EFFECTIVE RESEARCH METHODS SUPPORT TO LARGE SURVEYS; A SCENE FOR THE CLIMATE CHANGE, AGRICULTURE AND FOOD SECURITY ON A FIELD TYPOLOGY SURVEY

BY

JOASH MANGO

A project report submitted in partial fulfillment of the requirements for the degree of Master of Science in Research Methods

School of Mathematics, Statistics and Actuarial Sciences

MASENO UNIVERSITY

ABSTRACT

The main objective of this work was to find strategies to support science projects with limited research methods support. Many research projects face major challenges in the full implementation on the ground. This forms the basis of the research methods program, which intends to bridge the gap between science design and implementation. This study characterized the nature and magnitude of variability in field management through a survey that captured diversity at the farm level. Heterogeneity in land management is determined by settlement patterns and land access, management imposed on the land as this can create large differences in soil fertility over short distances within the same farms that influence crop production and fertility management interventions. This study supported the research project by advising on the best way to design the survey, then how to process and manage the data generated from the survey. This was accomplished through advising on efficient use of a data management tool (CS Pro software) especially for data entry and quality control and the final backing up of the data on external devices; final archiving of the data on a database. Besides these, there was one to one consultancy with the scientists and this encompass joint decision making on what variables to use for final analysis and the manner in which the results have been presented. Data analysis involved use of spatial analysis and statistics to do exploratory data analysis and further inferential tests to check the relationships that exist between the collected variables. Finally, this study presents a review of previous work and recommendations given on the methods to be used for future research work.

CHAPTER ONE

INTRODUCTION OF THE STUDY

1.1 Introduction

Maintaining productivity in most stallholder farming systems has relied in the past on shifting cultivation and fallowing. Due to human population increases, expansion of agricultural production through extensification has rendered these approaches impossible in most smallholder farming systems. Many smallholder farming systems have reached a limit in terms of intensity of land without inputs and are in transitions to intensification. This is the case with the smallholder farming systems in the study area in Nyando. Intensification requires increased use of nutrient inputs by managing crop residues, application of organic manures and inorganic fertilizer to achieve and sustain high land productivity.

At farm scale, however, it becomes evident that the processes leading to soil-fertility depletion are not homogeneously distributed in space. Variability in soil fertility arises from differences in underlying geology and geo-morphology, and due to a number of human induced mechanisms within the farming systems (i.e. farm management practices). Farmers manage several organic and mineral resources in order to attain their production goals. 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 (Vanlauwe *et al,* 2001). There is a need to revisit the strategies to deployment of nutrients management interventions in the heterogeneity in soil fertility within the household. Relevant scales need to be determined at which nutrient management interventions are being applied within the farm household system. This project recognizes that agriculture is very sensitive to climate change, and that climate change has the potential to transform food production, especially the pattern and productivity of crops, livestock, and fishery systems and to reconfigure food distribution and market access (Nelson, et al.,2009). In order to achieve implementation of research project this there is need to have proper research method support.

1.2 Statement of the problem

There has been limited capacity to offer research methods support to researchers and lack, in terms of numbers and competence of human resource capacity throughout sub- Saharan Africa. This limited support and inadequate capacity has weakened the quality of research and progress towards increasing food and nutritional security and alleviating poverty. This study was designed to bridge the gap between theoretical rigor and practical needs where the traditional areas of statistics and biometry have tended to emphasize mathematical statistics with little on the ground design and emerging development issues which require complementary competences. This has lead to improvement on the methodology of a field typology so as to replicate it in other sites.

1.3 Objectives of the study

This project report concentrates on characterization of nutrients management at a farm level within a smallholder farming system in the Lower Nyando site with the broad objective of contributing to understanding of how to improve data collection, management and analysis. The specific objectives are:

1. To quantify the magnitude of within-farm soil fertility gradients as affected by biophysical (e.g., variation in soil fertility within one farm) and socio-economic (e.g., household characteristic) conditions.
2. Identify the main variables for data analysis.
3. To identify the main factors affecting crop yield variability.

1.4 Significance of the study

Addressing soil depletion, food security and vulnerability in Sub-Saharan Africa require sufficient high quality data and background information that accessed by various researchers and stakeholders to design interventions. A field typology survey was considered to be suitable to gather information from the smallholder farmers at local levels about the various aspects of land use and nutrient management. Such information, supported by the relevant data, can be further used by researchers to investigate the complex relationships amongst these variables to see how best smallholder farmers can be helped to continue producing from their own farms even in the face of the ever changing climate. The findings are helpful to NGOs and other development practitioners, policy makers, and governments that are dealing with the trade-offs involved with moving towards integrated nutrient management in the Nyando farming system. The methodologies and research methods techniques employed in the field typology survey can also be adopted by researchers who intend to carry out similar studies.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter gives an in- depth background information of the underlying principles of data management, consultancy, communication skills, data analysis and revision of research proposals, which form the key components of the research project requirements.

2.1 Data organization and management

Data management includes all aspects of data planning, handling, analysis, documentation and storage, and takes place during all stages of a study (Bennet, 2001). Data used for analysis and reaching conclusions must be correct. If proper attention is not paid to data management, it is too easy to make mistakes in the processing, resulting in incorrect conclusions (Muraya & Chege, 2002).

It is recognized that surveys have long provided inexpensive means for collecting population based-information. Survey information can be collected in an extremely structured manner, or may be more informal, or a mixture of both approaches. Whatever tools used to collect the information, it is essential to maintain consistency throughout the exercise and to avoid errors arising from inadequately prepared tools. Most social scientists find that much time is taken in preparing the data for statistics, modeling that they want to complete and as the design and administration of surveys grows more complex, researchers are being challenged by the logistics in carrying out large-scale surveys and there is need to use a well-defined data management system for the success of the project (Massoud, 2009). A rich literature exists and continues to grow on the topic of survey data quality (Lyberg et al. 1997; Collins and Sykes 1999) and its management in national statistical agencies (Brackstone, 1999). Definitions of the concept proliferate, but clustered around the idea that the characteristics of the product under development meet or exceed the stated or implied needs of the user. Arondel & Depoutot (1998) suggests in their review that statistical organizations should break down quality into components or characteristics that focus around several key concepts: accuracy, relevance, timeliness, and accessibility. Accuracy is an important and visible aspect of quality that has been of concern to statisticians and survey methodologists for many years. It relates to the closeness between estimated and true (unknown) values. For many, accuracy means the measurement and reporting of estimates of sampling error for sample survey programs, but, in fact, the concept is much broader, taking into account no sampling error as well. Timeliness can refer to several concepts. First, it refers to the length of the data collection's production time—the time from data collection until the first availability of a product. Fast release times are without exception looked upon favorably by end users. Second, timeliness can also refer to the frequency of the data collection. Timely data are current data. Accessibility, as a characteristic of data quality, refers to the ability of data users to obtain the products of the data collection program. Data products have their most value—are most accessible—when they are easily available to end-users and in the forms and formats desired. Data products are of several types—individual micro data in user-friendly formats on different media, statistical tabulations on key survey variables, and analytic and descriptive analysis reports. Accessibility also implies the data products that include adequate documentation and discussion to allow proper interpretation of the survey results. Accessibility can also be described in terms of the efforts data producers make to provide technical assistance in using and interpreting the data products through consultation, training classes, etc. (Depoutot, 1998) suggests three other characteristics of data quality: comparability of statistics, coherence, and completeness. Comparability of statistics refers to the ability to make reliable comparisons over time; coherence refers to the ability of the statistical data program to maintain common definitions, classifications, and methodological standards when data originate from several sources; and completeness is the ability of the statistical data collection to provide statistics for all domains identified by the user community. To ensure that the quality of the data collected is not compromised in any capacity, the following key steps have been suggested by a number of researchers as a measure of complete data management strategies: Planning data management for the project, taking into account the objectives and planned outputs, the resources and skills available. At this stage, a data management plan should be set up to describe what data will be created, what policies will apply to the data, who will own and have access to the data, what data management practices will be used, what facilities and equipment will be required, and who will be responsible for each of these activities. The next step will involve checking of raw data and this will entail finding out if there are any missing values or if the variables are clearly labeled amongst others. Data entry and organization of computer files will follow whereby validation rules are put in place to minimize errors that would be made during the data entry process .Data entry process should be done promptly and simply.

