ANALYSIS OF SPATIAL VARIATION OF SOIL FERTILITY GRADIENTS IN VIHIGA AND SIAYA DISTRICTS OF WESTERN KENYA USING GEOSTATISTICAL TECHNIQUES

BY

JEREMIAH MOSIOMA OKEYO
N50/7412/2002

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE DEGREE OF MASTER OF ENVIRONMENTAL SCIENCE, KENYATTA UNIVERSITY

JANUARY 2006
DECLARATION

This thesis is my original work and has not been presented for a degree in any other university or any award.

Name: Okeyo Jeremiah Mosioma

Signature: ........................................

Date: 23/01/2006

DECLARATION BY SUPERVISORS

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Prof. Wellington N. Wamicha
Department of Environmental Science, Kenyatta University.

Signature: ........................................

Date: 23/01/2006

Prof. Chris A. Shisanya
Department of Geography, Kenyatta University.

Signature: ........................................

Date: 24/01/2006

Dr. Keith D. Shepherd
World Agroforestry Centre (ICRAF)

Signature: ........................................

Date: 23.01.06
DEDICATION

To my loving father Gerard Okeyo Omurwa whose constant effort and inspirational support kept me focused.
ACKNOWLEDGEMENT

The work reported in this thesis was made possible through effort put forth by a number of people. I would like to thank my supervisors Prof. Wellington Wamicha, Prof. Chris Shisanya and Dr. Keith Shepherd who continuously guided me throughout the study period. Their invaluable suggestions and constructive criticism contributed significantly to the success of this work.

I appreciate the support given by a number of TSBF-CIAT and ICRAF staff during various stages of the study. Specifically, I thank Dr. Bernard Vanlauwe and Dr. Ann Muriuki whose concern kept me afloat throughout the study period. I am also indebted to Laban Nyambega and Tom Owiyo with whom we spent long hours in the field during data collection and subsequent data analysis. Further thanks go to Ateku, Ichami, Thine and Sila for assistance in the spectral analysis. I also extend my gratitude to my colleagues: Aomo, Baaru, Karunditu, Kibisu, Mairura, Ming’ate, Ndirangu, Wangai and Waswa for their academic and moral support during my study period.

This study could not have been possible without the financial support from the Soil Fertility Gradients (SFG) project of TSBF-CIAT. Logistical support and linkage was provided by Dr. Daniel Mugendi. I also acknowledge the Department of Environmental Sciences, Kenyatta University for academic support.

God’s Blessings to you all
ABSTRACT

In western Kenya, several soil fertility management technologies have been developed in specific benchmark areas and then recommended to the rest of the farmers. Adoption of such technologies has been minimal at best, and, one of the reasons given for this low rate of adoption is that they did not take into consideration the existing spatial variations in biophysical and socio-economic conditions within which the local smallholder farmers operate. Against this background, a study was carried out to quantify the variability of soil fertility at different spatial scales and formulate domains for better targeting of soil fertility management recommendations.

Farms were selected using a hierarchical Y-frame sampling design and in each farm information on the main biophysical factors collected. Field measurements, observations and sampling were used to collect data on the biophysical conditions, while participatory rural appraisal (PRA) was used to collect socio-economic data. All fields in each farm were characterised and top soil samples collected at a depth of 0-20 cm. All the sample collection points were georeferenced using a GPS system. Exploratory data analysis techniques were used to assess the effects of biophysical and socio-economic parameters on soil fertility. Geostatistical techniques of semivariography and kriging were used to explore the spatial structure of soil fertility gradients. Mixed effects modelling was used to confirm relationships, while accounting for spatial correlation structures, and understanding the variance of predicted soil organic C at different spatial scales.

Predicted soil organic C was found to be spatially correlated and the spatial structure was modelled using experimental semivariograms fitted with spherical, exponential and ratio quadratic models. At the Y-level, using the exponential semivariogram model, spatial structures ranged from weak in Y3 (nugget/sill ratio > 0.75), moderate in Y2, Y5, Y7, Y8 and Y9 (nugget/sill ratio 0.25 < r < 0.75) to strong in Y1 and Y4 (nugget/sill ratio < 0.25). On average all the three variogram models gave a nugget/sill ratio of between 0.5-0.6 indicating moderate spatial correlation. The maximum range at which this spatial structure can be reliably predicted is up to 60 m beyond which correlation errors increase significantly. All the three model variogram
estimates had high nugget variances which imply that the micro-scale variation (i.e. variation below the minimum sampling interval) was large.

Analysis of the estimated variance components showed that the field (residual) effect accounted for the greatest percentage (62.5%) of the variation associated with random effects. After accounting for spatial variability all the other measured parameters (fixed effects) failed to explain the large local variability, thus, posing a challenge to making soil fertility management recommendations. Future soil fertility management strategies in western Kenya should target at explaining the large spatial variability of soil fertility within the smallholder farms.
# TABLE OF CONTENTS

DECLARATION ........................................................................................................... ii  
ACKNOWLEDGEMENT ............................................................................................... iv  
ABSTRACT .................................................................................................................. v  
TABLE OF CONTENTS .............................................................................................. vii  
LIST OF TABLES ........................................................................................................ x  
LIST OF FIGURES ....................................................................................................... xi  
LIST OF PLATES ......................................................................................................... xii  
LIST OF ACRONYMS ................................................................................................. xiii  

## CHAPTER ONE ................................................................. 1

### INTRODUCTION .......................................................... 1

1.1 Background information .................................................................................. 1  
1.3 Research questions ......................................................................................... 5  
1.4 Objectives of the study ................................................................................... 6  
1.5 Justification of the study ................................................................................ 6  

## CHAPTER TWO .......................................................... 8

### LITERATURE REVIEW .................................................. 8

2.0 Introduction ....................................................................................................... 8  
2.1 Nature of farming systems ............................................................................. 8  
2.2 Soil fertility degradation ............................................................................... 12  
2.3 Integrated soil fertility management ............................................................. 14  
2.4 Spectral analysis of soil ................................................................................ 15  
2.5 GIS and resource management ..................................................................... 17  
2.6 Geostatistics and spatial prediction ............................................................... 18  
2.6.1 Variogram Estimation .............................................................................. 19  
2.6.2 Fitting Variogram Models ................................................................. 21  
2.6.3 Kriging .................................................................................................... 22  
2.7 Data structure and modelling ...................................................................... 24  

## CHAPTER THREE .......................................................... 26

### RESEARCH METHODOLOGY ........................................ 26

3.1 Study area ...................................................................................................... 26  
3.1.1 Location .................................................................................................... 26  
3.1.2 Topography and Climate ....................................................................... 27
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.3</td>
<td>Geology and Soils</td>
<td>29</td>
</tr>
<tr>
<td>3.2</td>
<td>Design of the study</td>
<td>30</td>
</tr>
<tr>
<td>3.3</td>
<td>The sample and sampling procedure</td>
<td>32</td>
</tr>
<tr>
<td>3.4</td>
<td>Data collection</td>
<td>32</td>
</tr>
<tr>
<td>3.5</td>
<td>Spectral analysis of soils</td>
<td>33</td>
</tr>
<tr>
<td>3.6</td>
<td>Soil chemical analysis</td>
<td>34</td>
</tr>
<tr>
<td>3.7</td>
<td>Data analysis</td>
<td>35</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Geostatistical Analysis</td>
<td>36</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Mixed Effects Modelling</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>RESULTS AND DISCUSSION</td>
<td>40</td>
</tr>
<tr>
<td>4.0</td>
<td>Introduction</td>
<td>40</td>
</tr>
<tr>
<td>4.1</td>
<td>Farm Characterisation</td>
<td>40</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Field Characteristics</td>
<td>40</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Spectral Soil Analysis</td>
<td>47</td>
</tr>
<tr>
<td>4.2</td>
<td>Soil Organic Carbon</td>
<td>49</td>
</tr>
<tr>
<td>4.3</td>
<td>Variogram Estimations and Modelling</td>
<td>54</td>
</tr>
<tr>
<td>4.4</td>
<td>Spatial Interpolation -Kriging</td>
<td>65</td>
</tr>
<tr>
<td>4.5</td>
<td>Determination of Predicted Organic C</td>
<td>68</td>
</tr>
<tr>
<td>4.6</td>
<td>Mixed-Effects Modelling</td>
<td>74</td>
</tr>
<tr>
<td>4.6.1</td>
<td>The Spatial Model</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS</td>
<td>90</td>
</tr>
<tr>
<td>5.0</td>
<td>Introduction</td>
<td>90</td>
</tr>
<tr>
<td>5.1</td>
<td>Main findings</td>
<td>90</td>
</tr>
<tr>
<td>5.2</td>
<td>Conclusions</td>
<td>91</td>
</tr>
<tr>
<td>5.3</td>
<td>Recommendations</td>
<td>92</td>
</tr>
<tr>
<td>5.4</td>
<td>Further research areas</td>
<td>93</td>
</tr>
<tr>
<td>6</td>
<td>REFERENCES</td>
<td>94</td>
</tr>
<tr>
<td>7</td>
<td>APPENDICES</td>
<td>107</td>
</tr>
<tr>
<td>Appendix A.1: Field Characterisation (Form 1)</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>Appendix A.2: Field Characterisation (Form 2)</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>Appendix A.3: GPS Field Location Form</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>Appendix B.1: Linear mixed-effects model fit by REML</td>
<td>110</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B.2: Semivariogram Distance Estimates ............................................ 111
Appendix B.3: Fitted Ratio Semivariogram Model............................................. 112
Appendix B.4: Fitted Exponential Semivariogram Model................................. 113
Appendix B.5: Fitted Spherical Semivariogram Model ..................................... 114
LIST OF TABLES

Table 2.1: Characteristic of calculated semivariogram of spatial fertility. ............. 23
Table 4.2: Summary of Fields Sampled .................................................................... 40
Table 4.3: Summary of Field Areas and Distances ....................................................... 41
Table 4.4: Types of Land Use ...................................................................................... 43
Table 4.5: Location of the Fields on the Landscape ...................................................... 44
Table 4.6: Slope of Fields ............................................................................................ 44
Table 4.7: Farmers’ Rating of Field Soil Fertility .......................................................... 46
Table 4.8: Some physical field characteristics ............................................................... 46
Table 4.9: Measured Soil Properties ............................................................................ 48
Table 4.10: Soil Organic Carbon in each Y ................................................................. 49
Table 4.11: Parameter values of fitted semivariograms models .................................. 63
Table 4.12: Multible linear regression parameters for soil organic C ....................... 69
Table 4.13: Model Estimates and Random Effects for the Basic Mixed Effects Model .................................................................................................................. 74
Table 4.14: Summary of Fixed Effects of the Basic Mixed Effects Model ................. 75
Table 4.15: Parameter values for fitted model semivariograms .................................. 83
Table 4.16: Parameter values for fitted model semivariograms .................................. 86
Table 4.17: ANOVA Analysis between the fitted models ........................................... 86
Table 4.18: Model Semivariogram Estimates ............................................................... 87
LIST OF FIGURES

Figure 2.1: A Theoretical Semivariogram Model .................................................. 21
Figure 3.1: Map of the Study Area ................................................................. 29
Figure 3.2: The Y Sampling Frame ................................................................. 31
Figure 4.1: Correlation between field size and distance .................................. 42
Figure 4.2: Variation in average field slope ....................................................... 45
Figure 4.3: Boxplots showing the variation of soil organic C within-farms ......... 50
Figure 4.4: Boxplots showing the variation of soil organic C between-farms ...... 52
Figure 4.5: Spherical semivariogram models of Soil Organic C in each of the Ys .................................................................................................................. 57
Figure 4.6: Exponential semivariogram models of Soil Organic C in each of the Ys .................................................................................................................. 61
Figures 4.7 (a) – (n): Kriging predictions of soil organic C .................................. 67
Figures 4.8 (a) – (r): Plots of multiple regression models of soil organic C .......... 72
Figures 4.9 (a) – (d): Plots of multiple regression models of soil organic C .......... 73
Figure 4.10: Boxplots of within-group residuals; (a)– district level and (b)– Y level .... 76
Figure 4.11: Fitted model semivariogram at maximum distance ....................... 81
Figure 4.12: Fitted model semivariogram at 80 m ............................................. 82
Figure 4.13 Fitted Semivariogram Models; Exponential, Spherical and Ratio ..... 85
Figure 4.14: Fitted model semivariogram of normalized residuals .................. 88
LIST OF PLATES

Plate 4.1: Measuring the area of a field ................................................................. 42

Plate 4.2: Measuring soil reflectance values using a spectroradiometer .......... 47

Plate 4.3: Spatial variation of maize growth in a field ........................................ 51
**LIST OF ACRONYMS**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Criterion</td>
</tr>
<tr>
<td>BLUE</td>
<td>Best Linear Unbiased Model</td>
</tr>
<tr>
<td>CIAT</td>
<td>International Centre for Tropical Agriculture</td>
</tr>
<tr>
<td>DEM</td>
<td>Digital Elevation Model</td>
</tr>
<tr>
<td>DRS</td>
<td>Diffuse Reflectance Spectroscopy</td>
</tr>
<tr>
<td>DTM</td>
<td>Digital Elevation Model</td>
</tr>
<tr>
<td>ECEC</td>
<td>Effective Cation Exchange Capacity</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographical Information Systems</td>
</tr>
<tr>
<td>GPS</td>
<td>Geographical Positioning System</td>
</tr>
<tr>
<td>INM</td>
<td>Integrated Nutrient Management</td>
</tr>
<tr>
<td>ISFEIP</td>
<td>International Soil Fertility Evaluation and Improvement Program.</td>
</tr>
<tr>
<td>ISFM</td>
<td>Integrated Soil Fertility Management</td>
</tr>
<tr>
<td>LC</td>
<td>Land Cover</td>
</tr>
<tr>
<td>LISQ</td>
<td>Local Indicators of Soil Quality</td>
</tr>
<tr>
<td>Log (1/R)</td>
<td>Absorbance</td>
</tr>
<tr>
<td>LogLik</td>
<td>Log Likelihood</td>
</tr>
<tr>
<td>LU</td>
<td>Land Use</td>
</tr>
<tr>
<td>MSR</td>
<td>Mean Squared Residual</td>
</tr>
<tr>
<td>MSS</td>
<td>Multi-Spectral Scanner</td>
</tr>
<tr>
<td>NIRS</td>
<td>Near-Infrared Reflectance Spectroscopy</td>
</tr>
<tr>
<td>Org C</td>
<td>Organic Carbon</td>
</tr>
<tr>
<td>PLAR</td>
<td>Participatory Learning and Action Research</td>
</tr>
</tbody>
</table>
PPU Primary Production Unit
PRA Participatory Rural Appraisal
REML Restricted Maximum Likelihood
SFG Soil Fertility Gradient
SFI Soil Fertility Index
SMS Subject Matter Specialists
SOC Soil Organic Carbon
SSA Sub Saharan Africa
TISQ Technical Indicators of Soil Quality
TM Thematic Mapper
TSBF Tropical Soil Biology and Fertility
UNEP United Nations Environmental Programme
WRI World Resource Institute
CHAPTER ONE

INTRODUCTION

1.1 Background information

Soil fertility depletion is recognized as a principal biophysical factor associated with declining food security in smallholder farms in Sub-Saharan Africa (Sanchez and Jama, 2002). It is argued that no matter how effectively other constraints are remedied, per capita food production in Africa will continue to decline unless soil fertility depletion is effectively addressed. Also important are the underlying socio-economic causes of nutrient depletion. The causes of soil fertility depletion are multiple and are strongly interrelated including the interaction between biophysical and socio-economic factors. A holistic approach is thus required to ameliorate the soil fertility constraints (Murwira, 2003). There is a need to put more emphasis on the identification of productive and sustainable alternative land management techniques for the diverse farming circumstances in which the small scale farmers operate.

Until recently there has been little attempt to target nutrient management technologies in a complex farming environment (Deckers, 2002). In the past, many soil fertility improvement experiments have been carried out in experimental stations and the resulting technologies disseminated as blanket recommendations to the surrounding farming communities, such as recommended fertiliser rates. Basic principles of variability of the physical resources which underlie the farming system, were little understood, nor were the farmers’ socio-economic environment (Deckers, 2002). Also, plant nutrient management technologies were commonly recommended in isolation and therefore were not effective where other overriding limiting factors were overlooked.
The understanding of the spatial variability of soil fertility levels between and within farms is important for refining farm management practices and for assessing the impact of agriculture on the environment. The variability of soil properties within fields is often described by a classical method, which assumes that variation is randomly distributed within mapping units. Soil variability is the outcome of many processes acting and interacting across a continuum of spatial and temporal scales and is inherently scale dependent (Parkin, 1993). In addition, soil properties frequently exhibit spatial dependency, whereby, samples collected close to each other tend to be more correlated than those collected far apart. Therefore, parametric statistics are inadequate for analysis of spatially dependent variables because they assume that measured observations are independent inspite of their distribution in space (Hamlett et al., 1986).

Geostatistics provides a set of statistical tools for incorporating the spatial coordinates of soil observations in data processing, allowing for description and modeling of spatial patterns, prediction at unsampled locations, and assessment of the uncertainty attached to these predictions. In this study, geostatistics was used as an analytical tool to illustrate the spatial inter-relationship of collected soil data and to provide kriging weights for modeling such data. Geostatistical analyses have been used to estimate spatial variability of soil physical properties (Viera et al., 1981; Lascano and Hatfield, 1992), soil biochemical properties (Sutherland et al., 1991), and soil microbiological process (Aiken et al., 1991; Rochette et al., 1991). Values of soil properties obtained are used to predict values at unsampled locations.
Geostatistical data, also called random field data, consists of measurements taken at fixed locations (Kaluzny et al., 1998). It consists of variography and kriging. Variography uses variograms (semivariograms) to characterize and model the spatial variance of data, whereas kriging uses the modeled variance to estimate values at unsampled locations (Mohammadi, 2002). The process of geostatistical analysis can be divided into the following sections; estimating variograms, fitting theoretical variogram models, performing ordinary and universal kriging, and simulating geostatistical data (Kaluzny et al., 1998).

As more and more soil and environmental information becomes readily available through modern GIS technology, there is opportunity for better spatial targeting of technologies in relation to variation in the natural resource base. According to Deckers (2002) targeted balanced nutrient management systems in the soilscape of Sub-Saharan Africa is one of the corner stones of sustainable development. In order to address this issue, Dumanski and Craswell (1998) and Craswell et al. (1998) defined the concept of a Resource Management Domain as:

'a spatial (landscape) unit that offers opportunities for identification and application of resource management options to address specific issues. It is derived from geo-referenced biophysical and socio-economic information; it is dynamic and multi-scale and reflects human interventions in the landscape'.

According to Moore et al. (1993) some of the most interesting researches being carried out today aim at developing integrated biophysical-socioeconomic models, databases and information systems. When fully developed these decision support systems can easily be used by politicians, policy-makers, and managers who often lack intimate knowledge of the models being used. Thus, there is a need to develop an integrated approach to balanced nutrient management systems for better adoption of new technologies by smallholder farmers in Africa. The approach should also take
into account the spatial variations in terms of nutrient levels within farms, which can be highly variable.

1.2 Statement of the problem
As more research is being carried-out on better ways of addressing the problem of soil fertility, it is becoming more evident that there is a need to understand the spatial variations in existing biophysical and socio-economic conditions. To assess possible fertiliser strategies, Ruecker et al. (2003) recommended first assessing the spatially variable natural and socio-economic resource conditions of an area such as agricultural potential, population density and market access. Given that agriculture comprises an important part of many economies, it is important that both the consequences of expected environmental change, and imposed mitigation and adaptation strategies, can be reliably predicted.

