Department of Mathematics and Statistics

University of Kent at Canterbury, England.

An Investigation of Sample Survey Methodologies, Including the Application of Remote Sensing, for the Production of Agricultural Statistics in Tanzania

by

ABEIDI SIMBA MUSSA

A Thesis Submitted to the University of Kent at Canterbury

for the Degree of Doctor of Philosophy in Statistics

1992.

In Memory of

my parents,

(May God have mercy on them).

ACKNOWLEDGEMENTS

I thank Allah Almighty for keeping me alive, healthy and granting me patience and perseverance to the fulfilment of this study.

My greatest debt of gratitude is to my supervisor, Geoff Clarke, who offered me his precious time and guidance throughout the period of this study. His cordial and friendly relationship contributed immensely to the completion of the study.

I am sincerely grateful to Andrew Rutherford for introducing me to remote sensing technology, to Idris Kikula for providing me with basic reading material on remote sensing and to Francis Sichona for his enormous assistance and encouragement during my field work.

I am also grateful to officers of the Regional Remote Sensing Centre, Nairobi, John Baraza and Ibrahim Mohamed for kindly helping me acquiring satellite images and to officers of the Bureau of Statistics, Dar es salaam, for providing me the agricultural sample data for 1987/88.

This study is deeply indebted to the villagers of Igawilo, Sinoni, Ngarash and Nyankumbu in Tanzania. I would like to appreciate their kind hospitality, enthusiastic and earnest responses to my questions during the survey.

I would also like to profoundly thank my brother Abdulrahman and my friends, especially Slim and Kaimu, for their relentless efforts in forwarding missing information to me during the analysis of this study.

My heartfelt gratitudes to my wife Biubwa *d/o* Hamza Bachu and my children Yusuf and Fatuma who patiently and painfully suffered my long absence from home, and to other members of my family for their moral support.

I deeply appreciate the excellent cooperation I received from the staff of both the Department of Mathematics and Statistics and the Applied Statistics Research Unit. Particular appreciation to Mavis Swain, Edwina Bell and Gerry Latawiec for their untiring attention and assistance offered to me.

Grateful thanks are due to different authors whose books or articles are referenced in the text, for the excellent material they have provided me with. Also I am thankful to the librarian and library staff of the University of Kent at Canterbury for the valuable service they have rendered to me.

Finally, but not the least, I am grateful for the financial support by my University of Dar es salaam and British Council, without whom the study would not have been possible.

ABSTRACT

One of the many applications of remote sensing data, acquired by earth-orbitting resource satellites on a regular basis, is the estimation of crop areas in agricultural statistics. The evaluation of crop area statistics for traditional farming in sub-Sahara African developing countries using the rapidly advancing remote sensing data is aptly under consideration.

Because of operational problems in the developing countries in acquiring and exploiting images of all areas of interest, an approach has been proposed that combines remote sensing data, Landsat and SPOT, and ground sampling data in a flexible manner depending on materials, equipment available, expertise and experience of personnel involved. Under the prevailing circumstances of rather limited high-tech resources in the region, the visual interpretation of imagery has been used and emphasized.

A ratio-type estimator for the mean has been proposed using information on auxiliary variables for improvement at the estimation stage. The exact as well as asymptotic bias and variances of the estimator have been obtained. Comparison on infinite and finite populations with several other ratio-type estimators is done on the properties of bias and efficiency. The simplicity and suitability of the estimators in practical situations were also taken into consideration. The performances of the estimators have been examined using both hypothetical values and the data collected in 1990 in the field work performed during the study period.

Attention is turned to various methods of attaining yield rate. A crop-cut method and farmer's statements are discussed and tests performed on them with the Tanzanian agricultural sample survey data of 1987/88.

Finally a production estimate, that could easily be extrapolated over larger area, is obtained from the product of area estimate and yield rate as a result of integrating satellite data with the field ground sampling and farmers' interview data respectively.

SOME DEFINITIONS

Band: A selection of wavelengths.

Colour Composite: A colour image produced by assigning a colour to each of a number of images of a scene and optically or digitally superimposing the result.

Electromagnetic radiations: Energy propagated through space or through material media in the form of an advancing interaction between electric and magnetic fields.

Electromagnetic spectrum: The ordered array of electromagnetic radiation extending from short cosmic waves to long radio waves.

False Colour: The use of one colour to represent another.

Field: A piece of land in a parcel separated from the rest of the parcel by easily recognizable demarcation lines, such as paths, cadastral boundaries and/or hedges.

Holding: An economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form, or size.

Household: Briefly, a group of persons living together who make common provision for food or other essentials. For one person only, then it is a one-person household.

Parcel: Any piece of land of a holding entirely surrounded by other land, water, road, forest *etc.*, not forming part of this holding.

Pixel: A picture element having both spatial and spectral aspects.

Plot: A part or whole of a field on which a specific crop or crop mixture is cultivated.

Table of Contents

Chapter one

General introduction

1.1 Introduction: sub-Sahara African food crisis

Africa is a vast continent that accounts for approximately 28% of the earth's land surface (as compared with 17% for Europe and 20% for Asia). Sub-Sahara includes 46 out of 55 African countries (see Figure 1.1). At least 20, or half of these countries are experiencing emergency food supply (Standard 1990). Figure 1.2 indicates the seriousness of the situation.

In general the countries of Africa are trapped between a population growth, an annual increase of up to 3%, and agricultural production of less than 2% annual increase. The Food and Agriculture Organization (FAO) announced that among all developing countries of the world, Africa has the highest number of undernourished (Standard 1990). FAO estimated that 86 million people in sub-Saharan Africa were undernourished during 1969-71, representing 32.6% of the total population. During 1979-81, 110 million people (30.6%) were undernourished, and by 1983-85 approximately 142 million, more than one out of every three Africans (35.2%) were undernourished. The result is, of course, a growing need to import food. Sandbrook (1985) cautioned that by the year 2000, tropical Africa will probably import one-third of its food requirements.

However, agriculture remains the main occupation of the people, therefore it is the backbone of their economy. It is only through greater food production and keen environmental protection that the present dismal condition on agricultural growth can be overcome. Therefore, it is important to ensure improvement in the quality of data collection in this sector. As such, collection of data poses a problem because small land holding is the trend in the continent.

SUB-SAHARA AFRICAN DEVELOPING COUNTRIES				
ANGOLA	CONGO	KENYA	NIGER	TANZANIA
BENIN	COTE D'IVORE	LESOTHO	NIGERIA	TOGO
BOTSWANA	DJIBOUTI	LIBERIA	RWANDA	UGANDA
BURKINA FASO	EQT. GUINEA	MADAGASCAR	SAO T. & PR.	ZAIRE
BURUNDI	ETHIOPIA	MALAWI	SENEGAL	ZAMBIA
CAMEROON	GABON	MALI	SEYCHELES	ZIMBABWE
CAPE VERDE	GAMBIA	MAURITANIA	SIERRA LEONE	
C.A.R.	GHANA	MAURITIUS	SOMALIA	
CHAD	GUINEA	MOZAMBIQUE	SUDAN	
COMOROS	GUINEA BISSAU	NAMIBIA	SWAZILAND	

C.A.R. (Central African Republic)

EQT. (Equatorial)

SAO T. & PR. (Sao Tome and Principe)

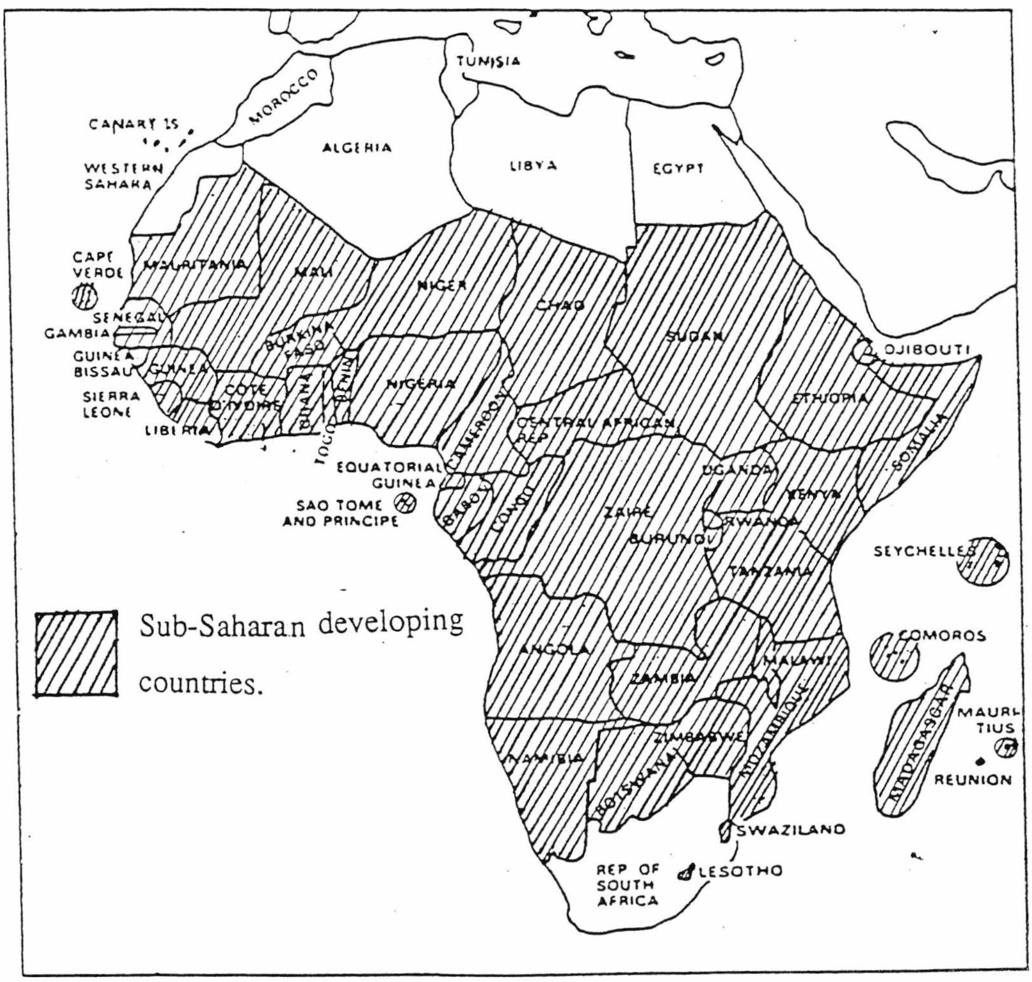
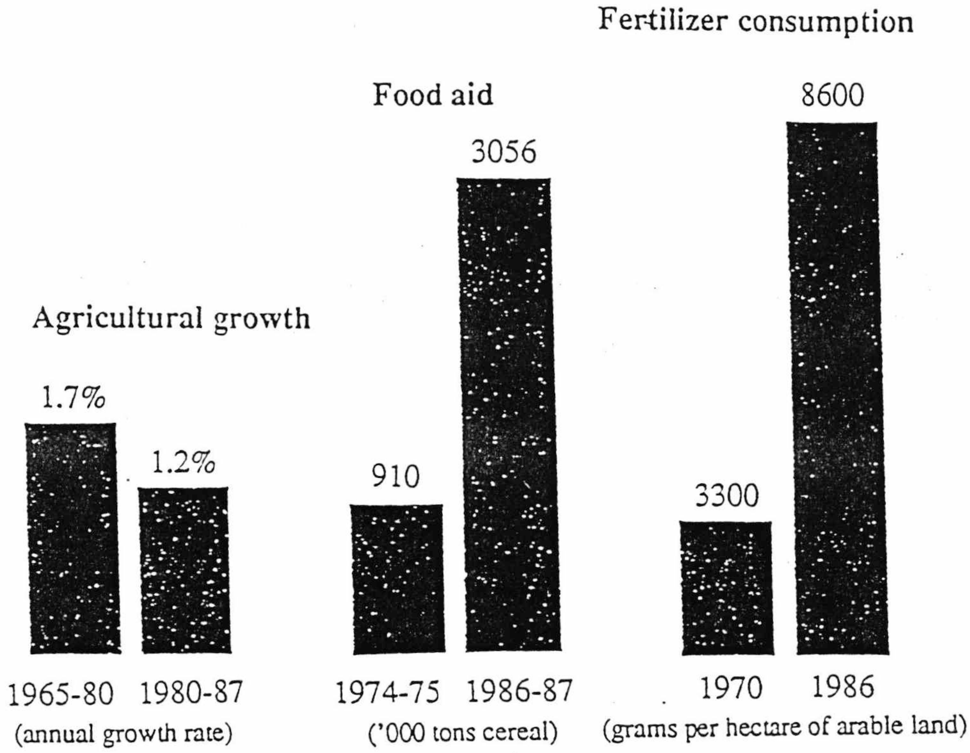


Figure 1.1: Sub-Saharan African developing countries



Average daily calorie intake remains at its 1965 level of 2100

Figure 1.2: Sub-Saharan Africa's food crisis

Source: FAO says it has solution to Africa's hunger, malnutrition (adapted from Standard 1990)

Two-thirds of all land holdings in Africa are under two hectares (*ha*) and nearly 96% are fewer than 8 *ha* (Daily News 1989). It is not that there is lack of land but the means for land utilization are limited. The main tool used is still the hoe, mechanization has not yet reached a majority of farmers in most countries, or else remains confined to donkey/oxen-powered carts and scratch ploughs (Bartholome 1991). Enormous effort is therefore needed on agricultural research in these countries for improved farming. When improved farming techniques like seed, irrigation, fertilizer, machinery, insecticide *etc.*, are recommended and made available, proper land use becomes important. On the other hand, agricultural planning is necessary for development, and the requirement of the statistical information of a country is essential for efficient planning.

In sub-Sahara African developing countries there is likely to be an inventory of crop area and production with some degree of accuracy for those crops which are marketed entirely through a few outlets or where cultivation is limited to a few large estates (cotton, tea, cocoa, rubber *etc.).* Normally office records of the marketing organization or the estates would supply such information. For most crops, especially staple food crops, the situation is different. The crops are grown by millions of small farmers on scattered holdings. A complete enumeration, even if it were technically feasible, would cost a prohibitive amount. In order to offset such costs, well planned sample surveys are needed.

Sample surveys on intention to planting, and estimates of area, help farmers plan their plantings, serve as direct measures of land utilization, and are prime indicators of the future demand for various farm supplies and labour. In Africa, generally these surveys are the responsibility of the agricultural ministries. The population census provides the frame for sampling in Malawi, Tanzania, Ethiopia, Morocco and Zaire. Pre-test, pilot and post enumeration surveys have been used to improve the quality of the data in Ghana, Malawi and Ethiopia. Tanzania and Ethiopia have used sampling in their surveys of the traditional or peasant farms, and complete enumeration of the commercial or state farms. Farm management surveys are also conducted to devise methods of improving subsistence farming.

This study concentrates on area and production estimates of crops. The estimates are essential for agricultural programs not only in Africa but throughout the world, because they provide basic data for research, program planning and administration. FAO is actively engaged in formulating improved methods and in promoting them in developing countries.

1.2 Agriculture and remote sensing

While agricultural planning personnel are devising information gathering and processing schemes which make best use of all sources of data, remote sensing technology is making rapid advances for developing and/or improving agricultural statistics. The increasing economic and social requirements for more accurate and timely information regarding the supply of natural resources and status of land cover, has assured the contribution of remote sensing in data collection and the inventory process in general. According to Kikula (1987), the new generation of the earth resources satellites seemed to have made possible more detailed analyses and monitoring of the environment. Apparently remote sensing systems capability has been exaggerated or over simplified (Taylor 1985). While it is true to speak of their high technical advancement, it is evident that these innovations cannot substitute totally for conventional methods of acquiring resource information, especially from the developing countries. Latham and Som (1991) emphasised that approaches are necessary that will combine the best aspects of remote sensing and ground based survey techniques. Colwell (1977) stressed that remote sensing, especially satellite remote sensing, need not and generally cannot provide all of the information usually required for crop surveys. Other sources of information, the author emphasised, are necessary and should be developed to complement the remote sensing techniques. Indeed, if these other sources of information can be established with better methodology and reliability, intergration with satellite data can be of much help in general optimal planning of resources utilization.

1.3 Aims and objectives

As it is a difficult and time-consuming task to estimate land cover of several square kilometres (*sq km*), remote sensing data from earth-monitoring sensors like Landsat and SPOT not only make the task relatively easy, but provide the potential for improving such estimates over extensive areas. By providing totals of large areas, these data eliminate sampling error associated with the estimated parameter. The study therefore intends to implement the use of remote sensing data as an auxiliary variable and attempts to find out in what circumstances it will be appropriate to use this auxiliary information for estimating a particular resource parameter. Also the study will attempt to assess the capabilities and limitations of remote sensing in practical situations, when it comes to be applied in sub-Sahara African developing countries. Generally, in an endeavour to attain an agricultural data collection system that can produce reliable and objective information quickly, the study will make use of people and the limited resources in the relevant countries.

Hence the main aims of the study are:

- to explore the possibility of effective integration of remote sensing data with some ground sought data in the estimation of crop area,
- to develop a statistical procedure based on these data sets for attaining a more efficient and relatively low-biased statistic,
- to establish a reliable procedure in the process of achieving a trusted and appropriate yield rate of a crop of interest

and

- to calculate production values accurately and quickly.

Thus, the study will use data, collected by the author, in a field survey in sampled villages from selected regions of Tanzania, one of the sub-Saharan countries in Africa. The selected regions represent about 20% of the whole country's land.

1.3.1 Plan for collection of data

The initial plan for the field work, which was expected to take about 32 weeks, consisted of the following programme with approximate numbers of weeks for each item shown in brackets.

1. Preparation of research materials, collection of funds, check on transport and routes. (4 weeks)
2. A trip to Nairobi, Kenya for the collection of satellite images. (6 weeks)
3. Discussion with officers at the Bureau of Statistics and Ministry of Agriculture. Final preparation for field work. (2 weeks)
4. Survey work in the selected villages. (16 weeks)
5. Initial data checking and validation. (4 weeks)

A random sample of 30 households per village was taken. Each holding area was measured. The FAO (1986) definitions of a holding and a household were closely followed. There were area measurements on the houses (household compound) and fields/plots (according to crops grown). Each holder was asked to give the production of their crops from the respective measured fields/plots in units which they are familiar with. That means farmers were asked to state the production in their local units or units of their understanding. Later the units were converted into standard units by appropriate conversion measures. Initially all crops were recorded but the analysis will concentrate on maize, beans, the combination of the two and the combination of maize with other crops. In conjuction with that, Landsat and/or SPOT data of the villages will be interpreted and analysed and finally will be combined with the collected field data to achieve the outlined aims and objectives.

1.3.2 Contents of this report

The study will review, in chapter two, different existing methods for data collection and the kind of data available. The chapter aims to compare the representativeness of data produced by different organisations in a country. Data produced by the Ministry of Agriculture and the Central Bureau of Statistics in Tanzania will be investigated.

Chapter three will introduce remote sensing and will descibe its linkage and usage with the earth surface. The chapter will outline the information flow from the satellite sensors to the observed characteristic on the earth to the production of data, referred to as remote sensing data in general and satellite data in particular. The chapter will also review some previous work on the application of remote sensing data in sub-Saharan Africa.

In chapter four, the application of remote sensing for providing information concerning agricultural cropped land in sub-Sahara African developing countries will be attempted. The chapter will also outline guidelines for visual interpretation and its merit for area estimation. Also the chapter will discuss the practical situation in the usage of satellite data in the villages where I performed the survey in particular, and the capabilities and limitations of its application by African users in general.

A ratio-type estimator will be proposed in chapter five. The model, based on its efficiency and bias, will be tested with some other models by using both hypothetical and real data that will be obtained from satellite and ground survey.

In chapter six attempts will be made, first to look into the relationship of reported and measured areas from the Agricultural Sample Survey (Agsasu), performed by Bureau of Statistics in 1987/88 and second to analyse the integrated data from physical measurement of crop areas at village level from the field work performed by the author in 1990, Mussa Field Survey (MFS) and the interpreted data from the satellite images.

Chapter seven will review the sources of error regarding yield as obtained from farmers and from the crop-cut method. The chapter will also compare yields reported by the farmers and the crop-cut method from the same fields/plots using Agsasu data. Finally villages' production will be calculated using area data from an integration of ground survey and satellite interpretation and yield data from the farmer's statements as obtained from the Mussa Field Survey.

Chapter eight will review the study results and give general conclusions. Also the chapter will attempt to suggest some measures for further improvements in the application of remote sensing as will be observed from the analysis.

Chapter two

Data collection: methods and organizations

2.1 Introduction

Area and production estimates for crops are an essential part of agricultural programmes throughout the world because they provide basic data for research, programme planning and administration. Surveys on intention to planting and estimates of area help farmers plan their plantings, serve as direct measures of land utilization, and are prime indicators of the future demand for various farm supplies and labour. The successive data needed annually include prospective plantings, actual plantings, area for harvest and actual harvested area and the respective production. It is of great importance to ensure improvement in the quality of data collected.

Quality of data depends on, among other things, precision, representativeness, validity and timeliness. Lack of these characteristics invites criticisms on the deficiencies of quality data which is often quite pronounced in developing countries (Eklof 1984, Casley and Lury 1985). A knowledge of agricultural data is considered as a priority for many countries in sub-Saharan Africa; essentially because the commodities produced are indispensable for the survival of consumers. But it is also one of the areas where, for a number of reasons, observation is particularly difficult. Agriculture is very spread out in spaces, it is also diffuse, in that it consists of a multitude of producers, often small, or even tiny. The organization of these producers among themselves is weak and in some instances non-existent. Governments recognize the great importance of reliable and timely food and agricultural statistics. Hence programmes must be set with aims geared towards raising nutritional levels and living standards. Therefore more effort is needed to improve the efficiency of agricultural statistics for adequate distribution of food and agricultural products, and to better the overall condition of the rural population. A methodology, in the survey techniques and sample designs, must therefore be aimed at for the benefit of the producers, who are also consumers, in the rural areas in order to achieve timely, objective and accurate data within the constrained resources available.

2.1.1 Accurate data

By accurate data we mean the estimate value generated is close to the actual value being estimated. Usually the closeness of an estimator and the true value is a measure of total error. In a case of sample survey, it is expressed as a difference between estimate and actual value, that is;

$$\text{Total error} = |\text{Estimate value} - \text{Actual value}|$$

$$= \text{Random error} + \text{Systematic error (bias)}$$

and

$$\text{Random error} = \text{Sampling error} + \text{Non-sampling error.}$$

Controlling these errors helps make data more useful and reliable.

Sampling error

Perhaps the only part of the error in sample surveys that is in the hands of a statistician, and is measurable and gives a guide to the minimum level of error present, is sampling error. It may be reduced considerably if the design is improved.

Non-sampling error

As for non-sampling error, Scott (1989) argued that there is no unified concept of non-sampling error and certainly no integrated theory comparable to sampling theory. He pointed out that the term covers an extremely diverse set of phenomena occurring at all stages of survey work and thought that the most useful basis for explaining non-sampling errors is to classify them by the stage of the survey at which they occur.

The following is one such classification.

Survey design

Questionnaire:	design, production
Sampling frame:	coverage – area level – household/dwelling level. rules of association – mistaken rules – mistaken application
Sampling:	design, application.
Non-response:	refusal, non-contact, loss of data, response from wrong unit.
Response error:	due to respondent, due to interviewer, other sources (language, proxy respondent *etc.*).
Coding:	by interviewer, by coder.
Data entry and editing:	manual, computer

Tabulation

Analysis and interpretation

Reproduction/publication.

In many cases there are possible techniques for measuring the impact of such errors and there are strategies for reducing them. The author further suggested that those strategies have to be tailored to each specific type of error and hinted that they should constitute the maintenance of constant spirit of self-criticism and scepticism, backed by a wide-ranging policy of checking and monitoring at all levels and a habit of meticulous attention to detail. The success in controlling non-sampling errors lies therefore in concentrating on concepts, procedures, training and measurement techniques deployed.

A balance between the need to eliminate non-sampling error and timely reporting of the results depends very much on experience in practical sample surveys.

Bias

Biases have different possible sources but can generally be classified into three main sources:-

1. Technical bias

   This is the type most often discussed by mathematical statisticians. Technical biases occur when the functional form of the estimator is such that the average over all possible samples is not equal to the true parameter value. Ratio and regression estimators are generally biased this way. But if attention is given to the usually known technical bias characteristics, this bias source should not present great difficulty, or at least its magnitude is assessed by standard or formulated estimators.

2. Bias due to measurement error

   Measurement error is one source of bias whose effects can be substantial. In fact, there is virtually no limit to the difficulties that can be brought about by measurement error (Williams 1978). The real difficulty is easy to understand. The measurement process should measure a value y but manages to give a different value. While the situation is simply explained and seems to be straight forward, it can be difficult in practice, because the errors can be introduced in subtle ways. Lack of instruments calibration is a major source. For example, no two 30*m* tapes are the same, the differences between them increase with use. So there should be regular checkings; tapes expand, stretch and gradations become scratched. In the same way compasses need to be checked regularly, spring balances used for weighing go wrong through dirt, rust and wear and tear. Few countries have a programme of regular checking, if they even check at all (Rutherford 1989, personal communication). Furthermore these errors can also be introduced by human observers when reading or recording a reading. Measurement error, concluded Williams, creates bad data which can be a very serious source of bias.

3. Selection bias

   This can occur when sample units are thought to have been drawn into the sample with one set of probabilities but actually are unknowingly drawn in with a different set. Such a difficulty can be associated with a failure to implement a sampling design properly or with the specific problem of non-response, and also with "out-of-date" frames and wrongly constructed frames.

These errors can be reduced to a great extent if the surveyor adheres to strictly organised procedures, training and measurement techniques.

2.1.2 Timely data

Agricultural data should be made available when they are needed. Therefore, timely data mean that information is reported when it is useful. Soon after the basic data are gathered, the relevant statistics should be available. So timely reporting implies that both the data collection and summarization are programmed to allow dissemination of relevant information against rigid deadlines with the limited manpower and budgetary resources available.

2.1.3 Objective data

Objective data imply that the results do not depend on vested interests of some group of individuals. A random procedure or a systematic one with an element of randomness should be followed to select those areas planned for data collection and the results should be considered only if this has been done. This should protect the published results from serious bias.

In sub-Sahara African developing countries, measurement of crop production with any degree of accuracy is only likely for those crops which are either marketed entirely through few outlets (marketing boards, national milling) or where cultivation is limited to few large estates where there is a tendency to have fairly good records on the size of the farm, inputs and yields. Moreover in developing countries large scale farming generally concentrates on the production of exportable cash products earning exchange of national currency (cotton, tea, cocoa, rubber *etc.*). For most crops, especially staple food crops, the situation is very different. Traditional farming produces most of the food that feeds the nation, but the statistics produced for this sector are unreliable. There is always contradiction among the various bodies and organizations which collect them.

The study intends to review some of the methodologies which are used to achieve the statistics with the main objective of forming a methodology that will produce timely and reliable statistics for traditional farming so that governments can take steps to deal with surpluses or can replenish the deficiencies in time. The proposed methodology should also be able to meet other objectives of area and production estimation of crops which include:-

- Knowledge of surface area and its contents: It is vital for proper planning and dissemination of agricultural inputs. For example fertilizer, ploughs, hoes *etc.* in addition to environmental protection.
- Land management: The rapidly expanding population, and the increasing demand for food, places strong stress on the use of land as well as mineral, vegetation and water resources.
- Reviewing the establishment of industries: Industrialization has added a substantial stress on the usage of the country's resources and environment during the past three decades, when most countries have gained their political independence.
- World community awareness: For trade purposes and for food security system; the global information and early warning system.

In order to achieve effective systematic programmes, developing countries are encouraged to set up regular sample survey procedures for agriculture (FAO 1982, Petricevic 1988).

2.2 Methods of data collection

There are several methods for assessing crop statistics through sample surveys. For the purpose of this study, I will concentrate on two major parts:

(i) Data collection on areas,

(ii) Data collection on yields.

2.2.1 Data collection on areas

Data collection on areas has always been a problem in developing countries because of shifting cultivation and lack of adequate knowledge and proper tools. Lack of titles to the land, for the majority of farmers, which would make them know their areas in some units, contributes greatly to the problem of data collection on areas. Several methods have been put into practice to change the situation. The methods include some described below.

2.2.1.1 Reports by administrative units

This method is based on regular reports for each administrative unit by extension workers, village chiefs or any other official person. It is a low cost and quick method which provides data of usually unknown reliability. To most of these people, collection of agricultural data can be an extra routine activity or can be performed on request.

The reports are based on personal judgement but sometimes can be strengthened from the farmers' opinions who normally use number of paces, amount of seeds sown, *etc*. The method is also an integral part of the eye-estimate method which is common in data collection not only on areas under cultivation but also on yield. The investigator normally works on arbitrary selected samples of farms. Personality may influence the recording; a friendly local extension officer is liable to over-estimate while many will tend to exhibit values towards an expected normal, that is, a likely under-estimate for good or bumper year and over-estimate for bad or deficit year.

2.2.1.2 Report by agricultural holder

This is a self enumeration in which a farmer completes a questionnaire and mails it back. This method is not normally practised for small holders who are scattered deep into the interior of the country where postal services are scarce. Also it cannot be used in those areas where many locals have different languages with no common one.

2.2.1.3 Interview method

This is probably the most widely used method in developing countries for assessing crop areas and yields. Essentially enumerators equipped with questionnaires visit selected holdings and interview the farmers. With this kind of method the difficulties lie in the design of questionnaire. Often the questionnaire is designed in such a way that the farmer should know the answer in units that can be easily converted into standard units. Such kind of situation is difficult to achieve in Africa where there are numerous kinds of measuring instruments each having different units. Most of the developing countries use this method alone or together with other methods.

Other methods, in which actual area measurement is involved, are explained in the following section.

2.2.1.4 Actual area measurement method

An actual measurement as applied by various statistical organizations is generally based on ground measurements and is considered to give the most reliable data. This is, strictly speaking, the only method applicable in countries where farmers are not able to provide data. The method is also recommended to all other countries for checking the quality of data collected by other methods provided that it is carried out to the required standard with appropriate instruments.

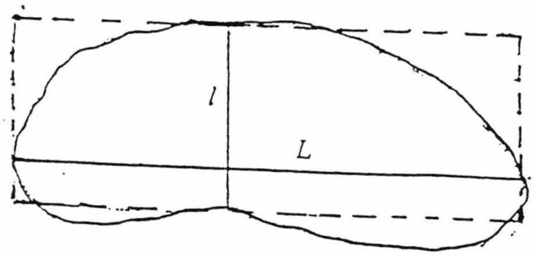
As a result of observations made by Panse (1954) on some deficiencies on other methods (non-actual measurements), actual area measurements made by trained enumerators are now widely used in developing countries.

The actual measurement of crop areas for estimation using up-to-date cadastral maps is an expensive operation and not an easy one. Few developing countries have such maps, and for those which have they are almost always out of date. To estimate crop areas using cadastral survey maps requires a scale of around 1:10000 if any degree of accuracy is to be achieved. Few countries maintain such maps for general services, for example, roads, rail and urban development. They are virtually useless when it comes to shifting agriculture in rural areas, as commonly found in sub-Saharan Africa. The methods for these types of surveys require expensive instrumentation. There should also be considerable time for setting up the instruments. Cadastral survey methodology is not an option easily available for agricultural surveys.

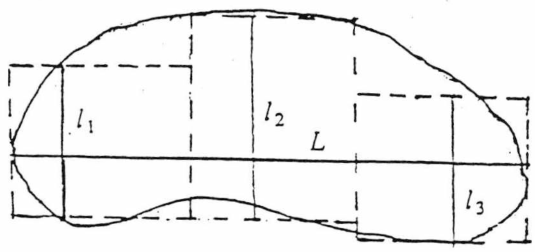
Agricultural statisticians in collaboration with engineers and cadastral surveyors have devised simple methods using relatively inexpensive instruments. Enumerators, after training, should be able to produce estimates of sufficient accuracy for individual crop areas. Three of these commonly used methods are outlined below.

2.2.1.4.1 Method of rectangulation

Land parcels and crop fields in many developing countries have irregular boundary lines which are not necessarily straight. The method measures two representative values, length and width of the farm. The use of rectangulation requires hypothetically covering an area with a rectangle which is judged to be the same area as that covered by the crop. Bits of the rectangle will lie outside the crop and *vice-versa*. The enumerator has to use his/her judgement when measuring the length and width of the rectangle before the computation of an area.

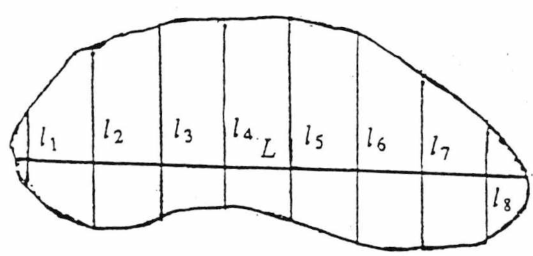


(a) $A = L \times l = 22.3$



(b) $A = L \times \bar{l} = 19.5$

$$\bar{l} = \frac{1}{3}(l_1 + l_2 + l_3)$$



(c) $A = L \times \bar{l} = 20.6$

$$\bar{l} = \frac{1}{8}(l_1 + l_2 + \cdots + l_8)$$

Figure 2.1: Area by rectangulation

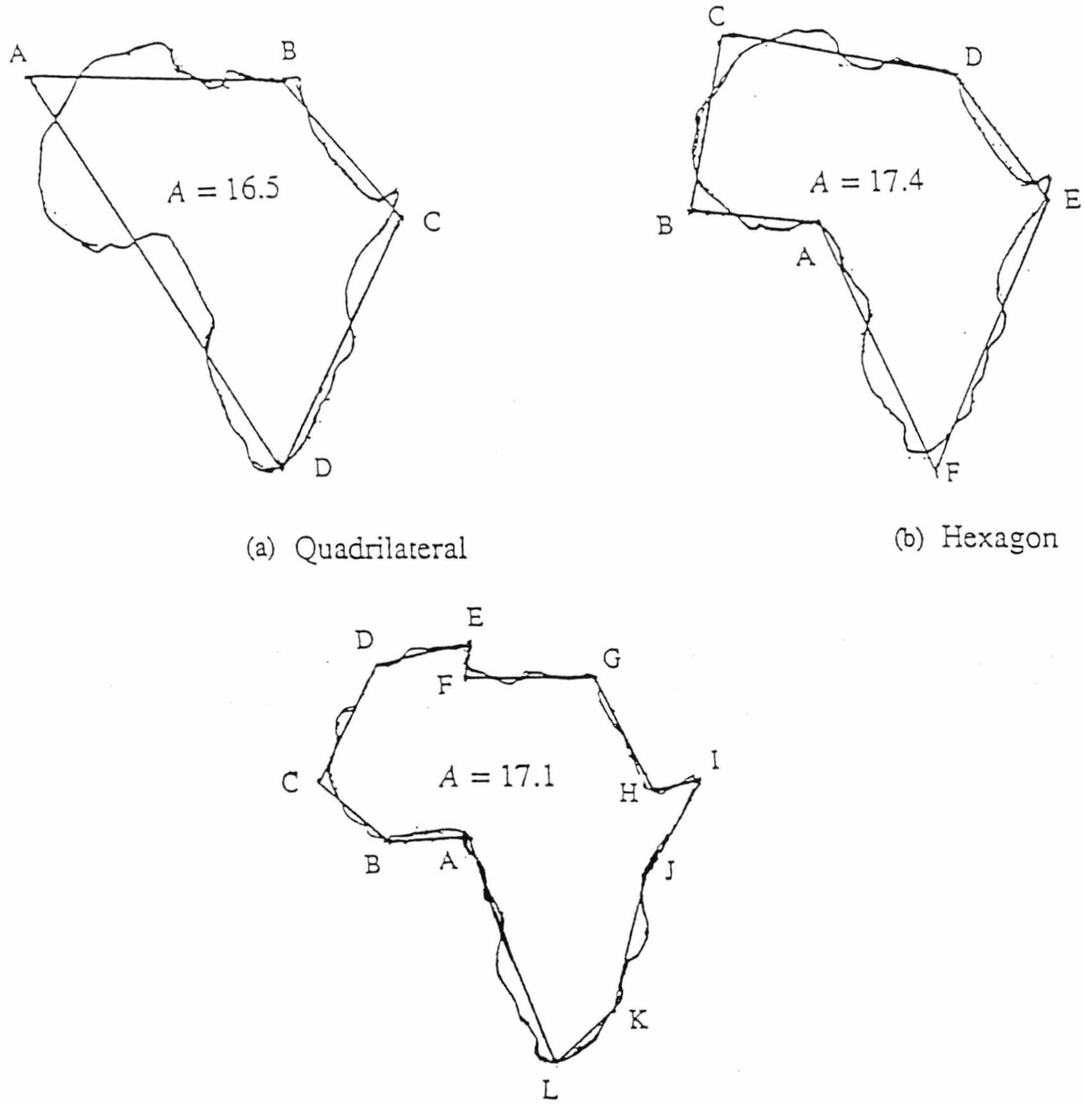
(adapted from FAO 1982)

The method can be slightly improved by having several measurements of widths which should be equidistant in between and thus dividing the total area into a number of rectangles or more precisely a number of trapezia, (Figure 2.1a, b and c). The method has four potential disadvantages.

1. The enumerator has to use judgement about balancing the areas included and excluded. For a person with long experience this may be easy, but it is difficult for an inexperienced one.
2. The method can only be applied during the early days of cultivation, but before the full crop development, because when crops are fully grown it becomes difficult to move through with measuring instruments.
3. Measurements of lengths and widths must be made at right angles to each other, and laying out lines accurately in this way is not easy. Departure from right angles produces different areas.
4. For a field with very irregular boundaries, there will be an infinite number of trapezia.

When the fields have very irregular shapes, which is common in countries with shifting cultivation, an alternative to rectangulation is to transform the field into a rectilinear closed figure and to demarcate on the ground the vertices of the equivalent polygon by poles or pegs. The operation is carried out in such a way that an equal portion of areas left out is compensated.

As an illustration of measuring the area of a field, the contour of the African continent is considered to represent the curvilinear boundaries in Figure 2.2 (many crop fields in developing countries have even more complicated shapes). In the figures (a, b and c) different attempts to produce regular polygons, namely a quadrilateral, a hexagon and a dodecagon respectively, have been made to arrive at a better estimate.



(a) Quadrilateral

(b) Hexagon

(c) Dodecagon

Figure 2.2: Area by different polygons

(adapted from FAO 1982, Petricevic 1988)

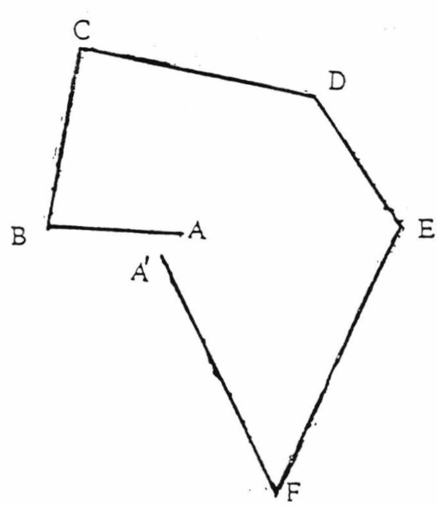


Figure 2.3: Closing error AA'

(adapted from FAO 1982, Petricevic 1988)

The percentage error in the estimated area because of this operation is not very high: 4 percent in the case of the quadrilateral and less than 1 percent in the case of the dodecagon (Petricevic 1988). The disadvantages to this alternative method are similar to the previous ones.