Once all the files have been well organized in the computer system, the data should be backed up outside the computer to take care of any eventuality that might arise that would otherwise lead to the loss of the data. Backing up of the raw data can be done in DVDs, CDs, USB drives (flash disks) or any external portable memory outside the computer. Another important aspect of data management is processing the data for analysis. This mainly involves arranging the variables in rows and columns that can be read by the software to be used, removing blanks spaces and relabeling the variables in a readable form by the software. Re checking the processed data is equally important to ensure that error occurrence is minimized and from this, it is appropriate to maintain data processing log, indicating changes or adjustments made to the dataset before analysis. Once all these steps have been ensured, the data is archived in a well-defined database for future use.

2.2 Data analysis

Analysis of data is a procedure of processing, cleaning, transforming and modeling of data with the aim of extracting useful information, suggesting conclusions and supporting decision making (Tabachnick, 2007). The process has multiple phases and approaches, encompassing diverse techniques under a variety of names, in different science and social science domains. In statistical applications, data analysis has been divided into three principal components; descriptive statistics, exploratory data analysis, and confirmatory data analysis. In statistical context, exploratory data analysis (EDA) is an approach to data analysis that employs a variety of techniques (mostly graphical) to maximize insight into a dataset, uncover underlying structure, extract important variables, detect outliers and anomalies, test underlying assumptions, develop parsimonious models and determine optimal factor settings (Hoaglin et al., 2000). At this stage, the statistician would want to look at the "shape" of the data, to see where most values lie, check whether they are clumped around a central value, and if so, to see if there are roughly as many above this value as below it. Here, we also look at the distribution for each variable to determine which analyses would be most appropriate. Sometimes, it is necessary to examine distributions of data partitioned by other key variables and this can be checked using histograms, box plots or scatter plots to check the pattern of association between any two quantitative variables. Data exploration would also reveal if there are any missing values where in surveys, missing values correspond to skipped questions or unendorsed options and outliers which represent "weird" values in the dataset. A discussion between analyst and client should take place in determining how missing values should be handled. However, in some cases, missing values for important variables might exclude a record from certain analyses. Sometimes, it is appropriate to place normalized values in place of missing values. Data exploration is an important step in data analysis as it will lead to further data cleaning which is extremely important when data collection method allows inconsistencies and this may involve labeling missing values and removal of outliers.

Descriptive statistics on the other hand, aims at describing the basic features of the data in a study (McPherson, 2001). They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data. Descriptive statistics are used to present quantitative descriptions in a manageable form. In a research study, there may be lots of measures and descriptive statistics helps in simplifying large amounts of data in a sensible way. Each descriptive statistic reduces lots of data into a simpler summary. For univariate analyses i.e. analyses involving the examination across cases of one variable at a time, there are three major characteristics of a single variable that we tend to look at through descriptive statistics and these include; the distribution of the data. The distribution is a summary of the frequency of individual values or ranges of values for a variable. The simplest distribution would list every value of a variable and the number of individuals who had each value and these can be presented in terms of percentages or mere counts. The second proponent will comprise of the central tendency which tries to estimate the "center" of a distribution of values (Trochim, 2008). The central tendency would want to establish the mean, median and mode. Also to descriptive statistics, it would be important to look into the dispersion aspect i.e. spread of values around the central tendency. The two common measures of dispersion are the range and the standard deviation. The range implies the difference between the highest and the lowest value in the dataset. On the other hand, the standard deviation is considered a more accurate and detailed estimate of dispersion because an outlier can greatly exaggerate the range. The standard deviation shows the relation that set of scores has to the mean of the sample and all these components must be checked into during descriptive analysis.

Once a clear data exploration and descriptive analyses are achieved, what follows is the inferential analysis that marks the climax of any data analysis process. Inferential statistics tests hypotheses about the data and may permit generalization beyond the dataset. Examples include comparing means (averages) for a given measurement between several different groups or for the same individuals across time. In his review, McPherson (2001) identifies three common aspects of inferential analyses that are widely used by many researchers depending on the nature of the study and these comprise of; Trends Analysis which mainly applies to studies involving repeated measurements at different points in time and this can help find changes or patterns over time. Second to his category, is Analysis Of Variance Models: Analysis of variance is used to compare average scores for different groups. Many different ANOVA models are available to tease out the effect of covariates (factors expected to relate to the response variable) and handle multiple dependent variables (outcomes). Most common to detailed inferential analysis is Multiple Regression & Correlation: Correlation measures the strength of the relationship between different variables in the data while on the other hand, multiple regression examines how well one set of variables predicts an outcome variable and specifies the unique contributions of each predictor. Canonical correlation analysis relates one set of variables to another set. In data analysis, it's important to be in a position to interpret the analysis results and decide which of the results to include in the final write up. Also of importance, the researcher should have a very clear idea on how best the results can be presented in the report for ease of understanding and interpretability by the end user.

2.3 Consultancy and communication

Consultancy is a structured process for helping an individual or a team to think more expansively about a particular, concrete dilemma. Outside perspective is critical to this protocol working effectively; therefore, some of the participants in the group must be people who do not share the presenter's specific dilemma at that time (Grove, 2009). In statistical terms, (Bland, 1998) defines consultancy as a collaboration of a statistician with a scientist for the purpose of devising solutions to research problem. Such consultation may involve brief interactions to respond to very specific questions or long term associations as team members on on-going projects. Statistical consultants practice their crafts in a variety of settings though the process of consultancy remains virtually the same (Stegman, 1995). One important aspect of consultancy is that a statistician must be able to effectively communicate with researchers and practitioners and be conversant in their functional areas. Consultancy skills should therefore be an important aspect of statistical training (Ruberg, 1998; Khamis, 1994; Rangecroft & Wallace, 1998; van Belle, 1982). It is important for every consultant to be able to use interviewing techniques, diagnose client's problems and purposes, structure and plan work to be done, enlist the client's collaboration, communicate with the client, share information and knowledge with the client, prepare proposals and conclusions orally or in writing.

To conduct an effective consultancy session, three key stages have to be envisaged. First is to establish rapport which usually begins with exchanging of pleasantries and small talk. In his consultancy experience, Zan (1995) outlines that first-time clients are always apprehensive because they do not know what to expect and may feel defensive about their statistics background. He goes further to suggest that it is important to allay client's apprehension by exhibiting a caring attitude that encourages open sharing of information. The second stage of consultation is concerned with identification of the research problem. During the early part of this stage, the client does most of the talking and the consultant asks questions to clarify point if necessary and the responses from the client should be objective enough and closely related to the area of study. Early in the second stage, it is important to determine if the person the consultant is talking to is the real client and if other people are involved in the research. Isenberg (1996) in his observation notes that there is a small group of researchers who, because they are busy attempt to consult through an intermediary and such people receive inferior advice and that is why the consultant should insist on seeing the real person involved in the research. It is also important to understand the significance of the client's research and see how well it fits into the knowledge base of a discipline (Bland, 1985). In the second stage of a consulting session, the consultant should obtain an accurate and relatively complete understanding of the various aspects of the research that have implications for the design and analysis. Incorrect assumptions about how the data had been collected should not be allowed to vitiate the analysis and interpretation of an experiment or a survey. In the third stage of consultancy, once the consultant has obtained all the information required, he/she will provide statistical advice to the client on how to go about the research problem based on the facts obtained and even from previous experiences.

