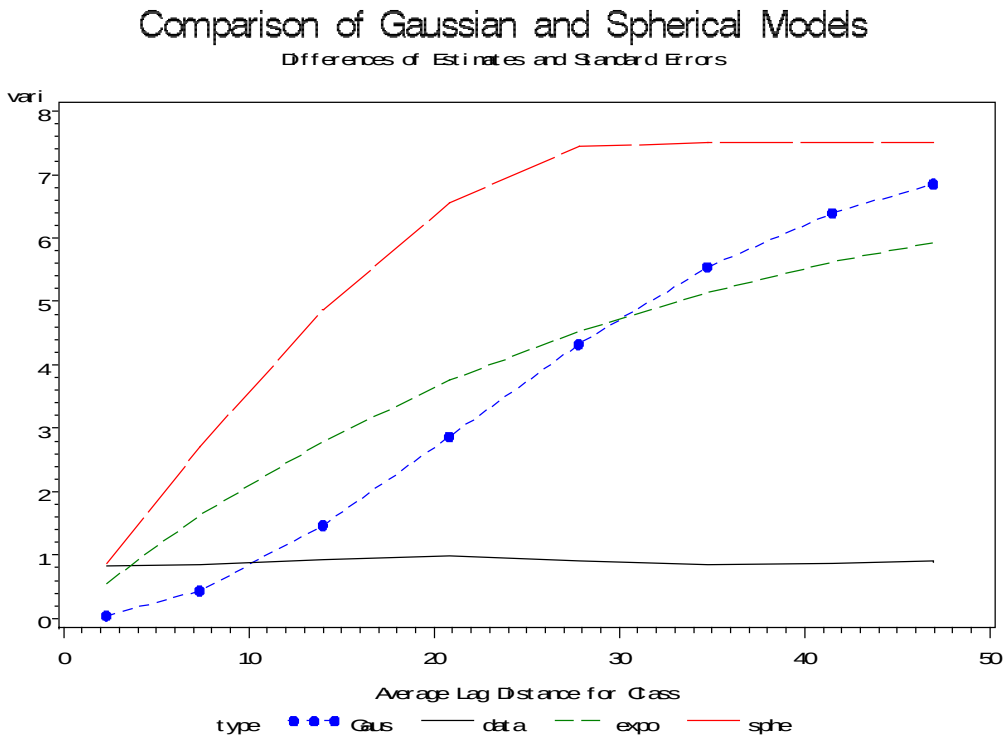
for potato yield data from Kulumsa site and computes the kriged surface using these parameter and grid choices. The kriged surface plot result is plotted in Figure 4.9, and the result for associated standard errors are plotted in Figure 4.10 that shows the standard errors are smaller where more data are available.

Table 4.20. standard error of kriging estimates

Obs	LAG	COUNT	VARIOG	type
1	-1	931	0	Gaussian
2	-1	931	0	regular
3	-1	931	0	robust
4	0	15020	0.60048	Gaussian
5	0	15020	0.82760	regular
6	0	15020	0.80333	robust
7	1	97158	4.27679	Gaussian
8	1	97158	0.85834	regular
9	1	97158	0.82769	robust
10	2	117368	7.14893	Gaussian
11	2	117368	0.92400	regular
12	2	117368	0.89349	robust
13	3	82649	7.49146	Gaussian
14	3	82649	0.98053	regular
15	3	82649	0.95735	robust
16	4	59063	7.49996	Gaussian
17	4	59063	0.91391	regular
18	4	59063	0.89524	robust
19	5	38748	7.50000	Gaussian
20	5	38748	0.84806	regular
21	5	38748	0.83430	robust
22	6	19705	7.50000	Gaussian
23	6	19705	0.87148	regular
24	6	19705	0.86308	robust
25	7	3204	7.50000	Gaussian
26	7	3204	0.90746	regular

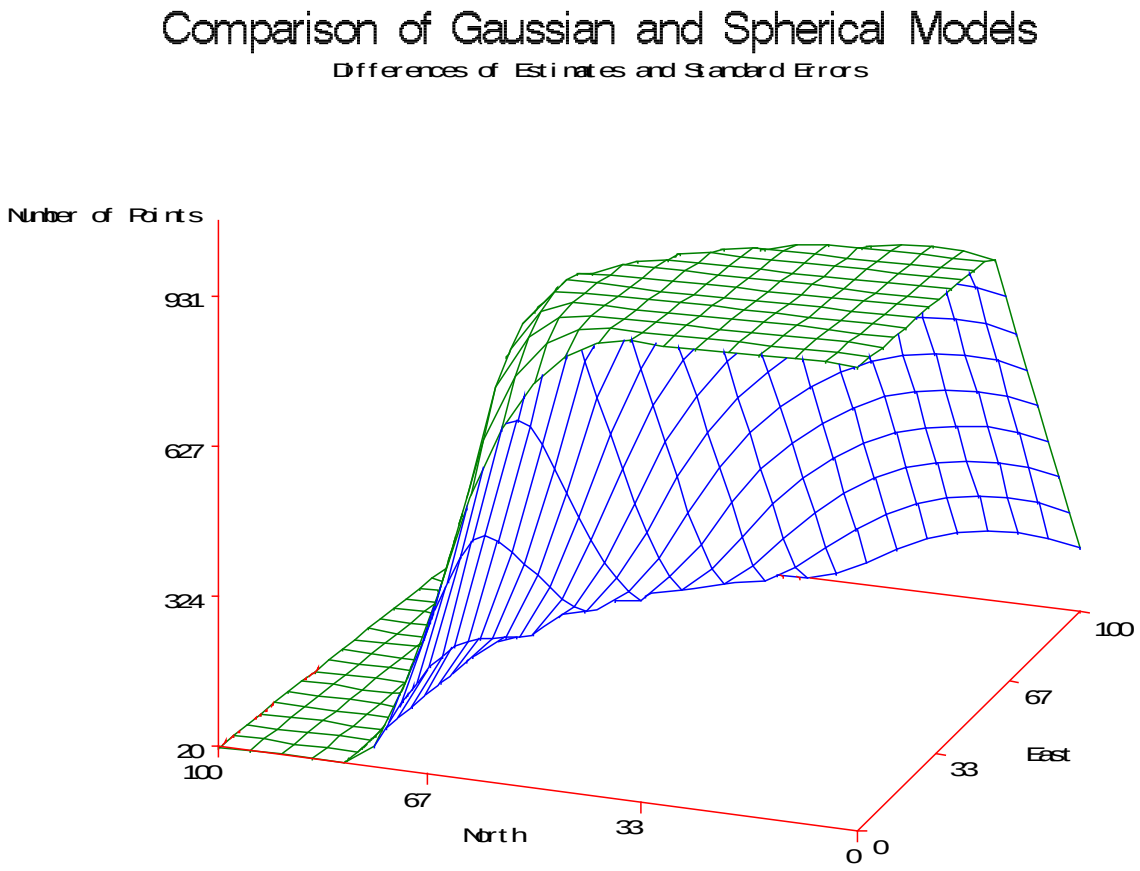
27	7	3204	0.88227	robust

Figure 4.9. Comparison of Gaussian, exponential and spherical Model estimates for standard errors of residuals on variances each drown aganist against average lag distance.



The predicted values at each of the grid locations do not differ greatly for the two variogram models. However, the standard error of prediction for the spherical model is substantially larger than the Gaussian model. The Gaussian model is larger than the exponential model for average lag distance of greater than 30, equal approximately at 30 and less for <30.

Figure 4.10. Surface plot of standard error of kriging estimates drown for number of points against Northing



In this plot the edge points have far fewer points contributing to their estimates, generally resulting in higher standard errors than non-edge points. However, we can see that the most central location appears to have the most neighbors, but is an elevated point in the standard errors surface. Additional points do not always improve estimates, especially if the additional points are outside the range of spatial correlation. The predicted values at each of the grid locations do not differ greatly

for the two variogram models. However, the standard error of prediction for the spherical model is substantially larger than the Gaussian model.

In general, the chapter attempts to show the spatial analysis techniques can considerably increase our capacity of understanding the spatial patterns associated to agronomic yield data from field trial based on potato yield from Hollota and Kulumsa agricultural research centers. Explanatory techniques like Moran's and Geary's indicates existence of spatial autocorrelation among plot yield. Thus, proper experimental design that goes with the pattern of soil heterogeneity, specifically spatial autoregressive statistical model fitting is becoming a powerful tool that can detect spatial variation in agricultural field trial. Estimating a variogram may be sensible to outlying values but when predicting observations, an alternative way of dealing with them is needed. One of the methods of treating is applying random spatial location effect for regular lattice. In our case, predicted values at each of the grid locations do not differ greatly for the classical and robust variogram models. Thus, the spatial effect, orientation of plot designs and blocks should be done carefully as they have to go with the pattern of soil heterogeneity and topography of the field.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Conclusion

In a potato yield, the growth of each plant is obviously influence by its neighbors. Because of limitations on space and nutrients, the plants tend to compete among themselves rather than to cooperate. It is of interest to measure the competition or effect between the plants. To do this first the yield from specific plot size is measured and a diagnosis for complete spatial randomness and detection for spatial pattern among potato yield is practiced. Lastly, an appropriate statistical model to these yield data are fitted. Essentially responses from n plots in agricultural field trial may be considered to measure the influence of one response up on another in some specific way.

In this study, some of the techniques of spatial analysis was applied on potato yield data collected from Hollota and Kulumsa agricultural research centers on 1.5 ha of land in each center in year 2001. First, the Monte Carlo test for complete spatial randomness is applied using the ranked statistics that compare the observed value and the calculated values and the result shows that no complete spatial randomness is detected in the potato yield data. This implies that the model fitted should be either spatially (regularly) patterned or clustered.

