Addis Ababa University

College of Natural Sciences

Office of Graduate Studies

Department of Statistics

Application of Spatial Mixed Model in Agricultural Field Experiment

By: Dibaba Bayisa

Advisor: Girma Taye (PhD)

June 2010

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

OFFICE OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

Application of Spatial Mixed Model in Agricultural Field Experiment

By

Dibaba Bayisa

A Thesis submitted to the Office of Graduate Programs of Addis Ababa University in partial fulfillment of the requirement for the Degree of Masters of Science in Statistics (Biostatistics)

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

OFFICE OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

Application of Spatial Mixed Model in Agricultural Field Experiment

By

Dibaba Bayisa

Approved by the Board of Examiners:

_____	_____
Department Head	Signature
M. K. Sharma	M. K. Sh
Examiner	Signature
Prof. Eshetu Wencheko	
Examiner	Signature

ACKNOWLEDGEMENT

First of all thanks to the almighty GOD for his indescribable help in all aspect. This paper has been accomplished with the help of many people to whom I am indebted. I express my deepest appreciation to my advisor and instructor Girma Taye (Ph.D), who introduced me to research in spatial model and for his generous contribution to the accomplishment of this thesis work and for his consistent and stimulating advice during the whole course of this study. The discussion that I had with him stimulated my intension of writing on this topic decisively, shaping my thought and building my confidence apart from making this thesis worthwhile. Generally, his material and moral supports were invaluable.

I am highly indebted to the staff members of Department of Statistics, Addis Ababa University for giving me an opportunity to pursue my study and their kind assistance in many ways.

My warm thanks goes to my father Bayisa Gemechu, my mother Elisabeth Nigatu, my sisters Dibabe Bayisa, and Dureti Bayisa, my brothers Dagaga Bayisa and Bikila Yohanis and the entire family and relatives specially my uncles Esayas Nigatu and Dereje Nigatu and his wife Almaz Niguse, whose endless encouragement, moral and financial support were the sources of inspiration to me, during my entire academic career, including the period I was pursuing the M.Sc. degree.

Special thanks go to my colleague Merga Belina for helping me in editing this paper and giving me different comments and constructive ideas while working this thesis.

I am very much beholden to my best friends Dabala Chemeda, Mestawot Kitila, Eyasu Tesfaye, Dereje Waktola, Dejene Kebede, Hawani Eshetu and all others, for their incessant support and encouragement throughout my MSc study.

My classmates Bacha Edosa, Muluneh Sorissa, Yabebal Ayelaw, Birhan Fetene, Nibret Alene, and others, also deserve my special appreciation for their comfortable and hospitable attitude throughout the academic period.

TABLE OF CONTENTS

Contents **Page**

LISTS OF TABLES

LISTS OF FIGURES

ABSTRACT

The objective of this study was to evaluate the efficiency of spatial statistical analysis in field trials and, particularly, we have demonstrated the benefits of the approach when experimental observations are spatially dependent. We have compared (i) classical randomized complete block model with independent and identically distributed errors (RCBDiid); (ii) the most common spatial models: Exponential, Spherical and Gaussian (with and without nugget effect); (iii) spatial models of (ii) with and without block effects; (iv) complete random model (CR). The data used in this study were obtained from separate trials for Bread wheat and Durum wheat which was coded as BW99RVTII (Regional Varity Trial Two of Bread Wheat conducted in the year 1999 Eth.C.) and DW99RVTII (Regional Varity Trial Two of Durum Wheat conducted in the year 1999 Eth.C.), respectively. The restricted maximum likelihood (REML) method was used for estimation of spatial covariance parameters. The denominator degree of freedom of F test for treatment effects were computed using the Kenward-Roger method for all models (Kenward and Roger, 1997). Semivariogram were used for the initial estimate of the parameters of the spatial covariance structure. Akaike's Information criterion (AIC) and Corrected Akaike's Information criterion AICc and the Likelihood ratio test were used for model comparison. The result showed that in all of the four trials, the estimated residual from RCBiid model were significantly spatially correlated, providing evidence of field heterogeneity within blocks that could make the RCBiid method less powerful than a method that incorporated the spatial correlation. Furthermore, the finally selected spatial model for all data set showed that blocking seemed to be unnecessary if these selected spatial model were used for each of the trials. The results also showed that the spatial models provided a smaller p-value and standard error than the classical analysis of variance models. We have concluded that randomization and blocking does not completely remove spatial variation. Hence, we recommended that researchers should take into account spatial variation in field experiment in situations were spatial dependence among the observations is significant.

Key words: Spatial Models; Semivariogram

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the study

Agricultural field experiments are used to identify high yield crop varieties, treatments with high performance or improved technologies for recommendation to farmers. These experiments are conducted in plots of various sizes, for which the extent of variability in the soil is not known a priori. Field experiments in agronomy and related disciplines have traditionally been affected by soil heterogeneity. This is especially of concern when treatment effects are small and soil variability is high, as this inflates the error term. Intrinsic soil variability is the result of the geological, hydrological, and biological mapped suggests that areas can be identified that are relative uniform, but more recent research suggests that soils generally constitute a continuum with variability at different scales (Van Es, 2002).

The structure of soil variability has important implications for the design of experiments. Most agronomic field experiments are based on the concepts of replication, local control (Blocking) and randomization (Atkinson and Bailey, 2001).

Replication allows for estimation of the experimental error by applying treatments to different plots under the same experimental conditions. Sufficient replication needed to distinguish treatment effects from background variability. Blocking is used in field experiments to control the adverse effects of soil heterogeneity. The use of randomization has been justified in many ways. Its basic purpose is to remove bias from the estimation of treatment effects (Atkinson and Bailey, 2001), and to equalize the error over all treatment differences (Yates, 1939; Fagroud and van Meirvenne, 2002).

Randomization is often considered the best protection and assurance against malicious manipulation of plot layout. Randomization is also believed to better justify the assumption of normal errors. A concern with randomization is the possibility of undesirable out comes such as treatments being repeatedly located in the same location in different blocks, and treatment pairs being repeatedly located in adjacent positions. This poses no concern when variability is truly random and stationary, but agricultural scientists often admit to minor adjustments to randomized designs when treatment allocations appear undesirable. The quality of result from

such experiments, however, depends on the strength of statistical methods used in accounting for field variability.

The best strategy of increasing production of crops is by increasing productivity per unit area using improved production technology. On the other hand, reliable improved technology can only be achieved if proper designing and modeling is done by accounting for spatial variability. Spatial variability occurs when observations that are close together tend to be more alike than those that are farther apart. It can result from physical characteristics such as field topography or temporal or temperature gradient in a greenhouse, or from process such as the spread of an infection. Failure to deal meaningfully with spatial variability greatly increases the risk of modeling or erroneous inference. A search for a method that account for field variability has been going on since the early days. A review methods developed over time to account for field variability are presented to recognize the contributions, identify the gaps and indicate future direction.

Selection of high yielding and disease resistant cultivars is a typical activity in variety trial programs. Crop breeding programs run at different stages, often from nursery to national variety trial with the following major objectives:

1. To generate new cultivars which are superior to the standard ones
2. To explore and adopt new technologies to increase food production
3. To study the effect of different factors of production on yield and quality and to identify the best combination of these factors which can optimize the yielding ability of crops
4. To generate information about basic understanding of the factors of production to be used by farmers and researchers
5. To develop recommendations regarding adaptability of selected varieties in different agro-ecologies

To fulfill these objectives research needs to cover several crops grown in a range of agro-ecologies each under different conditions. In the light of scarce resources and high field variability, therefore, effort should be exerted to achieve the above goals through improved design and modeling approaches that account for spatial variation.

The selection of experimental material and analysis techniques should be carefully considered in order to achieve optimal results. In crop improvement programs the major interest is in detecting differences among varieties of treatments. This is again governed by the amount of variability available in the field experiments and by the method of modeling employed. To optimize the use of resources, the best method of describing and accounting for field variability should be used so that variety effects and difference between them are measured as precisely as possible. This requires improvement in both design and modeling.

In Ethiopia, with a population of nearly 70 million and a growth rate of 3.3%, agriculture is the backbone of the economy employing 85% of the population. Agriculture accounts for 49% of the gross domestic product and 90% of the national export earnings. Total crop area has increased from 7,689,400 ha in 1994 to 9,035,700 ha in 1995 and then dropped to 8,020, 900 ha in 1996. This expansion of crop area has been achieved at the expense of natural forest and grazing land and this affects the environment and production of livestock. Unfortunately, there will not be an increase in crops' area in the future due to the limited area of arable land. Therefore, the best strategy of increasing production of crops is by increasing productivity per unit area using improved production technology. On the other hand, reliable improved technologies can only be achieved if proper designing and modeling is done by accounting for spatial variation. Therefore, the main concern of this study is to fit different spatial models (spatial linear fixed effect model and spatial linear mixed model) and compare them with conventional or traditional models such as linear fixed effect model and linear mixed effect model for four different data sets obtained from separate trials for Bread wheat and Durum wheat which was coded as BW99RVTII and DW99RVTII which were obtained from Ethiopian Agricultural Research Organization (EARO). Specifically, the aim of this study is to fit and analyze each data set using the following models: (i) RCBDiid; (ii) the most common spatial models: Exponential, Spherical and Gaussian (with and without nugget effect); (iii) spatial models of (ii) with and without block effects; (iv) complete random model (CR).

1.2. Statement of the problem

The best strategy of increasing production of crops is by increasing productivity per unit area using improved production technology. On the other hand, reliable improved technologies can only be achieved if proper designing and modeling is done by accounting for spatial variation. In

case of our country, Ethiopia, to our knowledge, little effort has been taken to use spatial model in field experiments. Since spatial model is also a new technique, it many researchers in our country Ethiopia are not familiar with it. Hence, the statements of the problems of this study are:

- ✓ Why is it needed to model and study the spatial model in the field trials?
- ✓ How do we include spatial dependence in the model?
- ✓ What is the implication of observing significant spatial dependence in field experiment?
- ✓ Is that really fitting spatial model is beneficiary for the researcher?
- ✓ What is the best variogram model for the data from field experiment?
- ✓ How do we interpret the estimated parameters of the variogram model?
- ✓ How do we check the significance of the fitted spatial model?

1.3. General objectives of the study

The general objective of this study is to fit the spatial model (Spatial linear mixed model and spatial linear fixed effect model) to the data under consideration and compare them with conventional or traditional linear model (linear mixed model and linear fixed effect model).

The specific objectives are:

- ✓ To compare the following models: (i) RCBDiid; (ii) the most common spatial models: Exponential, Spherical and Gaussian (with and without nugget effect); (iii) spatial models of (ii) with and without block effects; (iv) complete random model (CR).
- ✓ To examine the effect of spatial model on p values of the fixed effect,
- ✓ To test the hypothesis about treatment effect based on the selected model.
- ✓ To evaluate parameter estimates in the different models.
- ✓ To recommend efficient model to be used by researchers.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Introduction

The main objective of variety trials is to obtain precise yield estimates of variety means and variety differences that affect performance under farm conditions. Modeling approaches require fitting of several models and subsequently choosing of the most suitable model for a data set. However, eliminating spatial trend and pattern physically through proper planning of experiments is efficient, if accomplished successfully. The randomization approach in complete block design is attractive though it incurs bias and imprecision under field situation. This is because the soil characteristics are typically non-random and show fertility trend, spatial autocorrelation or periodicity. In the same way that spatial modeling is getting popular, robust designs which utilize spatial information are now common. Spatial variation in fertility, moisture, intercepted light, and other environmental factors can bias variety contrasts and inflate residual variation. Therefore, it is important to control, either by design or by analysis, the residual variation not accounted for by variety effects.

Fisher (1926) designed field experiments to account for spatial structure by randomized complete block design. He was seeking ways of improving field experiments in agriculture. Blocking has been introduced to consider the mixed effects of inherent spatial dependence of the environmental variables and the spatial dependence of the response variable to it (Underwood, 1997). Block designs (including complete and incomplete blocks and lattices) attempt an "a priori" reduction of residual variation by considering spatial heterogeneity. The experimenter however is faced with the problem of constructing blocks that are sufficiently homogeneous in the "expected" response of the varieties. Block size, shape, and orientation can seldom be determined by studying the spatial variation of the response variable. Management considerations frequently determine the experimental layout of a variety trial. As a result, most field experiments are laid out on a regular grid of, say, r rows and c columns and with groups of contiguous plots forming a rectangular block. Analysis of uniformity trials have shown that this regular spatial arrangement of blocks seldom provides efficient control of spatial variation (Mendez, 1970; Lin *et al.*, 1993).

The classical approach based on the randomization theory neutralizes the spatial correlation but is usually less efficient than spatial models. Simulation studies and randomization studies of uniformity trials have shown that spatial approaches usually estimate variety contrasts more precisely than the classical analysis based on uncorrelated observations (Grondona, *et al.,* 1996).

2.2. Development of Spatial model

The basis for the validity of classical designs is randomization (Fisher, 1926). Papadakis responded to Fisher's blocking methodology by suggesting a nearest-neighbor approach (NN). Papadakis introduced adjustment procedures to compensate for soil variation and advocated the case of covariate adjustments for the neighbouring plots which was later evaluated by Bartlett (Bartlett, 1938; Papadakis, 1937). According to Bartlett, the method of adjustment suggested Papadakis was approximately valid and sometimes useful in experiments where the number of plots per block is large. However, Bartlett was critical of the fact that a covariate was generated from the same set of data which represented the observed values. Williams (1952) proposed a technique which used serial correlations in one-dimension to obtain efficient estimate of treatment effects and their differences. In 1969 Atkinson compared the property of the one-dimensional Papadakis technique with that of maximum likelihood estimation assuming the autocorrelation model used by Williams. Both procedures were found to be similar in nature, but the Papadakis covariate was preferred because of its simplicity. In 1978 Bartlett brought forth Papadakis' nearest neighbor approach for field trials. Nearest neighbor approaches have been compared to standard blocking methods by Stroup, Tamura and their colleagues' (Stroup et al., 1994; Tamura et al., 1988) provide another alternative to Fisher's blocking scheme by inserting a polynomial trend variable (T_{ij}) into the familiar ANOVA model $Y_{ij} = \mu + T_{ij} + \varepsilon_{ij}$. Brownie *et al.,* (1993) also attempted to compare the NN approach to the polynomial trend regression suggested by Tamura.

Gleeson and Cullis (1987) proposed to sequentially fit a class of autoregressive-integrated-moving average models (ARIMA) to the plot errors in one direction (rows or columns). They also extended the previous model to two directions (rows and columns) assuming that, in the field, rows and columns are regularly spaced. However, differencing in this two dimensional analysis was prone to discard treatment information (Kempton et al., 1994).

Grondona *et al.* (1996). used two dimensional spatial analysis procedure based on separable ARIMA processes, proposed by Cullis and Gleeson to analyze 35 Cereal yield trials with incomplete block designs.

Zimmerman and Harville (1991) proposed a random field approach that can account for spatially correlated errors (CE) among spatially geo referenced observations, and thus may provide a more efficient estimation of treatment comparisons.

Basu and Reinsel (1994) considered regression models for two-dimensional spatial data when the errors follow a spatial unilateral first-order autoregressive moving average (ARMA) model in their study. The authors give details on the convenient computation of the generalized least squares (GLS) estimator of the regression parameters in the presence of spatially correlated errors, and compare the GLS estimator to the ordinary least squares (OLS) estimator in some special cases. They also considered the restricted maximum likelihood estimator of the spatial correlation model parameters, which may be preferred over the maximum likelihood estimators. For the special case of the spatial unilateral first-order AR model, the authors also gave details of the maximum likelihood as well as the restricted maximum likelihood estimation.

Brownie and Gumpertz (1997) investigated the validity and robustness of CE analysis and concluded that the method could achieve substantial increases in precision compared with classical RCBiid analysis when errors are spatially correlated while still maintaining type I error rates close to the desired significance level.

Van Es and Van Es (1993) evaluated the spatial nature of randomized arrangement of plots in RCB designs, and determined its effect on the outcome of experiments. The authors have demonstrated that when spatial dependence occurs, the statistical tests associated with contrasting treatments in plots nearer together have higher probabilities of type II error, which consists of different treatments appearing to be identical.

Kinfemikael Alemu (2000) applied spatial statistical methods to test spatial autocorrelation among 'neighbouring' plots over the experimental farm and compare 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 fertility pattern.

Duarte and Vencovsky (2005) evaluated the efficiency of spatial statistical analysis in the selection of genotypes in a plant breeding program and, particularly, they demonstrate the benefits of the approach when experimental observations are not spatially independent by making use of a yield trial of soybean lines with five check varieties (of fixed effect) and 110 test lines (of random effects), in an augmented block design. Their results showed that, a residual autocorrelation of significant magnitude and extension (range), which allowed a better discrimination among genotypes (increase of the power of statistical tests, reduction in the standard errors of estimates and predictors, and a greater amplitude of predictor values) when the spatial analysis was applied. Furthermore, the spatial analysis led to a different ranking of the genetic materials, in comparison with the non-spatial analysis, and a selection less influenced by local variation effects was obtained.

Stroup (2002) in his study used mixed model methods to assess power and precision of different designs in the presence of spatial variability and to compare competing design and analysis strategies. He used Satterthwaite's method and Kenward-Roger methods for approximating degrees of freedom.

Werneck and Maguire (2002) demonstrated a method for modeling spatial autocorrelation within a mixed model framework using data on environmental and socioeconomic determinants of the incidence of visceral leishmaniasis (VL) in the city of Teresina, Piauí, Brazil. The authors result indicates that, a model with a spherical covariance structure indicated significant spatial autocorrelation in the data and yielded a better fit than one assuming independent observations. Their findings support the hypothesis of spatial dependence of VL rates and indicate that it might be useful to model spatial correlation in order to obtain more accurate point and standard error estimates.

The recently developed REML approach (Cressie, 1993; Stroup *et al.*, 1994; Schabenberger and Pierce, 2002) include estimating empirical semivariograms, and then using these parameter estimates as priors in a regression model to identify spatial error covariance structures. Lambert *et al.* (2002) used the REML-geostatistics approach to analyze yield monitor data. Patterson and Thompson (1971) indicated that the score equation for the variance components could be solved iteratively using the Fisher scoring (FS) algorithm. For many applications this method presents computational difficulties due to the large size of matrices to be inverted and multiplied

(Gilmour *et al.*, 1995). Thompson (1977) presented an overview of the methodology with particular reference to animal breeding applications and showed how some of the computational burdens of FS can be overcome and suggested that the use of observed information rather than the expected one is sometimes preferable. Gilmour *et al.* (1995) proposed average information REML (AIREML) as an alternative to the FS REML method. They noted that although the AIREML algorithm is the same as the one proposed by Thompson in 1977, it avoids evaluation of the traces of large matrices which appear in both observed and expected (REML) information matrices. Unlike Thompson and others who proposed similar solutions for computational problems, the AIREML uses the average information to simplify computations of matrices. Gilmour and his colleagues illustrated the convergence performance of the AI algorithm relative to FS algorithms using a number of real data set. The Genstat engine uses REML or average score (ASREML) algorithms and both FS and AI algorithms are available as an option. The algorithm can be used for balanced as well as unbalanced data as models are fitted like regressions. REML (hence AIREML) algorithm enjoys a variety of applications and is particularly useful when one has to extract information on variance components from different levels or strata of population. It has the property of providing efficient estimates of treatment effects in unbalanced designs with more than one source of error.

Nan Hong *et al.* (2005) noted that failure to account for spatially correlated errors when present in the classical randomized complete block (RCB) analysis may cause inefficient estimation of treatment significance. They have discussed methods for selecting a covariance model in RCB analyses in the presence of spatial correlation. They have suggested the use of three types of statistical tools for Model Selection (semivariogram of the residuals from fitting a model with just fixed effects, the likelihood ratio test, and Akaike Information Criterion for model selection). To illustrate the procedure, the authors have analyzed winter wheat (Triticum aestivum L.) forage and corn (Zea mays L.) grain yield in the presence of spatial heterogeneity within blocks from a site-specific N management study and compared the selected covariance models to the RCBiid models and to other spatial models with respect to the estimation of treatment significance. The authors also noted that, the procedure can be extended to any experiment with fixed effects, or with both fixed and random effects, and which may potentially have spatially correlated errors. They have warned that their procedure is systematic and readily implemented; however, it remains difficult to evaluate whether an adequate covariance model has been

selected. Girma T. (2005) also discussed various methods of spatial data analysis, methods of fitting variogram models (prediction) and methods of modeling spatial variation.

Gonçalves *et al.* (2007) described a strategy to improve the precision of statistical data analysis of grapevine selection trials through the use of mixed spatial models. The authors compared the efficiency of mixed spatial models with that of a classical randomized complete block model (with independent and identically distributed errors). The comparisons were based on yield data from three large experimental populations of clones of the Arinto, Aragonez (Tempranillo) and Viosinho grapevine varieties. The fit of the spatial mixed models applied to yield data was significantly better than that of the classical approach, resulting in a positive impact on selection decisions and increasing the accuracy of genetic gain prediction.

Hoeting *et al.* (2006) considered the problem of model selection for geostatistical data. They considered AIC as applied to a geostatistical model and provided simulation results that shows using AIC for a geostatistical model is superior to the standard approach of ignoring spatial correlation in the selection of explanatory variables.

Werneck and Maguire (2002) demonstrated a method for modeling spatial autocorrelation within a mixed model framework, using data on environmental and socioeconomic determinants of the incidence of visceral leishmaniasis (VL) in the city of Teresina, Piauí, Brazil. A model with a spherical covariance structure indicated significant spatial autocorrelation in the data and yielded a better fit than the one assuming independent observations. The authors compared a model with spherical covariance and the independent model (that ignored the presence of spatial correlation), using the Likelihood Ratio Test (based on -2REML Log Likelihoods). The result of their finding indicates that, the model with a spherical covariance structure fit the data better than the independence model ($p < 0.0001$, Likelihood Ratio Test).

Xiyuan and Joachmim (2009) investigated the consequences of mixed linear models with various spatial covariance structures on line statistical tests and selections of plant cultivar trials by using the Satterthwaite and Kenward-Roger methods for approximating degrees of freedom. They have investigated different covariance structures such as: exponential, spherical, Gaussian, linear, linear-log, anisotropic power and anisotropic exponential, as well as first order autoregressive. They also found out that the spatial models provided a smaller standard error than the classical

analysis of variance models. They have also noted that only the optimally fitting spatial model could control spatial variability more effectively with less loss of error degrees of freedom and a higher power for line tests than the classical analysis of variance of block designs. They also showed that the Kenward-Roger method effectively provide a correction of the standard error for line contrasts and a reasonable power of spatial model analysis. They concluded that, selecting the covariance structures and using the Kenward-Roger method for approximating degrees of freedom in the spatial analysis of cultivar field trials using mixed linear models is necessary.