2.4 Planning and design of projects

Planning involves initially figuring out what you want to do, how you are going to do it and in projects how you are going to evaluate its performance. It involves identifying priority areas- a research problem, discussing and testing the various possible courses of action, choosing the most appropriate one (or ones), agreeing what you can expect to achieve, calculating the human and material resources needed to reach your objectives, anticipating possible problems and getting agreement among all concerned about clear targets and timetables for the work in view (IFRS, 2000). This can be done using many research strategies among them mono-disciplinary, multidisciplinary, interdisciplinary or trans-disciplinary, which also describe the team's interaction process. However the one most commonly used is the trans-disciplinary approach whereby experts from multiple disciplines collaborate as equal partners to create new knowledge theory that results in synergy that breaks disciplinary boundaries to enable effective investigation on the complex health and environmental interactions, taking into account cultural, economic, ecological, bio-medical, governance, gender, social and political context with the participation of stake holders. Planning is done to increase the likelihood that a project will be implemented efficiently, effectively and successfully (IFRS, 2000). There are several techniques used to plan projects; the Gantt chart is used in planning and scheduling complex projects; critical path technique (CRT) is for more complex projects; log frames; outcome mapping focuses on the changes in behaviour and progress towards impact; Theory of change (TOC) which is a graphical depiction of how a project would evolve and change through time while PRINCE2 (Projects IN Controlled Environments) and PMBOOK(Project Management Body Of Knowledge) are used for planning large projects or programs (Mind tools, 2010).

2.5 Reviewing reports and proposals

The process of reviewing research reports involves critical and objective analysis of historical work, seeing how it was done and most possibly identifying ways in which it would have been done in a much better way. This will involve a peer review which is defined as an objective process of subjecting an author's scholarly work, research or ideas to the scrutiny of others who are experts in the same field. According to (Ware, 2008) peer review is an important way of ensuring that evidence can better inform policy and practice. It is also done to improve on quality of publications, to increase relevance and/or appropriateness to the field by filtering out bad work, poorly conceived and minimum originality. With regards to the researcher it may be a way of motivating researchers to improve on their research and for them to recognize standards.

There are several ways of conducting a review. The most common one according to (Weiger, 2002) are Inspection, which is the most systematic and rigorous type of peer review. It follows a defined multistage process with roles defined to individuals. It is more effective at finding defects; Team review, which is less formal and less rigorous, participants roles may be combined, or simplified; Walkthrough, here the author simply describes it to colleagues and solicits comments; Peer desk check, this is where only one person besides the Author examines the work.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter focuses on the design of the study, site selection criteria, sampling schemes and the overall implementation of a field typology survey. It goes further to shed more light on how the various research methods tasks were accomplished. Amongst the tasks widely discussed in this chapter include: data management, data analysis, project planning and consultancy services offered to the research participants during the project phase.

3.2 Research design

The field typology survey was designed to collect information on biophysical, socio-economic and management aspect of each plot using a sample of farms that represent the study site. Socio-economic and farm management information included characteristics of the household head (name and gender), map of the farm, land use patterns and production. The different fields of each farm were identified with aid of a map drawn by the farmer. Centre and perimeter of each field were geo-referenced by means of a GPS (global positioning system). The area of each field was determined with GPS. Biophysical information was collected on field by field basis and included field characteristics (flooding, erosion etc.) and management (nutrient input use, soil conservation measures, etc.).

3.2.1 Research sites

The field typology survey was conducted in Lower Nyando. The site used is one of the benchmark sites of CCAFS which followed the criteria described below.

3.2.2 Site selection criteria

Table 1: Criteria used to select the research sites

Criterion
A set of research locations representing key biophysical and agro-ecological gradients of the respective regions

Research locations that represented the key socio-economic and (where relevant) demographic gradients for the region, including extent of urbanization and gendered participation in different agricultural production systems
Research sites that lied along gradients of anticipated temperature and precipitation change
Research sites that lied along gradients of current and anticipated land use pressure
Research sites that represented different institutional (e.g. land tenure) arrangements. Similarly, gradients of significant difference in political and governance history
Sites that had significant but contrasting climate-related problems and opportunities for interventions
High potential sites, i.e. where impact was likely to be achieved: sites that built on ongoing CGIAR and national research infrastructure and research sites, and thus had good existing data on historical weather records; characterization of the natural resource base; detailed, longitudinal data on agricultural production; detailed, longitudinal socio-economic and demographic data at the household and village settlement/district level; data on the food system; and data on historical events and shocks experienced in relation to food security in the site
Governance and institutional capacity that favored the likelihood of scaling up and generating transferable results
Local champions knowledgeable about the site and available for/committed to collaborative work with CCAFS
A network of regional partners that would facilitate scaling up
Sites that had mitigation/carbon sequestration potential
Sites that were safe to work in, i.e. had good security for research teams
Research sites that were physically accessible and had the minimum logistical comforts for conducting research
Marginal sites with high vulnerability where impact would be difficult to achieve but where the need for innovative solutions to poverty and climate change vulnerability may be greatest

3.2.3 Sampling scheme for research grids, villages and household selection

A multistage sampling technique was used where 3 layers in a hierarchy: a grid or 10 x 10 km block, villages within a block (20) and households within each village (10). This scheme did not refer explicitly to administrative hierarchies.

3.2.3.1 Block selection

'Site' meant the district or set of adjacent districts that were selected based on the set of criteria above (Table 1).Within that larger site; a 10 x 10 km block was selected for the field typology survey.

3.2.3.2 Village selection

A random sample of 20 villages was taken from all those in the block. If there were less than 20 villages in the block, then all were selected.

3.2.3.3 Household and sample size selection

10 households were randomly chosen from each of the 20 villages to constitute a total of 200 households. The proposed sample sizes were:

- 1 block
- 20 villages per block
- 10 households per village: 20 x 10 = 200 households.

Figure 1: A Map showing the Lower Nyando site.