Based on this result, it was attempted to apply Moran's index and Geary's coefficients in particular to test for spatial autocorrelation between combined plots

of potato yield data separately for both agricultural research centers. The result gained shows a significant positive spatial autocorrelation with approximate value of 0.5 and 0.46 was detected for plot size of $12m^2$ in Hollota and Kulumsa research centers respectively and its effect is slightly greater in Hollota than that of Kulumsa. The result for Hollota agricultural field trials on plot size of approximately 12 m^2 through 36 m^2, produce very less variability, and the yield gained from plots are spatially dependent. Similarly, for potato yield trial in Kulumsa, plot of size 48 m^2 or more will minimize or eliminate spatial autocorrelation between adjacent plots. Hence, it can be concluded that even increasing plot size will not generally make the observed spatial autocorrelation insignificant.

The spatial autoregressive response (AR) model analysis approach, that is adjusted for (take in to account) spatial autocorrelation effect on the row and column, was applied on the same potato yield data where significant positive spatial autocorrelation is detected and the result is compared with the OLS-based analysis. The result shows there is significant row effect in both AR- based and OLS models. Furthermore, the result shows that applying AR-based model is more efficient than OLS with 6% for potato yield data from Hollota site and 86% from that of Kulumsa. Similarly, the amount of explained variation (R^2) in grain yield increases from about 41% for OLS –based ANOVA to about 71% for AR-based ANOVA for Hollota center and about 66% for OLS –based ANOVA to about 69% for AR-based ANOVA for Kulumsa center.

The effect of the weighted autocorrelation parameter indicating that AR- response model is non linear. Spatial auto-regression model are appropriate alternatives to traditional regression for spatial data because they eliminate the dependencies that cause problems, and are sensitive to the specification of the weighted matrix, as it is demonstrated in the above application on potato yield data.

The classical and robust variogram models were used to produce a map of predicted potato yield data, taking in to account plot variation from the model prediction method called Kriging for Kulumsa agricultural research center. The result shows the predicted values at each of the grid locations do not differ greatly for the two variogram models. However, in the comparison of Gaussian and Spherical models the standard error of kriged prediction for the spherical model is subsequently larger than the Gaussian model for potato yield data.

Recommendations

- It is recommend for agronomists to utilize spatial statistics in analyzing auto-correlated experimental field crop yield data using two-step approach. First, use the Moran's autocorrelation to see if a crop response variable contains spatial autocorrelation. Second, apply spatial statistics if significant autocorrelation is detected. Therefore, it is important to develop further improved field plot techniques for potato cultivation to increase precision of experiments through use of optimum plot dimension and replication.

- The method of spatial autoregressive model is recommended to be adoptable to agronomic field data because the results of regression are identical to the approach used in generalized linear models. Moreover, regression models allow the establishment of the relationship variables, taking in to account the spatial effect, in the case, the explaining power of the model can benefit from significant gains.

- The theoretical foundation of statistics relay on the set of assumptions and sampling procedures that are often more applicable to experimental datasets than purely observational data. Very few problems addressed by spatial

analysis fall into the category of truly experimental research, hence, spatial analysis is often the start of further investigation into process and model building, and rarely an end it self.

- The SAS computer code for computing the autocorrelation, estimating an AR model and prediction can be modified to accept any regular lattice structure.

Limitations of the study

- Spatial analysis confronts many fundamental issues in the definition of its objects of study, in the construction of analytic operations to be used, in the use of computers for analysis, and in the presentation of analytic results. Many of these issues are active subjects of modern research. The most fundamental of these is the problem of defining the spatial location of the entities being studied. Other issues in spatial analysis include the limitation of mathematical knowledge, the assumption required by existing statistical techniques, and problems in computer packages based statistical analysis.
- Another limitation of our method is that it only considers fitting one parametric variogram model to change for empirical variogram. It might be useful to allow the actual variogram model to change for different spatial location from an exponential variogram to a spherical variogram or a linear variogram, for example.
- The results presented for the prediction purpose are limited to the case of variogram model fitting and ordinary kirging methods. That is, results may differ if a different fitting procedure were used.

References:

Arlinghaus, S.L. (1996). *Practical Handbook of Spatial Statistics.* CRS press, Inc. USA.

Atkinson, P.(1997). *Simulating Locational Error in field- basedmeasurements of "Reflectance", in geo ENVI-Geostatistics for Environmental Application.* Amesterdam, Kluwes.

Bartlett. M.S. (1937). *Properties of Sufficiency and Statistical tests.* Proceedings of the Royal Society A, 160, 128-282.

Barteles, R. (1997). *In the use of Limit Theorem Arguments in Economic Statistics,* Vol. 5 JAI press, Greenwich, CT.

Bartlett. M.S. (1978). *Nearest neighbor models in the analysis of field experiments.* Journal of the Royal Statistical Society b, 40, 147-158.

Besag, J.E. and Kempton, R. (1986). *Statistical of field Experiments using neighboring plots.* Biometrics.

Birkhoff, G. (1967). *Lattice Theory, 3rd ed. American Mathematical Society*, Providence, RI.

Browine, C.; Bawman D.T.; and Burton, J.W. (1993). *Estimating Spatial Variation in analysis of data from field trial. Argon. J.*

Briand. Reply, (1981). *Spatial Statistics*; John Wiley& sos, Inc, New York.

Chhidda Singh, (1996). *Modern techniques of Raising Field Crops*.Oxford & IBH Publishing Co. Pvt.Ltd, New Delhi, Calcutta.

Christakos, G. (1992). *Random Field Models in Earth Science*. New York, Acadamic Press.

Clark, P.J. and Evans, F.C (1954). *Distance to nearest neighbor as as a measure of spatial relationship in populations*. Ecology, 35, 445-453.

Clff,A.D. and Ord, J.K. (1973). *Spatial Autocorrelation*. Pion, London.

Clff,A.D. and Ord, J.K. (1975). The *comparison of means when samples consists of Spatially auto-correlated observations*. Env. Plan. A. 7:725-734.

Clff,A.D. and Ord, J.K. (1981). *Spatial Process: Models and Application*. Pion, London.

Cressie, N. (1991). *Statistics for spatial data*. Jhon Wiley, new York.

Cressie, N. rev.ed. (1993). *Statistics for spatial data*. Jhon Wiley, new York.

Crissie, N. and Hawkins, D.M. (1980). *Robust estimation of the variogram*, I. Journal of the International Association for Mathematical Geology, 12,115-125.

David, M. (1977). *Geostatistical ore Reserve estimation*. Elsevier, Amster dam.

Delfiner, P. (1976). *Linear estimation of of non-stationary spatial phenomena In Advanced geostatistics in the mining Industry*, M. guarascio M. David, and C. Huijbergts, eds. Redial, dordrecht, 49-68.

Durrett,R. (1994). *Stochastic Spatial Models*. Forefronts (News letter of the cornell Theory center)9(#4,spring), 4-6.

Fairfield Smith, H. (1938). *An Empirical low describing heterogeneity in the yield of agricultural crops*. Journal of Agricultural science 9Cambridge), 28, 1-23.

Fisher, R.A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

Fedrer, W.T. (1955). *Expermental Designes*. New York,Macmillan.

Gautwirth, J.L.,and Rubin, H. (1975). *The behavior of robust estimators on dependent data analysis of Statistics* ,3,1070-1110.

Geary, R.C. (1954). *The Contiguity Ratio and Statistical mapping*, The In corporate Statistician, 5, 115-145.

Gilmour, Cullis, B. R., Welham, S. J. & Thompson, R. (2004). An efficient computing strategy for prediction in mixed linear models. Comput.Statist. Data Analy. 44,571-586.

Girma, T.; Yohannes, T.; G/Medhin, W.G.; Tefera, A.and Peter, n.(2002). *Optimum plot and Block Dimensions and effect of plot shape on significance test in Potato*. Journal of the Ethiopian Statistical association. 12:39-59.

Girma, T.; Amsale, T. and Tanner D.G. (2000). *Estimation of Optimum Plot Dimension and Replication number for wheat Experimentation in Ethiopia.* African Crop Science Journal, 8:1:11-23.

Girma, T. A. (2005). *Using Spatial Modeling Techniques to Improve Data Analysis from Agricultural field Trial*. Unpublished PhD thesis University of KwaZulu-Natal, Pietermartizburg.

Gomez, K.A. and Gomez, A.A. (1984). *Statistical Procedures for Agricultural Research*, Wiley, New York.

Gower, J.C. (1988). *Statistics and Agriculture*. Journal of the Royal Statistical Society A. 151/ 179-200.

Green, P.J.; Jonnison, C.; and Seheult, A.H. (1984). *Analysis of Field experiment by least square smoothing*. J. Roy. Statist. Sec. Ser. B. 47(2): 299-315.

Griffith, D. A. (1988a). *Advanced Spatial Statistics*, Advanced Studies in Theoretical and Applied Economic Klower Academic Publishers, Dordrecht, The Netherlands.

Griffith, D. A. (1988b). *Estimating Spatial Autoregressive model parameter with Commercial statistical Packege*, Geogr. Annl, 20 (2): 176-186.

Griffith, D. A.(ed.) (1990). *Spatial Statistics: Past present, and future of Mathematical Geography*. Ann Arbl.Ml.

Griffith, D. A.(ed.) (1993). *Spatial Regression Analysis on the PC*. Spatial Statistics Using SAS. Assoc.Am. Geo. Washington D.C.