CHAPTER THREE

3. DATA AND METHODOLOGY

3.1. Data

The data used in this study were obtained from separate trials for Bread wheat and Durum wheat which were coded as BW99RVTII (Regional Varity Trial Two of Bread Wheat conducted in the year 1999 Eth.C.) and DW99RVTII (Regional Varity Trial Two of Durum Wheat conducted in the year 1999 Eth.C.), respectively. Three data sets were taken from the first type of wheat (BW99RVTII) which was grown at different locations (Location: Four, Five and Six) in the same year. In this case for the fourth location the block effects were assumed to be random and for the last two locations (Five and Six) the block effect were assumed to be fixed. Additionally one data was taken from the second type of Wheat (DW99RVTII) which is from location one, according to the code of EARO. In this case the block effects were assumed to be random. Thus, for this study we have considered three data sets of Bread wheat yield and one data of Durum wheat yield. For simplicity, the four data sets were coded as BW99RVTII4, BW99RVTII5, BW99RVTII6 and DW99RVTII1 for Bread wheat data from location four, Bread wheat data from location five, Bread wheat data from location six and Durum wheat data from location one, respectively. The first three trials (BW99RVTII4, BW99RVTII5 and BW99RVTII6) were carried out in a randomized complete block design (RCBD) with four replications (blocks) and 20 bread wheat varieties. Each block consisted of twenty 2.5 X 1.2 meter (m) plots with the spacing between plot and block equal to 0.40m and 1.00m, respectively. The last trial (DW99RVTII1) was also carried out in a randomized complete block design (RCBD) with four replications (blocks) but with nineteen Durum wheat varieties. Each block consisted of 19 2.50m X 1.20m plots with the spacing between plot and block equal to 0.40m and 1.00m, respectively *(See Appendix I)*. Each data set from the trials was analyzed using the following models: (i) RCBDiid; (ii) the most common spatial models: Exponential, Spherical and Gaussian (with and without nugget effect); (iii) spatial models of (ii) with and without block effects; (iv) complete random model (CR). For the last two data sets the block effects were taken as random. All the analyses were conducted using SAS software of Version 9.2. Randomized Complete Block Design (RCBD) is one in which the number of experimental units per block is equal to the

number of treatments and every treatment occurs once and only once in each block, the order of treatments within a block being randomized.

3.2. Methodology

3.2.1 The Classical linear models

Linear models are used to describe how a variable of interest (outcome) depends on other variables (effects). Model components used in agricultural field experiments include either fixed or both fixed effect and random effect.

3.2.1.1. Fixed effect linear model

Fixed effects are effects which are attributable to a limited set of levels of a factor which are there because we are interested in them. A model that contains only fixed effects and no random effects is known as a fixed effects model.

Fixed effects assumptions

- ✓ Includes the entire population of studies to be considered; do not want to generalize to other studies not included (including future studies).
- ✓ All of the variability between effect sizes is due to sampling error alone. Thus, the effect sizes are only weighted by the within-study variance.
- ✓ The collected data represent random samples from the same population
- ✓ Effect sizes are independent

In this study the data that were used are from the trials which are carried out in a randomized complete block design. Hence, the suitable fixed effect RCBDiid model for the data under study is given by

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, \ldots\ldots, a \\ j = 1, 2, \ldots\ldots, b \end{cases} \quad [1]$$

where y_{ij} the yield of i^{th} variety in the j^{th} block, a is the total number of variety, b is the total number of block, μ is an over all mean, τ_i is the effect of the i^{th} variety, β_j is the effect of j^{th} and block, ε_{ij} is independently and identically distributed random error term with mean zero and variance σ^2. Here both treatment and block effects were assumed to be fixed.

In matrix form, the general linear fixed effect model can be expressed as:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \qquad [2]$$

where **Y** stands for an $nx1$ vector of observations, **X** is an nxp matrix of constants associated with the fixed effects of the design, $\boldsymbol{\beta}$ is a $px1$ vector of unknown fixed effects (p is a number of unknown parameter plus 1), and $\boldsymbol{\varepsilon}$ is a vector of random residual errors with assumed distribution $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \boldsymbol{R})$.

Method of parameter estimation

Under the assumption that $\boldsymbol{R} = \sigma^2\mathbf{I_n}$, the OLS estimator is BLUE for $\boldsymbol{\beta}$. Thus, given $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \sigma^2\mathbf{I_n})$, where $\mathbf{I_n}$ denotes the identity matrix of dimension n and σ^2 represents the common error variance. The OLS estimate of $\boldsymbol{\beta}$ is obtained by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}, \qquad [3]$$

The relationship represented by equation [1] is assumed to be universal or constant across the geographic area.

3.2.1.2. Random effect linear model

In a random effect model factor levels are randomly selected from the set of all possible factor levels. The model includes only random effects and no fixed effects are termed a random effects model.

Random effects assumptions

- ✓ It contains only a sample from the entire population of studies to be considered. As a result, we do want to generalize to other studies not included in the sample (e.g., future studies).
- ✓ Variability between effect sizes is due to sampling error plus variability in the population of effects.
- ✓ In contrast to fixed effects models, there are two sources of variance.
- ✓ The study consists of random samples of some population in which the underlying (infinite-sample) effect sizes have a distribution rather than having a single value.

- ✓ Effect sizes are independent.

3.2.1.3. Linear Mixed Model

A model that contains both fixed effects and random effects is known as a mixed effects model. The mixed model equation can be regarded as a method of dealing with a complicated set of variances and covariances which we know affect the data. The suitable mixed effect RCBDiid model for the data under study is given by

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, \ldots\ldots, a \\ j = 1, 2, \ldots\ldots, b \end{cases} \quad [4]$$

where y_{ij} the yield of i^{th} variety in the j^{th} block, a is the total number of variety, b is the total number of block, μ is an over all mean, τ_i is the effect of the i^{th} *variety*, β_j is the effect of j^{th} random block which is normally and independently distributed with mean zero and variance of σ_b^2, and ε_{ij} is independently and identically distributed random error term with mean zero and variance σ^2. In matrix form, the linear mixed model can be expressed as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}, \quad [5]$$

where **Y**, $\boldsymbol{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are as defined in equation [2], **Z** is an $n \times q$ matrix of constants associated with the random effects, and $\boldsymbol{u}$ is a $q \times 1$ vector of unknown random effects parameters. It is assumed that:

- ✓ $E(\boldsymbol{u}) = \mathbf{0}$ and $Var(\boldsymbol{u}) = \boldsymbol{G}$
- ✓ $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\varepsilon}) = \boldsymbol{R}$,
- ✓ $Cov(\boldsymbol{u}, \boldsymbol{\varepsilon}) = \mathbf{0}$, and
- ✓ Both $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ are normally distributed.

The variance of $\boldsymbol{Y}$ is

$$\boldsymbol{V} = \boldsymbol{ZGZ'} + \boldsymbol{R}, \quad [6]$$

and can be estimated by setting up the random-effects design matrix **Z** and by specifying covariance structures for **G** (qxq) covariance matrix of $\boldsymbol{u}$ and **R** (nxn) covariance matrix of $\boldsymbol{\varepsilon}$. (Little *et al.* 1996).

Method of Parameter Estimation

The ordinary maximum likelihood estimator of variance component takes no account of the degrees of freedom used in estimating fixed effects and hence is biased. This introduces downward bias which increases as the number of fixed effects in the model increases. Therefore, it is more intuitive to use variance components which take account of the degrees of freedom for estimating treatment effects. Hence, we will use restricted maximum likelihood (REML) estimation method in order to estimate the variance components. REML is the processes of applying ML to the linearly transformed response data vector. The transformation is done in such a way that linearly transformed data vector contains none of the fixed effects.

Restricted Maximum Likelihood (REML), developed by Patterson and Thompson (1971), which in contrast to ML accounts for the loss in the degrees of freedom due to fitting fixed effects. To account for the loss in the degrees of freedom due to fitting of fixed effects, REML, in contrast to ML, maximizes only part of the likelihood of the data vector $\mathbf{Y}$ which is independent of the fixed effects. That is, rather than using $\mathbf{Y}$ (the data vector) directly, REML is based on the linear combinations of elements of $\mathbf{Y}$, chosen in such a way that those combinations don't contain any fixed effects, no matter what their value.

Now suppose that we have an $(n-p)\ x\ n$ matrix $\mathbf{K}$ of rank $n-p$ such that

$$\mathbf{K}'\mathbf{X} = \mathbf{0}, \qquad [7]$$

Then

$$\mathbf{K}'\mathbf{Y} = \mathbf{K}'\mathbf{X}\boldsymbol{\beta} + \mathbf{K}'\boldsymbol{\varepsilon} = \mathbf{K}\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{K}'\mathbf{V}\mathbf{K}), \qquad [8]$$

The distribution of $\mathbf{K}'\mathbf{Y}$ doesn't have any $\boldsymbol{\beta}'s$ in it. To estimate the parameters in $\mathbf{V}$, we use $\mathbf{K}'\mathbf{Y}$.

The REML Equations

With $\mathbf{Y} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ we have, for $\mathbf{K}'\mathbf{X} = \mathbf{0}$, $\mathbf{K}'\mathbf{Y} \sim \mathbf{N}(\mathbf{0}, \mathbf{K}'\mathbf{V}\mathbf{K})$. Then the logarithm of the likelihood function of the REML profile is

$$l_p = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{P}'\mathbf{V}^{-1}\mathbf{P} - \frac{n-p}{2}\log(2\pi) \quad [9]$$

where

$$\mathbf{P} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} \quad [10]$$

The next step is differentiating l_p with respect to **V**. Let φ be the arbitrary element of **V** which needs to be estimated. Mathematically,

$$\frac{\partial l_p}{\partial \varphi} = 0, \qquad \forall\, \varphi \epsilon\, \mathbf{V} \quad [11]$$

Iterative method must be used to find estimates for φ. The Newton-Raphson optimization method which requires first and second derivatives were used. Hence, the parameters $\boldsymbol{\beta}$ and **u** were estimated using iterated procedure (Newton – Raphson) with the following set of model equations (Henderson, 1984):

The following procedure consisted of resolving the mixed model equations (Henderson, 1984):

$$\begin{bmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'\hat{R}^{-1}\mathbf{Y} \\ Z'\hat{R}^{-1}\mathbf{Y} \end{bmatrix} \quad [12]$$

where $\hat{R}$ and $\hat{\mathbf{G}}$ are the estimate of **R** and **G**, respectively. Hence the estimates of fixed effects and the Best Linear Unbiased Prediction (BLUP) are, respectively

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y \quad [13]$$

and $$\tilde{u} = \hat{\mathbf{G}}Z'\hat{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta}) \quad [14]$$

3.2.2 The Spatial models

A spatial experiment is a comparative experiment in which the experimental units occupy fixed locations distributed throughout a region in *d* dimensional space. A feature common of most spatial experiments is the presence of systematic heterogeneity among the experimental units. Typically, the nature of this heterogeneity is such that there is appreciable correlation among neighboring units. The analysis of variance associated with the classical randomized designs ignores this correlation, relying on the randomization to neutralize any disruption that the

correlation may have on the estimation of treatment contrasts. In deed, the neutralization of spatial correlation was one of the early arguments used to support experimental randomization (Yates, 1938). Recently, however, there has been much interest in methods for analyzing spatial experiments that take spatial correlation into account (Bartlett, 1978; Kempton and Howes, 1981; Wilkinson et al., 1983; Green et al., 1985; Besag and Kempton, 1986; Williams, 1986a; Gleeson and Cullis, 1987; Curriero and Lele, 1999; Zimmerman and Harville, 1991; Stroup, 2002; Duarte and Vencovsky, 2005). Spatial Statistics differs 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 feature that distinguishes spatial statistics from classical statistics is that, 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, (Jayet *et al.,* 2004). Observations are correlated strictly due to their relative location position (referred to us 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 georeferenced data that remains dormant in statistical analysis of traditional data, composed of independent observations.

3.2.2.1. Types of spatial Data

The following discussions about spatial data are taken from Cressie (1991) book and lecture notes on spatial model (Fuentes, 2008; Murali, 2010).

Spatial data may come in various ways, 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 spatial statistical theory practitioners have recently gained powerful new ability to obtain, organize, and study complex relationships between farming and its natural recourse base, like soil fertility pattern and topography of the agricultural field trial. Application of this new capability 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 lead to numerous questions relating to the interpretation, manipulation, analysis of spatial data. There are three broad categories of spatial data: Point pattern, lattice data, and Geo-statistics.

I. **Geostatistics:** Geostatistics began with mining type applications and the prefix "geo" indicates that geostatistics dealt initially with statistics pertaining to the earth. Geostatistical data are measurements collected at discrete sample locations separated by some distance. Studies that rely on sample quadrates distributed (randomly or however) across a study area generate this kind of data. While the location of the sample points is known and of interest, it is the values of the measurement taken at each point that are of primary concern. Geostatistical data involve continuous variations of feature attribute. Here, the basic question is, "Are samples that are close together also similar with respect to the measured variable?" By extension, the scaling issue is to identify the distance(s) over which values tend to be similar. Geostatistics has applications in climatology, environmental monitoring, and geology. Geostatistical applications have expanded original mining applications to include modeling soil properties, ground water studies, rainfall precipitation, public health, etc. Most of the literatures (Curriero and Lele, 1999; Zimmerman and Harville, 1991; Stroup, 2002; Duarte and Vencovsky, 2005) assumes that the data from agricultural field experiments are geostatistics. So, the data we have considered in this thesis are categorized as geostatistics. Therefore, all the methods and assumptions that are studied in this thesis were limited to this category of spatial data.

II. **Lattice Data:** Nontrivial observations are taken at a finite number of sites whose whole constitutes the entire study region. Unlike geostatistics, there is no possibility of a response "between" data locations. Data locations are regions. Lattice data represent a region of given area, and can be intuitively defined by the notion of "neighbors". The scaling question with lattice data tends to focus on the question, "At what scale(s) do adjoining regions tend to have similar values?" Lattice type data provide the closest analogue to time series data. In time series data sets, observations are typically obtained at equally spaced time points. For lattice data, spatial data is obtained over a regularly spaced set of points (irregular lattice type data is a possibility as well). Lattice data represent discrete variation in space based on regular or irregular units such as disease prevalence per farm. Often this type of data are based on information aggregated within areas defined by administrative boundaries. Spatial dependence in this type of data is quantified by producing a contiguity or spatial weight matrix. This matrix is the basis for analysis of spatial correlation and multivariate spatial

correlation as well as for spatial principal components and regression analysis incorporating spatial effects

III. **Point Patterns:** Data analysis of point patterns corresponds to studies where interest lies in where events of interest occur. A fundamental question of interest in this context is whether or not the points of interest are occurring at random, or do the points cluster in some manner, or perhaps the points of interest are occurring with some sort of regularity. A familiar example is a map of all trees in a forest stand, where in the data consists of a list of trees referenced by their (x, y) locations.

3.2.2.2. The general spatial models

Suppose that the data $\{z(s_1), \ldots, z(s_n)\}$ observed at known spatial locations $\{s_1, \ldots, s_n\}$ are modeled as a realization of random process $\{Z(s): s \in D \subset R^d\}$. Here, the spatial index set **D** could have positive Lebesgue measure, be a countable set of points of R^d, or be a finite subset of R^d.

$$\{Z(s): s \in D \subset R^d\} \qquad [15]$$

Sometimes [16] is written as $\{Z(s; \omega): s \in D;\ \omega \in \Omega\}$, where $(\Omega, \mathcal{F}, P)$ is a probability space; the realization $\{Z(s): s \in D\}$ would correspond to a particular value of ω, say $\omega = \omega_0$.

3.2.2.2.1. Test of spatial independence

The application of a geostatistical model to the analysis of the dataset is justified if the observations are spatially related. In order to check whether there is a spatial dependence within a given data, we have used a semivariogram. Thus, if the semivariogram is not constant this implies that there is spatial dependence in a given data (Cressie, 1991).

3.2.2.2.2. Test for the assumption of Stationarity of the spatial process

The first step in the analysis of spatial model is checking for the assumption of Stationarity and Isotropy. In most of the spatial analysis the assumption of stationarity might not be fulfilled for that matter only we will check for the second order stationarity assumption.

Test for stationarity

In statistics it is common to assume that a random variable is stationary, i.e. its distribution is invariant under translation. In the same way, a stationary random function is homogeneous, and its distribution is invariant under translation. In its strictest sense stationarity requires all the moments to be invariant under translation but since this cannot be verified from the limited experimental data, we usually require that only the first two moments (the mean and the covariance, which is sufficient for linear estimations) to be constant. This is called "Weak" or second order stationarity. In other words we require:

i. The expected value (or mean) of the function Z(**s**) is constant for all points **s**. That is, $E(Z(\boldsymbol{s})) = \mu(\boldsymbol{s}) = \mu$ which is independent of s.

ii. The covariance function between any two points **s** and **s** + **h** is independent of the point **s**. It depends only on the vector **h**. That is,

$$E[Z(\boldsymbol{s}) - \mu][Z(\boldsymbol{s}+\boldsymbol{h}) - \mu] = E[Z(\boldsymbol{s})Z(\boldsymbol{s}+\boldsymbol{h})] - \mu^2 = C(\boldsymbol{h}) \qquad [16]$$

where **s** and **s** + **h** refer to points in a d-dimensional space where d could be 1, 2, or 3.

In particular, when **h** = **0**, the covariance comes back to the ordinary variance of Z(**s**) which must also be constant. **h**, is a lag-distance (ranging from zero to the maximum lag = 1 + 3.22log(n), where n is the total number of observation) between two points.

For the data under consideration stationarity is not an appropriate assumption to impose on the (assumed) stochastic mechanism that generated the data. This is because different treatments are being applied to different plots resulting in perhaps different means across the plots. Under such condition, one of the most common practices in geostatistical applications is to separately estimate the large scale structure (i.e. $\boldsymbol{\beta}$ *and* $\boldsymbol{u}$ by OLS or GLS or REML estimator which is based on the type of model under consideration, followed by estimation of the variogram ($i.e., 2\gamma(\cdot; \boldsymbol{\theta})$) based on the resulting residuals. This is simply a two-step procedure. Thus, to estimate the spatial dependence for the data, the mean structure will be (temporally) removed. After the removal of large scale structure (treatment effect) the next step is to analyze the residuals as if they are a sampling from a second order stationary random process. To check the stationarity of the fitted residuals we use the semivariogram. Thus, for a mean-stationary process

the semivariogram must level out and reach a sill (O. Bereke, 1999). This means, semivariogram behavior at large distances indicates the degree of stationarity of the process.

3.2.2.2.3. Estimation of variogram from estimated residuals

In geostatistics spatial dependence is commonly modeled with the semivariogram or variogram. Variogram is also the primary tool in the geostatistical analysis. The variogram measures the mean variability between two point's **s** and **s** + **h** as a function of their distance **h**. Thus the objective of variogram is estimating how spatial dependence varies with distance and also modeling distance decay. The variogram of an intrinsic random function is defined by the equation:

$$\gamma(\boldsymbol{h}) = \frac{1}{2} Var\left[\hat{e}_{(s+h)} - \hat{e}_{(s)}\right] \qquad [17]$$

By assumption the mean of $\hat{e}_{(s+h)} - \hat{e}_{(s)}$ is zero, then $\gamma(\boldsymbol{h})$ is just the mean square value of the difference. That is,

$$\gamma(\boldsymbol{h}) = \frac{1}{2} E[\hat{e}_{(s+h)} - \hat{e}_{(s)}]^2 \qquad [18]$$

where **s** and **s** + **h** refer to points in a d-dimensional space where d could be 1, 2, or 3, $\hat{e}_{(s+h)}$ and $\hat{e}_{(s)}$ is an estimated residual at point **s** + **h** and **s**, respectively.

3.2.2.2.3.1. Method-of-Moments Estimator of Variogram

The classical estimator of the variogram proposed by Matheron (1963) under the constant-mean assumption is given by:

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} \left[\hat{e}(\boldsymbol{s_i}) - \hat{e}(\boldsymbol{s_j})\right]^2 \qquad [19]$$

where the sum is over $N(\boldsymbol{h}) \equiv \{(i,j): \boldsymbol{s_i} - \boldsymbol{s_j} = \boldsymbol{h}\}$ and $|N(\boldsymbol{h})|$ is the number of distinct elements of $N(\boldsymbol{h})$. It is unbiased; however, it possesses very poor resistance properties. It is badly affected by a typical observations due to the $(.)^2$ term in the summed of Equation [19], and its cousin confounds skewness with atypical behavior.

3.2.2.2.3.2. Robust *Estimation of the Variogram*

The word robust is used to describe inference procedures that are stable when model assumptions depart from those of a central model. Cressie and Hawkins present a more robust approach to the estimation of variogram by transforming the problem to location estimation for approximately symmetric distribution (Cressie and Hawkins, 1980). In order to use the available knowledge of robust estimation, Cressie and Hawkins take fourth roots of squared differences, yielding robust (to contamination by outliers; see Cressie and Hawkins, 1980) estimators:

$$2\bar{\gamma}(\boldsymbol{h}) = \left\{ \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} \left| \hat{e}(\boldsymbol{s}_i) - \hat{e}(\boldsymbol{s}_j) \right|^{1/2} \right\}^4 \Big/ \left(0.457 + \frac{0.494}{|N(\boldsymbol{h})|} \right) \qquad [20]$$

The robust estimator of variogram automatically down-weights outlying data.

3.2.2.2.4. Isotropy

Isotropic process assumes that the spatial correlation structure is the same in all directions, or isotropic, i.e. when $\boldsymbol{\gamma(h)}$ depends only on the magnitude and not the direction of the vector **h**, the process is isotropic. In other word, suppose the process is second order stationary with semivariogram $\gamma(\boldsymbol{h}), \boldsymbol{h} \in \mathbb{R}^d$. If $\gamma(\boldsymbol{h}) = \gamma(\|\boldsymbol{h}\|)$ for some function, i.e. if the semivariogram depends on its argument **h** only through its length $\|\boldsymbol{h}\|$, then the process is isotropic. Some of isotropic semivariograms are Exponential, Spherical, Gaussian, Linear, Rational quadratic, Wave and so on.