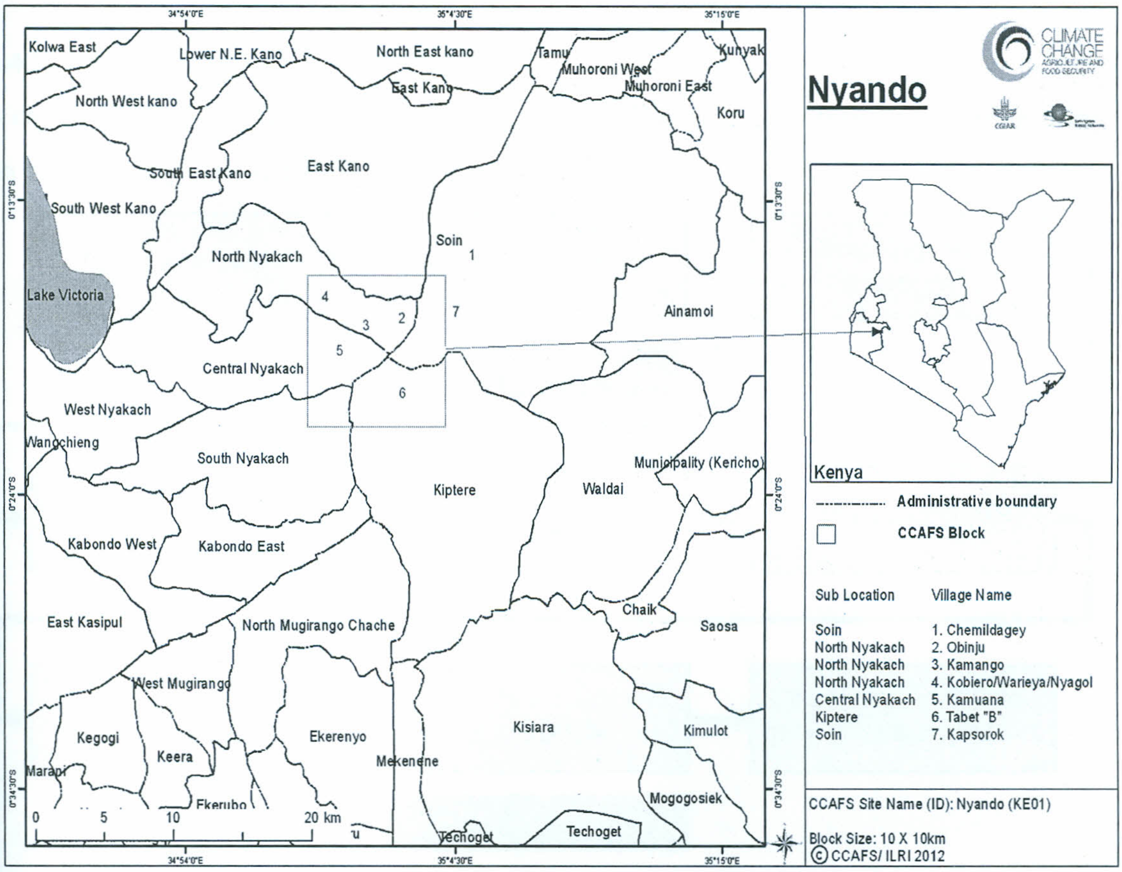


Figure 1: Location of CCAFS benchmark Nyando site, Kenya (Source: Onyango et. al., 2012)

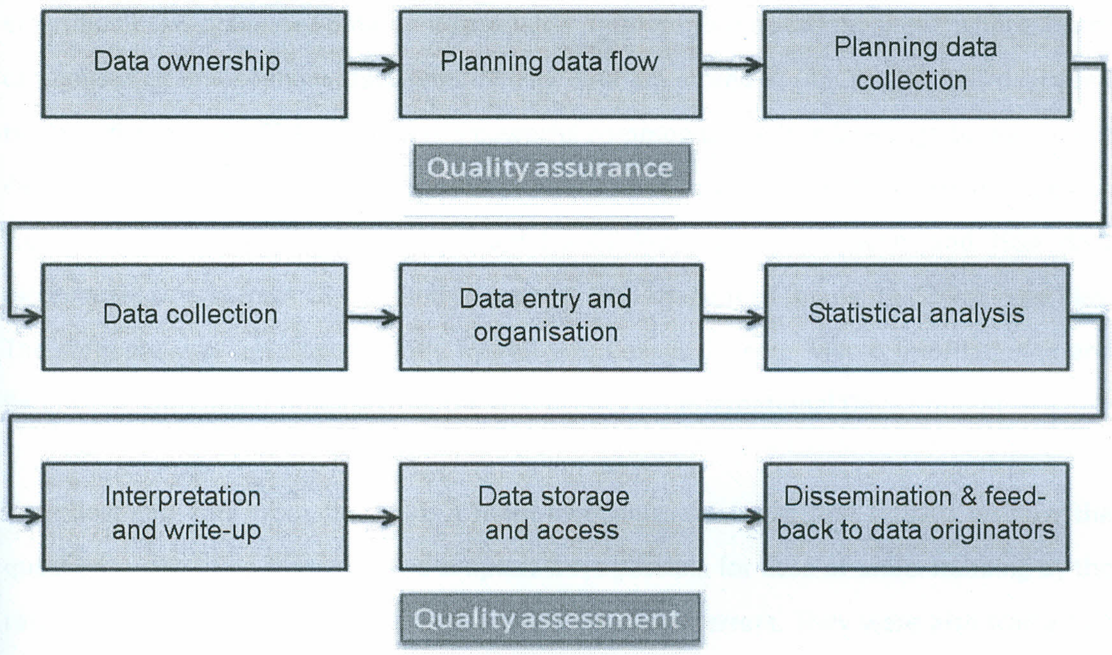
3.3 Data collection Procedures

3.3.1 Questionnaires

The field typology survey employed questionnaires to collect information. Questions on the relative importance of various production activities, management practices and resource allocation patterns were put to the farmer. This is a method whereby the interviewer is free to probe, it allows flexibility in asking additional questions and seeks clarification and elaboration on the answers given, and it is useful in building confidence to the respondent when responding to questions from the researcher. The questionnaires used for the survey were mainly structured in nature with closed ended questions to regulate the responses given by the interviewees. The GPS coordinates of the households and fields were recorded in Universal Transverse Mercator (UTM) system using the GPRS gadget so as to clearly map out the areas under study for ease of traceability during the next visit. Detailed information of the data collection procedures is covered under data management of this document.

3.4 Data Management

A complete data management plan for the Climate Change, Agriculture and Food Security on field typology survey adopted the following model.



Author: Gerald W. Chege and Peter K. Muraya (2001).

To effectively manage the data generated from the field typology survey, a complete data management plan had to be set up before the execution of the whole survey. Proper documentation of the plan was made available to all the team players, especially to the field supervisors to ensure that quality assurance in the data flow cycle was observed at all stages to avoid mix up of contents and loss of important information. The main components of the plan chiefly comprised of the objectives of the study, types of variables that were to be collected, the manner in which such variables were to be recorded by the enumerators, entered into the computer system, checked for quality control, backed up and archived in a permanent database for ease of access and for future use. Also, definition of the research activities had to be made clear with well-defined roles of the various participants of the survey. As part of the plan, Census Survey processing system commonly referred to as CS Pro was used to design the questionnaires that were used to collect the data from the field. CS Pro is a software package for entry, editing, tabulation, and dissemination of census and survey data. CS Pro combines the features of the Integrated Microcomputer Processing System (IMPS) and the Integrated System for Survey Analysis (ISSA) in a Windows environment. The software has a facility that lets the user to create, modify, and run data entry, batch editing, and tabulation applications from a single, integrated development environment. It processes data on a case basis (one or more questionnaires), where a case can consist of one or many data records. The data are stored in (ASCII) text files described by data dictionaries. CS Pro contains a powerful common procedure language to implement data entry control and edit rules. It also provides tools to view data and other text files, to view tables and thematic maps created by CS Pro, to convert IMPS and ISSA data dictionaries to and from CS Pro, and to convert ERSI shape files (maps) to CS Pro map files. The software was developed jointly by the U.S. Census Bureau, Macro International, and Serpro, SA, with major funding from the U.S. Agency for International Development.

To collect the data from the field, a team of enumerators was first trained on how the questions were to be framed in the simplest ways possible for ease of understanding by the respondents so as to avoid occurrences of non-response errors. They were also trained on how to probe the respondents to provide relevant information that would best answer the objectives of the study. All these activities called in for a pilot survey on a small sample size to evaluate the suitability of the data collection tool and on the efficiency of the enumerators.

On approval of the effectiveness of the data collection mechanisms, the enumerators were given a go ahead to collect data across the survey sites and these datasets had to be checked and verified by the supervisor as soon as they were brought from the field. The reason for this was to ensure that any necessary clarifications could be made with the enumerators whilst their minds were still fresh about the nature of the data submitted. The data were immediately entered into the computer system. The data were then backed up in an external computer hard drive and in USB disks to minimize chances of data loss in case of computer break down.