Grondona, M.O. and Cressie, N. (1991). *Using Spatial considerations in the analysis of experiments*. Technomet:33 (4):381-392

Hammersley, J. M. and Mazzarino, G. 919830. *Markov fields, correlated Percolation, and the Ising model*. Springer, New York.

Haning, (1980). *Statistical test and process generators for random field models*. Geographical analysis, 11, 45-64.

Haning, R.P. (1978). *Spatial Model for high planes agriculture*. Ann. Ass.Am. Geog. 68:(4): 493-504).

Hart, J.F. (1954). *Central tendency in areal distributions*. Economic Geography.

Hatheway, W.H. and Williams, E.J. (1961). *Efficient Estimation of the relationship between plot size and the variability of crop yields*. Biometrics.

Hawkins, D.M. and Crisse, N. (1984). *Robust Kriging a proposal*. Journal of the International Association for Mathematical Geology, 16: 61-75.

Huberl P.J. (1972). *Robust Statistics*. A review, Analysis of Mathematical Statistics, Wiley, New York.

Huberl P.J. (1981). Robust Statistics. Wiley, new York.

Jay,A., Cressie, N. and Paul, R. (2004). *Journal of Computational and graphical Statistics*, 13:2

Journel, A.G. Huijbregt, Ch.J. (1978). *Mining Geostatistics*. New York,Acadamic Press.

Kiefer, J. and Wynn, H.P. (1981). *Optimum balanced block and Latin square design for correlated observations.* Annals of Statistics, 9, 737-757.

Kinfemika'el, A. (2000). *Application of Spatial Statistical method, to the study of soil fertility pattern*. Addis Ababa University.

Koch, E.J.and Rigeney, J.A. (1951). *Method of estimating Optimum plot size from Experimental Data*. Agronomic Journal:43:17-21.

Krige. D.G.A. (1951). *A Statistical Approach to some mine valuations and sallied problem.* Watawatesr and Master's thesis, University of Witewatesrand.

Lewis A. and Hutchinson,M. F.(2000). *From Data Quality using Spatial Statistics to predict the implication of Spatial Error in Point Data*. Ann Arbor Press.

Matheron, G. (1963b). *Principles of Geostatistics*, Economic geology, 58, 1246-1266.

Matheron, G. (1965). *La theorie des variables Reginalisees et ses Applications*. Masson,Paris.

Mercer, W.B. and hall, A.D. (1911). *The experimental Error of Field trials*, Journal of Agricultural Science (Cabridge), 4, 107-132.

Moran, P.A.P. (1948). *The Interpretation of statistical maps*, journal of Royal Statistics society BIO, 243-251.

Moran, P.A.P. (1950). *Notes on Continuous stochastic phenomena*. Biometriika.37.

Morista, M. (1959). *Measuring of the dispersion and analysis of distribution patterns*. Memories of faculty of Science, Keyushu university, series E. Biology 2, 251-235.

Narayana, R. and Ramanta , C. (1982). *Effect of Plot Shape on Variability in Smithes Variance Law*. Expermental Agriculture: 18:333-38.

Narayana, R. and Ramanta , C. (1985). *Experimental Plot size and shape based on the data from uniformity trial with dry land in pigonpea*. Indian Journal of Agricultural Science

National Scince Foundation 91991). *Solicitation*: National center for Geographic information and Analysis. Biological, and Social Sciences directorate, Washington, D.C.

National Reaserch Council (board on Mathematical Science). *Renewing U.S. Mathematics*: A Plan for the 1990s national accadamy Press, washgton, D.C.

Odland, J., (1988). Spatial Autocorrelation.Sage, Beverly Hills, CA.

Olson, K. R.; Carmer, S.G.; and Olson, G.W. (1985). *Assessment of effects of soil variability on maximum alfalfa yields*, GEDMAB, New York.

Papadakis, J.S. (1937). *Me'thods Statitique pour des exp'eriences sur champ*. Bulletin de l'Institute Am'el. Plants a'sa lonque 23.

Patterson, H.D. & Thompson, R. (1991). *Recovery of interblock information when block sizes are unequal*. Biometrika 31,100-109.

Pielou, E.C. (1959). *The use of Point-to-plant distance in the pattern of plant populations*, Journal of Ecology, 47, 607-613.

Ripley, B. (1981). Spatial Statistics. Wiley, New York .

Robert Haning. (1990). *Spatial Data Analysis in Social and Environmental Science*; Cabridge University press.

Serra, J. (1980). *The Boolean Model and Random sets*. Computer graphics and Image processing, 12, 99-126.

Smith, H. F. (1938). *An Empirical Low Describing Heterogeneity in the yields* . Journal of Agricultural Science, 28:1-23.

Sokal, R. R., and Oden, N.L. (1998). *Spatial Autocorrelation in Biology*. Methodology, “Biological Journal of the Linnean Society”:10:19.

Stetler, F. (1982). *Specifying Weights in spatial forecasting* models. The result of some experiments . Environment and Planning A, 14,571-582.

Tobler, W. (1997). *Cellulae geography. In philosophy in Geography*, edited by S. Gale and G. Olsson, pp. 379-86, Reidel, Dordrecht.

Upton,G.J. and Fingleton,B. (1985. *Spatial data analysis by example*, Volume1: *point pattern and quantitative data*. Wiley, Toronto, Singapore.

Val Es, H.M. and Van Es, C.L. (1993). *The Spatial number of randomization and its effect on the outcome of Field Experments*. Agon J. 85:420- 428.

Yanli, Z. and Melanie, M. W. (2004). Journal of Computational and Graphical Statistics. 13:3: 719-738.

Yates, F. (1938). *The Comparative advantage of systematic and randomization arrangement in the design of agricultural and a biological experiments*. Biometrika, 30,444-466.

Zinmerman, D.L. and Harville, D.A. (1991). Arandom *Field approach to the analysis of Field plt experiment and other spatial Experments*.Biometrcs 47:223-239