An easier way of calculating the area, if the boundaries of a crop field are rectilinear and the field is a plane polygon with well identified vertices, is a triangulation method.

2.2.1.4.2 Method of Triangulation

In this method a plane can be subdivided into a number of triangles with common vertex (a condition which is not indispensable but useful). The area of the field can be calculated as the algebraic sum of these triangles. The common vertex of the triangles can be one of the vertices of the polygon, a point outside the field. A refinement is to have two vertices and measure two sets of triangles. Then a check calculation can be carried out.

The method is safer than rectangulation because it eliminates much of the judgement left to the enumerator, and setting out right angles is not involved. However the method can only be used before crop development, that is when all the vertices of the polygon representing the field are seen from one particular point and when all the distances between the fixed point and the vertices are easily measured without trampling the crop.

2.2.1.4.3 Method of traverse

When neither of the other methods can be used, the traverse method becomes a useful alternative. This requires a polygon to be laid down around the grown crop, such that each apex of the polygon can be viewed from its adjacent vertices. The measuring exercise requires two persons, let us say an enumerator and an assistant to the enumerator. The starting point, where the enumerator erects a pole holding one end of a tape, is recommended to be the most north westerly vertex for ease of compass reading and recognizing direction when walking clockwise. The assistant walks to the adjacent vertex as mentioned in a clockwise direction with the other end of the tape, and erects a pole. The enumerator takes the bearing of the assistant's pole with a compass. This is referred to as forward bearing. Then the enumerator walks to the assistant's pole, records the distance and takes the bearing back to the first pole. This is referred to as backward bearing. This is done in order to cross check the bearings before recording the bearing of the first point to the second. In a perfect situation the difference between these two bearings should be 180 degrees, but a range of difference between 178 and 182 is practically tolerable. The process is repeated from the second vertex to the third, and so on, to the last vertex and the first. Finally the lengths of all sides of the polygon are measured with their corresponding angles between the sides.

When all the sides and bearings are measured and the polygon is plotted at an appropriate scale, it almost invariably happens that the figure does not close, that is, two end points (first and last), which should coincide, are some distance apart (see Figure 2.3). This distance is called "closing error" or error of closure. It is expressed as a percentage of either the perimeter or the longest diagonal. It has an acceptable range depending on the required accuracy of the data. The upper limit of the acceptable closing error varies between 3 to 5 percent of the perimeter which would produce an error in the estimated area. The level can be raised for those countries with inexperienced enumerators and less precise instruments. When a level is accepted the polygon is closed using different methods of adjusting the sides and/or the angles before the area is calculated (Petricevic 1988). One of the means of adjustment is to review the forward and backward bearings.

All cadastral measurements are concerned with measuring the horizontal projection of the land area. Measuring surface distance will produce over-estimates of the horizontal projection. While it would be desirable to achieve horizontal measurement, enumerators tend to ignore slope. If the slope is small, the horizontal plane length can be neglected and the length of the side on the slope can be used. For slopes of 5°, 10°, 15° and 20° the relative errors introduced by using the length of the side on the slope are 0.4%, 1.5%, 3.4% and 6.0% respectively.

2.2.1.4.4 Measuring tools

Apart from transport problems to the villages, in the developing countries, there is also a problem of movement from one farm to another, therefore the entire measuring operation and instruments must be as simple as possible. Also the enumerators, in general, carry out multipurpose activities, they are at the same time extension workers, interviewers, surveyors and so on, and with the cost involved, the Food and Agriculture Organization of United Nations (FAO) has continously been experimenting with different types of measuring instruments. Only those instruments with an acceptable level of accuracy, easy to use and not too expensive are recommended for developing countries.

A simple metallic or plasticized tape is a low cost instrument for measuring distances. The tapes, mainly graduated in metres, decimetres and centimetres are available in different lengths. A preferable one is of 50 metres. Tapes have advantages over chains in that they are easier to handle and flexible. However after long use they loose their graduation markings and get stretched. They also become difficult to use during strong winds.

The instrument generally used in agricultural statistics for measuring bearing is a compass. Many types of compasses, with different price tags have been used in the last decade. However lately FAO has been recommending a relatively low-cost compass. It is compact, light and flat; it has no adjustable parts, which makes its operation quite simple, and it will stand up to heavy duty. Handled properly, the instrument will give readings with an accuracy of half a degree, approaching performance of very expensive compasses (Petricevic 1988).

Given the measurements of sides and bearings of any polygon, programs can be prepared for the handy programmable pocket calculators. The instructions for the use of the calculator are quite simple and require very little training to be mastered. FAO has prepared such programs, together with the mode of operation, for a large number of the available pocket calculators. Typical programs and the instructions for the calculators type fx-795P and fx-790P are enclosed in Appendix I.

2.2.2 Data collection on yields

Data collection on yields is normally done some time later after collection of data on areas. There have been several methods in the collection of yield statistics. In this chapter we will review eye-estimate methods, farmer's statements and the crop-cut method. The last two methods will be compared and discussed in detail in chapter seven.

2.2.2.1 Eye-estimate method

Reports by eye-estimate as outlined in section 2.2.1 by administrative units has been an early method common in developing countries. In spite of considering factors like crop condition, growth of crop, yield characteristics, this method involves less time, since other operations such as actual harvesting, threshing, cleaning and drying are not involved.

The serious handicap for this method is that only a specialised investigator who has a good knowledge and experience on the crops of the district can make a reliable estimate. However there is still doubt on this method as it is subject to personal bias/opinion. Moreover the method is discarded because of lack of those personnel who have thorough knowledge of the sampled areas. Panse (1954) experienced that the difficulties of using eye-estimates are greater under African conditions than they are in Europe, particularly in areas where mixed crops are grown.

In developing countries where there is a wide range of agricultural and economic conditions, different cultural practices, lack of proper storage facilities, frequent attacks by pests and plant diseases *etc.* have aggravated the inability of the eye-estimate method in estimating yields to provide reliable results from inexperienced reporters.

2.2.2.2 Farmer's statements

Unlike the eye-estimate method where a reporter assesses the yield, in this situation it is the farmer who assesses the yield under the interview method. There can be no one who knows the field better than the farmer. Units of measurements, known by the farmer, that can be converted to recognised national and international level must be available if this method is to be useful. New findings (Verma *et al.* 1988 and Scott *et al.* 1989) have shown that farmer's statements are as reliable as any of the new methods for estimating yields. This method will be discussed in detail in chapter seven when a comparison with the crop-cut method is made.

2.2.2.3 Crop-cut method

In this method the yield is measured by selecting sample fields and locating sample plots within the fields. In the sampled plots, subplots are hopefully randomly selected and the crop in them is harvested, threshed, dried and/or otherwise processed, and the produce weighed.

The proper choice of the size and shape of the subplot, which also determines the method of marking it, is a matter of considerable importance. The whole procedure can be summarized as:-

- selecting the sample fields and plots;
- locating the sample subplots within the plots;
- measuring the crop density (optional);
- harvesting and processing the crop;
- weighing the produce at different stages; and
- estimating the yield.

The estimation of yield is described in chapter seven.

There are basically three shapes of the subplots and numerous sizes depending on the crop and other conditions which will also be discussed in detail in chapter seven. In this section we will consider subplots of two kinds; a square of size $25m^2$ and a circle of size $10m^2$.

2.2.2.3.1 Square subplots

Square subplots have been very popular in many African surveys. They are demarcated using string or wire separated, at four equal spacing of $5m$ and one spacing of $7.07m$, by knots or rings respectively. Figure 2.4 illustrates the positioning of the square.

2.2.2.3.2 Circular subplots

Circular subplots are considered to be more accurate than square (Anderson and Holmberg 1988, Olsson 1990) and are now introduced in some developing countries (*eg*. Tanzania). They use solid tools which make the borderline inclusion less complicated. The subplots can be demarcated using an instrument similar to a pair of compasses. The instruments are manufactured in such a way that they can be folded to ease movement problems. Figure 2.5 illustrates the instrument.

2.2.2.4 Other methods

Other methods based on weather factors and observations on plant characteristics (*eg*. height of tiller or plant, weight of grain, length of panicle *etc*.) cannot be used for survey purposes and can only be conducted in research stations where meteorological and laboratory equipment and experts are available.

There are basically three shapes of the subplots and numerous sizes depending on the crop and other conditions which will also be discussed in detail in chapter seven. In this section we will consider subplots of two kinds; a square of size $25m^2$ and a circle of size $10m^2$.

2.2.2.3.1 Square subplots

Square subplots have been very popular in many African surveys. They are demarcated using string or wire separated, at four equal spacing of $5m$ and one spacing of $7.07m$, by knots or rings respectively. Figure 2.4 illustrates the positioning of the square.

2.2.2.3.2 Circular subplots

Circular subplots are considered to be more accurate than square (Anderson and Holmberg 1988, Olsson 1990) and are now introduced in some developing countries (*eg*. Tanzania). They use solid tools which make the borderline inclusion less complicated. The subplots can be demarcated using an instrument similar to a pair of compasses. The instruments are manufactured in such a way that they can be folded to ease movement problems. Figure 2.5 illustrates the instrument.

2.2.2.4 Other methods

Other methods based on weather factors and observations on plant characteristics (*eg*. height of tiller or plant, weight of grain, length of panicle *etc*.) cannot be used for survey purposes and can only be conducted in research stations where meteorological and laboratory equipment and experts are available.

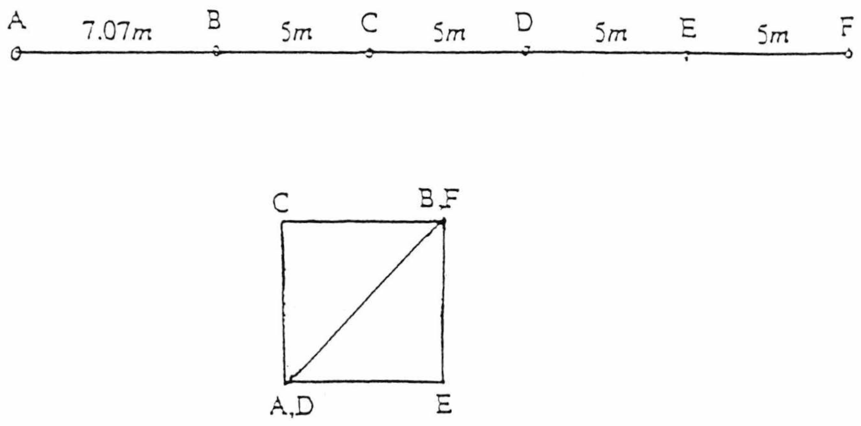


Figure 2.4: A crop cutting string or wire

(adapted from Takwimu 1988, Olsson 1990)

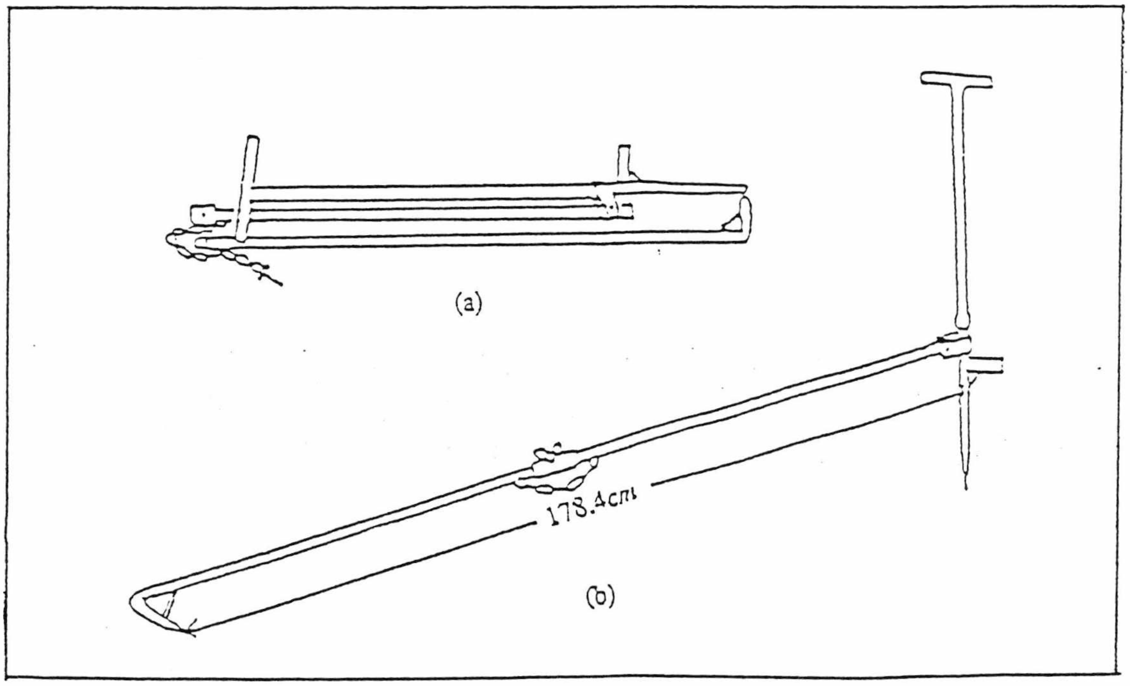


Figure 2.5: Circular crop cutting tool

(a) When folded for transport

(b) On the field

(adapted from Olsson 1990)

2.3 Tanzanian data situation

Tanzania is essentially an agricultural country in as much as 80% of the population is engaged in agriculture and allied activities. Such a vast majority of population, however, accounts for about 60% of the gross domestic product, and provides about 75% of foreign exchange earnings (MALD 1989). With current economic problems facing the country, as in any other developing countries, the need for statistical information to facilitate well informed decision making is not only the more important and obvious, but extremely fundamental to the whole planning process. It is on this background that we review the statistical situation in the existing organizations to suggest what measures could be taken to improve the situation.

2.3.1 Existing data collection systems

Tanzania is a union made up of two countries, Tanganyika, now referred to as Tanzania Mainland and Zanzibar and Pemba, now referred to as Tanzania Islands. There are some ministries which are common to both countries and others operate independently. Ministry of Agriculture is in the latter category. As such, agricultural data are published separately. Therefore there are two estimates, one from each ministry, and although they may be obtained through different methods, it is their sums that give national figures. This study will investigate existing data and the collection system of the Tanzania (Mainland) only. Tanzania (Islands) estimates are not included. The collection of agricultural statistics in Tanzania (Mainland) is based on two major independent institutions. These are the Ministry of Agriculture and Livestock Development (MALD) and the Central Bureau of Statistics (CBS) under the planning commission in the President's Office.

2.3.1.1 MALD

In MALD there are five sections/departments which collect agricultural data. These organs and their functions are briefly mentioned below.

1. The Statistics and Farm Management Unit (SFMU) of the Planning and Marketing Division.

It collects data from the districts and regions through the Agriculture Extension Services. This is a routine reporting system whereby at present the bulk of agricultural data are collected and transmitted to the ministry through the Regional Agricultural Development Officers (RADOs). It also collects data through sample surveys known as Current Agricultural Sample Survey (CAS). Perhaps this is the only interesting section as far as our study is concerned. There are laid down statistical procedures in the sample selection.

From a selected region, a list of all villages in their respective districts is formed. Villages are thus selected one in every k, using a random start for 1 to k and subsequently adding k to whatever random number was picked out. The exercise is stopped when the desired number of villages is attained (systematic sampling). From each selected village a list of all holders, persons who are supposed to be heads of households, is formed. A number of holders (sample size) is selected using random number tables. The selection of a region is however arbitrary. The survey uses actual measurements, that is, physical field area measurements and crop cutting to determine area and yield of a field respectively. There is a yearly current survey report. The exercise started in 1984/85.

2. The Crop Monitoring and Early Warning System (CM&EWS)

This department makes food crop production forecasts every year to monitor the food situation in the country. The exercise is based on agro-meteorology information received from various centres in the country and runs over the rainy seasons. The project collaborates with the meteorology directorate particularly in respect of weather information. It is rather unfortunate that I could not get hold of any literature that explains the procedure.

3. The Marketing Development Bureau (MDB)

This is also under the planning and marketing division like SFMU. This department among others, makes crop estimates on its own using CM&EWS data and publishes the results in series of years in a book *Basic Data Agriculture and Livestock Sector.*

4. The Project Preparation and Monitoring Bureau (PPMB)

This bureau collects data especially for the formulation of agricultural projects. It also collects information for its own use in monitoring and evaluation of agriculture programmes and projects.

5. The National Food Strategy Unit (NFSU)

This collects food data including agricultural input data for the formulation as well as reviewing of national food strategy including the strategic grain reserve (SGR).

2.3.1.2 CBS

The second major institution in the collection of agricultural data is the Central Bureau of Statistics, under the Ministry of Economic Planning, President's commission, who are performing a yearly agricultural sample survey (Agsasu) which is designed to give only national estimates. Agsasu functions under the umbrella of National Master Sample (NMS), a framework for integrating and systematizing household based surveys. The NMS, which was created in 1985, divides the country (Mainland only) into five groups. These groups were made by a combination of the twenty existing regions; hence each group forms a zone. The use of administrative units, particularly wards and villages, was adopted for the sake of organizational convenience as well as reduction of costs. The five zones comprised 150 strata. The Agsasu design is built from the 150 strata collapsed into 50 superstrata. From each superstratum a ward (primary sampling unit) is selected. Finally one village (secondary sampling unit) is selected from each ward. Only rural areas were included. As far as the NMS in general is concerned, the selection of villages completes the building of the design.

From the selected villages, Agsasu constructs a list of all farming households from which between 45 and 82 of them are selected through a systematic sampling procedure. The Agsasu first survey to be carried out within this frame was in 1986/87. The objective was and still is to obtain national estimates of crop area and yield. There has been only one survey for a regional estimate for one of the regions, Morogoro region, for 1986/87 whose results have not been published. The Agsasu survey for the national estimate is conducted in four stages in each of the two agricultural seasons; the short rain period, between September/October and February/March (*vuli*); and long rains period between March/April and July/August (*masika*). The first and last stages of the survey involved interviews on identification and particulars of members of households, their crops grown, area and harvest. The second and third stages involved measurements by enumerators of crop area and yield respectively. Agsasu focuses on traditional or peasant farming to cover production of the main food crops among others.

2.3.2 Comparison on output data

Because traditional farming is important for the food security of the nation, developing countries should therefore ensure proper methods for improving data collection. The devised methodologies on food and agriculture compilation have to meet estimates of acceptable precision.

The current problems with the set up of the outlined sections/departments are that there is no coordination whatsoever between them and a lot of the available data are not reliable (Makusi 1990). The situation is worse on estimates of food crops produced by small holders (traditional farmers).

Since traditional farming is important for the food security of the nation, it should be emphasised here that unreliability of estimates may cause disharmony in the inventory system. Tables 2.1 and 2.2 show some results from surveys organized by different organs from the same ministry which seem to give quite different values of the same estimate.

Table 2.1: Area and production estimates for Arusha region in Tanzania, 1985/86

	Crop	CAS	RADO	CM&EWS	MDB
Area ('000 ha)	Maize	145.44	129.92	124.7	110.53
	Paddy	2.77	2.35	2.15	na
	Sorghum	2.08	25.56	na	na
	Wheat	25.42	31.02	28.47	na
Production ('000 metric ton)	Maize	392.58	288.69	202.00	365.09
	Paddy	6.68	5.14	3.21	na
	Sorghum	2.19	30.19	na	na
	Wheat	81.76	49.54	50.34	na

na = not available

Table 2.2: Area and production estimates for Mbeya region in Tanzania, 1987/88

	Area ('000 ha)		Production ('000 tonnes)	
Crop	CAS	RADO	CAS	RADO
Maize	131.88	184.34	231.70	313.69
Paddy	40.75	38.84	108.86	80.23
Sorghum	9.10	16.50	13.97	21.33
Wheat	0.91	1.04	0.94	1.50

Table 2.3 gives crop planted area from the two major agricultural surveys in the country. One was performed by the Ministry of Agriculture (CAS), and the other by the Central Bureau of Statistics of the President's Office (Agsasu), for the same year 1987/88.

Table 2.3: Planted area by crop according to two different surveys in Tanzania. ('000 ha)

Crop	Agsasu	CAS
Maize	1541	1626
Cassava	567	707
Paddy	299	351
Sorghum	259	416
*B. Millet	259	308

* CAS estimate includes also Finger Millet.

Key:

CAS	Current Agriculture Survey
RADO	Regional Agriculture Development Officer
CM&EWS	Crop Monitoring and Early Warning System
MDB	Marketing Development Bureau
Agsasu	Agricultural Sample Survey

Sources: (i) Central Bureau of Statistics, Dar es salaam, Tanzania.

(ii) Ministry of Agriculture and Livestock Development Dar es salaam, Tanzania.

Land information, in terms of both area and production, is a major part of agricultural statistics. It is needed at least at three different levels: at the local level for the purpose of physical planning and administration, at the national level for the overall resource policy and management, and finally at international level for comparative description and analyses of national usage.

It seems that there is more data collection than data usage, let alone the unreliability of collection processes. The seriousness of the problem of unreliability and inconsistency in the production of data estimates at high level of management can be demonstrated by the following example of maize production estimates in Table 2.4.

Table 2.4: Maize production estimates 1986/87 for Tanzania ('000 metric ton)

	Source	Estimate
1.	Basic Data Agricultural and Livestock sector (SFMU) (1982/83-1986/87), MALD 1988	2787.33
2.	Food Security Bulletin MALD June, 1989	2358.50
3.	Annual Review of Agricultural Marketing (MDB), MALD 1988	2359.00
4.	Speech made by the Minister (MALD), budget session, 1989/90	2707.30

Source: The need of coordination of Agricultural statistics, MALD. Makusi (1990).

One of the main sources of variation in all these data is the methods used. In view of that it is preferable to streamline the methods and rationalize the activities in order to reduce these differences and attain reliable estimates at optimum costs. Better still, the costs of collection of different sets of data should be amalgamated to allow a larger sample to be included in the annual data set, thus improving precision.

Developing countries should be aware of the new technologies and try them to help improve agricultural statistics. The use of the remote sensing technique, among others, provides land information in terms of land cover. The main aim of this study is to make use of remote sensing in order to eventually incorporate this new type of data collection method into the usual agricultural surveys. With the aid of this modern technology, this study intends to integrate the ground information, by interviewing farmers regarding particulars of their households to propose a methodology for estimating area and production of crops at regular and reliable sample surveys.

Chapter three

Remote Sensing

3.1 Introduction

Remote sensing has been defined in various ways by different authors (Fischer 1975, Curran 1985, Lechi 1988, *etc.*), but in almost all definitions an important aspect which is more pronounced is the detection or measurement of physical properties of an object without having physical contact with it. This covers a wide range of possibilities including use of sound waves, light waves *etc.* From the agricultural statistics point of view, remote sensing refers to the collection of data about the earth's surface by a distant detector which makes use of the energy of electromagnetic waves (see section 3.3.1). The detected properties of the earth's surface give information in pictorial form or as statistics. The detector is of various kinds of instrument, from simple cameras to sophisticated scanners. In its simplest term, remote sensing technology, with its systems for data reception, processing and evaluation, can be considered as a technical imitation and extension of the human eye-brain system.

The history of using remote sensing for developing mankind can be traced back to the nineteenth century with simultaneous birth and development of photographic techniques, which improved the recording of sensing characteristics of the human observer and allowed for the visual information records. Photography was used in making topographic maps, at first by the use of ballons and later using aeroplanes. The latter is what is known as aerial photography. The first recorded photographs were taken from an aeroplane piloted by Wilbur Wright in about 1909 when he took motion pictures over Centocelli, Italy (Taherkia 1985). Ever since there have been rapid innovations of photographic instruments.

3.2 Development in remote sensing

Aerial photography has a history going back to the early 1900's. The aerial photographs were useful during war times when they were used in the analysis of bomb damage, surveys and analysis of bridge and dam structures with a view to their destruction. Visual interpretation of black and white aerial photographs parallelled research into the use of data from the new aircraft in both content and organization. Developments were followed very closely through reports, symposia and journals. By the late 1960s a wide range of photographic emulsions were being used regularly. The last two decades have been swept by new innovations and techniques that has brought a wide interest in the field of remote sensing, not only to the developed countries but to developing countries as well.

3.2.1 Space era

Perhaps the greatest turning point in the development of space technology was in October, 1957 when the then Union of the Soviet Socialist Republics (USSR) launched their first artificial satellite, Sputnik 1. A satellite is simply a technological object in orbit around a celestial body. In the case of Sputnik 1, the object was the size of a football and it carried a radio transmitter which sent signals to the earth. It orbited the earth every 96 minutes at heights of 215 to 930 *km*. It had been launched by a three-stage vehicle mounted one on top of the other. It ended its life during the first month of 1958. A month later United States of America (USA) launched Explorer 1. It also had a radio transmitter as well as a geiger counter for detection of electrically charged particles such as cosmic rays.

A series of space craft launchings followed. For example, USSR launched three space craft in 1959. The Luna 1 was launched in January. It passed within 750*km* of the moon. Luna 2 was launched in September and Luna 3 followed later. More followed as some made controlled landings back on earth. The USA launched a series of Rangers in which the seventh sent more than 4000 pictures back to earth.

The first photographs in more than one spectral band, multispectral photographs, from space were taken by USA's Apollo 9 in 1968. Three distinct spectral bands were obtained covering the visible and near infrared part of the electromagnetic spectrum (section 3.3.1). In it, were four Hasselbald cameras equipped with green, red and infrared filters boresighted and synchronized. Three of the cameras were loaded with black and white films and the fourth with colour infrared films.

3.2.2 Manned space ships

Major Yuri Gagarin, a USSR air force officer, made a significant start by entering the space in his space craft, Vostok 1, on 12th April, 1961. On 5th May, 1961, the USA launched their first man space flight in which Allan Shepard made to a peak of the altitude of 187 *km*.

A giant step to mankind was observed in July, 1969 when USA's Neil Armstrong and Aldrin landed the "Eagle" on the moon. At 2.55 GMT on 21st July, 1969, Armstrong became the first man to step on the moon.

Besides the lunar programme, there had been several earth orbital observations for meteorological purposes and earth resources. Meteorological satellites have been in use for a long time and are always being developed. Environmental scientists have found greatest utility in data collected by sensors on board the three polar orbiting satellites of Television and Infrared Observation Satellite (TIROS), National Oceanic and Atmospheric Administration (NOAA) and Nimbus and the geostationary Synchronous Meteorological Satellites (SMS). The scientists also collected data from satellites which include Geostationary Operational Environmental Satellites (GOES), Meteosat and Himawari (see Barrett and Curtis 1982 and Curran 1985 for details).

3.2.3 Earth resources satellite

Following the success of the manned space missions, the USA's National Aeronautics and Space Administration (NASA) and the Department of Interior Affairs developed experimental earth resources satellites series to evaluate the utility of images collected from the unmanned satellites (Curran 1985). The USA satellites launched for the purpose of studying earth resources are called "Landsats". The first on the series was called Landsat 1 (originally ERTS-1 for Earth Resources Technology Satellite) which was launched on 23rd July, 1972. It carried two types of sensors; a four waveband Multispectral Scanning System (MSS) and three Return Beam Vidicon (RBV) television cameras specially designed to acquire data about earth resources in a systematic and repetitive way. This satellite operated until 6th January, 1978 when it ceased, after exceeding its expected life by about four years. The Landsat 2 (formerly ERTS-2) was launched on 22nd January, 1975 and was retired in July, 1983. A third set of these series, the Landsat 3, was launched on 5th March, 1978 and retired in September 1983. Two more advanced versions of Landsats were launched on 16th July, 1982 and 2nd March, 1984 as Landsat 4 and 5 respectively (see Appendix II for details).

3.2.4 Second generation, unmanned earth resources satellites

The French satellite "Systeme Probatoire del'Observation de la Terre" (SPOT), which when translated, literally means the earth observation test system, was the first earth resources satellite to be launched from Europe (Curran 1985). SPOT-1 was launched on 22nd February, 1986 and SPOT-2 was launched in February 1991. The SPOT satellite (Appendix III) is operated by the French Centre National d'Etudes Spatiales (CNES) with participation from both Belgium and Sweden. India has a project to launch a satellite similar to SPOT. The second generation of earth resources satellites is advancing rapidly. Two Japanese satellites, the Marine Observation Satellites (MOS-1) and the Earth Resources Satellites (ERS-1) are expected to be launched soon. Landsats will be followed by satellites number 6 and 7 to be launched in the near future. European Space Agency and Canada are also planning to launch joint observation satellites.

It seems therefore that there will be plenty of images around and stimulating new technology because of the satellite's greater potentials and the benefits attained earlier.

The development of informatics makes it possible to imagine systematic treatments of data which cover agricultural fields exhaustively.

3.3 Information flow and remote sensing system

A flow of information must exist between the object and the observer. This is necessary for an observer to acquire knowledge about a remote object. There are many factors which can affect the usefulness, cost-effectiveness and a degree of accuracy for producing information on agricultural resources. These factors include:

- the characteristic of the scene
- the data collection and processing strategies
- the sampling strategy
- the timeliness of the data.

The first factor and the basics of the second will be examined in this section, while the details of practical data collection and other factors will be discussed in the following chapter.

Characteristics of the scene

Information on crop resources emanates from plant leaves. A leaf is built of layers of structural fibrous organic matter within which are pigmented, water-filled cells and air spaces. Each of the three features, that is, pigmentation, physiological structure and water content have an effect on the reflectance, absorbance and transmittance properties of a green leaf (Curran 1985). Leaf thickness affects both pigment content and physiological structure which contributes to the major differences in leaf reflectances between species. For example, a thick wheat flag leaf will tend to transmit little and absorb much radiation whereas a flimsy lettuce leaf will transmit much and absorb little radiation (Curran 1980). Plant leaf contains a family of pigments known as chrolophyll which absorbs visible light for photosynthesis. This phenomenon is the key role in the identification and measurement of canopy reflectance as will be explained in section 3.4.

Generally the spectral reflectance of a vegetation canopy behaves in a similar manner which apparently creates variability in scene characteristics. Fundamentally, there are a limited number of parameters that can cause variation in the reflectance of an agricultural vegetation canopy. The basic ones listed by Colwell (1977) included: the radiometric properties (reflectance and transmittance) of all the components of the vegetation canopy, (including live and dead leaves, stalks and fruiting bodies); the area, orientation and distribution of the components; the soil reflectance; and the solar zenith angle, the sensor look angle and the azimuth angle. Atmospheric conditions is another parameter that can have considerable effect on reflectance. The accuracy with which remote sensing can produce agricultural statistics is therefore quite variable. A manifestation of the above consideration is an additional parameter of crops grown in an area, whether one or two or as many as ten. For example, it will be impossible to tell with certainty which crops we have, if the reflected energy comes from the overlap region of corn and soybeans (Wigton and Huddleston 1978). In chapter four, we will address this problem which faces sub-Sahara African countries and discuss the mode to be determined in the application of remote sensing.

The data collection and processing strategies

The carrier for the information is electromagnetic radiation (EMR). The main elements in the process of data collection in remote sensing are, the objects to be studied, the observer or sensor, the EMR that passes between the two and the source of the EMR (Bird 1991). As EMR appears in many stages of the overall data acquisition and analysis, the author emphasised that understanding its basics is an important factor when dealing with the principles of remote sensing.

Curran (1985) and Abiodun and Oesberg (1988) managed to break down distinctly the remote sensing system into four basic elements.

(i) The source

(ii) The transmission path

(iii) The target/object

(iv) The sensor.

3.3.1 The source

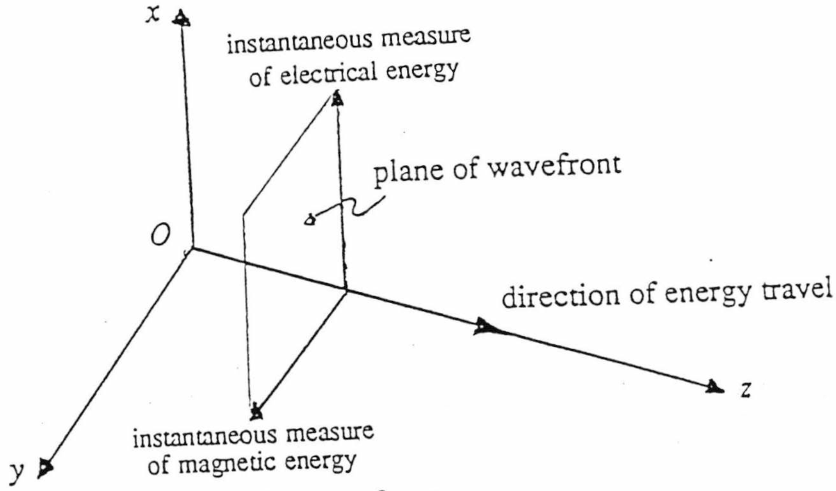
The source of Electromagnetic Radiation (EMR) may be natural like the sun's reflected light or it may be the heat emitted from the earth or even man made, like microwave radar. Actually all objects at temperature above absolute zero (0 K or $-273\,^{\circ}$C) radiate energy as EMR. The concept of EMR therefore describes the way in which energy is transferred from one object to another through space. The discovery of the existence of non-visible radiations opened the way to the thought that the sun emits energy which was beyond the detection capacity of the human eye (Lechi 1988). This made the sun the main source of the earth's radiant energy. Its surface temperature is approximately 6000 Kelvin. Only about one part in two billion of solar energy reaches the earth (Abiodun and Oesberg 1988).

In the process of energy transfer, that reflected, scattered or radiated from the earth's surface is primarily of importance in remote sensing. While there are several models describing the characteristics of EMR, only the wave model will be discussed in this study.

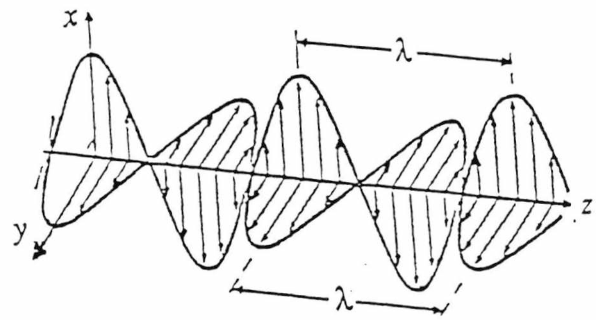
Wave model

The energy behaves in accordance with the basic wave theory and is composed of coupled time-varying electric and magnetic force fields which are orthogonal to each other. Mathematically EMR is represented by a smooth wave form (see Figure 3.1) with an associated wavelength, λ, frequency f and constant speed c. In space c is known as the speed of light and has a value of $3{\times}10^8\, m\, s^{-1}$. The relationship between c, λ, and f is given as

$$\lambda f = c \ \ (m\, s^{-1}).$$



(a) The wavefront



(b) The sinusoidal variation of electrical and magnetic radiation

Figure 3.1: Wave model of electromagnetic radiation

Electromagnetic radiations occur as a continuum of wavelengths and frequencies from short wavelengths and high frequencies called cosmic waves, to long wavelengths and low frequencies which are the radio waves. The term Electromagnetic Spectrum (EMS) is hence used to describe this vast range of short, medium and long wave radiations. Figure 3.2 shows the spectrum range and its respective wavelengths. In remote sensing one finds that the EMS is referred to in terms of wavelengths whose commonly employed units include nanometres $(n\,m)$, micrometres $(\mu\,m)$, centimetres $(c\,m)$ *etc.*

Since EMR from earth's surface is recorded in different wavelengths of the electromagnetic spectrum, the wavelengths that are of greatest interest in remote sensing study, are visible and infrared radiations in waveband $0.4 - 0.7\,\mu\,m$, and $0.7 - 14\,\mu\,m$ respectively, and microwave radiation in $10 - 200\,mm$. Specific terminologies for the infrared are near-infrared, middle-infrared and far-infrared (or thermal infrared) with their respective wavebands of $0.7 - 1.3\,\mu\,m$, $1.3 - 3.0\,\mu\,m$ and $3.0 - 14\,\mu\,m$. Boundaries between wavelength bands are gradational (Abiodun and Oesberg 1988). Figure 3.2 shows some overlapping to emphasize gradations.

The main regions of the spectrum used in remote sensing are summarized in Table 3.1.

Table 3.1: Regions of the spectrum used in remote sensing

Name	Wavelength range
Visible	$0.4\mu m - 0.7\mu m$
Near infrared	$0.7\mu m - 1.3\mu m$
Short wave infrared	$1.5\mu m - 2.5\mu m$
Mid infrared	$3.0\mu m - 5.0\mu m$
Thermal infrared	$8.0\mu m - 14.0\mu m$
Microwave	$1.0cm - 20.0cm$

Source: Principles of remote sensing

Bird (1991)

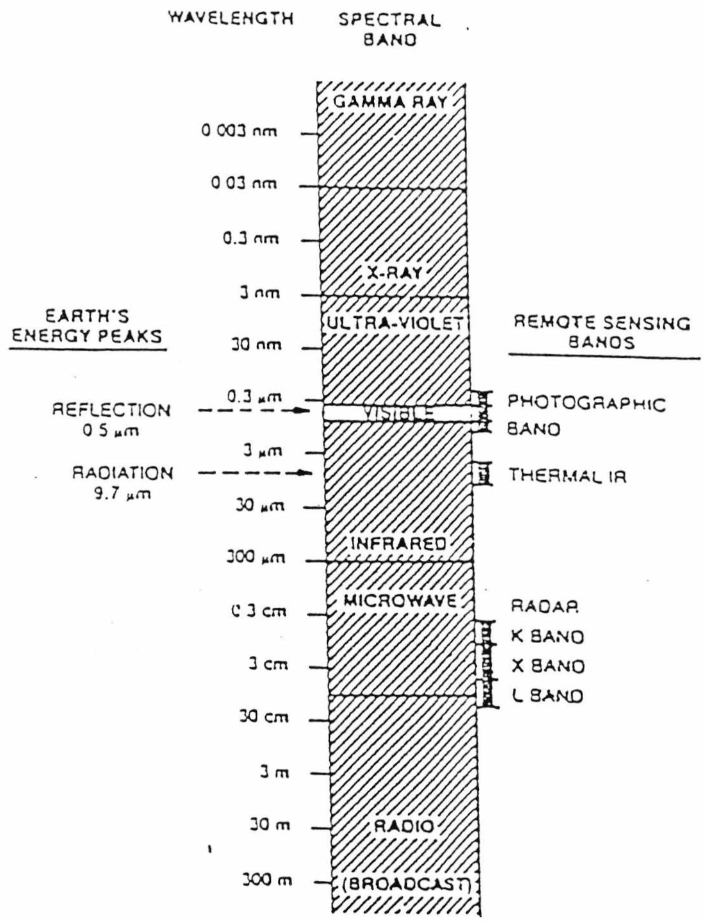


Figure 3.2: Electromagnetic spectrum

(from: Abiodun and Oesberg 1988)

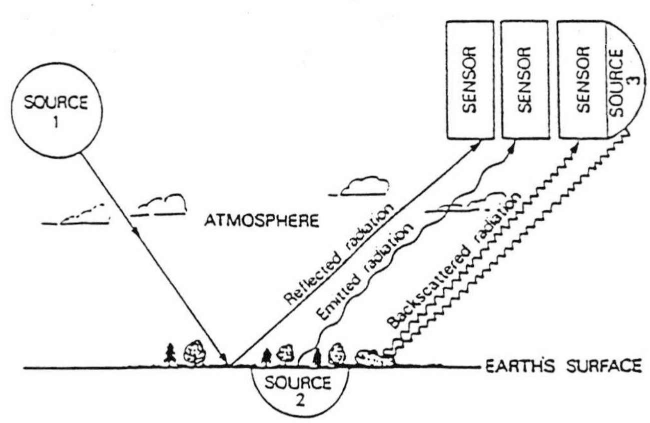


Figure 3.3: Remote sensing system

(from: Curran 1985)

The human eye perceives only a small part of the EMS, because the eye is not uniformly sensitive to all wavelengths of light. The average person can detect EMR between 0.4 and 0.7 μm whereas a photographic film can respond to about 0.3 to 1.2 μm. This spectral band is about three times as broad as that seen by the human eye.