Much has been done and is being done in sub-Saharan Africa to address issues of declining soil fertility, but there is consensus that the results remain limited in relation to the scale of the problems and that widely replicable and sustainable approaches have yet to be identified. One of the fundamental constraints for a more widespread adoption of soil fertility management strategies at the national level is the huge complexity of the natural resource and the socio-economic systems at lower levels of scale. Thus, effective soil fertility management requires stratification of the smallholder farms into homogenous units or domains within which management decisions should be more or less the same.

In western Kenya, however, no actual studies have been carried-out to try and quantify or map the spatial structure of the existing biophysical conditions as well as
the emerging socio-economic patterns. This has not taken place against a backdrop of many soil fertility management projects that have been carried out in the area ranging from better agroforestry practices such as improved fallows to nutrient replenishment products such as Minjingu rock phosphate. All these studies have tended to concentrate on limiting nutrients and at specific experimentation sites but have not considered how the same nutrients vary spatially from field to field and across farms.

Secondly, it has been observed that though the fertility levels are generally declining at the farm level, there is much variation in rates of soil fertility change, which can be positive as well as negative. For example, concentration of household waste and manure in fields near homesteads or in cattle enclosures (bomas) can result in soil nutrient 'hot' spots, whereas more marginal or distant fields may be severely degraded (Vanlauwe et al., 2002). If such spatial patterns can be unraveled and effectively modeled, then better targeting of Integrated Soil Fertility Management (ISFM) strategies can be achieved.

1.3 Research questions

The study was guided by the following research questions:

1. What are the main biophysical factors associated with variation in soil fertility at different spatial scales from plot to district level?

2. Is variation in soil fertility at plot and farm scales as great as at district scale?

3. What biophysical factors could form the basis for resource management domains in terms of (a) factors that are amenable to management, and (b) factors not amenable to management?
1.4 Objectives of the study

The main objective of the study was to formulate a framework for better targeting of soil fertility management recommendations to smallholder farms in Vihiga district of western Kenya that could be implemented using geostatistical techniques. To achieve this broad objective, the following specific objectives were looked at;

1. To quantify the variability in soil fertility level at different spatial scales from plot level to district level.
2. To identify the main biophysical factors affecting within-farm soil fertility gradients.
3. To suggest resource management domains that could be used to improve targeting of soil fertility management recommendations at plot to district scales.

1.5 Justification of the study

The non-linearity and spatiotemporal heterogeneity of agricultural ecosystems suggests that application of homogenous soil fertility management strategies is theoretically flawed and practically formidable. One of the fundamental stumbling blocks for more effective policy dissemination is the spatial complexity of biophysical and socio-economic conditions at different spatial scales. According to Nandwa (2003), there are two main requirements for sustainable soil fertility management. First, the capacity to manage soil fertility is dependent on understanding the biological processes regulating nutrient flux, organic matter dynamics and soil structure modification. Secondly is that effective soil fertility management can be achieved by the integration of the contributory soil processes with other factors regulating ecosystem dynamics including those of resource availability and access, and the farmers’ decision making processes. Thus, for research and extension efforts to develop and widely disseminate integrated nutrient management strategies there is
need to incorporate the spatial variability of the underlying biophysical conditions in technology testing and development.

A farm-level scale incorporating landscape patterns that the local farmers understand is likely to be a better approach to soil fertility management than the scientific rigour that farmers tend to avoid. Also, in an attempt to develop a methodology for studying the relationship among pattern, process and scale and for extrapolating information across heterogeneous landscapes, there is a need to take into account spatial autocorrelation of structures which most of the classical statistical methods used in agriculture largely ignore. It is through such approaches that it is possible to map and delineate soil fertility management domains in interactively complex ecosystems such as those occupied by smallholder farmers in western Kenya.
CHAPTER TWO
LITERATURE REVIEW

2.0 Introduction

This chapter gives a general overview of previous work done in Sub-Saharan Africa in effort to alleviate the problem of food insecurity through integrated soil fertility management while taking into account spatial variations. It is divided into five sections as follows; the first two sections (2.1 and 2.2) address the nature of farming systems in smallholder farms and the resultant soil fertility degradation occasioned by negative nutrient balances. Section 2.3 deals with integrated soil fertility management and its applicability in Sub-Saharan Africa. The last two sections (2.4 and 2.5) show how GIS and data modeling have been used for better resource management by taking into account spatial variability in agro-ecosystems.

2.1 Nature of farming systems

A farming system can be defined as a unit consisting of a human group (usually a household) and the resources it manages (land, labour, and capital) in its environment, involving the direct activities of producing plant and/or animal products with or without contribution of off-farm activities (Beets, 1990). The system is always a part of a larger social, economic, cultural and political environment that impacts on everything happening within the farming system (Nandwa, 2003). The type and complexity of a farming system is determined by both natural and socio-economic factors.

Variations in climatic, soil, water and biological resources result in different agroecological zones, beyond which generalisations about soil fertility are not very meaningful. It is essential that soil related potentials and constraints be taken into account when designing measures to improve soil fertility at the farmer’s level (Dudal, 2002). Measures aiming at the enhancement of soil fertility obviously need to be based on a more precise characterisation of the environment and the socio-economic conditions in the country concerned. According to Hamblin (1992) agroecological changes, when monitored systematically and analysed using models of agricultural systems, can provide information that is vital to effective policy development and adaptation of strategies. In addition, Lynam and Herdt (1989)
suggested that the strategies need to be clear and focused on specific issues at an operational level. These models are likely to be particularly important in defining and achieving sustainable agricultural systems.

Sustainable agricultural intensification requires prudent long-term management of the natural resource base on which agriculture fundamentally depends (Barrett et al., 2002). Recent years have seen dramatic increase in per capita food production in many areas at a global scale (Scholes et al., 1994). Despite these improvements, the global community has been concerned that these gains cannot be sustainable (York, 1988), mainly because of the recognition of both environmental problems related to the new technologies widely adopted in more favored agricultural regions and the relative neglect of serious problems associated with agricultural expansion and intensification in areas for which the new technologies were not well-suited (Pretty, 1997).

For instance, in Europe, following the end of World War II, two approaches to soil nutrient status were practiced. One attempted to build soil fertility levels by application of organic matter and amendments so that crops would not experience nutrient limitations (Seim, 1997). The second was to supply the nutrients needed (or removed by each crop) regardless of the soil nutrient status. Reduced profit margins subsequently brought crop production managers back to a more practical approach of supplying only those elements that limited yield. Environmental concerns also made blanket applications of nutrients, pesticides or water politically, socially and biologically incorrect. Improvement in system understanding and predictive ability requires integration of model development, field and laboratory experimentation, and performance monitoring of the system studies (White et al., 1993).

Farming systems in SSA are diverse and range from traditional farming systems, through a range of changing phases of true shifting cultivation to permanent and intensive arable cultivation with livestock and range production, at times with integration with wildlife enterprises (Duckham and Masefield, 1971). It is important to emphasize that in all areas, traditional farming systems have evolved and continue to evolve as coping strategies for the environment, and its changing biophysical and socio-economic circumstances (FAO, 1995b). This in essence means that current
farming systems are a mixture of aspects of traditional farming systems, emergent and modern farming systems or agriculture (Scherr, 1997). Despite such initiatives the problem of food insecurity remains most acute in Africa as compared with other developing countries (Cleaver and Schreiber, 1994). The challenge of overcoming soil productivity decline is exacerbated by the fact that soil fertility status is not a static phenomenon (temporary variability), and that soil fertility can be fairly complex and heterogeneous from the spatial point of view (Smaling et al., 1997).

In Kenya, like many other sub-Saharan Africa countries, soil fertility and hence productivity is declining at an alarming rate because areas of high agricultural potential are densely populated and in most cases farm holdings are less than one hectare (Gikonyo and Smithson, 2004). The average annual loss in soil nutrients is among the greatest in Africa, mainly because farmers practice intensive continuous cropping with limited or no replenishment of nutrients (Smaling, 1993). Nutrient depletion at the farm-scale results from imbalance of nutrient inputs and losses over time and reaches critical proportions when land is continuously cultivated without the addition of adequate external nutrient inputs (Woomer et al., 1999). At the field scale, nutrient depletion is caused by the internal flows of organic resources, particularly crop residues harvested as livestock feed and subsequent allocation of manures to higher-value crops and/or kitchen gardens.

The western Kenya region is one of the most densely populated areas in Kenya, supporting between 500 and 1200 inhabitants per km² (Hoekstra and Corbett, 1995). Despite the high agricultural potential, food production is low. Farming is mainly subsistence, with smallholder farmers growing two crops per a year, with little or no fertilizer inputs (Ojiem et al., 2004). Over the years, the practice has resulted in soil degradation and nutrient depletion due to continuous cultivation, removal or burning of crop residuals, loss of nutrients through erosion, overgrazing between seasons and inadequate use of inorganic fertilizers (Stoorvogel et al., 1993). Although judicious application of inorganic fertilizers is recognized as the most effective amendment for overcoming soil fertility decline or alleviating nutrient deficiencies, their high cost, inaccessibility, and generalized recommendations resulting in low, erratic and unprofitable crop responses limit their use, particularly on smallholder farms in
eastern Africa (Nandwa and Bekunda, 1998). However, the soil fertility problem is not uniform both at regional and farm scales (KARI, 2003).

In the 1990s the availability of GPS and GIS to agricultural applications made it possible to manage very small units (some fields within a farm or subsections of the same field) rather than managing the field as an average (Seim, 1997) leading to the precision agriculture paradigm. Reawakened interest in the Law of the Minimum and soil testing held out the prospect of increasing yield while decreasing costs. In many instances soil differences and yield monitors justified the expense of the GPS and GIS technology, as increased yields were achieved with fewer chemical inputs. Additionally, the environmental cost of over-application was lessened especially in the case of nitrate and phosphorus fertilisation.

An important outcome of precision agriculture was to perceive potential benefits of crop management by zones within fields rather than whole fields for increased profitability and environmental protection (Robert, 2002). At the same time, the new technologies of microcomputers, GIS and GPS have made the acquisition, processing, and utilization of spatial field data possible. Olson (1998) defines precision agriculture as the application of a holistic management strategy that uses information technology to bring data from multiple sources to bear on decisions associated with agricultural production, marketing, financing, and personnel. The scenario calls for the introduction of modern technologies to improve crop yields, provide information to enable in-field management decisions, reduce chemical and fertilizer costs through more efficient application, permit more accurate farm records, increase profit margins and reduce pollution (Seelan et al., 2003). The variable rate fertilizer technology (VRT) is receiving attention because of its potential benefit to the economics of agricultural production and its potential minimization of adverse environmental impacts caused by excessive rates of fertilizers and other agricultural chemicals (Gangloff et al., 2001).

Precision agriculture is not just addition of new technologies but is rather an information revolution, made possible by new technologies that result in a higher level, a more precise farm management system (Robert, 2002). Proper use of spatial and multivariate statistics may be a key to successful implementation of precision
farming strategies (Gangloff et al., 2001). However, adoption of the concept, including crop nutrition, has been slower than initially thought because of several kinds of significant challenges in terms of socio-economic, agronomical, and technological issues of agricultural production (Robert, 1999). When these barriers are overcome, site-specific practices, including crop nutrition, will be an essential aspect of the agricultural system of the 21st century because it offers a variety of potential benefits in profitability, productivity, sustainability, crop quality, food safety, environmental protection, on-farm quality of life, and rural economic development.

2.2 Soil fertility degradation

A key measure of the long-term productive capacity of an agroecosystem is the condition of its soil. Natural weather processes and human management practices can both affect soil quality. Sustaining soil productivity requires that soil degrading pressures be balanced with soil-conserving practices (World Resource Institute, 2002). The concept of nutrient depletion is derived from simple quantification of nutrient flows resulting in negative nutrient balances and/or stocks (Nandwa, 2003). However, nutrient differences on the landscape and resource management challenges arise from more complex variations in settlement history, past history of degradation, the mix of crop, perennial and livestock components, and the mix of commercial and subsistence enterprises (Scherr, 1999).

Nutrients flow at every spatial agro-ecosystem level (country, region, district, catchment area, farm, plot and soil solution), but the basic unit of study in agriculture is commonly the farm, which is fundamentally, a socio-economically defined system of production (Nandwa, 2003). A majority of the nutrient budget studies in SSA have been conducted primarily at the plot/farm level (Elias et al., 1998; Baijukya and Desteenh, 1998; Wortmann and Kaizzi, 1998; Harris, 1998; Bosch et al., 1998; Shepherd et al., 1996; Shepherd and Soule, 1998; Brouwer and Powel, 1998), with only a few studies to also covering the village/community level (Defoer et al., 1998; Brand and Pfund, 1998). Similarly, a majority of the above-cited studies have used full nutrient balances i.e. farm in-flows (mineral fertiliser, organic manure, deposition, biological N-fixation and sedimentation) minus farm out-flows (harvested crops, crop residues, leaching, denitrification and water erosion). In this context, the farm is
conceptualised as a set of dynamic units which depending on the management, form the source and/or destination of nutrient flows and economic flows (Stoorvogel, 1993).

For example, Templeton and Scherr (1997) found empirical evidence that the relationship between population growth and resource quality in hills and mountains was influenced by rainfall (mainly by affecting crop product choice, risks of soil degradation, and land use intensity), topography (affecting spatial distribution of production systems), and soil characteristics (through crop choice cropping frequency and input use). According to Dudal (2002), rural people have a deep understanding of these physical, biological and socio-economic components of their environment. They most often do not willingly and consciously degrade their resource base. When recommended improved practices are not adopted, it is not because of ignorance or stubbornness but often because the new technologies are not feasible in a specific setting or do not provide a return of the required investment (Enters, 1998; Scherr, 1993).

In general, the factors that cause environmental degradation, such as soil erosion are well known (McCloy, 1995). However, the magnitude and contribution of each of these factors, in specific situations is rarely known with any accuracy. The detailed information necessary to conduct management and operational control tasks is thus rarely available to resource managers and hence they cannot determine the best way to address specific issues (Pender et al., 1999).

Management of degradation, is therefore, integrally involved with community perceptions on the extent and the severity of the problem. McCloy (1995) further argues that degradation usually occurs in small increments that are unevenly distributed over time and space, their impact are neither uniform nor readily recognised, delaying community awareness of their severity. Thus, Nandwa (2003) concluded that there is a need for creation of a holistic framework for closer interaction between soil fertility subject matter specialists (SMS), economists, extensionists and policy makers (multi-disciplinary approach) in strategies for compacting nutrient depletion in a manner that involves increased farmer participation and all stakeholders. Hence the concept of integrated soil fertility management.
2.3 Integrated soil fertility management

Despite the diversity of approaches, efforts, solutions, and investment of resources and time by a wide range of institutions, soil fertility degradation in SSA remains the single most important constraint to food security (Sanchez et al., 1997). Essentially, integrated soil fertility management (ISFM) has come in as an adoption of a systematic participatory and holistic approach to research on soil fertility that embraces the full range of driving factors on the landscape such as biological, physical, chemical, social, economic and political aspects of soil fertility degradation (Kimani et al., 2003). The approach advocates for careful management of soil fertility aspects that optimise crop production potential through incorporation of a wide range of adaptable soil management principles, practices and options for productive and sustainable agroecosystems (Hilhorst and Toulmin, 2000).

Recent trends to make ISFM a knowledge intensive approach have integrated local and scientific knowledge through the development of Participatory Learning and Action Research (PLAR), as a process to help farmers to improve their soil fertility management strategies (Baltissen et al., 2000). A good example is the development of a guide for identifying, classifying and integrating local indicators of soil quality (LISQ) and technical indicators of soil quality (TISQ) to guide the development of sustainable management principles and strategies for ISFM (Barrios et al., 2000). The approach helps to overcome some of the drawbacks of past conventional research methods which have often assumed that conditions within individual fields are uniform resulting in generalisation of recommendations when dealing with soil fertility problems in heterogeneous gradients and niches (Kimani et al., 2003).

A good example of such variability is from a case study carried out in Kabras division, Kenya, where seven niches were identified under a single smallholder farm which tended to vary tremendously in soil fertility (Ojiem et al., 2000). Some of the important sources of these variability were the differential long-term management of the different fields of a farm, creating zones of high and low soil fertility due to concentration of agricultural produce and organic wastes in some (Scoones and Toulmin, 1999, Smaling et al., 1997). It can be said then that management decisions by the farmer induce the establishment of soil fertility gradients associated with resource allocation. The magnitude of the variation from farm to farm as well as for
different regions depends on the possible combinations of the deterministic factors, affected by both biophysical and socio-economic conditions (Tittonell, 2003).

The principal challenge facing most agricultural production systems is how to optimize the use of scarce nutrient resources to achieve sustainable production (Schlecht and Hiemaux, 2004). The key factors affecting nutrient flows must be critically assessed, together with the question if and how improved nutrient management in one production component affects nutrient availability to other components (Feller 1993). The concept of modelling can be used to optimize nutrient balance calculations by allowing for the integration of processes organized at different time scales, while the use of GIS allows integration of processes that are operative at different spatial scales (Stoorvogel et al., 1995). Since most soil management decisions are made at the farm level (Vanlauwe, 2004), it is therefore preferable to restrict the assessment of flows and balances to well-defined domains relevant to particular objectives (Schlecht and Hiemaux, 2004).

Although much is known on soil properties and processes and on individual cases of technology performance, there still is a serious data crisis on the impact on soil productivity of diverse INM components found in different farming systems in SSA (Nandwa and Bekunda, 1998). The extension of sustainable solutions is essential if the necessary wider impacts with many farmers are to be achieved (Murwira, 2003). Dissemination of technology using participatory approaches is key to extending recommendations from plot to farm level and beyond, and it requires that first we investigate on what local communities know and have. Of greater significance in scaling up is development of appropriate institutional arrangements and resource management information systems that are needed to ensure longer-term sustainability of the systems (Dumanski and Craswell, 1998).

2.4 Spectral analysis of soil

Near-infrared reflectance spectroscopy (NIRS) is a nondestructive analytical technique for studying interactions between incident light and a material’s surface (Chang et al., 2001). The technique was first developed more than three decades ago for rapid moisture analysis of grain (Ben-Gera and Norris, 1968) and has been widely
used in individual applications. In the resent past, diffuse reflectance spectroscopy (DRS) has been used as a rapid, non-destructive method for characterization of a wide range of agricultural materials based on their particular reflectance, as a function of wavelength in the electromagnetic spectrum (Davies and Giangiacomo, 2000; Shepherd and Walsh, 2002). This approach has been proposed to provide a rapid prediction of soil physical, chemical and biological properties (Janik et al., 1998). Some success has been reported in sensing soil organic matter in the field (Sudduth and Hummel, 1993) as well as in discriminating soil types from satellite multi-spectral data (Coleman et al., 1993).

Assessing soil spatial variability requires dense sampling to be adequately characterized (Tittonell, 2003). However, soil analyses are expensive and time consuming, making broad-scale quantitative evaluation difficult to achieve (Dent and Young, 1981). NIRS provides new opportunity for quantifying soil properties and variability for the development of risk-based systems of soil interpretations that are designed to quantify prediction uncertainty so that users may be able to employ such information in decision-making (Dewayne Mays, 1996; McKenzie et al., 2000). More recently, Shepherd and Walsh (2002) developed a scheme for the development and use of soil spectral libraries for rapid estimation of soil properties based on analysis with this technique, using a library of over 1000 archived topsoils from eastern and southern Africa. Using a multivariate regression approach they calibrated 10 different soil properties to soil reflectance and developed screening tests for various soil fertility constraints using classification trees.