3.2.2.2.4.1. Checking *for Isotropy*

Prior to fitting a parametric model to the sample semivariogram, we need to determine how the semivariogram depends on the relative orientation of data locations. We want to investigate whether isotropy would be a reasonable assumption for the data. The most popular method is graphical diagnostic. Thus, semivariogram will be computed in different directions to ensure that the spatial variation is isotropic (Cressie, 1991). Clark also stated that, isotropy or direction invariance can be deduced if the semivariograms for several directions coincide (Clark, 1979).

3.2.2.2.4.2. Properties *of variogram*

- ✓ Vanishes at $\mathbf{0}$, i.e., $\boldsymbol{\gamma(0) = 0}$.
- ✓ Evenness, i.e. $\boldsymbol{\gamma(-h) = \gamma(h)}$.
- ✓ The variogram $2\gamma(\boldsymbol{h})$ must satisfy a property called conditional negative definiteness (Matheron, 1971b), namely,

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j 2\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) \leq 0,$$

For any finite number of spatial locations $\{s_i : i = 1, 2, \ldots, m\}$ and a real numbers $\{a_i : i = 1, 2, \ldots, m\}$ satisfying $\sum_{i=1}^{m} a_i = 0$. If this condition is satisfied, we are assured of getting non negative estimates of the mean squared prediction errors. This condition is also correspond to the fact that the variance, var(X), of $X = \sum_{i=1}^{m} a_i Z(x_i)$ is given by the negative of this double sum and must be nonnegative.

- ✓ $\lim_{\|\boldsymbol{h}\|\to\infty}\{\gamma(\boldsymbol{h})/|\boldsymbol{h}|^2\} = 0$.

3.2.2.2.4.3. Variogram *terminology*

Nugget effect: If $\gamma(\boldsymbol{h}) \to c_0 > 0, as\ \boldsymbol{h} \to \mathbf{0}$, then c_0 has been called nugget effect by Matheron (1962). This is because it is believed that microscale variation (small nuggets) is causing a discontinuity at the origin.

Sill: Another important characteristic of the semivariogram deals with the fact that, as the separation distance h increases, this function eventually approaches a constant value, known as the sill and the difference between the sill and the nugget (sill-nugget) is called the '***partial sill***'.

Range: When a semivariogram has a sill, it means that there is a distance beyond which the correlation between variables is zero; this distance is called the range i.e.it is the value of $\boldsymbol{h}$ at which $\gamma(\boldsymbol{h})$ first attains the sill. Graphically it can be illustrated as follows:

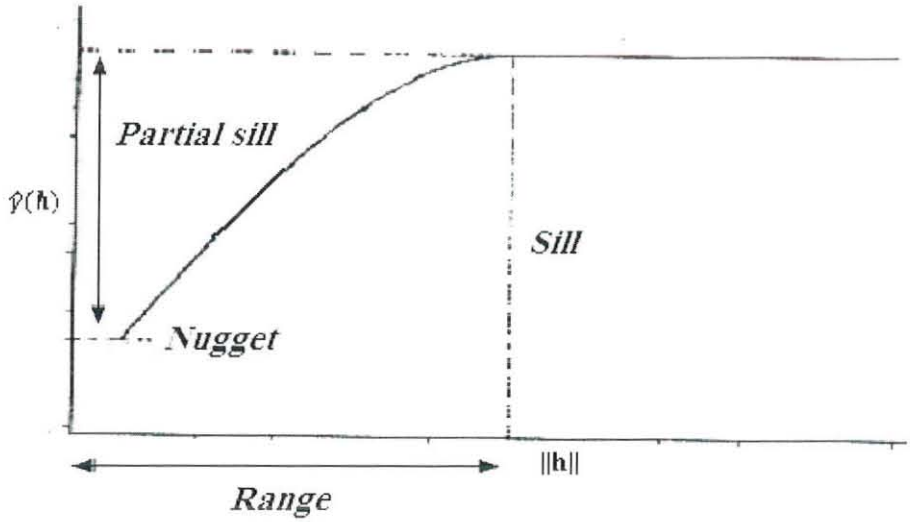


Figure 1: *Graphical illustration of components of Semivariogram*

3.2.2.2.5. Variogram model fitting

The above methods ($2\bar{\gamma}(\boldsymbol{h})$ and $2\hat{\gamma}(\boldsymbol{h})$) of variogram estimation are invalid variograms,

Why aren't we satisfied with just the sample semivariogram itself?

- ✓ The sample semivariogram may violate the required property of conditional negative definiteness. i.e. they are not necessarily conditionally negative-definite.
- ✓ For various purposes (e.g. kriging) we may require an estimate of the semivariogram at a lag not represented in the data.
- ✓ The sample semivariogram may be quite bumpy. A smoothed version may be helpful for understanding the nature of the spatial dependence.

The idea then is to search for a valid variogram that, as a measure of spatial dependence, is closest to the spatial dependence present in the data. One possible approach to solve this problem is to approximate the variogram by any parametric model which is known to be valid. The idea is to search, among the families of valid variograms, for one that best approximates the underlying spatial dependence of the available sample data. Thus, the next step of the geostatistical method is to select and fit a parametric family of models to the sample semivariogram (Menezes *et al.*, 2005).

3.2.2.2.6. Selection of a valid semivariogram model $\gamma(\mathbf{h};\boldsymbol{\theta})$

Once the sample semivariogram ($\bar{\gamma}(\boldsymbol{h})$ and $\hat{\gamma}(\boldsymbol{h})$) is computed the next step is to find a negative-definite function which, as a measure of spatial dependence, is in some sense closest to the sample data. At this stage we are concerned with questions such as the choice of exponential versus spherical families of distributions. The most common methods used to pick a valid family are based on graphical tools, with model selection reduced to approximating the estimated variogram curve by one distribution from the valid family. In this graphic representation, estimated values of semivariance $\hat{\gamma}(\boldsymbol{h})$ *or* $\bar{\gamma}(\boldsymbol{h})$, are plotted against their respective distances h, resulting in a scatter plot (sample variogram). The estimated (empirical) variogram will be plotted only up to distances equal to about half the dimensions of the observation region, because at regular distances the empirical variogram is very unreliable (Stein, 1999). For the vast majority of identified approaches, this search is restricted to the space of parametric families. This is the set of all analytic functions that depend on parameters in such a way that two members of the same family can be distinguished by their parameters values, and all members are conditionally negative-definite. The most commonly utilized variogram functions are the so-called spherical, exponential and Gaussian models (Grondona & Cressie, 1991; Zimmerman & Harville, 1991; Vieira, 2000; Nan Hong *et al.*, 2005) and are defined as follows:

Gaussian model:

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = 0 \\ c_0 + c_g\left(1 - e^{-\frac{\|\mathbf{h}\|^2}{a_g{}^2}}\right) & \mathbf{h} > 0 \end{cases} \qquad [21]$$

$$\boldsymbol{\theta} = (c_0, c_g, a_g)', \text{where } c_0 \geq 0, c_g \geq 0, \text{and } a_g \geq 0.$$

Spherical model (valid in $\mathbb{R}^1$, $\mathbb{R}^2$, and $\mathbb{R}^3$):

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} 0 & \mathbf{h} = 0 \\ c_0 + c_s\left\{\left(\frac{3}{2}\right)\left(\frac{\|\mathbf{h}\|}{a_s}\right) - \left(\frac{1}{2}\right)\left(\frac{\|\mathbf{h}\|}{a_s}\right)^3\right\}, & 0 < \|\mathbf{h}\| \leq a_s \\ c_0 + c_s & \|\mathbf{h}\| \leq a_s \end{cases} \qquad [22]$$

$$\boldsymbol{\theta} = (c_0, c_s, a_s)', \text{where } c_0 \geq 0, c_s \geq 0, \text{and } a_s \geq 0.$$

Exponential model (valid in $\mathbb{R}^d, d \geq 1$):

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} 0, & \mathbf{h} = 0 \\ c_0 + c_e\left\{1 - \exp\left(-\frac{\|\mathbf{h}\|}{a_e}\right)\right\}, & \mathbf{h} \neq 0 \end{cases} \quad [23]$$

$$\boldsymbol{\theta} = (c_0, c_e, a_e)', \text{where } c_0 \geq 0, c_e \geq 0, \text{and } a_e \geq 0.$$

where: c_0 is a nugget effect, c_g, c_s, c_e are partial sills for Gaussian, Spherical and Exponential model, respectively, and a_g, a_s, a_e are ranges for Gaussian, Spherical and Exponential models, respectively.

3.2.2.2.7. Method of Estimating Parameters of Variogram Model

Let $\boldsymbol{\gamma}(\boldsymbol{h};\ \boldsymbol{\theta})$ denote the parametric model to be fitted to the sample semivariogram and let $\boldsymbol{\Theta}$ denote the parameter space for $\boldsymbol{\theta}$ under the two most important assumptions of the process, intrinsic stationarity and isotropy. Once the appropriate variogram model has been selected the next step is to estimate the parameter of the model. This can be done through different method. In this thesis we will consider only the most applicable method of estimation in field experiment that is a restricted maximum likelihood (REML) estimator (Littell *et al.,* 1996).

3.2.2.3.Spatial linear fixed effect model.

Consider a linear fixed effect model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad [24]$$

where $\mathbf{Y}$ is the vector of observations, $\mathbf{X}$ the design matrix, and $\boldsymbol{\varepsilon}$ the vector of residuals, whose variance-covariance matrix is denoted as $\boldsymbol{\Sigma}$. The spatial dependence can be explained in the error term $\boldsymbol{\varepsilon}$. We need to estimate the components of $\boldsymbol{\beta}$ and $\mathbf{R}$ which, under the intrinsic stationarity hypothesis of the process, will depend on a vector of parameters $\boldsymbol{\theta}$, that is, $\mathbf{R} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is an nxn matrix whose elements are known function of $\boldsymbol{\theta}$(Exponential, Gaussian and Spherical). The estimation of both sets of parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, is carried out using an iterative procedure by means of restricted maximum likelihood (REML) method, which has certain advantages over other common methods like the maximum likelihood or weighted least squares. Hence, the

estimate of the parameter $\boldsymbol{\beta}$ is given as: $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{Y}$ and $\widehat{\boldsymbol{\theta}}$ is an estimator of $\boldsymbol{\theta}$.

3.2.2.4. Spatial mixed linear model

Consider a model under Equation [5]:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}, \qquad [25]$$

where $\mathbf{Y}$, $\boldsymbol{X}$, and $\boldsymbol{\beta}$ are as defined in equation [1], $\mathbf{Z}$ and $\boldsymbol{u}$ are as defined in equation [5], and $\boldsymbol{\varepsilon}$ is a vector of unobserved random errors in which the spatial dependence can also be explained in it. Here it is assumed that:

- ✓ $E(\boldsymbol{u}) = \mathbf{0}$ and $Var(\boldsymbol{u}) = \boldsymbol{G}$
- ✓ $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $Var(\boldsymbol{\varepsilon}) = \boldsymbol{R} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is an nxn matrix whose elements are known function of $\boldsymbol{\theta}$(Exponential, Gaussian and Spherical).
- ✓ $Cov(\boldsymbol{u}, \boldsymbol{\varepsilon}) = \mathbf{0}, and$
- ✓ Both $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ are normally distributed. Thus,

$$\begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim Gau\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{G} & 0 \\ 0 & \boldsymbol{\Sigma}(\boldsymbol{\theta}) \end{bmatrix}\right)$$

The estimation of both sets of parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, is carried out using an iterative procedure by means of restricted maximum likelihood (REML) method, that has certain advantages over other common methods like the maximum likelihood or weighted least squares.

3.2.3. Method of model comparison

For the comparison of spatial model with that of conventional models, the information criterion ($AIC\ or\ AIC_c$) and the likelihood ratio x^2 test will be used in this study. An overview of the two methods are discussed below.

3.2.3.2. Akaike's Information criterion (AIC) and Corrected Akaike's Information criterion ($\mathbf{AIC_c}$)

The AIC values are computed as

$$AIC = -2Log(l(\hat{\theta})) + 2q \qquad [26]$$

where l is the residual log likelihood of restricted maximum likelihood (REML) is implemented and q is the number of random effect parameters being estimated using REML. The AIC performs poorly if there are too many parameters in relation to the size of the sample (Sugiura, 1978). Thus, a bias adjustment was included leading to the formation of the modified AIC criterion (Hurvich and Tsai, 1989) defined as

$$AIC_c = -2Log(l(\hat{\theta})) + 2q\left(\frac{n}{n-q-1}\right),$$

This simplifies to

$$AIC_c = AIC + \frac{2q(q+1)}{n-q-1}, \qquad [27]$$

where n is the sample size. The smaller AIC values indicate better fit. We chose to use both AIC and AICc in preference to the other information criteria available in SAS Proc MIXED, Schwarz' Bayesian Information Criterion (BIC) (Schwarz, 1978). Guerin and Stroup (2000) compared the performance of AIC and BIC on covariance model selection for repeated measures and stated that AIC tended to result in a more complex model but with better Type I error rate control than BIC.

To compare two or more non-nested models with different global and local covariate, and with the same fixed effect and same parameterization, an AIC_c will be used. A model that had AIC_c close to zero will be selected.

3.2.3.3. Likelihood Ratio χ^2Test

The other statistical tool used for model selection is the likelihood ratio χ^2 test (LRT). This test, however, is only applicable when comparing two nested covariance models with the same X matrix (fixed effects). Consequently, when testing a hypothesis that involves comparison of non-nested covariance models (still with the same X matrix), AIC_c is used (as described above). When models are nested, hypothesis testing may be possible with the LRT statistic (λ) which approaches a χ^2 distribution with (df1 − df2) degrees of freedom, where df1 and df2 are the degrees of freedom of the full model and the reduced model, respectively and which is defined as:

$$\lambda = 2(l_2 - l_1) \sim \chi^2(\text{df1} - \text{df2}) \quad [28]$$

where l_1 and l_2 denote the residual log likelihoods of the reduced model and the full model, respectively (Littell *et al.*, 1996).

In practice, there is a condition that must be met to apply LRT. The hypothesized value of the parameter being tested must not be on the boundary of the parameter space. For example, by definition a variance component must be non negative, therefore the value zero is on the boundary of the parameter space. In this case, the asymptotic distribution of λ does not follow a χ^2 distribution. When the models being compared differ by such parameters, e.g., variance components or spatial ranges, the distribution of the test statistic consists of mixtures of χ^2 distributions (Verbeke and Molenberghs, 2000). If the models being compared differ by only one such parameter, the appropriate p-value for the LRT is one-half the reported p-value from the χ^2 distribution with one degree of freedom (consistent with dropping one parameter from the full model), hereafter referred to as LRT_{01} (Littell et al., 1996). If the models being compared differ by more than one such parameter, then the actual distribution of is unknown and we use AIC_c values (as described above) to determine which model has the better goodness of fit. The use of AIC_c to determine which model has a better fit is the same as an LRT but uses a critical value of twice the difference in the number of random effect parameters between the two models under consideration. Hence, if the two models have the same number of parameters, the AIC_c just selects the model with higher likelihood. If the two models differ by one random effect parameter, the AIC_c uses a "critical value" of 2, which is less conservative than an LRT, which has a "critical value" of 3.84 (the 95th percentile of χ_1^2) or LRT_{01}, which has a "critical value" of 2.71 (the 95$^{\text{th}}$ percentiles of LRT_{01}) (Morgan and Gumpertz, 1996). Consequently, while the use of LRT or LRT_{01} can be seen as adding complexity to model selection compared with simply using AIC_c values, it also adds a degree of robustness and confidence in the selection process that are absent when using AIC_c values alone (Nan Hong *et al.*, 2005).

3.2.4. Statistical test of treatment effects

Inference on treatment effects involves either hypothesis testing or interval estimation of estimable functions of the form $\mathbf{K}'\boldsymbol{\beta}$. When the variance and covariance parameters (e.g., the

error variance σ^2 and, where applicable, the block variance σ_b^2 or the spatial covariance range parameter, r) are known, the sampling distribution of the estimate of $\boldsymbol{K'\beta}$ is

$$\boldsymbol{K'\hat{\beta} \sim MVN\left(K'\beta, (K'CK)\right)}, \quad [29]$$

where $\boldsymbol{C = (X'V^{-1}X)}$ **and** $\boldsymbol{V = ZG'Z + R}$. The Wald statistic for testing H_0: $\boldsymbol{K'\beta = 0}$,

$$Wald(\boldsymbol{K'\beta = 0}) = \boldsymbol{\hat{\beta}'K(K'CK)^{-1}K'\hat{\beta}}, \quad [30]$$

has an approximate χ^2 distribution with rank (**K**) as degrees of freedom and noncentrality parameter $\boldsymbol{\beta'K(K'CK)^{-1}K'\beta}$. In practical situations, the variance and covariance parameters are not known and their estimates must be used. In such cases, an approximate F statistic

$$F(\boldsymbol{K'\beta = 0}) = \frac{\boldsymbol{\hat{\beta}'K(K'CK)^{-1}K'\hat{\beta}}}{rank(\boldsymbol{K})} \quad [31]$$

is used, where $\boldsymbol{C = (X'\hat{V}^{-1}X)^-}$ and $\boldsymbol{\hat{V}}$ is the estimate of **V** using the estimated variance and covariance parameters. This has an approximate F distribution with $rank(\boldsymbol{K})$ numerator degrees of freedom, v denominator degrees of freedom, and noncentrality parameter $\emptyset = \boldsymbol{\beta'K(K'CK)^{-1}K'\beta}$ (Stroup, 2002).

The denominator degrees of freedom depends on the estimate of $\boldsymbol{K'CK}$. In a simple blocked design v is obvious from the form of the analysis of variance, whereas in spatial mixed v will be approximated using two different methods:

(1) ***Satterthwaite's method:*** this approximation was advocated by Giesbrecht and Burns (1985) and Fai and Cornelius (1996) based on Satterthwaite (1941), in which the degrees of freedom for the test of treatment effects are computed based on the estimated covariance structure rather than on the number of plots as in ANOVA (Verbeke and Molenberghs, 1997).

(2) ***The Kenward-Roger method (Kenward and Roger 1997):*** This approximation also uses the basic idea of Satterthwaite (1941). Its extension relative to the Satterthwaite method of Giesbrecht and Burns (1985) and Fai and Cornelius (1996) is an asymptotic correction of the estimated standard error of fixed effects due to Kackar and Harville (1984) in small and or unbalanced data structures.

In this study, only the Kenward-Roger method of estimating denominator degree of freedom was used.

This approximation uses the equation

$$v = \frac{2E[(\boldsymbol{K}'\boldsymbol{C}\boldsymbol{K})]^2}{V[\boldsymbol{K}'\boldsymbol{C}\boldsymbol{K}]}. \qquad [32]$$

For a single degree of freedom estimable function, that is, when **K** is a vector, this reduces to

$$v = \frac{2E(\boldsymbol{K}'\boldsymbol{C}\boldsymbol{K})^2}{V[\boldsymbol{K}'\boldsymbol{C}\boldsymbol{K}]} \qquad [33]$$

The denominator of this expression can be approximated using gradients and the asymptotic covariance matrix of the estimates of the variance and covariance components.

CHAPTER FOUR

4. RESULT AND DISCUSSION

4.1. Introduction

The purpose of this chapter is to assess the presence of spatial dependence in the data sets considered in this study and fit different spatial models and RCBDiid models for each data set. An attempt has been made to identify the best fitting model for each of the data based on different model selection tools such as information criteria. Furthermore, in this chapter, the effect of the spatial models have been assessed by using p-values and standard errors. Of the four data sets considered in this thesis, for the first two both block and treatment/variety effects are assumed to be fixed and for the second two block effect are assumed to be random and treatment /variety effects were assumed to be fixed and for the second two data sets only the block effects were assumed to random. For all of the data sets the assumption of normality has been assessed using graphical method (QQ plots) and the formal test (Shapirio Wilks test). We start our data analysis by checking the common assumptions for all of the data sets; we then proceed to model fitting and selection for each of the data sets; and thereafter we proceed to see the impact of the fitted model on p-values of the test and standard errors. The yield from all trials was taken as kg/plot and is be summarized in Table 1.

Table 1: Total v*ariety number, plot shape, effects, spacing between block and plot, mean and range of the yield data of the trials.*

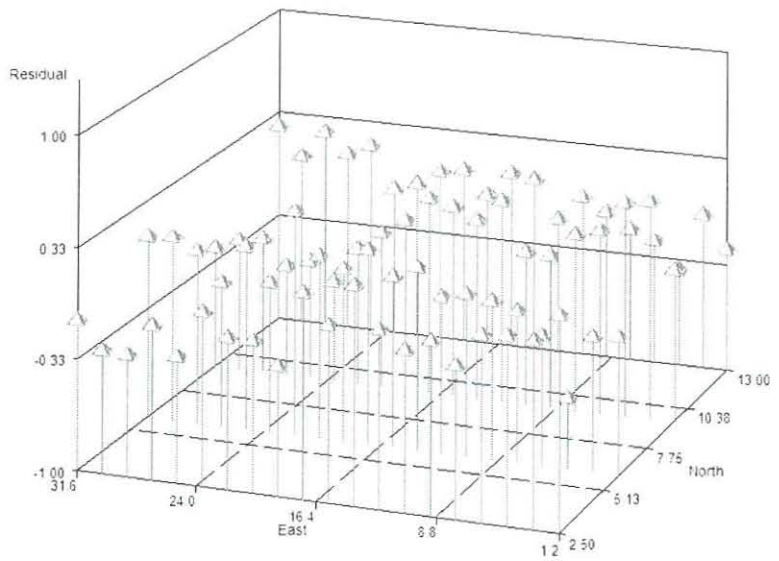
Trials	Design						Yield	
	Totalv ariety no.	Plot shape in meter,m	Effects		Spacing b/n		Mean (Kg/plot)	Range (Kg/plo t)
			Block	Variety	Block	Plot		
BW99RVTII5	20	2.5 X 1.2	Fixed	Fixed	1.00m	0.4m	0.806	1.003
BW99RVTII6	20	2.5 X 1.2	Fixed	Fixed	1.00m	0.4m	0.590	0.819
BW99RVTII4	20	2.5 X 1.2	Random	Fixed	1.00m	0.4m	1.292	1.023
DW99RVTII1	19	2.5 X 1.2	Random	Fixed	1.00m	0.4m	1.052	0.938

4.2. The common assumption for all data sets

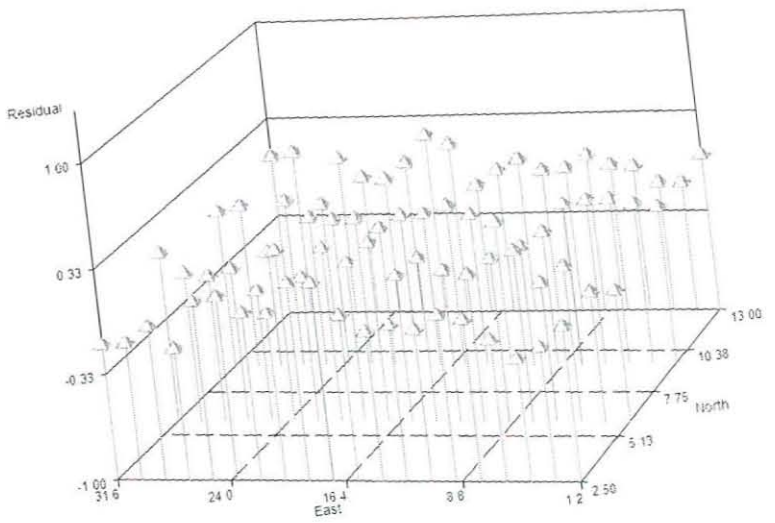
For all data sets, the RCBDiid were fitted and the assumption of normality for residuals from the fitted model was assessed by graphical method (QQ plots) and the formal test (Shapirio Wilks test). There was no strong evidence of non normality for the residuals for all trials, so the data were analyzed in the original scale *(See Appendix II)*.