APPENDEX I

Original potato yield data Hollota

2.68	2.80	2.40	3.04	2.00	2.90	2.62	2.88	2.76	2.88	1.70	2.30	2.80	2.48	2.30	2.50	2.10	1.70	2.50	1.70	1.90	2.50	1.96	2.52	2.88	2.86
1.60	1.40	2.50	2.00	2.62	2.10	2.90	2.30	2.50	2.10	2.40	2.20	1.94	2.84	3.30	2.10	2.50	1.90	2.40	2.10	2.20	2.92	2.80	3.00	1.50	1.30
2.00	1.78	2.80	2.10	1.91	2.90	2.10	3.00	2.10	2.60	3.30	2.20	2.60	2.96	2.84	2.20	2.70	2.20	1.50	2.80	2.00	1.78	2.82	3.40	1.90	2.54
1.60	1.80	2.30	2.70	2.50	1.68	3.50	3.00	3.30	2.31	2.40	2.61	2.36	2.60	2.50	2.61	2.30	2.50	2.90	2.00	2.30	2.70	2.00	2.90	2.50	2.60
3.74	2.56	4.30	3.10	3.60	3.80	2.50	3.40	2.80	2.78	2.47	2.31	2.40	1.50	3.72	2.40	1.70	2.60	2.80	4.00	2.40	3.10	3.30	3.55	2.80	2.30
2.30	2.90	2.80	2.30	2.70	2.78	2.10	3.20	2.60	2.31	2.04	2.02	2.60	2.20	2.24	1.60	1.70	1.80	1.90	1.50	2.10	3.00	2.30	2.60	1.68	2.40
1.94	2.64	3.10	1.80	2.72	3.30	3.20	2.40	2.30	2.80	1.94	2.70	2.80	3.20	2.40	2.60	2.30	2.10	3.00	2.40	2.90	1.72	1.60	1.90	3.60	2.30
2.40	2.30	3.20	2.60	1.60	3.00	2.90	4.10	2.26	2.60	1.80	2.00	3.60	2.60	3.70	2.80	3.20	3.30	2.40	2.90	2.70	2.80	2.50	3.08	3.30	2.40
2.40	2.40	2.50	2.86	2.30	2.60	2.30	2.00	2.10	2.80	2.20	2.20	1.80	2.06	2.00	2.10	1.90	1.80	1.90	1.50	1.74	1.60	2.70	1.50	3.00	1.64
2.98	3.20	2.90	2.80	2.90	3.00	22.50	2.80	2.90	2.40	2.56	2.42	3.00	3.20	3.44	3.00	3.00	4.70	3.10	2.52	3.52	3.46	2.40	3.40	2.80	2.00
3.10	3.42	3.80	3.60	3.04	2.70	3.20	2.52	3.02	3.10	3.00	2.20	2.30	1.96	3.70	2.30	2.96	2.50	2.06	1.60	2.00	2.20	2.30	2.50	2.80	3.16
2.70	2.94	3.04	2.00	2.64	3.40	2.90	2.76	2.20	3.00	2.40	2.30	2.40	2.50	2.16	2.60	2.70	3.10	3.00	2.30	3.00	2.10	2.30	2.50	2.50	3.90
2.80	2.90	2.00	2.50	2.70	2.30	2.00	2.60	2.80	2.60	2.80	2.60	2.80	2.40	3.51	3.00	2.50	2.30	2.60	3.02	2.80	2.34	2.20	3.96	2.10	2.90
3.20	4.10	2.40	2.30	2.80	2.50	3.50	2.50	2.80	4.10	3.00	2.20	2.30	2.40	2.52	2.84	2.70	3.04	2.80	1.70	2.50	3.20	2.40	3.06	2.60	2.40
2.40	3.00	2.60	2.00	1.82	3.30	2.20	1.70	3.00	3.80	2.30	2.50	2.30	2.80	2.70	1.40	1.86	2.32	3.20	2.40	2.04	2.05	2.50	2.50	3.50	3.60
2.60	3.30	2.98	2.00	3.10	2.10	2.40	2.70	2.46	2.80	2.85	2.74	2.80	2.20	1.80	2.40	2.50	2.00	2.86	2.20	2.00	2.20	2.50	2.40	3.30	3.16
4.70	2.68	2.00	3.01	2.70	3.00	3.30	2.90	3.40	2.20	1.20	3.20	2.20	2.30	2.10	2.92	2.90	2.30	2.50	1.96	2.20	1.80	2.20	3.00	2.10	2.50
2.90	2.14	2.20	3.70	3.70	3.30	2.30	2.18	2.00	1.96	2.21	2.70	2.90	2.60	2.40	3.00	1.90	2.20	2.64	2.30	2.04	2.70	2.86	2.42	2.50	3.29
3.60	2.00	3.60	2.00	3.30	2.30	3.06	2.80	2.10	3.50	3.60	2.20	2.60	2.90	1.62	2.10	3.50	2.60	2.50	2.94	1.60	1.96	3.10	2.00	3.60	1.50
3.30	3.40	3.44	4.00	3.50	3.20	1.70	2.90	2.74	2.88	1.90	1.60	2.90	4.20	3.70	2.70	3.06	2.20	3.20	2.80	3.00	2.80	3.00	2.52	2.10	2.30
4.06	2.80	3.40	2.72	3.30	3.20	2.60	2.80	2.30	4.16	3.62	3.40	2.30	3.28	2.60	3.10	2.70	2.50	3.00	1.62	2.30	3.30	2.70	3.38	3.60	4.01
2.40	2.26	2.34	1.92	1.92	1.74	1.90	3.40	2.54	2.04	2.20	1.84	2.20	1.50	1.76	2.14	2.30	1.70	1.80	2.20	2.20	1.90	1.94	2.06	3.16	2.52
2.36	2.24	2.08	2.30	1.90	1.36	1.42	2.72	2.62	1.90	2.14	1.64	1.70	1.20	1.90	2.18	2.20	2.50	2.40	1.24	1.52	1.30	2.30	1.50	1.54	1.66
1.62	1.10	2.00	1.72	1.80	1.40	1.90	1.80	1.20	1.80	1.12	1.70	1.96	1.92	1.00	1.80	2.00	1.16	1.72	1.50	1.80	1.20	2.02	1.80	1.40	1.60
1.70	1.90	2.54	2.64	2.52	2.20	2.32	2.06	2.90	1.60	2.40	1.00	2.00	2.10	2.62	3.40	1.64	2.10	1.94	2.24	2.56	2.56	2.74	2.24	2.80	2.10
1.90	1.80	2.64	2.22	1.90	2.36	2.50	1.80	2.70	1.50	2.00	1.70	1.00	1.40	1.44	2.40	1.04	1.62	1.86	1.92	1.60	1.80	2.22	2.06	2.20	2.00
2.00	2.06	2.10	1.90	1.96	2.40	2.50	1.60	1.34	1.00	2.24	1.62	2.70	2.12	1.70	1.60	2.02	2.60	2.90	2.60	2.00	2.40	2.20	2.60	2.50	1.20
1.98	1.70	2.00	2.20	2.00	2.36	2.00	1.90	1.40	1.60	2.20	1.60	1.80	2.30	2.00	2.70	1.50	2.90	2.70	2.54	2.42	1.70	2.00	1.90	2.00	1.40
1.16	1.22	1.42	1.20	1.30	2.00	1.84	1.70	1.30	2.98	2.38	1.80	1.20	1.80	1.90	2.10	1.70	1.90	2.30	2.00	1.50	3.30	2.00	2.10	1.00	1.80
1.40	1.90	1.78	1.70	1.30	1.90	1.50	2.16	1.58	1.70	1.60	1.50	1.52	1.80	1.00	1.02	2.60	2.16	1.40	1.10	2.10	1.30	1.00	1.70	2.40	1.20

2.40	2.68	2.20	2.64	1.52	3.31	2.66	2.94	3.00	3.04	2.60	2.66	2.20	2.40	3.60	2.10	2.54	3.20	3.42	2.00	3.54	3.42	2.80	2.70	2.50	3.70
2.50	1.60	2.30	2.80	2.70	2.20	2.80	2.92	2.36	2.70	3.00	2.10	2.60	2.70	2.16	2.08	2.40	3.70	3.24	2.20	3.40	3.10	2.90	3.11	3.00	2.40
2.10	2.74	2.70	2.18	2.92	2.50	2.12	1.94	2.50	2.50	2.48	2.38	2.50	2.70	2.00	2.10	2.20	2.90	2.14	2.70	2.50	2.60	2.40	1.90	1.90	2.10
2.84	2.80	2.40	2.60	2.78	2.96	2.70	1.54	3.10	2.10	2.20	2.98	2.50	2.10	2.30	2.90	3.70	2.50	2.70	3.00	2.94	3.36	2.46	2.30	1.60	1.66
2.60	1.78	2.10	2.80	1.20	2.04	2.60	2.30	2.10	2.88	2.60	2.50	3.40	2.50	2.52	2.70	2.10	2.00	3.00	3.20	2.40	2.70	2.30	2.72	3.30	4.66
3.40	3.00	3.80	2.70	2.60	2.41	2.60	3.00	3.78	2.98	2.30	2.10	2.66	2.10	1.80	3.60	2.70	2.80	2.90	3.06	2.76	2.80	2.36	3.40	2.96	2.56
3.00	3.10	3.30	1.70	2.20	2.00	3.30	2.00	1.68	3.38	2.92	2.60	2.62	2.50	2.00	3.30	2.20	3.20	2.90	3.40	3.20	3.40	2.60	1.90	2.20	2.80
3.50	2.80	2.86	2.00	2.70	3.10	3.00	3.60	2.00	2.70	2.10	2.40	2.30	2.50	2.90	3.20	3.04	2.56	2.96	3.26	2.70	3.30	1.78	3.10	3.20	3.00
2.92	3.10	2.30	2.46	2.08	3.06	2.80	2.40	2.30	1.86	2.60	2.66	1.80	2.94	2.66	2.90	2.40	2.54	2.52	3.38	2.70	2.10	2.80	2.98	3.36	2.96
3.24	2.90	2.78	3.50	2.68	2.60	3.00	3.40	3.30	2.98	2.10	2.00	2.70	2.60	4.00	2.30	3.42	2.90	1.50	2.90	2.76	2.40	2.60	2.70	3.62	2.50
2.34	2.60	3.20	3.10	2.90	2.56	2.90	2.60	2.76	2.84	3.50	2.80	2.00	2.20	2.90	2.60	3.00	2.60	3.20	4.00	3.50	2.28	2.70	2.80	2.40	3.00
3.64	3.50	3.40	3.40	2.80	2.20	3.80	4.00	2.86	3.30	3.50	2.50	3.60	3.94	3.40	2.50	2.70	2.50	2.78	4.58	2.10	2.40	3.30	2.88	3.30	3.00
1.86	1.30	1.64	1.14	1.96	1.86	1.42	1.20	2.98	1.76	2.46	1.72	2.30	3.32	3.00	1.65	1.66	2.22	2.60	3.62	4.08	3.04	3.60	2.96	2.70	3.72
2.68	1.44	0.86	2.46	1.72	2.24	2.24	2.11	2.30	2.56	1.90	2.84	2.98	1.18	2.76	1.94	2.52	2.64	2.50	2.39	3.50	2.54	3.50	2.40	2.78	2.24
2.24	1.88	1.08	2.00	1.80	2.12	1.62	1.90	2.30	3.00	2.76	2.14	2.70	1.92	2.54	2.70	2.52	2.96	3.80	3.66	2.30	2.50	2.82	2.60	3.00	2.50
2.90	2.12	1.54	2.10	1.90	2.96	2.60	2.40	2.60	2.63	2.50	2.61	2.40	3.20	2.70	2.80	3.60	2.40	2.50	2.40	2.50	2.46	2.38	3.06	2.00	3.80
3.00	3.40	2.90	2.60	3.60	2.14	2.54	3.16	2.88	2.98	2.30	3.06	2.20	2.74	2.30	1.14	3.70	1.80	2.90	1.40	2.80	2.70	2.09	2.70	2.54	4.08
3.50	1.80	2.00	1.46	2.84	2.26	1.70	3.16	1.40	2.30	2.16	2.34	2.20	2.84	2.26	2.84	2.60	2.20	3.18	2.40	2.28	3.30	2.04	1.60	2.60	1.40
2.74	2.28	1.40	2.40	2.44	1.90	2.54	2.12	2.98	2.00	2.50	2.40	2.84	2.50	2.66	2.20	1.60	2.06	1.40	2.74	2.24	2.10	2.54	1.56	2.60	1.60
2.04	1.70	1.84	0.90	2.80	2.50	2.68	1.52	2.20	1.72	2.96	2.40	1.70	2.18	2.20	2.22	2.56	2.16	2.36	1.92	2.50	1.50	1.50	2.12	3.21	2.90
2.88	2.46	2.84	2.00	2.26	1.60	2.20	2.56	2.30	1.40	2.80	3.50	1.60	1.70	4.00	3.86	1.40	1.94	2.76	2.50	2.70	2.64	2.34	2.42	2.60	2.24
3.30	3.10	2.52	2.68	3.40	3.80	4.20	2.60	1.62	4.30	1.92	2.52	3.38	2.73	2.40	2.44	2.80	2.74	2.44	2.90	2.44	3.00	2.50	2.12	2.72	3.82
3.04	2.50	2.60	3.18	2.70	3.10	2.00	1.56	3.20	2.66	3.86	3.10	2.28	2.86	2.50	2.18	2.60	2.50	2.08	2.00	2.30	2.20	3.46	2.80	2.40	2.94
3.32	2.88	2.42	3.30	3.70	3.40	2.60	1.80	1.83	3.20	1.80	1.98	3.50	2.70	2.10	2.98	1.68	1.98	2.90	2.30	2.90	3.44	2.54	2.50	2.90	6.24
2.70	3.70	2.56	2.70	2.20	3.66	2.30	3.30	2.50	3.40	3.08	2.60	1.90	2.60	2.10	3.56	2.80	2.70	2.10	2.50	2.30	2.36	2.16	2.60	2.30	2.60
3.14	2.90	3.50	2.70	4.00	2.90	2.70	2.80	2.20	3.08	2.70	3.08	4.20	2.24	2.20	2.50	2.60	3.24	3.00	2.10	2.50	2.10	3.10	2.84	1.80	2.30
4.08	3.50	4.24	3.90	2.20	2.94	3.26	2.30	2.20	4.20	2.90	4.10	3.30	2.60	3.40	2.50	1.50	1.70	1.24	2.10	1.50	2.60	1.90	2.90	3.00	3.60
3.50	2.00	4.00	3.80	2.90	3.60	2.60	2.90	2.80	3.50	2.50	2.10	2.80	3.50	3.20	1.80	3.00	3.00	3.10	2.70	4.00	2.40	3.40	4.70	1.80	2.96
2.98	2.70	2.50	2.80	3.10	3.10	2.60	2.70	3.20	2.80	1.90	2.50	2.90	3.50	2.80	2.20	2.70	2.90	1.90	2.20	1.80	2.26	2.32	2.40	2.80	4.00
3.90	3.60	2.90	2.44	2.06	2.90	1.50	2.60	2.34	3.60	2.50	2.60	3.90	3.02	2.10	1.90	3.40	2.40	2.60	3.48	2.40	1.96	1.50	2.30	2.70	1.50
3.60	2.60	3.30	3.42	2.52	2.50	2.30	2.42	2.52	2.26	2.58	2.56	2.66	3.30	2.80	2.50	2.00	3.10	2.20	2.30	2.90	2.04	2.50	2.30	2.80	4.32
4.14	3.20	3.10	3.30	2.90	2.88	3.08	2.40	3.80	3.00	2.80	2.90	2.80	2.62	2.52	3.58	1.16	2.34	2.86	3.40	2.10	1.50	3.86	2.10	2.40	2.80
3.60	3.66	2.40	3.24	1.60	3.10	2.80	2.78	2.98	3.18	3.70	2.70	3.20	2.50	1.80	3.20	2.10	1.96	3.98	3.34	2.64	2.20	2.70	3.20	3.70	2.90