The ability to rapidly and nondestructively characterize soils using reflectance spectroscopy permits thorough sampling of the variation within a target population of soils (Stenberg et al., 1995). In a study conducted by Dalal and Henry (1986) using three wavelengths' absorbance (log (I/R)), they were able to predict total organic C, total N, and moisture content by multible linear regression ($r^2 = 0.86, 0.86$ and 0.92 respectively). Using similar approaches, Morra et al. (1991) concluded that NIRS could be used to predict soil C and N content. An important observation was made by Sudduth and Hummel (1996) that the predictions of soil properties became less accurate as the geographical range of samples increased. The statement underlies the basic principle of geostatistics that as the distance between sampling points increases,
the samples become more dissimilar or less autocorrelated (Cressie, 1993). Any calibration procedure requires that the variability in the target population is adequately sampled for robust prediction.

2.5 GIS, mathematical modeling and resource management

In recent years, the application of modern geographical information technology for handling spatial data has become widely used in developing site-specific policies and land management strategies (Dumanski and Craswell, 1998; Wood and Pardey, 1998). Geographical Information Systems (GIS) can be defined as a set of computer tools for collecting, storing, retrieving, transforming and displaying spatial information from the real world (Burrough and McDonnell, 1990). GIS has been particularly useful for visualising, querying and analysing spatial patterns of various geographical phenomena. New possibilities have been arising as GIS is increasingly integrated with remote sensing, spatial statistics and spatial simulation models (Ruecker et al., 2003). Supplementing application-oriented aspects such as spatial analysis to identify policy options, GIS can be characterised as a decision support system involving the integration of spatially referenced data in a problem solving and planning environment at different spatial scales (Kenneweg, 1992; Cowen, 1988).

One of the main problems in natural resource management is how to integrate natural resource data with socio-economic data to analyse complex agricultural regions for targeting policy and land management strategies (Ruecker et al., 2003). This raises the question of how to model the complex interactions among various natural and socio-economic process components for certain spatial domains. The other problem, is identification of suitable alternative land uses, either because the existing land use pattern currently provides an acceptably low standard of living, or because it is damaging to some aspect of the environment (UNEP, 1997).

Computer-based, mathematical models that realistically simulate spatially distributed, time-dependent environmental processes in nature are increasingly recognised as useful tools for, quantitative assessment for complex environmental issues of local, regional and global concern (Steyaert, 1993). These environmental simulation models provide diagnostic and predictive outputs that can be combined with socio-economic
data for assessing local and regional environmental risk or natural resource management issues. The simple act of defining a land use in terms of a sequence of activities undertaken to produce goods or services creates a very powerful tool in this respect (UNEP, 1997). By arranging the activities in tabular form, together with their associated inputs, it becomes a simple matter to define any land-use or production system quantitatively, carry out an economic analysis or an assessment of management and capital requirements, model environmental impact or map land use. It also provides a framework for modelling alternatives and improvements to the production system and land use. However, lack of reliable sources of spatial data often impedes such approaches.

2.6 Geostatistics and spatial prediction

Ecological systems are spatially heterogeneous, exhibiting considerable complexity and variability in time and space (Gustafson, 1998). The quantification of such environmental heterogeneity has long been an objective in ecology (Patil et al., 1971; Pielou, 1977), but efforts to develop appropriate quantification methods began around two decades ago (Romme, 1982; Burrough, 1986). It's only in recent years that the process has accelerated so that there are now literally hundreds of quantitative measures of landscape pattern that have been proposed to quantify various aspects of spatial heterogeneity (Baker and Cai, 1992; McGarigal and Marks, 1995). Geostatistics is one such method which provides descriptive tools such as semivariograms to characterize the spatial pattern of continuous and categorical soil attributes (Goovaerts, 1999).

As a rapidly evolving branch of mathematics, geostatistics originated in the mining industry in the early 1950s to help improve ore reserves calculation (Wackernagel, 1998). The first steps were taken in South Africa, with the work of the mining engineer Krige and statistician Sichel (Cressie, 1993). Webster (1985) reported that geostatistics is a method to analyze spatial variability of properties. It provides the basis for describing spatial variation in soil quantitatively, estimating soil properties and mapping them, and for planning rational sampling schemes that makes the best use of manpower and are cost effective (Kamaruzaman, 2004). Generally, the understanding of the spatial distribution of soil chemical properties particularly total nitrogen (N), available phosphorous (P), and exchangeable potassium (K) are
important for refining farm management practices and for assessing the impact of agriculture on the environment (Eltaib et al., 2002).

The variability of soil properties within fields is often described by assuming that variation is randomly distributed within mapping units. In addition, often only one soil profile is taken as being representative of a soil mapping unit and few if any additional point samples are usually reported. Soil variability is the outcome of many processes acting and interacting across a continuum of spatial and temporal scales and is inherently scale dependent (Parkin, 1993). In addition, soil properties frequently exhibit spatial dependency (Eltaib et al., 2002). Therefore, parametric statistics are inadequate for analysis of spatially dependent variables because they assume that measured observations are independent in spite of their distribution in space (Hamlett et al., 1986) Geostatistical analyses have been used to estimate spatial variability of soil physical properties (Viera et al., 1981; Lascano and Hatfield, 1992), and in soil biochemical properties (Bonmati et al., 1991; Sutherland et al., 1991).

Geostatistical data, also called random field data, consists of measurements taken at fixed locations (Kaluzny et al., 1998). Data analysis consists of variography and kriging. Variography uses variograms (semivariograms) to characterize and model the spatial variance of data, whereas kriging uses the modeled variance to estimate values at unsampled locations (Mohammadi, 2002). The process of geostatistical analysis can be divided into the following sections; estimating variograms, fitting theoretical variogram models, performing ordinary and universal kriging, and simulating geostatistical data (Kaluzny et al., 1998).

2.6.1 Variogram Estimation
Geostatistical data typically exhibit small-scale variation that may be modeled as spatial autocorrelation and incorporated into estimation procedures. The variogram provides a measure of spatial correlation by describing how sample data are related with distance and direction (Kaluzny et al., 1998). Geostatistical analysis of soil properties is based on the assumption that a variable, z, measured at a location, x, may be treated as a realization of a random function, denoted by Z(x) (Lark, 2000). The analysis is possible if the random function is intrinsic, that is if
\[ E[Z(x) - Z(x + h)] = 0, \]  
and  
\[ 2\gamma(h) = E[\{Z(x) - Z(x + h)\}^2] \]

depends only on the spatial separation or lag \( h \). The function \( \gamma(h) \) is the variogram.

An estimate of the variogram is needed in geostatistics for estimation (kriging) (Burgess and Webster, 1980), simulation (Papritz and Webster, 1995), and the design of optimal sampling strategies (McBratney et al., 1981). It is usually obtained by estimating \( \gamma(h) \) for discrete lags, then fitting a suitable function of lag to these estimates (McBratney and Webster, 1986). The most widely used estimator of the variogram is due to Matheron (1962):

\[ 2\hat{\gamma}_M(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \{z(x_i) - z(x_i + h)\}^2, \]  
where \( N(h) \) pairs of observations among the available data are separated by lag \( h \). This estimator is asymptotically unbiased for any intrinsic random function (Cressie, 1993), but because it is based on squared differences among data, it is very sensitive to outlying values of \( z \). A single outlying datum can distort the estimate of the variogram since it occurs in several paired comparisons over many or all lag intervals. This in turn affects calculated kriging variances, the variability of simulated data and the spacing of sampling grids designed using the variogram (Lark, 2000).

Variogram estimation is a crucial stage of spatial prediction, because it determines the kriging weights. It is important to have a variogram estimator which remains close to the underlying variogram, even if outliers (faulty observations) are present in the data (Genton and Furrer, 1998). A robust estimator was developed by Cressie and Hawkins (1980). It forms the basis of the variogram estimator \( \hat{\gamma}_{CH}(h) \) where:

\[ 2\hat{\gamma}_{CH}(h) = \left\{ \frac{1}{N(h)} \sum_{i=1}^{N(h)} \{z(x_i) - z(x_i + h)\}^2 \right\}^{\frac{4}{494}} \cdot \frac{0.494}{0.457} + \frac{0.045}{N(h)} \cdot \frac{0.457}{N^2(h)}. \]  
The estimate of location (in braces) is rescaled to the scale of the variogram by the fourth power. The denominator of the equation above is a correction based on the assumption that the underlying process to be estimated has normally distributed differences over all lags. The estimator has been used in some published geostatistical...
analyses (Shouse et al., 1990; Sullivan, 1991; Birrel et al., 1996), and some of the analyses of soil data.

2.6.2 Fitting Variogram Models
The experimental variogram is replaced by a theoretical variogram function essentially for the reason that the variogram model should have a physical meaning (Wackernagel, 1998). Variogram fitting is thus, another crucial stage of spatial prediction, because it also determines the kriging weights (Genton and Furrer, 1998). According to Kaluzny et al. (1998), the main goal of variogram analysis is to construct variograms that best estimate the autocorrelation structure of the underlying stochastic processes. Most variograms are defined through parameters; namely, the nugget effect, sill, and range. The parameters are depicted on the generic variogram shown in Figure 2.1 and are defined as follows:

- nugget effect – represents micro-scale variation or measurement error. It is estimated from the empirical variogram as the value of $\gamma(h)$ for $h=0$.
- sill – the $\lim_{h \to \infty} \gamma(h)$ representing the variance of the random field.
- range – the distance (if any) at which data are no longer autocorrelated.

![Figure 2.1: A Theoretical Semivariogram Model](image)

The use of a theoretical variogram guarantees (using weights subject to a certain constraint) that the variance of any linear combination of sample values is positive (Wackernagel, 1998).
2.6.3 Kriging

Kriging is a generic name adopted by the geostatisticians for a family of generalized least-squares regression algorithms (Goovaerts, 1999). It is a linear interpolation method that allows predictions of unknown values of a random function from observations at known locations (Kaluzny, 1998). It is possible to define the best domain in which to interpolate (i.e. the extent to which data should be considered in order to get an optimal interpolation at a given point), it defines the shape and orientation for optimal interpolation, estimates the \( \lambda_i \) weights in a more rigorous way than a mere function of distance and also makes it possible to estimate the error associated with each interpolated value (Rute, 2003). The \( \lambda_i \) weights are estimated through the general interpolation formula:

\[
x_p = \sum_{i=1}^{n} \lambda_i x_i ; \sum_{i=1}^{n} \lambda_i = 1.
\]

Kriging is called an optimal interpolation method because the interpolation weights \( \lambda_i \) are chosen to provide for the value at a given point the Best Linear Unbiased Estimate (BLUE). Kriging rests on the recognition that the spatial variation of any geological, soil, or hydrological property, i.e. any "regionalized variable", is too irregular to be modeled by a smooth mathematical surface, but these phenomena can be better described by a stochastic surface (Burrough, 1990).

Ordinary kriging is the most widely used kriging method. It serves to a value at a point of a region for which a variogram is known, using data in the neighborhood of the estimation location (Wackernagel, 1998). Weights are chosen to ensure that the average error for the model is zero and that the modeled error variance is minimized (Isaaks and Srivastava, 1989). The ordinary kriging variance is the sum of the simple kriging variance (assuming a known mean) plus the variance due to the uncertainty about the true value of the mean. When the weight of the mean is small, the sum of the weights of simple kriging is close to one and ordinary kriging is close to the simple kriging solution, provided the variance of the kriged mean is also small (Wackernagel, 1998).

To account for the accuracy of the soil map information in mapping, Heuvelink and Bierkens (1992) proposed a weighted average of soil map predictions and kriged
estimates where more weight is given to the information with the smallest prediction error variance. A case study showed this heuristic method to produce a more accurate map of mean water table than either kriging or soil map prediction as long as the soil map accuracy is correctly estimated and point observations are scarce (Goovaerts, 1999). In general, mapping is most efficient if survey is done on a regular grid in the sense that the maximum kriging error is minimized (Webster and Oliver, 2000).

An example of the use of geostatistical analysis of soil properties is a study in Malaysia, where research was carried out to determine the spatial variability of N, P, and K in a rice field in the Sawah Sempadan region (Eltaib et al., 2002). Results showed that soil chemical properties measured at the two depths (0-20 and 20-30 cm) had high spatial dependencies even within small distances. The semivariogram values for total N, available P, and exchangeable K are shown in Table 2.1.

Table 2.1: Characteristic of calculated semivariogram of spatial fertility.

<table>
<thead>
<tr>
<th>Soil Property</th>
<th>Depth</th>
<th>Model</th>
<th>Range (A)</th>
<th>Nugget Variance (C0)</th>
<th>Sill variance (C0+C)</th>
<th>Ratio nugget/sill</th>
<th>Spatial dependence level</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>0-20</td>
<td>spherical</td>
<td>12</td>
<td>0.00002</td>
<td>0.0172</td>
<td>98</td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td>20-30</td>
<td>spherical</td>
<td>13</td>
<td>0.00043</td>
<td>0.0164</td>
<td>97</td>
<td>Weak</td>
</tr>
<tr>
<td>P</td>
<td>0-20</td>
<td>spherical</td>
<td>12</td>
<td>13.00000</td>
<td>229.2000</td>
<td>94</td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td>20-30</td>
<td>spherical</td>
<td>38</td>
<td>109.80000</td>
<td>260.0000</td>
<td>59</td>
<td>Moderate</td>
</tr>
<tr>
<td>K</td>
<td>0-20</td>
<td>spherical</td>
<td>31</td>
<td>61.00000</td>
<td>477.6000</td>
<td>72</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>20-30</td>
<td>spherical</td>
<td>32</td>
<td>127.00000</td>
<td>1075.0000</td>
<td>88</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

Source: Eltaib et al. 2002.

The nugget (C0) varied among the two soil layers with the values of 0.0002% and 0.0004% of total N, 13.0 and 109.8 mg/kg for available P and 61.0 and 127.0 mg/kg for exchangeable K, for top- and sub-soil layers, respectively. The values of the sill variance for all nutrients varied between 0.01724 and 0.01636 % for total N, 229 and 260 mg/kg for available P and 47 and 1075 mg/kg for exchangeable K for top- and sub-soil layers, respectively. The range also varied between 12 and 38m for available P and 31 and 32 m for exchangeable K for both surface and sub-soil layers, respectively. The results showed that high nugget effects were found for all the
to quantify relief form analysis in order to define basic relief units for geomorphological and pedological mapping. The main topographic attributes used to define these relief units were slope, plan curvature, and profile curvature. This approach provides a systematic and methodological basis for delineation of complex relief units. It may be possible to use these relief units to stratify the measured soil attributes and separate the micro- and meso-scale spatial variabilities that are important in delineating soil fertility domains.

Land cover (LC) is a critical component of most surface/subsurface models, whether they deal with plant growth, hydrology, erosion, or salinisation. Land cover is a more useful variable than land use (LU) because land use is an economic concept and a given land use can be associated with a variety of land covers (Gersmehl et al., 1987). In addition, land cover is a dynamic quantity that often exhibits great seasonal variation. Land cover data is most commonly obtained from satellite imagery such as Landsat Thematic Mapper (TM) and Multi-Spectral Scanner (MSS) or SPOT 1 and visual photo interpretations (Moore et al., 1993).

In recent years, spatial dependence models of geostatistics have gained popularity as they allow the quantification of landscape spatial structure from point-sampled data. One such model that has received much attention and will be used in this study is the variogram (Cresse, 1993). The variogram reveals the randomness and structured aspects of the spatial dispersion of a given variable and is a plot of the average squared differences between the values of a spatial variable at pairs of points separated by a lag distance against the lag (Davidson and Csillag, 2003). The empirical variogram describes the overall spatial pattern of sample data (Fortin, 1999) and a variety of theoretical variogram models can be fitted on it to describe spatial structure of a landscape attribute. These then provide powerful capabilities which can be used to analyse realistically the complex spatial relationships in ecological systems.
CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Study area

3.1.1 Location

This study was conducted in two districts; Vihiga and Siaya in western Kenya as shown on the map (Figure 3.1). Although, the two districts are adjacent to each other, they vary distinctly in terms of climatic, physical, demographic and administrative factors. Vihiga is one of the six districts in Western province, situated to the north-east of Siaya and largely covers the upper parts of the Lake Victoria Basin extending further north to border Kakamega forest. Siaya district on the other hand covers the lower parts of the basin extending to the lake and is administratively located in Nyanza province.

Geographically, Vihiga district lies between longitude 34° 30' and 35° 00' east, and between latitude 00° 00' and 00° 15' north. The Equator cuts across the southern tip of the district. It borders Kakamega district to the north, Nandi district to the east, Kisumu and Siaya districts to the south (GOK, 1997). The district is 33 km wide from east to west and 19 km wide from north to south and occupies a total area of 541 sq. km. Siaya district on the other hand lies between longitude 33° 58' and 34° 33' east, and latitude 0° 26' south and 0° 18' north. The total area of the district is 3,523 km², out of which about 1,005 sq. km is lake water, under the lakes Sare, Kanyaboli and parts of L. Victoria (GOK, 1997). It is bordered by Busia district to the north, Vihiga and Kakamega districts to the north-east, Kisumu to the south-east, Homa-Bay across the Winam Gulf to the south and L. Victoria to the west.
3.1.2 Topography and Climate

Vihiga district lies on the eastern fringes of the L. Victoria's lake basin. Its altitude ranges from 1300 to 1500 m above sea level and slopes gently from west to east. It is characterized by undulating hills and valleys with a fast network of streams and brooks which are the tributaries of the two main rivers, namely, River Esalwa (Edzawa) and River Yala. These rivers flow along the general gradient of the district from north east to south west and drain into Lake Victoria. The district's annual rainfall ranges from 1800 to 2000 mm and it is described as being adequate and reliable. It is bimodaly distributed with peaks between April and June (long rains) and September to November (short rains). Temperatures in the district range from 14° to 32° C. Diurnal temperature changes are low.

On the other hand, Siaya district lies within the lake basin extending to the northern shores of the lake. The land is mainly a peneplain and slopes very gently from east to west (Jaetzold and Schmidt, 1982). Its altitude ranges from 1,140 to 1,500 m above sea level. Yala Division is located at the higher end of this range. The landscape is characterised by undulating and rolling uplands, with slopes varying from 2 to 16% (Mango, 1999). The district is traversed by the Yala and Nzoia rivers both of which drain southwards through the district and enter L. Victoria through the Yala swamp. The district is drier to the south along the shores of L. Victoria but progressively gets wetter towards the hinterland as the altitude increases. On the highlands, the annual rainfall ranges from 1800 to 2000 mm, while on the lower areas the rainfall ranges from 800 to 1600 mm. The rainfall pattern is bimodal with the long rains occurring from March to June and the short rains from September to December. The temperature varies with altitude. The higher lands to the north-east have a mean of
21° C, while to the south-west on the shores of L. Victoria it is much higher with an average of 22.5° C.