Any discrepancies identified were verified from the questionnaires in hard copies and the right values keyed in. The data was further exported to excel sheet and SPSS and filtered into column by column to identify any strange or odd values. In SPSS, preliminary cross tabulations were done to check for unexpected values and if any was identified, corrections were done in the original master copy. An audit trail (data log) was kept to track changes made in the data. Once the data was well checked and approved to be error free, the whole set was exported to SPSS statistical software in readiness for analysis. "The final archiving of the field typology data was done with a metadata fully describing the data for ease of access and utilization by future researchers."

3.5 Data analysis

The data collected from the field typology survey was subjected to analysis using SPSS statistical package software where an overall syntax was developed to ease site analyses in the area under study. Data exploration was first done to identify the pattern and shape of the continuous variables and see where most values clustered. Diagnostic tests were also performed on continuous variables (Land size) with the production of residual plots to see whether the data complied with the assumptions of normality, equal variance (homoscedacity), independence of the groups and linearity. This exploration was equally important in identifying any outliers in the data that would otherwise make it violate the assumptions or otherwise exaggerate the estimates of the standard error. For the qualitative variables such as fertilizer use, frequency tables were used to summarize the results.

Fertilizer use was calculated from the amounts as reported by the farmer in kilograms applied to each field and this was expressed as application rates (kg ha^{-1}) by considering the area of each particular field.

Farmers indicated quantities of the harvest in kilograms or bags (90 kg of maize). Many of the values in kg given to local units were taken from previous work in the region, especially for food like maize or cassava (Impactlite 2012). Crop parameters such as dry matter content from own assessment and measurement.

In order to determine the dry matter content (DMC, %) of the various crop products, the dry matter weight (DW_{out}) of the outputs from the field that is maize was calculated. The dry matter content of the maize was calculated by dry weight of the maize over an average harvest index divided by the field harvest area; this was according to the following equation:

$$\text{TBY} \ ({}^{t}\text{DM}^{\text{ha-1}}) = [(\text{DW}_{\text{out}}/\text{HI})\ \text{i} + (\text{DW}_{\text{out}}/\text{HI})\ 2 + \ldots\ (\text{DWOU}/\text{HI})\ \text{n}]\ /\text{Field area}$$

Where DW_{out} and HI are the dry matter harvested of the crop grown in a particular field. The yield (${}^{t}\text{DM}^{\text{ha-1}}$) of each individual crop was calculated as the DW_{out} divided by the field area times an estimation of the fraction of total area shared by each particular crop in the field.

Logistics regression models were developed to explain maize yield variability at different field types by including different fertilizer types.

3.6 Developing a field typology

A field typology was developed by considering previous approaches (e.g. Smaling et al., 1997; Mapfumo and Giller, 2001) and by studying the results of the resource flow maps, adopting a terminology that can describe the variability found at the different field types. This typology aimed at discriminating resource allocation patterns and internal (within farm) nutrient flows that may have a short and/or long-term effect on the development of soil fertility gradients. In principle, land use and distance from the house were considered as the main criteria for this typology, which worked acceptably for most fields. Other criteria for classifying fields included type and number of crops that are grown, number of subplots, type of inputs used, soil management activities and sequential order within the farm.

CHAPTER FOUR

ANALYSIS RESULTS, PRESENTATION AND INTERPRETATION OF FINDINGS

4.1 Introduction

This chapter gives the analysis results of the data collected from the field typology survey. Three relevant field types were identified using the proposed methodology (see in Chapter 3: Developing a field typology).

4.2 Field Types

Different field types were identified within a farm, varying in production activities, resource allocation and management practices, as revealed by the farm transects (see example in Fig. 1). The field type 1 were the small fields around the homestead, used for a variety of crops sharing small pieces of land or intercropped (see example in Table 2 and Fig 2) which were often absent in the field type 3. The field types 1 were normally managed often and were the first fields to be planted and weeded, receiving kitchen wastes and the sweepings from the house.

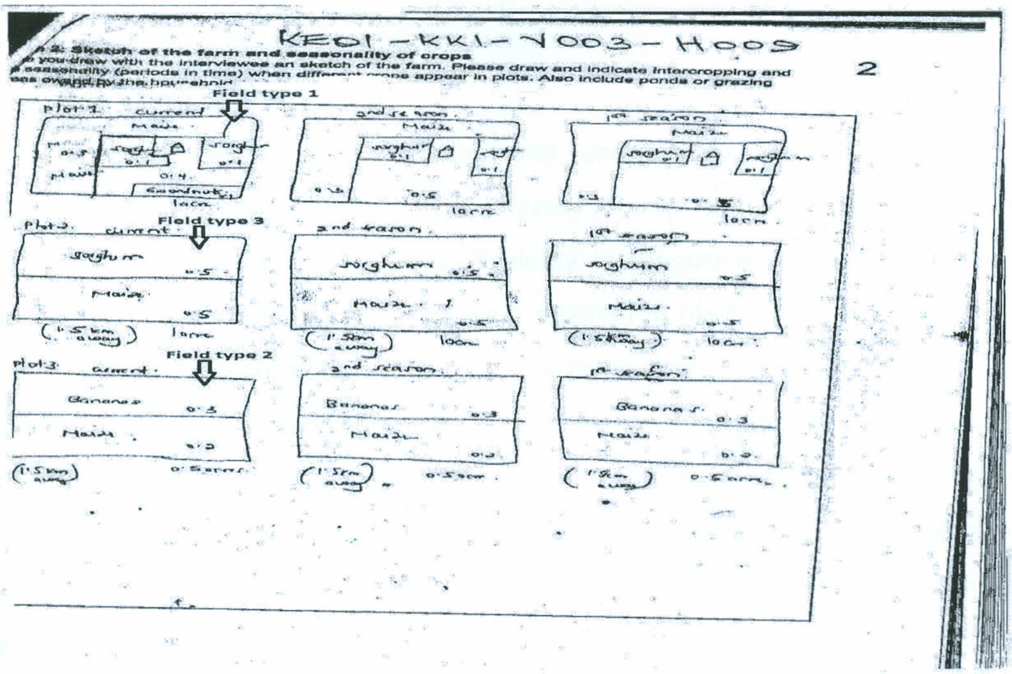


Fig.2. Example of farm sketch drawn during the visit to the farms in Lower Nyando (scanned from the original field sketch). The sketch was drawn in collaboration with farmers while walking in the farm and discussing with the aid of a semi-structured checklist

Table 2: Most frequently grown crops for the different field types, averaged over all field types at Nyando (*n*=407)

	Field Type		
Crops	**1**	**2**	**3**
maize	92	76	31
beans [a]	70	27	7
cassava	10	4	2
cowpeas	0	5	1
groundnuts	10	6	1
mangoes	0	1	0
millet	10	3	3
napier	12	2	0
bananas	28	4	0
vegetables [b]	8	6	1
soyabean	0	2	0
onions	2	2	0
sorghum	36	39	22
sugarcane	16	11	2
green grams [b]	8	7	2
sweet potatoes	5	7	18
tea	2	0	2
tomatoes	0	4	0

[a] Intercropped with other crops around homestead for domestic use.
[b] Intercropped for commercial purposes

The field type 2 and 3 were mid-distance and remote fields respectively in which more extensive crops were grown, though the diversity of crop types decreased with increasing distance from the homestead (Table 2). Typically, less inputs were applied (e.g. fertilisers, improved seeds) and the lowest yields were attained in the field type 3. The field type 3 were distant and the crop produce more prone to be stolen, particularly in areas of steep slopes. In this type of field, associated with poor quality land, farmers planted their woodlots or crops that are known to produce under conditions of poor soil fertility, such as sweet potatoes or sorghum. In field type 2 an intermediate management situation was found, strongly affected by the farm type. In wealthy farms they were managed in a similar way to the field type 1, though input use was less intense. Elements for field typology grouping are illustrated (Table 3).