APPENDEX II Original potato yield data Kulumsa

5.04	4.8	6.7	6.5	4.2	5.7	5.3	5.3	5.7	6.5	5.7	6.5	7.7	4	5.4	5.2	6.9	5.9	3.3
5.24	6.9	6.1	5.2	3.9	6	6	7.3	4.8	5.8	5.3	5.7	4.8	5.5	7.2	4.9	5.2	5.3	6.1
7.94	5.1	6.9	4.7	5.1	3.4	5.7	4.3	6	5.2	6	6	5.6	5.8	6.5	5.8	4.9	5.5	3.1
6.6	4	5.5	4.3	4.7	6.2	5.8	4.9	4.4	7.5	6.2	5.5	5.9	3.9	5.4	6.2	3.3	5.9	6.3
5.46	7.6	5.7	5.3	5.2	6.5	5.6	5.4	6.4	5.8	4.4	5	4.7	4.9	5.7	4.8	5.1	4.5	3.4
7.1	5.7	5	4.7	6.1	6.5	5.2	5.9	6.5	5.9	5.7	6.4	7.1	5	5.6	5	6.7	4.2	6.7
7.1	6.4	4.9	5.3	7.2	4.4	6.3	6.7	3.5	4.2	4.7	6.6	6.4	4.6	6.4	6.7	8.2	3.8	4.9
5.7	5.2	6.3	5.5	5.3	5.6	6	6.2	6.8	5.1	5.3	6.7	4.5	4.3	6.9	4.9	6.2	6.6	4
5.64	6.7	4.8	4.9	5.7	5.2	6.2	4	4.9	5	5	5.5	6.9	4.3	4.7	6.3	4.7	4.7	6.4
5.88	5.2	5.1	4.4	5	6.6	4.6	5.5	4.1	6	4.3	5.4	6.6	4.1	6.3	4.3	5.5	6	5.3
3.76	5.5	5.7	6.1	5.6	4.5	6.1	5.3	6.2	5.1	6.8	5.9	4.8	5.9	3.9	5.3	5.6	5.9	7.4
6.16	5.2	5.5	4.4	6.7	3.9	7.6	6.1	6.8	7.5	5.5	7.1	4.9	6.4	4.4	7.5	4.1	5.3	5
3.14	5.7	6.4	5.4	4.3	5.6	4.8	4.8	6.3	6.6	5.3	5	4	5.1	5.2	5.1	5.9	7.7	6.2
7.54	5.4	4.3	5.9	4.6	5.9	5.6	5.6	4	4.9	3.7	5.8	6.4	6.4	6.1	6.4	5	4.7	6.3
4.1	5.9	5.5	6.6	4.5	6.3	7.2	6	6	6.3	5.1	3.8	5.4	3.5	4.9	4.9	5.5	5.2	5.1
5.5	5.1	4.9	4.3	5.7	4.5	5.5	5.6	4.9	7.9	3.7	6.1	6.1	5.5	4.2	5.2	3.9	6.5	6.2
5.7	5	4.1	7	5.7	6	7.1	6.3	5.9	5.2	4.6	6.6	5.3	5.1	5.4	4.9	5.9	5.3	4.4
7.3	5.7	5.3	6.1	5.2	2.9	5.2	5.1	4.6	5.9	5.4	5.1	5.5	6.3	5.1	5.8	4.1	5.7	5.1
6.46	6.6	4.1	6.7	5	7	5.1	4.4	6.2	6.6	3.9	5	5	5	8.5	5	4.7	4.3	6.6
4.66	6.3	4.6	5.7	6.2	5.9	4.8	5.9	6.7	4.8	8.6	4.2	5	5.4	4.5	4.4	5.2	6	4.9
3.88	4.9	4.8	4.9	4	3.6	5.7	5	5	3.7	2.6	6.3	4.8	6.6	5.3	4.4	5.7	4.9	5.5
6.64	7.1	4.6	4.7	4.9	3.3	5.1	5.5	5.1	4.4	4.5	4.5	4.9	5.2	4.5	4.5	3.5	5.7	6
5.42	4.2	5.2	5.2	3.8	5	3.8	6	5.6	4.5	4.5	3.8	2.7	5.8	4	4.2	4.6	5.6	6.4
5.12	4.3	5.9	5.8	2.6	3.2	5.2	5.3	6	5	6	5.6	3.3	3.4	4	4.2	4.2	5.2	6.7
5.34	4.7	3.5	5.8	3.8	4.9	5.2	4.8	5.2	3.7	5	4.7	6.3	4.7	5.3	5	4.1	4.4	4.3

6.18	4.1	3.7	5.5	5.4	4	5.2	4.1	4.9	4.3	5.2	4.2	4.9	5.7	4.8	5.5	4.4	3.4	5.5
5.08	5.3	5.4	6	4.1	4.4	4.6	4	4.1	5.2	3.7	4.6	3.2	3.2	4.2	5.1	8.2	5.4	6.1
4.14	5.2	4	5.5	4.5	4.8	5.3	5.3	5.1	5.5	4.3	5.8	6.2	4.7	4.4	5.7	4.9	3.4	4.3
5.58	4.3	4	5.8	4.2	3.8	3.9	4.3	4.9	4	5.3	3.6	4.3	5.8	5.4	4	3.9	4.2	5.1
4.4	4.6	4.5	5.6	5	4.1	5.5	5.5	4.1	3.7	4.7	6.2	4.9	5.2	5.5	6.2	5.5	4.7	5.2
7.1	4.2	5.4	5.9	4.9	5	4.3	5.1	4.7	5	5.3	4.6	4	5.8	5.1	5.7	3.3	5.5	4.5
4.5	5	4.9	4.6	5	4.3	5.3	4.5	4.5	4.9	5.9	6.9	7.2	6.2	5.6	5.5	4.4	4.2	2.6
4.9	5.9	4.4	5.7	5.1	4.3	4.3	5.6	5.2	5.3	5	6.7	3.6	5	5.4	5.4	4.5	5.4	4.1
6.56	5.9	4	6	6.2	5.8	4.3	5.9	5.5	4.5	4.5	6.5	4.6	5.4	5.9	5.6	4.7	5.1	3.7
3.64	5.3	4.3	6.2	5.1	4.9	6.1	4.4	5	4	4.1	5.4	4	5.7	4.3	6.5	3.9	4.2	4.9
6.88	5.8	4.8	5.3	4	5	5.5	6	5.8	6	5.4	6.6	5.8	5	5.2	5.7	6.2	5.2	6.8
6.02	5.04	5.6	4.9	5.2	6.2	5.3	4.5	5.5	6.5	5.6	7.3	6.4	5.4	6.2	4.9	5.7	5.1	6.9
6.1	5.82	5.2	5.5	6	6.3	4.6	5	5.4	4.9	6.3	5.3	4.9	5.8	5.4	6.2	5.4	4.7	5.2
5.04	4.82	6.2	6.4	5.5	6.2	4.8	6.7	5.5	5	6.3	6.8	5.6	5.2	3.7	5.7	6.2	4.5	6.7
6.12	4.99	7.4	3.8	5.4	5.1	5	5.9	5.5	5.9	5.8	4	4.9	4.3	5.5	6.2	5.7	5.1	6
4.15	4.88	4.4	6.2	5.1	5.9	5.3	4.7	5.3	5.2	5.9	8	4.7	3.4	6.8	6.4	5.7	5.4	7.5
6.22	3.38	5.2	5.4	4.6	6.1	4.6	4.8	6.1	4.8	5.4	6	5.6	4.3	6.1	4.4	4.3	3.9	5.3
6.41	5.66	7	5.4	5.3	5.5	5.2	7.3	5.5	4.8	5	6.2	5.3	6.3	5.2	4.5	5.2	6.5	6.1
5.58	4.6	5.9	4.5	5.2	5.1	5.7	5.4	4.8	5.8	5.5	7.2	5.5	5.1	3.2	4.5	5.1	5.6	5.5
6.38	5.59	4.6	7	4.7	5.6	6.1	4.6	4.5	4.6	6.5	4.8	5.3	5	6.1	4.6	4.8	6.3	5.6
5.42	6.06	5.1	5.1	4.9	6.3	4.9	5.5	5.3	5.8	7	5.3	5.2	7.4	6.1	3.7	4.2	7.4	6.1
4.6	5.04	5.9	5.8	5.6	4	5.6	5.5	4.9	4.2	6.3	6.2	6.1	4.2	5.1	5.2	6.2	4.7	6.4
6.02	5.62	4.5	5.5	4.8	5.5	4.7	6.9	5.7	4.4	3.8	5.1	5.6	7.8	5.2	4.5	4.3	5.2	5.1
6.94	5.9	6.1	5.6	5.2	5	5	4.3	5.4	5	6.3	7.1	5.9	6.9	5.9	7.2	4.9	5.2	5.9