Figure 3.1: Map of the study area showing the location of the nine Y-frame sampling units.
3.1.3 Geology and Soils
The geological formation of Vihiga district comprises mainly the Nyanzian rock system. These rocks are granitic and are found in parts of Tiriki East and West, Emuhaya and Vihiga divisions. The soils in the district include the well drained, dark red friable soils partially covered with humic top soils derived from both volcanic and basement complex rocks and the yellow red loamy sands derived from both sediments and basement rocks. The dominant soils are the humic Nitisols, the dystro-mollic Nitisols and the nitro-humic Ferralsols (FAO, 1995a). Vertisols occur as isolated pockets along rivers and valley bottoms. The district is divided into two main agricultural zones. They are the Upper Midland zone (UM1) and Lower Midland zone (LM1) (Jaetzold and Schmidt, 1982). The Upper Midland zone has fertile, well drained, dark red soils which support the growing of tea, coffee, finger millet and cassava. The zone covers the western slopes of Nandi escarpment and parts of Central Sabatia, Vihiga and Tiriki Divisions. In total the upper midland zone covers 90% of the district. The Lower Midland zone (LM1) covers the western parts of Emuhaya. The soils are loamy sands which have been derived from sediments and basement rocks. These soils are suitable for the growing of sugarcane, maize, coffee, beans, finger millet and sorghum.

In Siaya district, most soils are underlain by plinthite (murrum) at a shallow depth, resulting in low moisture retention. Dominant soils are developed on basic igneous rocks. Humic Gleysols dominate the lower regions including the Yala swamp, while Ferralo-orthic Ferralsols, orthic Acrisols and chromic Luvisols are the most common soils on the higher areas (FAO, 1995a). The district is divided into four major agroecological zones that are delineated across from north-south towards the lake (Jaetzold and Schmidt, 1982). The lower midland sugarcane zone (LM1) with a long
cropping season followed by medium or intermediate rains. It has a very good yield potential. The second is the marginal sugarcane zone (LM2) with a long cropping season followed by a weak medium to short rains. It has a fairy good yield potential. The third zone is the lower midland cotton zone (LM3) with a medium to long cropping season followed by a (weak) short or very short one. The forth agroecological zone extends up to the lake, the marginal cotton zone (LM4) with a (weak) medium to short cropping season and intermediate rains.

3.2 Design of the study

The study was carried out in nine sub-locations selected randomly from the two districts, five from Vihiga and four from Siaya. In each of the selected sub-locations, a Y-frame (Figure 3.2) sampling design was used to select ten (10) farms where actual sampling was done for the various biophysical characteristics and socio-economic interviews. The three arms of the Y-frame were orientated in such a way that it covered the largest part of the sub-location, while the angle between adjacent arms of the Y was 120 degrees. Farms were located at constant lag distances along the arms of the Y. This design gave the optimal arrangement for geostatistical analysis in terms of generating a range of distances between farms with the minimum number of plots.

The centre of the Y-frame formed the central farm where sampling for each Y began. The exact location of the central farms was determined using the global positioning system (GPS). Consequently, the direction of each arm was given in terms of degrees in reference to the True North. In each farm, both biophysical and socio-economic information was collected.
The Y sampling Frame

Y_C3

Y_C2

Y_C1

Y_0

Y_B1

Y_B2

Y_B3

Y_A3

Y_A2

Y_A1

Key:
Y_0 - Shows farms on the sampling frame.
A, B & C - Each of the 3 arms of the Y separated by 120°.
Scale - Distance from one farm to the next

Y arm Scale:

0m 100m 200m 600m

Figure 3.2: The Y Sampling Frame
3.3 The sample and sampling procedure

Along each of the three arms of the Y-frame, three farms were selected systematically at intervals of 100, 300 and 900 metres from the centre of the Y. In total, ten farms were selected for sampling in each of the nine Ys making a total of 90 farms. Each farm was in turn divided into fields depending on type of crops grown and how it is managed. Such that two areas with the same crop but with different management interventions were treated as different fields and sampled separately.

The exact locations of all the farms on the Y-frame were determined using GPS coordinates. All fields (primary production units) (PPUs) within a farm were sampled and the farmer (household head) or key informant interviewed.

3.4 Data collection

A combination of physical survey techniques and participatory rural appraisal (PRA) were used for data collection. Data on biophysical aspects of the farm was obtained mainly using field observations, ground measurements and soil sampling. The basic unit for soil sampling in each field was a slope-directed transect located within a 5 x 5 m plot. The location of this sampling plot within a field was determined using a table of computer generated random numbers showing direction in degrees and percent distance from centre of the field. The random number allocation was designed to be proportional to the relative areas of a plot represented at different distances. This procedure ensures that edge effects are not under-sampled. At the centre of the plot, soil was sampled using a cylindrical Eddman auger at two depths, 0-20 cm and 20-50 cm. Two other soil samples were obtained at 2 metres up- and down-slope at depths of 0-20 cm only. All the soil samples collected were air-dried, crushed, passed
through a 2 mm sieve, and weighed before being taken to the laboratory for further analysis.

3.5 Spectral analysis of soils

Diffuse Reflectance Spectrometry (DRS) was used for rapid assessment of soil quality as outlined by Shepherd and Walsh (2002) and using the optical set-up described by Shepherd et al. (2003). Reflectance spectra were determined on all samples but formal laboratory analysis was conducted only on a subset of samples for the validation of soil fertility indices. The reflectance spectra were recorded for all the soil samples using a FieldSpecTM FR spectroradiometer (Analytical Spectral Devices, Boulder, Co) at wavelengths from 0.35 to 2.5 μm with a spectral sampling interval of 0.001 μm. During scanning enough soil from each sample was placed in a 7.4 cm diameter Duran glass Petri dish to give a sample thickness of about 1 cm. The samples were scanned through the bottom of the Petri dishes using a high-intensity source probe (Analytical Spectral Devices, Boulder, Co). The probe illuminates the sample (4.5W halogen lamp giving a correlated color temperature of 3000K; Welchalllyn, Scaneattles Falls, NY) and collects the reflected light from a 3.5 cm diameter sapphire window through a fiber-optic cable.

To sample within-dish variation, reflectance spectra were recorded at two positions, by successively rotating the sample dish through 90° between readings. Reflectance readings for each wavelength band were expressed relative to the average of the white reference readings (Shepherd et al., 2003). Two soil fertility indices were derived from the reflectance spectra: (1) a soil fertility index, based on a multivariate distribution of laboratory-measured soil properties, and (2) a soil condition index, based on Mahalanobis distance of the soil spectra from a population of sediment spectra taken from the Winam Gulf in Lake Victoria. The latter index is closely
related to erosion-deposition processes over the past 40 years. Both indices have been
developed from a diverse soil spectral library of western Kenyan soils
(http://www.worldagroforestrycentre.org/sites/program1/specweb/home.htm). Based
on the recorded spectral signatures, representative samples were selected for
validation of spectral indices using standard laboratory soil chemical analysis
techniques.

Apart from soil, other biophysical factors in the field were characterised. These
included slope (%), observations on soil erosion and hardsetting, soil conservation
structures, flooding, and rock cover. Participatory rural appraisal (PRA) using a guide
questionnaire was used to collect information on the socio-economic aspects of the
farm (see appendix 3).

3.6 Soil chemical analysis
A representative number of soil samples were selected based on the predicted soil
fertility index (SFI) and analyzed using standard methods widely used for tropical
soils (Shepherd and Walsh, 2002). Soil pH was determined in water using a 1:2.5
soil/solution ratio. Samples were extracted with 1 M KCL using a 1:10 soil/solution
ratio, and analyzed by NaOH titration for exchangeable acidity and by atomic
absorption spectrometry for exchangeable Ca and Mg, and exchangeable Na by flame
photometry (ISFEIP, 1972; Yurimaguas Experiment Station Staff, 1989). Samples
with a pH ≥ 5.5 were assumed to have zero exchangeable acidity and samples with pH
< 7.5, zero exchangeable Na. Samples were extracted with 0.5 M NaHCO₃ + 0.01 M
EDTA (pH 8.5, modified Olsen) using a 1:10 soil solution ratio and analyzed by
flame photometer for exchangeable K and colorimetrically (molybdenum blue for
extractable P (ISFEIP, 1972; Yurimaguas Experiment Station Staff, 1989). Organic C was determined colorimetrically after H$_2$SO$_4$ dichromate oxidation at 150°C (Heanes, 1984). Particle-size analysis was determined using the hydrometer method after pretreatment with H$_2$O$_2$ to remove organic matter (Gee and Bauder, 1986). Effective cation-exchange capacity was calculated as the sum of exchangeable bases, and ECEC$_{clay}$ was calculated as ECEC divided by the clay fraction.

3.7 Data analysis.

Conventional statistical methods are generally inadequate to describe data that are spatially correlated. Thus, this study used regionalized variable theory, popularly known as geostatistics (Matheron, 1971), a methodology for the analysis of spatially correlated data. The data obtained was analysed using exploratory statistics, geostatistics, linear regression, and mixed-effects modelling. The four basic steps were: data entry and checking, exploratory analysis (cross-tabulations, scatter plots, correlation matrix, box and whisker plots, frequency distributions), geostatistical analysis (semivariograms), statistical analysis, and modelling.

Exploratory analysis was aimed at helping to identify outliers and non-normally distributed variables. It was also used to identify trends among the variables and as a screening tool for identification of variables to be included in further analyses. Geostatistical analysis was used to examine the spatial correlation structure of the soil fertility indicators. The maximum lag indicates the range of spatial influence, beyond which samples are spatially independent. Mixed-effects model was used to confirm relationships between soil fertility index and the different site characteristics (e.g. slope, market distance, and management factors), accounting for within-group
correlation at different spatial scales, and to study the variance in soil fertility at different spatial scales. MS-Excel was used for data handling, while GenStat and S+Plus for statistical analysis.

3.7.1 Geostatistical Analysis

Geostatistical analysis of soil properties is based on the assumption that a variable, \( z \), measured at a location, \( x \), may be treated as a realization of a random function, denoted by \( Z(x) \). The analysis is possible if the random function is intrinsic, that is if

\[
E[Z(x) - Z(x+h)] = 0, \tag{6}
\]

and

\[
2\gamma(h) = E[(Z(x) - Z(x+h))^2] \tag{7}
\]

depends only on the spatial separation or lag \( h \). The function \( \gamma(h) \) is the variogram.

An estimate of the variogram is needed in geostatistics for estimation (kriging) (Burgess and Webster, 1980), and simulation modeling. Although the classic estimator is the most commonly used and is asymptotically unbiased for any intrinsic random function (Cressie, 1993), it was however not used in this study because it is very sensitive to outlying values. Instead the more robust estimator \( \hat{\gamma}_{CH}(h) \) was used as in the formula below,

\[
2\hat{\gamma}_{CH}(h) = \left\{ \frac{1}{N(h)} \sum_{i=1}^{N(h)} \left| z(x_i) - z(x_i + h) \right|^2 \right\}^4, \tag{8}
\]

The advantage of this robust estimator is that the effect of outliers is reduced, without removing specific data points from a data set. Data from each Y-frame sampling region was modeled using the above robust estimator to produce an experimental
semivariogram which was then fitted with both the spherical and exponential semivariogram models. The spherical function is given by

$$\gamma(h) = c_o + c \left\{ \frac{3h}{2a} - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right\},$$

with the parameter values $c_o$, $c$ and $a$ being fitted from the developed experimental variograms. For the exponential models, the formula is given as

$$\gamma(h) = c \left\{ 1 - \exp \left( -\frac{h}{r} \right) \right\},$$

with sill, $c$, and a distance parameter, $r$, that defines the spatial extent of the model. The function has an important place in statistical theory and it represents the essence of randomness in space (Webster and Oliver, 2000). It is the variogram of first-order autoregressive and Markov processes.

Kriging was only done on the Y-frame data sets whose spatial variability ranged within the Y-frame sampling area. Ordinary kriging in 2-dimensions was performed in S+SPATIALSTATS module of S+plus. It assumes that the mean is unknown and estimates the value of a random variable, $Z$, at an unsampled point, $x_o$, by $\hat{Z}(x_o)$, with the same support as the data, by

$$\hat{Z}(x_o) = \sum_{i=1}^{N} \lambda_i z(x_i),$$

a weighted average of the data, where $\lambda_i$ are the weights. To ensure that the estimate is unbiased, the weights are made to sum to 1,

$$\sum_{i=1}^{N} \lambda_i = 1,$$

and the expected error is $E[\hat{Z}(x_o) - Z(x_o)] = 0$. The estimate variance is

$$\text{var} [\hat{Z}(x_o)] = E \left[ \left( \hat{Z}(x_o) - Z(x_o) \right)^2 \right]$$
\[ 2 \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_0) - \sum_{i=1}^{N} \sum_{j=i}^{N} \lambda_i \lambda_j \gamma(x_i, x_j), \]  

where \( \gamma(x_i, x_j) \) is the semivariogram of \( Z \) between the data points \( x_i \) and \( x_j \), and \( \gamma(x_i, x_0) \) is the semivariogram between the \( i \)th data point and the target point \( x_0 \).

### 3.7.2 Mixed Effects Modelling

Mixed-effects models provide a powerful and flexible tool for analyzing grouped data. The models are primarily used to describe relationships between a response variable and some covariates in data that are grouped according to one or more classification factors. A mixed-effects model incorporates both fixed and random effects. Fixed effects are factors or parameters that are associated with an entire population, or with repeatable levels of experimental factors, and in this study; they comprise all the biophysical and socio-economic factors measured at the farm or field level. Random effects on the other hand are parameters associated with experimental units drawn at random from a population. In this study, sub-locations (\( Y_s \)), farms and fields within farms were treated as random effects.

A multilevel linear mixed-effects (LME) model was used to analyse the data. The model was a three-level nested model as outlined by Pinheiro and Bates (2000) and is shown in equation (14). In the 3-level model, the response for the \( k \)th level-3 unit within the \( j \)th level-2 unit within the \( i \)th level-1 unit is written as;

\[ y_{ijk} = x_{ijk} \beta + z_{ijk} b_i + z_{ijk} a_i b_i + z_{ijk} a_i b_{ijk} + \varepsilon_{ijk}, \]  

\[ i = 1, \ldots, M, \ j = 1, \ldots, M_i, \ k = 1, \ldots, M_{ij}, \]

where the fixed effects model matrices are \( x_{ijk} \), \( i = 1, \ldots, M \), \( j = 1, \ldots, M_i \), and \( k = 1, \ldots, M_{ij} \), of size \( n_{ijk} \times p \); the first-, second-, and third-level random effects are \( b_i, b_{ij}, \) and \( b_{ijk} \).
and $b_{ijk}$ of length $q_1$, $q_2$ and $q_3$, with the corresponding model matrices $z_{i,j,k}$, $z_{i,j,k}$ and $z_{i,j,k}$ of sizes $n_i \times q_1$, $n_i \times q_2$, and $n_i \times q_3$. The within-group errors $e_{ijk}$ are assumed to be independent for different $i$, $j$ or $k$ and to be independent of the random effects.

The parameters were estimated using the restricted maximum likelihood (REML). The within-group error covariance structure was modeled by combining the variance correlation structures and variance functions. Correlation structures are used to model within-group correlations that are not captured by the random effects. These are generally associated with temporal or spatial dependencies (Pinheiro and Bates 2000), but in this study only spatial dependencies were dealt with at the district, Y (sub-location), and farm. The variance functions were used to model heteroscedasticity in the within-group errors.

The spatial correlation structure observed in the data was represented by their semivariogram, instead of their correlation function. Since the empirical autocorrelations at larger lags tend to be less reliable, because they are estimated with fewer residual pairs, the number of lags with which to calculate the empirical autocorrelations was controlled using the $\text{maxdist}$ (maximum distance with reliable estimates) argument. The semivariogram was then modeled using the spherical, exponential and ratio-quadratic models.
CHAPTER FOUR
RESULTS AND DISCUSSION

4.0 Introduction
In this chapter the results of all the data analysed are presented and discussed in seven sections. The first two sections give the general summary of the data using simple quantitative techniques of means and variances, and the results presented in tables, graphs and box plots. Section three shows the spatial trends in the data and how it’s modelled using semivariograms. In section four, the spatial trends observed in section three are interpolated using kriging techniques. Some of the factors that are responsible for generating such spatial trends are considered in section five, while, the last two sections introduce the mixed-effects model and show how it accounts for variances at different spatial scales.

4.1 Farm Characterisation

4.1.1 Field Characteristics
All the data obtained from the two districts using the Y sampling frame was cross-checked and quality controlled before further analysis was done. Table 4.2 gives a summary of the farms and fields sampled and their distribution in the different study sites. A total of 745 fields were sampled in 90 farms from the two districts.

Table 4.2: Summary of Fields Sampled

<table>
<thead>
<tr>
<th>District</th>
<th>Y</th>
<th>No. of Farms</th>
<th>No. of Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vihiga</td>
<td>Y1</td>
<td>10</td>
<td>75</td>
</tr>
<tr>
<td>Vihiga</td>
<td>Y2</td>
<td>10</td>
<td>79</td>
</tr>
<tr>
<td>Vihiga</td>
<td>Y3</td>
<td>10</td>
<td>85</td>
</tr>
<tr>
<td>Vihiga</td>
<td>Y4</td>
<td>10</td>
<td>73</td>
</tr>
<tr>
<td>Siaya</td>
<td>Y5</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>Siaya</td>
<td>Y6</td>
<td>10</td>
<td>82</td>
</tr>
<tr>
<td>Siaya</td>
<td>Y7</td>
<td>10</td>
<td>97</td>
</tr>
<tr>
<td>Siaya</td>
<td>Y8</td>
<td>10</td>
<td>99</td>
</tr>
<tr>
<td>Vihiga</td>
<td>Y9</td>
<td>10</td>
<td>75</td>
</tr>
</tbody>
</table>
On average, each farm had 7.8 fields in Vihiga district while Siaya district had an average of 9.0 fields per farm. Thus, farms in Siaya district had one or more fields as compared to those of Vihiga. The average area of fields per farm per Y is given in Table 4.3, with the corresponding distances between the fields. Generally farms in Vihiga district had smaller fields (< 0.25 acres) as compared to those of Siaya which were relatively bigger (> 0.25 acres). The average farm size in Vihiga district was 1.5 acres while in Siaya district it was 2.7 acres in line with a similar study carried out by Gikonyo and Smithson (2004).

**Table 4.3: Summary of Field Areas and Distances**

<table>
<thead>
<tr>
<th>Y</th>
<th>Ave. Area (acres)</th>
<th>Ave. Distance (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>0.15</td>
<td>63.0</td>
</tr>
<tr>
<td>Y2</td>
<td>0.24</td>
<td>94.2</td>
</tr>
<tr>
<td>Y3</td>
<td>0.23</td>
<td>81.5</td>
</tr>
<tr>
<td>Y4</td>
<td>0.17</td>
<td>71.7</td>
</tr>
<tr>
<td>Y5</td>
<td>0.31</td>
<td>99.4</td>
</tr>
<tr>
<td>Y6</td>
<td>0.31</td>
<td>63.7</td>
</tr>
<tr>
<td>Y7</td>
<td>0.30</td>
<td>105.2</td>
</tr>
<tr>
<td>Y8</td>
<td>0.28</td>
<td>134.3</td>
</tr>
<tr>
<td>Y9</td>
<td>0.11</td>
<td>47.8</td>
</tr>
</tbody>
</table>

The average distance between fields of the same farm in each of the Ys was also calculated (Table 4.3) and found to be closely correlated with the area ($r^2 = 0.66$) as shown in Figure 4.1. Measurement of field area and distances was done automatically using the GPS as shown in Plate 4.1.
The relationship in Figure 4.1 means that bigger fields are located further away from the homestead and according to Vanlauwe et al. (2001) such fields tend to receive less attention from the farmer in terms of management and application of inputs such as organic manures obtained from the farm.