In remote sensing, it is the near-infrared band that is of interest as we will observe later, because it corresponds to the band of high reflectance for vegetation. Infrared radiation is also important because it is predominantly the radiation emitted by the earth and the atmosphere. A Radio Detecting And Ranging (RADAR) provides its own source of electromagnetic energy in the microwave part of the spectrum and records the object's response. We will however not be concerned with this kind of teledetection.

3.3.2 Transmission path

The transmission path is that route in which radiation travels. Before and after it has interacted with target/object, EMR has to pass through a path prior to its detection. The path is the atmosphere, whereby it attenuates the radiation's speed, frequency, intensity, spectral distribution and direction because of scattering, absorption and refraction.

Atmospheric scattering primarily affects the direction of visible radiation but can also alter the spectral distribution of visible and near visible wavelengths. One type of scattering, known as Rayleigh, affects short visible wavelengths and results in haze. Unlike scattering, absorption affects wavelengths that are both shorter and longer than those of visible light. The regions of the spectrum in which atmospheric absorption is low are called atmospheric windows, and it is through these windows that remote sensing of the earth's surface takes place. A window is therefore a region of the electromagnetic spectrum where the atmosphere is effectively transparent, that is a band which offers maximum transmission and minimal attenuation. Refraction occurs when EMR passes through a stratified atmosphere depending on its stability.

3.3.3 The target/object

The target/object is the term used for all the elements of the earth's surface. An object reflects, absorbs and transmits radiation. Objects are identified by the extent to which they reflect incident radiation or emit their own. How much is reflected, absorbed, transmitted and/or emitted depends both on the material itself and other characteristics of the object. It is thus possible to identify features on the earth's surface on the basis of their spectral properties.

The proportion between the reflected, absorbed and even scattered electromagnetic energy on a vegetative surface varies with wavelength. The basic form of the spectrum is the same for most vegetation types but each is characterized by its own fine structure. The features happen to have unique patterns in different portions of EMS that can be easily and accurately separated by multispectral analysis (Lillesand and Kiefer 1987, Curran 1985 and Hall-Konyves 1988).

3.3.4 The sensor

A sensor is any device that receives EMR and presents it in a form suitable for obtaining information. The device is sensitive to various parts of the spectrum, thus simultaneous images of the same scene can be acquired in various wavelength regions. Diagramatically Figure 3.3 summarizes the remote sensing system.

3.4 Representation of measurements

Representations of remote sensing measurements are of two kinds; analog form (photograph) and digital form.

3.4.1 Representation in analog form

The lens system of the camera passes electromagnetic energy in the field of view (FOV) onto the surface of photographic film so that the spatial distribution of the energy reaching the film matches that coming from the object/target in the FOV of the camera; that is, it focuses the image onto the film. The surface of photographic film is covered by sensitive chemicals which respond progressively to the amount of electromagnetic energy reaching the surface. The developed film is an analogue showing the spatial variation of reflected energy coming from within FOV of the camera - photograph (Taylor 1991).

Panchromatic film is sensitive to visible wavelengths and produces photographs with a tonal range from black (low reflectance) to white (high reflectance) with grey tones for intermediate levels.

Colour photographs are produced using complex films which have in-built filters to separate the incoming energy into different wavebands. The film captures three images of the subject which are held in separate layers corresponding to three pseudo-primary colour ranges:

blues ($0.4 - 0.5\mu m$)

greens ($0.5 - 0.6\mu m$)

reds ($0.6 - 0.7\mu m$).

When the film is processed the response of each image is coded with the appropriate coloured dye. That is to say, in the image representing particular colour wavelengths, high reflectance gives bright colour and it is darkish for low reflectance. In the processed photograph the three responses are viewed together and the colours mix so that the combined response produces a colour picture.

Filters are used to modify the light falling on the subject or passing through the camera lens. They affect the colour of light that passes by blocking some of the undesired colour. Blocking visible blue light produces images of green, red and near infrared wavelengths inside the colour film sensitive to infrared. The pseudo-primary colour dyes are then used so that the image representing green light is dyed in shades of blue; the image representing red light is dyed in shades of green; and the infrared image which is invisible to the human eye is dyed in red shades. The combined effects of leaf pigments and physiological structure give all healthy green leaves their characteristic reflectance properties: low reflectance of red and blue light, medium reflectance of green light and high reflectance of near infrared radiation. Figure 3.4 summarizes the reflectance properties. This is the effect of colour shift that makes vegetation which appears green on colour film as it naturally reflects more green than blue or red to appear red as it reflects more near-infrared than green or red. The combination produces a false colour picture, referred to as False Colour Composite (FCC).

3.4.2 Representation in digital form

In contrast to photographs, a representation of measurement in digital form makes use of incoming electrical energy, converted into digital mode. The digital conversion is made into a finite range of discrete values depending on the sensor system. The range is normally from zero to 255 corresponding to minimum and maximum energy levels, but in some cases it may be from zero to 1023 (Taylor 1991).

The sensor systems which are designed to produce imagery in digital format are recorded from the scan line, divided into units of the instantaneous field of view (IFOV) which dictates the spatial resolution. The smaller the IFOV the better the spatial resolution. An image covering the whole scene is produced by systematically scanning the whole area with the sensor, thus producing a picture made up of elements (pixels). The digital image is a very large table of numbers, a number representing a pixel, each in the correct relative location and having a numerical value corresponding to the amount of electromagnetic energy coming from the pixel area (Taylor 1991).

Panchromatic pictures can be produced by mapping the digital values onto a visual display unit of a computer system so that each number represents a different grey tone. Low numbers are black or dark grey, high numbers are light grey or white.

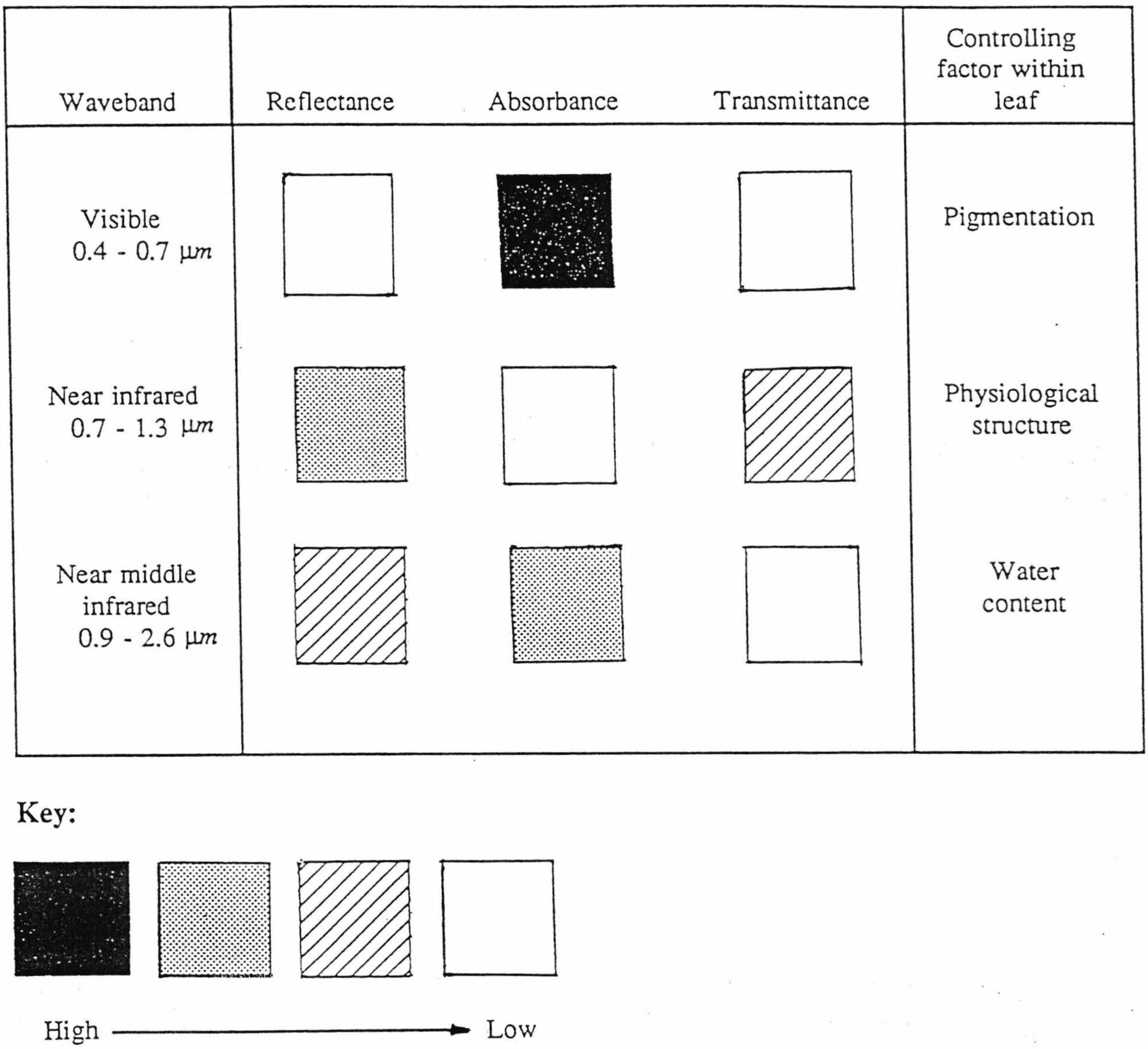


Figure 3.4: The reflectance, absorbance and transmittance properties of a green leaf in visible, near infrared and near middle infrared wavebands (adapted from Curran 1985)

Colour pictures are produced on a colour monitor where images representing three wavebands may be similarly coded separately in tones of red, green and blue on each of the colour guns. The resulting mixture produces the colour picture by viewing the image simultaneously. When images representing blue, green and red wavelengths are coded respectively in blue, green and red tones, the result is an image with natural colour. When the image representing green, red and near infrared are coded in blue, green and red respectively, the result is a false colour image analogous to that produced by colour infrared photographic film.

Figure 3.5 shows the entire operational process up to the output level, that of a digital or pictorial form, which is the main interest to the user.

3.5 Landsat sensors

The sensors that had been carried by Landsat 1, 2 and 3 were MSS and RBV while the recent satellites of this family Landsat 4 and 5 carry on board together with MSS a new sensor system Thematic Mapper (TM) which has seven bands.

Satellite orbit progresses slightly westward with each day just overshooting the orbit pattern of the previous day. The orbit of the landsat is sun-synchronous, that is the satellite keeps pace with the sun's westward progress as the earth orbits. This explains the satellite crossing at the equator at the same local sun time (9.42*am*) on each pass. The significance of the sun-synchronous orbit is that a specific area can be imaged under similar sun illumination conditions during specific seasons.

3.5.1 Multispectral Scanner (MSS)

The principal feature of MSS system is, as its name implies, the multispectral scanner. This consists of a mirror which oscillates through about 12° which reflects light on to detectors. The effect of the oscillating mirror is to scan a piece of ground 185*km* wide at right angles to the satellite's direction of motion (see figure 3.6). As the satellite moves quickly around the earth, it is necessary to record 6 scan lines at once.

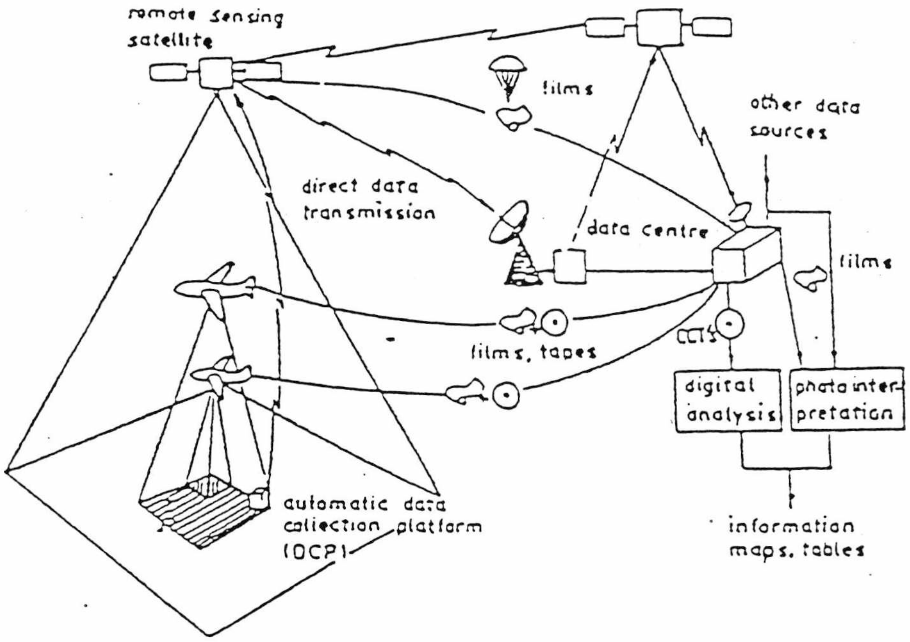


Figure 3.5: Scheme of a complex system for multi-stage acquisition and analysis of remote sensing data
(from: Abiodun and Oesberg 1988)

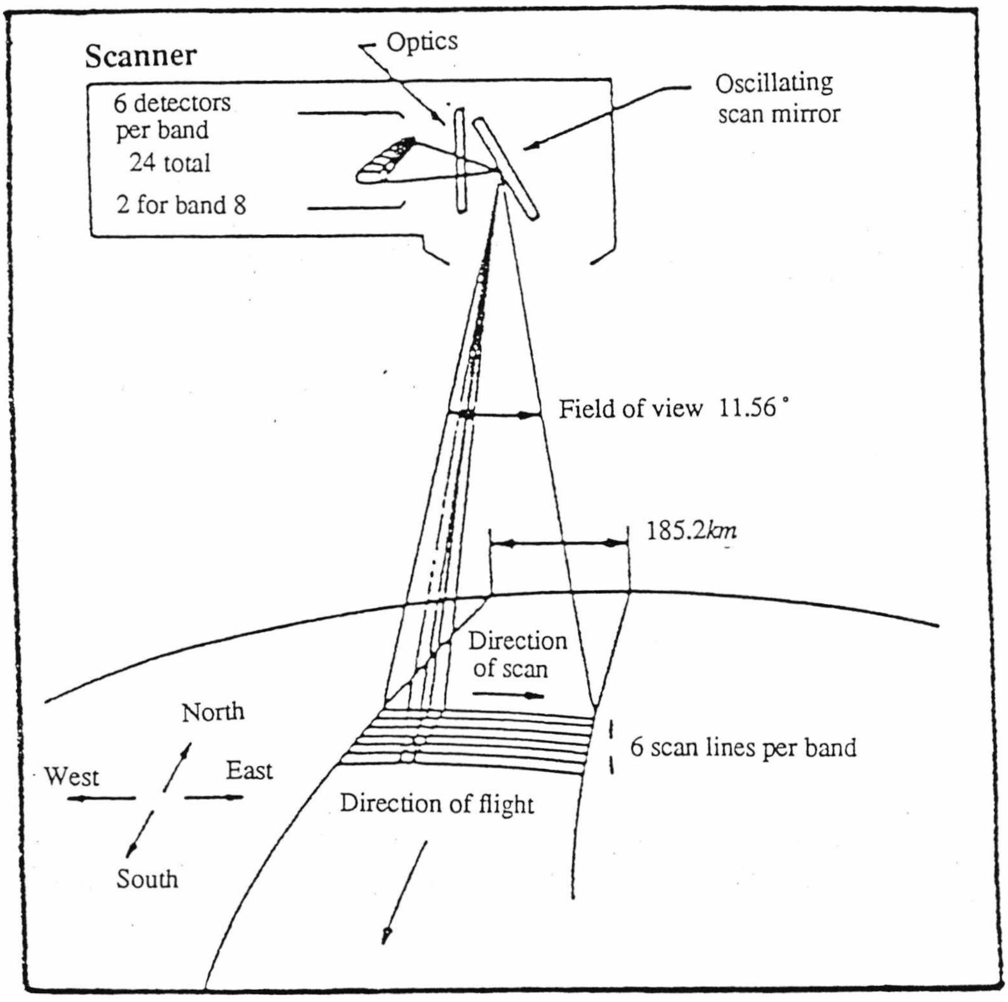


Figure 3.6: Landsat MSS scanning arrangement

(adapted from Lechi 1988)

For the four bands therefore, 24 detectors are in use. MSS has been operating in two groups of Landsat family.

Landsat 1, 2 and 3

The MSS records four images of a scene, each covering a ground area of 185*km* by 185*km* at nominal ground resolution of 79*m* (Curran 1985). These four images, each for a waveband in selected portions of the visible and near-infrared (NIR) wavelengths are listed as:

Band 4 $0.5 - 0.6\,\mu m$ (green)

Band 5 $0.6 - 0.7\,\mu m$ (red)

Band 6 $0.7 - 0.8\,\mu m$ (NIR)

Band 7 $0.8 - 1.1\,\mu m$ (NIR).

Landsat 3 had an additional Band 8 $10.4 - 12.6\,\mu m$ (thermal infrared) which failed shortly after launch. (Curran 1985, Taherkia 1985).

The approximate values for orbital parameters and coverage cycle which undoubtedly vary due to perturbation in satellite orbits and other factors for Landsats 1, 2 and 3 as outlined by Lechi (1988) are as following.

Altitude : 900 – 950*km*.

Inclination : 99°.

Period : 103 minutes.

Time of descending mode equator : 9.42*am*.

Distance between successive tracks : 2760*km* at equator, 2129*km* at 40°(14 orbits/day).

Distance between adjacent tracks : 158*km* at equator, (sidelap = 27*km* or 14%), 122*km* at 40°(sidelap = 63*km* or 34%) (18 days for repeat coverage).

Swath width : 185*km*.

Landsat 4 and 5

There have been slight changes in MSS for this family of Landsats. The first change was the addition of Band 8 in Landsat 3 which we have noted to have failed shortly after launch. The second change involved lowering the altitude of these Landsats which, as a result made the spatial resolution to be $82m$. The third change involved shifting the number of MSS bands from 4, 5, 6 and 7 to 1, 2, 3 and 4 respectively.

3.5.2 Return beam vidicon (RBV)

RBV television cameras were carried on Landsat 1, 2 and 3. On 1 and 2 three cameras were used, each filtered into a different waveband to view the same $185km$ by $185km$ ground area simultaneously as the MSS, and had ground resolution of $80m$ (Curran 1985). The respective bands are detailed below (Lechi 1988).

Band 1	$0.475-0.575\,\mu m$	(blue - green)
Band 2	$0.580-0.680\,\mu m$	(orange - red)
Band 3	$0.690-0.830\,\mu m$	(red - infrared).

As for Landsat 3, it carried two identical cameras with waveband $0.505-0.750\,\mu m$.

3.5.3 Thematic mapper (TM)

A remarkable innovation in the family of Landsats is the introduction of TM and the removal of RBV. The TM is a seven band scanner which is designed to maximize vegetation and rocks type discrimination. Bands 1 - 5 and Band 7 have ground resolution of $30m$ and Band 6 of $120m$. This represents a substantial improvement in spatial resolution over the previous landsat systems. Wavelengths of different bands are listed below.

Band 1	$0.45-0.52\,\mu m$	(green)
Band 2	$0.52-0.60\,\mu m$	(green)
Band 3	$0.63-0.69\,\mu m$	(red)
Band 4	$0.76-0.90\,\mu m$	(NIR)

Band 5 1.55 – 1.75 μm (near-middle infrared)

Band 6 10.4 – 12.5 μm (thermal infrared)

Band 7 2.08 – 2.35 μm (middle infrared).

Landsat 4 was inserted into a different path orbit from the previous ones. It circles the earth every 99 minutes at an altitude of about 705*km* resulting in 14½ orbits per day. The Landsat 5 has all characteristics similar to those of Landsat 4 but only that its cycle is offset 8 days from the latter.

3.5.4 SPOT

The first SPOT satellite, SPOT-1 has a near-polar, sun synchronous, 832*km* high orbit which repeats every 26 days. There are three bands in the visible and NIR portions of the spectrum. Curran (1985) thought that it was unlikely that a sensor with finer tolerance than Landsats 4 and 5 could be manufactured in the near future. The ground resolution for SPOT is 30*m*, while in a broader spectral band (black and white), it is 10*m*.

One of the key features of the SPOT instrument package is the provision for off-nadir viewing. That is, the observed region need not necessarily be centred on the ground track, it is possible to observe any region of interest within a 950*km* wide strip centred on the satellite ground track. The device introduced is the pushbroom scanner, a new generation of MSS, called High Resolution Visible (HRV) scanner.

The pushbroom scanner has overcome the problem of rotating mirrors which reduced accuracy, hence it contains no moving parts and records each scan line at one go by means of line detectors - one detector for each area sampled on the ground. This means there are many more detectors to calibrate than in MSS, a disadvantage in the pushbroom scanner. Another disadvantage is the inability to sense wavelengths longer than NIR. However the advantages in the weight, size, reliability, accuracy and higher spatial ground far outweigh the disadvantages.

The basic design for SPOT encourages its application to the small scale subdivision of agricultural lands in many parts of the world and also to cartography.

3.6 Literature review

Remote sensing of the earth's surface has many applications, both military and civil. It is relatively recently that the civil use of this technology has become available at reasonable cost. This study concentrates on those aspects which yield information on the development of agriculture, especially in relation to the cultivation of land and the assessment of crop area and production.

Estimation demands an accurate and reliable method of determination of crop area and yield. The Large Area Crop Inventory Experiment (LACIE) constituted an early effort to perform crop area estimations (MacDonald and Hall 1980) whereby satellite-borne sensors were used to determine the reflected solar radiation in various wavelength intervals (Idso *et al.* 1977). Images were used to measure the area of wheat grown, while other meteorological factors such as temperature and rainfall were also taken into account to estimate yield per unit area. The aim was to estimate production within 10% of the true value 90% of the time and to obtain accurate monthly estimates from early in the season through harvest (Taylor 1985).

As part of the 1980 worldwide agricultural census organized by FAO, Helden (1981) used the technique known as Area Frame Sampling. He studied the potential application of area frame sampling to demonstrate the possible use of Landsat data for stratification purposes, on a project started for collection of national agricultural statistics on a sample basis, in Niger. The author concluded that a complete area frame sampling can be constructed for all areas of Niger needing an estimated time of one to two man-years.

Moreira *et al.* (1986) estimated wheat area in Brazil by means of aerial photographs and Landsat MSS data. Using the Gaussian maximum likelihood classifier (see section

4.2), the author concluded that digital classification of single date Landsat MSS data provided reasonably accurate results for wheat area estimation (7.5% over-estimation) and that improvements in area estimation accuracy could be expected using Landsat TM and SPOT data. McCloy *et al.* (1987) monitored rice areas in New South Wales. From the regression analysis between estimated and observed rice area, the authors reported coefficient of determination around 90%. Hall-Konyves (1988), who disregarded the problem of mixed pixels, used remote sensing data to investigate its application in Swedish agriculture and found out that winter wheat and rye were strongly over-estimated while the estimated area of rape seed and sugar beet tallied very well with the actual area. The over-estimation was caused by classification of non-agriculture land. Application of non-agriculture land cover masks prior to classification might have increased the area estimation accuracy of rye. Winter wheat and barley are often spectrally confused and associated with rather low classification accuracies. On the other hand, spectral data of rape seed and sugar beet differed significantly from each other as well as from other crops on several acquisition dates. The yellow colour of the rape seed flowers and the unique phenological curve of sugar beet probably, the author thought, had an influence on the significant difference between these crops as well as on the area estimation accuracy. The author therefore concluded that for cereals like winter wheat, rye and barley a low area estimation possibility existed while the area estimation potential of rape seed and sugar beet was high in the studied provinces.

Spectral properties of individual crop species vary during their growth cycle. This together with the variation between species especially those with differing growth cycles, means that spectral differences between crops change with time. The fields in Africa's traditional farming vary and are often irregular. Crop land, frequently with mixed crops, is interspersed with non crop land which causes individual pixels to contain mixed reflective values. In such a complex situation, crop identification is therefore unreliable and misclassification cannot be ruled out.

Aerial photography has been in wide use for area estimation. An interesting example was that of Fiji. The estimate of area under rainfed rice in the Central Division in 1978 from the census was $1028ha$ and from post enumeration (checks) current survey was $707ha$. Aerial photography indicated that the area was $2083ha$, a figure even higher than that of the census. The processing of the aerial photographs was carefully checked and found to be correct. Also as a means of checking the validity of the method, the estimates of rice area in main irrigation schemes was compared with the known figure compiled by Drainage and Irrigation Division, and found to be in close agreement. Hence it was reasoned that if the area obtained from the aerial photographs ($2083ha$) were of right magnitude, then the ground sample census estimate of $1028ha$ and the survey estimates of $707ha$ would have to be investigated. It was argued that either the estimates of number of farmers growing rice (census 1422, survey 950) were too low or the estimates of average rice area per farm (census 0.72 *ha*, survey 0.74 *ha*) were too low or both.

Three aerial photograph blocks of rice land were therefore selected, one in each of the three main rice growing provinces, each block consisting of at least 20 farmers in which exactly 20 were listed. Hence the resulting 60 farmers were classified according to whether or not their names were located on the compiled rice frame. Only 29 of the 60 farmers listed in the exercise belonged to the frame.

Area measurements from those 60 farms were taken by 1980. The farms in the frame ($n = 29$) gave an average area per farm of $1.29ha$ while those not in the frame ($n = 31$) gave $2.50ha$, and for all farms ($n = 60$), the average area per farm was $2.22ha$. The average of $2.22ha$ was much higher than that obtained in the survey ($0.74ha$).

Discussion between all enumerators and officials responsible for both the above exercise and the survey to try and find an explanation for the discrepancy, revealed that in the aerial photo validation exercise the assistance of the local rice field officer was obtained in ascertaining the locations of the rice fields whereas in the survey assistance was obtained from the farmer. It was suggested that the farmer provided information only on the rice fields located in the vicinity of his/her house whereas the field officer advised also on rice fields located at some distance away. Unfortunately no reason was given for the farmer's behaviour. So it was decided to seek the assistance of the local field officers in finding additional rice fields operated by the farmers who were selected for validation survey. The validation survey was focused on a sub-sample of 10% of the post enumeration survey sample. This turned out to be 16, but because of the confusion over the fact that one farm operator had two farms and one of his farms was enumerated as nil by the census enumerator but was measured as $0.89ha$ in the post enumeration check, led the validation survey to locate rice on 12 of the 16 farms in post enumeration checks. The comparison of totals of areas measured from the sub-sample were $7.92ha$ from the census and $8.73ha$ from post enumeration survey. The findings of the validation survey were that on one such farm, additional fields of $5.59ha$ were thereby located bringing the total area of the 12 farms reporting rice to $14.32ha$ raising the average area per farm to $1.19ha$.

To know how many farmers might be unlisted, a supplementary listing was done in which an additional 119 farmers or 7% of the total 1638 previously listed were recorded. A further 5 aerial photo blocks of rice land were therefore selected and a listing of 60 farmers were carried out with classification on whether or not their names were located on the rice frame. Table 3.2 shows the results of the two series of aerial photographing.

Table 3.2: Classification of rice farmers according to census frame

		Number of rice farmers		
			listed in the census frame	
Series	Number of photo blocks selected	listed in the blocks	included	omitted
1	3	60	29	31
2	5	60	46	14
Total		120	75	45

Source: Report on the census of Agriculture 1978

Rothfield and Kumar (1980).

The total number of farmers omitted from the frame was thus estimated as $1638 \times 45/75 = 983$. Thus the estimated population of rainfed rice farmers in the division was 2621.

The area under cultivation of main season rainfed rice in the division for 1980 was estimated as follows:-

(i) area estimated from post enumeration survey, unadjusted : $707ha$.

(ii) correction factor due to area measurements: $14.32/7.92 = 1.808$, the ratio was of measured area in two post enumeration visits to area measured in the census for the sub-sample of 16 farmers.

(iii) correction factor due to incompleteness of frame: $2621/1638 = 1.6$, which was the estimated population (by the aid of the 8 photo blocks) to the original listing.

The revised estimated area was therefore $707 \times 1.808 \times 1.6 = 2045ha$, which was very close to the aerial photography of $2083ha$ (Rothfield and Kumar 1980).

The famines, droughts, under-production in Africa and consequent aid efforts have given rise to a need for close monitoring, accurate estimation, proper land management and effective rural development schemes. The cost of constant aerial photography is prohibitive. The cost of aerial photography on an area equal to one Landsat MSS scene is 100 - 200 times more expensive, though the latter resolution of ground objects will be 10 times coarser (Howard 1988). The ability of remote sensing to provide crop inventory and prediction statistics, and to map the most needy locations, makes it a potentially invaluable tool (Taylor 1985).

In an effort to integrate remote sensing in the developing countries, estimates of percentage cropped land area carried out in Kano State, northern Nigeria were made in a pilot study by Silsoe College (Taylor 1985). The study for four locations, each of 1,200 square kilometres area, showed that estimates from satellite images using maximum likelihood classification were similar to independent measurements from aerial photography. Table 3.3 shows the comparison of percentage cropped land area.

Table 3.3: Comparison of cropped land area estimates using satellite imagery and airphoto interpretation in Kano State, Nigeria.

	Percentage area	
Location	Satellite	Airphoto
Wudil	74	82
Gaya	58	47
Hadejia	53	53
Kazaure	63	59

Source: Agricultural Remote Sensing

Taylor (1985)

Remote sensing has also been applied in several other projects in sub-Saharan Africa. In Kenya, the main areas of application have been in agriculture, forestry and geology. There were also studies applying remote sensing technique on savanna vegetation in Kenya and Bushland coverage in East Africa (see Lamprey 1985). In Mali, a national remote sensing committee has been set up, accountable to Ministry of Transport. Sudan declared and adopted a policy on the use of remote sensing in conducting surveys, coordination of different organizational activities and training. In Nigeria several projects concerning river basins, valleys and petroleum exploration and uranium mining have been carried out.

Other applications of remote sensing included:-

- Desertification in Kordofan, Sudan (Helden 1984, Olsson 1985b)
- Population census in Sudan (Stern 1985)
- Fuelwood resources and land degradation in Sudan (Olsson 1985a)
- Forestry, land use and soil erosion in Ethiopia (Helden 1987)
- Drought impact monitoring in Ethiopia (Helden and Eklundh 1988) and
- The impact of climate and man on land transformation in Sudan (Ahlcrona 1988).

In the developed world the need is to increase the precision of inventory estimates, to reduce costs and to improve timeliness. Similarly but with a different emphasis, the developing world need is to reduce the overall margin of errors and to be able to produce estimates of cultivated area at least without resorting to complex and time consuming observational procedures. In most developing countries, in particular sub-Saharan Africa, the resource constraints on government institutions enforce limits on the amount of funds, staff and time spent. Basically what is needed in general is mutual cooperation between the two *worlds*. Sudan has gone through one of the most severe drought periods in her recent history. From various international organizations, Sudan received more than 500 million of dollars worth of food aid. As a result, United States Agency for International Development (USAID), contracted Regional Centre for Services in Surveying, Mapping and Remote Sensing (RCSSMRS) in Nairobi, Kenya, to acquire satellite images, obtain aerial photographs of the country and to provide technical assistance for crop estimation survey and analysis of the data (SERISS 1987).

The project was successful on areas where mechanized farming was practised. However in traditional farming areas the project ran up against difficulties. For some crops like sorghum and groundnuts the area estimation by Sudan government differed by a factor of two from that produced by RCSSMRS (see SERISS 1987). Table 3.4 gives the comparison of the estimates.

Table 3.4: Comparison of estimates: traditional agriculture in Sudan

	Planted area		Harvested area		Production ('000 tonnes)		
	('000 feddan)						
	Govt.	ASF	ASF	C.V.	Govt.	ASF	C.V.
Millet	3485	5662	5006	14%	254	221	16%
Sorghum	1243	638	527	29%	137	33	32%
Sesame	1056	1989	1788	21%	74	42	24%
Ground nuts	874	1126	1086	28%	160	135	29%

Source: Sudan Emergency and Recovery Information and Surveillance System (SERISS 1987)

SERISS (1987) observed that the difficulties of estimating production in the areas of traditional agriculture were more complex than previously realized. The conclusion was that estimates from traditional farming require further study, more detailed analysis and much attention in future crop surveys.

Fields in developed countries are mainly pure stand, large and more homogeneous. Climate is usually more favourable, and there is mechanized farming. The successful application of remote sensing which made estimates from scene characteristics less problematic can be attributed to these factors. On the other hand in most of sub-Saharan Africa, fields are not like those of the developed countries. With the exception of Niger, field sizes mostly varied within the range 0.1 and 1.0 hectares (Verma *et al.* 1988). In the United Kingdom (UK) farms of less than $6ha$ fall under the category of statistically insignificant contributions. The remote sensing problems in dealing with traditional agriculture are hence very different from those of the developed countries. One pixel, for example, can be the size of a plot in Africa in a mixed farm which makes it difficult to classify into a single crop type; whereas in developed countries crop cover can be classified for each pixel. This enables digital processing and statistical model fitting like regression of area on number of pixels, to be used straightforwardly in the estimation of crop areas (see Battese and Fuller 1981, Smith and Ramsey 1982, *etc.*). A great disadvantage to the developing countries, is that a computer is needed to record and restore images, even to pictorial form. This, observed Taylor (1985), has meant that only limited number of people have been able to use the imagery in a form other than standard photographic products.

However a wide body of disciplines is already making use of satellite data. With problems in its implementation still present in developing countries, its use is slowly gaining recognition (Alonzo 1988). Malingreau (1981) proclaimed that there was evidence that in conjunction with other approaches, the right use of remote sensing can contribute to a better understanding of the environment and a more successful implementation of integrated projects.

This study does not intend to make use of remote sensing data by area sampling frame (ASF), whereby crops are statistically sampled based on stratified areas because of lack of enough images for the desired segments. A country like Tanzania would need about 45 up-to-date Landsats $185km \times 185km$ images, and assuming the technical facilities are available, complete area sampling frame, as observed earlier according to Helden (1981), would need an estimated time of one to two man-years. Moreover area sampling frame created a misrepresentation for the case of traditional farming in Sudan. One reason was that crops were sometimes replanted when rains were late. In many cases there was a need for additional aerial surveys (SERISS 1987). It is clearly therefore, necessary to define systems where satellite information is appropriately integrated with the existing conventional methodologies for optimum applications. In the next chapter we will review a possibility of incorporating this technology in agricultural surveys in sub-Saharan Africa, taking into account the limitations on available technology, manpower skills and financial capabilities.

Chapter four

The application of remote sensing to sub-Saharan Africa

4.1 Aerial photography and satellite data

Aerial photography has been proposed as an alternative to field surveys by Ottichilo *et al.* (1985) for example. The authors' procedure is based on parallel transects along which vertical photos are taken with a small format camera. Crop area estimates are then derived from each photo by counting the number of points of a sampling grid which fall into cultivated fields. Aerial photographs were taken over the agricultural portions of the districts. The authors stressed that unnecessary photography and aircraft time could be avoided if the agricultural lands were to be properly stratified.

In the estimation of yield, Peden *et al.* (1985) pointed out that the radiometer, which is an instrument for measuring quantitatively the intensity of EMR in a band of wavelengths in any part of EMS, is potentially cost effective when compared with ground-based sampling alone. The authors nevertheless pointed out that it would be advantageous to replace radiometer data with even less expensive and less dangerous methods. Flying at $70m$ above ground level requires very competent pilots with exceptionally well maintained aircraft. Moreover, the efficiency of the method was demonstrated for regions with simple cropping systems dominated by one crop only (maize in this case). The situation is very different for regions of mixed cropping.

Research perspectives are under way for utilizing the satellite sensing technology. It is not that satellite data, sometimes termed remote sensing data, have not been used before. The application of remote sensing has been in great use in the USA and in Canada as well as in Europe. The necessity to pay more attention to rainfed foodcrops (more specifically to cereals) in sub-Saharan Africa, and a realistic approach to constraints related to remote sensed data acquisition and processing, needed a completely different approach.

4.1.1 The implementation of remote sensing

For the objective of regional delineation and estimation of regional surfaces devoted to cultivation, the Landsat MSS's seem quite sufficient. They are the most readily available satellite imagery of less developed countries (King 1982). Land cover characterization generally can be achieved for detailed analysis of regional delineation. Also TM imagery is substantially available (Figure 4.1) and SPOT to a lesser extent (Figure 4.2). It is therefore important to note that it is possible to estimate the total gross horizontal area under cultivation of a country to a relatively small margin of error from the remote sensing data.

4.2 Analysis and interpretation of remote sensing data

Basically there are two methods of analysing remote sensed data, namely visual interpretation and digital interpretation. The latter uses a computer and could be classified into supervised and unsupervised classification. With supervised classification, extrapolation is done from areas with known ground information. There is considerable interaction with the analyst who guides the classification by identifying areas on the image that are known to belong to each category (*eg*. cultivated and uncultivated areas). There are several classification strategies; among the common ones is that of choosing a class that would maximize the likelihood of a correct classification given the information in the training data. The strategy, known as maximum likelihood classification, does not only use the mean or other average values in assigning classification, but also the variability of brightness values in each class. Unless one is prepared to use many sites in a training set, the true variability of pixels within a class will not be reliably estimated from the training set data (Kershaw 1987).