4.2.1. Diagnosis for the trend effect

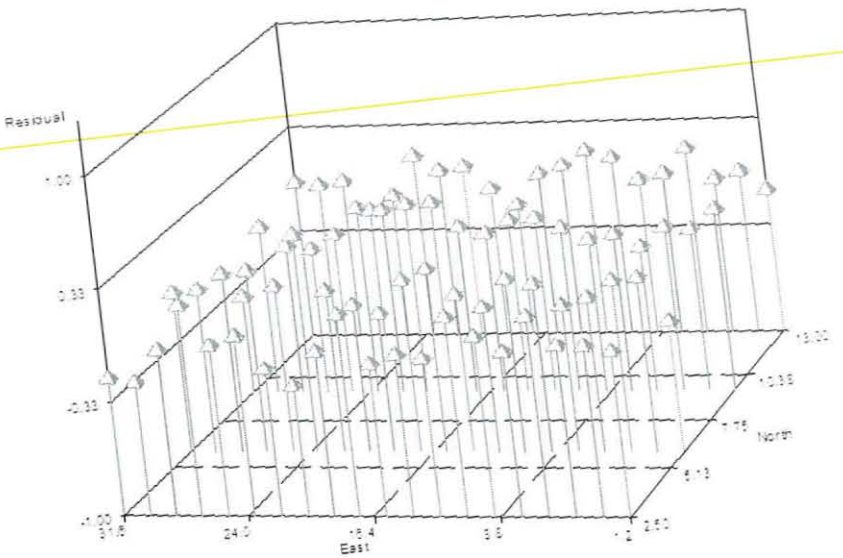
One of the fundamental assumptions underlying variogram modeling is the absence of a spatial trend. In other words, the mean value of the process should be constant throughout space. This is a crucial step for removing any obvious surface trend effect before checking for spatial dependence (the semivariogram model). Since different varieties were applied to different plots in all data sets, we should first remove the effect of varieties and also the block effect from all data sets. This can be done by first fitting the RCBDiid model for all data sets and obtain the residuals for this fitted model. Hence, further the trend effect must be checked for the estimated residual from the RCBDiid model for each of the data sets. This can be checked by a 3-dimensional (3d) surface plot which is a three dimensional plot of the measured variable of the estimated residuals for each of the four data sets with their respective location.



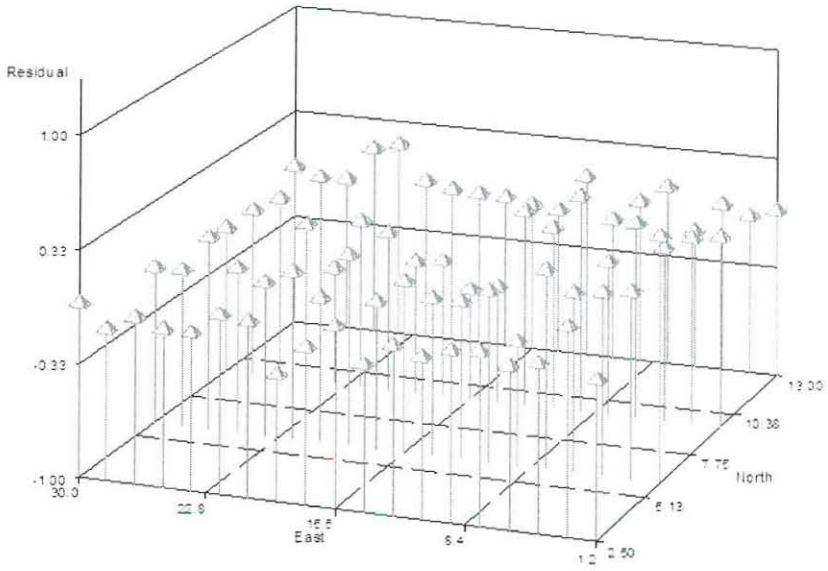
(a)


Residual
1.00
0.33
-0.33
-1.00
31.6
24.0
16.4
8.8
1.2
East
2.50
5.13
7.75
10.38
13.00
North

(b)


Residual
1.00
0.33
-0.33
-1.00
East
North

(c)



(d)

Figure 2: *(a) Three dimensional scatter plot of measurement locations for the estimated residuals from RCBiid model of BW99RVTII5 data set, (b). Three dimensional scatter plot of measurement locations for the estimated residuals from RCBiid model of BW99RVTII6 data set. (c) Three dimensional scatter plot of measurement locations for the estimated residuals from RCBiid model of BW99RVTII4 data set, and (d) Three dimensional scatter plot of measurement locations for the estimated residuals from RCBiid model of DW99RVTII1 data set.*

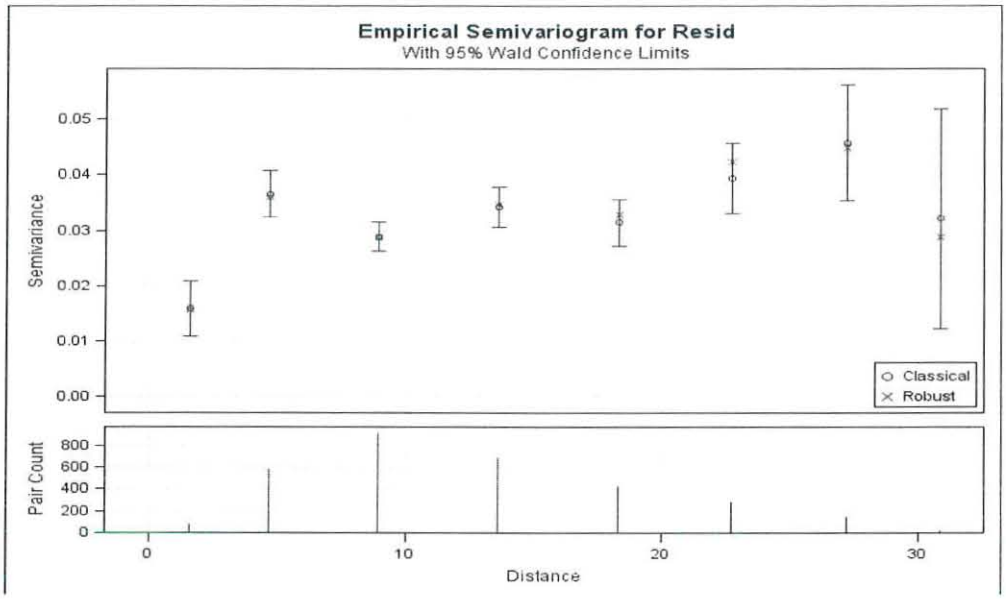
As it can be seen from Figure 2 above, there is no trend effect in the estimated residual for all of the data sets under study. Hence, we can work with the estimated residuals without any doubt about trend effect. The reason for this result is that, at the beginning since different plot are applied with different treatment we have removed the effect of variety and block. In other words the data were already detrended and we are working with only the estimated residuals.

4.2.2 Estimation of Empirical and Robust Semivariogram

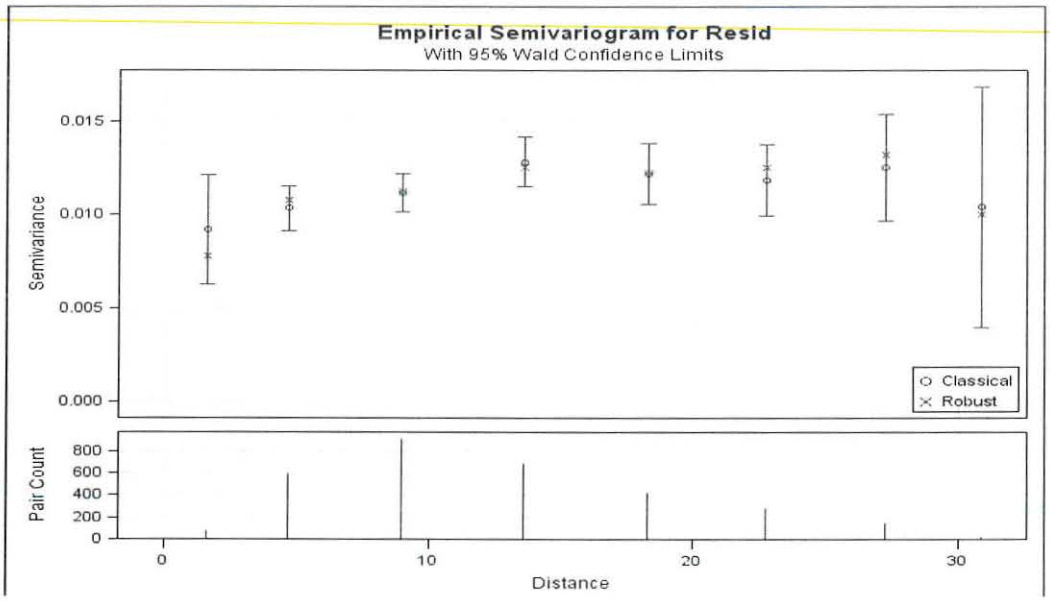
The semivariogram is the main important tool in the analysis of spatial data. To estimate empirical (robust) variogram we should first divide the data into different lags based on their location. For plotting and estimations purposes, it is desirable to have as many points as possible for the plot of semivariogram against lag distance. This corresponds to having as many distance intervals as possible. If the lag distance is set too small, there may be too few points in one or

more of the intervals. On the other hand, if the lag distance is set too large value, the number of point pairs in the distance intervals may be much greater than those needed for estimation precision, thereby "wasting" point pairs at the expense of variogram points. Hence, there is a tradeoff between the number of distance intervals and the number of point pairs within each interval. In SAS VERSION 9.2. MANUAL, it is stated that, optionally, one can use a rule of thumb in order to determine the number of lags. Hence, in our case for this study the number of lags was determined by the rule of thumb i.e. number of lags = 1 + 3.32log (n), where n is the number sample observations. In this case the number of lags that have been utilized for this study was about 7. Distributions of pairwise distance for all of the data sets are given in *Appendix IV*. The maximum lag for each of data sets were determined by the SAS version 9.2 software. For BW99RVTII4, BW99RVTII5 and BW99RVTII6 data sets, the estimated maximum lag was 4.59m and for DW99RVTII1 data set, the estimated lag distance was 4.38m. In all data sets the lag distances were measured in meter.

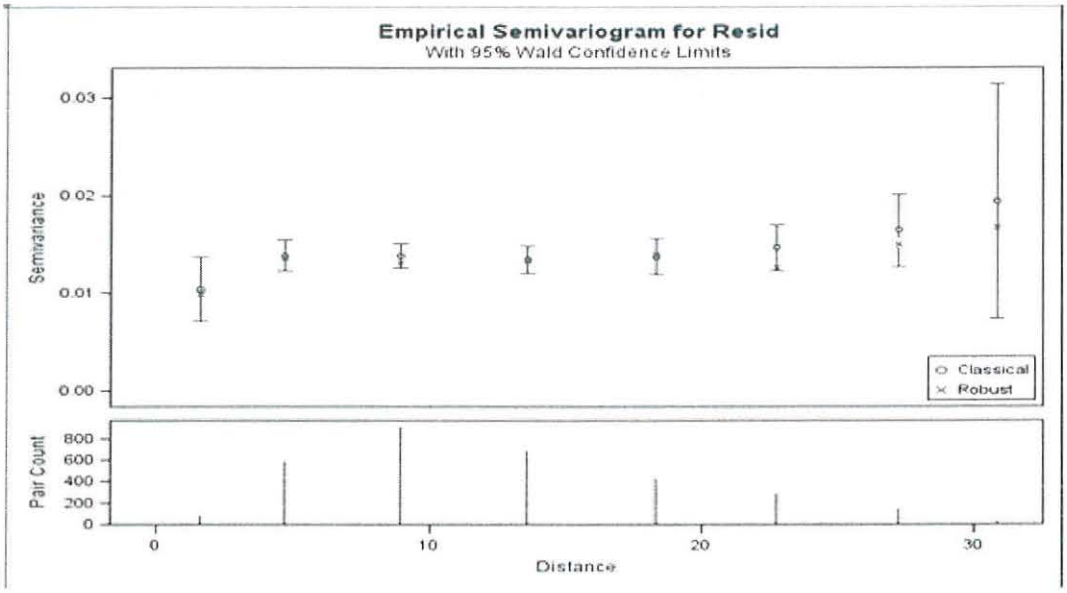
For all of the data sets considered in this study, both empirical (classical) and Robust semivariogram for the estimated residuals of the RCBiid model were fitted with their corresponding Wald confidence limits. If there is an outlier in the data sets, the robust semivariogram is preferred to the classical semivariogram. But in our case as it can be seen from the Figure 3 below, almost for all of the data sets both the empirical (classical) and the Robust semivariogram are the same. This implies that, there are no as such extreme outliers in the estimated residuals of RCBDiid model for each of the four data sets. By default the SAS version 9.2. software computes the Wald confidence limits for the estimated residual semivariance *(See Appendix V)*. Here also as it can be seen from Figure 3 both the Robust and the classical semivariance values at each lag lies within a 95% Wald confidence limits. This result also shows that the fitted semivariances for the estimated residuals are significant at 5% level of significance.


Empirical Semivariogram for Resid
With 95% Wald Confidence Limits
Semivariance
0.05
0.04
0.03
0.02
0.01
0.00
Classical
Robust
Pair Count
800
600
400
200
0
0
10
20
30
Distance

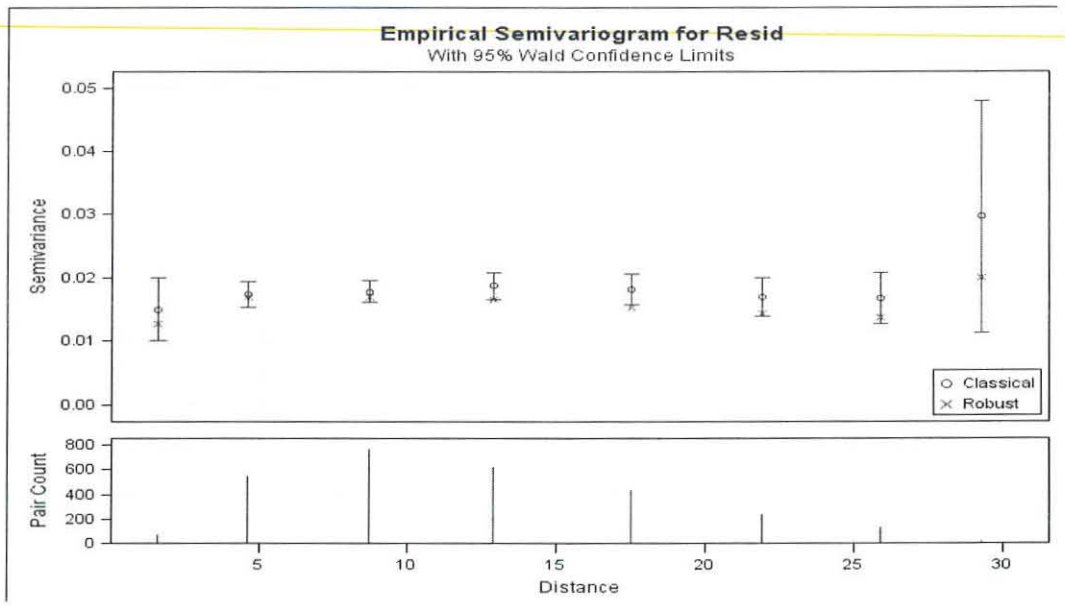
(a)


Empirical Semivariogram for Resid
With 95% Wald Confidence Limits
Semivariance
0.015
0.010
0.005
0.000
Classical
Robust
Pair Count
800
600
400
200
0
0
10
20
30
Distance

(b)



(c)



(d)

Figure 3: *(a). The Empirical (Classical) and robust semivariogram for the estimated residuals of the RCBiid model for BW99RVTII5 data set, (b). The Empirical (Classical) and robust*

semivariogram for the estimated residuals of the RCBiid model for BW99RVTII6 data set. (c). The Empirical (Classical) and robust semivariogram for the estimated residuals of the RCBiid model for BW99RVTII4 data set and (d). The Empirical (Classical) and robust semivariogram for the estimated residuals of the RCBiid model for DW99RVTII1 data set. The histogram like graph under each of estimated semivariograms represents the distribution of pairwise distance for each of estimated residuals.

4.2.3. Diagnosis for the assumption of stationarity

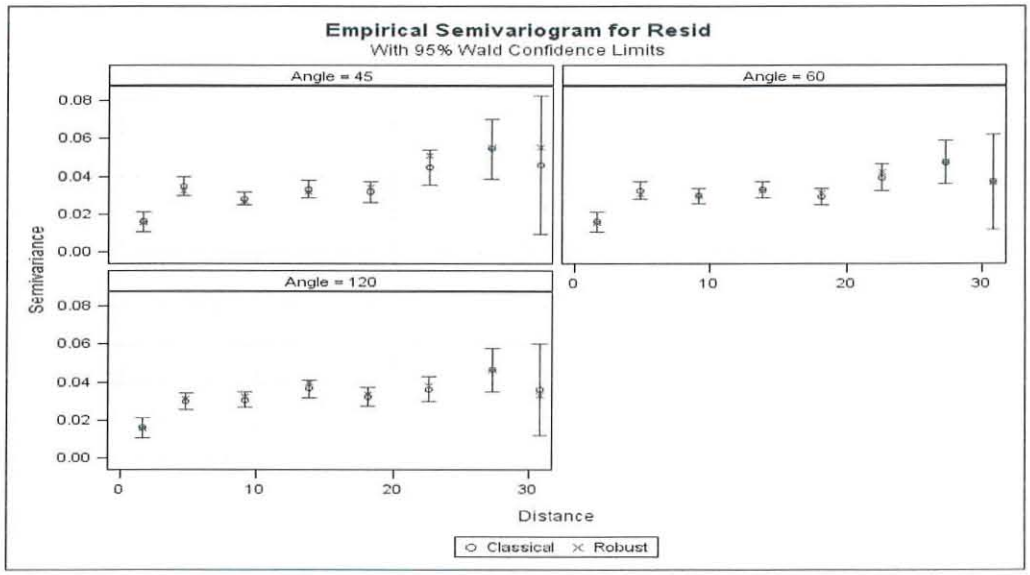
For a second order stationary process, the semivariogram must level out and reach a sill (O. Bereke, 1999). As it can be seen from the above Figure 3, for all of the data sets the semivariogram level out and reach sill. Hence, we conclude that the stationarity assumption has been met for all of the four data sets.

4.2.4. Diagnosis for the presence of spatial dependence

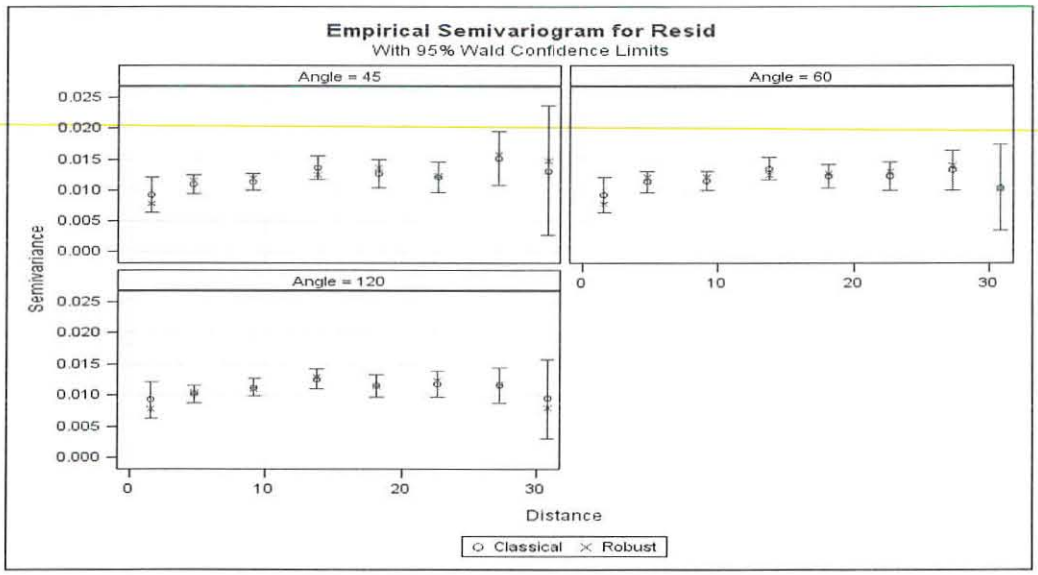
As it has been clearly stated in Chapter three in the methodology part, if the fitted empirical or robust semivariogram for a given data set is constant, then there will be no spatial dependence among the observations. Clearly, as it can be seen from the Figure 3 above, the fitted empirical and robust semivariogram for all data sets indicate positive spatial correlation among the observations that is the observations that are closer together tend to be more similar than ones farther apart (semivariance increased as lag distance increased). This indicates that there is a spatial dependence in the estimated residuals of the RCBiid model for both data sets.