Plate 4.1: Measuring the area of a field using a GPS system

The three types of land uses; commercial, subsistence and mixed-use were identified in the study area as shown in Table 4.4. Over 77% of all fields were used for...
subsistence farming only. Fields used for mixed farming accounted for 20%, while only 3% of the fields were used for commercial purposes. According to Scherr (1999) the type of land use tends to influence resource management decisions of the farmer. Fields used for commercial purposes get more inputs in terms of organic and inorganic resources while those for subsistence receive less or are neglected. The results also conform to those of Stoorvogel et al. (1993) and Ojiem et al. (2004) who also found out that smallholder farmers in western Kenya mainly practice subsistence farming with inadequate use of organic and inorganic fertilisers. Moreover, domination of a farming system by subsistence agriculture means that the farmers’ ability to purchase more inputs to replenish nutrients is limited (Stoorvogel, 1993) resulting in continuous soil degradation (Nandwa and Bekunda, 1998). The farmers are then forced into a vicious cycle between degraded soils and low soil fertility on one hand and reduced crop productivity and weakened ability to maintain soil fertility on the other.

**Table 4.4: Types of Land Use**

<table>
<thead>
<tr>
<th>Y</th>
<th>Commercial</th>
<th>Subsistence</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>3</td>
<td>59</td>
<td>13</td>
</tr>
<tr>
<td>Y2</td>
<td>12</td>
<td>54</td>
<td>13</td>
</tr>
<tr>
<td>Y3</td>
<td>0</td>
<td>52</td>
<td>33</td>
</tr>
<tr>
<td>Y4</td>
<td>11</td>
<td>45</td>
<td>17</td>
</tr>
<tr>
<td>Y5</td>
<td>0</td>
<td>60</td>
<td>20</td>
</tr>
<tr>
<td>Y6</td>
<td>0</td>
<td>55</td>
<td>27</td>
</tr>
<tr>
<td>Y7</td>
<td>0</td>
<td>89</td>
<td>8</td>
</tr>
<tr>
<td>Y8</td>
<td>1</td>
<td>94</td>
<td>4</td>
</tr>
<tr>
<td>Y9</td>
<td>1</td>
<td>66</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>574</td>
<td>143</td>
</tr>
<tr>
<td>% Total</td>
<td>3.8</td>
<td>77.0</td>
<td>19.2</td>
</tr>
</tbody>
</table>

The fields were also characterized depending on their position on the landscape as shown in Table 4.5. Most fields (over 75%) were located on the midslope, about 14% on the bottomland, 8% on the upslope and a few (3%) were located on the
drainage lines. The concentration of many fields on the mid-slopes has a direct impact on soil fertility degradation (McCloy, 1995), as such fields are prone to soil erosion.

Table 4.5: Location of the Fields on the Landscape

<table>
<thead>
<tr>
<th>Y</th>
<th>Upslope</th>
<th>Mid-slope</th>
<th>Bottomland</th>
<th>Drainage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>10</td>
<td>52</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Y2</td>
<td>16</td>
<td>55</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Y3</td>
<td>12</td>
<td>57</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>Y4</td>
<td>10</td>
<td>57</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Y5</td>
<td>0</td>
<td>77</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Y6</td>
<td>9</td>
<td>63</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Y7</td>
<td>2</td>
<td>63</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>Y8</td>
<td>1</td>
<td>80</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>Y9</td>
<td>0</td>
<td>58</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>562</td>
<td>103</td>
<td>20</td>
</tr>
<tr>
<td>% Total</td>
<td>8.1</td>
<td>75.4</td>
<td>13.8</td>
<td>2.7</td>
</tr>
</tbody>
</table>

The average slope was found to be 9.8% for Vihiga district and 5.7% for Siaya district (Table 4.6). Most of the large percentage slopes were observed in Vihiga district, while most of the lower slopes were observed in Siaya district.

Table 4.6: Average slope of Fields per Y

<table>
<thead>
<tr>
<th>District</th>
<th>Y</th>
<th>Slope (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vihiga</td>
<td>Y1</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>Y2</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Y3</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Y4</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Y9</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td><strong>Average</strong></td>
<td><strong>9.8</strong></td>
</tr>
<tr>
<td>Siaya</td>
<td>Y5</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>Y6</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>Y7</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>Y8</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td><strong>Average</strong></td>
<td><strong>5.7</strong></td>
</tr>
</tbody>
</table>

The slope differences between the two districts are clearly distinguishable (Figure 4.2). Vihiga district represented by Y1, Y2, Y3, Y4 and Y9 shows slopes well above 8
% while Siaya represented by Y5, Y6, Y7 and Y8 has slopes below 8%. Slope as a factor of topography acts as one of the factors that affect the spatial distribution of production systems (Templeton and Scherr, 1997). Steeper slopes such as those observed in a number of farms in Vihiga district, directly predispose such farms to increased soil erosion when holding other factors constant. Most fields in these farms are likely to suffer from reduced soil fertility due soil erosion if management interventions do not take measures of controlling soil erosion and conserving the soil.

![Slope per district](image)

**Figure 4.2:** Variation in average field slope per district

Assessment of soil fertility of individual fields by the farmers themselves rated only 13% of their fields as of high fertility, 52% as of medium soil fertility and 35% as of low fertility (Table 4.7). Although soil fertility can be fairly complex and heterogeneous from the spatial point of view, farmers know how to rate their farms (Baltissen et al., 2000; FAO, 1995; Smaling, 1997). Their assessment was based on local indicators such as soil color, structure, absence or presence of certain weeds and performance of crops on such fields.
Table 4.7: Farmers’ Rating of Field Soil Fertility

<table>
<thead>
<tr>
<th>Y</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>18</td>
<td>21</td>
<td>36</td>
</tr>
<tr>
<td>Y2</td>
<td>14</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>Y3</td>
<td>10</td>
<td>46</td>
<td>29</td>
</tr>
<tr>
<td>Y4</td>
<td>12</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td>Y5</td>
<td>7</td>
<td>40</td>
<td>33</td>
</tr>
<tr>
<td>Y6</td>
<td>11</td>
<td>49</td>
<td>22</td>
</tr>
<tr>
<td>Y7</td>
<td>17</td>
<td>48</td>
<td>32</td>
</tr>
<tr>
<td>Y8</td>
<td>7</td>
<td>83</td>
<td>9</td>
</tr>
<tr>
<td>Y9</td>
<td>0</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>96</td>
<td>387</td>
<td>262</td>
</tr>
<tr>
<td>% Total</td>
<td>12.9</td>
<td>51.9</td>
<td>35.2</td>
</tr>
</tbody>
</table>

Other field characteristics that were measured include number of soil conservation structures, number of seasons fields have been left fallow, soil hard setting, percentage rock/stone cover and their sizes (grade), and duration of flooding in a year where applicable (Table 4.8). These are some of the factors that determine whether agroecosystems are sustainable or soil degradation will take place (Hilhorst and Toulmin, 2000; Kimani et al., 2003). No soil conservation structures (terraces, vegetative e.g. napier, or crop residues) were found in 444 fields, whereas 206 fields had only one soil conservation structure, 67 fields had two and 18 fields had 3 or more structures (Table 4.8).

Table 4.8: Some physical field characteristics

<table>
<thead>
<tr>
<th>Category</th>
<th>Structures</th>
<th>Fallow</th>
<th>Hardsetting</th>
<th>Rock cover</th>
<th>R. Grade</th>
<th>Flooding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>444</td>
<td>636</td>
<td>616</td>
<td>703</td>
<td>39</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>206</td>
<td>30</td>
<td>124</td>
<td>24</td>
<td>7</td>
<td>732</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>33</td>
<td>5</td>
<td>9</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>6</td>
<td>2</td>
<td>631</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>18</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Structures = Soil conservation structures, R. grade = Rock grade

A total of 118 fields were fallow between 1-7 seasons, although most these were still being used for temporary grazing of animals. Permanent soil hard setting was only
observed in five fields, temporally hard setting in 124 fields, while the rest had no signs of hard setting. The percentage rock cover per field was absent or minimal at 0-5% in 703 fields, with higher percentages occurring only in fewer than 42 fields, mainly concentrated in Y3 of Vihiga district. Over 732 fields had no sign of flooding, while 13 were prone to flooding for a period of 4 or more months in a year.

4.1.2 Spectral Soil Analysis

Three soil properties and two indices were successively predicted using the soil spectral libraries developed by Shepherd and Walsh (2002) for rapid estimation of soil properties in east and southern Africa. The spectroradiometer used in determining the reflectance values of soil samples through scanning is shown in Plate 4.1.

Plate 4.2: Scanning soil samples using a spectroradiometer

Three soil properties, soil organic C, extractable P and total N, and two indices, soil fertility index (SFI) and soil degradation index (dseds) were determined (Table 4.9). In all the Ys, SFI had negative values with an average of -5.1. The negative values imply that soil fertility is generally declining across all fields and farms (Shepherd and Walsh, 2002). The soil degradation index had an average of 0.55 with soil prone
to degradation having higher values while those which are more resistant having lower values. The average values for percentage organic carbon, extractable P and total N are 0.56, 2.79 and -0.8 respectively. The low values obtained for the three measured soil properties indicated that the soils were generally low in fertility (Dalal and Henry, 1986; Morra et al., 1991).

Table 4.9: Measured Soil Properties

<table>
<thead>
<tr>
<th>Statistic</th>
<th>SFI</th>
<th>dseds</th>
<th>Org C</th>
<th>Extr P</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>-4.3</td>
<td>0.68</td>
<td>12.25</td>
<td>1.86</td>
<td>0.97</td>
</tr>
<tr>
<td>Y2</td>
<td>-6.8</td>
<td>0.49</td>
<td>16.90</td>
<td>1.60</td>
<td>1.24</td>
</tr>
<tr>
<td>Y3</td>
<td>-5.5</td>
<td>0.61</td>
<td>11.38</td>
<td>6.36</td>
<td>1.06</td>
</tr>
<tr>
<td>Y4</td>
<td>-8.6</td>
<td>0.75</td>
<td>13.95</td>
<td>4.65</td>
<td>1.27</td>
</tr>
<tr>
<td>Y5</td>
<td>-4.5</td>
<td>0.51</td>
<td>15.60</td>
<td>5.45</td>
<td>1.43</td>
</tr>
<tr>
<td>Y6</td>
<td>-4</td>
<td>0.47</td>
<td>16.43</td>
<td>1.39</td>
<td>1.55</td>
</tr>
<tr>
<td>Y7</td>
<td>-1.2</td>
<td>0.49</td>
<td>12.68</td>
<td>2.68</td>
<td>1.33</td>
</tr>
<tr>
<td>Y8</td>
<td>-6.5</td>
<td>0.46</td>
<td>13.45</td>
<td>1.75</td>
<td>1.28</td>
</tr>
<tr>
<td>Y9</td>
<td>-5</td>
<td>0.50</td>
<td>14.32</td>
<td>5.63</td>
<td>1.42</td>
</tr>
<tr>
<td>Average</td>
<td>-5.1</td>
<td>0.55</td>
<td>14.11</td>
<td>3.49</td>
<td>1.28</td>
</tr>
</tbody>
</table>

Since the main objective of this study was to look at the spatial variability of soil fertility gradients, then, it was prudent to use only one parameter which had been predicted with good accuracy for the analysis. Equally important, soil organic carbon (SOC) has been widely used as an index or indicator of soil quality and sustainable land management in East and Southern Africa (Woomer et al., 1994; Dewayne Mays, 1996; McKenzie et al., 2000). Hence, organic C which gave good calibrations ($r^2 > 0.75$) was chosen for further analysis. According to Shepherd and Walsh (2002) this level of prediction accuracy is sufficiently high for studies in which spatial or temporal variability of an attribute is large relative to the accuracy of its measurement, as typically found in large-area applications and farm advisory work.
4.2 Soil Organic Carbon

The average soil organic C for all fields was 14.11 g/kg with a standard deviation of 1.63 and a coefficient of variation of 11.7 % (Table 4.10). The large coefficients of variation mean that soil organic C varies widely in each of the Ys.

Table 4.10: Soil Organic Carbon per Y

<table>
<thead>
<tr>
<th>Y</th>
<th>Mean</th>
<th>variance</th>
<th>std dev</th>
<th>CV#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>12.25</td>
<td>6.34</td>
<td>2.52</td>
<td>20.56</td>
</tr>
<tr>
<td>Y2</td>
<td>16.90</td>
<td>3.09</td>
<td>1.76</td>
<td>10.40</td>
</tr>
<tr>
<td>Y3</td>
<td>11.38</td>
<td>0.97</td>
<td>0.99</td>
<td>8.67</td>
</tr>
<tr>
<td>Y4</td>
<td>13.95</td>
<td>4.14</td>
<td>2.04</td>
<td>14.59</td>
</tr>
<tr>
<td>Y5</td>
<td>15.60</td>
<td>1.37</td>
<td>1.17</td>
<td>7.52</td>
</tr>
<tr>
<td>Y6</td>
<td>16.43</td>
<td>4.26</td>
<td>2.06</td>
<td>12.57</td>
</tr>
<tr>
<td>Y7</td>
<td>12.68</td>
<td>3.53</td>
<td>1.88</td>
<td>14.82</td>
</tr>
<tr>
<td>Y8</td>
<td>13.45</td>
<td>0.93</td>
<td>0.97</td>
<td>7.19</td>
</tr>
<tr>
<td>Y9</td>
<td>14.32</td>
<td>1.69</td>
<td>1.30</td>
<td>9.07</td>
</tr>
<tr>
<td>Average</td>
<td>14.11</td>
<td>2.93</td>
<td>1.63</td>
<td>11.71</td>
</tr>
</tbody>
</table>

# CV- Coefficient of Variation

The range of variation in soil organic C content within and between farms in each of the Ys is shown by the boxplots in Figures 4.3 and 4.4. Such variation in turn reflects the type and complexity of the farming systems within the study area as determined by both biophysical and socio-economic factors. Similar results were reported by Scoones and Toulmin (1999) and Smaling et al. (1997) although they associated the variation with differential long-term management of fields creating zones of high and low soil fertility. According to Nandwa (2003), such large variations in soil fertility at lower scales of individual niches or fields in farms and village settings are due to a number of factors including differences in soil texture, landuse/fallow history, soil management, and microclimatic differences.
This large variability of soil organic C occurs both within fields of the same farm and between farms, with no clear cut trends or gradients, consequently posing a problem to targeted soil fertility management initiatives. According to Barrett et al. (2002), Sub-Saharan Africa’s extraordinary biophysical variability limits the geographical scope over which any particular natural resource management (NRM) practice proves effective. Even within a particular region, microvariability in hydrology, soils, climate, etc., can render techniques found effective on some farms ineffective on others. The spatial variation can also be feasible in some fields by the differential growth of crops such as the maize field shown in Plate 4.3. The maize in the foreground and in the background to the right near the bananas looks healthier and is bigger in size than the one in the middle ground and to the left.
The adoption of improved NRM practices techniques occurs as a result of decisions made by a wide range of people, each influenced by the incentives and constraints they encounter. It is also important to emphasize that in all areas, traditional farming systems (including their pertinent soil management practices) were evolved and continue to evolve as coping strategies for the environment, and its changing biophysical and socio-economic circumstances (Nandwa, 2003)
These results to some extent support the popular assumption that soil fertility levels in different fields within a farm are influenced by the distance of that field from the homestead. Fields further away from the homestead tend to receive fewer inputs and are managed less intensively than those fields near the homestead, and hence are low in fertility levels (Vanlauwe et al, 2002). Similar farm-level variation has been observed in the ring management systems in semi-arid West Africa, where inner circles near the farms and villages are more intensively used and managed (Nandwa, 2003). The fields near the homesteads receive substantial amounts of nutrients from animal manure and household wastes. As a consequence, soil productivity in this part of the farm remains at a relatively high level, while fields further away suffer from declining soil fertility. Results from long-term soil fertility trials indicated that losses of up to 0.69 t carbon ha\(^{-1}\) yr\(^{-1}\) in the soil surface layers is common in SSA even with
high levels of organic inputs (Nandwa, 2000). One of the reasons why nutrient depletion has been given little recognition in SSA, is that the issue has been perceived differently at various spatial scales. For example, it is difficult to convince farmers and policy makers to react proactively to agro-ecosystems with negative nutrient balances (depleted soils), which have been continuously cultivated till organic matter contents can no longer buffer nutrient depletion.

According to Fresco and Kroonenberg (1992), in ecological and geological time-scales, equilibrium situations in nutrient budgets hardly exist, as climate change, volcanism and biodiversity development all have their more or less gradual impact on agro-ecosystems. When these are coupled with variability in farmers’ management, which in turn is influenced by socio-economic conditions, it is not surprising that soil fertility gradients emerge at different spatial scales. Such complex interactions have led to the high variability of soil organic C observed in within and between smallholder farms of western Kenya.

Although the variation in soil fertility is being perceived as a hindrance to technological interventions and better management systems such as ISFM, Brouwers (1993) had a different view. He argued that smallholder farmers exploit the microvariability within their farms in such a way that during the different seasons as conditioned by rainfall amounts, there are always pieces of land where crops can perform well. Hence, farm and field heterogeneity should be regarded as an asset to the farmer rather than a problem. This is likely to be the case in western Kenya, since the main goal of farmers is to adequately meet subsistence needs rather than have a bumper harvest. Such microscale variation can only be more useful if extrapolated to
macroscale levels by use of appropriate data upscaling techniques such as GIS and spatial modelling.

4.3 Variogram Estimations and Modelling

The spatial variation in soil fertility across the study area was modeled using the experimental semivariograms fitted with theoretical models. Soil organic carbon (Org C) was used for these estimates because it was determined well from spectral reflectance values and is closely correlated with soil fertility levels in the field. According to Haining (1990), data are often correlated in space creating spatial structure. The resulting correlation or covariance structure can be evaluated and used to increase the accuracy of modeling and prediction efforts. Robust experimental variograms for each Y were determined and then fitted with spherical and exponential models as shown in Figures 4.5 (a-i) and 4.6 (a-i) respectively. The advantage of the robust estimator is that the effect of outliers is reduced, without removing specific data points from the data set (Kaluzny et al., 1998; Mohammadi, 2002).

The spherical models in Figures 4.5 (a-i) show the spatial correlation of soil organic C in each of the Y- sampling sites (Y1 - Y9). The models of Y1, Y4 and Y6 show that spatial autocorrelation in these sites continues beyond the maximum distance covered by the study. All the other Ys showed spatial autocorrelation which ended within the maximum distance covered by the Y sites. All samples collected at distances greater than those given by the model variograms’ range were spatially independent. The essence of fitting empirical variograms with theoretical variogram functions was to ensure that the variance of predicted values were positive. In addition, variogram
models should at least have physical meaning (a random function with the given type of variogram that can exist) (Wackernagel, 1998).

(a)

(b)

(c)
(d) Y4

(e) Y5

(f) Y6
Figures 4.5(a-i): Spherical semivariogram models of Soil Organic C per Y.
The two theoretical variogram models, the spherical and the exponential models were used for the data as they best fit its structure. Gaussian models were avoided because they are generally unrealistic and lead to unstable kriging systems and artifacts in the estimated maps (Wackernagel, 1995). The models were fitted iteratively to determine the optimal distance within which spatial correlation was evident with optimum variogram parameter values i.e. nugget effect, sill and range. These variogram parameters are essential in determining kriging weights (Genton and Furrer, 1998) as it will be shown in section 4.4.