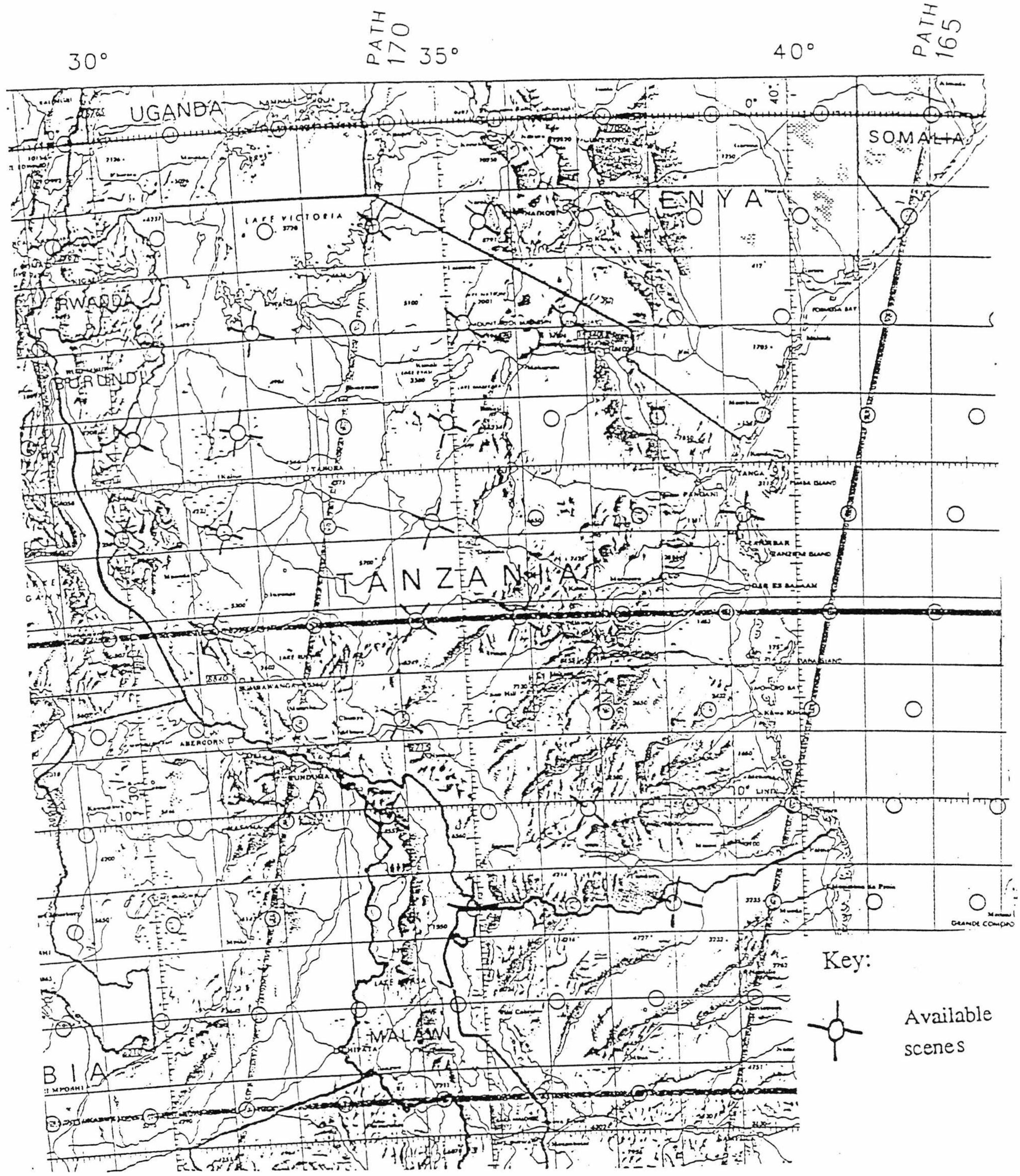


Figure 4.1: Scene coverage with Landsat data over Tanzania

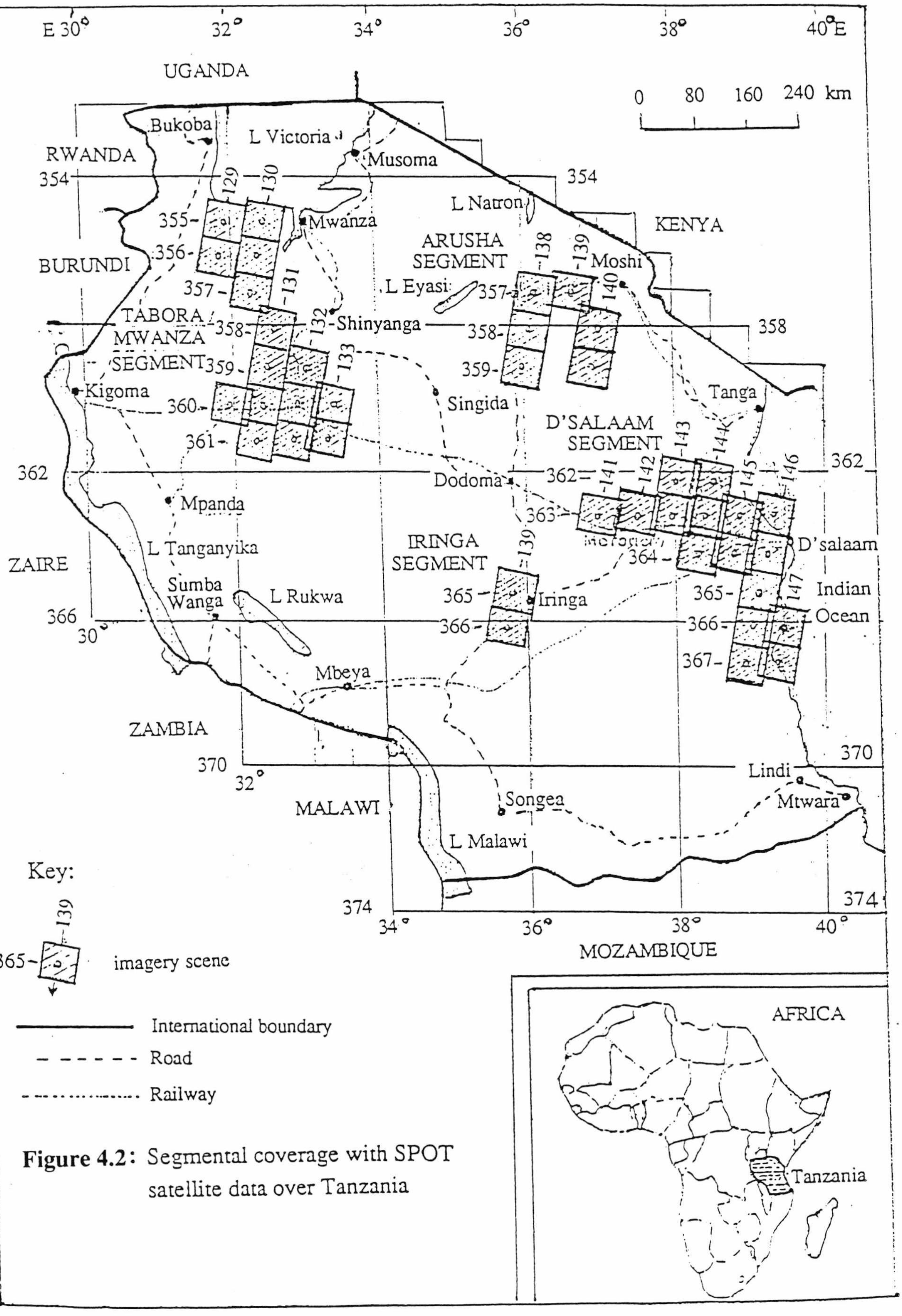


Figure 4.2: Segmental coverage with SPOT satellite data over Tanzania

The unsupervised classification is based on the signatures appearing on the imagery, that is, it uses only the statistical properties of the image as a basis for classification (Sabins 1986). Therefore there is minimal interaction with the analyst in a search for natural groups of pixels present within the image. This method is potentially useful for classifying images where the analyst has no independent information about the scene. The term classifier refers loosely to a computer program that implements a specific procedure for image classification. Although the distinction between the supervised and unsupervised classification is useful, the two strategies are not as clear as they are defined, for some methods like hybrid classifiers share characteristics of both (Schowengerdt 1983).

Generally most remote sensing identification utilizes basic elements of air photo or image interpretation. The stimuli for image interpretation, both computer and human, are indicated on these elements in Figure 4.3. The dotted line in the computer side indicates difficulties in the reverse direction.

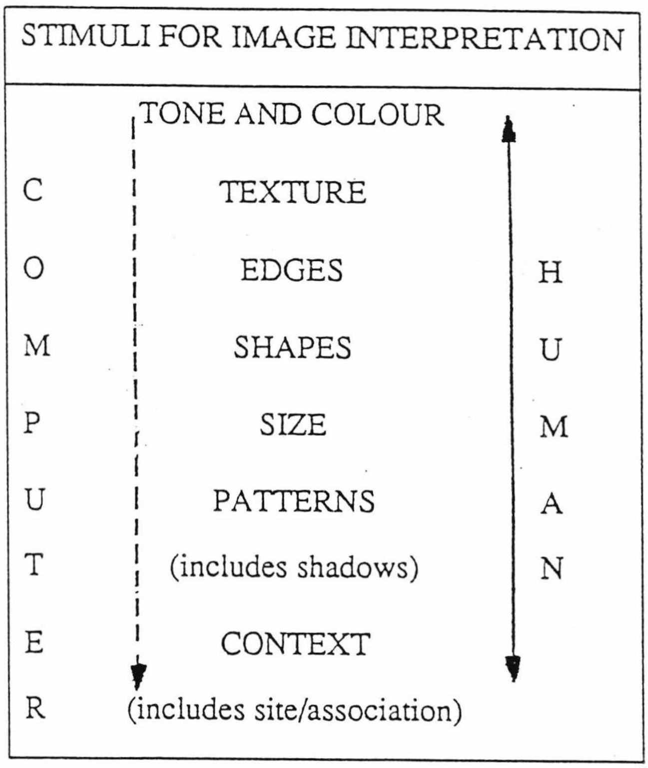


Figure 4.3: Stimuli for image interpretation

Let us consider each of these characteristics in turn.

Tone:

Tone refers to the relative brightness of objects on the photo or image. Depending on the form of the remote sensing data, this tone could be in grey tone or colour. For example, vigorous vegetation is bright red, rangeland is a much darker red, while sparsely vegetated, barren or disturbed ground (may be harvested vegetation) is white to tan. Thus the intensity of the tone could provide an idea on what a particular feature on the terrain could be. However, the same terrain feature will have different tones depending on a number of other variables like phenology of vegetation at different localities, sun angle, interference of the electromagnetic spectrum, photographic production, hydrological conditions, time of the year *etc.* Also complications arise because different features may depict similar tones on the image.

Absolute values of tones in themselves will not identify a feature (King 1984) because feature detection depends upon relative tone values in comparison with other basic and other association elements. Even relative tones may be inconsistent depending on other variables. Tone boundaries are more significant than tone values by themselves.

An example for general postulates for soil interpretation in black and white photos is presented in Table 4.1.

Table 4.1: General guides to soil interpretation in black and white airphotos

Tone	Feature
Dark	Poor drainage e.g. clay also larva and burning.
Light	Good drainage, usually sand. Also sand stone, quartzite, quartz and granite, surface limestone (calcrete) and gypsum.

Table 4.2 gives examples of tone on images for different terrain features in infrared colour composites.

Table 4.2: Generalised interpretation of infrared colour composites

colour	Terrain feature
White or yellow	Soil erosion. Very dry grassland. Grassland or bare cultivation fields with light topsoil. Leafless escarpment woodland.
Light blue/green	Leafless savanna. Grassland, recent cultivation or fallow with dark topsoil. Shallow or sediment-laden water. Towns.
Dark blue	Light old burning
Brown	Partly leafless woodland. Heath.
Pink	Green grassland (moderate cover)
Reddish orange to light red	Woodland. Bamboo thicket Forest with glades. Swamp vegetation Crops.
Dark red	Forest
Grey or greyish green	Leafless woodland with shrub under storey
Black	Intense or recent burning Deep sediment-free water.

Source: Remote Sensing Manual

King (1984).

Shape

Shape refers to the general form, configuration or outline of individual objects. Shape is very important during interpretation and because it is obvious to recognition, it is often taken for granted. Its recognition however is only possible if there are tonal differences which then produce the boundary shape. For example, cultivation is distinguished from bareground patches by its characteristic shape. Similarly airports can be recognized by the characteristic shape of the runways. Also the meandering linear shape of rivers is very characteristic. Isolated clouds can cause confusion. But a closer examination will quickly reveal that the shape of the cloud (normally white in colour) is identical to a black spot (its shadow) occuring at some distance depending on the sun angle. Thus knowledge of the object's shape should help the interpreter to start getting clues on identity of an object.

Position/site

Position or site is the location of objects in relation to other features. Relative vertical position is normally much more important than horizontal position. This element can be very useful during interpretation, for example, a patch of mountain forest remnant could be distinguished from a swamp from its position on landscape. Both features could have the same tone, size and even shape. Also, an artificial dam could be distinguished by its position on a drainage line.

Pattern

Pattern relates to spatial arrangement of objects. The repetition of certain general forms or relationship is characteristic of both natural and man-made phenomena. Pattern is an important feature in remote sensing that is easily recognizable. For example contour farming and landscapes with vegetation growing patterns, like tea, sisal *etc.* are clearly recognized. In satellite images, however, the pattern is depicted by the tone of the objects.

Shadow

The significance of shadow to interpreters is that it both facilitates and obstructs interpretation depending on the circumstances. If we can use the example of clouds again, it is possible to determine the sun's angle by comparing the relative positions of the cloud and its shadow. This is how shadow can facilitate interpretation. On the other hand the landscape features can seriously be obstructed by shadow because it reflects little light and is difficult to discern on photos or images. In a rugged landscape image, the effect of shadow is severe.

Texture

Texture is a product of individual shape, size, pattern, shadow and tone. Texture becomes progressively finer and ultimately disappears as changes are made from large to small scale. On images, a dense woodland will have a different texture from that of an open woodland. The tone of a patch of green woody vegetation may be quite close to that of green and uniform grassland. The latter however may be distinguished by its smooth texture. The woody vegetation will often have a coarse texture.

Other criteria

So far we have attempted to discuss each element of remote sensing data interpretation separately. This however, is done for convenience only. In practice a combination of these clues, to identify a feature presented on a photo or image, is employed. Quite often an interpreter's skill is stretched beyond those outlined criteria. For example one notices on an image a tiny circular feature surrounded by bare ground which may look like it has been overgrazed all around. During interpretation it is probably natural to establish whether or not there are livestock in the area. If the presence of livestock is confirmed, and also that the bare groundness is actually because of overgrazing, then the next step would be to establish why there should be overgrazing only around the particular feature. A skilled image interpreter could guess that the circular feature is most likely an artificial source of water for the livestock in the area. One has therefore tried to identify the feature in the context of what would be happening in the area. These and many other examples are what might be called "contextual interpretation".

In addition to the foregoing example, many other times an interpreter is called upon to use supplementary information for effective identification. For example, an open area surrounded by dense vegetation may not have an obvious interpretation. But comparison of the image and maps of the area may indicate that the open area is a settlement. In many cases the road network may not be fully plotted from an image alone. It is greatly simplified when a map of the area is used. Another example is that during interpretation one comes across a tone which could be either dense woodland or dense bushland. In the absence of ground information, the interpreter is rescued by supplementary information, a topographical map for instance. These maps may indicate predominance of trees or shrubs in the different areas. If the map shows that the area in question is dominated by trees then the tone denotes dense woodland. On the other hand if shrubs dominate the area then the tone represents dense bushland.

4.3 The proposed methodology

The proposed methodology (Figure 4.4), based on the existence of a hierarchy in levels of administrative landscape organization, is intended to utilize remote sensing in the delineation of area land cover and structure as shown by the straight line arrows for the reasons that will be described in section 4.4. The exercise is proposed yearly, which means acquisition of images for different villages every year. However it is considered appropriate to use images of two to three years old. This is valid if there were no disruptive activities like war, famine and projects like the Tanzanian villagization programme (for example 1973-76). The proposed methodology is also hoped, in the near future, to be implemented for stratification of the whole country which would lead to the selection of appropriate samples.

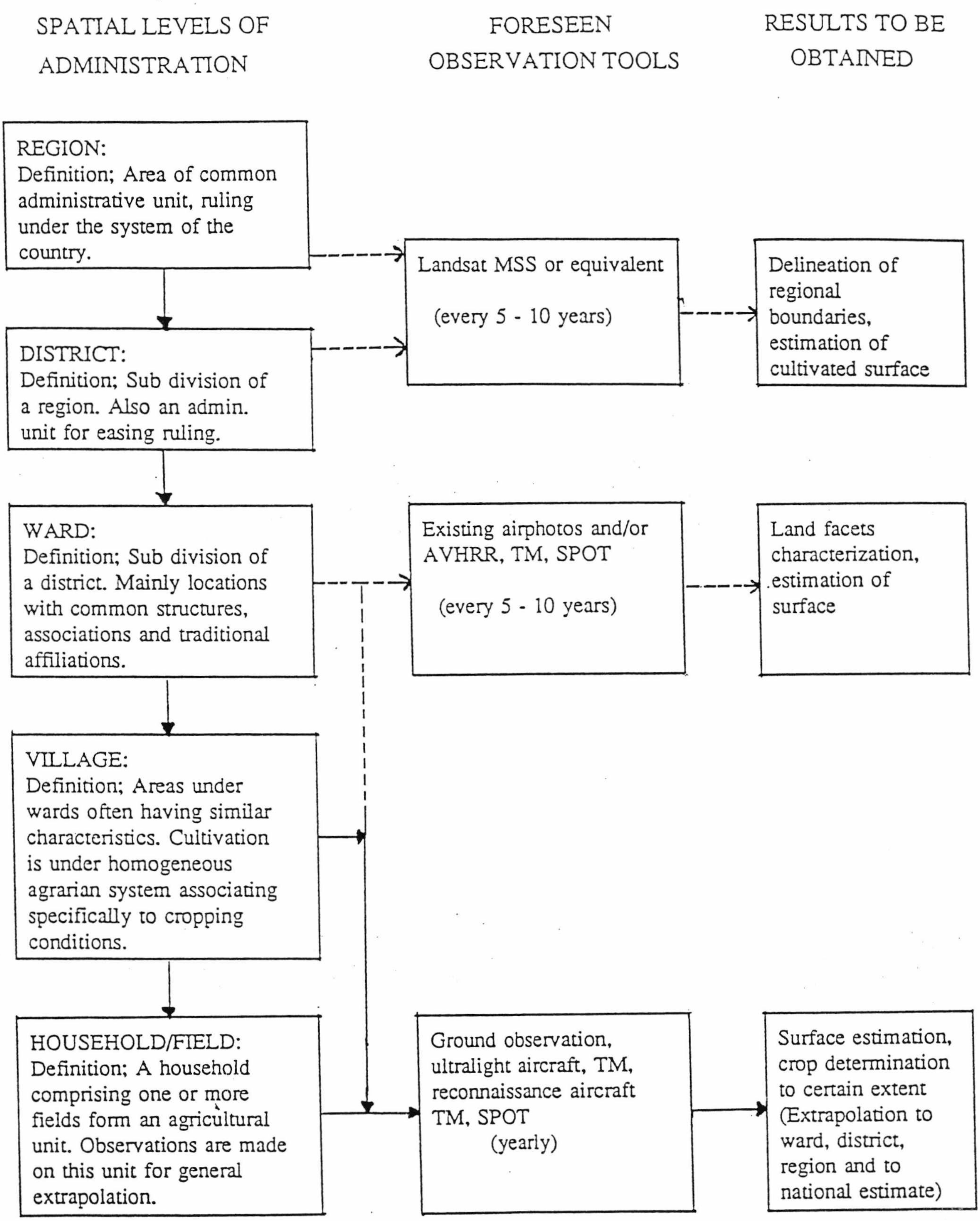


Figure 4.4 : Flow chart of the proposed methodology

The dotted lines indicate such procedures on large areas which would make the stratification possible. Such exercises are proposed to be done after every 5 to 10 years period for some significant multitemporal comparison such as land settlement (including villagization).

First level: Region

The Region is created by a nation for mainly political and administrative reasons. In some instances population growth and distribution might play a role in the creation or formation of these regions.

Second level: District

A district is a sub-division of a region. Although politics may still prevail in the process of establishing these districts, however surrounding structures like mountains and rivers are taken into consideration for the ease of defining boundaries. The population arrangement, with its own culture and tradition emerges. Nevertheless, the divisions are based on actual situations, rather than on potential productions.

Third level: Ward

A ward is a sub-division of a district. It is in this level that human factors, the life of the population with its own culture and tradition are pronounced. Land facet can be characterized by association of specific climatic conditions, land barriers (rivers, ridges, lakes *etc.*) and occasionally cultural techniques.

Fourth level: Village

The question of traditional farming is even more pronounced at this level where two or more villages form a ward. The location of cultivated areas, the type of field pattern, the type of crops and the sequence mode of crop rotation in relation with cultural practices are distinct. Physical features of land form natural boundaries. These features (valleys, ridges, rivers, mountains, forests *etc.*) often restrict accessibility and hence a traditional behaviour and attitude are maintained. The proportion of each ground cover can vary locally. It happens that a given land cover type depending on certain conditions is not occupied by agriculture, either because of low population or because of cultural practices. In order to have better estimates of land representation and utilization in a sample design, several villages should be included in any stratified segment.

Fifth level: Household/Field

The household is the smallest population unit apart from the individual and apart from the non-geometric and irregular landscape structures, its corresponding field/plot is the smallest spatial unit in which homogeneity is hoped to be found, as far as the nature of crops, cultural practices and the resulting yield are concerned.

4.4 Image analysis

When an image is to be utilized it is frequently necessary to make corrections in brightness and geometry if the accuracy of interpretation, either manually or by machine, is not to be prejudiced. Satellite images stored on computer compatible tapes (CCTs) are read into a computer which manipulates the data and the results of manipulations are displayed. The stages of reading, processing and displaying of the images can be performed using a mainframe computer, a micro-computer with graphics or a purpose built digital image processor (Curran 1985). Under the digital interpretation with supervised or unsupervised classification, the required categories of land cover can be estimated. Lack of such facilities will make it necessary to use visual interpretaion.

A simple way to utilize satellite information is to produce photo-like images directly on film or paper and study the images in much the same way as conventional air photos. These images can be standard products or tailored (preprocessed) images, where certain terrain features, details, structures *etc.*, are enhanced to make the interpretation possible.

The biggest advantage with visual interpretation of satellite images is that it can use the same equipment that is used for conventional interpretation and mapping from air photos. The technique is an eye-balling one, where the image, tracing paper and marker pen are the basic requirements. Also needed are a light table and when available, a colour additive viewer (CAV), which makes it possible to generate different colours at the interpreter's convenience (see Figure 4.5). Often when CAV is not available, analysis is performed by the use of the fixed colour balance of an infrared colour composite (false colour composite).

The study intended to estimate the gross area of the country under cultivation and subsequently set ground sample surveys on relatively small samples that would allow the total area to be sub-divided between various crops or crop categories. However there was no possibility of acquiring images to cover the entire country (about 45) which would make selection of several villages possible for ground survey. The study based its evaluation on three out of the twenty regions in the country (Mainland). The choice of the regions is mainly geographical. It was decided to cover different topographical features of the country.

One region was taken in the north inclined east (Arusha region), where it covers the mixed, mountainous lands and the dry grassland features of the low areas. Another was in the south (Mbeya region), which represents the highlands features of the southern part of the country, while the last one was in the north west (Mwanza region), which encompasses the scattered cultivations and the rocky features surrounding the big lake basin areas. Figure 4.6 shows the map of Tanzania with the selected regions (shaded).

Arusha region

In Arusha region agricultural land occupies approximately 5% of the total land area; grazing land occupies approximately 85%. The remainder is montane forest, wildlife reserves and lakes (RCO 1981). The total area of the region is about 82,000*sq km*. The dominant economic activities are subsistence herding and rainfed agriculture.

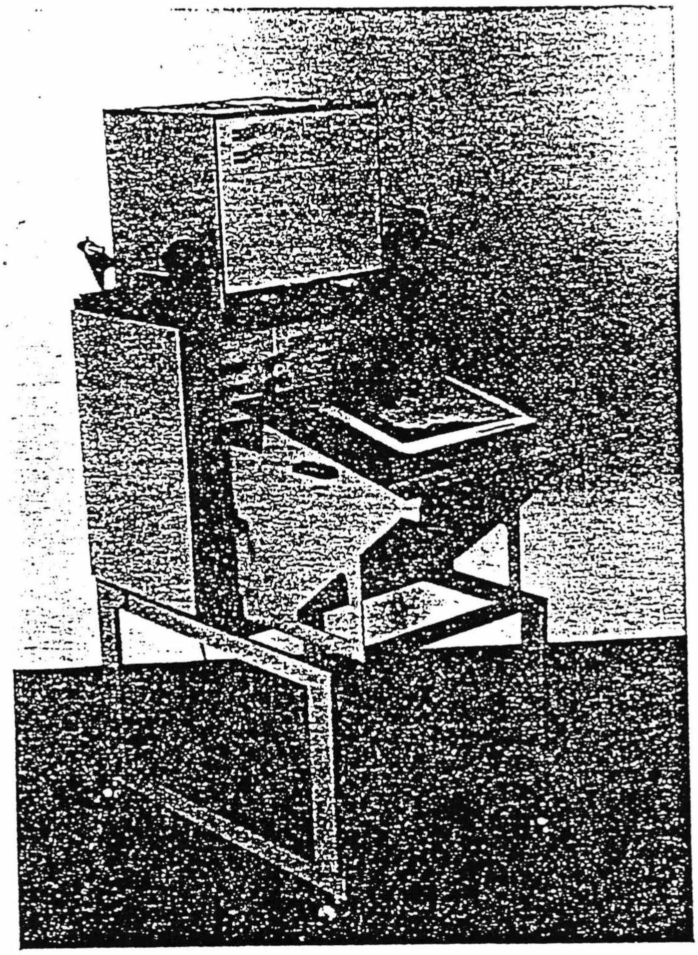


Figure 4.5: International Imaging Systems CAV (colour additive viewer) (King 1984)

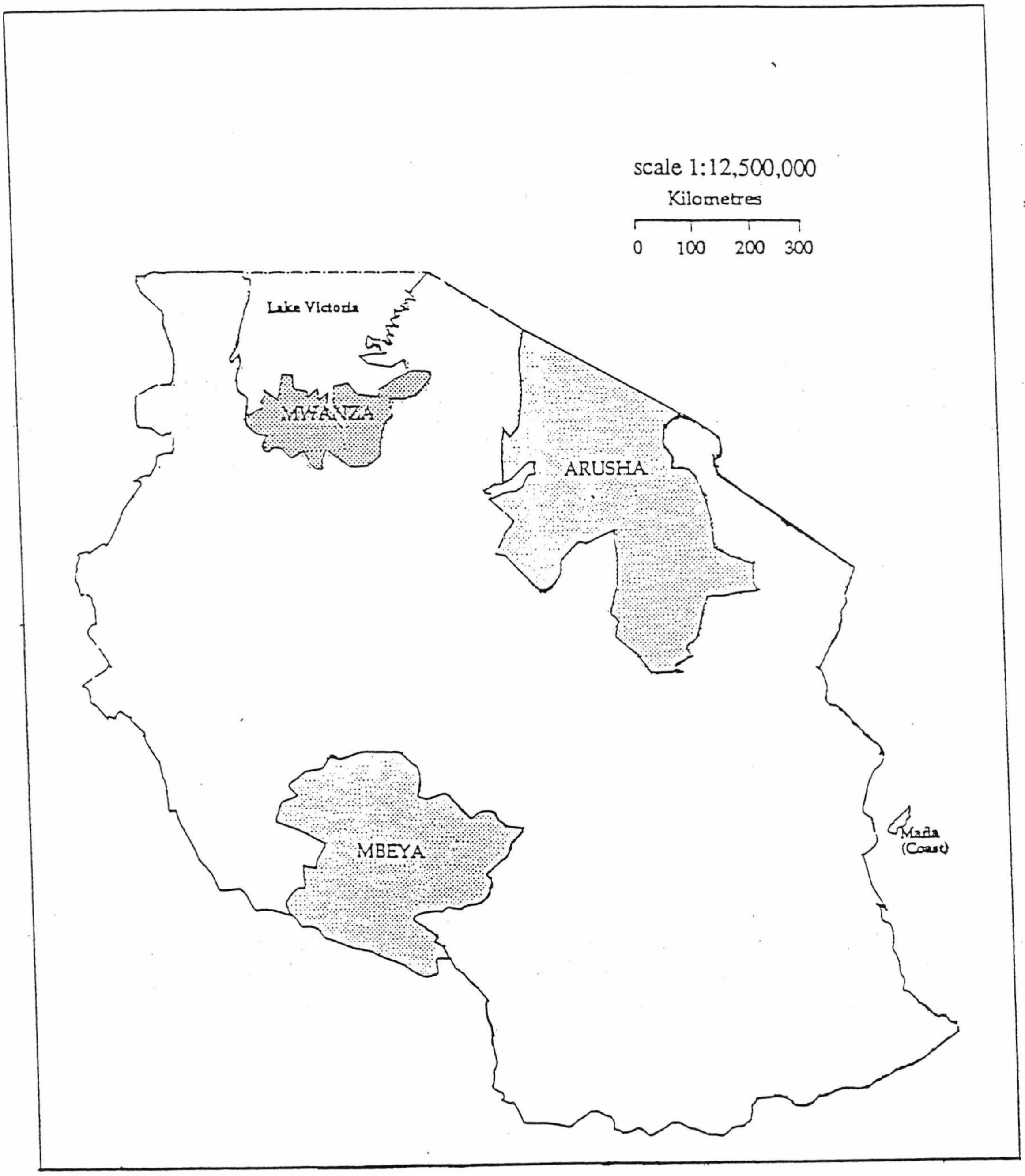


Figure 4.6: Selected regions for 1990 field survey, Tanzania

(Mussa Field Survey)

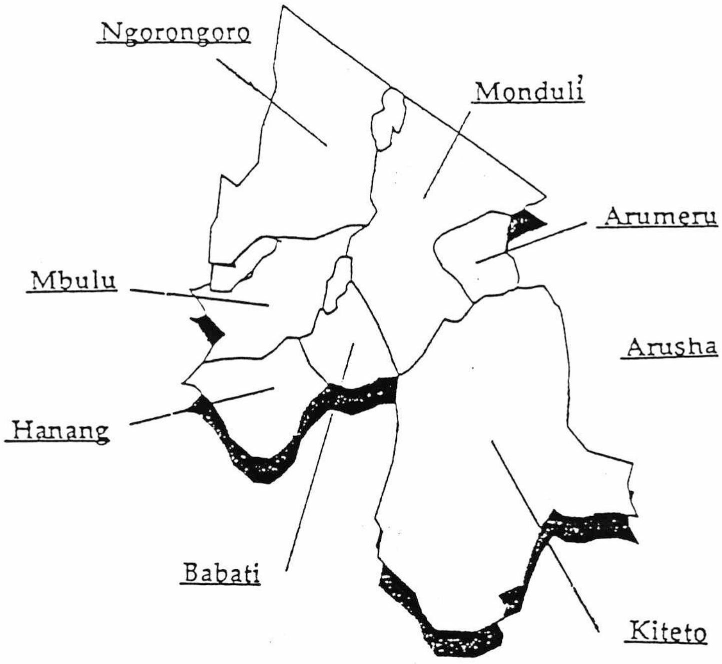
The region is divided into eight districts as can be seen in Figure 4.7a. The major food crops grown are maize, wheat, millet, paddy, sorghum and beans. Coffee is the main cash crop grown in the region. More than 80% of maize is produced in the four districts of Babati, Hanang, Kiteto and Mbulu. About 80% of the wheat is from Hanang, while Mbulu and Arumeru account for the remainder. Kiteto district leads in the production of millets followed by Babati. The other districts growing millet are Monduli, Mbulu and Arumeru. Paddy is produced mostly in Babati. Only Babati, Kiteto and Mbulu produce a small quantity of sorghum. Beans as a subsistence crop are often interplanted with other crops. Wheat is also widely grown by the national food corporation and private estates. The region has potential in tanzanite mining.

Arusha town, which is about $300km$ from Nairobi (Kenya), is situated along a strategic major road from Dar es salaam to Nairobi. The great north-south-road from Cairo (Egypt) to Capetown (South Africa) also passes through Arusha town. It is the centre of the northern tourist attractions where most of Tanzania's game and national parks are found as well as a host to various national and international meetings and conferences. It is flourishing with several light and consumer goods industries, ranging from textile milling to car tyres and electronics.

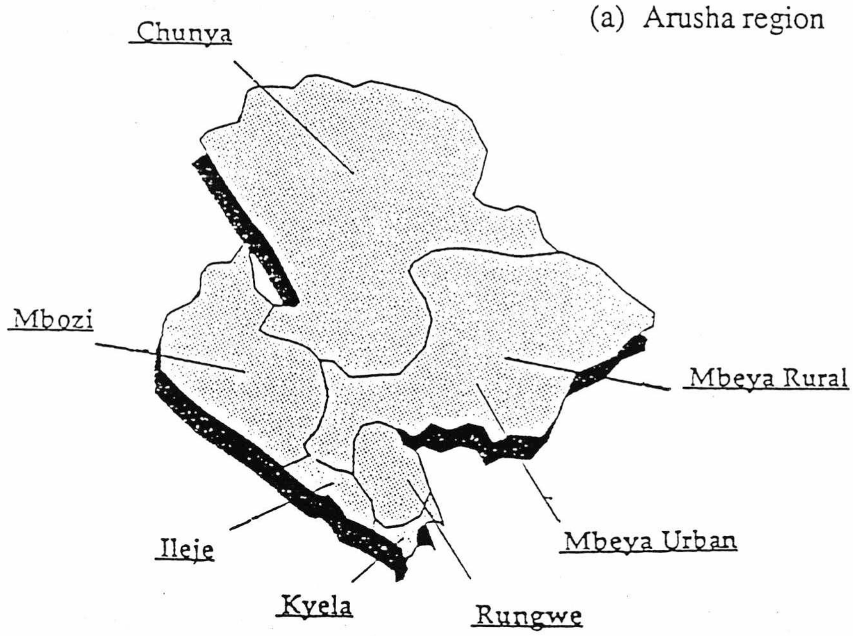
Mbeya region

The region is divided into seven districts as indicated in Figure 4.7b. The total area of the region is about $60,000sq\ km$. The major food crops grown are maize, paddy, beans sorghum, round (Irish) potatoes and bananas. Millet, sweet potatoes and wheat are also grown but on a small scale. The cash crops grown are coffee, tea, pyrethrum and cotton.

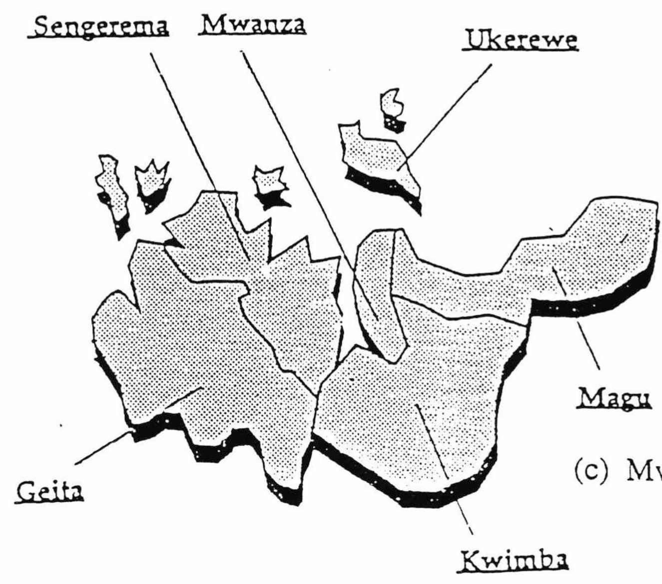
Maize is the main food crop which covers almost half the total area under food crops. Mbozi and Mbeya Rural districts represent the greatest share of maize production. Apart from a few locations, maize is always interplanted with other crops. This creates problems in area measurements, especially when the pattern of farming is complicated.



(a) Arusha region



(b) Mbeya region



(c) Mwanza region

Figure 4.7: Districts of the selected regions

For example, in Kyela district, farmers usually grow maize in boundaries of their paddy farms as the demarcation of their farms (MALD 1989).

However estimation of cultivated land is important as land under cultivation is expanding in Chunya, Ileje, Mbeya Rural and Mbozi where there is vacant or empty land. In other districts like Kyela and Rungwe where there is scarcity of land, area measurement is also important and even more difficult because of changing farming patterns for improving production. Paddy is grown by traditional farmers, but it is also grown in large quantity by the national food corporation as well as private estates mainly in Mbeya Rural. Sorghum is a subsistence food crop which is also used for brewing. Cultivation of wheat is limited to a number of rainfed areas in the highlands. The region has substantial quantities of gold, coal and iron ore deposits.

Mbeya town is at the crossroad to Zambia and Malawi from Dar es salaam.

Mwanza region

Mwanza region is along the southern shores of Lake Victoria, consisting of six districts as shown in Figure 4.7c. The main food crops grown in this region of about 20,000*sq km* are maize, sorghum, paddy and cassava. Along with these crops, cotton as cash crop is widely produced. For example, in 1989 cotton contributed to the country about 16.5% of the foreign exchange earnings of which Mwanza region supplied about half the production.

Cattle rearing is another extensive agricultural activity of the region. Fishing is also a major activity because of the proximity of the lake. The geological structure of the areas surrounding the lake make gold and salt mining essential businesses of the region.

Several light manufacturing and consumer good industries operate in the town of Mwanza. Industries in textile, bottling, fishnets manufacturing, soap and chemicals flourish with many transactions across the national borders. The town is a centre of road transport to Musoma (near Kenya), Bukoba (near Uganda) and Karagwe (near Rwanda). It also provides landing and dock facilities for ships and steamers plying Lake Victoria to the Uganda and Kenya ports of Jinja and Kisumu respectively. It is also a major railway station along the central railway line to Kigoma (near Burundi and Zaire). The railway line is a main link between this town and Dar es salaam, the capital.

The selected regions represent about 20% of the entire country's land. It was aimed to have two or more villages in each of these regions. As it was difficult to obtain readable images on prearranged dates and because of the constraints of acquiring ordered images for a particular location, it was thus necessary to work on the available images. Therefore the villages in these regions were chosen because of the availability of the images.

The available images were provided in analog form (pictorial prints $40cm \times 50cm$) from positive negatives, already enhanced for user's direct application. The images, obtained from Regional Remote Sensing Centre in Nairobi, Kenya were:

1. Mbeya; Landsat TM-image path/row:169/066 Date:2/8/84

Interpretation of this image has not been difficult because the area has been imaged just after harvest for many fields. This has enabled the identification of boundaries of the building structures and permanent forests, also the road and railway paths were outlined distinctly.

2. Arusha; Landsat TM-image path/row:168/062 Date:17/10/88

The scene has been imaged much later after harvest, by that time burning was immense in many parts of the village. Features like valleys, small rivers and cattle paths which normally make village boundaries were difficult to identify. However the forests on the high mountain slopes and the coffee plantations were easily identified.