4.2.5. Diagnosis for the assumption of isotropy

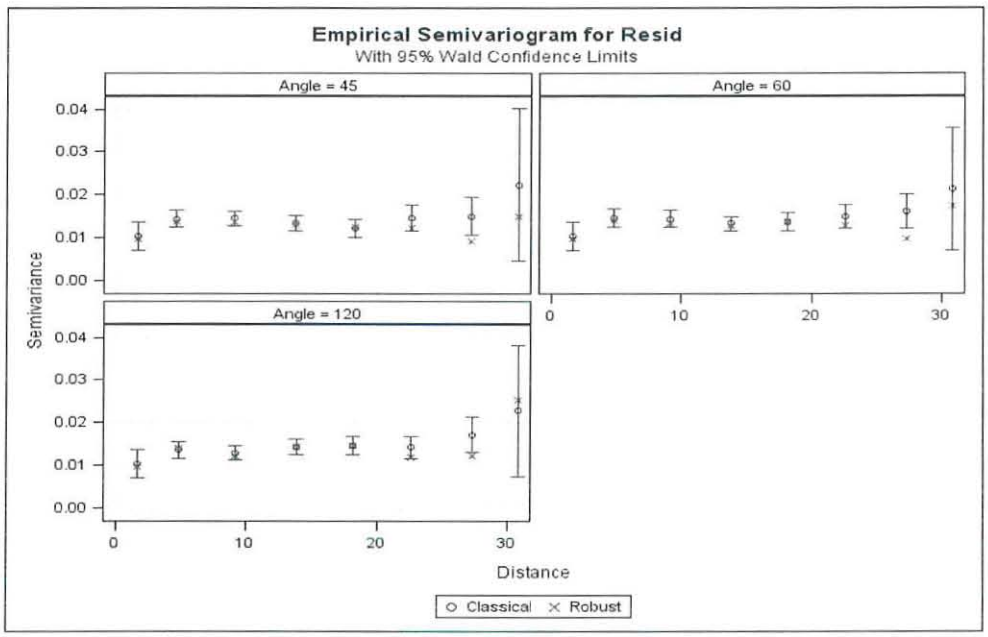
As it has been clearly explained in the methodology part of Chapter three, the empirical variogram is not a valid variogram; hence, we should search for the spatial model which satisfies the assumption of variogram model. Particularly the assumption of isotropy of the spatial data must be satisfied so as to search for the best spatial model among the isotropic variogram models. The isotropic assumption of the spatial data can be checked by fitting semivariogram in different directions. In our case by arbitrary selecting three directions (degrees 120^0, 60^0 and 45^0) we have fitted a semivariogram.


Empirical Semivariogram for Resid
With 95% Wald Confidence Limits
Angle = 45
Angle = 60
Angle = 120
Semivariance
0.00
0.02
0.04
0.06
0.08
0
10
20
30
Distance
Classical
Robust

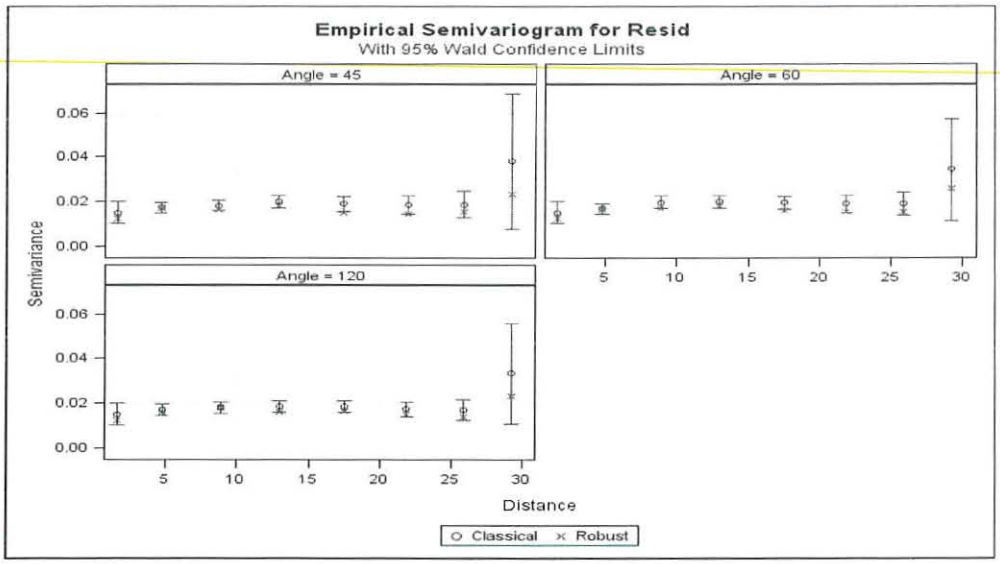
(a)


Empirical Semivariogram for Resid
With 95% Wald Confidence Limits
Angle = 45
Angle = 60
Angle = 120
Semivariance
0.000
0.005
0.010
0.015
0.020
0.025
0
10
20
30
Distance
Classical
Robust

(b)



(c)



(d)

Figure 4: *(a). A three dimensional variogram for estimated residuals of the RCBiid model for BW99RVTII5 data set and (b). A three dimensional variogram for estimated residuals of the RCBiid model for BW99RVTII6 data set, (c). A three dimensional variogram for estimated*

residuals of the RCBiid model for BW99RVTII4 data set, and (d). A three dimensional variogram for estimated residuals of the RCBiid model for DW99RVTIII data set.

As it can be seen from Figure 4 above, the fitted semivariogram for estimated residuals of the RCBiid model for all of the data sets were fairly similar in all three directions, suggesting that the assumption of isotropy holds for the estimated residual of the RCBiid model for all of the four data sets.

4.3. Results

For all of the four data sets, the semivariogram fitted were used to estimate the starting parameters for the candidate models. It was also assumed (based on Figure 3) that the spatial covariance of the errors would be best modeled with either a Gaussian, Exponential, or Spherical function either with or without a nugget effect. These are the three most common spatial covariance structure fitted in field trials (Zimmerman and Harville, 1991).

4.3.1. Model selection for the first two data sets (BW99RVTII5 and BW99RVTII6)

For both of the data sets considered in this section, both block and variety effects were assumed to be fixed. Hence, those models that will be fitted for these data sets are called fixed effect models and particularly those models which take into account for spatial variation are called spatial fixed effect models. We have used only AIC and AICc as a model selection tool. Ideally, the LRT were used to compare the CR model and spatial candidate models.

4.3.1.1. Model selection for BW99RVTII5 data set

The semivariogram for the estimated residuals of the RCBiid model for BW99RVTII5 data set (Figure 3 (a)) indicates positive spatial correlation in that observations that were closer together tended to be more similar than ones farther apart (semivariance increased as lag distance increased). The semivariogram also indicated that spatial covariance structures with nugget effects would be appropriate, as the semivariance remained quite high because the lag distance approached zero. The BW99RVTII5 data set candidate covariance models and their related statistics and covariance parameters are summarized in Table 2. A nugget exponential and a

nugget spherical (both with and without block effect) did not reach convergence in this data fitting. This shows that these models are not suitable for this data set. Apart from these models, the other spatial models (both with and without block effects) and the RCB model had markedly smaller AIC and AICc values than the CR model. This result is also supported by the Likelihood ratio test. Thus, apart from the two models, Exponential with nugget and Spherical with nugget, the LRT indicates that the other spatial models (both with and without block effects) and the RCB model had a very significant difference in comparison with the CR model. This result suggested that spatial variability is significant in this trial, and that the use of both blocking and spatial model analysis were useful for this trial. But all of the candidate spatial models, both with and without blocking (apart from Exponential with nugget and Spherical with nugget models) had markedly smaller AIC and AICc values than the RCB model showing that spatial models were more useful than the RCB model. Among a without block effect no-nugget spatial candidate models, the no-nugget exponential model was selected because it had the smallest AIC and AICc values compared with a without block effect no-nugget Gaussian and spherical models. On the other hand among a no-nugget spatial candidate models with block effect a no-nugget Gaussian model was selected because it had the smallest AIC and AICc over the other models under this category. Between a without block effect no-nugget exponential and no-nugget Gaussian model with block effect, a without block effect no-nugget exponential model with smaller AIC and AICc value was chosen. Up on the comparison of without block and nugget spatial candidate models, a nugget Gaussian model was selected according to AIC and AICc. In comparing with block effect nugget spatial candidate models, a nugget Gaussian model was selected according to AIC and AICc. Between without block effect nugget Gaussian and with block effect nugget Gaussian models, with block effect nugget Gaussian model with smaller AIC and AICc value was chosen. Finally, between without block effect no-nugget exponential and with block effect nugget Gaussian models, a without block effect nugget Gaussian model with smaller AIC and AICc was chosen to be the best model for BW99RVTII5 data set (Table 2).

Table 2. BW99RVTII5 data set spatial candidate covariance models and related statistics and covariance parameters including randomized complete block with iid errors (RCBiid).

BW99RVTII5				MODELS						
				CR/RCBD	**No nugget Exponential**	**Nugget Exponential**	**No nugget Spherical**	**Nugget Spherical**	**No nugget Gaussian**	**Nugget Gaussian**
Fitting Statistics	**Without Block(1)**	q		1	2	-	2	-	2	3
		-2lnLL		22.7	-1.5	-	1.7	-	0.7	-5.7
		AIC		24.7	*2.5*	-	5.7	-	4.7	**0.3**
		AICc		24.8	*2.7*	-	5.9	-	4.9	**0.7**
		P value		0.0013	<0.0001	-	<0.0001	-	0.0003	0.0004
	With Block(2)	q		1	2	-	2	-	2	3
		-2lnLL		21.8	3.8	-	7.3	-	*3.5*	-1.1
		AIC		23.8	7.8	-	11.3	-	*7.5*	4.9
		AICc		23.8	8.0	-	11.5	-	7.8	5.4
		P value	Block	0.0079	0.3565	-	0.8221	-	0.0908	0.2852
			Trt	0.0003	0.0003	-	<0.0001	-	0.0003	0.0004
Covariance parameter estimates	**Without Block(1)**	Range, m		NA	3.4843m	-	19.3578m	-	1.9158	3.2793m
		Sill		*0.05388*	0.05854	-	0.1689	-	0.04761	0.054401
		Nugget		NA	NA	-	NA	-	NA	0.009721
	With Block(2)	Range, m		NA	2.8473m	-	19.371m	-	1.8362m	3.0841m
		Sill		*0.04614*	0.05146	-	0.1751	-	0.04318	0.049903
		Nugget		NA	NA	-	NA	-	NA	0.009753

Likelihood ratio tests	**(1) to CR**	d.f.	-	1	-	1	-	1	2
		LRT	-	24.2	-	21	-	22	28.4
		P value	-	<0.0001	-	<0.0001	-	<0.0001	<0.0001
	(2) to CR	d.f.	-	1	-	1	-	1	2
		LRT	-	18.9	-	15.4	-	19.2	23.8
		P value	-	<0.0001	-	0.0001	-	<0.0001	<0.0001

d.f. denotes difference of number q of covariance parameters between the two compared models
LRT denotes difference of -2lnLL between two compared models
- denotes failure of model to converge and the optimally-fitting spatial model is in bold type
NA = Not Applicable, the estimated value of residual variance for CR/RCBD model is in Italic.

This last for this data set showed that blocking seemed to be unnecessary if a Gaussian covariance structure with nugget effect were used for this trial. As it can also be seen from the above Table 2, except for no nugget Gaussian model all spatial models without block effects had smaller AIC and AICc than a spatial model with block effect. This result is also supported by the result of p-value for block effect ($p > 0.05$) when spatial models with block effect are modeled.

4.3.1.2. Model selection for BW99RVTII6 data set

Like the BW99RVTII5 data set, the semivariogram for the estimated residuals of the RCBiid model for BW99RVTII6 data set (Figure 3 (b)) indicated positive spatial correlation, in that observations that which were closer together tended to be more similar than those farther apart (semivariance increased as lag distance increased). The semivariogram also indicated that spatial covariance structures with nugget effects would be appropriate, but unlike for the above data set the semivariance for this data set is not quite high as lag distance approached zero. The BW99RVTII6 data set candidate covariance models and their related statistics and covariance parameters are summarized in Table 3. From the result of Table 3, all the spatial models (with or without block effect) and RCB models had a smaller AIC and AICc than CR model. The result of LRT also indicates that the spatial models (with or without block effects) had a significant difference in a comparison with the CR model. Like the result of BW99RVTII5 data set, this result also suggested that a significant spatial variability existed in the estimated residuals of the RCBiid model for BW99RVTII6 data set, and that the use of both blocking and spatial model analysis was useful for this trial. Unlike for BW99RVTII5, all the spatial models are not better than the RCB model. As it can be observed in Table 3, a without block no-nugget Spherical model (with or without block effect) and without block no nugget Gaussian model had a greater AIC and AICc than the RCB model. Apart from these three models, all the other spatial models (with or without block effects) had smaller AIC and AICc values than the RCB model (Table 3).

Table 3. BW99RVTII6 data set spatial candidate covariance models and related statistics and covariance parameters including randomized complete block with iid errors (RCBiid).

BW99RVTII6				MODELS						
				CR/RCBD	**No nugget Exponential**	**Nugget Exponential**	**No nugget Spherical**	**Nugget Spherical**	**No nugget Gaussian**	**Nugget Gaussian**
Fitting Statistics	**Without Block(1)**	q		1	2	3	2	3	2	3
		-2lnLL		-19.8	-45.8	-47.9	-39.9	-47.9	-34.0	-47.4
		AIC		-17.6	-41.8	***-41.9***	-35.9	-41.7	-30.0	-41.4
		AICc		-17.7	-41.6	***-41.7***	-35.7	-41.4	-29.8	-40.9
		P value		0.0024	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	With Block(2)	q		1	2	3	2	3	2	3
		-2lnLL		-38.3	-45.1	-45.8	-34.3	-45.4	-42.4	-45.9
		AIC		-36.3	-41.1	-39.8	-30.3	-39.4	-38.4	*-39.9*
		AICc		-36.2	-40.9	-39.4	-30.0	-38.9	-38.2	*-39.5*
		P value	Block	<0.0001	0.0071	0.0596	0.5016	0.0482	0.0002	0.017
			Trt	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Covariance parameter estimates	**Without Block(1)**	Range, m		NA	3.3681m	11.7301m	9175.22M	16313	1.6632	5.0026
		Sill		*0.02652*	0.0274	0.041801	39.8151	22.09549	0.024	0.026347
		Nugget		NA	NA	0.006571	NA	22.0874	NA	0.009477
	With Block(2)	Range, m		NA	1.755m	3.384m	12256m	12.6082m	1.3625m	3.5541m
		Sill		*0.01609*	0.01734	0.01822	53.5488	0.020088	0.0161	0.017567
		Nugget		NA	NA	0.005964	NA	0.008978	NA	0.008741

Likelihood ratio tests	**(1) to CR**	d.f.	-	1	2	1	2	1	2
		LRT	-	26	28.1	20.1	28.1	14.2	27.6
		P-value	-	<0.0001	<0.0001	<0.0001	<0.0001	0.0002	<0.0001
	(2) to CR	d.f.	-	1	2	1	2	1	2
		LRT	-	25.3	26	14.5	25.6	22.6	26.1
		P-value	-	<0.0001	<0.0001	0.0001	<0.0001	<0.0001	<0.0001

d.f. denotes difference of number q of covariance parameters between the two compared models
LRT denotes difference of -2lnLL between two compared models
- denotes failure of model to converge and the optimally-fitting spatial model is in bold type
NA =Not Applicable, the estimated value of residual variance for CR/RCBD model is in Italic.

On the other hand, except for a with block effect no-nugget Gaussian model, all without block effect spatial candidate model had smaller AIC and AICc than with block effect spatial candidate models. But unlike for BW99RVTII5, this conclusion is not strongly supported by the p-values for block effect which for some of them it is less than 0.05. Based on Table 3 above, among a without block effect no-nugget spatial candidate models, the no-nugget exponential model was selected because it had the smallest AIC and AICc value compared with a without block effect no-nugget spatial candidate models. Among a with block effect no-nugget spatial candidate models, a no-nugget Exponential model was selected because it had the smallest AIC and AICc values among others models under this category. Between a with or without block effect no-nugget exponential model, a without block effect no-nugget exponential model with smaller AIC and AICc values was chosen. Among a without block effect nugget spatial candidate models, a nugget Exponential model was selected because of its smaller AIC and AICc values. Of those nugget spatial with block effect candidate models, a nugget Gaussian model with smaller AIC and AICc values was chosen. Between without block effect nugget Exponential model and with block effect nugget Gaussian model, a without block effect nugget Exponential model with smaller AIC and AICc values was chosen. Finally, from a without block effect no-nugget or nugget exponential model, a without block effect nugget Exponential model with smaller values of AIC and AICc was chosen to be the best model for BW99RVTII6 data set. This last model for this data set showed that blocking seemed to be unnecessary if a nugget Exponential covariance structure were used for this trial.

4.3.2. Model selection for the second two data sets (BW99RVTII4 and DW99RVTII1)

Unlike the first two data sets (BW99RVTII5 and BW99RVTII6) in which both block and variety effects are assumed to be fixed, for both of data sets (BW99RVTII4 and DW99RVTII1) considered in this section the block effects are assumed to be random and the variety effect is again fixed. Since there is a random and fixed effect components in the fitted model, those models that were fitted for the two data sets taken under this section are called a mixed effect models, and particularly those models which took spatial variation into account are called spatial

mixed effect models. The model selection procedures for each of the data sets taken in this section are shown under the following sub section.

4.3.2.1. Model selection for BW99RVTII4 data set

As it can be seen in (Fig. 3 (c)), like the above two data sets the semivariogram of the estimated residuals of the RCBiid model for *BW99RVTII4* data set indicates positive spatial correlation, in that observations that were closer together tended to be more similar than those farther apart (semivariance increased as lag distance increased). The semivariogram also indicated that spatial covariance structures with nugget effects would likely be appropriate, as the semivariance remained quite high as the lag distance approached zero. The BW99RVTII4 data set candidate covariance models and their related statistics and covariance parameters are summarized in Table 4. For this data set, a no-nugget Exponential (with or without block effect) models did not reach convergence in this data fitting. This shows these models are not suitable for this data set. Hence, a no-nugget exponential model was excluded. Apart from these models, the other spatial models (both with and without block effects) and the RCB model had markedly smaller AIC and AICc values than the CR model. This result is also supported by the Likelihood ratio test. Thus, apart from the two models, no-nugget Exponential (with or without block effect) models, the LRT indicates that the other spatial models (both with and without block effects) and RCB model had a very significant difference in comparison with the CR model. This result suggested that a significant spatial variability existed in this trial, and that the use of both blocking and spatial model analysis was useful for this trial. But all of the candidate spatial models, both with and without blocking (apart from no-nugget Exponential (with or without block effect) and nugget Spherical with block effect models) had markedly smaller AIC and AICc values than the RCB model showing that spatial models are more useful than the RCB model.

Table 4. BW99RVTII4 data set spatial candidate covariance models and related statistics and covariance parameters including randomized complete block with iid errors (RCBiid).

BW99RVTII4			MODELS						
			CR/RCBD	**No nugget Exponential**	**Nugget Exponential**	**No nugget Spherical**	**Nugget Spherical**	**No nugget Gaussian**	**Nugget Gaussian**
Fitting Statistics	**Without Block(1)**	q	1	2	-	2	3	2	3
		-2lnLL	-13.8	-38.0	-	-33.3	-35.5	-32.3	-37.3
		AIC	-11.8	**-34.0**	-	-29.3	-29.5	-28.3	*-31.3*
		AICc	-11.8	**-33.8**	-	-29.1	-29.1	-28.1	*-30.9*
		P value	0.0729	<0.0001		<0.0001	<0.0001	0.0018	0.0011
	With Block(2)	q	2	3	-	3	4	3	4
		-2lnLL	-32.0	-38.8	-	-33.3	-32.0	-38.8	-39.6
		AIC	-28.0	-32.8	-	-29.3	-26.0	-32.8	-31.6
		AICc	-27.8	-32.4	-	-29.1	-25.6	-32.4	-30.9
		P value	0.0037	0.0010		<0.0001	0.0037	0.0028	0.0013
Covariance parameter estimates	**Without Block(1)**	Range, m	NA	3.527m	-	5318.33m	30.42m	1.9251	3.4495m
		Sill	*0.02929*	0.03210	-	25.7692	0.075028	0.02760	0.029225
		Nugget	NA	NA	-	NA	0.007768	NA	0.0011
	With Block(2)	Range, m	NA	2.0397	-	5318.31m	160.0m	1.6813m	2.6076
		Block	0.01015	0.007482	-	1.12E-15	0.01015	0.009292	0.008386
		Sill	0.01914	0.02207	-	25.7691	0.01914	0.02039	0.021632
		Nugget	NA	NA	-	NA	0.01914	NA	0.006742

Likelihood ratio tests	**(1) to CR**	d.f.	-	1	-	1	2	1	2
		LRT	-	24.2	-	19.5	21.7	18.5	23.5
		p-value	-	<0.0001	-	<0.0001	<0.0001	<0.0001	<0.0001
	(2) to CR	d.f.	1	2	-	2	3	2	3
		LRT	18.2	25	-	19.5	18.2	25	25.8
		p-value	<0.0001	<0.0001	-	0.0001	0.0004	<0.0001	<0.0001
	(2) to (1)	d.f.	1	1	-	1	1	1	1
		LRT01	18.2	-0.8	-	0	3.5	-6.5	-2.3
		p-value	<0.0001	-	-	0.50	0.0307	-	-

d.f. denotes difference of number q of covariance parameters between the two compared models
LRT denotes difference of -2lnLL between two compared models
- denotes failure of model to converge and the optimally-fitting spatial model is in bold type
NA = Not Applicable, the estimated value of residual variance for CR/RCBD model is in Italic.