The exponentially fitted model variograms are shown in Figures 4.6 (a-i) and they give the spatial correlation of soil organic C in each of the Ys. The models of Y1, Y4 and Y6 appear as straight lines which increased continuously without reaching a sill meaning that in these regions spatial autocorrelation continued beyond the maximum distance covered by the study. The models of Y2, Y5, Y7, Y8 and Y9 show spatial autocorrelation which reaches the sill at some distance within the maximum distance covered by the study. An exception to all the other Ys is Y3 with a straight horizontal line, a condition called pure nugget effect. This means that there is no spatial correlation between the sampling points and point values can be treated as being independent.
(a) Y1

(b) Y2

(c) Y3

Objectives:
- Y1: objective = 0.0098
- Y2: objective = 0.0114
- Y3: objective = 0.0015
(d)

Y4

![Graph](image)

Objective: 0.0061

(distance)

(e)

Y5

![Graph](image)

Objective: 0.0026

(distance)

(f)

Y6

![Graph](image)

Objective: 0.0035

(distance)
Figures 4.6(a-i): Exponential semivariogram models of Soil Organic C per Y.
All the model variograms in Figures 4.5 (a-i) and 4.6 (a-i) exhibited large positive nugget values attributable to such variability as short scale variability (between sampling points), random and inherent variability, and sampling error. The causes of extreme local variability in crop growth, across distances of even few meters, are still poorly understood (Voortman et al., 2002). It has been attributed to differences in soil chemistry (e.g. Scott-Wendt et al., 1988a, 1988b; Kretschmar et al., 1991; Wendt et al., 1993; Stein et al., 1997), but also correlates with differences in local topography (Brouwer and Powell, 1998). In Western Kenya, however, other factors such as farm size, type of management and socio-economic status of the farmer play a great role in determining soil fertility levels as well as their variability within the farm.

The differences in model parameter estimates between the spherical and exponential models are shown in Table 4.11. In general, spherical models fitted better than exponential models as evidenced by the smaller mean squared residuals (MSR). However, both models were generally similar for all the data subsets. The nugget-to-sill ratio is used as a criterion to classify the spatial dependency of soil properties. According to Sun and Zhao (2002), a variable is considered to be having strong spatial dependence if the ratio is less than 25%, and has a moderate spatial dependence if the ratio is between 25 and 75%, otherwise, the variable has a weak spatial dependence. Both the spherical and exponential models show strong spatial dependence in Y1 and Y4 (between 0 and 0.001), as shown in Table 4.11. In Y2, Y7 and Y8 both models show moderate spatial dependence (between 0.24 and 0.65), while in Y3 (0.92) they show weak spatial dependency. For the remaining Ys i.e. Y5, Y6 and Y9, the two models classify spatial variability differently. The spherical model classifies the spatial dependence in Y5 and Y9 as weak (nugget/sill ratio of
0.85 in both cases), while the exponential model classifies them as moderate (nugget/sill ratio of 0.60 and 0.64 respectively).

Table 4.11: Parameter values of fitted semivariograms models

<table>
<thead>
<tr>
<th>Y-</th>
<th>Model</th>
<th>co</th>
<th>c</th>
<th>a(m)</th>
<th>co/c</th>
<th>MSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>Spherical</td>
<td>6.20E-02</td>
<td>4.83E+03</td>
<td>1.33E+08</td>
<td>1.30E-05</td>
<td>0.0098</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>6.20E-02</td>
<td>1.11E+02</td>
<td>1.98E+06</td>
<td>0</td>
<td>0.0098</td>
</tr>
<tr>
<td>Y2</td>
<td>Spherical</td>
<td>1.90E-02</td>
<td>6.50E-02</td>
<td>2.82E+02</td>
<td>0.3</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>1.60E-02</td>
<td>7.00E-02</td>
<td>1.18E+02</td>
<td>0.24</td>
<td>0.0114</td>
</tr>
<tr>
<td>Y3</td>
<td>Spherical</td>
<td>4.50E-02</td>
<td>4.90E-02</td>
<td>2.15E+02</td>
<td>0.92</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>4.90E-02</td>
<td>0.00E+00</td>
<td>0.00E+00</td>
<td>4.52E+07</td>
<td>0.0015</td>
</tr>
<tr>
<td>Y4</td>
<td>Spherical</td>
<td>5.30E-02</td>
<td>2.10E+02</td>
<td>1.24E+07</td>
<td>0</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>5.30E-02</td>
<td>4.80E+01</td>
<td>1.90E+06</td>
<td>0.001</td>
<td>0.0061</td>
</tr>
<tr>
<td>Y5</td>
<td>Spherical</td>
<td>2.20E-02</td>
<td>2.60E-02</td>
<td>6.84E+02</td>
<td>0.85</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>2.20E-02</td>
<td>3.70E-02</td>
<td>5.00E+02</td>
<td>0.6</td>
<td>0.0026</td>
</tr>
<tr>
<td>Y6</td>
<td>Spherical</td>
<td>3.10E-02</td>
<td>6.14E+03</td>
<td>2.78E+08</td>
<td>5.14</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>3.10E-02</td>
<td>1.53E+02</td>
<td>4.46E+06</td>
<td>0</td>
<td>0.0035</td>
</tr>
<tr>
<td>Y7</td>
<td>Spherical</td>
<td>4.50E-02</td>
<td>7.70E-02</td>
<td>5.84E+02</td>
<td>0.58</td>
<td>0.0124</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>3.20E-02</td>
<td>9.30E-02</td>
<td>2.21E+02</td>
<td>0.35</td>
<td>0.0136</td>
</tr>
<tr>
<td>Y8</td>
<td>Spherical</td>
<td>1.80E-02</td>
<td>2.80E-02</td>
<td>3.51E+02</td>
<td>0.65</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>1.40E-02</td>
<td>3.40E-02</td>
<td>1.47E+02</td>
<td>0.42</td>
<td>0.0012</td>
</tr>
<tr>
<td>Y9</td>
<td>Spherical</td>
<td>4.20E-02</td>
<td>5.00E-02</td>
<td>6.56E+02</td>
<td>0.85</td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>3.90E-02</td>
<td>6.20E-02</td>
<td>3.50E+02</td>
<td>0.64</td>
<td>0.0075</td>
</tr>
</tbody>
</table>

# co = nugget effect, c = sill, a = range and MSR = mean squared residual.

Analysis of the fitted variogram models for each of the data subsets indicated that the spatial autocorrelation structure of soil organic carbon ranged from small distances to beyond the maximum distance covered by the study i.e. beyond 900 meters. In the Y1, Y4 and Y6, spatial autocorrelation structure goes beyond the maximum distance covered by the sampling design. In these three subset areas, the range of spatial dependence is expected to hold up to 1.334e+08 m in Y1, 1.24e+07 m in Y4 and 2.78e+08 m in Y6. In reality these may not be true and the large nugget effects experienced by the models indicate unexplained variance between the sampled points. Beyond these distances, the sample values are then expected to be independent and not influenced by spatial structure. According to a study carried out by Voortman et al., (2002), only a small portion of the variation in crop yields was explained by soil
macronutrients N, P and K, and manure application rates. A large portion of the yield differences was explained by spatial dependence or autocorrelation i.e. by the yield values of neighboring observations.

If the slope of the dissimilarity function changes abruptly at a specific scale, this suggests that an intermediate level of the variation has been reached at this scale (Wackernagel, 1998). Such variation occurs in Y2 and Y8, where the spherical variograms reach an abrupt sill. The behavior at the very small scales, near the origin of the variogram, is of importance, as it indicates the type of continuity of the regionalized variable. All the variograms in Figures 4.5 and 4.6 with both nugget and sill variance are said to be differentiable, while those which increase without a sill are said to be continuous but not differentiable, and only one, Y3 (Figure 4.6c) is said to be discontinuous i.e. has a zero slope indicating no spatial structure, a condition termed as pure nugget effect. This phenomenon can be explained by the biophysical conditions of the area dominated by rugged granite rocks which rises above 1950 m. These granites have heterogeneous lithologies which vary from granodiorites, through adamellites to tonalities (Shiozaki, 1983; Opiyo-Akech, 1988; Mathu and Nyambok, 1993). The rocks cover large proportions in each field, hence limiting farming activities and potential soil erosion. Thus, soil organic C levels on these farms remain more or less the same indicating no spatial structure. Further analysis of the spatial structure of predicted soil organic C is given in section 4.6 where the spatial component is incorporated into the mixed effects model.
4.4 Spatial Interpolation - Kriging

The main application of geostatistics in soil science has been the estimation and mapping of soil attributes at unsampled points. Kriging as a method of spatial prediction uses generalised least-squares regression algorithms to make predictions. It addresses the problem of estimation based on a continuous model of stochastic spatial variation, making the best use of existing knowledge by taking account of the way that a property varies in space, using the variogram model. The kriging estimates are weighted linear combinations of the data. The weights are allocated to the sample data within the neighborhood of the points to be estimated in such a way as to minimize the estimation or kriging variance, and ensure that the estimates are unbiased (Webster and Oliver, 2000).

In this study, ordinary kriging was used for predicting the nearby unsampled locations. A random function model (Equation 11) of spatial correlation was used to calculate a weighted linear combination of available samples in each Y sampling area. For kriging to be possible, the three parameters of a variogram model i.e. the nugget effect, range and sill must be precisely defined and entered in the kriging equation. Kriging was not possible in three Ys (Y1, Y4 and Y6) because spatial correlation was low as indicated by the nugget/sill ratio in Table 4.11 in above. Also the kriging weights were chosen to ensure that the average error for the model is zero and that the modelled error variance is minimized (Isaaks and Srivastava, 1989). The kriging predictions for the other Ys are shown in Figures 4.7 (a-f) as contour lines with the respective values of soil organic C as indicated.
(a) Y2 Kriging Predictions

(b) Y3 Kriging Predictions

(c) Y5 Kriging Predictions
Figures 4.7 (a–f): Kriging predictions of soil organic C
Better spatial predictions are possible in areas where strong spatial correlation of soil organic C exists such as in Y5, Y7 and Y9. Strong spatial correlation exists where the percentage nugget/sill ratio is high. It should also be understood that the kriging variance is primarily a measure of the density of information around the estimation point. Therefore, a higher density of sampled points gives better kriging variances leading to better kriging estimates.

4.5 Determination of Soil Organic C

Multible linear regression was used to account for (predict) the variance in the dependent variable (soil organic C), based on linear combinations of independent variables. The factors used as independent variables are mainly the biophysical factors measured in the field whose results are presented in section 4.1 under field characterisation. These include factors such as slope of the field, number of soil conservation structures, soil hard-setting, percentage rock cover, rock grade, place of field in the landscape, type of land use, number of seasons left fallow if field is not cultivated and fertility level assessment by the farmer. Other factors used included the georeferenced position of the field, the farm and field codes, distance between fields, and field areas.

Using these variables, a multiple linear regression model was run for each of the nine Y-frame sampling regions to establish a set or sets of factors (independent variables) that explain a proportion of the variance in the dependent variable by significant levels. The multiple regression equation takes the form;
where $b_1$, $b_2$ and $b_n$ are the regression coefficients, representing the amount the dependent variable $y$ changes when the independent changes by 1 unit; and $c$ is the constant, where the regression line intercepts the $y$ axis, representing the value of the dependent variable $y$; when all the independent variables are set to 0. The standardized versions of the $b$ coefficients are the beta weights, and the ratio of the beta coefficients is the ratio of the relative predictive power of the independent variables. Results of the fitted models are presented in Table 4.12 showing the $R^2$, residual standard error (RSE), $F$-statistic, p-values and degrees of freedom (DF).

Table 4.12: Multiple linear regression parameters for soil organic C

<table>
<thead>
<tr>
<th>Y Region</th>
<th>$R^2$</th>
<th>RSE</th>
<th>DF</th>
<th>p-value</th>
<th>$F$</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>0.73</td>
<td>0.203</td>
<td>41</td>
<td>0</td>
<td>3.29</td>
<td>33/41</td>
</tr>
<tr>
<td>Y2</td>
<td>0.76</td>
<td>0.198</td>
<td>37</td>
<td>0.001</td>
<td>2.966</td>
<td>40/37</td>
</tr>
<tr>
<td>Y3</td>
<td>0.53</td>
<td>0.249</td>
<td>42</td>
<td>0.369</td>
<td>1.11</td>
<td>42/42</td>
</tr>
<tr>
<td>Y4</td>
<td>0.58</td>
<td>0.252</td>
<td>36</td>
<td>0.17</td>
<td>1.379</td>
<td>36/36</td>
</tr>
<tr>
<td>Y5</td>
<td>0.62</td>
<td>0.178</td>
<td>40</td>
<td>0.019</td>
<td>1.999</td>
<td>32/40</td>
</tr>
<tr>
<td>Y6</td>
<td>0.61</td>
<td>0.207</td>
<td>50</td>
<td>0.002</td>
<td>2.563</td>
<td>30/50</td>
</tr>
<tr>
<td>Y7</td>
<td>0.53</td>
<td>0.268</td>
<td>62</td>
<td>0.005</td>
<td>2.115</td>
<td>33/62</td>
</tr>
<tr>
<td>Y8</td>
<td>0.66</td>
<td>0.149</td>
<td>62</td>
<td>0</td>
<td>3.721</td>
<td>33/62</td>
</tr>
<tr>
<td>Y9</td>
<td>0.57</td>
<td>0.209</td>
<td>44</td>
<td>0.019</td>
<td>1.977</td>
<td>30/44</td>
</tr>
<tr>
<td>Average</td>
<td>0.49</td>
<td>0.229</td>
<td>630</td>
<td>&lt;0.001</td>
<td>5.279</td>
<td>113/630</td>
</tr>
</tbody>
</table>

# RSE- residual standard error; DF- degrees of freedom

Using multiple linear regression it was possible to explain between 49 - 76 % of the total variation in soil organic carbon ($p<0.05$) in the top soil for all the locations that were sampled and presented as Y-frames. However, in two of the sub-locations, Y3 and Y4, the linear model was not significant at $p< 0.05$ and this can be explained by the biophysical setting of these areas. The sites are characterized by granite batholiths of western Kenya (commonly known as Maragoli Hills) which have heterogeneous lithologies varying from granodiorites, through adamellites to tonalities (Shiozaki,
1983; Opiyo-Akech, 1988; Mathu and Nyambok, 1993). The two sites to some extent distort the general models for Vihiga district and that of the two districts combined.

The F-test was used to test the significance of $R^2$, which is the same as testing the significance of the regression model as a whole. If prob(F) was < 0.05, then the model was considered significantly better than would be expected by chance and the null hypothesis of no linear relationship between $y$ (predicted soil organic C in this case) and the independent variables we rejected. The geographical position of farms and fields in terms of northing and easting seemed to play a major role in explaining the variability in predicted soil organic C together with the farm and field coding system. Other factors that also played a significant role included slope, position or place of field on the landscape (upland, midland or bottomland), distance of fields from the homestead, type of land use, fertility, soil hard setting and number of years since first cultivation took place. The highest percentage of the variation explained was 73 and 76% in Y1 and Y2 respectively (Table 4.12). Plots of the fitted models and their residuals are shown in Figures 4.8 (a-r).
Figures 4.8 (a–r): Plots of multiple regression models of soil organic C

The main factors that explained the variance of soil organic C across the entire study included fertility level description, position of the field and farm on the landscape
(place), slope of the field, and the geographical location of the farm in terms of northing and easting. Others include distance of the field from the homestead, soil structure, soil surface hard setting, type of land use, number of years since field was first cultivated, and size of field (area). Residual plots for all the models are randomly distributed meaning that the entire model fits where adequate. On a larger scale, the regression models were fitted for the two districts separately (Figures 4.9 (a-d)).

![Vihiga Plot of Org C vs Fit](image)

![Plot of Residuals vs Fit for Vihiga](image)

![Siaya Plot of Org C vs Fit](image)

![Plot of Residuals vs Fit for Siaya](image)

**Figures 4.9 (a–d):** Plots of multiple regression models of soil organic C

The ability to adequately predict soil organic C at the district level is slightly higher in Siaya than Vihiga by at least 4% (p<0.001). This variation between the two districts can be attributed, on one hand, to the different climatic conditions, size of farmland per household and population density. In addition, Siaya district has more or less homogenous field conditions in terms of biophysical factors as compared to Vihiga district.
4.6 Mixed-Effects Modelling

The normal multiple linear regression analysis done in section 4.5 above, though valid, did not take into account the groupings within the sample data, its unbalanced nature within the groups, and the existing spatial correlations. Mixed-effects models provide a flexible and powerful tool for the analysis of such grouped data, which may arise in many diverse areas (Pinheiro and Bates, 2000), and were used for further analysis of the data. Linear mixed-effects models were fitted to the data as an iteration process to determine the main factors that contributed significantly to an optimal model. All the parameters that were measured in the field were treated as fixed-effects. The grouping factors, that is, the districts, Ys within a district, and farms within a Y, were all treated as random factors while the fields within a farm were treated as the residual variance. The Eastings and Northings represented latitudes and longitudes respectively and were used to account for spatial variability of the data.

After several iterations, the parameter combinations shown in Table 4.13 were found to give better model predictions for the data. The p-values were used to eliminate the most insignificant fixed effect factors, leaving the significant ones.

Table 4.13: Model Estimates and Random Effects for the Basic Mixed Effects Model

<table>
<thead>
<tr>
<th>Random term</th>
<th>Std Deviation</th>
<th>Variance Comp.</th>
<th>% Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>District</td>
<td>0.090</td>
<td>0.008</td>
<td>2.6</td>
</tr>
<tr>
<td>Y</td>
<td>0.225</td>
<td>0.051</td>
<td>16.5</td>
</tr>
<tr>
<td>Farm</td>
<td>0.238</td>
<td>0.056</td>
<td>18.4</td>
</tr>
<tr>
<td>Field/residual</td>
<td>0.438</td>
<td>0.192</td>
<td>62.5</td>
</tr>
</tbody>
</table>

The model output included the values of the *Akaike Information Criterion* (AIC) (Sakamoto, Ishiguro and Kitagawa, 1986) and the *Bayesian Information Criterion*
(BIC) (Schwarz, 1978), which is also sometimes called *Schwarz's Bayesian Criterion* (SBC). These were the model comparison criteria evaluated as;

\[
\text{AIC} = -2\log \text{Lik} + 2n_{\text{par}},
\]

\[
\text{BIC} = -2\log \text{Lik} + n_{\text{par}} \log(N),
\]

where \(n_{\text{par}}\) denotes the number of parameters in the model and \(N\) the total number of observations used to fit the model. Under these definitions, “smaller is better”. That is, if we are using AIC to compare two or more models for the same data, the model with lowest AIC is preferred. Similarly, when using BIC the model with the lowest BIC is preferred. The model shown in Table 4.13 gave AIC and BIC values as 1111.6 and 1203.4 respectively, the lowest values obtained meaning it was the best model fit. A summary of the fixed effects from the same model is given in Table 4.14, showing only those factors that showed more or less significant effects on model determination.