A SPOT mosaic-image path/row:129/356 and 130/356 of 2/6/88 for Nyankumbu village, in Geita, Mwanza region, was made available for interpretation and analysis only within the premises of the Institute of Resources Assessments of the University of Dar es salaam. This scene has been imaged during the time of late harvest. The nature of the area, that is, scattered cultivation with bushes and trees has made easy the identification of features like roads, buildings and rivers.

4.4.1 Area measurement

A basic problem in the analysis of remote sensed data is to estimate the proportion of a given scene devoted to a particular land cover type. The usual approach to solve this type of problem digitally, using a computer, is to design a classifier for identifying the land cover of interest. Once the scene is classified, the number of pixels it constitutes becomes the desired estimate of the land cover of interest. The Large area crop inventory experiments (MacDonald and Hall 1980) are an example of the approach. The difficulty due to crop classes overlapping on the experiment led to a mixture model approach (Lennington *et al* 1984 and Chhikara 1984, 1986). The mixture model is a mathematical statement that the overall distribution of feature values for a set of data can be characterized as being equal to a linear combination of simpler distributions corresponding to a class of interest. From the problem experienced on scenes of developing countries (section 4.5.1) even the mixture model seems to be difficult to apply for a particular crop of interest. In view of that, this study employed manual classification based on visual interpretation.

4.4.2 Visual interpretation

The choice of a method to be adopted is influenced by a number of factors.

Some of them are:-

- objective of the study
- location of the study area
- type of data available
- level of detail required in the study
- manpower available.

While it is stressed that remote sensing is most effective in developing countries where the data are sparse, these countries are least able to afford sophisticated equipment. The lack of facilities in Tanzania made participation in digital analysis difficult (Dunford *et al.* 1983). The same is true for the majority of sub-Sahara African developing countries. Hence, often the only alternatives are, as King (1981) put them, to rely on analysis by foreign experts, who are sometimes little aware of surrounding conditions or, to take the less expensive but usually more productive option of visual and intelligent interpretation. The rather high resolution of the image products, however, favoured the development of computerized data analysis, based on overseas technical and human resources which somehow becomes too complex for practical applications in the developing countries. The local knowledge and competence concerning the land and its resources in most cases were unused in the development process. Many studies so far have been too academic and have had a limited influence on the practical applications and the development process (Stromquist *et al.* 1988). This has also been realized, for example by Dunford *et al.* (1983) in their use of remote sensing techniques for land use information for rural development planning in Arusha region, Tanzania, where they could easily utilize the facilities of digital image analysis of laboratory at University of Arizona, but they were persuaded by two factors to rely on visual interpretation. First, there was the cost, time and uncertainty of obtaining acceptable, deliverable products from digital analysis. Secondly, the regional staff needed to participate in the information interpretation to facilitate their use of the results in planning activities. This is a great advantage of using visual interpretation of satellite images; the possible use of the local competence in the country of study instead of equipment and staff at overseas.

Stromquist *et al.* (1988) summarized some African studies with their respective authors using visual interpretation of satellite data as shown in Table 4.3. Arguments for using visual interpretation of satellite data put forward by these authors included low cost, rapid access to achieve accuracy and involvement of local staff.

Table 4.3: African examples of studies using visual interpretation of satellite data

Author	Year	Aims	Image type	Area
Stromquist	1976	Land systems, soil erosion	B&W images of individual spectral bands	Central Tanzania
FAO	1977	Landsystems, soil erosion	B&W,FCC images	Morocco
Johansson & Stromquist	1978	Vegetation and geomorphology	B&W images of individual spectral bands	Central Tanzania
King	1982	Land systems, soil erosion	FCC multitemporal images	Tanzania
Haack	1983	Vegetation studies	FCC images	Swaziland
Dunford et al	1983	Rural dev. planning	FCC image	Tanzania
Edwards et al	1983	Grassburns	B&W, Band 5	South Africa
Zietsmann	1984	Land-use	FCC images	South Africa
Larsson & Hanson	1985	Water dev.	FCC bitemporal images	Ethiopia
Stromquist et al	1986	Soil erosion	FCC bitemporal images	Lesotho
Mushala	1986	Land systems	Single band transp FCC multitemporal images	Tanzania

Key:

B&W	Black and white
FAO	Food and Agriculture Organization
FCC	False Colour Composite

Source: An evaluation of the SPOT imagery potential for land resources inventories and planning; Stromquist *et al.* (1988).

In order to make effective and rewarding use of remote sensing techniques in developing countries in the near future, the ambition should therefore be to introduce and apply methods which have an immediate impact and rapidly become operational.

The relevance and functional efficiency of a certain technology has to be judged in local circumstances. In the specific case of expensive information systems technology the question of cost/benefit comes immediately to mind. The large amount of foreign currency or binding international loans may indeed be directed to more short term economically justifiable alternatives. Even if such equipment is donated, it should always be remembered that acquisition cost is not the only cost of equipment ownership. Equipment maintenance and depreciation are ongoing costs that have a way of exceeding a budget.

Costly equipment is difficult to justify unless it does a job that cannot be done any other way, or actually produces results more quickly and accurately than other methods. High volume of work is often an essential requirement for cost effectiveness. Hence the introduction of regional centres for remote sensing in Ouagadougou, Burkina Faso and in Nairobi, Kenya is a healthy centralization measure for sub-Sahara African developing countries, whereby many projects might utilize the well maintained equipment at appropriate times and within budgetary limits.

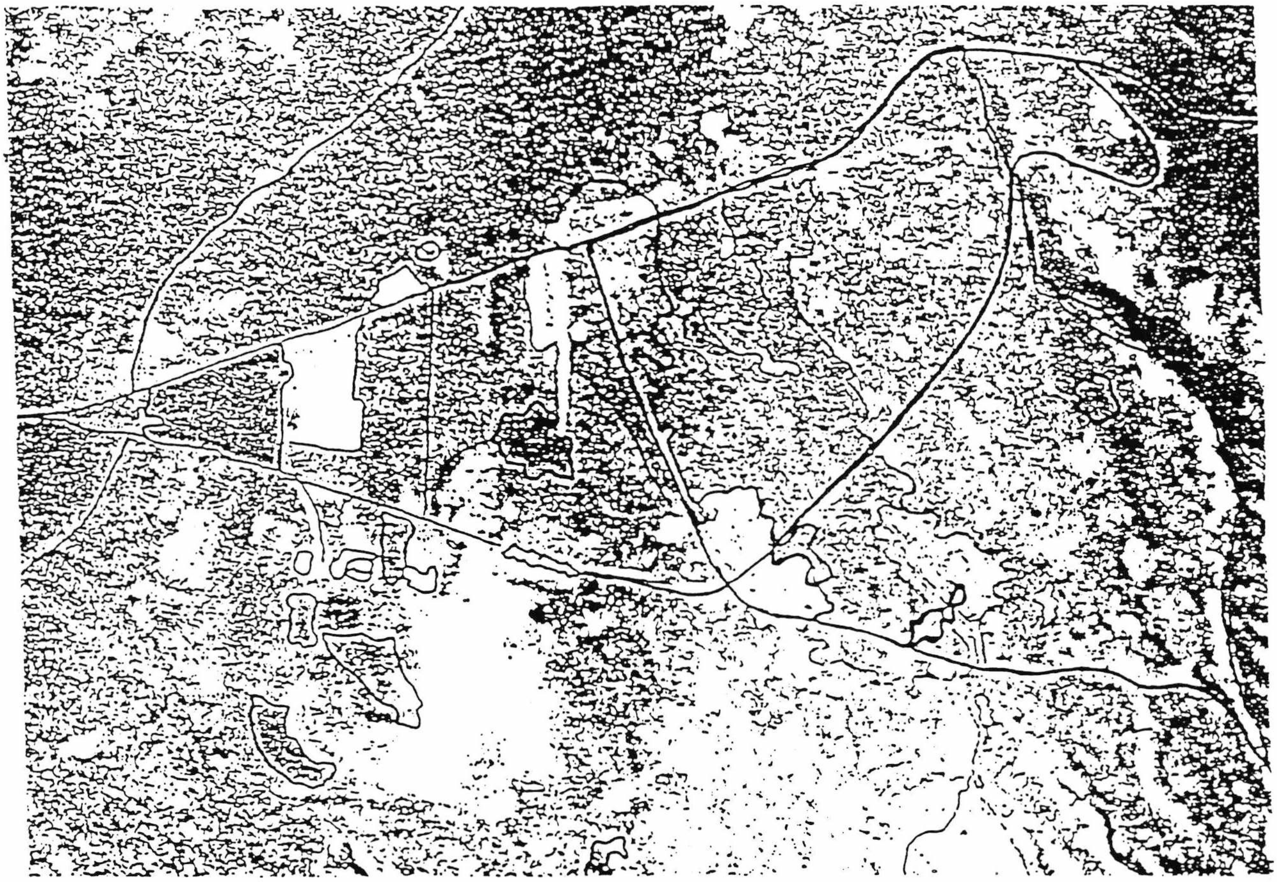
However in a decentralized perspective, very few users will ever have the opportunity to work with digital data processing equipment. Although they should certainly be aware of the existence of the equipment and its application, it still seems that in order to make effective and rapid operation, visual interpretation is very useful. Human interpreters should therefore be depended upon, when they can perform a task without sophisticated instruments and that task arises infrequently. Hence the technical approach should rely, to a high degree, on existing instrumentation and facilities, and it must be capable of being used by professionals from different fields after only a short period of training. This experience equips them with a new dimension in planning and enables staff at district, regional or central level to identify problems and potentials and to reach their own decisions on necessary actions to be taken. Obviously requirements and capabilities will change according to places and circumstances. In view of that, for the moment the visual analysis method seems to be more realistic and readily applicable.

4.4.3 Image interpretation

The manual classification of land cover type, as an alternative approach to the ones mentioned, does not consider a particular crop of interest but the total cultivated land of known boundaries clearly delineated. From the (analog) images, land cover types seen as combination of colour, texture and pattern were delineated on clear overlays. Local knowledge on uncultivated bushes, quarries, collection of houses in a village *etc.* and interpreters' experience on confusion of tones, texture *etc.* made use of labelling those delineated surfaces. The TM delineation for Igawilo area in Mbeya is shown in Figure 4.8a and topographical map on Figure 4.8b, both at the same scale of 1:50,000. A key for general features for all topographical maps is given in Appendix IV. Figure 4.9a at the scale of 1:100,000 and Figure 4.9b at the scale of 1:50,000 show image delineation and topographical map for Monduli area in Arusha respectively. As explained earlier, the region has been imaged at the late dry season where burning is more pronounced (black, in the false colour image), and the just cleared land from burning, that makes bare cultivation fields (white). (The whites on the top centre are clouds). The SPOT image delineation for Nyankumbu area in Geita district, Mwanza, is shown in Figure 4.10a with its respective topographical map on Figure 4.10b at the scale of 1:50,000.

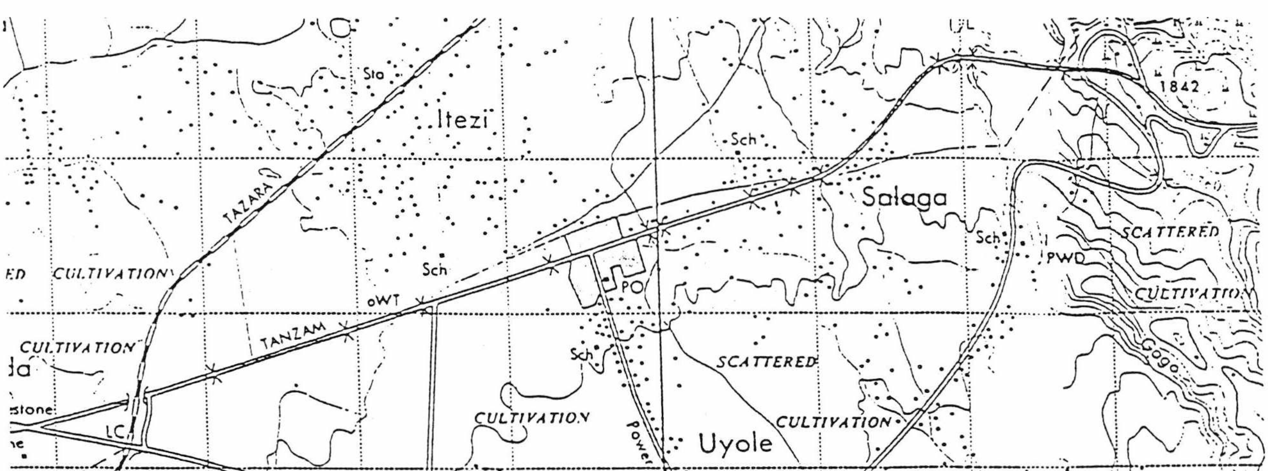
At least three methods can be employed to obtain area measurements of the desired cultivated land in a village.

1. Using Planimeter
2. Scaled dot grid
3. Cutting and weighing of photo area.



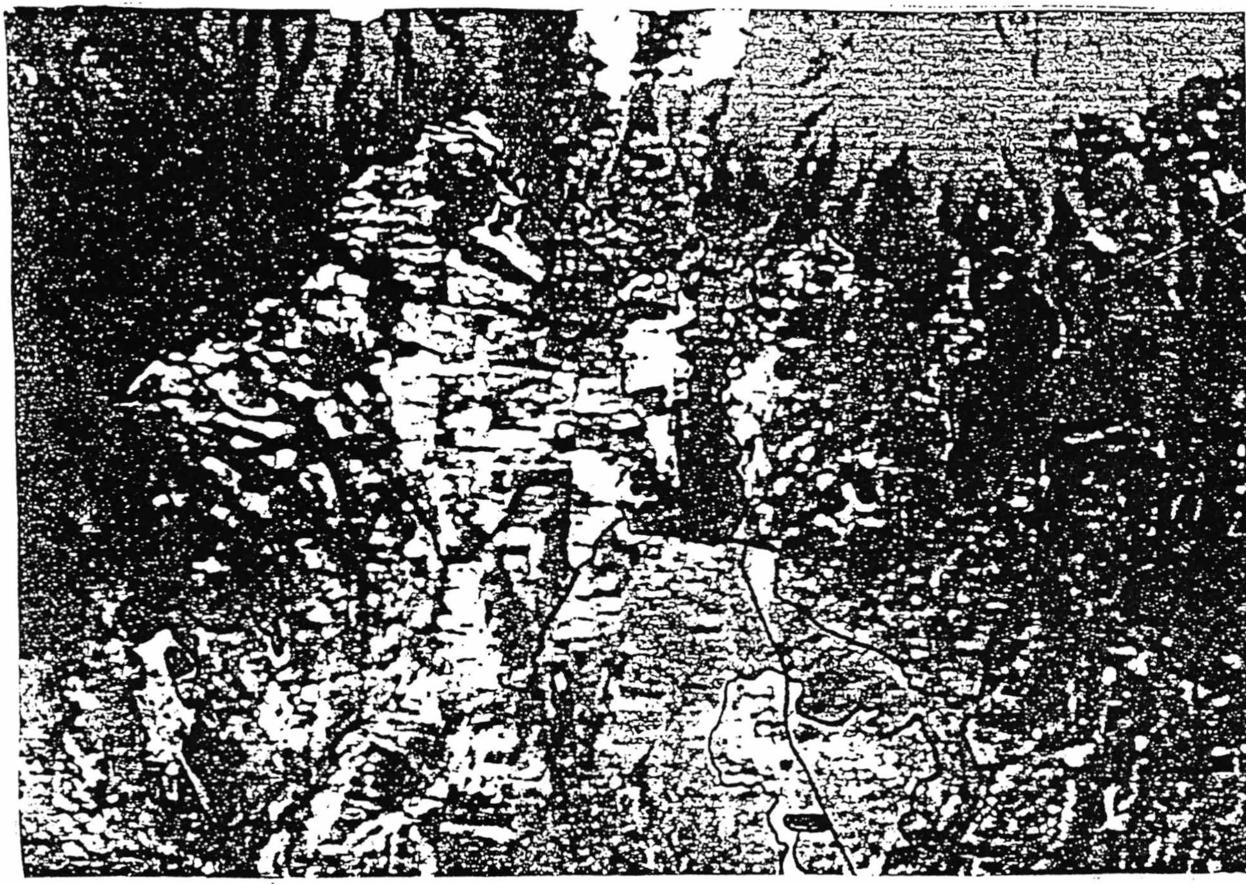
scale 1:50,000

Figure 4.8a: Image for Igawilo village, Mbeya region, Tanzania



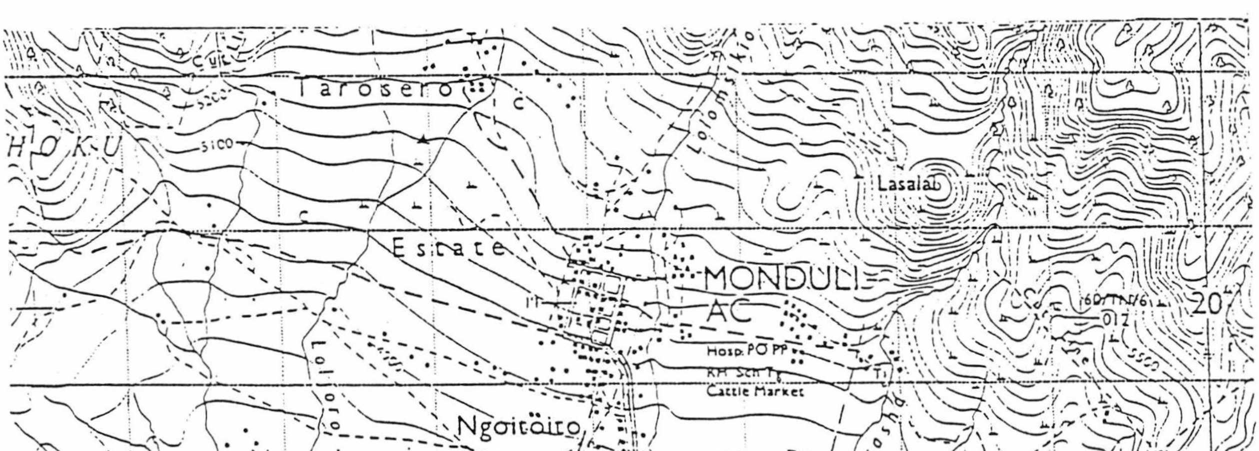
scale 1:50,000

Figure 4.8b: Topographical map for Igawilo village, Mbeya region, Tanzania



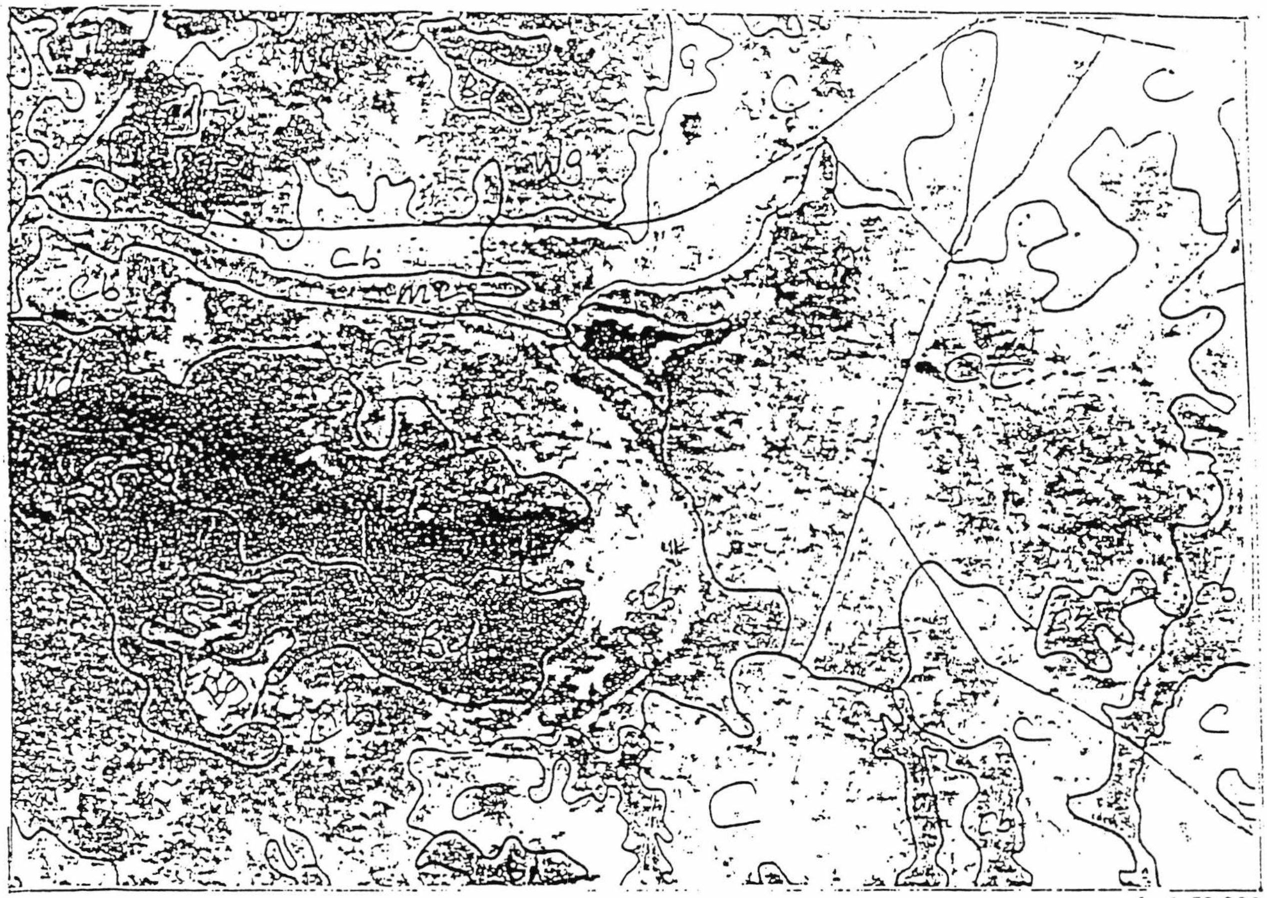
scale 1:100,000

Figure 4.9a: Image for Sinoni and Ngarash villages, Arusha region, Tanzania



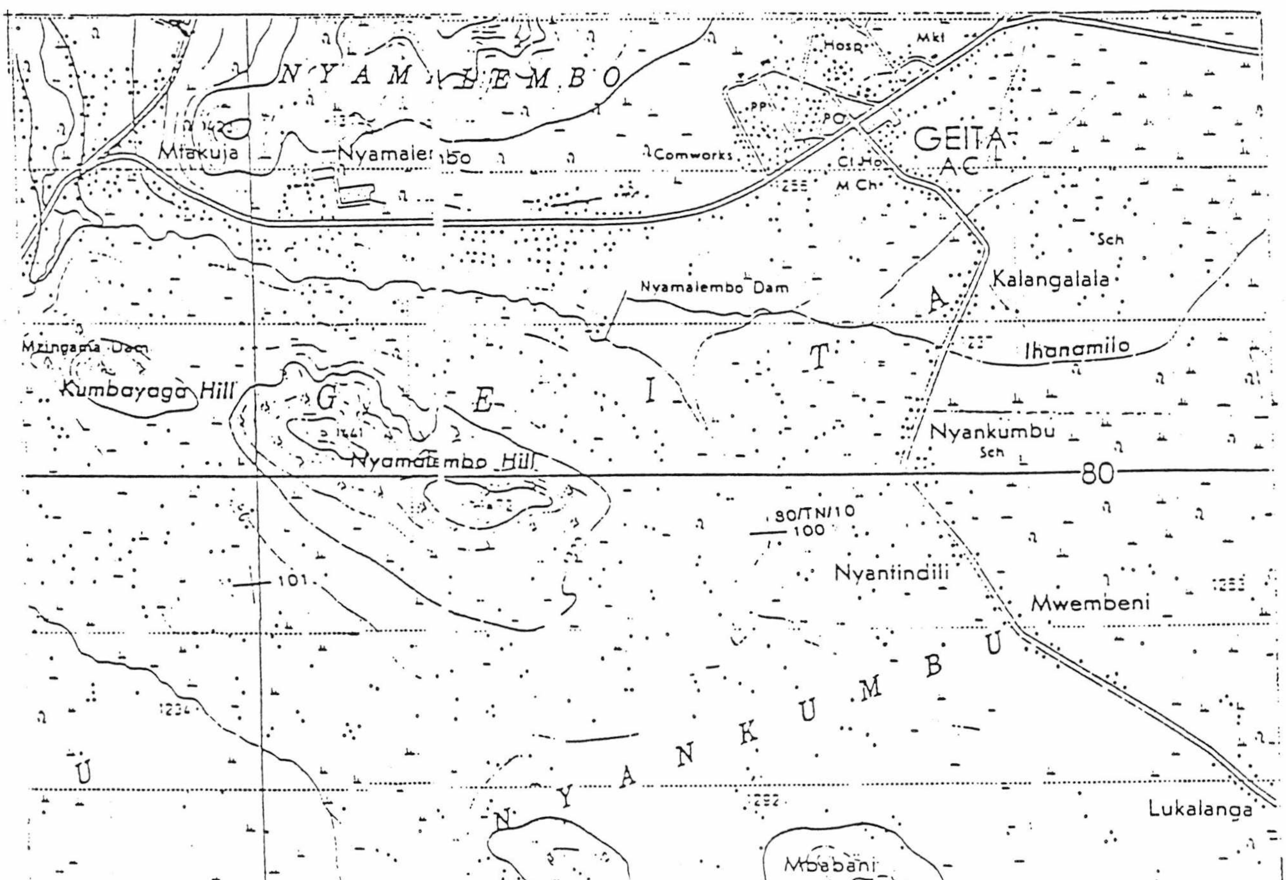
scale 1:50,000

Figure 4.9b: Topographical map for Sinoni and Ngarash villages, Arusha region, Tanzania



scale 1:50,000

Figure 4.10a: Image for Nyankumbu village, Mbeya region, Tanzania



scale 1:50,000

Figure 4.10b: Topographical map for Nyankumbu village, Mbeya region, Tanzania

The planimeter method is the most accurate of these three. There are different kinds of planimeters, from the simple and less expensive to the complex and expensive. The scaled dot grid is the fastest, though less accurate than the planimeter, but is particularly useful for calculating the proportions of two or more types of terrain features (such as agricultural and non-agricultural land) of a large geographic area. Another method is to cut out areas of a particular land use on an image or a map and accurately weigh them. Comparison of these weights with that of a known area cut from the same imagery or a map will allow individual areas to be calculated. The method is delicate and requires highly sensitive scales for weighing the cut pieces. Accurate cutting is also difficult, essentially if boundaries are neither straight lines nor smooth curves.

In this study because the planimeter was not available, the scaled dot grid has been used instead. The method consists of a transparent grid divided into square spaces with dots depending upon the type of sampling to be done and the degree of accuracy desired. The dot grid is placed over imagery about the delineated sampled area and dots falling on each type of required terrain are counted. This is easily done by recording the number of complete squares within, plus the number of dots for incomplete squares. Each square (and therefore dot) has a certain value depending upon the scale of imagery. The dot area grid permits area estimates of acceptable accuracy to be made in one third to one sixth of the time required for planimetry. An example for Igawilo village in Mbeya, is given below for the calculation of its cultivable area.

Example

Figure 4.11a shows the village delineation, the collection of houses (H), the clay quarry (CQ), school and dispensary compound (SD) and the sacred forest (F) have to be eliminated from the village area. Figure 4.11b portrays the desired area to be measured superimposed on the scaled dot grid.

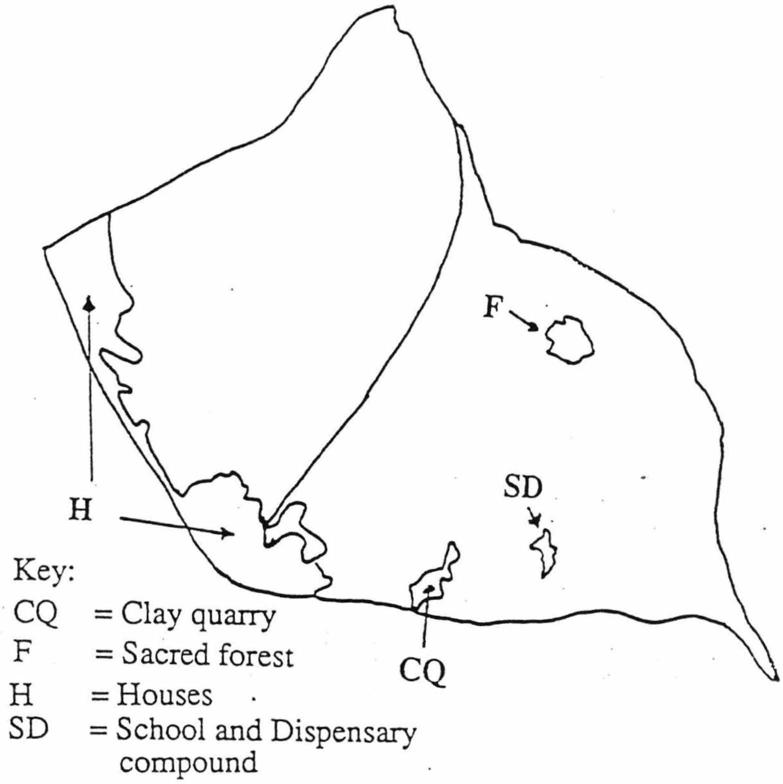


Figure 4.11a: Igawilo village, boundaries delineated

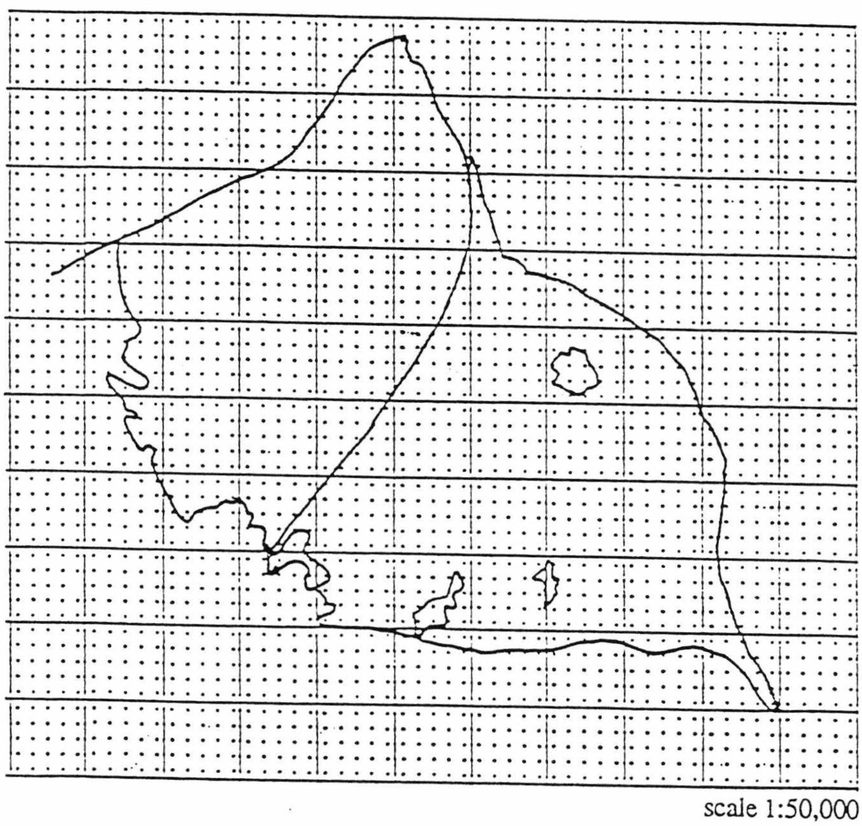


Figure 4.11b: Igawilo village, desired area

The number of squares which are completely within the boundaries of the desired land to be measured is 21. The number of dots where only part of a square is within the boundaries is 413, which make 16.52 squares (since each square in this grid has 25 dots). Therefore total number of squares is 37.52. Each square represents $0.25\ sq\ km$ from the scale 1:50,000 of the image. The desired area is

$$37.52 \times 0.25\ sq\ km = 9.38\ sq\ km = 938\ ha,$$

which is the total cultivated area of the Igawilo village.

4.5 Some disadvantages of satellite images

Remote sensing does not seem able yet to allow detailed observation at the field level, but can outline groups of fields, particularly with the use of TM and SPOT under the conditions of traditional agriculture as observed in the sub-Saharan developing countries. If we consider, for example, the situation of the fields which spread on the slopes, Bartholome (1986) argued that such fields are very significant by their surface as well as by their yield, which is considered by the peasants as very good when compared with less "complicated" locations.

It is possible to construct a complete area sampling frame for all areas of a country covered by topographical maps and aerial photos with an acceptable reliability and also stratification based on land use and interpretation of satellite FCC. But as was seen in the case of Sudan in chapter three (section 3.6), there can be some problems. The basis of any interpretation is the formulation of the exact aim of the study. This study focuses its interest in traditional agriculture at household/field level. For that purpose remote sensing can be considered as an effective operational tool for crop statistics determination in the categorization of locations (groups of fields). That ensures overall representation of those fields which for one reason or another do not appear in the listing frame. Bartholome (1986) noted that temporary fields scattered throughout the savanna are often insufficiently observed, because their a priori importance and location are not known by classical approaches.

While investigating the feasibility of using remote sensing techniques in surveys one should consider the related economic and social aspects of doing so. From the start, expectations have been high that this new technology may provide quick estimates in terms of both area and production. Satellite technology is highly sophisticated and requires large and advanced computer resources in order to handle the recorded data. The rather limited high-tech resources in the sub-Sahara African developing countries must therefore be considered in the application and practical use. There are few advanced image analysis systems available and also very few trained local experts to handle satellite data. It is difficult to foresee any drastic improvement in the present situation in the near future (Stromquist *et al.* 1988). For these countries, the primary basis for economic development lies in their natural resources. Yet, these nations with very few exceptions, do not have thorough knowledge about the nature, quantity and locations of these resources, to harness them effectively for the welfare and progress of their people. A large percentage of these is still inadequately mapped (Ouedraogo 1980).

On the other hand, it is obvious that these countries will have much to gain from remote sensing technology in many applications. Most countries have yet to determine the extent and condition of their arable land, forest, rangeland and water resources. One of the crucial elements at the present time of world food and energy shortages and of spreading environmental deterioration is the need to forecast and estimate crop production, to detect land degradation and to recognize alterations in land use. Remote sensing has proved a great assistance, as indicated in the previous chapter in knowing the threat of desertification (Hellden 1984, Olsson 1985b), population concentration (Stern 1985), land resources, use and degradation (Olsson 1985a, Hellden 1987, Ahlcrona 1988) and drought impact monitoring (Hellden and Eklundh 1988).

At present, aerial photography is by far the most widely used technique in most developing countries. The main major advantages of conventional aerial photography are:-

- the high resolution
- the wide choice of methods
- the availability of instruments and trained personnel, *etc.*

The disadvantages include:-

- the uneconomic repetitive coverage
- the relatively high cost per km^2 compared to satellite images (Pacheco 1980).

The satellite imagery has certain advantages when compared with aerial photography. The provision of a synoptic view has been one of the biggest advantages of satellite sensing. Before the advent of satellite technology, air photo mosaics were made to provide a limited, synoptic view of an area. With regards to type identification the satellite technology brought in an added advantage of multiband differentiation. Different parts of the electromagnetic spectrum behave differently when they interact even with the same feature on the landscape, a concept of scene characteristic explained in chapter three (section 3.3). Similarly a particular part of the electromagnetic spectrum will behave differently depending on the particular aspect of the environment that it encounters on the landscape. This phenomenon is best illustrated by the spectral reflectance curve in Figure 4.12. The phenomenon has been capitalized upon to effect type identification of land cover resources. The repetitive coverage of images for a particular area provides the possibility of multitemporal analysis of images making it possible to monitor deforestation among many other environmental trends.

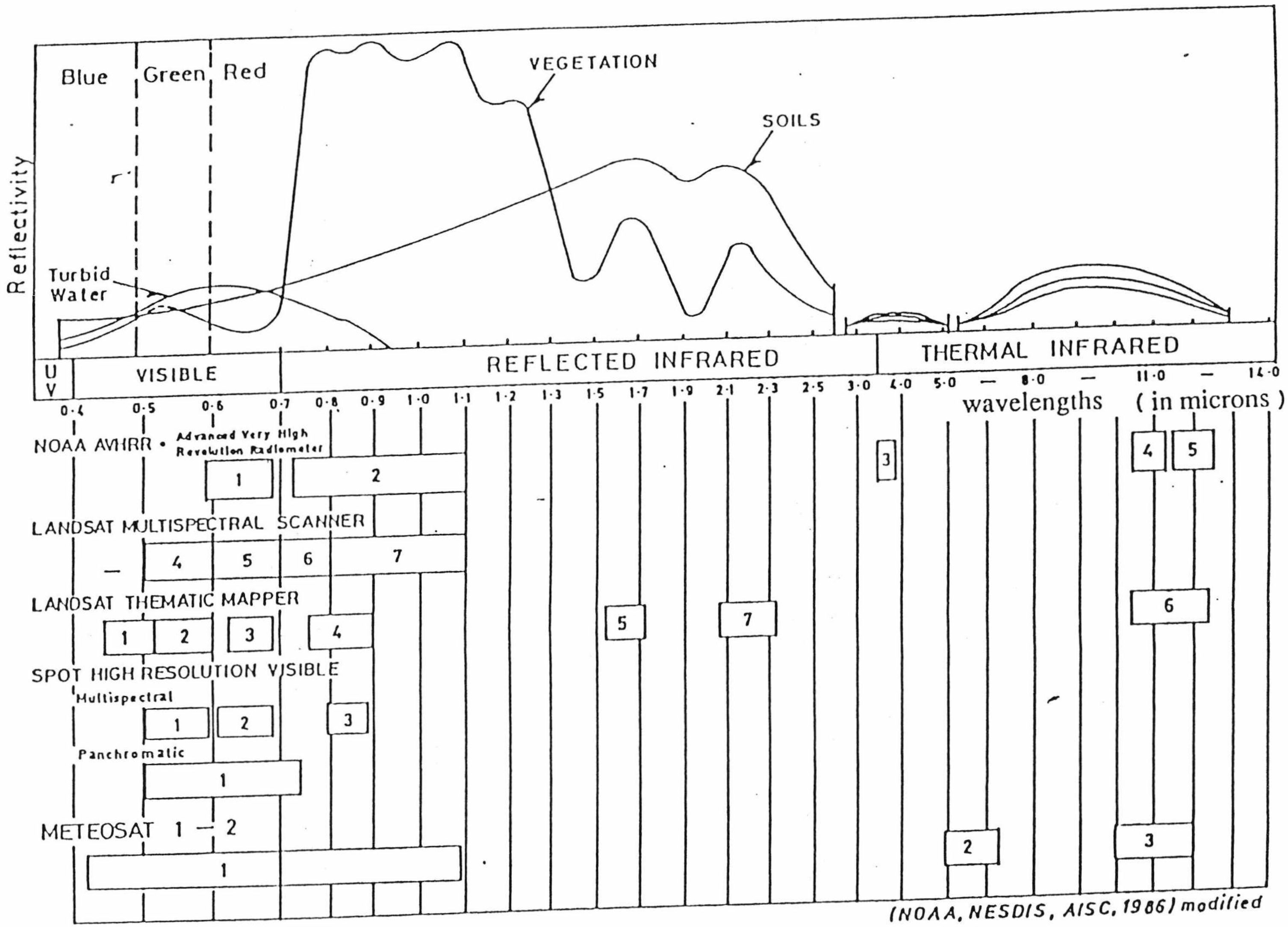


Figure 4.12 : Spectral reflectance curve for vegetation and soils

However, limitation of resolution compared to aircraft imagery, which sometimes can be up to $3m$, is a great setback in the utilization of satellite imagery. There is an effect of occasional cloud cover which becomes severe during the wet season. Also, to acquire satellite images, there must be some foreign currency involvement which is strictly used for most priority activities in many developing countries.