Among a without block effect no-nugget spatial candidate models, the no-nugget exponential model was selected because it had the smallest AIC and AICc values. Among a no-nugget spatial with block effect candidate models, both no-nugget Exponential and Gaussian models were selected because they had equal smaller AIC and AICc values than a no-nugget spherical model. Since two models were selected among the block effect no-nugget spatial model, a comparison between these models with a selected without-block effect no-nugget spatial model we will use AIC and LRT_{01} methods. To compare without block effect no nugget exponential model and with-block effect no-nugget Gaussian model, AIC and AICc were used because they were based on different spatial covariance structures. In deed a no-nugget Exponential model was selected because it had smaller AIC and AICc values. To compare a with- and without-block effect no-nugget exponential model, the LRT_{01} were used because these models were based on the same spatial covariance structure, thus they were nested. The value of λ for the comparison of these models was -0.8. Clearly, the LRT_{01} was not significant and again a without-block effect no nugget Exponential model was selected. Hence, among a no-nugget spatial model (with- or without-block effect) a without-block effect no nugget Exponential model was selected. On the other hand among a without-block effect nugget spatial candidate models, a nugget Gaussian model were selected because it had a smaller AIC and AICc values than a nugget Spherical model. Among a with-block effect nugget spatial candidate models, a nugget Gaussian model was selected because it had a smallest AIC and AICc value than nugget spatial candidate model. To compare a with and without block effect nugget Gaussian model, again the LRT_{01} criterion was used because these models were based on the same spatial covariance structure, thus they were nested. The value of λ for the comparison of these models was -2.3. Clearly, the LRT_{01} was not significant and a without-block effect Gaussian model was selected. Finally among the selected nugget spatial candidate models and no-nugget spatial model, a without block effect no nugget Exponential model was selected as the best model for BW99RVTII4 data set. Like the first two data sets, this finally selected model for this data set showed that blocking seemed to be unnecessary if a no-nugget Exponential covariance structure were used for this trial.

4.3.2.2. Model selection for DW99RVTII1 data set

As it can be seen in (Figure 3 (d)), like the above three data sets, the semivariogram for the estimated residuals of the RCBiid model for DW99RVTII1 data set indicates positive spatial

correlation. The DW99RVTII1 data set candidate covariance models and related statistics and covariance parameters are summarized in Table 5. A with-block effect no-nugget Spherical model did not reach convergence in this data fitting. This shows that this model is not suitable for the data set. Apart from this model, the other spatial models (both with- and without-block effects) and RCB model had markedly smaller AIC and AICc values than the CR model. The semivariogram (Figure 3 (d)) for the estimated residuals of the RCBiid model for DW99RVTII1 data set seemed to be nearly pure nugget effect. This indicates that if the semivariogram for the estimated residuals of the RCBiid model had been as the only criterion used to select the covariance model for DW99RVTII1 data set analysis, one might have selected the RCBiid model. But using the same principle of selection procedure for the above data sets, a without-block effect no nugget Exponential model was the optimally fitting model for the DW99RVTII1 data set. Furthermore, as it can be seen from the result of LRT_{01} that the spatial models with block effects were not significantly different from those spatial models without block effects (LRT of (2) to (1) in Table 5) showing that the blocking seemed to be unnecessary if spatial models were already used for these trials.

Table 5. DW99RVTII1 data set spatial candidate covariance models and related statistics and covariance parameters including randomized complete block with iid errors (RCBiid).

DW99RVTII1			MODELS						
			CR/RCBD	**No nugget Exponential**	**Nugget Exponential**	**No nugget Spherical**	**Nugget Spherical**	**No nugget Gaussian**	**Nugget Gaussian**
Fitting Statistics	**Without Block(1)**	q	1	2	3	2	3	2	3
		-2lnLL	-18.9	-24.5	-25.4	-12.7	-24.5	-22.0	-25.1
		AIC	-16.9	**-20.5**	-19.4	-8.7	-18.5	-18.0	-19.1
		AICc	-16.9	**-20.3**	-19.0	-8.5	-18.1	-17.8	-18.6
		P value	0.0538	0.0075	0.0386	0.0002	0.0338	0.0446	0.0103
	With Block(2)	q	2	3	4	-	4	3	4
		-2lnLL	-20.9	-25.2	-27.3	-	-27.3	-22.7	-27.1
		AIC	-16.9	-19.2	-19.3	-	-19.3	-16.7	-19.1
		AICc	-16.9	-18.8	-18.6	-	-18.6	-16.2	-18.3
		P value	0.0343	0.0069	0.0015	-	0.0067	0.0460	0.0120
Covariance parameter estimates	**Without Block(1)**	Range, m	NA	1.7224m	3.4197	6956.07m	18.8833m	1.3685m	3.8055m
		Sill	*0.02649*	0.02782	0.02822	47.0448	0.03023	0.02643	0.02746
		Nugget	NA	NA	0.01145	NA	0.01874	NA	0.01576
	With Block(2)	Range, m	NA	1.6719m	7.1466m	-	20.1628m	1.2969m	7.5206m
		Block	0.002213	0.001892	0.002794	-	0.002933	0.001771	0.002914
		Sill	*0.02430*	0.02643	0.02729	-	0.02318	0.02484	0.02542
		Nugget	NA	NA	0.01538	-	0.01734	NA	0.01856

Likelihood ratio tests	**(1) to CR**	d.f.	-	1	2	1	2	1	2
		LRT	-	5.6	6.5	-6.2	5.6	3.1	6.2
		P value	-	0.0180	0.0388	-	0.0608	0.0783	0.0450
	(2) to CR	d.f.	1	2	3	-	3	2	3
		LRT	2	6.3	8.4	-	8.4	3.8	8.2
		P value	0.1573	0.04290	0.0384	-	0.0384	0.1496	0.0421
	(2) to (1)	d.f.	1	1	1	-	1	1	1
		LRT_{01}	2	0.7	1.9	-	2.8	0.7	2
		p-value	0.07865	0.2014	0.08405	-	0.04715	0.2014	0.07865

d.f. denotes difference of number q of covariance parameters between the two compared models
LRT denotes difference of -2lnLL between two compared models
- denotes failure of model to converge and the optimally-fitting spatial model is in bold type
NA = Not Applicable, the estimated value of residual variance for CR/RCBD model is in Italic.

4.4. Semivariograms, Spatial Ranges, and Selected Covariance Models

For BW99RVTII5 and BW99RVTII6 data sets, the estimated ranges of spatial dependence of the errors from the selected model (Without block effect nugget Gaussian and without block effect nugget Exponential respectively, Tables 2 and 3) were approximately 3.28m and 11.73m respectively, which was less than the maximum distance within a block (~31.8 m) but greater than the plot dimensions (~2.8 m). This indicated that spatially correlated errors exist within and among blocks, probably due to the omission of one or more unknown covariates from the model. In this case, the block effect would not be expected to be fully effective in explaining the spatial structure. Additionally, semivariograms for the estimated residuals of the RCBiid model for both of these data sets indicate that a spatial covariance structure should include a nugget effect. These facts are consistent with the selected model for these data sets, and indicated that the spatial correlation should have been accounted for by ignoring blocking and including an estimation of the spatial covariance structure of the errors in the model.

For the BW99RVTII4 data set, the estimated range of spatial dependence of the errors from the selected model (Without block effect no nugget Exponential, Table 4) was ~3.53m, which was less than the maximum distance within a block (~31.8 m) but greater than the plot dimensions (~2.8 m). This indicates that, like the first two data sets, spatially correlated errors exist within and among blocks, probably due to the omission of one or more unknown covariates from the model. In this case, the random block effect would not be expected to be fully effective in explaining the spatial structure.

For the DW99RVTII1 data set, the estimated range of spatial dependence of the errors from the selected model (Without block effect no nugget Exponential, Table 5) was ~1.72m, which is slightly less than the plot size, indicating that spatially correlated errors were present within the scale of adjacent plots. In such a case, a random block is not expected to be effective in accounting for spatial heterogeneity. In this case, a spatial covariance function alone would be better to account for the spatial correlation.

4.5. Model Impact on P-Values

In all of the four trials, the errors were spatially correlated, the estimated range for the selected spatial model was less than the block size, and as it can be seen in Table 6 below, the RCBiid method tended to give higher estimates of the standard errors of estimated varieties than the other model.

Table 6: Minimum, maximum and average standard error (SE) and the degrees of freedom (D.F.) of estimated varieties in a selected model for each of data set and their corresponding RCBDiid model.

		Selected model for the each of data sets and their corresponding RCBDiid model							
		BW99RVTII5		BW99RVTII6		BW99RVTII4		DW99RVTII1	
		RCBDiid	WBNG	RCBDiid	WBNNE	RCBDiid	WBNNE	RCBDiid	WBNNE
SE	Max.	0.1519	0.1120	0.0897	0.0621	0.0978	0.0891	0.1103	0.1082
	Min.	0.1519	0.0962	0.0897	0.0564	0.0978	0.0791	0.1103	0.1006
	Aver.	0.1519	0.1045	0.0897	0.0594	0.0978	0.0849	0.1103	0.1040
D.F.	Max.	57	43.8	57	54.3	57	51.5	54	54.8
	Min.	57	32.6	57	42.5	57	45.6	54	44.8
	Aver.	57	37.61	57	49.0	57	48.7	54	49.1

WBNG denotes a without-block effect nugget Gaussian model
WBNNE denotes a without-block effect no-nugget Exponential model

Consequently, for all of the data sets, the RCBiid and the selected covariance model resulted in different conclusions regarding treatment significance. For instance, the selected spatial covariance model for DW99RVTII4 and DWRVTII1 resulted in a smaller significance level for variety comparison, $p < 0.0001$ and $p = 0.0112$ compared with $p = 0.0037$ and $p = 0.0356$ for the RCBiid model, respectively. The data in Tables 3 to 5, indicate that the varieties p-value associated with any given candidate covariance model, except for BW99RVTII6 data set, is very sensitive to the form of that model. This shows that, arbitrary selection of a covariance model can result in dramatically different treatment conclusions. The inclusion or omission of block effect and nugget effect in the models, except for BW99RVTII6 data set, also affected the p-value of the F-test of different varieties (Tables 3-5). For instance, for the DW99RVTII1 data set, the p-values for models with-block effects and no-nugget spatial model range from 0.0069 to 0.0460. But for a nugget spatial candidate models with-block effect, the p-values range from

0.0015 to 0.0.0120. If no-nugget and no-block effects were included, the p-value dropped dramatically to 0.0002 for the no-nugget Spherical model without block effect.

4.6. The impact of the selected spatial model on detecting differences among treatments or variety contrasts

As can be seen in Table 7 below, again the superiority of spatial models is clearly shown in detecting a difference among treatment contrasts.

Table 7. Statistical test for some of the treatment contrasts on the bases of selected spatial model and RCBDiid model for each data set.

Trials	Source[1]	RCBDiid			Selected spatial model		
		DDF[2]	F[3]	P value	DDF	F	P value
BW99RVTII4	1 vs 3	57	1.6245	0.2076	49.9	7.1013	0.0103
	1 vs 4	57	2.329	0.1325	50.3	6.0821	0.0171
	1 vs 10	57	3.6759	0.0602	45.6	13.2736	0.0006
BW99RVTII5	1 vs 2	57	0.0829	0.7744	38.7	4.9480	0.0320
	1 vs 3	57	1.624	0.2076	42.1	6.5230	0.0143
	1 vs 5	57	0.6052	0.4398	41.7	5.9991	0.0185
BW99RVTII6	1 vs 8	57	8.8485	0.0042	51.9	11.4461	0.0013
	1 vs 15	57	3.6246	0.0619	51.3	7.7845	0.0073
	2 vs 18	57	6.4444	0.0138	48.3	9.6277	0.0031
DW99RVTII1	2 vs 12	54	4.7630	0.0334	43.4	7.5008	0.0089
	4 vs 18	54	4.2603	0.0438	49.9	12.3267	0.0009
	5 vs 12	54	8.4238	0.0053	43.9	9.2539	0.0039

[1] the selected treatment contrasts for the illustration from each trials.
[2] Denominator degrees of freedom
[3] F calculated

4.7. Discussion

We have used a semivariogram for the initial estimate of the parameters of the spatial covariance structure for each of the data sets in this study. Some of the spatial candidate models did not converge and these all are removed from the analysis with the assumption that these models are not suitable for the given data sets. Exponential with nugget and Spherical with nugget models, a no-nugget Exponential (with or without block effect) models and with block effect no-nugget Spherical model did not reach convergence for BW99RVTII5, BW99RVTII4 and DW99RVTII1 data sets, respectively. Some of the estimated parameters of the spatial covariance are also unrealistic, thus, the estimated range for some of the spatial candidate models are unrealistic. For instance, for the BW99RVTII6 data set, an estimated range for a nugget spherical and no-nugget spherical model both without-block effect were 16313m and 9175.22m, respectively. These are greater than the block dimension (Table 3) and also for the BW99RVTII4 data set, an estimated range for a without block effect no-nugget spherical model and a no-nugget spherical model with block effect were 5318.33m and 5318.31m, respectively (Table 4). If we also consider the DW99RVTII1 data set, the estimated range for a without-block no-nugget spherical model was 6956.07m (Table 5). The corresponding estimated sill was also unrealistically large. The model selection tool also indicates that these models have poor fit spatial models among the candidate spatial models.

A without block effect nugget Gaussian model and nugget Exponential model were selected for the first two data sets, BW99RVTII5 and BW99RVTII6, respectively and the no nugget exponential without block effect was selected for the second two data sets, BW99RVTII4 and DW99RVTII1. It was found that in all of the four trials, the estimated residual from RCBiid model were significantly spatially correlated, providing evidence of field heterogeneity within blocks that could make the RCBiid method less powerful than a method that incorporated spatial correlation. These finally selected models for the data sets showed that the blocking seemed to be unnecessary if these selected spatial model were used for each of the trials. For instance, for the first two data sets (BW99RVTII5 and BW99RVTII6) in which block effects are constant, the p-value is significant for block effect after spatial variation were taken into account. This also supports the above idea. For each of the data sets the estimated parameters of the selected semivariogram model did match with the visual inspection of the semivariograms.

We have also showed that the p-value for the fixed treatment effect was smaller under selected spatial models than the traditional RCBD models for each of the data sets except for BW99RVTII6 (Tables 2 - 5). For instance, for the DW99RVTII1 data set, the p-values for models with-block effects and no-nugget spatial model range from 0.0069 to 0.0460. But for a nugget spatial candidate models with-block effect, the p-values range from 0.0015 to 0.0.0120. If no-nugget and no-block effects were include, the p-value dropped dramatically to 0.0002 for the no-nugget Spherical model without-block effect.

The selected spatial models also have a smaller standard error than the RCBDiid model (Table 6). The selected spatial models for all of the four data sets under consideration also have higher power of detecting difference among treatment contrasts (Table 7). For instance, for the BW99RVTII4 data set, while the analysis under RCBDiid model did not detect any difference among the 1^{st} and 3^{rd} , the 1^{st} and 4^{th} and the 1^{st} and 10^{th} varieties (p value > 0.05), The selected spatial model analysis for this data set showed that two of the three contrasts were significant (p-value <0.025). In general, if data sets have a significant spatial dependence, fitting the model by taking spatial variation into account is important.

CHAPTER FIVE

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

The importance of spatial variability to be expected from a logical and subject-related perspective is confirmed in a variety of experiments. As presented in much of the literature, spatial analysis may lead to improvement in efficiency with regard to standard error for the estimation of fixed effects than non-spatial analysis provided that spatial variability is present. This study focused on comparisons of linear fixed and mixed models with various spatial covariances using Kenward-Roger methods for approximating denominator degrees of freedom. With this framework of the fixed and mixed models, spatial analysis is convenient in practice with regard to providing different covariance structures for a wide range of spatial variability, model selection by likelihood-based criteria, and the availability of programs for realization, e.g., SAS software. Semivariogram were used as a special tool for spatial analysis. Most of the investigated spatial models showed better fit and smaller standard error for treatment contrasts than the RCBDiid model. A suitable model can take spatial variability into account well, and effectively adjust the location effect for treatment effect estimation. Therefore, the selected use of an optimal model is important for analyzing plant cultivar trials. In general, from this study we found that:

- Randomization does not completely remove spatial correlation.
- Blocking also does not completely control the spatial variation.
- Spatial-modeling approach can give more precise inference.
- The selected model for all data set showed that the blocking would be unnecessary if these selected spatial model were used for each of the trials.
- In the presence of spatial dependence among observations of variety trials, the spatial modeling provides an increased power in statistical tests, reducing standard errors of treatment/variety estimates.

5.2. Recommendations

- Since the spatial models have greater efficiency than the conventional analysis of variance, researchers should take into account spatial variation in field experiment in situations were spatial dependence among the observations is significant.
- Since the arbitrary selection of a covariance model can result in dramatically different treatment conclusions, agronomist is advised to use the procedure of model selection that have been used in this study. Thus, first, use the semivariogram to check the presence of spatial dependence in the data and identify the possible covariance structure of the data; second, use this semivariogram as an initial estimate of the covariance parameter for the selected possible covariance structure in the first step; finally, use the information criterion and LRT to select an optimally fitting spatial model.
- For all data sets used in this study, block effects become insignificant after the inclusion of spatial covariance structure in the model, this result raises the question "Does really the design matters?". Hence, we recommend that further study be undertaken to better understand the effect of spatial model taking different designs other than RCBDiid.
- We have only used information criterion and Likelihood ratio test criterion for model selection. But there are also some other criteria used to select a valid variogram model such as cross validation, so, we recommend that further studies should be done in this area.

REFERENCES

Atkinson, A. C. and Bailey, R. A. 2001. One hundred years of the design of experiments on and off the pages of "Biometrika". *Biometrika*, **88** (1): 53-97.

Bartlett, M.S. 1978. Nearest Neighbor Analysis Models in the Analysis of Field Experiments. *Journal of the Royal Statistical Society* **40**(2): 147-174.

Bartlett, M.S. 1938. The approximate recovery of information from field experiments with large blocks. *Journal of Agricultural Science* **28**: 418-427.

Basu, S. and Reinsel, G. C. 1994. Regression Models with Spatially Correlated Errors. *Journal of the American Statistical Association*, **89**, (425): 88- 99.

Berke, O. 1999. On Spatiotemporal Prediction for Online Monitoring Data. Research Report 96/9, University of Dortmund, Department of Statistics. Submitted for publication to Communications in Statistics – Theory and Methods.

Besag, J. E. and Kempton, R. A. 1986. Statistical Analysis of Field Experiments Using Neighbouring Plots. Biometrics, 26: 327-333.

Brownie, C., Bowman, D. T. and Burton, J. W. 1993. Estimating Spatial Variation in Analysis of Data from Yield Trials: A Comparison of Methods. *Agronomy Journal* ***85***: 1244- 1253.

Brownie, C. and Gumpertz, M. L. 1997. Validity of spatial analyses for large field trials. *J. Agric. Biol. Environ. Stat.*, **2**:1–23.

Clark, I. 1979. Practical Geostatistics. Applied Science Publishers, Essex, England.

Cressie, N. A. C. and Hawkins, D.M. 1980. Robust estimation of the variogram I. *Journal of the international association of Mathematical Geology* **2**: 115-125

Cressie, N. A. C. 1991. Statistics for Spatial Data. John Wiley & Sons: New York.

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

Cullis, B. R. and Gleeson, A. C. 1991. Spatial analysis of field experiments – an extension to two dimensions. *Biometrics* **47**, 1449—1460.

Curriero, F. C. and Lele, S. 1999. A composite likelihood approach to semivariogram estimation. Journal of Computational and Graphical Statistics. **4**:9–28.

Doreian, P. 1981. Estimating Linear Models with Spatially Distributed Data. *Sociological Methodology*, **12**: 359 – 388

Duarte, J. B. and Vencovsky, R. 2005. Spatial statistical analysis and selection of genotypes in plant breeding. Pesq. agropec. bras., Brasília, **40**: 107-114.

Faground, M. and Merivenne, M.V. 2002. Accounting for Soil Spatial Autocorrelation in the Design of Experimental Trials. *Soil Sci. Soc. Am. J.*, **66**: 1134–1142.

Fai, A. H. T. and Cornelius, P. L. 1996. Approximate F-tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments. *Journal of Statistical Computation and Simulation* **54**: 363–378.

Fisher, R. A. 1926. The arrangement of field experiments. *J. Min. Agric. G. Br.*, **33**: 503–513.

Fuentes, M. 2008. SPATIAL STATISTICS COURSE: Handout with examples *http://www.stat.ncsu.edu/people/fuentes/courses/barcelonaspatial/lectures/handout.pdf.* Universitat Politecnica de Catalunya.

Giesbrecht, F. G. and Burns, J. C. 1985. Two-Stage Analysis Based on a Mixed Model: Large-sample Asymptotic Theory and Small-Sample Simulation Results. *Biometrics*, **41**: 477–486.

Gilmour, A. R., Cullis, B. R. and Verbyla, A. P. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric. Biol. Environ. Stat.* **2**: 269—293.

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

Gleeson, A. C. and Cullis, B. R. 1987. Residual maximum likelihood (REML) estimation of a neighbor model for field experiments. *Biometrics* **43**: 277-288.

Gonçalves, E. St.Aubyn, A. and Martins, A. 2007. Mixed spatial models for data analysis of yield on large grapevine selection field trials. *Theor Appl Genet* **115**: 653–663.

Grondona, M. O. and Cressie, N. A. C. 1991. Using Spatial Considerations in the Analysis of Experiments. *Technometrics*, **33**, (4): 381-392.

Grondona, M. O., Crossa, J., Fox, P. N. and Pfeiffer, W. H. 1996. Analysis of Variety Yield Trials Using Two-Dimensional Separable ARIMA Processes. *Biometrics* **52**, (2): 763 – 770

Green, P. J., Jennison, C. and Seheult, A. H. 1985. Analysis of field experiments by least squares smoothing. *Journal of Royal Statistical Society, Series B,* **47:** 299-355.

Guerin, L. and Stroup, W.W. 2000. A simulation study to evaluate PROC MIXED analysis of repeated measures data. Proc. Ann. Conf. Applied Stat. Agric., 12th, Manhattan, KS. Kansas State Univ., Manhattan, KS.

Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding, University of Guelph.

Hoeting, J. A., Davis, R. A., Metron, A. A. and Thompson, S.E. 2006. Model selection for geostatistical models. *Ecological Applications* **16**: 87–98.

Hurvich, C. M. and Tsai, C. L. 1989. Regression and Time Series Model Selection in Small Samples. *Biometrika*, **76**: 297–307.

Jayet, H., Legallo, J. and Anselin, L. 2004. Spatial panel econometrics, Chapter 18 of the volume, Baltagi, B. ed.

Kackar, R. N. and Harville, D. A. 1984. Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association* **79**: 853–862.

Kempton, R. A. and Howes, C. W. 1981. The use of neighboring plot values in the analysis of variety trials. *Applied Statistics* **30**: 59-70.