**Table 4.14: Summary of Fixed Effects of the Basic Mixed Effects Model.**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Value</th>
<th>Std.Error</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.814</td>
<td>0.149</td>
<td>640</td>
<td>12.165</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Landuse1</td>
<td>0.084</td>
<td>0.048</td>
<td>640</td>
<td>1.737</td>
<td>0.0829</td>
</tr>
<tr>
<td>Landuse2</td>
<td>0.048</td>
<td>0.021</td>
<td>640</td>
<td>2.286</td>
<td>0.0226</td>
</tr>
<tr>
<td>Fertility1</td>
<td>-0.008</td>
<td>0.028</td>
<td>640</td>
<td>-0.297</td>
<td>0.7662</td>
</tr>
<tr>
<td>Fertility2</td>
<td>-0.021</td>
<td>0.014</td>
<td>640</td>
<td>-1.454</td>
<td>0.1466</td>
</tr>
<tr>
<td>Place1</td>
<td>-0.04</td>
<td>0.034</td>
<td>640</td>
<td>-1.203</td>
<td>0.2293</td>
</tr>
<tr>
<td>Place2</td>
<td>-0.013</td>
<td>0.02</td>
<td>640</td>
<td>-0.646</td>
<td>0.5187</td>
</tr>
<tr>
<td>Place3</td>
<td>0.009</td>
<td>0.027</td>
<td>640</td>
<td>0.332</td>
<td>0.7398</td>
</tr>
<tr>
<td>Rock1</td>
<td>0.08</td>
<td>0.062</td>
<td>640</td>
<td>1.279</td>
<td>0.2015</td>
</tr>
<tr>
<td>Rock2</td>
<td>0.011</td>
<td>0.057</td>
<td>640</td>
<td>0.201</td>
<td>0.8409</td>
</tr>
<tr>
<td>Rock3</td>
<td>-0.111</td>
<td>0.086</td>
<td>640</td>
<td>-1.28</td>
<td>0.2009</td>
</tr>
<tr>
<td>Rock4</td>
<td>-0.043</td>
<td>0.057</td>
<td>640</td>
<td>-0.758</td>
<td>0.4484</td>
</tr>
<tr>
<td>Rock5</td>
<td>0.039</td>
<td>0.042</td>
<td>640</td>
<td>0.936</td>
<td>0.3498</td>
</tr>
<tr>
<td>Rockgrade1</td>
<td>0.112</td>
<td>0.114</td>
<td>640</td>
<td>0.985</td>
<td>0.325</td>
</tr>
<tr>
<td>Rockgrade2</td>
<td>-0.09</td>
<td>0.049</td>
<td>640</td>
<td>-1.841</td>
<td>0.0661</td>
</tr>
<tr>
<td>Rockgrade3</td>
<td>0.006</td>
<td>0.024</td>
<td>640</td>
<td>0.245</td>
<td>0.8065</td>
</tr>
</tbody>
</table>
Before making further inferences about the fitted mixed-effects model, it was necessary to check whether the underlying distributional assumptions appear valid for the data. The within-group residuals were used as primary quantities estimated from Best Linear Unbiased Predictions (BLUPs) of the within-group errors, and the results plotted separately for the two main grouping levels as shown in Figures 4.10 (a & b).

![Boxplots of within-group residuals; (a) district level and (b) Y level](image)

**Figure 4.10:** Boxplots of within-group residuals; (a) district level and (b) Y level

These plots were useful in verifying that the errors are centred at zero (i.e., $E[\varepsilon] = 0$), have constant variance across groups (districts, Ys and farms) ($\text{Var}(\varepsilon_i) = \sigma^2$), and are independent of group levels. The plots in Figure (4.10) above indicated that the residuals are centred at zero, thus fulfilling one of the basic assumptions of mixed-effects models' inference, that the within-group errors should be independent and identically normally distributed with mean zero and variance $\sigma^2$.

Analysis of the estimated variance components (Table 4.13) showed that the district effect accounts for little (2.6%) of the variation in predicted organic C as compared to the Y or farm effects. The Y and farm effects accounted for 16.5 and 18.4%, respectively.
respectively, of the total variance observed in predicted soil organic C. Farms within a
Y region exhibited between-farm variability compared to that between Ys but with
better estimates. This can be explained by the fact that farms within the same Y region
show spatial autocorrelation, hence the value of predicted soil organic C in adjacent
farms can be favourably predicted with minimal standard errors.

The within-farm or between-fields (residual) variability, accounts for the greatest
percentage (62.5%) of the variation associated with random effects. Thus, the
devolution of an individual value of predicted soil organic C in a field is the deviation
of that field from the average of the farm, Y and district where it is located. It
encompasses all of the unexplained variation from field to field within a farm, such as
local environmental effects (soil type, other biophysical factors not accounted for),
management interventions, and measurement error. According to Voortman et al
(2002), the causes of extreme local variability in crop growth across distances of even
a few meters are still poorly understood. The variance of 0.192 means that the
standard deviation of the field-to-field variation is \( \sqrt{0.192} = 0.438 \). Thus, there is
more variability of predicted soil organic C within individual farms (approximately up
to 63%) than the higher grouping levels.

In a study carried out by Lin et al., (2005), assessment of soil spatial variability at
multiple scales showed that majority of the variability (over 50% in most cases) for all
the three soil properties measured was at the local point scale. These results suggest
that careful examination of short-range soil property variability should not be
overlooked. In Lin’s case, they associated the variability to a number of factors
ranging from climate at the basin scale to localised differential infiltration and runoff.
caused by the differences in landscape positions and soil characteristics. In western Kenya, however, the very high local variability cannot be associated with any specific factor such as slope or landscape position as most of the measured parameters failed to significantly explain the observed variations. But, the variation may be attributed to management factors which are known to vary from one farmer to the next as influenced by the prevailing socio-economic conditions. According to a study carried out in western Kenya by Crowley and Carter (2000), spatial and social heterogeneity in soil management implied that farmers did not necessarily intensify utilization of their land uniformly but, rather, gave preference to some areas over others and some crops over others. Similarly, differences in socio-economic circumstances may produce consistently divergent soil management strategies between households of the same community and corresponding differences in the soil fertility status of their farms.

One of the main reasons why the variation of predicted soil organic C is large within the farm, is the fact that farmers manage several organic and mineral resources differently in order to attain their production goals in different fields. According to Vanlauwe et al (2001), the net flow of resources is not equal for the various fields belonging to a single farm household but varies substantially, creating areas with carbon and nutrient accumulation and depletion. Since, soil fertility is closely related to soil organic C, then, variations of predicted soil organic C in this study imply variations in soil fertility or soil fertility gradients. Therefore, the existence of such gradients within smallholder farms must be considered when designing integrated soil fertility management strategies. These results are similar to the work of other researchers (Scoones and Toulmin, 1999; Smaling et al., 1997), who also showed the
existence of positive and negative carbon and nutrient balances for different fields within a farm. Richards (1983) noted that even where extensive cultivation is the predominant practice, as the case in western Kenya, farmers often manage some parts (fields) of their farms more intensively than others.

Apart from land use, most of the other factors treated as fixed effects (Table 4.14), had minimal influence on the general model as shown by the p-values which are mostly large, $p > 0.05$. The type of land use, either subsistence or commercial, was the single most important field specific-factor that had an influence on predicted soil organic C. Land use type two (subsistence use), was the most significant with $p < 0.02$, while type one (commercial use) had a significance of $p < 0.08$. Over 77% of all fields sampled were for subsistence use, an important indicator of the socioeconomic status of the farmers’ and their ability to manage soil fertility levels within their farms. However, when the spatial structure was taken into consideration, landuse ceased to have any significant influence on the predicted model, reducing the p values from 0.02 to 0.05 and 0.08 to 0.21 respectively. According to a recent study carried-out in western Kenya by Tittonell (2003), soil fertility gradients within farms were found to originate from a host of processes ranging from the inherent productivity of soils to the effect of the differential management practices that the farmers consequently apply to them.

Other fixed effects that had some influence on the model included qualitative fertility classification by the farmers, place on the landscape, rock type and rock grade. The farmers were able to classify their fields in terms of fertility as being high, medium or low. Fields classified as of medium soil fertility had a better correlation with predicted
soil organic C (p<0.15) as compared to the other classes. Actually, farmers are often aware of the existence of soil fertility gradients within their farms and use local terms to ascribe different soil quality features to different fields (TSBF, 2001). Place of the field on the landscape either upslope, mid-slope, or bottomland also had some minimal influence on the model. The effect of rock type and rock grade was less significant apart from rock grade 2 which was almost significant (p<0.07). Other measured biophysical factors such as distance from the homestead, slope, area and number of conservation structures within the field that were hypothesized to influence predicted soil organic C and within farm soil fertility, however, had no significant effect.

Although this study concentrated mainly on the effects of biophysical factors on the spatial variations of soil fertility, it was evident that short-term and micro-scale processes including socioeconomic endowment of the farmers have, also had a substantial impact. The net effect of these factors over the long run is a gradual build-up of nutrient rich micro-niches at the expense of a gradual decline in fertility over a much wider area, as already observed in western Kenya. According to Crowley and Carter (2000), such processes are easily missed in studies that aggregate data to the farm and higher system levels or assume an equal distribution of nutrients across the landscape. Therefore, it is important to note that micro-variations in soil fertility and other soil properties are essential in farmers’ choices of crops per locale and the variable impact of technologies in space. If this is the case, then, modeling of the spatial variations of soil fertility attributes in combination with the socioeconomic conditions of the farmers is an appropriate approach towards better targeting of ISFM technologies.
4.6.1 The Spatial Model

Variogram analysis and modelling in Section 4.3 showed that predicted soil organic C displays spatial correlation. It is therefore necessary to account for such spatial correlation structures in the context of mixed-effects models. The variogram method for the linear mixed-effects (lme) class estimates was used to fit an experimental semivariogram of the residuals as shown in Figure 4.11, using averaged lag distances at arbitrary constant increments. The different combinations of lag distances used for plotting the experimental semivariogram in Figure 4.11 are shown in appendix B.2.

The semivariogram represents a nested random function $Z(x)$, and according to Wackernagel (1998) a nested variogram model is a sum of spatial components characterising different spatial scales, i.e. reaching different sills of variation ($b_u$) at different scales, except maybe for the last coefficient ($b_s$), which could represent the slope of an unbounded variogram model. The semivariogram is therefore a product of several other individual semivariograms optimised at different lag distances. The large nugget variance showed that the sampling interval was coarse and a small-scale variance component can only be identified if the sampling grid is sufficiently fine.

![Figure 4.11: Fitted model semivariogram at maximum distance.](image)
Since the data was collected using an irregular sampling scheme, an imbalance in the distance measurements was created. As a result of such imbalanced distance measurements, the number of residual pairs used at each lag distance varied considerably, making some semivariogram estimates more reliable than others. In general, the number of residual pairs used in the semivariogram estimation decreases with distance for irregularly sampled data, making the values at larger distances unreliable (Pinheiro and Bates, 2000). To avoid unreliable semivariogram estimates, the maximum distance was controlled using the \textit{maxDist} (maximum distance) argument in the S-plus analytical software. A reliable semivariogram was produced at a maximum distance of 80 m as shown in Figure 4.12. A nugget effect of about 0.6 was obtained with a semivariogram which appears to approach 1 around a distance of 76-80 m.

\textbf{Figure 4.12}: Fitted model semivariogram at 80 m.

The maximum distance was determined by an iterative process until an optimal distance was reached where the experimental semivariogram fairly smoothed out reaching a sill (Figure 4.12). To describe the spatial component of the experimental model, three isotropic variogram models (the exponential, spherical and rational
quadratic) were fitted using REML in ordinary least squares. The model parameters obtained are shown in Table 4.15, and were compared based on the information criteria AIC and BIC.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>1</td>
<td>22</td>
<td>1109.6</td>
<td>1210.7</td>
</tr>
<tr>
<td>Spherical</td>
<td>1</td>
<td>22</td>
<td>1109.9</td>
<td>1210.9</td>
</tr>
<tr>
<td>Ratio</td>
<td>2</td>
<td>22</td>
<td>1108.0</td>
<td>1208.0</td>
</tr>
</tbody>
</table>

The rational quadratic model fit has the smallest AIC and BIC and hence gave the best fit among the spatial correlation models considered. However, plots of the fitted models (Figure 4.12) showed that the exponential model better fitted the data as compared to the spherical and ratio models. All the three model variogram estimates had high nugget variances which imply that the micro-scale variation (the variation below the minimum sample distance) was large and can only be accounted for by sampling at finer distances. At the farm level, micro-scale variation is a factor of different management practices which differ from field to field, variation in biophysical aspects such as soil type and fertility, and the sampling error. At the Y-level, micro-scale variation can only exist if sampled farms are adjacent to each other such as those around the centre of the Y. In such cases, micro-scale variation can also be a factor of different management aspects between farmers which are in turn influenced by farmers’ socio-economic status.

According to Webster and Oliver (2000), the nature of an experimental variogram depends on the spatial scale over which it was measured. If a large extent is covered with wide sampling intervals then all of the variance might appear as nugget. Alternatively, if small intervals are chosen to resolve the short-range variance then the
sampling required to estimate the contribution to large distances may be too costly. The Y-frame sampling adopted for this study combined both short- and long-range sampling intervals ranging from the farm, Y to district levels.

However, both sampling intervals lacked the density necessary to account for most of the variance observed at those levels. On one hand, the unaccounted short-range variance is indicated by the large nugget effects (Figures 4.11, 4.12 and 4.13) occurring even when the maximum distance was limited. On the other hand, the long-range variance is shown by the wavy pattern of the experimental semivariogram (Figure 4.11). If the slope of the dissimilarity function changes abruptly at a specific scale (Figure 4.12b and c), this suggests that an intermediate level of the variation has been reached at this scale (Wackernagel, 1998). The spherical and ratio quadratic models reach an abrupt sill. In general, the sampling design was more of a nested sampling design with unbalanced classes but due to a limited sampling density and irregular sampling, the spatial structure failed to elicit all the spatial scale(s) of variation in predicted soil organic C.
Figure 4.13: Fitted Semivariogram Models (a) Exponential model (b) Spherical model and (c) Ratio model

Though the models fitted in Table 4.15 allow considerable flexibility in the specification of the random-effects structure, they restrict the within-group errors to independent, identically-distributed, random variables with mean zero and constant variance. However, the data showed that the within-group errors had unequal
variances (i.e. was heteroscedastic) and correlated. Spatial correlation structures were used to model the dependence observed in the data. The models were fitted using the generalized least squares (gls) modeling function of REML as provided by the non-linear mixed effects (nlme) library.

Table 4.16: Parameter values for fitted model semivariograms

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>1089.8</td>
<td>1177.1</td>
<td>-525.9</td>
</tr>
<tr>
<td>Spherical</td>
<td>1090.4</td>
<td>1177.6</td>
<td>-526.2</td>
</tr>
<tr>
<td>Ratio</td>
<td>1094.7</td>
<td>1181.9</td>
<td>-528.3</td>
</tr>
</tbody>
</table>

The exponential model provided the best estimates for modeling the within-group error covariance structure since it had the lowest AIC and BIC values. A comparison between the basic mixed-effects model which does not account for spatial variability, and the final mixed-effects model which does account for spatial correlation structures, indicated a strongly significant improvement of the model as shown by the anova analysis (Table 4.17). Thus, accounting for spatial variability greatly improves the basic model significantly (p<0.0001) to arrive at a more stable final model. However, when looking at the actual change in absolute values of the AIC and BIC functions, the change was minimal (Tables 4.15 and 4.16), an indication that the mixed effects model had accounted for most of the observed spatial variation. The model factored in the Y-frame sampling design used in this study as random effects, hence controlling most of the variability which would have otherwise occurred due to spatial separation of sampled areas.

Table 4.17: ANOVA Analysis between the fitted models

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic model</td>
<td>1</td>
<td>1111.6</td>
<td>1203.4</td>
<td>-535.8</td>
</tr>
<tr>
<td>Final model</td>
<td>2</td>
<td>1089.8</td>
<td>1177.1</td>
<td>-525.9</td>
</tr>
</tbody>
</table>

Test | L.Ratio | p-value |
-----|---------|---------|
1 vs 2 | 19.79   | <.0001  |
The large value of the likelihood ratio (L. Ratio) test statistic gives strong evidence that spatial correlation exists and individual sample values cannot be said to be independent. This means that the influence of the fixed effects on the model with spatial correlation structures accounted for, is significantly different from one in which they have not been accounted. Further evidence on existence of spatial correlation structures is given by the nugget variance and range estimates (Table 4.18) of the variogram models. The nugget/sill ratio for all the three semivariogram models fitted lied between 50 to 60 %, indicating moderate spatial correlation.

Table 4.18: Model Semivariogram Estimates

<table>
<thead>
<tr>
<th>Model</th>
<th>Range</th>
<th>Nugget</th>
<th>Sill</th>
<th>Nugget/Sill (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>1089.8</td>
<td>0.534</td>
<td>1</td>
<td>53.4</td>
</tr>
<tr>
<td>Spherical</td>
<td>2708.5</td>
<td>0.512</td>
<td>1</td>
<td>51.2</td>
</tr>
<tr>
<td>Ratio</td>
<td>697.1</td>
<td>0.590</td>
<td>1</td>
<td>59.0</td>
</tr>
</tbody>
</table>

In section 4.3, majority of the Ys had moderate spatial variability agreeing with the findings given in this section using spatial correlation structures in mixed-effects modeling. After accounting for the spatial correlation structures in the data (Table 4.18) the exponential model of normalized residuals was fitted again to the experimental semivariogram (Figure 4.14) and the model appeared to be bounded, reaching a sill of close to one at a range of around 60 m. Beyond the range of 60 m, the variogram tended to decrease again, an indication that there was some intermediate variation at those lag distances which could not be accounted for.
Therefore, it can be deduced from Figure 4.14 that the fitted exponential semivariogram model can only be used to make spatial predictions for soil organic C reliably up to a maximum range of 60 m beyond which prediction errors are expected to increase significantly. The variogram still show a large nugget effect indicating a wide range of unaccounted short-range variation. Such behavior at the very small scales, near the origin of the variogram, is of importance, as it indicates the type of continuity of the regionalized variable. For predicted soil organic C, this indicates that its values change significantly over short distances. If the nugget effect from the above semivariogram models is linked to the variance components of the mixed-effects model, then at small lag distances near the origin, the nugget effect is high. Also the variance component associated with the lowest random-effects level (the field or residual) accounts for most of the variability observed.

The magnitude of such spatial variability, strongly influenced by site-specific biophysical factors and socio-economic conditions, may be sufficiently large to affect within-farm soil fertility levels. Therefore, the short-range spatial correlation
structures need to be recognized when designing soil fertility management strategies. Micro-scale spatial variations occur mainly at small distance intervals which can only be attributed to the lowest variance component, the field level. Since fields occur within farms, then, the most appropriate intervention point at which the problem of soil fertility gradients should be addressed is the farm level.
CHAPTER FIVE

SUMMARY OF FINDINGS, CONCLUSIONS AND
RECOMMENDATIONS

5.0 Introduction
This chapter gives a summary of the main findings of the study in four sections as outlined below. Section 5.1 summarises the main findings of the study, while, section 5.2 outlines some of the conclusions that can be drawn from the study. Recommendations of the study are stated in section 5.3, whereas areas for further research work conclude the chapter in section 5.4.

5.1 Main findings
The results analysed and discussed in chapter four showed that soil organic C is a highly variable soil property with large variations both within fields and between farms with no distinct or clear-cut pattern. However, spatial modelling showed that organic C contains spatial autocorrelation structures. Three semivariogram models, the spherical, exponential and ratio quadratic models were used to model the observed spatial structure. All the three models had a nugget/sill ratio of between 50-60%, indicating moderate spatial correlation.

The exponential semivariogram model provided the best estimates for spatial structure and when fitted in the mixed effects model, it gave the lowest AIC and BIC values (1089.81 and 1177.05 respectively) signifying the model’s high prediction capability. When compared with the initial model which did not account for spatial structure, it was significantly better (p < 0.0001). The maximum range at which the model can be effectively applied is 60m, beyond which prediction errors increase significantly.
Also, high nugget variances (nugget effects) were observed in all the semivariogram models, an indication that a wide range of micro-scale variation which is unaccounted for still exists. Similarly, kriging estimates showed that spatial interpolations of soil fertility predictions are only possible in areas where strong spatial correlation structures exist such as in Y5 and Y9. In the other Y-regions kriging errors are significantly large to allow meaningful spatial interpolation.