4.5.1 Rainfed agriculture and remote sensing

From a look at satellite imagery, the small fields appear as a scatter of lighter patches across the landscape. The basic element satellite data earth's surface is classified for each pixel, an acronym for picture element, as explained in chapter three section 3.3.1, is the unit for which MSS and TM information is recorded and is about 0.45 and 0.09 hectares in area respectively. The corresponding SPOT unit of measurement is 0.04 for colour composite and 0.01 for panchromatic. This makes it highly probable that the content of a pixel will not be homogeneous; it means that two or more land covers characterize the ground resolution. Therefore even with such high resolution, extraction of detailed land information, that is, identification of a field in traditional agriculture becomes tedious.

A field is normally divided into plots; the number of plots in the fields depends on the farmer's planting plans. A plot can either contain one kind of crop or mixed crops. Primarily, the plot sizes are often small, and exceptionally greater than the pixel size. The average land use by economic unit of agricultural production (holding) is 1.1 hectares and 43% of these holdings are less than 0.5 hectares (Takwimu 1989).

Trees are found in nearly every field, mango trees, baobab *etc.* which render identification of cultivation land difficult. The practice of mixed cropping (maize and beans, maize and others, *etc.*) makes identification rather complicated by spectral reflectance parameters.

The percentage of soil covered by the vegetation is always low and even lower than the natural vegetation, except in some localized favoured areas, or those around inhabited areas which are intensely manured. In that case, a dense vegetation can grow, not only mixed cropping but also weeds. Thus in the common situation, the nature and the hydric state of the soil will have prominent influence on the spectral behaviour of the cultivated space.

The within field growing pattern is often irregular, and crops grow well only where there is good soil. Elsewhere, the terrain is washed out eventually together with the seeds. This is likely in those fields located on steep slopes, where crops grow on the talus, in small patches of soil between blocks. Such situation and location give rise to many specific problems in remote sensing. Again because of the varying soil coverage it is expected to get a strongly mixed spectral signature (vegetation + soil + rocks). Even the slopes influence dramatically the brightness of the pixel where fields are not grouped, but tend to spread on the slopes. These fields turn back into fallow after some number of years.

It is in these intricate situations that agricultural statistics seems to be hardly described only by remote sensing. In conjuction with other approaches, however, the right use of remote sensing information can contribute to a better understanding of the land facets and a more successful implementation of integrated projects.

4.6 Seasonal changes and image acquisition

Remote sensing data from different times of the year reveals seasonal changes and from different years, records long term changes. Seasonal changes can lead to dramatic variability of tones on images. Under such circumstances, the variability is caused by the following reasons.

1. Differences in vegetation growth which are more difficult to distinguish in the uniform rank growth of wet and early dry season.
2. Deciduous trees losing their leaves.

Both aspects have far reaching influence on the interpretability of an image. During the wet season and shortly thereafter, all woodland vegetation has a dark tone, whereas grassland has a noticeably lighter tone. As the dry season progresses, some trees lose their leaves, the grass becomes drier and more of it is burnt. Woodland reflectances vary and it becomes more difficult to distinguish grassland from woodland. Linear features such as roads, railways and even strand lines can be seen more clearly on early dry season imagery. Similarly, relief contrast is usually stronger early in the dry season because of the atmospheric haze which obscures late dry season imagery. Later in the dry season, grassland becomes distinguishable because it is burnt more often and burning is more visible. Thus wooded hills surrounding a grassland plain are easily discriminated. Nevertheless burning late in the dry season is a less reliable indicator of grassland than the tonal differences at the beginning of the dry season because burning can be confused with water if swamp vegetation is burnt.

Thus three main factors govern changes in seasonal reflectance as revealed by colour composites;

(i) vegetation growth

(ii) burning

(iii) atmospheric haze.

Burning and haze increase as the dry season progresses, while vegetation growth decreases. The purpose of the image and location of the study area may suggest the time of imaging. However there are other factors influencing the ecology of different locations that if not reckoned with could be misleading.

It is generally recommended to use early to mid-dry season's imaging where images seem to provide the most information when vegetation density contrasts are revealed and there is not yet too much haze obscurity. In the comparison of Multispectral scanner images between the two seasons King (1981) recommended mid-dry seasons. It was his experience that colour at that time reveals the greatest vegetation density contrasts, which are often associated with relief.

In a case where interest is in finding the area of a particular crop or crops combination, it thus becomes necessary to integrate this technique with the ground work. That is, ground survey data can be used with the total area to estimate the cultivated area of the crop of interest with greater accuracy.

Information on a small scale of say y, the character of our interest, that is the area of a particular crop on a household, and say x, an auxiliary character, that is the total area of the holding related to y, can be collected in a usual ground survey method. This information is then to be integrated with X, the total cultivable area in the village obtained from remote sensing, to estimate Y, the total area of the crop in the village by making use of the ratio-type estimators. Although ratio-type estimators are known to be biased they are simple and easy to use. In addition there has been a great deal of effort to modify and make them unbiased or with relatively low and insignificant bias. In the next chapter we will review some of these ratio-type estimators and the effect of their efficiencies and bias in an effort to achieve a better estimator for incorporating remote sensing data with the ground survey data.

Chapter five

Ratio-type estimators

5.1 Introduction

The use of auxiliary information in the finite population total or mean is a common occurrence in practice. It is widely used for example at the stage of designing for the purpose of stratification, or to choose another sampling design such that the probability of including a unit in the sample is proportional to the size of the auxiliary variable. Auxiliary information is frequently employed at the stage of estimation in the form of ratio, regression, product and difference estimators, because of their simplicity and efficiency. Such estimators take advantage of the correlation ρ_{yx} between the characteristic y and the auxiliary variable x. These estimators, under certain conditions, give more reliable estimates of the population value under study than those based on simple averages (Sukhatme and Sukhatme 1974). These estimators belong to a general expression $\bar{y} + h(\bar{X} - \bar{x})$, where $\bar{y}, \bar{x}$ are the arithmetic means of the sample of n pairs y_i, x_i drawn from a population of N pairs respectively, and $\bar{X}$ is a population mean, supposedly known for the variable. The random variable h is a function of sample pairs (x_i, y_i) which converges to some finite quantity, say H, which takes the value

(i) $r = \dfrac{\bar{y}}{\bar{x}}$, as a simple ratio estimate such that $\bar{y}_r = r\bar{X}$ is a customary ratio estimator.

(ii) b_{yx}, a regression coefficient such that $y_{lr} = \bar{y} + b_{yx}(\bar{X} - \bar{x})$ is a regression estimator.

(iii) $p_r = \bar{y}\,\bar{x}$, a product form such that $\bar{y}_p = \dfrac{p_r}{\bar{X}}$ is a product estimator, and

(iv) d, a constant, then $\bar{y}_d = \bar{y} + d(\bar{X} - \bar{x})$ is a difference estimator.

5.1.1 Simple ratio estimator

Ratio estimators are among the most commonly used. For example, to estimate population total, Y, or population mean, $\bar{Y}$, it is customary to take a simple random sample of size n and estimate $\hat{Y} = rX$ and $\hat{\bar{Y}} = r\bar{X}$ respectively. A well known defect of these estimators is the fact that they are usually biased. However by taking advantage of the correlation between y and x, the ratio estimator comparatively provides a more reliable estimate of the population value than that based on simple arithmetic mean.

5.1.1.1 Bias of ratio estimate

Since both $\bar{y}$ and $\bar{x}$ are unbiased estimates of $\bar{Y}$ and $\bar{X}$ respectively, then

$$R = \frac{Y}{X} = \frac{E(\bar{y})}{E(\bar{x})} = \frac{E(r\bar{x})}{E(\bar{x})}.$$

Bias of r is expressed as

$$\begin{aligned} Bias\,(r) &= E(r) - R = E(r) - \frac{E(r\bar{x})}{E(\bar{x})} \\ &= \frac{E(r)E(\bar{x}) - E(r\bar{x})}{E(\bar{x})} \\ &= \frac{-cov(r,\bar{x})}{\bar{X}}. \end{aligned}$$

Because $cov(r,\bar{x}) = \rho_{r\bar{x}}\,\sigma_r\,\sigma_{\bar{x}}$, an upper bound to the bias of the ratio estimate r can therefore be obtained as

$$\left|Bias(r)\right| \le \frac{\sigma_r\,\sigma_{\bar{x}}}{\bar{X}} = \sigma_r \sqrt{\frac{N-n}{N}\,n}\; C_x \tag{5.1}$$

where $C_x = \dfrac{S_x}{\bar{X}}$ is the coefficient of variation of x.

In (5.1) if n is sufficiently large, $Bias(r)$ is negligible relative to the standard deviation. In section 5.3 various approximations to the bias in ratio-type estimators will be obtained.

5.1.1.2 Approximate variance of ratio estimator

The usual approximation for the variance of r in a simple random sampling without replacement for relatively large n is taken as

$$v(r) = E[r - E(r)]^2 \approx E(r - R)^2$$

$$= E\left[\frac{\bar{y} - R\,\bar{x}}{\bar{x}}\right]^2 \approx \frac{E(\bar{y} - R\,\bar{x})^2}{\bar{X}^2}.$$

Considering the variable $u_i = y_i - R\,x_i$, the following can thus be obtained.

$\bar{u} = \bar{y} - R\,\bar{x}$ and $\bar{U} = \bar{Y} - R\bar{X} = 0$

which makes

$$v(r) = \frac{1}{\bar{X}^2} E(\bar{u} - \bar{U})^2$$

$$= \frac{1-f}{n\bar{X}^2} \sum_{i=1}^{N} \frac{(u_i - \bar{U})^2}{(N-1)}$$

$$= \frac{1-f}{n\bar{X}^2} \sum_{i=1}^{N} \frac{(y_i - R\,x_i)^2}{N-1}.$$

This result as shown in Cochran (1977, equation 6.4) leads to equation (5.2) below, also shown by Cochran in equation 6.12

$$v(r) = \frac{1-f}{n\bar{X}^2}(s_y^2 + r^2 s_x^2 - 2\,r\,s_{yx}) \tag{5.2}$$

which is the approximate variance of the ratio estimator.

s_{yx} is the sample covariance between y and x which is given as

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}. \tag{5.3}$$

Des Raj (1964) proved that $v(r)$ understates the true variance $V(r)$. He showed that the understatement as a proportion of $v(r)$ exceeds three times the relative variance of $\bar{x}$ if the distribution of $\bar{x}$ is symmetrical about the mean and would be higher in case where $\bar{x}$ is negatively skewed. Consequently, for most practical purposes, approximate variance results are equally valid for comparison of precision.

Since $\bar{y}_r = r\bar{X}$, then (5.2) can be written as

$$v(\bar{y}_r) = \frac{1-f}{n}(s_y^2 + r^2 s_x^2 - 2r\, s_{yx}).$$

The relative efficiency of $\bar{y}_r$ compared to $\bar{y}$ is given as

$$\textit{Efficiency} = \frac{s_y^2}{s_y^2 + r^2 s_x^2 - 2r\, s_{yx}}$$

$$= \frac{1}{1 + \left[\dfrac{C_x^2}{C_y^2}\right] - 2\rho_{yx}\left[\dfrac{C_x}{C_y}\right]}.$$

It follows that in relatively large samples the ratio estimate will be more efficient than the corresponding sample estimate based on the simple arithmetic mean if the denominator is less than 1, that is, if $\left[\dfrac{C_x^2}{C_y^2}\right] < 2\rho_{yx}\left[\dfrac{C_x}{C_y}\right]$ or $\rho_{yx}\dfrac{C_y}{C_x} > \dfrac{1}{2}$.

5.2 Some ratio-type estimators

It has been observed in section 5.1.1.1 that often in practical surveys, the bias is a small fraction of the standard deviation of the estimate and can be neglected. There is however an important class of survey designs in which this bias may become considerable. With small samples drawn from each of a large number, L, of strata, then if the bias in each stratum has the same sign, which according to Goodman and Hartley (1958) does often happen, the bias in the estimate of the population total will be approximately L times that of an individual stratum, while the standard deviation only multiplies by $\sqrt{L}$ (Cochran 1977). Therefore mean square error will be of the order of magnitude L^2. Had it been that the estimate in each stratum total was unbiased, the order of magnitude would have been L. It is in these situations that the use of an unbiased ratio estimator is of great advantage. Hence there has been considerable interest in developing estimates of the ratio-type estimators that are unbiased or subject to smaller bias than the customary ratio estimator.

Quenouille (1956) managed to reduce bias from the terms of order n^{-1} to order n^{-2}, that is, from $O(n^{-1})$ to $O(n^{-2})$ and introduced a class of estimators of r of the kind $t_Q = 2r - ½(r_1 + r_2)$ by splitting the sample at random into two groups each of size $n/2$ when n is even where $r_j = \dfrac{\bar{y}_j}{\bar{x}_j}$, $(j = 1,2)$.

r is a simple ratio estimate introduced in section 5.1(i). $\bar{y}_j$, $\bar{x}_j$ are means of y, x respectively obtained from the j^{th} half-sample.

It might be thought that this reduction in bias could possibly be achieved at the expense of a corresponding increase of variance. But Quenouille managed to show that any such increase in variance is of small order in n compared with the variance itself. The argument has been followed by Durbin (1959) who demonstrated that for the class of estimators, Quenouille's device actually reduces the variance, and proved that $v(t_Q)$ is smaller than $v(r)$ in spite of the fact that for sufficiently large n, t_Q has a smaller bias than r.

Hartley and Ross (1954) proposed an unbiased ratio-type estimate

$$\bar{y}_{HR} = \bar{r}\bar{X} + \frac{n(N-1)}{N(n-1)}(\bar{y} - \bar{r}\bar{x}) \tag{5.4}$$

where $\bar{r} = \dfrac{1}{n}\sum\limits_{i=1}^{n} r_i = \dfrac{1}{n}\sum\limits_{i=1}^{n}\dfrac{y_i}{x_i}$.

Using (5.4) above Mickey (1959) derived an estimate

$$\bar{y}_M = r'_- + \frac{n(N-n+1)}{N\bar{X}}(\bar{y} - r'_-\bar{x})$$

where r'_- denotes the mean of r'_j obtained by omitting the j^{th} pair, in turn, from the sample, so that $r'_j = \dfrac{\sum y}{\sum x}$ over the remaining $(n-1)$ members.

Lahiri (1951) showed that the customary ratio estimate r is unbiased if the sample is drawn with probability proportional to $\sum x_i$. Along the same line Midzuno (1951) suggested that the simplest method was to draw the first member of the sample with probability proportional to x_i. The remaining $(n - 1)$ members of the sample are drawn with equal probability. From then, several different ratio-type estimators were developed. The study will review some, which are approximately unbiased yet computationally comparable and which will be more efficient than the customary ratio estimator.

In a comparative study of ratio estimators, Tin (1965), on the assumption of simple random sampling, compared several estimators regarding their properties on bias, efficiency and approach to normality. He considered the following.

Simple ratio estimate $r = \dfrac{\bar{y}}{\bar{x}}$

Quenouille's ratio estimator $t_Q = 2r - ½(r_1 + r_2)$

Beale's ratio estimator $t_B = \dfrac{r(1 + \eta C_{xy})}{(1 + \eta C_x^2)}$

and

Modified ratio estimator $t_{MOD} = r\left\{1 + \eta(C_{xy} - C_x^2)\right\}$

where $C_{xy} = \dfrac{s_{xy}}{\bar{x}\,\bar{y}}$ and s_{xy} the sample covariance as given in equation (5.3) in section 5.1.1.2. $\bar{y}$, $\bar{x}$, r_1, r_2 and C_x are as previously defined, and $\eta = (\dfrac{1}{n} - \dfrac{1}{N})$. He concluded that the three ratio estimators t_Q, t_B, and t_{MOD} were attractive from theoretical as well as computational points of view. However when variances of these estimators in finite populations were compared, t_Q appeared to be more efficient than r while t_B and t_{MOD} were equally more efficient than t_Q to $O(n^{-2})$. In the case of x and y having a bivariate normal distribution, t_{MOD} was the most efficient estimator among the four with t_B next, followed by t_Q to $O(n^{-2})$. At first sight, the modified estimator t_{MOD}

appeared to have been derived from t_B when terms of order n^{-2} were neglected, but Tin (1965) argued that it was actually obtained by a direct effort to reduce the bias in r.

Srivastava (1967) was the first to consider a generalized ratio and product estimator for estimating the population mean $\bar{Y}$, in a form of class estimators

$$\bar{y}_{\alpha} = \bar{y}\left\{\frac{\bar{x}}{\bar{X}}\right\}^{\alpha}$$

where the optimum value of α is $-\rho_{yx}\dfrac{C_y}{C_x} = -Q$, say. Among the estimators incorporated in this class are the following as listed by Srivastava, Jhajj and Sharma (1986)

$$\bar{y}_C = (1-W)\bar{y} + W\bar{y}\left[\frac{\bar{X}}{\bar{x}}\right]$$

$$\bar{y}_V = (1-a)\bar{y} + a\bar{y}\left[\frac{\bar{x}}{\bar{X}}\right]$$

and

$$\bar{y}_W = \bar{y}\frac{\bar{X}}{A\bar{x} + (1-A)\bar{X}}.$$

The values of the constants which minimize Mean Square Error (MSE) up to term of order n^{-1} of the respective estimators were $W = Q$, $A = Q$ and $a = -Q$. Srivastava (1980) defined a large class of estimators of $\bar{Y}$ of which $\bar{y}_C$, $\bar{y}_V$ and $\bar{y}_W$ are members. The class for the case of a single auxiliary variable was $\bar{y}_h = \bar{y}\,h(\frac{\bar{x}}{\bar{X}})$ where h(.) is a parametric function such that h(1) = 1 and satisfies certain conditions, like h(.) should be continous and bounded in a region. He obtained the optimum values of the parameters in h(.) which minimize MSE of the estimator up to terms of $O(n^{-1})$. The optimum values of the parameters so obtained were dependent upon Q only.

All such estimators thus involve Q which is a function of population parameters.

When the value of Q is not known, it is required to be estimated from the given sample. Reddy (1978) has shown that the value of Q is fairly stable in repeated surveys. He established that the value of Q is more stable than other population parameters such as linear regression coefficients (say B family). Srivastava, Jhajj and Sharma (1986) compared three estimators of the Q kind and two estimators of B in five different populations and found that the MSEs of both Q and B estimators were minimum when estimated without the knowledge of population parameters.

Chakrabarty (1979) proposed two ratio-type estimators

$$\bar{y}_{C1} = (1-W)\bar{y} + W\bar{y}_r$$

and

$$\bar{y}_{C2} = (1-W)\bar{y} + W\, t_Q \bar{X} \qquad \text{for } all\ \ W \geq 0.$$

The author compared the two estimators with $\bar{y}_r$ and Srivastava's (1967) $\bar{y}_S = \bar{y}\left\{\dfrac{\bar{X}}{\bar{x}}\right\}^W$.

He concluded that the three ratio-type estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_S$ were preferable to both $\bar{y}$ and $\bar{y}_r$, and that their efficiencies were the same in large samples and were practically of the same order in small samples. He noted that computationally $\bar{y}_{C1}$ was simplest and that the bias of $\bar{y}_{C2}$ was least.

In a similar approach Sisodia and Dwivedi (1981) proposed the ratio-type estimator

$$\bar{y}_{SD} = (1-b)\bar{y} + b\,\bar{y}\left[\frac{\bar{X}}{\bar{x}}\right]^{\alpha}. \tag{5.5}$$

In (5.5), if $\alpha = 1$ then $\bar{y}_{SD}$ reduces to $\bar{y}_{C1}$, if $b = 1$ then $\bar{y}_{SD}$ reduces to $\bar{y}_S$ and if $b = 1$ and $\alpha = 1$ together, then $\bar{y}_{SD}$ reduces to the customary ratio estimator $\bar{y}_r$.

In the estimation of the population statistics, for example mean $\bar{Y}$, the study will also consider an unbiased ratio-estimator or one with a bias of $O(n^{-2})$ with as low a variance as possible and compare with Chakrabarty's two estimators which proved to be efficient, convenient to compute, and almost unbiased.

It has been observed that the estimator t_Q featured relatively well in Tin's (1965) comparison and Durbin (1959) showed that it is more efficient than r for x under the normal and gamma distributions. Also Chakrabarty (1979) used this estimator and achieved least bias; without loss of generality this study will confine to this estimator. A comparison of the estimator t_Q, in the form of Chakrabarty's $\bar{y}_{C2}$, with t_{MOD}, which is more convenient computationally than t_B, will be made, though both are identical and equally efficient.

Therefore t_{MOD}, in Srivastava's class of estimators, is proposed in this study as

$$\bar{y}_{MOD} = (1-W)\bar{y} + W\, t_{MOD}\bar{X}.$$

Along with the two estimators, that is $\bar{y}_{C1}$ and $\bar{y}_{C2}$ the study will also compare $\bar{y}$ and $\bar{y}_r$, the sample mean and customary ratio estimators respectively.

5.3 Asymptotic bias

It is obvious that $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ are consistent, but in general biased like the ratio estimator $\bar{y}_r$. The sampled values are assumed to be positive in order to determine their biases and that the sample size n sufficiently large so that

$$|\delta_{\bar{x}}| = \left| \frac{\bar{x} - \bar{X}}{\bar{X}} \right| < 1 \tag{5.6}$$

which is not unreasonable in an actual sample survey where x's are usually positive or can be adjusted to become positive. However the assumption is valid for all values of x (David and Sukhatme 1974). Also in order to facilitate the asymptotic expansions the following result is needed (see Kendall and Stuart 1963, Tin 1965 and Wu 1982).

$$E(\bar{x} - \bar{X})^r (\bar{y} - \bar{Y})^s = \begin{cases} O(n^{-\frac{1}{2}(r+s)}) & r+s \quad even \\ O(n^{-\frac{1}{2}(r+s+1)}) & r+s \quad odd. \end{cases}$$

Since all terms of order n^{-2} in the calculations will be ignored, then all expectations with combinations of $(r+s) \geq 4$ when $(r+s)$ is even or $(r+s+1) \geq 4$ when $(r+s)$ is odd will be included in the term of $O(n^{-2})$.

Using (5.6) it can be shown that

$$r = \frac{\bar{y}}{\bar{x}} = \frac{\bar{y} - \bar{Y} + \bar{Y}}{\bar{x} - \bar{X} + \bar{X}} = \frac{\bar{Y}}{\bar{X}} \left\{ \frac{1 + \dfrac{\bar{y} - \bar{Y}}{\bar{Y}}}{1 + \dfrac{\bar{x} - \bar{X}}{\bar{X}}} \right\}$$

$$= R\,(\frac{1 + \delta_{\bar{y}}}{1 + \delta_{\bar{x}}}) = R\,(1 + \delta_{\bar{y}})\,(1 + \delta_{\bar{x}})^{-1}$$

where $R = \dfrac{\bar{Y}}{\bar{X}}$, the population ratio. Expanding the above by Taylor's series,

$$r = R\{1 + \delta_{\bar{y}} - \delta_{\bar{x}} + \delta_{\bar{x}}{}^2 - \delta_{\bar{x}}\,\delta_{\bar{y}} + \cdots\}. \tag{5.7}$$

Taking expectations on both sides gives

$$E(r) = R\left\{1 + E\,\delta_{\bar{x}}{}^2 - E\,\delta_{\bar{x}}\,\delta_{\bar{y}} + O(n^{-2})\right\}$$

$$= R\left\{1 + \frac{S_x{}^2}{n\bar{X}^2} - \frac{S_{xy}}{n\bar{X}\,\bar{Y}} + O(n^{-2})\right\}$$

$$= R\left\{1 + \frac{C_x^2}{n} - \frac{C_{xy}}{n} + O(n^{-2})\right\}$$

$$= R + \frac{R}{n}(C_x^2 - C_{xy}) + O(n^{-2}).$$

The bias of r is found as

$$\begin{aligned} Bias(r) &= E(r) - R \\ &= \frac{R}{n}(C_x^2 - C_{xy}) + O(n^{-2}). \end{aligned}$$

Therefore

$$\begin{aligned} Bias(\bar{y}_r) &= \bar{X}\,Bias(r) \\ &= \frac{R\,\bar{X}}{n}(C_x^2 - C_{xy}) + O(n^{-2}) \end{aligned}$$

$$= \frac{\bar{Y}}{n}(C_x^2 - C_{xy}) + O(n^{-2}).$$

In t_Q, since r_1 and r_2 are independent

$$E(t_Q) = 2E(r) - ½E(r_1) - ½E(r_2)$$
$$= 2\left\{R + \frac{R}{n}(C_x^2 - C_{xy})\right\} - ½\left\{R + \frac{R}{n/2}(C_x^2 - C_{xy})\right\}$$
$$- ½\left\{R + \frac{R}{n/2}(C_x^2 - C_{xy})\right\} + O(n^{-2})$$
$$= 2R + \frac{2R}{n}(C_x^2 - C_{xy}) - \left\{R + \frac{2R}{n}(C_x^2 - C_{xy})\right\} + O(n^{-2})$$
$$= R + O(n^{-2})$$

and

$$Bias(t_Q) = E(t_Q) - R$$
$$= 0 + O(n^{-2}).$$

Consequently the biases of $\bar{y}_{C1}$ and $\bar{y}_{C2}$ are respectively,

$$Bias(\bar{y}_{C1}) = W\,Bias(\bar{y}_r)$$
$$= \frac{W\bar{Y}}{n}(C_x^2 - C_{xy}) + O(n^{-2})$$

and

$$Bias(\bar{y}_{C2}) = W\bar{X}\,Bias(t_Q)$$
$$= 0 + O(n^{-2}).$$

Likewise

$$t_{MOD} = r\left\{1 + \eta(C_{xy} - C_x^2)\right\}$$
$$= r\left\{1 + \eta C_{xy} - \eta C_x^2\right\}$$
$$= r\left\{1 + \eta\frac{S_{xy}}{\bar{X}\bar{Y}} - \eta\frac{S_x^2}{\bar{X}^2}\right\}.$$

Substituting r from equation (5.7) the above can be re-written as

$$t_{MOD} = R\left\{1 + \delta_{\bar{y}} - \delta_{\bar{x}} + \delta_{\bar{x}}^{2} - \delta_{\bar{x}}\,\delta_{\bar{y}} + \cdots\right\}\left\{1 + \eta\frac{S_{xy}}{\bar{X}\,\bar{Y}} - \eta\frac{S_x^2}{\bar{X}^2}\right\}$$

$$= R\left\{1 + \delta_{\bar{y}} - \delta_{\bar{x}} + \delta_{\bar{x}}^2 - \delta_{\bar{x}}\,\delta_{\bar{y}} + \eta\frac{S_{xy}}{\bar{X}\,\bar{Y}} - \eta\frac{S_x^2}{\bar{X}^2} + \cdots\right\}.$$

Taking expectations on both sides gives us

$$E(t_{MOD}) = R\left\{1 + \frac{S_x^2}{n\bar{X}^2} - \frac{S_{xy}}{n\bar{X}\,\bar{Y}} + \frac{\eta S_{xy}}{\bar{X}\,\bar{Y}} - \frac{\eta S_x^2}{\bar{X}^2} + O(n^{-2})\right\}$$

$$= R\left\{1 + \frac{C_x^2}{n} - \frac{C_{xy}}{n} + \frac{C_{xy}}{n} - \frac{C_{xy}}{N} - \frac{C_x^2}{n} + \frac{C_x^2}{N} + O(n^{-2})\right\}$$

$$= R + \frac{R}{N}(C_x^2 - C_{xy}) + O(n^{-2}).$$

(The above derivation is given in detail in Appendix V).

Therefore the bias of t_{MOD} is expressed as

$$Bias(t_{MOD}) = \frac{1}{N}(C_x^2 - C_{xy}) + O(n^{-2})$$

and that of $\bar{y}_{MOD}$ as

$$Bias(\bar{y}_{MOD}) = \frac{W\bar{X}R}{N}(C_x^2 - C_{xy}) + O(n^{-2})$$

$$= \frac{W\bar{Y}}{N}(C_x^2 - C_{xy}) + O(n^{-2})$$

$$= \frac{W\bar{Y}}{n}\frac{n}{N}(C_x^2 - C_{xy}) + O(n^{-2}),$$

that is

$$Bias(\bar{y}_{MOD}) = \frac{n}{N}\,Bias(\bar{y}_{C1}).$$

The asymptoptic bias of $\bar{y}_{C2}$ is zero in the order of n^{-2} and hence much smaller than that of $\bar{y}_r$, $\bar{y}_{C1}$ and $\bar{y}_{MOD}$ in the same order. However the bias of $\bar{y}_{MOD}$ is smaller than that of $\bar{y}_{C1}$ and $\bar{y}_r$ and it can be approximately equal to that of $\bar{y}_{C2}$ when $N \gg n$. Also when $(C_x^2 - C_{xy}) = 0$, the regression of y on x passes through the origin and all the estimators become unbiased to the term of $O(n^{-2})$.

Again, when substituting the formula for exact bias of $\bar{y}_r$ from Hartley and Ross (1954), the exact bias of $\bar{y}_{C1}$ becomes as follows

$$Bias(\bar{y}_{C1}) = -W\,Cov(r,\bar{x})$$

where $Cov(r,\bar{x})$ as expressed in section 5.1.1.1 is the covariance of r and $\bar{x}$.

Since $Cov(r,\bar{x}) = \rho_{r\bar{x}}\,\sigma_r\,\sigma_{\bar{x}}$ and $\sigma_{\bar{y}_r} = \bar{X}\,\sigma_r$ then

$$\frac{|Bias(\bar{y}_{C1})|}{\sigma_{\bar{y}r}} \le \frac{W\,C_x}{\sqrt{n}} \tag{5.8}$$

giving the upper bound to the ratio of the bias to the standard error of $\bar{y}_r$. Thus if $\frac{W\,C_x}{\sqrt{n}} \le 0.1$, the bias of $\bar{y}_{C1}$ is negligible in relation to the standard error of $\bar{y}_r$. This also applies to the bias of $\bar{y}_{MOD}$ which is a fraction $\frac{n}{N}$ that of $\bar{y}_{C1}$.

5.4 Asymptotic variance

In defining the variance of estimators $\bar{y}_{C1}$ and $\bar{y}_{C2}$, Chakrabarty (1979) considered up to terms of $O(n^{-2})$ only. He therefore showed that the variances are identical and, omitting the finite correction factor, are given by

$$V(\bar{y}_{C1}) = V(\bar{y}_{C2}) = \frac{S_y^2}{n}\{1 + W\,K(W\,K - 2\rho_{yx})\} \tag{5.9}$$

where $K = \frac{C_x}{C_y}$ and that the value of W which minimizes this variance is

$$W_{opt} = \frac{\rho_{yx}}{K}.$$

The minimum variance was given as

$$V_{\min} = \frac{S_y^2}{n}(1 - \rho_{yx}^2) \tag{5.10}$$

which is equal to the variance of the linear regression estimator up to terms of $O(n^{-2})$.

The variance of $\bar{y}_{MOD}$ as derived in Appendix VI using a similar method suggested by Chakrabarty (1979) is

$$V(\bar{y}_{MOD}) = \frac{S_y^2}{n}\{1 + WK\{WK - 2\rho_{yx} + \frac{2n}{N}(K - \rho_{yx})\}\}$$

which is slightly greater than that of $\bar{y}_{C1}$ and $\bar{y}_{C2}$. However for large n it has been seen that $\rho_{yx}\frac{C_y}{C_x} > \frac{1}{2}$ which implies that $K < 2\rho_{yx}$ and hence the factor $\frac{2n}{N}(K - \rho_{yx})$ is always less than $\frac{2n}{N}\rho_{yx}$ and hence relatively very small. As for small n and $N \gg n$ again the factor disappears. Hence, considering the factor negligible, then the variance of $\bar{y}_{MOD}$ is equivalent to those of $\bar{y}_{C1}$ and $\bar{y}_{C2}$.

Substituting $W = 1$ in (5.9) the variance of $\bar{y}_r$ is expressed as

$$V(\bar{y}_r) = \frac{S_y^2}{n}\{1 + K(K - 2\rho_{yx})\}.$$

Thus the asymptotic efficiencies of $\bar{y}_{C1}$ ($\bar{y}_{C2}$ and $\bar{y}_{MOD}$) over $\bar{y}$ and $\bar{y}_r$ are respectively given by

$$E_1 = \frac{V(\bar{y})}{V(\bar{y}_{C1})} = \frac{1}{\left\{1 + WK(WK - 2\rho_{yx})\right\}}$$

and

$$E_2 = \frac{V(\bar{y}_r)}{V(\bar{y}_{C1})} = \frac{\left\{1 + K(K - 2\rho_{yx})\right\}}{\left\{1 + WK(WK - 2\rho_{yx})\right\}}.$$

The following is achieved as a result of the above

$$E_1 \geq 1 \text{ if } W \leq 2\frac{\rho_{yx}}{K} \quad \text{and}$$

$$E_2 \geq 1 \text{ if } \frac{(2\rho_{yx} - K)}{K} \leq W \leq 1$$

making estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ better than $\bar{y}$ and $\bar{y}_r$.

The estimators $\overline{y}_{C1}$, $\overline{y}_{C2}$ and $\overline{y}_{MOD}$ are better than $\overline{y}$ and $\overline{y}_r$ for a wide range of W values. The efficiencies of E_1 and E_2 of the estimators $\overline{y}_{C1}$, $\overline{y}_{C2}$ and $\overline{y}_{MOD}$ over $\overline{y}$ and $\overline{y}_r$ will depend on ρ, K and W.

Tables 5.1a and 5.1b give the estimators' efficiencies for selected values of ρ, K and W.

Table 5.1a: Efficiencies for selected values of ρ and K and $W = 0.25$

	$K = 0.5$		$K = 1.0$		$K = 1.5$		$K = 2.0$	
ρ	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2
0.1	101	116	99	178	94	277	89	400
0.2	104	109	104	166	101	268	95	400
0.3	106	101	110	153	109	257	105	400
0.4	109	93	116	139	119	244	118	400
0.5	112	84	123	123	131	229	133	400
0.6	116	75	131	105	145	210	154	400
0.7	119	65	140	84	162	187	182	400
0.8	123	55	150	63	185	157	222	400
0.9	126	44	163	33	215	118	285	400

Table 5.1b: Efficiencies for selected values of ρ and K and $W = 0.50$

	$K = 0.5$		$K = 1.0$		$K = 1.5$		$K = 2.0$	
ρ	E_1	E_2	E_1	E_2	E_1	E_2	E_1	E_2
0.1	99	114	87	157	71	209	56	256
0.2	104	109	95	152	79	210	62	262
0.3	110	104	105	147	90	211	71	271
0.4	116	99	117	141	104	213	83	283
0.5	123	92	133	133	123	215	100	300
0.6	131	85	153	123	151	219	125	325
0.7	140	77	182	109	195	224	167	367
0.8	150	68	222	89	276	234	250	450
0.9	163	57	286	57	471	259	500	700

(In Tables 5.1a-b values for E_1 have been taken from Chakrabarty 1979 and the other values have been computed for this study)

Tables 5.1a and 5.1b suggest the following.

(i) $W = ¼$ appears to be a good overall choice for $\overline{y}_{C1}$, $\overline{y}_{C2}$ and $\overline{y}_{MOD}$ for low correlation $(0.2 < ρ < 0.4)$ and/or $K > 1$.

(ii) $W = ½$ appears to be a good choice for moderate to high correlation ($\rho > 0.4$) and $K > 1$.

(iii) In cases of higher correlation ($\rho > 0.8$) and $K \leq 1$, $\bar{y}_r$ is preferable.

The strength of the correlation between y and x is an important advantage to the customary ratio estimator. The asymptotic variance given in (5.10) of the estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ with optimum value of $W = \rho_{yx}/K$ is equal to the asymptotic variance of the linear regression estimator given in section 5.1 (ii). These estimators are therefore asymptotically no more efficient than $\bar{y}_{lr}$ with constant weights ($W = ¼$ or $½$). However, the estimator $\bar{y}_{lr}$ suffers appreciably in the ratio of bias to standard error when the relationship is not linear, and Cochran (1977) has shown that the bias in $\bar{y}_{lr}$ is of order n^{-1} and hence it is more biased than $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ whose biases are of order n^{-2}. Thus in situations where freedom from bias is important, then $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ may be preferable to $\bar{y}_{lr}$.

5.5 The exact theory

The model used by Durbin (1959) and Rao and Webster (1966) to investigate the bias estimation of ratios is assumed in this study. Also Chakrabarty (1973) used a similar model when investigating the exact efficiency of the ratio estimator $\bar{y}$ and the stability of the variance of $\bar{y}_r$ relative to that of $\bar{y}$. He has shown that for $\rho_{yx} \geq 0.4$ and $K < 2\rho_{yx}$ the ratio estimator is generally more efficient than the unbiased estimator $\bar{y}$ even in small samples, and that the variance estimator of the ratio estimator is generally more stable than the variance estimator of $\bar{y}$. He also used the the same model (Chakrabarty 1979) when comparing the variances of estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_S$ and their exact efficiencies relative to that of $\bar{y}$ and preferred them in situations where the freedom from bias is less desirable. Also as his model III, Tin (1965) compared the estimators r, t_Q, t_B and t_{MOD} and concluded that t_{MOD} is the most efficient with t_B, t_Q next in that order, while Rao and Webster (1966) showed that although the bias of t_{MOD} is slightly smaller than the bias of t_Q (divided into g groups of size p instead of 2 groups of size $n/2$), t_{MOD} has a slightly larger MSE than t_Q for $m > 12$ and/or $\alpha = 0$ (that is when the bias of t_Q and r is zero). But they observed that the difference is so small that it is difficult to choose between t_{MOD} and t_Q on the basis of mean square error only.

The said model is

$$y_i = \alpha + \beta x_i + u_i \quad ; \quad \beta > 0$$

$$E(u_i \mid x_i) = 0, \quad E(u_i, u_j \mid x_i, x_j) = 0$$

$$v(u_i \mid x_i) = n\delta \quad (\delta \text{ is a constant of order } n^{-1})$$

where the variates $\frac{x_i}{n}$ have a gamma distribution with parameter h so that $\bar{x} = \frac{\sum x_i}{n}$ has the gamma distribution with the parameter $m = nh$.