Kempton, R. A. Gregory, R. S. Hughes, W. G. and Stoehr, P. J. 1986. The effect of inter-plot competition on yield assessment in triticale trials. *Euphytica* **35**: 257-267

Kempton, R. A., Seraphin, J. C. and Sword, A.M. 1994. Statistical analysis of two-dimensional variation in variety yield trials. *Journal of Agricultural Science* **122**: 335-342.

Kenward, M. G., and Roger, J. H. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**: 983–997.

Kinfemikael Alemu, 2000. Application of spatial statistical method, to the study of soil fertility pattern. Unpublished Msc Thesis. Addis Ababa University.

Lambert, D., Lowenberg-DeBoer, J. and Bongiovanni. R. 2002. Spatial Regression, An Alternative Statistical Analysis for Landscape Scale On-farm trials: Case Study of Variable Rate Nitrogen Application in Argentina. In: Proceedings of the 6th International Conference on Precision Agriculture. July 14-17, 2002, Bloomington, MN, (ASA-CSSA-SSSA, Madison, Wisconsin, 2002).

Lin, S. C., Lin, C. R., Gukovsky, I., Lusis, A. J., Sawchenko, P. E. and Rosenfeld, M. G. 1993. Molecular basis of the little mouse phenotype and implications for cell type-specific growth. **364**, (6434): 208-13.

Littell, R. C., Milliken, G. A., Stroup, W.W. and Wolfinger, R. D. 1996. SAS system for mixed models. SAS Inst., Cary, NC.

Matheron, G. 1962. Traite de Geostatistiquem Appliquee, Tome I. Memories du Bureau de Recherches Geologiques et Minieres. **14**: 9-34.

Matheron, G. 1963. Principles of geostatistics. *Econ. Geol.* **58**: 1246-1266.

Matheron, G. 1971. The Theory of Regionalized Variables and Its Applications. Cahiers du center de Morphologie mathematique. **5**: 245-260.

Mendez, I. 1970. Study of uniformity trials and six proposals as alternative to blocking for the design and analysis of field experiments. Institute of Statistics Mimeo Series 696, North Carolina State University, Raleigh.

Menezes, R., Garcia-Soidan, P. and Febrero-Bande, M. 2005. A comparison of approaches for valid variogram achievement. Computational Statistics **20**: 623-642.

Morgan, J.E. and Gumpertz, M.L. 1996. Random effects models: Testing whether variance components are zero. Proc. Joint Stat. Meetings, Chicago, IL. 4–8 Aug. 1996. Am. Stat. Assoc., Alexandria, VA.

Murali, H. 2010. Lecture note on Spatial Models. *http://www.astrostatistics.psu.edu/su10/ lectures /su 10notebook.pdf*

Nan Hong, J., White, G., Gumpertz, M. L. and Weisz, R. 2005. Spatial Analysis of Precision Agriculture Treatments in Randomized Complete Blocks: Guidelines for Covariance Model Selection. Agronomy Journal, Vol. 98.

Papadakis, J.S. 1937. Méthode Statistique Pour Des Experiences Sur Champs. Bulletin del'Institut de l'Amelioration des Plantes Thessalonique, 23.

Patterson, H. D. and Thompson, R. 1971. Recovery of interblock information when block sizes are unequal. *Biometrika,* **58**: 545-554.

SAS Institute Inc., 2008. SAS/STAT User's Guide, Version 9.2, Cary, NC: SAS Institute Inc

Satterthwaite, F. E. 1941. Synthesis of variance. *Psychometrika,* **6**: 309-316.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat*, **6**: 461–464.

Schabenberger, O. and Pierce, F. J. 2002. Contemporary Statistical Models for the Plant and Soil Sciences (CRC Press, Boca Raton, FL).

Stein, M. L. 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York: Springer.

Stroup, W. W., Baenziger, P. S. and Mulitze, D. K. 1994. Removing Spatial Variation from Wheat Yield Trials: A Comparison Of Methods. *Crop Science* **34**: 62-66.

Stroup, W. W. 2002. Power analysis based on spatial effects mixed models: a tool for comparing design and analysis strategies in the presence of spatial variability. *Journal of Agricultural, Biological, and Environmental Statistics*, **7**: 491–511.

Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Commu-nications in Statistics Part A Theory and Methods, 7:13- 26.

Tamura, R. N., Nelson, L. A. and Naderman, G. C. 1988. An Investigation of The Validity and Usefulness of Trend Analysis for Field Plot Data. *Agronomy Journal*, **80**: 712- 718.

Thompson, R. 1977. Estimation of quantitative genetic parameters. *Proc. Int. Conf. Quant. Genet.*, p 639.

Underwood, A. J. 1997. Experiments in ecology: their logical design and interpretation using analysis of variance. Cambridge University Press, Cambridge, UK.

Van Es, H.M. and Van Es, C.L. 1993. Spatial nature of randomization and its effect on the outcome of field experiments. *Agronomy Journal,* **85**: 420-428.

van Es, H.M., 2002. Sources of soil variability. In: Dane, J., Topp, C. (Eds.), Methods of Soil Analysis, Part 4: Physical Properties. *Soil Sci. Soc. Am.*, Madison WI.

Verbeke, G. and Molenberghs, G. 2000. Linear mixed models for longitudinal data. Springer, New York.

Vieira, S.R. 2000. Uso de geoestatística em estudos de variabilidade espacial de propriedades do solo. In: NOVAIS, R.F. (Ed.). *Tópicosem Ciência do Solo. Viçosa. SBCS*, p.1-54.

Werneck, G. L. and Maguire, J. H. 2002. Spatial modeling using mixed models: an ecologic study of visceral leishmaniasis in Teresina, Piauí State, Brazil. Cad. *Saúde Pública, Rio de Janeiro*, **18** (3):633-637.

Wilkinson, G. N., Eckert, S. R., Hancock, T. W. and Mayo, O. 1983. Nearest-neighbour (NN) analysis of field experiments (With Discussion). *Journal of Royal Statistical Society, Series B* **45**: 151-211

Williams, R. M. 1952. Experimental Designs for Serially correlated Observations, *Biometrika*, **39**: 151-167.

Williams, E. R. 1986a. A neighbour model for field experiments. *Biometrika, 73*: 279-287.

Xiyuan H. and Joachmim, S. 2009. Comparison of various spatial models for the analysis of cultivar trials. *New Zealand Journal of Agricultural Research*, **52**: 277-287.

Yates, F. 1938. The comparative advantages of systematic and randomized arrangements in the design agricultural and biological experiments. *Biometrika,* **30**: 444-466.

Yates, F. 1939. The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann. Eugen.,* **10**:317–325.

Zimmerman, D. L. and Harville, D. A. 1991. A random field approach to the analysis of field plot experiments and other spatial experiments. *Biometrics,* **47**:223–239.

APPENDICES

APPENDIX I

Table 8. The list of BW99RVTII4 data taken from Ethiopian Agricultural Research Organization (EARO).

Loc[1]	Entry[2]	Rep[3]	North[4]	East[5]	Trnyldkg[6]
4	1	1	2.5	1.2	1.19223
4	2	1	2.5	2.8	1.25214
4	3	1	2.5	4.4	1.10406
4	4	1	2.5	6	1.24683
4	5	1	2.5	7.6	1.1163
4	6	1	2.5	9.2	1.31448
4	7	1	2.5	10.8	1.38618
4	8	1	2.5	12.4	0.96372
4	9	1	2.5	14	1.05519
4	10	1	2.5	15.6	0.93897
4	11	1	2.5	17.2	1.24407
4	12	1	2.5	18.8	1.08006
4	13	1	2.5	20.4	0.87837
4	14	1	2.5	22	0.95181
4	15	1	2.5	23.6	1.12923
4	16	1	2.5	25.2	1.19439
4	17	1	2.5	26.8	1.37079
4	18	1	2.5	28.4	1.11795
4	19	1	2.5	30	1.06083
4	20	1	2.5	31.6	0.5982
4	17	2	6	1.2	0.96315
4	18	2	6	2.8	1.22655
4	2	2	6	4.4	1.33428
4	1	2	6	6	1.21092
4	11	2	6	7.6	0.9933
4	20	2	6	9.2	0.834
4	15	2	6	10.8	1.16139
4	10	2	6	12.4	0.96909
4	6	2	6	14	1.26492
4	13	2	6	15.6	1.26306
4	3	2	6	17.2	1.19709
4	9	2	6	18.8	1.00335
4	7	2	6	20.4	1.16799
4	14	2	6	22	1.11702
4	4	2	6	23.6	1.45665
4	16	2	6	25.2	1.24578
4	8	2	6	26.8	1.1958
4	19	2	6	28.4	1.40355
4	12	2	6	30	1.15674
4	5	2	6	31.6	1.18893
4	7	3	9.5	1.2	1.50873

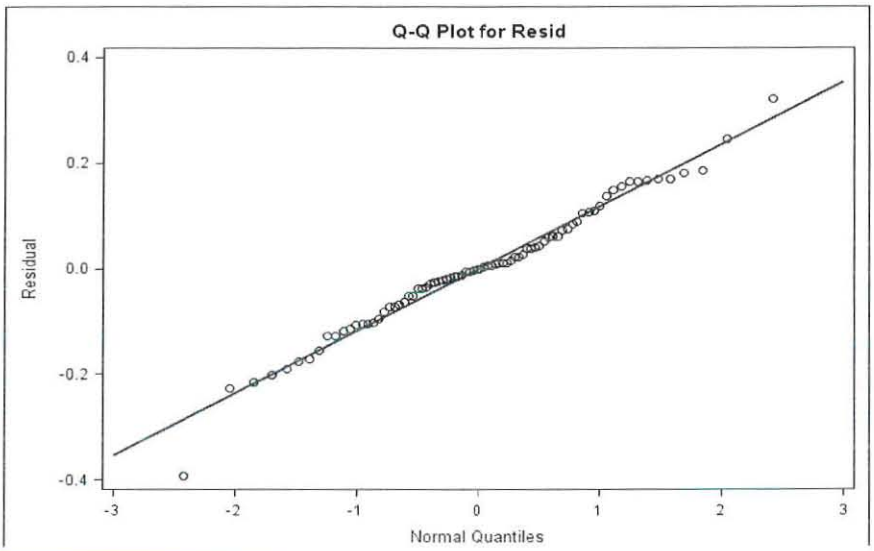
4	18	3	9.5	2.8	1.30551
4	20	3	9.5	4.4	0.96009
4	17	3	9.5	6	1.19313
4	13	3	9.5	7.6	1.24824
4	5	3	9.5	9.2	1.27143
4	12	3	9.5	10.8	1.2984
4	8	3	9.5	12.4	1.28733
4	10	3	9.5	14	1.28445
4	2	3	9.5	15.6	1.42425
4	6	3	9.5	17.2	1.46925
4	19	3	9.5	18.8	1.62126
4	16	3	9.5	20.4	1.53294
4	14	3	9.5	22	1.39626
4	9	3	9.5	23.6	1.43559
4	3	3	9.5	25.2	1.27431
4	15	3	9.5	26.8	1.14534
4	11	3	9.5	28.4	1.158
4	1	3	9.5	30	1.44753
4	4	3	9.5	31.6	0.86952
4	8	4	13	1.2	1.12695
4	11	4	13	2.8	1.27158
4	6	4	13	4.4	1.43334
4	20	4	13	6	1.13727
4	14	4	13	7.6	1.28595
4	3	4	13	9.2	1.27863
4	5	4	13	10.8	1.47174
4	10	4	13	12.4	1.41009
4	1	4	13	14	1.50213
4	12	4	13	15.6	1.32705
4	7	4	13	17.2	1.24656
4	19	4	13	18.8	1.38981
4	17	4	13	20.4	1.40769
4	9	4	13	22	1.36182
4	13	4	13	23.6	1.45791
4	4	4	13	25.2	1.18266
4	16	4	13	26.8	1.19025
4	15	4	13	28.4	1.26933
4	18	4	13	30	1.30791
4	2	4	13	31.6	1.45482

[1] Location of the experimental site
[2] The variety.
[3] Replication/Block
[4] Spatial position of each plot in the North direction
[5] Spatial position of each plot in the East direction
[6] Yield of varieties within each plots in kg.

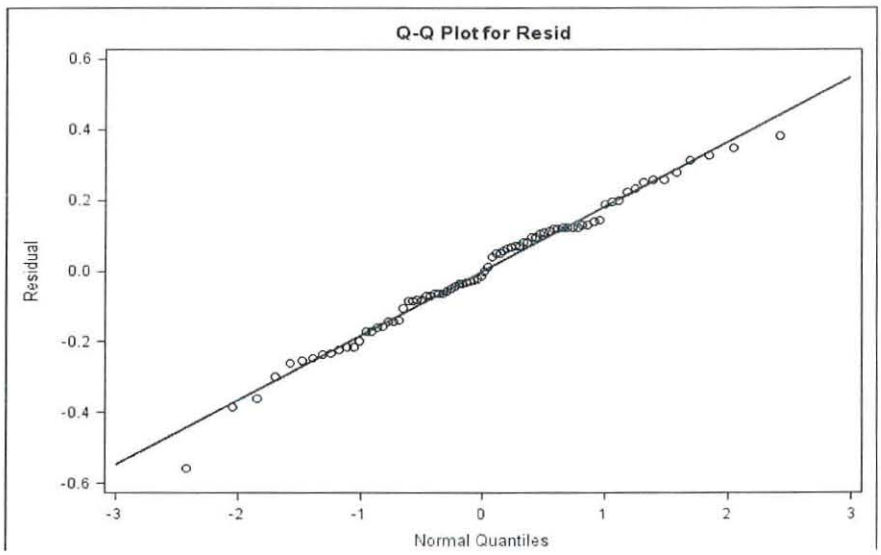
APPENDIX II

Diagnosis for the assumption of normality for each data sets.

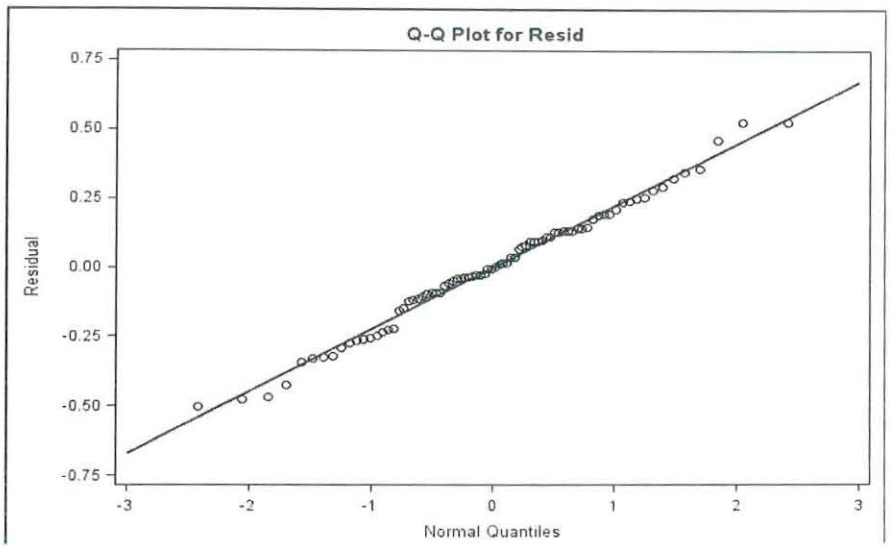
(A) QQ Plots



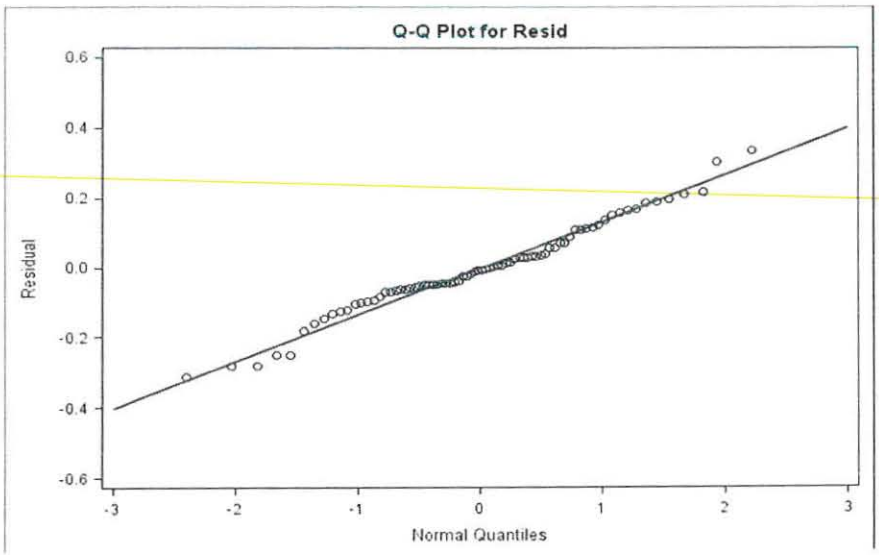
QQ plots for BW99RVTII4 data



QQ plot for BW99RVTII5 data



QQ plot for BW99RVTII6 data



QQ plot for DW99RVTII1 data

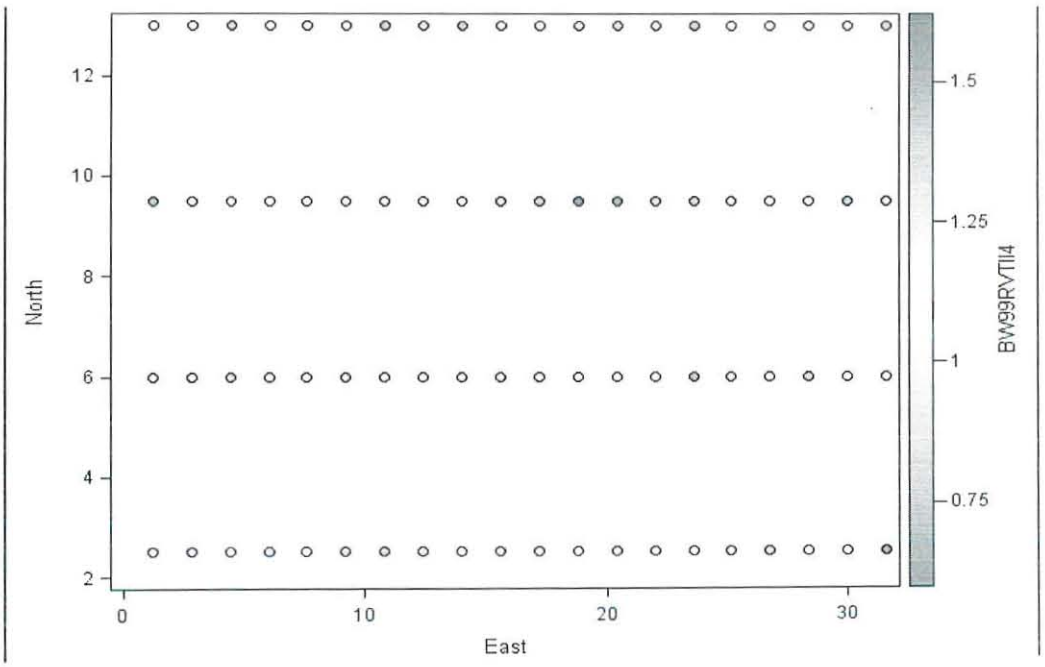
Figure 5. QQ plot for all of data sets

(B)

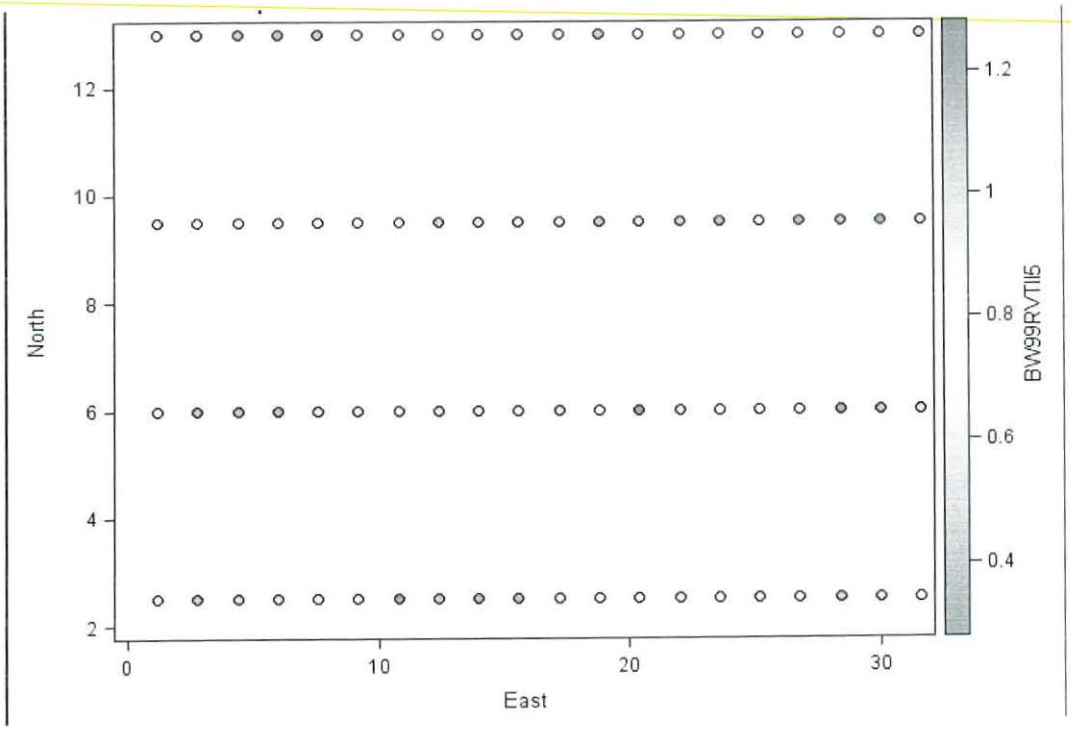
Table 9. The Shapirio Wilks test for the assumption of normality for each data set.