Table 3: Elements of a functional typology for household categorization applied in Nyando by Mango et al. (2013)

Field type	Resource allocation [a] and production orientation	Main characteristics' [b]	Field type index[c]
1	High to medium resource allocation, mainly for self-subsistence	Receive both organic and inorganic fertilizer, next to homestead, soil erosion is controlled. Crop rotation this is evidence by different variety of crops grown on small pieces of land.	>10
2	Medium resource allocation	Receive both organic and inorganic fertilizer. Plot not next to homestead.	<10
3	Low resource allocation	Does not receive both organic and inorganic fertilizer. Plot with only single crop planted.	<4

[a] Referring to resource endowment of the field (i.e. type of fertilizer applied, management of the fields etc.)
[b] These refer to the distance of the field from the homestead in relation to resource allocation
[c] These refer to different score obtained from fertilizer use, management aspect, number of subplots per field etc.

4.3 Fertilizer use

There were large differences in the use of organic and inorganic fertilisers between the different field types (Table 4). Mineral fertilisers were used with varying intensities in the different field types within the field. The wealthiest field types 1 and 2 applied them in all field types, and relatively high rates were used in the remote fields of these case study farms. For field type 3 no fertilizer was used as they were distant field and crops were deemed to be stolen as farmers did not want to put more resources on field they are not sure to have harvest.

The use of organic fertilisers (Table 4) decreased strongly with the distance from the homestead and field types, and differed between the crops. Vegetable crops grown in the field type 1 received most of the organic resources, followed by the grain crops grown in the field type 2. Virtually only 1% organic resources were applied to field type 3, due to the extra effort required to transport coarse materials to distant parts of the farm and almost all application rates for manure were less than 20 kg ha–^{1}.

Table 4: Type of fertilizer used per field type (frequency %)

	Field Type		
Type of Fertilizer	**1**	**2**	**3**
DAP	68	30	0
Manure	42	35	1
CAN	2	3	0
NPK2323	4	1	0
Urea	0	1	0
MEA	2	0	0

4.4 Crop residue management

Four main ways of managing residues were identified and defined: residues used as fodder, burnt, composted and incorporated (Table 5). The first three are *assumed* to extract the residues from the field after harvest. Crop residues used as fodder are mainly transported to grazing sites where the animals are tethered. Grazing of standing crop residues on the field is not widely practiced in field type 1 due to the intensive double cropping (all fields are planted in both rainy seasons due to proximity to the homestead).Grazing of standing crop is in field type 3 during the second rains only in those fields that are left as fallow - a practice mainly adopted by the large sized farms of field type 3 and to a less extent by field type 2.

Table 5: Four ways of managing crop residues within the field types.

	Field Type		
	1	**2**	**3**
Grazed	50%	72%	76%
Burned	20%	50%	66%
Removed for compost	80%	29%	0%
Incorporated	34%	12%	2%

Burning of crop residues is practiced as a way of clearing the fields before planting mainly in field type 2 and 3. In field type 1 and to a less extent in field type 2, crop residues are taken from the field to a compost pile or compost pit, where they are mixed with animal manure and kitchen wastes and used as organic fertilisers into planting holes. This biomass transfer implies that residues harvested from a certain field may be used to fertilize other fields. Residues incorporation is the most common practice in field type 3 as well as in certain field type 2. In many cases, however, residues are not evenly incorporated on the fields but accumulated as trash lines along their boundaries (they are often used to demarcate field boundaries).

4.5 Regression model to explain maize yield variability

Logistic regression model was developed to explain maize yield variability including different types of fertilizer. The preliminary screening of the data revealed wide differences across categories of variables (particularly for univariate and multivariate analysis) this helped to build regression models with all the variables pooled. For that reason, multivariate analyses were carried out for manure and DAP. This approach also helped in identifying variables that better explained the variability found in the maize yield (see table 6 and 7). The introduction of the different types of variables for the construction of the models was done stepwise, to study their relative explanatory power. The number of points used in the regression analyses varied with the type of explanatory variables, with a different number of observations available.

The Z score formula used in the analysis is;

$$z = (x - \mu) / \sigma$$

The variables in the z-score formula are:

z = z-score

x = raw score or observation to be standardized

μ = mean of the population

σ = standard deviation of the population

UNIVARIABLE RESULTS:

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
GrainTha1	238	7.93	.00	7.93	1.2409	1.33261
Valid N (listwise)	238					

Table 6: manure

```
Logistic regression                               Number of obs   =        213
                                                  LR chi2 (1)     =      20.26
                                                  Prob > chi2     =     0.0000
Log likelihood = -132.28024                       Pseudo R2       =     0.0711

------------------------------------------------------------------------------
    Manure_2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   GrainTha1 |    .506433   .0899943    -3.83   0.000     .3574881    .7174348
       _cons |   1.270161   .2718551     1.12   0.264     .8349767     1.93216
------------------------------------------------------------------------------
```

There was a significant decrease in graintha^{-1} for those using manure only OR (95% CI) = 0.50 (0.36 – 0.72).

```
Logistic regression                               Number of obs   =        212
                                                  LR chi2 (1)      =      20.82
                                                  Prob > chi2     =     0.0000
Log likelihood = -131.50525                       Pseudo R2       =     0.0734

------------------------------------------------------------------------------
    Manure_2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Highest yield|   .9982515   .0004679    -3.73   0.000     .9973348    .9991691
       _cons |   1.250605   .2638206     1.06   0.289     .8270956    1.890971
------------------------------------------------------------------------------
```

There was a significant decrease in highest yield for those using manure only OR (95% CI) = 0.998 (0.997 – 0.999).

Table 7: DAP

```
Logistic regression                               Number of obs   =        213
                                                  LR chi2 (1)     =      23.66
                                                  Prob > chi2     =     0.0000
Log likelihood = -130.12172                       Pseudo R2       =     0.0833

------------------------------------------------------------------------------
       DAP_2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  FieldType_1 |   .1675287    .065442    -4.57   0.000     .0779082     .360243
       _cons |   2.636364   .9335551     2.74   0.006     1.317009    5.277422
------------------------------------------------------------------------------
```

Note: Field type 1 mostly had DAP compared to field type 2.

Table 8: DAP

```
Logistic regression                               Number of obs   =        213
                                                  LR chi2 (1)     =      13.00
                                                  Prob > chi2     =     0.0003
Log likelihood = -135.45175                       Pseudo R2       =     0.0458

------------------------------------------------------------------------------
       DAP_2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   GrainThal |   1.549817   .2048226     3.32   0.001     1.196152     2.00805
       _cons |   .3731932   .0788723    -4.66   0.000      .246625    .5647163
------------------------------------------------------------------------------
```

There was a significant increase in grainTha^{-1} for those using DAP OR (95% CI) = 1.54 (1.20 – 2.01).

```
DAP highest yield
DAP

Logistic regression                               Number of obs   =        212
                                                  LR chi2 (1)     =      33.67
                                                  Prob > chi2     =     0.0000
Log likelihood = -124.16038                       Pseudo R2       =     0.1194

------------------------------------------------------------------------------
       DAP_2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Highest yield|   1.001996   .0004226     4.73   0.000     1.001168    1.002825
       _cons |   .2574203   .0594963    -5.87   0.000     .1636473     .404927
------------------------------------------------------------------------------
```

There was a significant increase in highest yield for those using DAP OR (95% CI) = 1.002 (1.001 – 1.003).