5.5.1 The exact bias

From the stated model, $\bar{y} = \alpha + \beta \bar{x} + \bar{u}$ and $E(\bar{y}) = \alpha + \beta m = \bar{Y}$.

For any sample size n, Chakrabarty (1979) showed that the biases of $\bar{y}_{C1}, \bar{y}_{C2}$ are respectively

$$Bias(\bar{y}_{C1}) = \alpha W/(m-1) \qquad \text{and}$$
$$Bias(\bar{y}_{C2}) = -2W\alpha/\{(m-1)(m-2)\}.$$

The bias of $\bar{y}_r$ was found by substituting $W = 1$ in the equation for bias of $\bar{y}_{C1}$ as

$$Bias(\bar{y}_r) = \alpha/(m-1).$$

In the similar way the bias of estimator $\bar{y}_{MOD}$ has been derived (Appendix VII) as

$$Bias(\bar{y}_{MOD}) = \alpha W/\{(m-1)(m+1)\}.$$

It can be seen that the biases of $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ are of the order n^{-2}, while those of $\bar{y}_r$ and $\bar{y}_{C1}$ are of order n^{-1}, since $m = nh$ in our model. Also the bias of $\bar{y}_{C1}$ is less than the bias of $\bar{y}_r$ if $W < 1$. Further, for the special case of linear regression through the origin, that is when $\alpha = 0$, the estimators $\bar{y}_r$, $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ are unbiased.

5.5.2 The exact variance

Chakrabarty (1979), using a similar method to that of Rao and Webster (1966), obtained the variances of $\bar{y}_{C1}$ and $\bar{y}_{C2}$ to be

$$v(\bar{y}_{C1}) = \frac{W^2 m}{(m-1)^2(m-2)}\alpha^2 + (1-W)^2 m\beta^2$$

$$+ [\frac{W^2 m^2}{(m-1)(m-2)} + \frac{W(1-W)(m+1)}{(m-1)} + (1-W)]\delta$$

$$- \frac{2W(1-W)m}{(m-1)}\alpha\beta$$

and

$$v(\bar{y}_{C2}) = \frac{W^2 m^2(m^2-6m+17)\alpha^2}{(m-1)^2(m-2)^2(m-4)} - \frac{2W(1-W)m(m-3)\alpha\beta}{(m-1)(m-2)}$$

$$+ (1-W)^2 m\beta^2 + [(1-W)^2 + \frac{W^2(m^2-7m+18)m^2}{(m-1)(m-2)^2(m-4)}$$

$$+ \frac{2W(1-W)m(m-3)}{(m-1)(m-2)}]\delta.$$

Likewise, the derivation of the variance of $\bar{y}_{MOD}$ is detailed in Appendix VIII and is given as

$$v(\bar{y}_{MOD}) = (1-W)^2\beta^2 m$$

$$+ \frac{W^2 m^3}{(m-1)(m-2)(m+1)(m+2)(m+3)}\left[\frac{(m^3+4m^2+8m-1)}{(m-1)(m+1)} + \frac{(n+1)}{(n-1)}\right]\alpha^2$$

$$+ \left[(1-W)^2 + \frac{2(1-W)Wm^2}{(m+1)(m-1)} + \frac{W^2 m^3}{(m-1)^2(m+1)(m+2)(m+3)}\left\{(m^2+4m+1) + \frac{(n+1)}{(n-1)}\right\}\right]\delta$$

$$- \frac{2(1-W)W\alpha\beta m^2}{(m-1)(m+1)}.$$

Putting $W = 1$ and $W = 0$ to the values of $V(\bar{y}_{C1})$, the variances of $\bar{y}_r$ and $\bar{y}$ are obtained as

$$v(\bar{y}_r) = \frac{m^2\alpha^2}{(m-1)^2(m-2)} + \frac{m^2\delta}{(m-1)(m-2)}$$

and

$$V(\bar{y}) = \delta + \beta^2 m$$

respectively.

It is noted (Chakrabarty 1979) that in terms of the model

$$\alpha = \bar{Y}[(K - \rho_{yx})/K]$$
$$\beta = \bar{Y}[\rho_{yx}/Km]$$
$$\delta = \bar{Y}^2[(1-\rho_{yx}^2)/K^2m)]$$

and
$$K = C_x/C_y.$$

The exact efficiencies of $\bar{y}_r$, $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ relative to that of $\bar{y}$ are given by

$$E_r' = \frac{v(\bar{y})}{MSE(\bar{y}_r)}$$
$$E_i' = \frac{v(\bar{y})}{MSE(\bar{y}_i)}$$

where

$$i = \begin{cases} 1 \text{ for } C1 \\ 2 \text{ for } C2 \\ 3 \text{ for } MOD \end{cases} \tag{5.11}$$

E_r' and E_i' ($i = 1,2$ and 3) can be expressed as functions of $K = C_x/C_y$, $m = nh$, ρ_{yx} and W. Since it is difficult to investigate analytically these efficiencies from the resulting expression, their values in percentages were therefore evaluated for selected values of ρ_{yx}, K and m. Tables 5.2a-c give such values for $W = ¼$ and Tables 5.2d-f give such values for $W = ½$.

Table 5.2a: Exact efficiencies with $W = 0.25$ for selected values of m and K

		$\rho = 0.2$				$\rho = 0.3$				$\rho = 0.4$			
m	K	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3
8	0.25	68	94	100	101	72	96	101	103	76	97	103	104
	0.50	61	95	100	103	68	98	104	105	77	101	107	108
	1.00	37	93	96	101	43	99	102	107	51	106	109	114
	1.50	21	87	88	96	24	95	93	104	28	104	103	114
16	0.25	85	98	101	102	90	100	103	103	95	101	104	105
	0.50	77	100	103	103	86	103	105	106	96	106	108	109
	1.00	49	99	102	103	56	105	107	109	66	111	114	115
	1.50	29	94	96	99	32	102	104	107	37	112	114	117
20	0.25	89	99	102	102	94	101	103	103	99	102	104	105
	0.50	81	100	103	103	90	103	106	106	100	106	109	109
	1.00	51	100	102	103	59	106	108	109	69	112	114	115
	1.50	30	96	98	100	34	104	106	108	39	114	115	118
32	0.25	94	100	102	102	100	102	103	103	105	103	104	105
	0.50	86	102	103	103	95	104	106	106	107	107	109	109
	1.00	55	102	103	104	63	107	108	109	74	114	115	116
	1.50	33	98	99	100	37	106	107	108	43	116	117	118

Table 5.2b: Exact efficiencies with $W = 0.25$ for selected values of m and K

		$\rho = 0.5$				$\rho = 0.6$				$\rho = 0.7$			
m	K	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3
8	0.25	79	98	104	105	83	100	105	107	86	101	106	108
	0.50	88	105	110	112	101	108	114	115	117	112	117	118
	1.00	62	114	117	121	78	123	126	130	105	133	136	139
	1.50	33	116	114	126	40	131	127	140	50	150	144	158
16	0.25	100	103	105	106	105	104	107	107	111	105	108	109
	0.50	109	109	112	112	126	112	115	115	148	116	118	119
	1.00	80	119	121	123	100	127	130	131	134	137	139	140
	1.50	44	124	126	129	53	139	140	143	67	157	157	161
20	0.25	104	103	106	106	110	105	107	107	117	106	108	109
	0.50	114	109	112	112	131	113	115	115	154	116	118	119
	1.00	83	120	122	123	105	128	130	131	140	138	139	140
	1.50	46	126	127	129	56	140	141	143	70	158	159	161
32	0.25	111	104	106	106	118	106	107	108	125	107	109	109
	0.50	121	111	112	112	140	114	115	115	164	117	119	119
	1.00	89	121	122	123	112	129	130	131	150	139	140	140
	1.50	50	128	129	130	61	142	143	144	76	160	160	162

Table 5.2c: Exact efficiencies with $W = 0.25$ for selected values of m and K

		$\rho = 0.8$				$\rho = 0.9$			
m	K	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3
8	0.25	89	102	107	109	91	104	108	110
	0.50	139	116	121	122	168	120	125	125
	1.00	160	142	148	150	324	160	161	163
	1.50	67	176	166	181	103	211	196	211
16	0.25	117	107	110	110	123	108	111	111
	0.50	178	119	122	122	222	123	126	126
	1.00	203	149	150	151	408	162	163	163
	1.50	90	181	180	183	138	214	210	213
20	0.25	123	108	110	110	131	109	111	112
	0.50	186	120	122	122	234	124	126	126
	1.00	212	149	150	151	425	162	163	163
	1.50	95	182	181	184	146	214	211	214
32	0.25	133	109	110	110	142	110	112	112
	0.50	199	121	122	123	252	125	126	126
	1.00	226	150	151	151	453	163	163	163
	1.50	103	183	183	184	159	214	213	214

Table 5.2d: Exact efficiencies with $W = 0.50$ for selected values of m and K

		$\rho = 0.2$				$\rho = 0.3$				$\rho = 0.4$			
m	K	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3
8	0.25	68	86	91	102	72	89	94	105	76	91	96	107
	0.50	61	85	89	101	68	91	95	107	77	97	102	114
	1.00	37	73	71	89	43	81	80	99	51	92	91	111
	1.50	21	55	50	69	24	63	57	79	28	73	66	91
16	0.25	85	95	100	103	90	98	102	106	95	101	105	109
	0.50	77	95	99	103	86	101	105	109	96	107	112	115
	1.00	49	84	87	93	56	94	97	103	66	105	109	115
	1.50	29	67	69	75	32	76	78	85	37	88	90	99
20	0.25	89	97	101	103	94	100	103	106	99	102	106	109
	0.50	81	97	100	103	90	102	106	109	100	109	113	116
	1.00	51	86	89	93	59	96	99	103	69	108	111	116
	1.50	30	69	71	76	34	79	81	86	39	91	93	100
32	0.25	94	99	102	103	100	102	105	106	105	105	107	109
	0.50	86	99	102	104	95	105	108	109	107	111	114	116
	1.00	55	90	92	94	63	99	102	104	74	112	114	117
	1.50	33	73	75	77	37	83	85	88	43	96	98	102

Table 5.2e: Exact efficiencies with $W = 0.50$ for selected values of m and K

		$\rho = 0.5$				$\rho = 0.6$				$\rho = 0.7$			
m	K	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3
8	0.25	79	94	99	110	83	96	101	112	86	99	103	115
	0.50	88	104	109	121	101	111	117	129	117	120	126	138
	1.00	62	106	105	127	78	125	125	148	105	152	152	177
	1.50	33	87	78	108	40	107	95	133	50	139	122	172
16	0.25	100	103	108	112	105	106	111	114	111	109	114	117
	0.50	109	114	119	123	126	122	127	131	148	131	136	140
	1.00	80	120	124	131	100	140	145	152	134	168	173	180
	1.50	44	105	107	117	53	129	131	143	67	167	168	185
20	0.25	104	105	109	112	110	108	112	115	117	111	115	118
	0.50	114	116	120	123	131	124	128	131	154	133	137	140
	1.00	83	123	127	132	105	143	147	152	140	171	175	181
	1.50	46	108	111	118	56	133	136	145	70	173	175	188
32	0.25	111	108	110	112	118	111	114	115	125	114	117	118
	0.50	121	119	121	123	140	127	129	131	164	136	138	140
	1.00	89	127	129	132	112	147	150	153	150	175	178	181
	1.50	50	114	116	120	61	140	142	148	76	181	183	191

Table 5.2f: Exact efficiencies with $W = 0.50$ for selected values of m and K

		$\rho = 0.8$				$\rho = 0.9$			
m	K	E'_r	E'_1	E'_2	E'_3	E'_r	E'_1	E'_2	E'_3
8	0.25	89	101	104	117	91	103	105	119
	0.50	139	129	135	147	168	140	146	158
	1.00	160	192	193	218	324	260	262	283
	1.50	67	198	167	242	103	340	264	403
16	0.25	117	112	117	120	123	115	120	124
	0.50	178	141	146	150	222	152	158	161
	1.00	203	208	213	221	408	274	279	285
	1.50	90	238	235	261	138	409	391	441
20	0.25	123	114	118	121	131	118	122	124
	0.50	186	143	147	150	234	155	159	162
	1.00	212	211	216	221	425	277	280	285
	1.50	95	245	245	265	146	422	410	447
32	0.25	133	118	120	122	142	121	124	125
	0.50	199	146	149	150	252	158	161	162
	1.00	226	216	218	222	453	280	283	285
	1.50	103	257	258	269	159	441	435	457

(In Tables 5.2a-f values for E'_r, E'_1, and E'_2 have been taken from Chakrabarty 1979 and the other values have been computed for this study)

From Tables 5.2a-c the following is observed.

(i) For low correlation ($\rho_{yx} < 0.4$), $K < 1$ and $m \geq 20$, the estimator $\bar{y}_r$ is less efficient than $\bar{y}$. But for higher correlations $\rho_{yx} > 0.6$ it is more efficient than estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ for $0.5 \leq K \leq 1$, $m \geq 16$.

(ii) The estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ are more efficient than both $\bar{y}$ and $\bar{y}_r$ for the following values of ρ_{yx}, K and m:

(a) $\rho_{yx} = 0.2$, $K \leq 1$, $m \geq 16$

(b) $0.2 < \rho_{yx} \leq 0.5$ for all values of K and m.

From our model, noting that $C_x = h^{-½}$, $C_{\bar{x}} = m^{-½}$ and $n \leq m$ if $h \geq 1$ then it may be concluded that $W = ¼$ appears to be a good choice for the estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ even in small samples. Moreover, the exact efficiencies of these estimators are of the same order as judged by their mean square errors.

From Tables 5.2d-f the following is observed.

(i) The estimators $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ are more efficient than $\bar{y}_{C1}$, $\bar{y}_r$ and $\bar{y}$ for $\rho_{yx} = 0.5$,$K \leq 0.5$,$m \geq 16$.

(ii) The estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ are more efficient than both $\bar{y}$ and $\bar{y}_r$ when:

(a) $\rho_{yx} = 0.4$,$K \leq 1$, $m \geq 16$

(b) $\rho_{yx} \geq 0.5$, $0.25 \leq K \leq 1.5$ and $m \geq 16$.

(iii) The estimator $\bar{y}_r$ is more efficient than others for $0.8 \leq \rho_{yx} \leq 0.9$ and $0.5 \leq K \leq 1$.

Thus $W = ½$ appears to be a good choice for estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ for moderate to high correlation $0.4 < \rho_{yx} < 0.8$. Once again, the exact efficiencies of $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ are generally of the same order. They approach asymptotic efficiency around $m = 32$. For example in Table 5.1a, when $\rho_{yx} = 0.4$,$K = 1$,$E_1 = 116$

whereby $E_1' = 114$,$E_2' = 115$ and $E_3' = 116$ in Table 5.2a.

In spite of the above observations, it seems that it is difficult to choose among the estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ on the basis of their exact mean square errors. However when considering the upper bound attained from (5.8), that is the values of absolute biases of estimators relative to their standard errors, which is considered as negligible under 0.1, a more preferable choice between the estimators may be provided. Using the i's as expressed in (5.11), the following can be defined.

$$B_r = \frac{\left|Bias(\bar{y}r)\right|}{[MSE(y_r)]^{½}}$$

$$B_i = \frac{\left|Bias(\bar{y}i)\right|}{[MSE(y_i)]^{½}}$$

whose calculated values are shown in Tables 5.3a-h.

Table 5.3a: $|\text{Bias}|/(\text{MSE})^{½}$, with $W = 0.25$ for selected values of m and K (in %)

		$\rho = 0.2$				$\rho = 0.3$			
m	K	B_r	B_1	B_2	B_3	B_r	B_1	B_2	B_3
8	0.25	1.66	0.49	0.16	0.05	1.71	0.49	0.16	0.05
	0.50	9.48	2.96	1.01	0.34	6.68	2.00	0.69	0.23
	1.00	19.74	7.80	2.64	0.90	18.57	7.05	2.38	0.81
	1.50	25.02	12.31	4.06	1.43	24.57	11.88	3.91	1.37
16	0.25	1.23	0.33	0.40	0.01	1.26	0.33	0.04	0.01
	0.50	7.03	2.00	0.29	0.11	4.94	1.35	0.20	0.08
	1.00	14.90	5.31	0.77	0.31	13.99	4.78	0.69	0.28
	1.50	18.56	8.42	1.22	0.50	18.22	8.10	1.17	0.48
20	0.25	1.10	0.29	0.03	0.01	1.13	0.29	0.03	0.01
	0.50	6.34	1.77	0.20	0.08	4.46	1.20	0.13	0.05
	1.00	13.50	4.71	0.53	0.22	12.66	4.42	0.48	0.20
	1.50	16.85	7.48	0.84	0.36	16.54	7.12	0.81	0.34
32	0.25	0.88	0.22	0.01	0.00	0.91	0.23	0.01	0.00
	0.50	5.08	1.38	0.09	0.04	3.56	0.93	0.06	0.02
	1.00	10.86	3.68	0.25	0.11	10.18	3.31	0.22	0.10
	1.50	13.62	5.86	0.39	0.17	13.36	5.64	0.38	0.17

Table 5.3b: $|\text{Bias}|/(\text{MSE})^{1/2}$, with $W = 0.25$ for selected values of m and K (in %)

		$\rho = 0.4$				$\rho = 0.5$			
m	K	B_r	B_1	B_2	B_3	B_r	B_1	B_2	B_3
8	0.25	5.26	1.49	0.51	0.17	8.99	2.50	0.85	0.28
	0.50	3.55	1.02	0.35	0.11	0.00	0.00	0.00	0.00
	1.00	17.29	6.24	2.11	0.71	15.87	5.39	1.81	0.61
	1.50	24.15	11.42	3.78	1.31	23.14	10.89	3.58	1.25
16	0.25	3.89	1.00	0.14	0.06	6.66	1.68	0.24	0.10
	0.50	2.62	0.68	0.10	0.04	0.00	0.00	0.00	0.00
	1.00	12.99	4.22	0.61	0.25	11.89	3.63	0.52	0.21
	1.50	17.91	7.77	1.12	0.46	17.63	7.42	1.06	0.44
20	0.25	3.51	0.89	0.10	0.04	6.01	1.49	0.16	0.07
	0.50	2.40	0.61	0.07	0.02	0.00	0.00	0.00	0.00
	1.00	11.77	3.74	0.42	0.18	10.74	3.22	0.36	0.15
	1.50	16.25	6.90	0.77	0.33	16.00	6.59	0.73	0.31
32	0.25	2.80	0.69	0.04	0.02	4.81	1.16	0.07	0.03
	0.50	1.88	0.47	0.03	0.01	0.00	0.00	0.00	0.00
	1.00	9.43	2.92	0.20	0.08	8.62	2.50	0.16	0.07
	1.50	13.12	5.40	0.36	0.16	12.91	5.15	0.34	0.15

Table 5.3c: $|\text{Bias}|/(\text{MSE})^{1/2}$, with $W = 0.25$ for selected values of m and K (in %)

		$\rho = 0.6$				$\rho = 0.7$			
m	K	B_r	B_1	B_2	B_3	B_r	B_1	B_2	B_3
8	0.25	12.86	3.53	1.20	0.40	16.86	4.57	1.56	0.52
	0.50	4.05	1.05	0.35	0.12	8.74	2.13	0.72	0.24
	1.00	14.28	4.47	1.51	0.51	12.44	3.49	1.17	0.39
	1.50	22.91	10.40	3.41	1.19	22.87	9.90	3.23	1.12
16	0.25	9.58	2.37	0.34	0.14	12.65	3.08	0.44	0.18
	0.50	2.99	0.70	0.10	0.04	6.48	1.43	0.20	0.08
	1.00	10.66	3.00	0.43	0.17	9.26	2.34	0.33	0.13
	1.50	17.45	7.06	1.01	0.42	17.42	6.68	0.95	0.39
20	0.25	8.65	2.10	0.23	0.10	11.44	2.72	0.30	0.13
	0.50	2.69	0.62	0.07	0.03	5.84	1.26	0.14	0.06
	1.00	9.63	2.66	0.29	0.12	8.36	2.07	0.23	0.09
	1.50	15.83	6.26	0.69	0.30	15.80	5.92	0.65	0.28
32	0.25	6.93	1.64	0.11	0.05	9.18	2.12	0.14	0.06
	0.50	2.15	0.48	0.03	0.01	4.67	0.98	0.06	0.03
	1.00	7.72	2.07	0.13	0.06	6.70	1.61	0.10	0.04
	1.50	12.77	4.89	0.32	0.14	12.75	4.61	0.30	0.14

Table 5.3d: $|\text{Bias}|/(\text{MSE})^{1/2}$, with $W = 0.25$ for selected values of m and K (in %)

		$\rho = 0.8$				$\rho = 0.9$			
m	K	B_r	B_1	B_2	B_3	B_r	B_1	B_2	B_3
8	0.25	20.95	5.62	1.91	0.64	25	10	6.69	2.27
	0.50	14.28	3.25	1.11	0.37	20.96	4.42	1.50	0.50
	1.00	10.22	2.43	0.81	0.27	7.27	1.27	0.42	0.14
	1.50	23.23	9.37	3.03	1.05	24.54	8.80	2.82	0.97
16	0.25	15.87	3.79	0.54	0.22	19.24	4.50	0.65	0.26
	0.50	10.66	2.18	0.31	0.13	15.88	2.96	0.42	0.17
	1.00	7.59	1.62	0.23	0.09	5.38	0.84	0.12	0.05
	1.50	17.71	6.28	0.89	0.37	18.78	5.84	0.82	0.34
20	0.25	14.38	3.35	0.37	0.16	17.48	3.99	0.44	0.19
	0.50	9.63	1.93	0.21	0.09	14.39	2.62	0.29	0.12
	1.00	6.84	1.43	0.16	0.06	4.85	0.74	0.08	0.03
	1.50	16.06	5.55	0.61	0.26	17.05	5.16	0.56	0.24
32	0.25	11.58	2.61	0.17	0.07	14.14	3.11	0.20	0.09
	0.50	7.72	1.50	0.10	0.04	11.59	2.03	0.13	0.06
	1.00	5.48	1.11	0.07	0.03	3.88	0.58	0.03	0.01
	1.50	12.97	4.32	0.28	0.13	13.79	4.00	0.26	0.12

Table 5.3e: $|\text{Bias}|/(\text{MSE})^{1/2}$, with $W = 0.50$ for selected values of m and K (in %)

		$\rho = 0.2$				$\rho = 0.3$			
m	K	B_r	B_1	B_2	B_3	B_r	B_1	B_2	B_3
8	0.25	1.66	0.93	0.32	0.11	1.71	0.95	0.32	0.11
	0.50	9.48	5.59	1.90	0.67	6.68	3.85	1.31	0.46
	1.00	19.73	13.77	4.53	1.68	18.56	12.75	4.20	1.56
	1.50	24.27	19.46	6.19	2.42	23.86	19.18	6.09	2.38
16	0.25	1.23	0.65	0.09	0.03	1.26	0.65	0.09	0.04
	0.50	7.03	3.89	0.56	0.25	4.94	2.67	0.39	0.16
	1.00	14.90	9.78	1.42	0.60	13.98	9.02	1.31	0.55
	1.50	18.56	14.17	2.05	0.88	18.22	13.95	2.01	0.86
20	0.25	1.10	0.57	0.06	0.02	1.13	0.58	0.06	0.02
	0.50	6.34	3.47	0.39	0.17	4.45	2.38	0.26	0.11
	1.00	13.49	8.75	0.98	0.43	12.65	8.07	0.91	0.39
	1.50	16.85	12.73	1.43	0.63	16.53	12.53	1.41	0.62
32	0.25	0.88	0.45	0.03	0.01	0.91	0.46	0.03	0.01
	0.50	5.07	2.73	0.18	0.08	3.56	1.87	0.12	0.05
	1.00	10.86	6.91	0.46	0.21	10.17	6.37	0.42	0.19
	1.50	13.61	10.13	0.68	0.31	13.35	9.97	0.67	0.31

Table 5.3f: $|\text{Bias}|/(\text{MSE})^{1/2}$, with $W = 0.50$ for selected values of m and K (in %)

		$\rho = 0.4$				$\rho = 0.5$			
m	K	B_r	B_1	B_2	B_3	B_r	B_1	B_2	B_3
8	0.25	5.26	2.89	0.99	0.34	8.99	4.89	1.67	0.58
	0.50	3.54	1.98	0.67	0.23	0.00	0.00	0.00	0.00
	1.00	17.29	11.64	3.85	1.42	15.87	10.41	3.45	1.26
	1.50	23.47	18.95	6.01	2.35	23.14	18.80	5.94	2.33
16	0.25	3.89	2.00	0.29	0.12	6.66	3.39	0.49	0.20
	0.50	2.61	1.37	0.20	0.08	0.00	0.00	0.00	0.00
	1.00	12.98	8.21	1.19	0.50	11.89	7.31	1.06	0.44
	1.50	17.90	13.77	1.98	0.85	17.63	13.64	1.96	0.84
20	0.25	3.51	1.78	0.20	0.08	6.01	3.01	0.34	0.14
	0.50	2.35	1.22	0.13	0.06	0.00	0.00	0.00	0.00
	1.00	11.74	7.33	0.82	0.36	10.74	6.52	0.73	0.32
	1.50	16.24	12.36	1.39	0.61	16.00	12.25	1.37	0.60
32	0.25	2.80	1.40	0.09	0.04	4.81	2.36	0.15	0.07
	0.50	1.88	0.96	0.06	0.02	0.00	0.00	0.00	0.00
	1.00	9.43	5.78	0.38	0.17	8.62	5.14	0.34	0.15
	1.50	13.12	9.83	0.66	0.30	12.91	9.73	0.65	0.30

Table 5.3g: $|\text{Bias}|/(\text{MSE})^{1/2}$, with $W = 0.50$ for selected values of m and K (in %)

		$\rho = 0.6$				$\rho = 0.7$			
m	K	B_r	B_1	B_2	B_3	B_r	B_1	B_2	B_3
8	0.25	12.86	6.97	2.36	0.83	16.86	9.03	3.06	1.08
	0.50	4.05	2.13	0.72	0.25	8.74	4.42	1.51	0.52
	1.00	14.28	9.04	3.00	1.09	12.44	7.46	2.49	0.89
	1.50	22.91	18.78	5.91	2.33	22.87	19.04	5.93	2.35
16	0.25	9.58	4.81	0.70	0.29	12.65	6.27	0.91	0.38
	0.50	2.99	1.47	0.21	0.08	6.48	3.04	0.44	0.18
	1.00	10.66	6.31	0.91	0.38	9.26	5.18	0.75	0.31
	1.50	17.45	13.62	1.95	0.84	17.42	13.79	1.97	0.85
20	0.25	8.65	4.28	0.48	0.21	11.44	5.58	0.63	0.27
	0.50	2.69	1.30	0.14	0.06	5.84	2.71	0.30	0.13
	1.00	9.63	5.63	0.63	0.27	8.36	4.61	0.51	0.22
	1.50	15.83	12.22	1.37	0.60	15.80	12.37	1.38	0.61
32	0.25	6.93	3.36	0.22	0.10	9.18	4.38	0.29	0.13
	0.50	2.15	1.02	0.06	0.03	4.67	2.12	0.14	0.06
	1.00	7.72	4.42	0.29	0.13	6.70	3.62	0.24	0.11
	1.50	12.77	9.71	0.65	0.30	12.75	9.82	0.65	0.30

Table 5.3h: $|\text{Bias}|/(\text{MSE})^{1/2}$, with $W = 0.50$ for selected values of m and K (in %)

		$\rho = 0.8$				$\rho = 0.9$			
m	K	B_r	B_1	B_2	B_3	B_r	B_1	B_2	B_3
8	0.25	20.95	11.17	3.77	1.33	25.10	13.35	4.48	1.59
	0.50	14.28	6.89	2.35	0.81	20.96	9.57	3.25	1.12
	1.00	10.22	5.60	1.87	0.66	7.27	3.25	1.08	0.37
	1.50	23.23	19.87	6.09	2.44	24.54	22.36	6.56	2.70
16	0.25	15.87	7.77	1.13	0.47	19.24	9.31	1.35	0.56
	0.50	10.66	4.74	0.69	0.28	15.88	6.58	0.95	0.39
	1.00	7.59	3.84	0.55	0.23	5.38	2.20	0.31	0.13
	1.50	17.71	14.38	2.04	0.88	18.78	16.17	2.25	0.98
20	0.25	14.38	6.92	0.78	0.33	17.48	8.30	0.93	0.40
	0.50	9.63	4.21	0.47	0.20	14.39	5.85	0.66	0.28
	1.00	6.84	3.42	0.38	0.16	4.85	1.95	0.21	0.09
	1.50	16.06	12.90	1.43	0.63	17.05	14.50	1.58	0.71
32	0.25	11.58	5.44	0.36	0.16	14.14	6.52	0.43	0.20
	0.50	7.72	3.30	0.22	0.10	11.59	4.58	0.30	0.14
	1.00	5.48	2.67	0.17	0.08	3.88	1.52	0.10	0.04
	1.50	12.97	10.23	0.68	0.31	13.79	11.49	0.76	0.35

(In Tables 5.3a-h values for B_r, B_1, and B_2 have been taken from Chakrabarty 1979 and the other values have been computed for this study)

From Tables 5.3a-d for $W = 1/4$, the following is observed.

(i) B_2 and B_3 are generally less than 1%; B_1 is less than 10% for $m = nh \geq 16$.

(ii) The customary ratio estimator is badly biased (> 10%) for $K \geq 1$.

From Tables 5.3e-h for $W = 1/2$ the following is observed.

(i) $B_2 < 1\%$ for $K \leq 1$ and less than 2.5% for $K > 1$ when $m \geq 16$.

$B_3 < 1\%$ for all K, ρ_{yx}, and $m \geq 16$.

(ii) B_r exceeds 10% and is considerably higher than B_i's at higher correlations. Thus, for $\rho_{yx} = 0.9$ and $0.5 \leq K \leq 1$, although $MSE(\bar{y}_r) \leq MSE(\bar{y}_i)$, estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ may be preferable in situations where the freedom from bias is desirable.

5.6 Conclusion

The estimators $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ have comparatively smaller bias than the customary ratio estimator $\bar{y}_r$, and also are more efficient for a wide range of values of ρ_{yx}, K and m, and with a ratio of absolute value of bias to standard error less than 10%. For $\bar{y}_{MOD}$ the ratio is less than 1% for most values of ρ_{yx}, K and m. The study will, therefore, make use of $\bar{y}_{MOD}$ in the estimation of area of the crop of interest.

Chapter six

Data analysis

6.1 Introduction

The Tanzania Mainland's agricultural sample survey (Agsasu), from the Bureau of Statistics, is designed so far to give only national estimates. The sampling scheme adopted by Agsasu does not allow much analysis in spite of a massive collection of data. The village estimates correspond to the superstrata estimates, there being only one village sampled per superstratum. The superstratum from the Agsasu design as detailed in chapter two (section 2.3.1.2) is the combination of three strata. The national mean of the survey is a weighted mean of the village means. The only estimate of the variance is that from households within villages. For example the population total Y is estimated by

$$\hat{Y} = \sum_{\forall h} \frac{A_h}{A_{hi}} \frac{M_{hi}}{M_{hij}} \frac{N_{hij}}{n_{hij}} \sum Y_{hijk}.$$

The only variable in the estimate is Y_{hijk} from measurement in the k^{th} household in the j^{th} village of the i^{th} ward selected for the stratum h. All the other factors $\frac{A_h}{A_{hi}}, \frac{M_{hi}}{M_{hij}}$, and $\frac{N_{hij}}{n_{hij}}$ are simply weighting factors. Yet to get anywhere with the analysis, the variances between villages within a superstratum have to be estimated. In other words there must be at least two villages, but better still three to five villages per superstratum if resources are available. It is then possible to calculate an estimate of variance between villages within superstrata. This will make Agsasu work on one hundred villages or more instead of fifty.

6.1.1 Variance estimates of means - Agsasu 1987/88

Let us consider a sample of n wards (psu's) in Agsasu data. The estimate of a village mean $\bar{y}_j$ is given by

$$\bar{y}_j = \frac{1}{m} \sum_{k=1}^{m} \bar{y}_{jk}$$

for the j^{th} village with m households.

Since each psu has only one selected village, the estimate of the overall (population) mean for n villages is therefore given by

$$\bar{y}_o = \frac{1}{n} \sum_{j=1}^{n} \bar{y}_j .$$

Given that the variance for within wards is

$$s_w^2 = \frac{\sum_{j=1}^{n} \sum_{k=1}^{m} (y_{jk} - \bar{y}_j)^2}{n(m-1)}$$

and the variance for between wards is

$$s_b^2 = \frac{\sum_{j=1}^{n} (\bar{y}_j - \bar{y}_o)^2}{(n-1)}$$

then the variance for overall mean $(\bar{y}_o)$ is

$$v(\bar{y}_o) = \frac{1-f_j}{n} s_b^2 + \frac{f_j(1-f_w)}{n\,m} s_w^2 \qquad (6.1)$$

where f_w and f_j are sampling fractions of the households and villages respectively.

In very general terms, villages vary more than households within villages. n must be less than $n\,m$ unless one household per village is selected so that the second term in (6.1) disappears. In the case of Agsasu the first term in (6.1) disappears, because of selecting one village per ward and only one ward in a stratum, yet the second term is likely to be very small, because $n\,m = 3106$, the total households sampled. If s_b^2 could be estimated, the first term would be relatively larger. An estimate of the variance based on the within villages variance is therefore likely to give a gross under-estimate.

Conversely, treating the sample as a simple random sample of households ignoring the design structure, the estimate of variance is likely to be a gross over-estimate.

The variance for a simple random sample is given as

$$s_{sr}^2 = \frac{\sum_{j=1}^{n}\sum_{k=1}^{m}(y_{jk} - \bar{y}_o)^2}{(n\,m - 1)} \tag{6.2}$$

The best that is supposed can be done is to calculate both (6.1) and (6.2) knowing that the true variance must lie somewhere between the two.

One possible way to approach the true estimate appropriate in this (Agsasu) case is to use one class of variance estimator known as the collapsed stratum variance estimator which has a positive bias if the true stratum totals of the collapsed pairs differ to a large extent (Cochran 1977). With L, the number of strata, being even, an estimate may be attempted by grouping the strata in pairs. From a typical pair of village's totals y_{i1}, y_{i2} where $i = 1,2,\ldots,L/2$, the respective stratum totals are $\hat{Y}_{i1} = N_{i1}\,y_{i1}$ and $\hat{Y}_{i2} = N_{i2}\,y_{i2}$. The variance is therefore given by

$$v(\hat{Y}_{cst}) = \sum_{i=1}^{L/2}(\hat{Y}_{i1} - \hat{Y}_{i2})^2 \tag{6.3}$$

and $v(\bar{y}_{cst}) = v(\hat{Y}_{cst})/N^2$, where $N = \sum_{h=1}^{L} N_h$. However when L is odd, at least one group must clearly be of size different from two. The extension of (6.3) to G groups of any chosen sizes $L_g \geq 2$ is

$$v(\hat{Y}_{cst}) = \sum_{g=1}^{G}\frac{L_g}{L_g - 1}\sum_{t=1}^{L_g}\left[\hat{Y}_{gt} - \frac{\hat{Y}_g}{L_g}\right]^2 \tag{6.4}$$

where $\hat{Y}_g$ is the estimated total for group g.

Although the Agricultural sample survey for 1987/88 has covered the entire country, the analysis has been restricted to the study area only, namely Arusha Region, Mbeya Region and Mwanza Region, for Tanzania Mainland. The sampled village populations in these regions and household sizes are presented in Table 6.1.

Table 6.1: Village population of Agsasu and household size.

Village	Population Male	Population Female	Population Total	Household Number	Household Av.Size
Himiti	1286	1201	2487	425	5.9
G-Lumbo	1550	1498	3048	485	5.8
Ngoha	867	922	1789	440	4.0
Matwiga	1389	1373	2762	571	4.8
Kanyelele	1360	1426	2786	433	7.0
Irunda	1025	1050	2075	373	6.3
Kagunga	1277	1350	2627	321	7.5
				Mean	5.9

Source: (i) National Census 1988, and

(ii) Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar-es-salaam.

Table 6.2 shows the following area variance estimators of the means for maize area as calculated using the Agsasu data 1987/88.

(i) An estimate of variance of the overall mean - $v(\bar{y}_o)$.

(ii) An estimate of variance treating the sample as simple random and ignoring design structure - $v(\bar{y}_{sr})$.

(iii) An estimate of variance based on collapsed strata - $v(\bar{y}_{cst})$.

Table 6.2: Variance estimators of the maize mean areas

$v(\bar{y}_o)$	$v(\bar{y}_{sr})$	$v(\bar{y}_{cst})$
0.0150652	0.564859	0.0330842

Source: Agricultural sample survey, 1987/88

Bureau of Statistics, Dar es salaam.

6.2 Farmer - Enumerator relationship on area - Agsasu 1987/88

In Agsasu, area planted (APL) is the information on the plot planted area as given by the farmer during the first stage of the survey. A corresponding information by enumerators is achieved at the second stage through physical measuring, that is, measured area (MAR). There is yet another area information of the same plot in the fourth and last stage of the survey work, given by the farmers after harvest, area harvested (AHV).

Correlation coefficients between these three independent variables for each village in the sampled wards were calculated. The table below shows the coefficients under different categories.

Table 6.3: Correlation coefficients for different areas by villages

Village	Sample Size	Coefficients between (APL - MAR)	(APL - AHV)	(MAR - AHV)
		Crops under all combinations		
Himiti	13	0.883	0.491	0.660
G-Lumbo	5	0.995	0.967	0.976
Ngoha	20	0.483	0.154	0.510
Matwiga	25	0.672	0.374	0.497
Kagunga	10	0.939	-0.323	-0.307
		Under maize		
Ngoha	19	0.475	0.136	0.496
Matwiga	11	0.463	-0.122	0.361
Kagunga	5	0.974	-0.314	-0.319
		Under maize+bean		
G-Lumbo	3	0.997	0.961	0.979
Matwiga	7	0.435	0.165	0.448
		Under maize+other		
Himiti	9	0.877	0.233	0.585
Matwiga	7	0.557	0.825	0.675
Kanyelele	12	-0.081	-0.529	-0.050
Irunda	3	0.922	1.000	0.922
Kagunga	5	0.924	0.094	-0.123

Source: Agricultural sample survey, 1987/88

Bureau of Statistics, Dar es salaam.