Data	Test	Statistics	P-Value
BW99RVTII4	Shapiro-Wilk	0.983667	0.4011
BW99RVTII5	Shapiro-Wilk	0.987121	0.6064
BW99RVTII6	Shapiro-Wilk	0.988063	0.6688
DW99RVTII1	Shapiro-Wilk	0.971242	0.0817

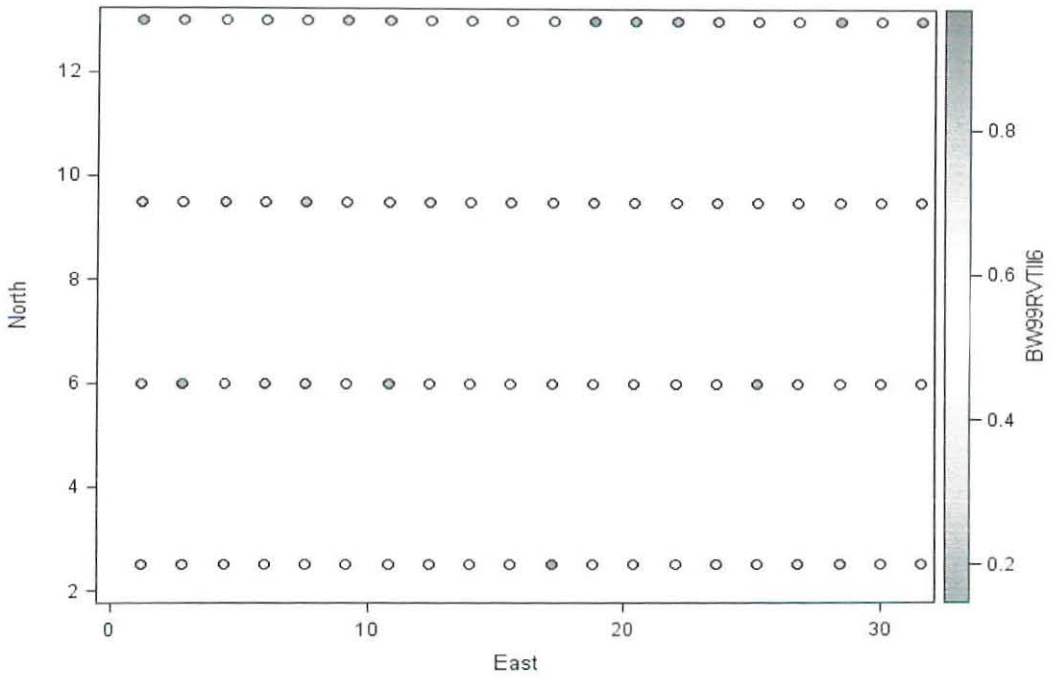
APPENDIX III.



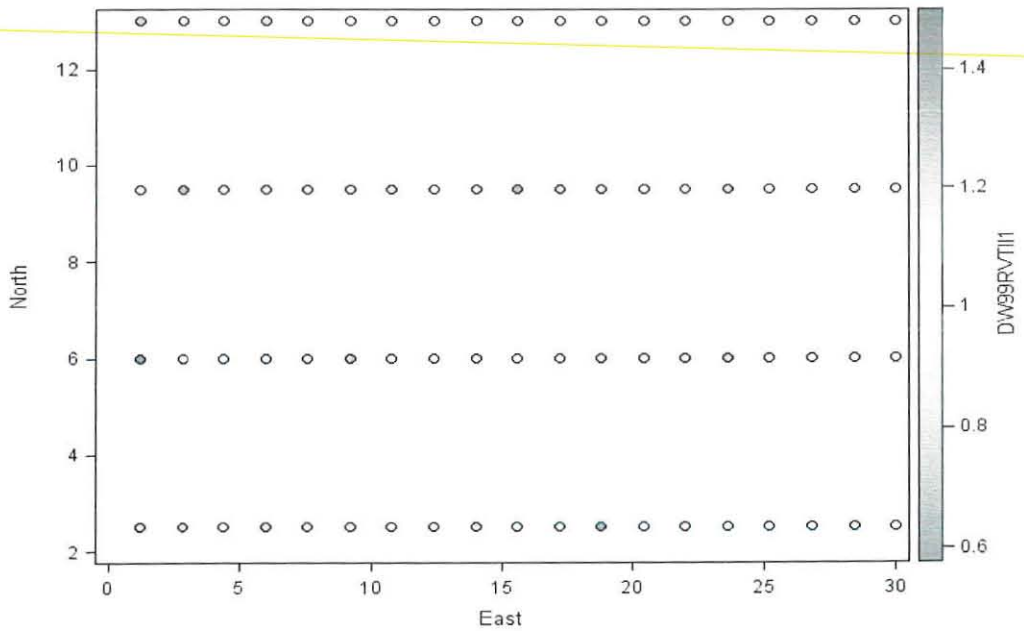
(a) Spatial distribution of BW99RVTII4 data set.



(b) Spatial distribution of BW99RVTII5



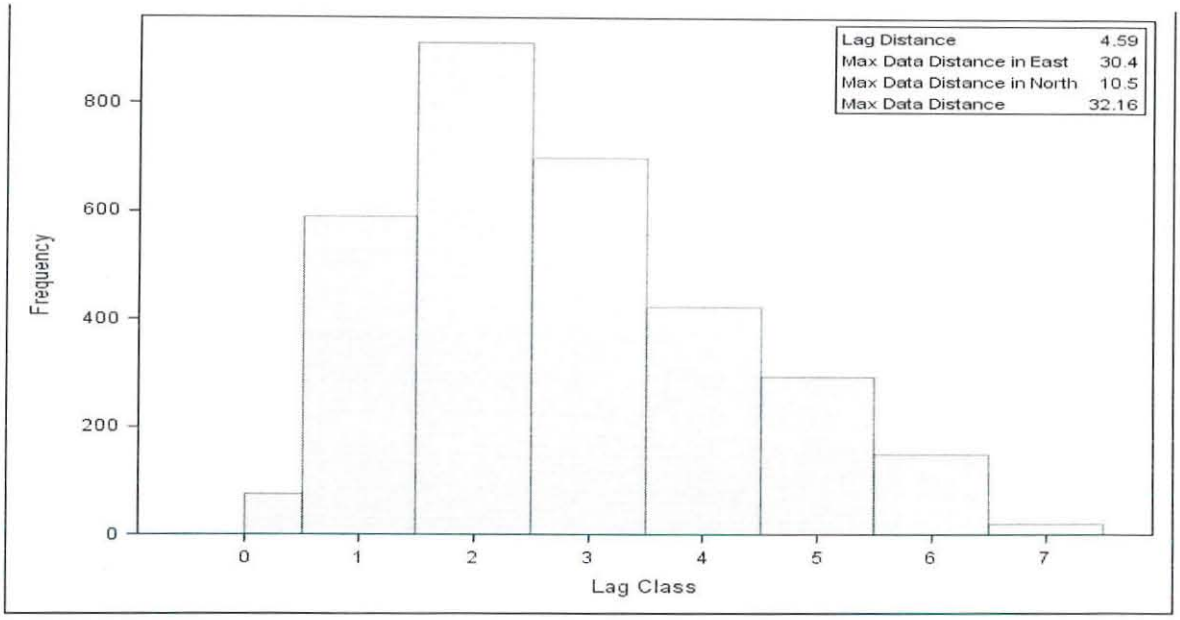
(c) Spatial distribution of BW99RVTII6



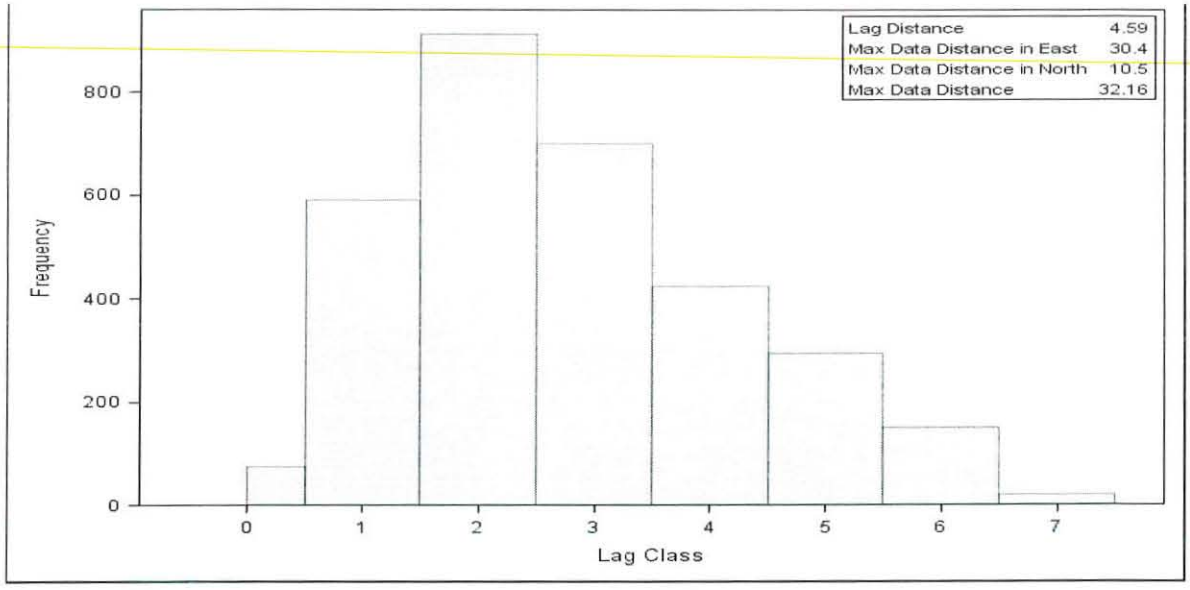
(d) Spatial distribution of DW99RVTII1

Figure 6. Spatial distribution of the data sets considered in this study

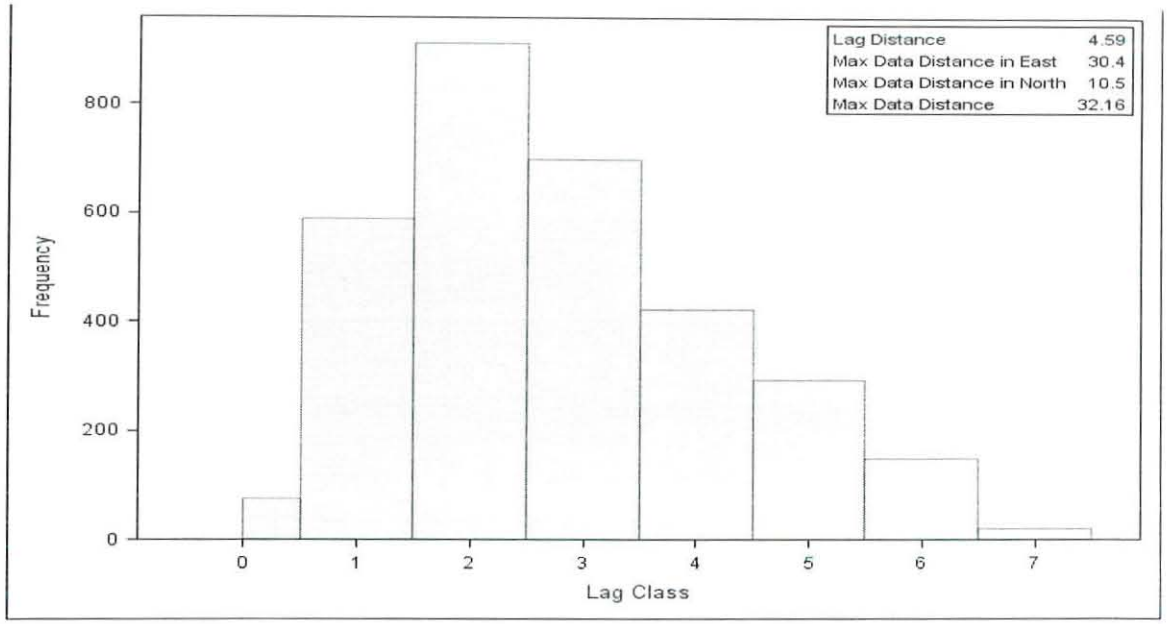
APPENDIX IV



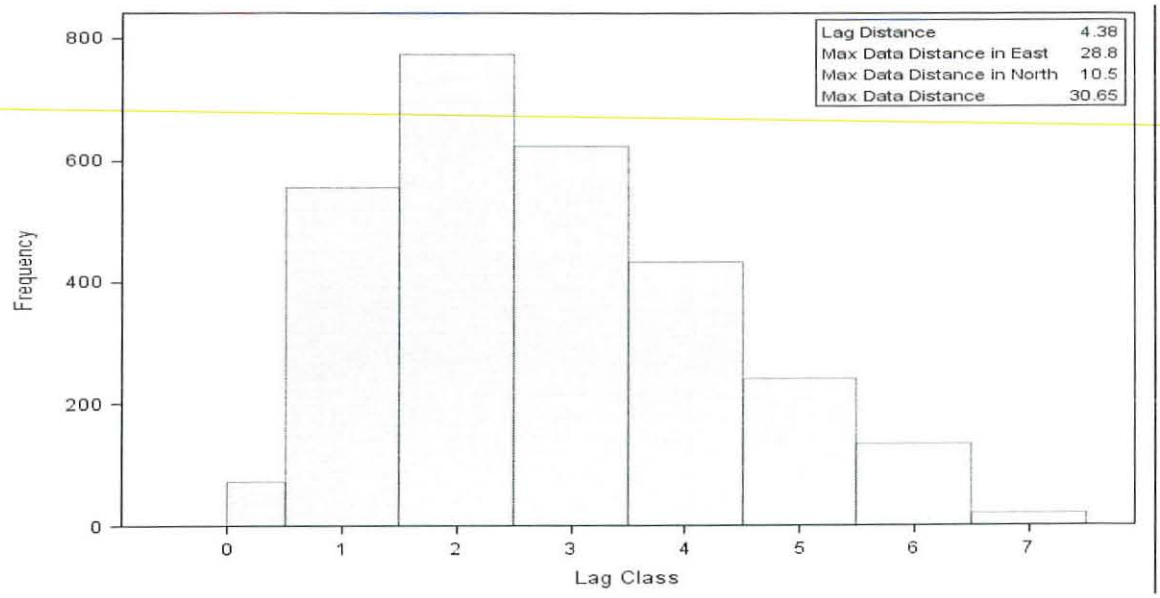
(a) BW99RVTII4



(b) BW99RVTII5



(c) BW99RVTII6



(d) DW99RVTII1

Figure 7. Distribution of pair wise distance for fitted residuals from the RCBDiid model for each data sets.

APPENDIX V

Table 10. Lag class, Pair count, Average distance, Semivarince for fitted residuals from the RCBDiid model for each data sets.

(a) BW99RVTII4

Lag Class	Pair count	Average distance	Semivariance				
			Robust	Classical	S.E.	95% Confidence Limits	
						LCL	UCL
0	76	1.6	0.00982	0.0104	0.001682	0.0071	0.0137
1	588	4.65	0.01349	0.0138	0.000807	0.0123	0.0154
2	912	8.94	0.01298	0.0138	0.000647	0.0125	0.0151
3	698	13.57	0.01332	0.0134	0.000718	0.012	0.0148
4	422	18.27	0.01397	0.0137	0.000945	0.0119	0.0156
5	294	22.72	0.01259	0.0146	0.001208	0.0123	0.017
6	150	27.24	0.01083	0.0164	0.001897	0.0127	0.0201
7	20	30.85	0.01672	0.0194	0.006128	0.0074	0.0314

(b) BW99RVTII5

Lag Class	Pair count	Average distance	Semivariance				
			Robust	Classical	S.E.	95% Confidence Limits	
						LCL	UCL
0	76	1.6	0.0157	0.0159	0.00258	0.0108	0.0209
1	588	4.65	0.0359	0.0365	0.00213	0.0324	0.0407
2	912	8.94	0.0288	0.0288	0.00135	0.0262	0.0315
3	698	13.57	0.0348	0.0342	0.00183	0.0307	0.0378
4	422	18.27	0.0328	0.0314	0.00216	0.0272	0.0356
5	294	22.72	0.0424	0.0393	0.00324	0.033	0.0457
6	150	27.24	0.0449	0.0458	0.00529	0.0354	0.0562
7	20	30.85	0.0288	0.0321	0.01014	0.0122	0.0519

(c) BW99RVTII6

Lag Class	Pair count	Average distance	Semivariance				
			Robust	Classical	S.E.	95% Confidence Limits	
						LCL	UCL
0	76	1.6	0.00778	0.00923	0.001497	0.0063	0.0122
1	588	4.65	0.01075	0.01034	0.000603	0.0092	0.0115
2	912	8.94	0.01127	0.01119	0.000524	0.0102	0.0122
3	698	13.57	0.01255	0.01284	0.000687	0.0115	0.0142

4	422	18.27	0.01229	0.01219	0.000839	0.0106	0.0138
5	294	22.72	0.01252	0.01186	0.000978	0.0099	0.0138
6	150	27.24	0.01324	0.01254	0.001448	0.0097	0.0154
7	20	30.85	0.01005	0.01043	0.0033	0.004	0.0169

(d) DW99RVTII1

Lag Class	Pair count	Average distance	Semivariance				
			Robust	Classical	S.E.[3]	95% Confidence Limits	
						LCL[2]	UCL[1]
0	72	1.6	0.0124	0.0148	0.002473	0.01	0.0197
1	555	4.64	0.0169	0.0174	0.001047	0.0154	0.0195
2	773	8.71	0.0167	0.0178	0.000904	0.016	0.0195
3	624	12.87	0.0162	0.0186	0.001051	0.0165	0.0206
4	432	17.48	0.0156	0.0183	0.001242	0.0158	0.0207
5	240	21.89	0.0145	0.0171	0.001563	0.0141	0.0202
6	134	25.87	0.0131	0.0167	0.002034	0.0127	0.0206
7	20	29.29	0.018	0.0296	0.009346	0.0112	0.0479

[1] Upper class limits for semivariance.
[2] Lower class limits for semivariance.
[3] Standard error of semivariance.

APPENDIX VI

The SAS codes for spatial data analysis.

A. SAS code for RCBDiid model fitting and Q-Q plots

```
/*RCBDiid model fitting----------------------------------------*/
        PROC MIXED DATA=;
        CLASS entry rep ;
        MODEL trnyldkg = entry rep /outp=expvar;
        RUN;
/*QQ plot for checking normality-----------------------------------*/
        PROC UNIVARIATE DATA=expvar NORMAL plot;
        VAR Resid;
        TITLE1 'QQ-PLOT OF RESIDUALS';
        QQPLOT Resid/NORMAL (L=1 MU=EST SIGMA=EST);
        RUN;
```

B. SAS code for surface (3D) plot of the residuals

```
        PROC G3D DATA=expvar;
        TITLE 'Surface Plot of estimated residual';
        SCATTER East*North=Resid / XTICKNUM=5 YTICKNUM=5
        GRID ZMIN=-2 ZMAX=2;
        SYMBOL1 V=dot COLOR=red;
        RUN;
```

C. SAS code for semivariance estimation and variogram plot

```
/* computing lag distance-------------------------------------------------*/

        PROC VARIOGRAM  DATA=expvar OUTDISTANCE=outd;
        COMPUTE NHC=7 NOVARIOGRAM;
        COORDINATES XC=East YC=North;
        VAR Resid;
        RUN;

/* computing semivariance and plot of semivariance--------------------*/

        PROC VARIOGRAM  DATA=expvar  OUTVAR=outv;
        COMPUTE LAGD=4.59 MAXLAG=7 cl ROBUST ATOL=45;
        COORDINATES XC=East YC=North;
        VAR Resid;
        RUN;

/* plot of semivariance through different direction (teast for isotropy) -------*/
```

```
PROC VARIOGRAM  DATA=expvar  OUTVAR=outv;
COMPUTE LAGD=4.59 MAXLAG=7 cl ROBUST ATOL=45;
COORDINATES XC=East YC=North;
DIRECTIONS  60, 45, 120;
VAR Resid;
RUN;
```

D. SAS code for fitting spatial models

```
/*Fitting a without block effect no nugget exponential model */;
        PROC MIXED METHOD =REML;
        CLASS entry;
        MODEL tyld = entry /ddfm=kr outp=predicted s ;
        REPEATED/SUBJECT= INTERCEPT;
        TYPE =SP(exp) (east north);
        RUN;
/* Fitting a nugget exponential model with block (fixed) effect*/;
        PROC MIXED METHOD = REML;
        CLASS entry rep;
        MODEL tyld = entry rep/DDFM = KR;
        REPEATED/SUBJECT = INTERCEPT LOCAL TYPE = SP(exp) (east north);
        /* Insert starting parameter values for PARTIAL SILL, RANGE and NUGGET
        PARMS (0.045) (3) (0.006);
        RUN;
/* Fitting a nugget Spherical model with block (fixed) effect*/;
        PROC MIXED METHOD = REML;
        CLASS entry rep;
        MODEL trnyldkg = entry rep/DDFM = KR;
        REPEATED/SUBJECT = INTERCEPT LOCAL TYPE = SP(sph) (east north);
        RUN;
/* Fitting a nugget Gaussian model with block effect*/;
        PROC MIXED METHOD = REML;
        CLASS entry rep;
        MODEL trnyldkg = entry rep/DDFM = KR;
        REPEATED/SUBJECT = INTERCEPT LOCAL TYPE = SP(gau) (east north);
        RUN;
/* Fitting a no nugget Gaussian model with block (fixed) effect*/;
        PROC MIXED METHOD = REML;
        CLASS entry rep;
        MODEL trnyldkg = entry rep/DDFM = KR;
        REPEATED/SUBJECT = INTERCEPT TYPE = SP(gau) (east north);
        RUN;
/* Fitting a no nugget Gaussian model without block effect*/;
        PROC MIXED METHOD = REML;
        CLASS entry;
        MODEL trnyldkg = entry/DDFM = KR;
        REPEATED/SUBJECT = INTERCEPT TYPE = SP(gau) (east north);
        RUN;
```

```
/* Fitting a no nugget Gaussian model with block (random) effect*/;
        PROC MIXED METHOD = REML;
        CLASS entry rep;
        MODEL trnyldkg = entry rep/DDFM = KR;
        Random rep;
        REPEATED/SUBJECT = INTERCEPT TYPE = SP(gau) (east north);
        RUN;
```

E. Computing chi-square p value

```
        DATA p_v_chisquare;
        p_value =(1-probchi(11,1));
        RUN;
        PROC PRINT  DATA= p_v_chisquare;
        FORMAT p_value 10.4;
        RUN;
```

DECLARATION

I, the undersigned, declare that the thesis is my original work, has not been presented for degrees in any other University and all sources of material used for the thesis have been duly acknowledged.

Name: Dibaba Bayisa

Signature: _____

Date: 08, July, 2010

This thesis has been submitted for examination with my approval as a University advisor.

Name: Girma Taye, PH.D

Signature: _____

Date: 8 July, 2010