Analysis of the estimated variance components showed that the between-fields (residual) variability accounts for up to 62.5% of all the variation associated with random effects. The Y and farm effects account for up to 16.5 and 18.4% of the total variance observed, respectively. The district effect only accounted for a minimal 2.6% of the variance observed. Therefore, most of the variability of predicted soil organic C experienced (approximately up to 63%) occurred within individual farms. On the other hand, all the measured variables considered as fixed effects failed to significantly account for the variations observed in predicted soil organic C. Hence, posing a serious challenge in making recommendations for soil fertility management in smallholder farms of western Kenya.

5.2 Conclusions

This study has shown that soil organic C in western Kenya is spatially structured and the structure was effectively modelled using semivariogram models. Most of the observed variance occurred at smaller intervals and was not effectively modelled but reported as residual variance. This residual (field) variance components accounted for almost ? (about 67%) of the total variance associated with random effects.

The large field variance components together with the high nugget variances observed imply that the microscale variations between fields are sufficiently large to
affect the basic soil processes that determine resource use efficiency by the plant. Their importance should be taken into account when designing and targeting soil fertility management interventions. However, it was not possible to unravel all the causes of the observed variances. A kind of field typology is proposed to try and determine the main factors generating most of the observed microscale variances.

The general multiple linear regression models applied in Section 4.5, showed that several factors contributed to the observed variations in predicted soil organic C, but failed to account for it spatially. The Y-frame sampling design provided a spatially hierarchical structure that was used in mixed-effects modelling and made it possible to identify the variability occurring at different spatial scales and its causes. The set of analytical approaches applied in this study, can be considered to be promising for on-site monitoring and data analysis in farmers’ fields, especially when local variation of soils and crop response are the solid criterion for spatially varying management operations.

5.3 Recommendations

A farm-level approach is likely to be the most appropriate entry point in managing the problem of soil fertility gradients within smallholder farms in western Kenya. The existing spatial correlation structures and the large variance components associated with individual fields are important factors to consider when designing soil fertility management strategies.

On-going and future research in western Kenya either on experimental plots or within farmers’ fields should not assume that the plots are homogenous and spatially independent. The spatial correlation structures within the region should not be
overlooked as well as the large micro-scale variation which needs to be carefully observed and factored into data analyzes.

5.4 Further research areas
A similar study should be carried out in the area using hierarchical sampling combined with grid or transect sampling within each hierarchy at small distances to try and account for the large micro-scale variation observed in this study. The research will also be helpful in providing better kriging estimates which can be used for linear interpolations.
REFERENCES


Smaling EMA, Nandwa SM and Janseen BH(1997) Soil Fertility in Africa is at Stake. In: Buresh RJ, Sanchez PA, and Calhoun F (Eds). _Replenishing Soil Fertility in Africa_, SSSA Special Publ. SI. SSSA, Madison, WI.


# APPENDIX A.1: FIELD CHARACTERISATION (FORM 1)

**FORM 1: Field level information**

<table>
<thead>
<tr>
<th>Name of household head:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Farm ID (District code/ Y code/farm code):</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Homestead Northing:</th>
<th>Easting:</th>
</tr>
</thead>
</table>

## SFG PROJECT- FARMER-LED FIELD CHARACTERISTICS RECORDING FORM (Farmer interview)

<table>
<thead>
<tr>
<th>Field code</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
<th>F12</th>
<th>F13</th>
<th>F14</th>
<th>F15</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Follow farm map)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Land use as?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial =C, subsistence =S, Mixed =M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>When was the field first cultivated?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flooded &gt; 4 months yr⁻¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Yes = ✓ / No = ✗)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>If under fallow, how may seasons?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Input use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fertilizer = F, Organic manure = OM, None = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Farmer assessment of soil fertility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High = H, medium = M, low = L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Crop code:
- Maize = m, Sorghum = s, Millet = mi, Cassava = c, Banana = ba, Beans = b, Groundnuts = g, Cowpea = cp, Green gram = gr, Dolichos = d, Pigeon pea = p, Napier = n, Pasture = pa, Vegetable = v (but specify), Tea = t, Coffee = cof, Cotton = cot, Others (specify)
APPENDIX A.2: FIELD CHARACTERISATION (FORM 2)

FORM 2: Field level information

<table>
<thead>
<tr>
<th>Field code (Follow farm map)</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>F11</th>
<th>F12</th>
<th>F13</th>
<th>F14</th>
<th>F15</th>
</tr>
</thead>
</table>

**Position**
- Upland = u, Mid-slope = m, Bottomland = b
- Drainage = d

**Slope (degrees)**

**Visible Erosion**
- None = 0, sheet = s, Rill = r, Gully/Mass = g

**Soil hard-setting**
- None = 0, temporary = t, Permanent = p

**Conservation Structures**
- None = 0, biological = b, structural = s

**# of conservation structures within field**

**Auger depth (cm)**

**Rock / Stone cover**

**Dominant rock grade**
- Gravel = g, stones = s, boulders = b

Rock scale: 1 = 0-5%, 2 = 5-25%, 3 = 25-50%, 4 = 50-75%, 5 = 75-95%, 6 = 95-100%
APPENDIX A.3: GPS FIELD LOCATION FORM

FORM 3: GPS location of fields

GPS RECORDING FORM

<table>
<thead>
<tr>
<th>Name of household head:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Farm ID (District code/ Y code/farm code):</td>
<td></td>
</tr>
<tr>
<td>Homestead Northing:</td>
<td>Easting:</td>
</tr>
</tbody>
</table>

Draw each field (Maximum of 3 fields per sheet) and label each field as in the farm map
Indicate the direction of slope and slope
Indicate the GPS at field corners and at center of the field
Indicate distance and orientation of 5 x 5m plot as (_______, ______ m)
Indicate structures along field borders i.e. Hedge = H, trees = T, conservation structures = C and species composition of vegetative conservation structures (maximum of 2 dominant species on each hedge)
Appendix B.1: Linear mixed-effects model fit by REML

AIC
1111.602

BIC
1203.436

LogLik
-535.8011

Random effects:
Formula: ~ 1 | District
  (Intercept)
StdDev: 0.08992693

Formula: ~ 1 | Y %in% District
  (Intercept)
StdDev: 0.2253581

Formula: ~ 1 | Level %in% Y %in% District
  (Intercept) Residual
StdDev: 0.237517 0.438147

Fixed effects: C ~ Landuse + Fertility + Place + Rock + Rockgrad

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.81</td>
<td>0.15</td>
<td>640</td>
<td>12.17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Landuse1</td>
<td>0.08</td>
<td>0.05</td>
<td>640</td>
<td>1.74</td>
<td>0.08</td>
</tr>
<tr>
<td>Landuse2</td>
<td>0.05</td>
<td>0.02</td>
<td>640</td>
<td>2.29</td>
<td>0.02</td>
</tr>
<tr>
<td>Fertility1</td>
<td>-0.008</td>
<td>0.02</td>
<td>640</td>
<td>-0.29</td>
<td>0.77</td>
</tr>
<tr>
<td>Fertility2</td>
<td>-0.02</td>
<td>0.01</td>
<td>640</td>
<td>-1.45</td>
<td>0.15</td>
</tr>
<tr>
<td>Place1</td>
<td>-0.04</td>
<td>0.03</td>
<td>640</td>
<td>-1.20</td>
<td>0.23</td>
</tr>
<tr>
<td>Place2</td>
<td>-0.01</td>
<td>0.02</td>
<td>640</td>
<td>-0.64</td>
<td>0.52</td>
</tr>
<tr>
<td>Place3</td>
<td>0.008</td>
<td>0.03</td>
<td>640</td>
<td>0.33</td>
<td>0.74</td>
</tr>
<tr>
<td>Rock1</td>
<td>0.08</td>
<td>0.06</td>
<td>640</td>
<td>1.28</td>
<td>0.20</td>
</tr>
<tr>
<td>Rock2</td>
<td>0.01</td>
<td>0.06</td>
<td>640</td>
<td>0.20</td>
<td>0.84</td>
</tr>
<tr>
<td>Rock3</td>
<td>-0.11</td>
<td>0.09</td>
<td>640</td>
<td>-1.28</td>
<td>0.20</td>
</tr>
<tr>
<td>Rock4</td>
<td>-0.04</td>
<td>0.06</td>
<td>640</td>
<td>-0.76</td>
<td>0.45</td>
</tr>
<tr>
<td>Rock5</td>
<td>0.04</td>
<td>0.04</td>
<td>640</td>
<td>0.94</td>
<td>0.35</td>
</tr>
<tr>
<td>R-grade1</td>
<td>0.11</td>
<td>0.11</td>
<td>640</td>
<td>0.99</td>
<td>0.33</td>
</tr>
<tr>
<td>R-grade2</td>
<td>-0.09</td>
<td>0.05</td>
<td>640</td>
<td>-1.84</td>
<td>0.07</td>
</tr>
<tr>
<td>R-grade3</td>
<td>0.006</td>
<td>0.02</td>
<td>640</td>
<td>0.24</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Standardized Within-Group Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Q1</th>
<th>Med</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2.155319</td>
<td>-0.6418257</td>
<td>-0.1122786</td>
<td>0.4784789</td>
<td>4.902486</td>
</tr>
</tbody>
</table>

Number of Observations: 745
Number of Groups:
District Y %in% District Level %in% Y %in% District
2 9 90
Appendix B.2: Semivariogram Distance Estimates.

<table>
<thead>
<tr>
<th>Category</th>
<th>Variogram</th>
<th>distance</th>
<th>No. pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.633</td>
<td>13.9</td>
<td>151</td>
</tr>
<tr>
<td>2</td>
<td>0.969</td>
<td>21.6</td>
<td>151</td>
</tr>
<tr>
<td>3</td>
<td>0.913</td>
<td>28.1</td>
<td>151</td>
</tr>
<tr>
<td>4</td>
<td>0.772</td>
<td>33.4</td>
<td>151</td>
</tr>
<tr>
<td>5</td>
<td>0.737</td>
<td>38.6</td>
<td>151</td>
</tr>
<tr>
<td>6</td>
<td>0.735</td>
<td>43.9</td>
<td>151</td>
</tr>
<tr>
<td>7</td>
<td>0.937</td>
<td>48.9</td>
<td>151</td>
</tr>
<tr>
<td>8</td>
<td>0.938</td>
<td>54.2</td>
<td>151</td>
</tr>
<tr>
<td>9</td>
<td>0.934</td>
<td>59.4</td>
<td>151</td>
</tr>
<tr>
<td>10</td>
<td>1.111</td>
<td>66.2</td>
<td>152</td>
</tr>
<tr>
<td>11</td>
<td>0.946</td>
<td>73.6</td>
<td>151</td>
</tr>
<tr>
<td>12</td>
<td>0.814</td>
<td>81.1</td>
<td>151</td>
</tr>
<tr>
<td>13</td>
<td>0.701</td>
<td>89.4</td>
<td>151</td>
</tr>
<tr>
<td>14</td>
<td>0.978</td>
<td>98.7</td>
<td>151</td>
</tr>
<tr>
<td>15</td>
<td>1.297</td>
<td>110.6</td>
<td>151</td>
</tr>
<tr>
<td>16</td>
<td>0.814</td>
<td>125.2</td>
<td>151</td>
</tr>
<tr>
<td>17</td>
<td>1.124</td>
<td>144.5</td>
<td>151</td>
</tr>
<tr>
<td>18</td>
<td>0.839</td>
<td>169.1</td>
<td>151</td>
</tr>
<tr>
<td>19</td>
<td>1.143</td>
<td>205.4</td>
<td>151</td>
</tr>
<tr>
<td>20</td>
<td>1.607</td>
<td>310.5</td>
<td>152</td>
</tr>
</tbody>
</table>
Appendix B 3: Fitted Ratio Semivariogram Model

Generalized least squares fit by REML
Model: C ~ Landuse + Fertility + Place + Rock + Rockgrad
Data: Test
  AIC    BIC    logLik
1094.669 1181.911 -528.3346

Correlation Structure: Rational quadratic spatial correlation
Formula: ~ Easting + Northing
Parameter estimate(s):
  range  nugget
  697.0788  0.5894541

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.89</td>
<td>0.135</td>
<td>14.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Landuse1</td>
<td>0.05</td>
<td>0.047</td>
<td>1.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Landuse2</td>
<td>0.04</td>
<td>0.021</td>
<td>2.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Fertility1</td>
<td>0.00</td>
<td>0.028</td>
<td>-0.16</td>
<td>0.87</td>
</tr>
<tr>
<td>Fertility2</td>
<td>-0.01</td>
<td>0.014</td>
<td>-1.04</td>
<td>0.30</td>
</tr>
<tr>
<td>Place1</td>
<td>-0.02</td>
<td>0.034</td>
<td>-0.61</td>
<td>0.54</td>
</tr>
<tr>
<td>Place2</td>
<td>0.00</td>
<td>0.020</td>
<td>0.17</td>
<td>0.87</td>
</tr>
<tr>
<td>Place3</td>
<td>0.02</td>
<td>0.027</td>
<td>0.69</td>
<td>0.49</td>
</tr>
<tr>
<td>Rock1</td>
<td>0.08</td>
<td>0.062</td>
<td>1.29</td>
<td>0.20</td>
</tr>
<tr>
<td>Rock2</td>
<td>0.01</td>
<td>0.055</td>
<td>0.20</td>
<td>0.84</td>
</tr>
<tr>
<td>Rock3</td>
<td>-0.11</td>
<td>0.084</td>
<td>-1.34</td>
<td>0.18</td>
</tr>
<tr>
<td>Rock4</td>
<td>-0.05</td>
<td>0.056</td>
<td>-0.92</td>
<td>0.36</td>
</tr>
<tr>
<td>Rock5</td>
<td>0.05</td>
<td>0.041</td>
<td>1.32</td>
<td>0.19</td>
</tr>
<tr>
<td>R-grade1</td>
<td>0.11</td>
<td>0.116</td>
<td>0.91</td>
<td>0.36</td>
</tr>
<tr>
<td>R-grade2</td>
<td>-0.10</td>
<td>0.048</td>
<td>-2.09</td>
<td>0.04</td>
</tr>
<tr>
<td>R-grade3</td>
<td>0.00</td>
<td>0.02</td>
<td>0.10</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Standardized residuals:
  Min  Q1  Med  Q3  Max
-2.14 -0.81 -0.15 0.46 5.14

Residual standard error: 0.571145
Degrees of freedom: 745 total; 729 residual
Appendix B 4: Fitted Exponential Semivariogram Model

Generalized least squares fit by REML
Model: C ~ Landuse + Fertility + Place + Rock + Rockgrad
Data: Test

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1089.8</td>
<td>1177.05</td>
<td>-525.90</td>
</tr>
</tbody>
</table>

Correlation Structure: Exponential spatial correlation
Formula: ~ Easting + Northing

Parameter estimate(s):
- range: 1176.533
- nugget: 0.5342801

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Value</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.90</td>
<td>0.142</td>
<td>13.37</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Landuse1</td>
<td>0.06</td>
<td>0.047</td>
<td>1.25</td>
<td>0.21</td>
</tr>
<tr>
<td>Landuse2</td>
<td>0.04</td>
<td>0.021</td>
<td>1.97</td>
<td>0.05</td>
</tr>
<tr>
<td>Fertility1</td>
<td>0.00</td>
<td>0.028</td>
<td>-0.16</td>
<td>0.88</td>
</tr>
<tr>
<td>Fertility2</td>
<td>-0.01</td>
<td>0.014</td>
<td>-0.88</td>
<td>0.38</td>
</tr>
<tr>
<td>Place1</td>
<td>-0.02</td>
<td>0.034</td>
<td>-0.60</td>
<td>0.55</td>
</tr>
<tr>
<td>Place2</td>
<td>0.00</td>
<td>0.021</td>
<td>0.16</td>
<td>0.88</td>
</tr>
<tr>
<td>Place3</td>
<td>0.02</td>
<td>0.027</td>
<td>0.74</td>
<td>0.46</td>
</tr>
<tr>
<td>Rock1</td>
<td>0.08</td>
<td>0.062</td>
<td>1.32</td>
<td>0.19</td>
</tr>
<tr>
<td>Rock2</td>
<td>0.02</td>
<td>0.055</td>
<td>0.30</td>
<td>0.77</td>
</tr>
<tr>
<td>Rock3</td>
<td>-0.11</td>
<td>0.084</td>
<td>-1.30</td>
<td>0.20</td>
</tr>
<tr>
<td>Rock4</td>
<td>-0.05</td>
<td>0.056</td>
<td>-0.95</td>
<td>0.34</td>
</tr>
<tr>
<td>Rock5</td>
<td>0.05</td>
<td>0.041</td>
<td>1.31</td>
<td>0.19</td>
</tr>
<tr>
<td>R-grade1</td>
<td>0.11</td>
<td>0.116</td>
<td>0.95</td>
<td>0.34</td>
</tr>
<tr>
<td>R-grade2</td>
<td>-0.10</td>
<td>0.049</td>
<td>-2.09</td>
<td>0.04</td>
</tr>
<tr>
<td>R-grade3</td>
<td>0.00</td>
<td>0.024</td>
<td>0.02</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Standardized residuals:
- Min: -2.118962
- Q1: -0.8061667
- Med: -0.1668528
- Q3: 0.45385
- Max: 5.034268

Residual standard error: 0.5824017
Degrees of freedom: 745 total; 729 residual
Appendix B 5: Fitted Spherical Semivariogram Model

Generalized least squares fit by REML
Model: C ~ Landuse + Fertility + Place + Rock + Rockgrad
Data: Test

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>1090.4</td>
<td>1177.6</td>
<td>-526.195</td>
</tr>
</tbody>
</table>

Correlation Structure: Spherical spatial correlation
Formula: ~ Easting + Northing
Parameter estimate(s):
- range 2708.47
- nugget 0.5148

Coefficient Value      Std.Error  t-value  p-value
(Intercept)  1.91      0.143       13.33   <.0001
Landuse1    0.06      0.047       1.26    0.21
Landuse2    0.04      0.021       2.03    0.04
Fertility1  -0.00      0.028      -0.13   0.89
Fertility2  -0.01      0.014      -0.90   0.37
Place1      -0.02     0.034       -0.61   0.54
Place2      0.00      0.020       0.17    0.87
Place3      0.02      0.027       0.73    0.47
Rock1       0.08      0.062       1.31    0.19
Rock2       0.01      0.055       0.26    0.79
Rock3       -0.11     0.084      -1.35   0.18
Rock4       -0.05     0.056      -0.93   0.35
Rock5       0.05      0.041       1.31    0.19
R-grade1    0.11      0.116       0.97    0.33
R-grade2    -0.11     0.049      -2.16   0.03
R-grade3    0.00      0.024      -0.05   0.96

Standardized residuals:

Min Q1 Med Q3 Max
-2.090868 -0.7914366 -0.1682617 0.4321174 4.892108

Residual standard error: 0.5984329
Degrees of freedom: 745 total; 729 residual

ANOVA Model Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>1111.6</td>
<td>1203.4</td>
<td>-535.8</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>1089.8</td>
<td>1177.1</td>
<td>-525.9</td>
</tr>
</tbody>
</table>

Test L.Ratio  p-value
1 vs 2 19.794 <.0001