APPENDEX III

```
/*SAS Code for Spatial Autocorrelation
*************************************************************;
* AUTOCORR.SAS - Spatial Autocorrelation Analysis Program.   *;
* Compute Moran's I and Geary's C spatial autocorrelation    *;
* measures. Generate asymptotic tests as well as Monte Carlo *;
* tests for significance. Requires as input, both a Weight   *;
* data set and a Response Data Set.                          *;
Filename Totalyld'C:/My Document/Totalyld;
Option linesize=72;
*************************************************************;
TITLE1 'SPATIAL AUTOCORRELATION';
/*
 * Here we have 120 items with a response X measured on each.
 * These items are linked together according to the weight
 * matrix in data WEIGHTS. Use Moran's I and Geary's C
statistics
 * to test for spatial autocorrelation among the items.
 */
DATA RESPONSE;
     Infile Totalyld;
*************************************************************;
* I is the location index and X is the response value.       *;
*************************************************************;
 INPUT I X @@;  /* X'S ARE ASSUMED IN THE CORRECT ORDER */
 LIST;
DATA WEIGHTS;
Infile Totalyld;
*************************************************************;
* I and J are the location indices for the pair and WT is    *;
* the weight assigned to the pair. Specify all pairwise      *;
```

APPENDEX III(continued)

```
* links between units considered to be directly correlated.  *;
************************************************************;
 INPUT I J WT;
 LIST;
PROC IML; /* SAS INTERACTIVE MATRIX LANGUAGE */
************************************************************;
* Use IML to do the actual computations. This lets us easily *;
* do the Monte Carlo test as well. For the Monte Carlo test  *;
* you will be given the list of results for each simulation  *;
* as well as the rank of each result. The final value will   *;
* be the observed value and rank for the data. If the rank   *;
* is extreme then you will reject Ho that the data are       *;
* distributed as under the simulation (randomly).            *;
* NOTE: The printing format for the weights might need to be *;
* changed if you are using non-integer valued weights. I used*;
* 3.0 to keep the amount of the output down.                 *;
************************************************************
START LOADDATA;
 USE RESPONSE VAR {X};
 READ ALL;
 CLOSE RESPONSE;
 N=NROW(X);
 W=J(N,N,0); /* INITIALIZE WEIGHT MATRIX TO ZEROS */
 USE WEIGHTS VAR {I J WT};
 READ ALL;
 /* PLACE WEIGHTS INTO THE WEIGHT MATRIX */
 DO K=1 TO NROW(I);
  W[I[K],J[K]]=WT[K];
  W[J[K],I[K]]=WT[K];  /* ASSUMING SYMMETRY IN WEIGHTS */
 END;
 /* DROP THE I J AND WT MATRICES */
 FREE I J K WT;
```

```
 PRINT / 'INITIAL DATA ' ,, 'RESPONSE',X,, 'WEIGHTS',
W[FORMAT=3.0];
   /* $$$    THE FORMAT MAY NEED TO BE CHANGED IF    $$$ */
   /* $$$     FRACTIONAL WEIGHTS ARE TO BE USED.     $$$ */
FINISH;
START STATS;
 XBAR=X[+]/N;
 Z=X-REPEAT(XBAR,N,1);
 ZSQ=Z`*Z;
 XVAR=ZSQ/(N-1);
 VMRATIO=XVAR/XBAR;
 Z2=Z#Z; Z4=Z2`*Z2;
 B2=N*Z4/ZSQ**2; /* KURTOSIS */
 FREE Z2 Z4;
 SUMW=0; S1=0;
 DO K=1 TO N;
  DO J=1 TO N;
   IF K <> J THEN
    DO;
      SUMW=SUMW + W[K,J];
      S1=S1 + (W[K,J]+W[J,K])**2;
    END;
  END;
 END;
 S1=S1/2;
 S2=W[,+]+W[+,]`;
 S2=S2`*S2;
 SUMW2=SUMW*SUMW;
 TEMP=N||XBAR||XVAR||VMRATIO||B2;
 RNAME={" "};
 CNAME={"   N" "   XBAR" "VARIANCE" "V/M RATIO" "KURTOSIS"};
 PRINT / 'SUMMARY STATISTICS ' TEMP[ROWNAME=RNAME
COLNAME=CNAME];
FREE XBAR XVAR VMRATIO K J RNAME CNAME TEMP;
```

```
FINISH;
START AUTOCORR;
 I=0; C=0;
DO K=1 TO N;
  DO J=1 TO N;
    IF K <> J THEN
     DO;
       I = I + W[K,J] * Z[K]*Z[J];
       C = C + W[K,J] * (X[K]-X[J])**2;
     END;
  END;
 END;
 I=(N/SUMW)*(I/ZSQ);
 C=((N-1)/(2*SUMW))*(C/ZSQ);
 FREE K J;
FINISH;
START AUTOSTAT;
 RNAME={" "}; CNAME={"I   " "C   "};
 TEMP=I||C;
 PRINT ,, 'SPATIAL AUTOCORRELATION STATISTICS'
      TEMP[ROWNAME=RNAME COLNAME=CNAME];
 FREE RNAME CNAME TEMP;
FINISH; /* AUTOSTAT */
START TESTOFI;
 EXPECT=1/(1.0-N);
 RNAME={" "}; CNAME={"  E(I)  " "OBSERVED"};
 TEMP=EXPECT||I;
PRINT "TEST OF I", "EXPECTED AND OBSERVED VALUES"
  TEMP[ROWNAME=RNAME COLNAME=CNAME];
 VAR=((N**2*S1-N*S2+3*SUMW2)/(SUMW2*N**2-1))-EXPECT**2;
 ZZ=(I-EXPECT)/SQRT(VAR);
P=1-PROBNORM(ABS(ZZ));
 CNAME={"VARIANCE" "   Z   " "  P>|Z|  "};
 TEMP=VAR||ZZ||P;
 PRINT  'TEST BASED ON  NORMALITY:'
```

```
         TEMP[ROWNAME=RNAME COLNAME=CNAME];
 VAR=(N*((N**2-3*N+3)*S1-N*S2+3*SUMW2)-
       B2*((N**2-N)*S1-2*N*S2+6*SUMW2))/
       ((N-1)*(N-2)*(N-3)*SUMW2) - EXPECT**2;
 ZZ=(I-EXPECT)/SQRT(VAR);
 P=1-PROBNORM(ABS(ZZ));
 TEMP=VAR||ZZ||P;
 PRINT  'TEST BASED ON  RANDOMIZATION:'
         TEMP[ROWNAME=RNAME COLNAME=CNAME];
 FREE EXPECT VAR ZZ P RNAME CNAME TEMP;
FINISH;
START TESTOFC;
 EXPECT=1.0;
 RNAME={" "}; CNAME={"  E(C)  " "OBSERVED"};
 TEMP=EXPECT||C;
 PRINT "TEST OF C", "EXPECTED AND OBSERVED VALUES"
TEMP[ROWNAME=RNAME COLNAME=CNAME];
 VAR=((2*S1+S2)*(N-1)-4*SUMW2)/(2*(N+1)*SUMW2);
 ZZ=(C-EXPECT)/SQRT(VAR);
 P=1-PROBNORM(ABS(ZZ));
 CNAME={"VARIANCE" "   Z   " "  P>|Z|  "};
 TEMP=VAR||ZZ||P;
 PRINT  'TEST BASED ON  NORMALITY:'
         TEMP[ROWNAME=RNAME COLNAME=CNAME];
 /* ZZ AND P BELOW ARE TEMPORARIES */
ZZ=(N-1)*S1*(N**2-3*N+3-(N-1)*B2)-(N-1)/4*S2;
 P=(N**2+3*N-6-(N**2-N+2)*B2)+SUMW2*(N**2-3-(N-1)**2*B2);
 VARC=ZZ*P/(N*(N-2)*(N-3)*SUMW2);
 ZZ=(C-EXPECT)/SQRT(VAR);
P=1-PROBNORM(ABS(ZZ));
 TEMP=VAR||ZZ||P;
 PRINT  'TEST BASED ON  RANDOMIZATION:'
         TEMP[ROWNAME=RNAME COLNAME=CNAME];
 FREE EXPECT VAR ZZ P RNAME CNAME TEMP;
FINISH;
```

```
START MONTE;
 ALLI=I; ALLC=C;
 DO REP=1 TO 99;
    R=UNIFORM(REPEAT(0,N,1));
    Y=X; X[RANK(R),]=Y;
    Y=Z; Z[RANK(R),]=Y;
    RUN AUTOCORR;
    ALLI=I//ALLI; ALLC=C//ALLC;
 END;
 PRINT / 'MONTE CARLO TEST - OBSERVED VALUE IS LAST VALUE';
 R=RANK(ALLI); ALLI=ALLI||R;
 CNAME={"  I " "RANK"};
 PRINT 'RANKS FOR I VALUES:' ALLI[COLNAME=CNAME];
 PRINT / 'MONTE CARLO TEST - OBSERVED VALUE IS LAST VALUE';
 R=RANK(ALLC); ALLC=ALLC||R;
 CNAME={"  C " "RANK"};
 PRINT 'RANKS FOR C VALUES:' ALLC[COLNAME=CNAME];
 FREE ALLI ALLC REP R Y ;
FINISH;
START SAVEW;
DO I=1 TO NROW(W);
 W[I,I]=2;
END;
CREATE WMATRIX FROM W;
APPEND FROM W;
CLOSE WMATRIX;
FINISH;
RESET NOLOG NONAME FW=10;
RUN LOADDATA;             /* READ THE DATA INTO THE MATRICES */
RUN STATS;                /* COMPUTE SUMMARY STATISTICS      */
RUN AUTOCORR;             /* COMPUTE AUTOCORRELATION STATS.  */
RUN AUTOSTAT;             /* PRINT COMPUTED STATS.           */
RUN TESTOFI;              /* COMPUTE TESTS OF I              */
RUN TESTOFC;              /* COMPUTE TESTS OF C              */
RUN MONTE;                /* PERFORM MONTE CARLO SIMULATIONS */
```

```
RUN SAVEW;                  /* SAVE WEIGHT MATRIX FOR MDS      */
QUIT;                       /* DONE WITH SAS/IML               */
/*
 * Use the weights as similarities among the units. Use MDS
 * to display units as linked together by the weights.
 */
DATA WMatrix;
 Set WMatrix;
 Id+1;
RUN;
Proc MDS Data=WMatrix SIMILAR=2 LEVEL=Ordinal DIM=2 PCONFIG
OUT=Config;
 ID Id;
 Var Col1-Col192;
Run;
%let plotitop = gopts = device = gif, Color = Black, Colors =
Black;
title2 'Plot of Weight Matrix Configuration';
%Plotit(Data=Config(Where=(_type_='CONFIG')), Datatype=mds,
           LabelVar=Id, VtoH=1.75);
 /* Add lines in graph to joined units. */
Data New;
 Set Weights;
 PType+1;
Id=I; Output;
 Id=J; Output;
 Keep PType Id;
Run;
Proc Sort Data=New;
 By Id;
Run;
Proc Sort Data=Config;
 By Id;
Run;
Data New;
```

```
 Merge New Config(Where=(_type_='CONFIG'));
 By Id;
Run;
Data New;
 Set Config(In=In1) New;
 If In1 Then PType=0;
Run;
GOptions Reset=Symbol Reset=Axis;
Title2 "Weight Configuration With Connections";
Proc GPlot Data=New;
 Plot Dim2*Dim1=PType / NoLegend VAxis=Axis1 HAxis=Axis2;
 Axis1 Order=(0 To 100 By 10) Length=100in Label=(A=90);
 Axis2 Order=(0 To 100 by 10) Length=100in;
 Symbol1 C=Black V=Dot I=None;
 Symbol2 C=Black V=None I=Join L=1 R=100;
Run;
Quit;
```