MULTIVARIABLE RESULTS:

Table 9: DAP

```
Logistic regression                               Number of obs   =        212
                                                  LR chi2 (4)     =      54.68
                                                  Prob > chi2     =     0.0000
Log likelihood = -113.65663                       Pseudo R2       =     0.1939

-------------------------------------------------------------------------------
        DAP_2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
  FieldType_2 |   .1527299   .0675773    -4.25   0.000     .0641651    .3635375
    GrainThal |    1.36975   .2823526     1.53   0.127     .9144908     2.05165
Highest yield |   1.001468   .0005769     2.55   0.011     1.000338    1.002599
       Aream2 |   .9999873   .0000332    -0.38   0.701     .9999223    1.000052
        _cons |   1.079818   .5340349     0.16   0.877     .4096193    2.846563
-------------------------------------------------------------------------------
```

In multivariate model there was a significant increase in highest yield for those using DAP OR (95% CI) = 1.001 (1.000 – 1.003) adjusting for area and field type.

There was also an increase in grainTha^{-1} however without significance, OR (95% CI) = 1.37 (0.91 – 2.05).

Table 10: MANURE

```
Logistic regression                               Number of obs   =        212
                                                  LR chi2 (4)     =      28.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -127.54011                       Pseudo R2       =     0.1013

------------------------------------------------------------------------------
    Manure_2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 FieldType_2 |   1.35406   .5691251     0.72   0.471     .5941136    3.086075
   GrainTha1 |  .6240316   .1464002    -2.01   0.044     .3940138    .9883295
Highest yield |  .9987823   .0006412    -1.90   0.058     .9975264     1.00004
      Aream2 |   1.00001   .0000307     0.34   0.735     .9999503    1.000071
        cons |  1.213834   .5785368     0.41   0.684     .4769359    3.089288
------------------------------------------------------------------------------
```

In multivariate model there was a decrease in highest yield for those using manure however without significance OR (95% CI) = 0.999 (0.997 – 1.000) adjusting for area and field type. There was a significant decrease in grainTha-1, OR (95% CI) = 0.62(0.39 – 0.99).

4.6 Field type variability

Analysis of variance was developed to explain field type variability including different types of variables. There is significant differences in the subplots between the field types (see table 8).

Table 11: Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Field Type	238	2	1	3	1.94	.520
Valid N (listwise)	238					

Table 12: Subplots

Field Type

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	59.831	10	5.983	28.997	.000
Within Groups	85.836	416	.206		
Total	145.667	426			

There is significant difference between the subplots in the field types. This difference is between the total numbers of subplot per each field type.

Further analysis of variance was performed between field type and woody cover, erosion, graze plot and burned crop residues. From the results below in field type and woody cover there is significant between field type and woody cover as the p value was set as $p \leq 0.5$ to be significant. For the field type against erosion, grazed and burned they were not statistically significant to the field type as shown in the results.

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Field Type * Woody cover	427	100.0%	0	0.0%	427	100.0%
Field Type * Erosion	354	82.9%	73	17.1%	427	100.0%
Field Type * Grazed	427	100.0%	0	0.0%	427	100.0%
Field Type * Burned	427	100.0%	0	0.0%	427	100.0%

Field Type * Woody cover

Field Type

Woody cover	Mean	N	Std. Deviation	Minimum	Maximum
0	2.35	71	.537	1	3
1	2.08	356	.584	1	3
Total	2.12	427	.585	1	3

ANOVA Table

			Sum of Squares	df	Mean Square
Field Type * Woody cover	Between Groups	(Combined)	4.518	1	4.518
	Within Groups		141.149	425	.332
	Total		145.667	426	

ANOVA Table

			F	Sig.
Field Type * Woody cover	Between Groups	(Combined)	13.604	.000
	Within Groups			
	Total			

Field Type * Erosion

Field Type

Erosion	Mean	N	Std. Deviation	Minimum	Maximum
0	1.99	86	.623	1	3
1	2.06	268	.533	1	3
Total	2.05	354	.556	1	3

ANOVA Table

			Sum of Squares	df	Mean Square
Field Type * Erosion	Between Groups	(Combined)	.367	1	.367
	Within Groups		108.910	352	.309
	Total		109.277	353	

ANOVA Table

			F	Sig.
Field Type * Erosion	Between Groups	(Combined)	1.186	.277
	Within Groups			
	Total			

Field Type * Grazed

Field Type

Grazed	Mean	N	Std. Deviation	Minimum	Maximum
0	2.13	246	.603	1	3
1	2.12	181	.561	1	3
Total	2.12	427	.585	1	3

ANOVA Table

			Sum of Squares	df	Mean Square
Field Type * Grazed	Between Groups	(Combined)	.010	1	.010
	Within Groups		145.657	425	.343
	Total		145.667	426	

ANOVA Table

			F	Sig.
Field Type * Grazed	Between Groups	(Combined)	.030	.862
	Within Groups			
	Total			

Field Type * Burned

Field Type

Burned	Mean	N	Std. Deviation	Minimum	Maximum
0	2.12	417	.588	1	3
1	2.20	10	.422	2	3
Total	2.12	427	.585	1	3

ANOVA Table

			Sum of Squares	df	Mean Square
Field Type * Burned	Between Groups	(Combined)	.063	1	.063
	Within Groups		145.605	425	.343
	Total		145.667	426	

ANOVA Table

			F	Sig.
Field Type * Burned	Between Groups	(Combined)	.183	.669
	Within Groups			
	Total			

CHAPTER FIVE

SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMENDATIONS

4.1 Introduction

Widely different approaches have been followed so far to assess the origin, the magnitude and the importance of soil fertility gradients. This chapter summarises and discusses them, inter-linking their results. Some final considerations are made about the methods used, their practicality and performance.

4.2 Discussion of findings

Fields within a farm were initially classified according to their relative position with respect to the homestead. This criterion, however, was not objective enough to distinguish management practices and was therefore complemented with a more general conceptualization. In principle, the distance from the homestead affected the allocation of production activities and resources when labour was scarce and/or when a biophysical gradient existed within the farm, affecting the distribution of production units (i.e. when the house was built in the upper part of a topographic gradient and the farthest fields were in a valley bottom).

A broader concept to classify fields is that in which the difficulty to access to or to implement crop husbandry practices in the fields is considered. Extremely remote are difficult to plough or to weed, while the valley bottoms below them are difficult to access carrying certain inputs like cattle manure. In areas of steep topography, Tabaita, sloping fields may also be near the homestead. Closer fields which are field type 1 were of good quality when they were next to the house.

4.3 Conclusions and recommendations

The integrated methodological approach followed in this work helped increasing the understanding of the management aspects of the household that affect the soil fertility gradients. By considering resource allocation in the characterization of within-farm variability, it was possible to identify resource allocation patterns in relation to soil fertility gradients. Such an approach could also be considered while targeting soil fertility management strategies and fine-tuning decision aids for resource allocation in smallholder farms.

Farmers manage their fields according to their distance from the homestead, so that the soil fertility gradients are in fact also management intensity gradients. Additionally, the heterogeneity in agricultural productivity, in terms of the intensity of nutrient depletion, and the allocation of resources and production activities to the different fields within the farm varied in magnitude between farm types. In areas of high subplots within the farm it appears to override the inherent biophysical properties in determining the pattern of resource allocation and the magnitude of the soil fertility gradients within a farm. Conversely, in areas of high variability in the inherent biophysical background, perceived land quality determines the resource allocation pattern emerging from farmers' management decisions. Since scarce resources and investments are preferably allocated to land units next to homestead, such a pattern results in increased within-farm soil fertility gradients.