APPENDEX IV

```
*THE SAS CODE FOR SPATIAL AUTOREGSSION MODEL*
FILENAME YIELDTOHOL'C:MY DOCUMENT/HOLLOTA;
OPTION LINESIZE=72;
*-----------------------------------------------------------------*
*DATA ARE FOR A SQRT(N)-BY-SQRT(N) REGULAR LATTICE, WHICH HAS *
*BEEN AUGMENTED TO A SQRT(N)-BY-(SQRT(N)+1) REGULAR LATTICE. *
*THEADDITIONAL LATTICE COLUMN IS ADDED TO THE RIGHT HAND SIDE,*
*AND CONTAINES MISSING VALUES. DATA ARE ENTERED IN TO      *
*"YIELD.DAT" BEGINNING WITH THE UPPER LEFT HAND CORNER OF THE
*LATTICE, AND MOVING FROM LEFT TO RIGHT. THIS RESULTS IN       *
*ACONCATENATION OF SUCESSIVE LATTICE ROW ENTRIES, WITH EACH*
*SUCESSIVE PAIR BEING SEPARETED BY A MISSING VALUE AREAL UNIT
*VALUE,*
*-----------------------------------------------------------------*;
DATA STEP1;
     INFILEYIELDTOHOL;
     INPUT NUMBER X ROW COLUMN V1 V2 V3 V4 V5 V6
          V7 V8 V9 V10 V11 V12 V13 V14 V15
          V16 V17 V18 V19 V20;
Y=LOG(X);
;
RUN;
*-----------------------------------------------------------------*;
*TWO DIRECTIONS OF SPATIAL LAG VALUES ARE DETERMINED. N4 MUST*
*BE LAGGED BY SQRT(N)+1, RESULTING IN CHANGE IN "LAG_".*
*-----------------------------------------------------------------*;
DATA STEP2;
     SET STEP1;
N2=LAG1(Y);
N4=LAG7(Y);
PROC SORT DATA=STEP2;
     BY DESCENDING NUMBER;
RUN;
```

```
*-------------------------------------------------------------*;
*THE REMAINING TWO DIRECTIONS OF SPATIAL LAG VALUES ARE    *
*DETERMINED. N3 MUST BE LAGGED BY SQRT(N)+1, RESOLUTION IN *
*ACHANGE IN "LAG_".*
*-------------------------------------------------------------*;
DATA STEP3;
     SET STEP2;
N1=LAG1(Y);
N3=LAG7(Y);
RUN;
*-------------------------------------------------------------*;
*DETERMINATION OF THE NUMBERS AND THE AVERAGE NEIGHBORING  *
*VALUES FOR EACH AREAL UNIT.*
*-------------------------------------------------------------*;
DATA STEP4;
     SET STEP3;
IF N1 EQ '.' THEN NBR1=0;ELSE NBR1=1;
IF N2 EQ '.' THEN NBR2=0;ELSE NBR2=1;
IF N3 EQ '.' THEN NBR3=0;ELSE NBR3=1;
IF N4 EQ '.' THEN NBR4=0;ELSE NBR4=1;
NEIGH=NBR1+NBR2+NBR3+NBR4;
IF N1 EQ '.' THEN N1=0;
IF N2 EQ '.' THEN N2=0;
IF N3 EQ '.' THEN N3=0;
IF N4 EQ '.' THEN N4=0;
WX=(N1+N2+N3+N4)/NEIGH;
IF Y EQ '.' THEN WY= '.';
RUN;
*-------------------------------------------------------------*;
*ESTIMATION OF THE SAR SPATIAL AUTOCORREGRESSIVE PARAMETER*
*(RHO). "ALPHA" AND "DELTA' ARE SPECIFIC TO SQRT(N), AND COME*
*FROM THE JACOBIAN APPROXIMATION TABULATED RESULTS.A1=0.143535
and d1=1.067932 are specific to the 20-by -6 regular lattice of
the variety trial*
*-------------------------------------------------------------*;
```

```
PROC NLIN METHOD=MARQUARDT MAXITER=500;
     PARMS RHO=0.5,B0=0,B1=0,B2=0,B3=0,B4=0,B5=0,
          B6=0,B7=0,B8=0,B9=0,B10=0,B11=0,B12=0,
          B13=0,B14=0,B15=0,B16=0,B17=0,B18=0,B19=0,B20=0
          ;
     BOUNDS -1.0<RHO<1.0;
     A1=0.143535;
     D1=1.067932;
     JHAT=EXP(A1*(2*LOG(D1)-LOG(D1+RHO)-LOG(D1-RHO)));
          ZY=Y*JHAT;
     MODEL ZY=(RHO*WY+(1-RHO)*B0+B1*R2+B2*R3+B3*V2+B4*V3+B5*V4+
          B7*V6+B8*V7+B9*V8+B10*V9+B11*V10+B12*V11+B13*V12+
          B14*V13+B15*V14+B16*V15+B17*V16+B18*V17+
B19*V18+B20*V19)*JHAT;
     DER.B0=(1-RHO)*JHAT;
     DER.B1=(ROW)*JHAT;
     DER.B2=(COLUMN)*JHAT;
     DER.B3=V2*JHAT;
     DER.B4=V3*JHAT;
     DER.B5=V4*JHAT;
     DER.B6=V5*JHAT;
     DER.B7=V6*JHAT;
     DER.B8=V7*JHAT;
     DER.B9=V8*JHAT;
     DER.B10=V9*JHAT;
     DER.B11=V10*JHAT;
     DER.B12=V11*JHAT;
     DER.B13=V12*JHAT;
     DER.B14=V13*JHAT;
     DER.B15=V14*JHAT;
     DER.B16=V15*JHAT;
     DER.B17=V16*JHAT;
     DER.B18=V17*JHAT;
     DER.B19=V18*JHAT;
     DER.B20=V19*JHAT;
```

```
     DER.B21=V20*JHAT;
     DER.RHO=((RHO*WX+(1-
RHO)*B0+B1*R2+B2*R3+B3*V2+B4*V3+B5*V4+B6*V5+B7*V6+B8*V7+B9*V8
          +B10*V9+B11*V10+B12*V11+B13*V12+B14*V13
          +B15*V14+B16*V15+B17*V15+B17*V16+B18*V17
          +B19*V18+B19*V18+B20*V19-Y)*(2*A1*RHO/(D1**2-
     RHO**2))+WY-B0)*JHAT;
     ID JHAT;
     OUTPUT OUT=TEMP1 P=XHAT; RUN;
     DATA STEP5;
          SET STEP1;
          SET STEP4;
          SET TEMP1;
          CORY = YHAT/JHAT;
          ME=((X-CORX)**2);
     RUN;
     PROC MEANS SUM;
          VAR ME;
          OUTPUT OUT=MS SUM=MSE;
     PROC PRINT DATA=MS;
               VAR MSE;
     RUN;
     ********************************************************
     *OBTAIN THE PSEUDO R2 FOR THE AR MODEL*
     ********************************************************;
     PROC CORR;
          VAR Y CORY;
     RUN;
     ********************************************************
     *PERFORM ANOVA WITH YIELD UNADJUSTED FOR AUTOCORRELATION*
     *AND YIELD ADJUSTED FOR AUTOCORRELATION.*
     ********************************************************;
     DATA STEP6;
          SET STEP5;
          RO=0.0726;
```

```
     MN=5.9302;
     ADJY =Y-(RO*WY-RO*MN);
RUN;
PROC GLM;
     CLASS ROW COLUMN;
     MODEL Y= ROW COLUMN;
RUN;
PROC GLM;
     CLASS ROW COLUMN;
     MODEL ADJY= ROW COLUMN;
RUN;
```

APPENDEX V

```
/*-------------------------------------------------------------*/
 /*       SAS code for Spatial Prediction & Krigiging
            */
 /*                 Using the SAS System
*/
/*-------------------------------------------------------------*/
/*************************************************************/
/*    NAME: SPATIAL.SAS
*/
 /*   TITLE: PROCS VARIOGRAM, KRIG2D CODS
*/
 /* PRODUCT: SAS/STAT  SAS/GRAPH BASE
*/
/*    KEYS: VARIOGRAM KRIGING SPATIAL PREDICTION
*/
 /*   PROCS: VARIOGRAM, KRIGE2D, GPLOT, PRINT
*/
 /*    DATA:   POTATO YIELD
*/
      FILENAME YIELDKULUMSA'C:YLD.DAT';
     OPTIONS LINESIZE=72;
/*************************************************************/;
data yld ;
     INFILE YIELDKULUMSA;
   input row column  yld @@;
run ;
proc print data=yld split='#' ;

var column row yld ;
   label column  = 'column# m'
      row = 'row# m'
      yld = 'potato#yield (kg/m2)'
```

```
      ;
   title 'Simulated potato yield' ;
   run ;
symbol1 value=dot height=.5 cv=black ;
proc gplot data=yld ;
   plot  row*column  ;
   label row = 'row-m.'
     column= 'column-m.';
   title 'Scatter Plot of Measurement Locations' ;
   run ;
proc g3d data=yld ;
   label row = 'row'
         column='column'
         yld = 'yield'
         ;
   scatter row*column=yld/ xticknum=5 yticknum=5 grid ;
   title 'Surface Plot of potato yield' ;
run ;
 /*- Produce an initial histogram of inter-pair distances ----*/
proc variogram data=yld outdistance = outd ;
   compute novariogram  ;
     coordinates xc=row yc=column  ;
     var yld ;
run ;
 /*- Print the outd data set   -------------------------------*/
proc print data=outd ;
 title 'OUTDISTANCE= Data Set Showing Distance Intervals' ;
run ;
 /*- Get a midpoint for charting  ----------------------------*/
data outd ; set outd ;
   mdpt = round((lb+ub)/2,.1) ;
   label mdpt = 'Midpoint of Interval' ;
run ;
 /*- Chart of initial histogram   ----------------------------*/
proc gchart data=outd ;
```

```
   vbar mdpt / type=sum sumvar=count discrete ;
   title 'Distribution of Pairwise Distances' ;
run ;

 /*- Produce last histogram of inter-pair distance------------*/
proc variogram data=yld outdistance = outd ;
compute novariogram nhc=20  ;
  coordinates xc=row yc=column ;
  var yld ;
run ;
 /*- Print the new outd data set   ----------------------------*/
proc print data=outd ;
   title 'OUTDISTANCE= Data Set Showing Distance Intervals' ;
run ;
 /*- Get a midpoint for charting  ----------------------------*/
data outd ; set outd ;
  mdpt = round((lb+ub)/2,.1) ;
  label mdpt = 'Midpoint of Interval' ;
run ;
 /*- Chart of last histogram   ---------------------------------
----*/
proc gchart data=outd ;
   vbar mdpt / type=sum sumvar=count discrete ;
   title 'Distribution of Pairwise Distances' ;
run ;
 /*- Compute the variogram   ---------------------------------*/
proc variogram data=yld
     outvar = outv
   ;
   compute lagdistance=7 maxlag=10 robust ;
   coordinates xc=row yc=column ;
   var yld ;
   run ;
 /*- Print the outv data set   -------------------------------*/
 proc print data=outv label ;
```

```
   var lag count distance variog rvario ;
   title 'OUTVAR= Data Set Showing Sample Variogram Results' ;
run ;
 /*- Rearrange data for plotting semivariogram   -------------*/
data outv ; set outv ;
  vari = variog ; type = 'regular' ; output ;
  vari = rvario ; type = 'robust'  ; output ;
run ;
axis1 minor=none label=(c=black 'lag distance') offset=(3,3) ;
axis2 minor=(number=1)
      label=(c=black 'Variogram') offset=(3,3) ;
symbol1 i=join l=1 v=star c=black  ;
symbol2 i=join l=1 v=square c=black  ;
/*- plot the semivariogram   ---------------------------------*/
proc gplot data=outv ;
  plot vari*distance=type / haxis=axis1 vaxis=axis2 ;
   title 'Standard and Robust Semivariogram for
potato yield Data' ;
run ;
 /*- Add in theoretical semivariogram for plotting  ----------*/
data outv3 ; set outv ;
c0 = 7.5 ; a0 = 30 ;
  vari = c0*(1-exp(-distance*distance/(a0*a0))) ;
  type = 'Gaussian'; output ;
  vari = variog ; type = 'regular' ; output ;
  vari = rvario ; type = 'robust'  ; output ;
run ;
symbol1 i=join l=1 v=star c=black  ;
symbol2 i=join l=1 v=square c=black ;
symbol3 i=join l=1 v=diamond c=black ;
 /*- Plot all the semivariograma   --------------------------*/
proc gplot data=outv3 ;
     plot vari*distance=type / vaxis=axis2 haxis=axis1 ;
   title 'Theoretical and Sample Semivariogram for potato yield
Data' ;
```

```
run ;
 /*- Do the kriging   ----------------------------------------*/
proc krige2d data=yld outest=est ;
   pred var=yld r=60 ;
   model scale=7.5
         range=30
         form =gauss;
  coord xc=row yc=column ;
  grid x=0 to 100 by 10 y=0 to 100 by 10 ;
run;
 /*- Plot the predicted values   ----------------------------*/
proc g3d data=est ;
 scatter gyc*gxc=estimate /
         grid
         ;
label gyc='column'
       gxc='row'
       estimate='yield'
       ;
   title 'Surface Plot of Kriged potato yield' ;
run ;
 /*- Plot the standard errors   -----------------------------*/
proc g3d data=est ;
 scatter gyc*gxc=stderr /
         grid
         ;
 label gyc='column'
       gxc='row'
       stderr='Std Error'
       ;
   title 'Surface Plot of Standard Errors of Kriging Estimates'
;
run ;
/*- Run KRIGE2D on original Gaussian model ------------*/
proc krige2d data=thick outest=est1;
```

```
pred var=yld r=60;
model scale=7.5 range=30 form=gauss;
coord xc=east yc=north;
grid x=0 to 100 by 10 y=0 to 100 by 10;
run;
/*- Run KRIGE2D using Spherical Model, modified range -*/
proc krige2d data=yld outest=est2;
pred var=yld r=60;
model scale=7.5 range=60 form=spherical;
coord xc=east yc=north;
grid x=0 to 100 by 10 y=0 to 100 by 10;
run;
data compare ;
merge est1(rename=(estimate=g_est stderr=g_std))
est2(rename=(estimate=s_est stderr=s_std));
est_dif=g_est-s_est;
std_dif=g_std-s_std;
run;
proc print data=compare;
title ’Comparison of Gaussian and Spherical Models’;
title2 ’Differences of Estimates and Standard Errors’;
var gxc gyc npoints g_est s_est est_dif g_std s_std
std_dif;
run;
```