Management decisions at farm scale, which are affected by both biophysical and socio-economic factors, have an important impact on the resulting soil fertility. To understand opportunities for sustainable intensification, an approach is needed to analyses the combined effect of management and current soil fertility on farm productivity, raising the scale of analysis to the livelihood level. Some operational decisions are often made on a day-by-day basis, contrasting with the strategic questions on sustainability of the farming systems in view of long-term changes of biophysical and socio-economic conditions.
Thus, there is a clear need to include the dynamics of the farming systems in the analysis of tradeoffs between factors affecting soil fertility management at multiple spatio-temporal scales.

4.2 Lessons learnt from the study

Data management is not just about having the data brought from the field and analyzed. A lot more goes into quality control where thorough cleaning must be done so that valid results and conclusions can be drawn from the collected data. Some of the quality control points that should be enhanced comprise of: proper training of the enumerators, having competent supervisors to do good supervisory work and make constant follow ups. The data collection tool (questionnaire) should be made simple for ease of understanding by the respondent, short and concise to capture the key variables of interest. As this was evident as enumerators did not understand the questionnaire well when administering some of the questions, more so when capturing GPS coordinates of the plots. Data on crop type and fertilizer use was captured well as was shown during analysis. Also, validation rules should be set up in the data entry system to limit erroneous entries and double data entry system put in place as a quality control mechanism. Proper choice of the data entry software is of paramount consideration as this will help in checking errors and reduce the time spent on data entry. In the whole process, constant backstopping support should be provided to the entire team by trained personnel, preferably, a research method professional who has better understanding of the area under study and has the technical expertise in data management.

Organizations and research institutions need research methods support and capacity building should be set as a major priority. Majority of those who participate in research projects lack the technical skills and expertise in designing data collection instruments, management of the data and data analysis.

REFERENCES

Arondel, P. and Depoutot, R. (1998). *Overview of Quality Issues when Dealing with Socio-Economic Products in the International Environment*. Paper presented at the 30th ASU meeting

Bamwerinde, W., Bashaasha, B., Ssembajjwe, W. and Place,F. (2006). *The Puzzle of Idle Land in the Densely Populated Kigezi Highlands of Southwestern Uganda*. *International Journal for Environment and Development* 3 (1): 1-13.Benhin, J. 2006 Climate Change and South African Agriculture: Impacts and Adaptation Options, CEEPA Discussion Paper No. 21, University of Pretoria, South Africa.

Bennet,S.M. (2001a). *Master data management.* London: Knowledge integrity

Bennett, S. M. (2001b). *Data Management for surveys and Trials.A practical primer using Epi data.* London: Epi Data Association

Bland,D.N.(1998).*Some Problems of Statistical Consultancy.*Available in: Journal of the Royal Statistical Society Vol. 133, No. 2, pp. 139-165: Blackwell Publishing for the Royal Statistical Society: UK. [Accessed on 10th February 2011 at 1500hours]

Brackstone, G. W. (1999a). *Statistics for Social Data analysis.* Itasca, Ill: Peacock.

Brackstone, G. (1999b). *Managing Data Quality in a Statistical Agency.* In: *Survey Methodology*, 25, pp. 139-149

CTA,(1997). *Strengthening Biometry and Agricultural Research*. Summary and recommendations of a CTA/University of Hohenhein workshop, 1996. CTA (Technical Centre for Agricultural and Rural Cooperation), Wageningen, The Netherlands

Collins, M. and Sykes,W.(1999). *The Problems of non-coverage and unlisted numbers in telephone interviews.* London Press: UK

David,C.(1999).*Data Management Systems in Surveys.* London: Defense Technical Information Center

Deressa, T. (2006).*Measuring the Economic Impact of Climate Change on Ethiopian Agriculture*: Ricardian Approach, CEEPA Discussion Paper No. 25, University of Pretoria, South Africa

Depoutot,A.W. (1998). *Harmonising Data in Official Statistics;Development Procedures and Data Quality.* Newburry Park: Manfred Ehling. Available at : http://www.gao.gov/special.pubs/pe10111.pdf. (n.d.)

Essortment,(2002).*Job training methods.* Available at: http://www.essortment.com/all/jobstraining_rshn.htm, [accessed 18/10/2010]

Grove, S.K. (2009).*The Practice of Nursing Research; conduct, critique and utilization.* St. Louis Missouri: USA

Hoagli,D.C.(2000).*Review of Exploring Data Tables, Trends and shapes.* In: *Trichonometrics, 28,399-400.* Available at: http://pubs.amstat.org/doi/abs/10.1198/tech.2000 (n.d)
http://www.worldagroforestry.org/ [Accessed on 10th November 2010]
http://www.ccafs.cgiar.org/[Accessed on 1st January 2011]

IFRS (2000): Mandatory IFRS Reporting around the World: Early Evidence on the Economic Consequences. Available at: http://onlinelibrary.wiley.com.[Accessed on 12th/02/2011 at 1430hours]

Isenberg, J., Bennet, C. and David, G. (1996).*The RICCI Flow, Techniques and Applications.* Rhodes Island: USA

Johanson, R. & Saint, W. (2007). *Cultivating knowledge and skills to grow African Agriculture*. Report to the World Bank, June 2007.

Khamis, H.J. (1994).*Criteria of Evaluating the Performance of Statistical Consultants.* In: The American Statistician, 35,216-269.Dayton: USA

Lyberg, L. and Biemer, P. (1994).*Introduction to Survey Quality.* Hoboken WJ: USA

Lobell, D. B., et al. (2008) .*Prioritizing climate change adaptation needs for food security in 2030.* Science 319.5863 ;607-10.

Massoud. (2009). *Good data Management Practices.* Manchester: London press.

McPherson (2001). *Preparing Data for Analysis.* New York: Wiley

Muraya, P. and Chege, G. W. (2002). *Research Data Management.* Nairobi: McMilan press.

Nelson, W.R. and R. Hendrickson (2009). *User Manual for STATPAC a General Purpose Program for Data Analysis and for Fitting Statistical Models to data*. Schenectady, NY 12345 US

Rangecroft, M. and Wallace, W. (1998). *Teaching Statistical Consulting Before Statistical Methodology* .John Wiley and sons press: UK

Ruberg, S.J. (1998). *The role of consultancy in university education and professional training in statistics*. In L. Brunelli & G. Cicchitelli (Eds.), Proceedings of the First Scientific Meeting of the International Association for Statistical Education (pp. 285-297). Perugia, Italy: International Association for Statistical Education.

Savery, J.R. (2006).*Overview of Problem-based Learning: Definitions and Distinctions.* The Interdisciplinary Journal of Problem-based Learning, volume 1.

Stegman, C.E. (1995).*Statistical consulting in the University: A Faculty member's Perspective.*In: Journal of Education and Behavoural Statistics. Available at: jeb.sagepub.com/content/10/3/269.shop.[Accessed on 3rd February 2011 at 0800 hours]

Tabachnick, B. and Linda, S. (2007).*Using Multivariate Statistics.* Boston: London Press

Trochim, C., Richard, P. M. and Brandie K. (2008). *Measuring Collaboration and Transdisciplinary Integration in Team Science.* Elsivier: USA

VanBelle, B. and Llyod, F. (1982).*A methodology for the Health Sciences.* Wiley: New York.Available at:http://www.getcited.org/pub.[Accessed on 20/12/2010 at 1540 hours]

Ware, J.E. (2008). *Health Survey. Manual and interpretation guide*. Chicago Press: USA Available at: http://www.PetrovskiyYY/article. [Accessed on 10th/01/2010 at 0825hours]

Weiger, K.E. (2002).*Peer Reviews in Software: A Practical Guide*. Cutter IT Journal. Available at: http://www.processimpact.com/articles/seven_truths.html.[accessed on 17/10/2010 at 0300hours]

Zan, H., Chen, H. and Zeng, D. (1995). *Applying Associative Retrieval Techniques to alleviate the Sparsity problem in collaborative filtering.* ACM New York: USA