The distribution of the sample correlation coefficient $\hat{\rho}$ is shown (Fisher 1958) to depend only on ρ the population correlation coefficient and n the sample size. However the distribution of $\hat{\rho}$ is far from normal, for large values of ρ, being sharply skewed in the neighbourhood of $\rho = \pm 1$. For large sample size $(n > 400)$ and ρ moderately large, $\hat{\rho}$ is approximately normal with mean ρ and variance $(1-\rho^2)^2/n$ (Chaudhary 1973). So the significance of $\hat{\rho}$ can be determined regarding the statistic

$$Z = \frac{(\hat{\rho}-\rho)\sqrt{n}}{1-\rho^2}$$

as standard normal. On the other hand when n is small and ρ large, the distribution of $\hat{\rho}$ is skewed, the more so as ρ increases. Fisher (1958) through his transformation, showed that if

$$z = ½\ln\frac{1+\hat{\rho}}{1-\hat{\rho}} = 1.1513\log_{10}\frac{1+\hat{\rho}}{1-\hat{\rho}} \tag{6.5}$$

then z is distributed normally with mean $\mu_z = ½\ln\frac{1+\rho}{1-\rho}$ and variance $\frac{1}{(n-3)}$.

A homogeneity test for the estimated correlation coefficients in Table 6.3 can be done using a χ^2-test by defining z_i's by relation (6.5). Therefore a statistic $u = \sum(n_i-3)(z_i-\bar{z})^2$ is approximately a χ^2-distribution with $(k-1)$ degrees of freedom (d.f.) (Chaudhary 1973), where

$$\bar{z} = \frac{\sum(n_i-3)\,z_i}{\sum(n_i-3)} \qquad i = 1,2,\cdots,k.$$

Table 6.4 summarizes the χ^2-test for the various categories of crop combination with maize.

The estimated correlation coefficients for APL-MAR and MAR-AHV under all combinations of maize, APL-MAR and APL-AHV under maize+other, were further tested because of their significant difference at $\alpha = 5\%$. A test for differences on pairs of villages' estimated correlation coefficients, using the normal distribution was performed and the results are shown in Table 6.5.

Table 6.4: Variance estimators of the means.

Crop	d.f.	APL - MAR	APL - AHV	MAR - AHV
All combinations	4	17.747**	9.998*	11.337**
Maize only	2	5.049	0.619	1.386
Maize+bean	1	6.202*	2.570	2.565
Maize+other	4	10.621**	22.365**	4.943

* Significant at $\alpha = 10\%$

** Significant at $\alpha = 5\%$

Table 6.5: Test for differences on pairs of villages' estimated correlation coefficients

Villages Pairs	Crops under all combinations of maize	
	APL - MAR	MAR - AHV
Himiti - G-Lumbo	-2.072*	-1.824
Himiti - Ngoha	2.164*	0.577
Himiti - Matwiga	1.507	0.649
Himiti - Kagunga	-0.691	2.252*
G-Lumbo - Ngoha	3.301*	2.197*
G-Lumbo - Matwiga	2.952*	2.248*
G-Lumbo - Kagunga	1.578	3.146*
Ngoha - Matwiga	-0.890	0.054
Ngoha - Kagunga	-2.678*	1.959*
Matwiga - Kagunga	-2.109*	1.988*
	Maize+other	
	APL - MAR	APL - AHV
Himiti - Matwiga	1.137	-1.448
Himiti - Kanyelele	2.739*	1.567
Himiti - Irunda	-0.222	-3.299*
Himiti - Kagunga	-0.310	0.175
Matwiga - Kanyelele	1.181	2.931*
Matwiga - Irunda	-0.871	-2.350*
Matwiga - Kagunga	-1.140	1.245
Kanyelele - Irunda	-1.597	-4.164*
Kanyelele - Kagunga	-2.171*	-0.874
Irunda - Kagunga	-0.011	3.026*

* Significant at $\alpha = 5\%$ $|z_{\alpha/2}| = 1.96$

Source: Agricultural sample survey, 1987/88

Bureau of Statistics, Dar es salaam.

6.2.1 Summary

The results obtained from Tables 6.3 - 6.5 can be summarized as follows.

1. There was significant difference, at $\alpha = 5\%$, between the coefficients on APL - MAR and MAR - AHV, and no difference on APL - AHV taken by villages.

The significant differences on APL - MAR both by village and maize + beans categories suggest that there is a great disparity on areas given by farmers compared to the measurements by enumerators. This is again reflected on the relationship between MAR and AHV. Rarely do farmers use tapes or any scientific instrument to measure their field/plot areas. Some normally use steps pacing around their fields/plots. Others associate daily work capacity to area. These methods lack proper conversion to standard units, as steps or working capacity differ from one farmer to another. Sometimes farmers use slightly better units of measures, like number of seeds or total seeds weight covering certain area, but as will be shown in Figure 6.1, that still is not a good enough measure to give accurate area. The farmer may unconsciously change planting pattern, or may favour particular spots while planting. The technique, though slightly better, is therefore not an appropriate device or mechanism of translation into standard units.

There seems to be a surprisingly systematic agreement between farmer's area planted, APL, and area harvested, AHV. This on one hand, suggests that farmers are consistent in their estimation, or would not bother much about harvested area when looking forward for produce, except when the produce does not happen as expected. Therefore they would give just a rough estimate compared to area planted. On the other hand, it casts some doubt on farmers' methods or ways of estimating land area. Under the rainfed farming, most inputs are occasional out of the control of the farmer, due to things like weather, fertilizer and even seeds; it is therefore evident that area harvested, AHV, is supposed to be less than or at most equal to area planted, APL. One has to take in mind the fact that the sowed surfaces are generally much larger than the harvestable ones (Bartholome 1986). The author expressed his experience in Sahelian countries which is true in most of African countries.

2. For area under maize only there was no difference between all the three relationships.

Maize plant spacing in most parts of the country is done uniformly in accordance with the advice of agricultural officers. Therefore, in the case of a field/plot of pure maize, under normal circumstances, it becomes simple for the farmer to convert the area by associating with plant rows or columns as the case may be, in contrast to bean or even sorghum where no rows or columns are distinguishable and no specific pattern is followed.

3. There was no difference on MAR - AHV but significant difference on the rest in the case of maize + other.

Maize + other category is as complicated to the enumerator as it is to the farmer. If training and supervision are not taken seriously, such results are bound to happen.

4. There was significant difference on APL - MAR and no difference on the rest in the case of maize + beans.

In contrast to maize + other, maize + bean category seems to be less complicated for enumerator and farmer, especially when there is proper plant spacing. This difference can be associated with the method applied in the area measurement.

Generally, the difference between APL and MAR suggests that there is still a need for making physical area measurements while the accuracy of the farmer is still questionable. When maize is mixed with several other crops, the farmer's estimate becomes even less reliable. Besides, as mentioned earlier, farmers are more keen on production rather than area, because of trading purposes. Only when a farmer contemplates selling a farm or renting is interest in area measurement developed.

6.3 The 1990 field survey

The author made a field survey (Mussa Field Survey) in the three regions of Tanzania from February to October 1990. The regions are outlined in chapter four (section 4.4) and shown in Figure 4.6. The time was chosen to coincide with the early growth stage of maize up to the harvesting time. Maize is the main staple food crop grown in most parts of the country. It is normally planted during the months of December to early February and harvested between May and August depending on the region and village. Although the survey has covered all crops grown in those selected areas, the analysis will be limited to pure maize stands, pure beans stands and the combination of maize with beans and with other crops. "Maize and beans" is a common mixed crop in many parts of the country. The proportion of maize/beans is not the same between farmers and does not remain the same within farmers between seasons.

6.3.1 Concepts and definitions

The concepts and definitions which have been used are the ones contained in the FAO's programme for the 1990 world census of agriculture (FAO 1986). On some occasions more probing was done before a definition was applied. This was necessary to avoid misrepresentation, especially in the concept and definition of a household.

An agricultural holding is defined as an economic unit of agricultural production under single management comprising all livestock kept and land used wholly or partly for agricultural production purposes, without regard to title, legal form or size. This is important when coming across farmers who either rent some land or hire for a farming season. In the former case it is not included in the holding. A household concept is one of the basic elements of national statistics systems. The concept, according to United Nations, is based on arrangements made by persons, individually or in groups, for providing themselves with food or other essentials for living. Various types of households exist; a one-person, multiperson, extended family household, *etc.* A better way to comply with a given definition is to consider explanation from members of households as well. This is but one way which may eliminate confusion in the otherwise universal problem of frame listing.

6.3.2 Methodology

The first stage of the survey was to locate the regions to be studied as explained in chapter four (section 4.4). Within a region a district is selected and finally a village. The villages, as also explained earlier, were chosen depending on the availability of the images. The images were obtained from the Remote Sensing Centre, Nairobi, Kenya, and at the Institute of Resources Assessment at the University of Dar es salaam.

The second stage was to list all holdings in the selected villages. This was done with the help of village leaders, as they had lists completed not more than two years ago. Therefore, there was a matter of updating the lists, which involved verifying from the ten cell representatives, known as *balozi*. The *balozi* is assigned to lead about 20 households acting as arbitrator for minor disputes. The *balozis* who keep particulars of all the households are unpaid officials without office, *etc.* They have to work within the structure of their villages, looking after their families, land, livestock, trading, shops, *etc.*, as any other villager. It is therefore difficult to arrange meetings with them during the working day. Making up-to-date frames can therefore be time consuming.

The third stage was selecting households for interview and area measurements. Thirty (30) households were selected in each village by a simple random process. The sample size was a deliberate one in order to accommodate the analysis in both situations of large and small samples. Table 6.6 shows the sampled village populations and household sizes.

The last stage was to visit the selected holders at their homes as well as at their farms for area measurements of their holdings. The fields may be at one place or scattered within the locations. This stage is even more difficult than locating *balozi*. At this stage it is necessary to have the assistance of a villager, well conversant with the locations, who guides and also participates in the area measurement process.

Table 6.6: Village population and household size for 1990 field survey.

Village	Location	Population Male	Population Female	Population Total	Household Number	Household Av.Size
Nyankumbu	Nyantorotoro	403	398	801	157	5.1
	Mbugani	806	796	1602	314	
Igawilo	Iganjo	1331	1499	2930	658	4.3
	Igawilo	2267	2555	4822	1121	
Ngarash	Juu	658	651	1309	262	5.0
	Kati	772	764	1536	307	
Sinoni	Bondeni	851	843	1694	339	5.0
	Milimani	622	615	1237	247	

Source: (i) National Census 1988, and

Bureau of Statistics, Dar-es-salaam.

(ii) Mussa Field Survey.

6.3.3 Survey work

The survey focused on area measurement for respective crops grown, household and others like fallow land, livestock land, *etc.* The method of traverse, explained in chapter two (section 2.2.1.4.3) was adopted in this survey because of its simplicity. Tape and compass were required. Because it was difficult to carry the ranging rods, ordinary long straight sticks available in the villages were used instead. Production from these fields provided by the farmers in the units which they understand was also recorded. This issue on production will be discussed in detail in chapter seven.

6.3.4 Problems in the survey

There were several problems in the survey process, the main one being transport. As seen in Figure 4.6 the selected regions are well over 500*km* from Dar es salaam and from each other. Public transport was used but was unreliable and difficult. Movement within villages was by walking, a daily walking distance of four to five kilometres.

Another problem was administration. This involved reporting to various administrative levels, from regional to village level in order to get letters of authority. The process could take from three days to a week, where local festivals or holidays were involved.

6.4 Method of analysis

As indicated earlier the analysis is limited to the two main staple food crops, maize and beans, for our practical purpose. The crops of following categories are carried out:-

- Maize
- Maize and beans
- Maize and others
- Beans.

The methods use for comparison were:-

(i) Customary ratio

(ii) Separate ratio

(iii) Combined ratio

(iv) Stratified (within village)

(v) Chakrabarty ratio

(vi) Modified ratio.

The customary ratio estimator is that which had been considered in chapter five from a ratio $r = \frac{\bar{y}}{\bar{x}}$, whose mean estimator is $\bar{y}_r = r\bar{X}$ which has a variance of

$$v(\bar{y}_r) = \frac{s_y^2}{n}\left\{1 + K(K - 2\rho_{yx})\right\}$$

omitting the finite population correction factor. K and ρ_{yx} are as explained in chapter five.

6.4.1 Separate ratio

In the separate ratio estimate, the total of each stratum is attained and then these are added. Given y_h, x_h as the totals in the h^{th} stratum, and that X_h is the stratum total, the total estimate of $\hat{Y}_{RS}$ is given by

$$\hat{Y}_{RS} = \sum_h \frac{y_h}{x_h} X_h = \sum_h \frac{\bar{y}_h}{\bar{x}_h} X_h.$$

In this estimator, there is no need to assume that the true ratio remains constant from stratum to stratum. Also the estimate requires a knowledge of the separate totals X_h.

Result (6.1);(Cochran 1977): If an independent simple random sample is drawn in each stratum and sample sizes are large in all strata, then the variance of the population estimate for separate ratio is

$$V(\hat{Y}_{RS}) = \sum_h \frac{N_h^2(1-f_h)}{n_h}(S_{yh}^2 + R_h^2 S_{xh} - 2R_h \rho_h S_{yh} S_{xh})$$

where $R_h = \dfrac{Y_h}{X_h}$ is the true ratio in stratum h.

From result (6.1), the estimated variance is given by

$$v(\hat{Y}_{RS}) = \sum_h \frac{N_h^2(1-f_h)}{n_h}(s_{yh}^2 + r_h^2 s_{xh}^2 - 2r_h \hat{\rho}_h s_{yh} s_{xh}).$$

It is to be expected, however that the sample in each stratum should be relatively large enough so that the approximate variance formula applies in each stratum. When the sample is small and strata are many, the mean square error will be relatively large as explained in chapter five (section 5.2).

An alternative estimate to the separate ratio is derived from a single combined ratio.

6.4.2 Combined ratio

From the sample data the followings are computed.

$$\hat{Y}_{st} = \sum_h N_h \bar{y}_h, \quad \hat{X}_{st} = \sum_h N_h \bar{x}_h.$$

These are the standard estimates of population totals Y and X respectively made from a stratified (hence subscript st) sample. The combined ratio population estimate $\hat{Y}_{RC}$ is therefore given by

$$\hat{Y}_{RC} = \frac{\hat{Y}_{st}}{\hat{X}_{st}} X = \frac{\bar{y}_{st}}{\bar{x}_{st}} X$$

where $\bar{y}_{st} = \hat{Y}_{st}/N$, $\bar{x}_{st} = \hat{X}_{st}/N$ are the estimated population means from a stratified sample. The population variance in relation to the result (6.1) above is given by

$$V(\hat{Y}_{RC}) = \sum_h \frac{N_h^2(1-f_h)}{n_h} (S_{yh}^2 + R^2 S_{xh}^2 - 2R\,\rho_h\, S_{yh}\, S_{xh})$$

whose estimated variance is given by

$$v(\hat{Y}_{RC}) = \sum_h \frac{N_h^2(1-f_h)}{n_h} (s_{yh}^2 + r^2 s_{xh}^2 - 2\,r\,\hat{\rho}_h\, s_{yh}\, s_{xh}).$$

The estimate $\hat{Y}_{RC}$ does not require a knowledge of the X_h but only of X. Also the combined estimate is much less subject to the risk of bias than the separate estimate (Cochran 1977). With only a small sample in each stratum the combined estimate is preferable unless there is a wide difference between the strata ratios R_h. When there are such differences and regrouping of the strata is possible so that each group will not differ much and has a large sample size, the separate ratio estimator should thus be used.

In both cases of separate and combined ratio, the variance of the estimate of their means is achieved by dividing with their respective squares of the sum of stratum totals ($\left[\sum_h N_h\right]^2$).

6.4.3 Stratified (within village)

The poststratification methodology will be used, that is, stratification after data have been collected. The stratification was based on geographical locations of the village. This is because the variable that was suitable for stratification, that is the area of a village occupied by crops was not known, before the data collection. The initial analysis of images was done during the data collection in the villages. The estimated variance of the mean for stratified (within villages) $v(\bar{y}_{st})$ is therefore given (Cochran 1977) as

$$v(\bar{y}_{st}) = \sum_h W_h^2 (1 - f_h) \frac{s_h^2}{n_h}$$

where W_h is the stratum proportion.

6.4.4 Chakrabarty and Modified ratios

The asymptotic variances of Chakrabarty (1979) estimates and Modified estimate as found in chapter five (section 5.4) are used with a correction factor $(1 - f_v)$, where f_v is the sampling fraction of the village. The values $W = \rho_{yx} C_x / C_y$ or 1, as theoretically the values of W must be between 0 and 1, and $K = C_x / C_y$. These values are shown in Table 6.7 with other relevant general values necessary for calculating the asymptotic variances, among other parameters, for Separate ratio, Combined ratio and Stratified estimates as calculated from the Mussa Field Survey data.

Table 6.7a: Maize general statistics by villages

Village	$\bar{x}$	$\bar{y}$	s_x	s_y	$\hat{\rho}_{yx}$	C_x	C_y	W	K
Iga89	1.14	0.48	0.81	0.41	0.77	0.70	0.83	0.91	0.84
Iga90	1.15	0.49	0.81	0.46	0.63	0.69	0.92	0.84	0.75
Nga89	4.44	1.42	2.27	1.88	0.85	0.50	1.30	1.00	0.39
Nga90	4.54	1.41	2.24	1.84	0.78	0.49	1.29	1.00	0.38
Sin89	1.28	0.22	0.81	0.44	0.41	0.62	1.94	1.00	0.32
Sin90	1.35	0.25	0.79	0.50	0.55	0.57	1.95	1.00	0.30

Table 6.7b: Maize+bean general statistics by villages

Village	$\bar{x}$	$\bar{y}$	s_x	s_y	$\hat{\rho}_{yx}$	C_x	C_y	W	K
Nya89	4.86	1.41	3.12	0.97	0.31	0.63	0.67	0.33	0.94
Nya90	4.93	1.54	3.03	0.77	0.46	0.61	0.49	0.37	1.25
Nga89	4.44	0.88	2.27	0.77	0.04	0.50	0.87	0.07	0.58
Nga90	4.54	0.91	2.24	0.81	0.07	0.49	0.87	0.12	0.56
Sin89	1.28	0.62	0.81	0.48	0.46	0.62	0.77	0.57	0.81
Sin90	1.35	0.57	0.79	0.42	0.15	0.57	0.72	0.19	0.79

Table 6.7c: Beans general statistics by villages

Village	$\bar{x}$	$\bar{y}$	s_x	s_y	$\hat{\rho}_{yx}$	C_x	C_y	W	K
Iga89	1.14	0.05	0.81	0.10	0.25	0.70	1.92	0.69	0.37
Iga90	1.15	0.05	0.81	0.11	0.41	0.69	1.96	1.00	0.35
Nga89	4.44	0.47	2.27	0.73	0.52	0.50	1.53	1.00	0.33
Nga90	4.54	0.43	2.24	0.69	0.49	0.49	1.59	1.00	0.31

Key:

Nya89;Nya90 ═ Nyankumbu 1989;Nyankumbu 1990
Iga89;Iga90 ═ Igawilo 1989;Igawilo 1990
Nga89;Nga90 ═ Ngarash 1989;Ngarash 1990
Sin89;Sin90 ═ Sinoni 1989;Sinoni 1990

Source: Mussa Field Survey.

The following tables give the area estimates of the asymptotic variance of means for different villages in two years, for different crop categories.

Table 6.8a: Maize area variance estimators of the means for Igawilo

Estimator	1989	1990
Ratio	0.002271	0.004282
Separate ratio	0.002616	0.006548
Combined ratio	0.002622	0.006577
Stratified	0.006273	0.009758
Chakrabarty	0.002241	0.004182
Modified	0.002252	0.004199

Table 6.8b: Maize area variance estimators of the means for Ngarash

Estimator	1989	1990
Ratio	0.055023	0.58981
Separate ratio	0.032809	0.047012
Combined ratio	0.046156	0.055954
Stratified	0.106156	0.105841
Chakrabarty	0.055023	0.058981
Modified	0.052840	0.057267

Table 6.8c: Maize area variance estimators of the means for Sinoni

Estimator	1989	1990
Ratio	0.005133	0.006040
Separate ratio	0.005946	0.005002
Combined ratio	0.005928	0.005196
Stratified	0.006881	0.006599
Chakrabarty	0.005133	0.006040
Modified	0.005114	0.005978

Table 6.8d: Maize+bean area variance estimators of the means for Nyankumbu

Estimator	1989	1990
Ratio	0.038142	0.025989
Separate ratio	0.021204	0.012631
Combined ratio	0.023018	0.014134
Stratified	0.029613	0.020584
Chakrabarty	0.026496	0.014589
Modified	0.027245	0.015440

Table 6.8e: Maize+bean area variance estimators of the means for Ngarash

Estimator	1989	1990
Ratio	0.023993	0.025655
Separate ratio	0.026365	0.027089
Combined ratio	0.026888	0.027881
Stratified	0.020401	0.022826
Chakrabarty	0.018652	0.020615
Modified	0.018695	0.020690

Table 6.8f: Maize+bean area variance estimators of the means for Sinoni

Estimator	1989	1990
Ratio	0.006599	0.007751
Separate ratio	0.007007	0.006382
Combined ratio	0.006860	0.006296
Stratified	0.007427	0.004157
Chakrabarty	0.005733	0.005453
Modified	0.005855	0.005508

Table 6.8g: Bean area variance estimators of the means for Igawilo

Estimator	1989	1990
Ratio	0.000311	0.000331
Separate ratio	0.001567	0.000297
Combined ratio	0.001565	0.000303
Stratified	0.001595	0.000389
Chakrabarty	0.000307	0.000331
Modified	0.000307	0.000330

Table 6.8h: Bean area variance estimators of the means for Ngarash

Estimator	1989	1990
Ratio	0.012877	0.011921
Separate ratio	0.012825	0.012438
Combined ratio	0.013344	0.012824
Stratified	0.017424	0.015895
Chakrabarty	0.012877	0.011921
Modified	0.012761	0.011831

Source: Mussa Field Survey.

The exact biases and variances of $\bar{y}_r$, $\bar{y}_{C1}$, $\bar{y}_{C2}$ and $\bar{y}_{MOD}$ as derived in chapter five, section 5.5.2 are shown in Tables 6.9a-c and 6.10a-c, and whose respective MSEs are shown in Tables 6.11a-c for maize, maize+bean and bean for different villages. Tables 6.12a-c reflect the exact efficiencies also for maize, maize+bean and bean for different villages.

Table 6.9a: Biases for maize area based on exact theory

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Igawilo	1989	0.000690	0.000630	0.000021	0.000010
	1990	0.001264	0.001062	0.000035	0.000017
Ngarash	1989	-0.014412	-0.014412	-0.000244	-0.000119
	1990	-0.011974	-0.011974	-0.000194	-0.000095
Sinoni	1989	-0.000803	-0.000803	-0.000021	-0.000010
	1990	-0.002418	-0.002418	-0.000053	-0.000025

Table 6.9b: Biases for maize+bean area based on exact theory

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Nyankumbu	1989	0.012670	0.0041770	0.000114	0.000055
	1990	0.012195	0.004506	0.000115	0.000055
Ngarash	1989	0.006881	0.000479	0.000008	0.000004
	1990	0.006429	0.000799	0.000013	0.000006
Sinoni	1989	0.003499	0.001970	0.000052	0.000025
	1990	0.005059	0.000958	0.000021	0.000010

Table 6.9c: Biases for bean area based on exact theory

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Igawilo	1989	0.000262	0.000180	0.000006	0.000003
	1990	-0.000132	-0.000132	-0.000004	-0.000002
Ngarash	1989	-0.002331	-0.002331	-0.000039	-0.000019
	1990	-0.002049	-0.002049	-0.000033	-0.000016

Table 6.10a: Variances for maize area based on exact theory

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Igawilo	1989	0.002295	0.002253	0.002220	0.002157
	1990	0.004391	0.004247	0.004187	0.004089
Ngarash	1989	0.057608	0.057608	0.057366	0.056411
	1990	0.062413	0.062413	0.062082	0.061087
Sinoni	1989	0.005289	0.005289	0.005226	0.005090
	1990	0.006273	0.006273	0.006215	0.006079

Table 6.10b: Variances for maize+bean area based on exact theory

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Nyankumbu	1989	0.040472	0.027222	0.027058	0.026946
	1990	0.027830	0.015169	0.015098	0.015006
Ngarash	1989	0.025761	0.019511	0.019489	0.019488
	1990	0.026564	0.020852	0.020814	0.020808
Sinoni	1989	0.007182	0.006115	0.006060	0.006001
	1990	0.008081	0.005510	0.005491	0.005485

Table 6.10c: Variances for bean area based on exact theory

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Igawilo	1989	0.000306	0.000296	0.000292	0.000287
	1990	0.000280	0.000280	0.000276	0.000267
Ngarash	1989	0.013543	0.013543	0.013440	0.013214
	1990	0.012677	0.012677	0.012584	0.012381

Table 6.11a: The MSEs for maize area

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Igawilo	1989	0.002296	0.002254	0.002220	0.002157
	1990	0.004392	0.004249	0.004187	0.004089
Ngarash	1989	0.057816	0.057816	0.057366	0.056411
	1990	0.062556	0.062556	0.062082	0.061087
Sinoni	1989	0.005290	0.005290	0.005226	0.005090
	1990	0.006279	0.006279	0.006215	0.006079

Table 6.11b: The MSEs for maize+bean area

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Nyankumbu	1989	0.0040632	0.027239	0.027058	0.026946
	1990	0.027978	0.015189	0.015098	0.015006
Ngarash	1989	0.025808	0.019511	0.019489	0.019488
	1990	0.026605	0.020853	0.020814	0.020808
Sinoni	1989	0.007194	0.006119	0.006060	0.006001
	1990	0.008106	0.005511	0.005491	0.005485

Table 6.11c: The MSEs for bean area

Village	Year	$\bar{y}_r$	$\bar{y}_{C1}$	$\bar{y}_{C2}$	$\bar{y}_{MOD}$
Igawilo	1989	0.000306	0.000296	0.000292	0.000287
	1990	0.000280	0.000280	0.000276	0.000267
Ngarash	1989	0.013549	0.013549	0.013440	0.013214
	1990	0.012681	0.012681	0.012584	0.012381

Table 6.12a: Efficiencies for maize area based on exact theory

Village	Year	E'_r	E'_1	E'_2	E'_3
Igawilo	1989	231	235	239	245
	1990	154	159	162	166
Ngarash	1989	196	196	198	201
	1990	176	176	177	180
Sinoni	1989	114	114	116	119
	1990	126	126	128	131

Table 6.12b: Efficiencies for maize+bean area based on exact theory

Village	Year	E'_r	E'_1	E'_2	E'_3
Nyankumbu	1989	73	109	110	110
	1990	68	125	126	126
Ngarash	1989	76	100	100	100
	1990	79	100	100	100
Sinoni	1989	106	124	125	127
	1990	69	102	102	102

Table 6.12c: Efficiencies for bean area based on exact theory

Village	Year	E'_r	E'_1	E'_2	E'_3
Igawilo	1989	100	103	105	107
	1990	114	114	116	120
Ngarash	1989	127	127	128	130
	1990	123	123	124	126

Source: Mussa Field Survey.

There is variation between the years for the same estimators as well as between the estimators for the same year. The variation between the years may be due to conditions imposed on the farmer which he (or she) may not be able to control. Depending on these conditions farmers may plant the whole field or part of it or not plant at all.

The availability of seed is a very important aspect in the determination of area planted. When the seeds are available, farmers use more land and unavailability restricts to a smaller portion. With the introduction of hybrid seed, there has been a massive use of different kinds of hybrid seeds which cost farmers considerably. If the farmer has money by the next farming season, he will certainly be able to buy the hybrid seed and continue planting. When the situation is different, that is by the approaching of the next planting season the farmer has no money to buy the hybrid seed, he will either re-use the seed, that is using his part of grain reserve as seed or borrow money to buy hybrid seed or combine the two alternatives. There are other alternatives, such as the farmer may have money but seeds could not be available in time because of poor distribution facilities. In such situations the farmer may decide to plant other kind of temporary crops whose seeds are easily available, like cabbage, tomatoes, potatoes *etc.* In all these circumstances a change of area on a particular crop is affected.

A farmer may fall sick and not be able to plant for that season. For a big responsible family this may not cause a big problem, or traditionally neighbours would organize a help scheme. Under any of the circumstances planting area is affected by the farmer not being able to plant himself or herself for that season. Also old age for those with no children around contribute to the variation, as area planting at a particular season will depend on the health condition.

Weather is the crucial factor conditioning the traditional farmer. Raining too early can disturb the planting pattern, which may change the area of the field/plot. Too much rain can cause erosion and loss of seeds. Rain after planting is also crucial to the development of a crop and thus to its production.

Land tenureship, though most farmers do not have titles, also contributes to the variation. Some farmers would let land to be tilled by others on cash terms. This normally happens on an agriculture yearly basis. This process causes variation of area for a particular crop and to a large extent in the production. It is very rare in such cases that land is kept fallow. Fallow land on the other hand is also a source of variation.

Sometimes variation is caused by new developments in a village. For example Table 6.8a, which is the maize variance estimators for Igawilo, shows greater variation than in other tables. This village has been absorbed in the town municipal council.

Farmers fear seizure of their land for town planning, some are subdividing land to their inheritors while others clear bushes for more land. Also some are changing agriculture planting from temporary crops to coffee, which is the permanent crop so that if the government decides to take the land, then it has to pay a big compensation for coffee plantations. Normally the government is reluctant to uproot coffee plants as it is one of the major foreign currency earners. A farmer therefore may retain his land by planting coffee.

As for variation between the estimates within a year in almost all the cases Chakrabarty and Modified estimators have least variances. Chakrabarty (1979) proposed two estimators with the same asymptotic variance and one having no bias. $\bar{y}_{C1}$ is always more biased than the modified estimator $\bar{y}_{MOD}$. As such $\bar{y}_{MOD}$ has a tendency to lower mean square error than $\bar{y}_{C1}$. Because these two estimators are equally simple to calculate, the study therefore opts to use $\bar{y}_{MOD}$ in the area estimation. On the other hand $\bar{y}_{C2}$ has no bias at least up to order n^{-2} compared to $\bar{y}_{MOD}$. This makes $\bar{y}_{MOD}$ competitive with $\bar{y}_{C2}$. In such a situation the choice of an estimator depends on its simplicity in the practical application. $\bar{y}_{C2}$ will pose a practical problem in developing countries when it comes for example to re-grouping the sample size into two parts, and involving calculations of r_1 and r_2 as explained in chapter five. Also to have better estimates of the variance between units within a stratum, more villages must be included in the survey which will force a smaller sample selection of number of households per village because of avoiding extra costs. If the sample size of household selected for survey is small relative to the total number of households, a case which is not uncommon in developing countries when it comes to the selection of agricultural sample size (Brooks 1955), then splitting the sample size further might give unreliable r_1 and r_2. It has been suggested that in the case of Agsasu it is probable that the size could be reduced to say about 6 to 15 without any serious loss in precision, depending on the relative size of s_b^2 and s_w^2 (Rutherford 1988). So for a small sample size of households relative to the village size $\bar{y}_{MOD}$ becomes almost as unbiased as $\bar{y}_{C2}$ to the order n^{-2}.

In the exact theory the biases and variances of $\bar{y}_{MOD}$ for all cases are smaller compared to $\bar{y}_{C1}$ and $\bar{y}_{C2}$ as can be observed in Tables 6.9a-c to 6.10a-c. The MSEs for $\bar{y}_{MOD}$ are relatively smaller than others which therefore effect greater efficiencies (Tables 6.12a-c). For these reasons and for the simplicity of attaining adequate sample size $\bar{y}_{MOD}$ is thus preferred for estimation of crop area.

6.5 Estimation of crop area

Estimation for crop areas for the three regions will be done in this section. These are the regions where I have collected the ground information and also performed the area interpretation from the satellite images. In a situation where there are several villages involved in a region, and under simple random sampling, the average of the mean crop area per household in a village would give a good estimate. From these estimates, there are different means of extrapolating to regional level. The sampling weights will depend on each circumstance, one may use number of villages in that region, number of farming households or in the case of Agsasu, number of people. In this study because of cost and time constraints I have area estimates from four villages in the three regions; that is, one village each in two regions and two villages in the remaining region. The cultivated areas of the villages as interpreted from the images are shown in Table 6.13.

Table 6.13: Cultivated areas of villages as interpreted from images

Village	Cultivated area X (ha) as interpreted from Remote Sensing
Nyankumbu	1102
Igawilo	938
Ngarash	1624
Sinoni	1012

The mean crop area per household for a village is calculated as

$$\bar{y}_{MOD} = (1 - W)\,\bar{y} + W\, t_{MOD}\,\bar{X}$$

where t_{MOD} is the Tin (1965) modified estimator outlined in chapter five (section 5.2), $\bar{X}$ is the mean cropped land of the household. Using general crop statistics given in Table 6.7a, b and c, respective crop area estimates for the villages are calculated as:-

$$\hat{Y} = N\,\bar{y}_{MOD}$$

where N is the total number of farming households in the village.

The standard errors of $\hat{Y}$ for different crop categories have been calculated from Tables 6.11a-b. They are in some cases large, because a particular crop in many households is little grown or not grown at all. It might even be because of the land distributional pattern on those comparatively large farms. Examples are, maize in Igawilo (1990) for the former case and maize and beans in Ngarash for the latter.

The estimates with their standard errors in brackets are shown in Table 6.14.

On the other hand, suppose areas of these villages were calculated without the aid of remote sensing data and using the arithmetic mean $\bar{y}$, then the estimates would be like those in Table 6.15, which are shown after Table 6.14.

Table 6.14: Area for the crops by village with the aid of remote sensing

Village	Year	Maize	Area (*ha*) Maize+bean	Bean
Nyankumbu	1989	na	540 (75.3)	na
	1990	na	587 (57.7)	na
Igawilo	1989	431 (80.1)	na	55 (29.2)
	1990	473 (113.8)	na	41 (29.1)
Ngarash	1989	575 (130.2)	471 (76.5)	175 (63.0)
	1990	551 (140.6)	495 (82.1)	156 (63.3)
Sinoni	1989	173 (40.2)	428 (43.7)	na
	1990	187 (45.7)	351 (43.4)	na

Table 6.15: Area for the crops by village based on land measurement only without the aid of remote sensing

Village	Year	Maize	Area (*ha*) Maize+bean	Bean
Nyankumbu	1989	na	647 (78.5)	na
	1990	na	725 (64.1)	na
Igawilo	1989	828 (128.0)	na	86 (31.2)
	1990	872 (148.1)	na	89 (35.4)
Ngarash	1989	778 (182.8)	482 (74.9)	258 (71.0)
	1990	802 (186.0)	518 (81.9)	245 (69.7)
Sinoni	1989	124 (44.1)	350 (48.1)	na
	1990	147 (52.1)	334 (43.7)	na

na = not available or not analysed because of insignificant number of households.

Most of the values in Table 6.15 with their standard errors are not only high but seem to be out of proportion with the total area of the villages. For example the average ratio of maize area for Igawilo village in the year 1990 as can be seen from Table 6.7a is $0.49/1.15$, and for the beans (Table 6.7c) is $0.05/1.15$. The average approximate ratio of the two is 0.47. This covers about $0.47 \times 938 = 441 ha$ of the total village.

From Table 6.15, the area of maize and that of beans amount to $961 ha$, which has exceeded the area of the entire village per season per annum. This is an over-estimation of about 120% compared to $\bar{y}_r$ and about 90% compared to $\bar{y}_{MOD}$. On the other hand Sinoni village gave lower estimates using $\bar{y}$ compared to $\bar{y}_r$ and $\bar{y}_{MOD}$ but with higher standard errors. The village, being on highland areas, practises intensive cultivation; the estimate achieved through $\bar{y}$ can be considered as low (an under-estimation of about 20% and 10% respectively), because rarely is land left uncultivated. The use of remote sensing data has therefore eliminated the discrepancy of land misrepresentation of between -10% and 90%, while integrating it with the modified mean estimator $\bar{y}_{MOD}$ managed to reduce variance.

There is however a problem of mixed cropping as observed in the case of maize+bean. In the development of area statistics for mixed cropping, one of the main concern is the apportionment of the areas among the various crops in the mixture. The following section tries to single out various possibilities and problems involved in the estimation of single crop area estimation from mixed cropping.

6.5.1 Crop proportion estimation in mixed cropping

Plots are observed in fields from sampled households in a village to estimate parameters for the different crops grown. Crops are to be identified and appropriate measures labelled by their area. Crops are grown together in many instances because of some advantages. This may be because of either exploiting resources better together than singly or ensuring food production throughout the year by different crops which grow well under different weather conditions. Crops may not be necessarily sown at the same time. Undercropping or relay cropping is a term used when another crop is introduced before the first is harvested. Sometimes a leguminous species will be grown with the main crop to improve yields. This is the case for crop category known as "maize and beans" which is grown as alley cropping. A mixture of say maize and paddy is often referred to as undersowing or mixed cropping. Pearce *et al.* (1988) emphasized that diversified cropping is not necessarily intercropping. With a small area of land and a need for several crops, some to supply food and some to supply cash, the farmer will grow them intermingled, filling in gaps as may seem appropriate. The authors stressed that unless some benefit can be claimed for growing species in close association, the result should not be claimed as intercropping. The complex cropping patterns, particularly in sub-Saharan Africa, present special problems not only in terms of terminology but for measurement as well. Lack of adequate spatial separation between the crops and inherent cultural growth patterns special to the location, special to certain crops, among other factors, may cause misproportioning of some fields, thus resulting in an inaccurate estimate of the discriminant parameter. It is not to be expected that the same statistical approach will suit all problems (Pearce *et al.* 1988).

Poate and Casley (1985) gave two basic types of crop mixture.

1. One crop is occupying space within the plot that would otherwise be occupied by another, so that each crop is grown at a lower density than would occur if they were grown separately.

2. One crop is added between the rows of another crop which has been planted at its normal density (sometimes referred to as interplanted cropping).

In either case presentation of crop area as a simple sum of all land on which the crop appears, irrespective of mixture, will be misleading unless it is supported by other information. There have been several different suggested alternatives in an attempt to overcome this problem. Among them in the case of two crops category is the bivariate method, where the variates are analysed together giving equal weight to each. Pearce *et al.* (1988) noted that bivariate analysis is not available unless it can be assumed that the correlation coefficient between the two yields is the same for all treatments. Where spacing is involved and in some other cases also, the authors cautioned, that could be a risky assumption. Standardizing crop areas to a common base, and preparing specially constructed tables seem to be other alternatives.

While the construction of tables can work in a controlled environmental or experimental situation in project monitoring and evaluation, standardization seemed to be more appropriate in a survey situation.

Different countries follow different ways of presenting crop area according to the following procedures.

(i) The area is divided equally to all the crops grown.

(ii) The area is divided proportionally to each crop.

(iii) Whole area is reported to each crop.

(iv) Whole area is reported to the main crop only.

(v) The area is alloted by seed rate, plant density, *etc.*

Poate and Casley (1985) argued that these procedures do not only involve tortuous calculations but also do have a weakness in choosing a standard to act as the denominator. They argued further that sole crop densities, yields and seed rates vary considerably from year to year and from one region to another. They, however, recommended an approach of presenting the areas of crops in at least two levels of detail:

1. The overall land area on which the principal crops are grown together with crop yields (*ie*. (iv) above).
2. For each crop, a breakdown of the area into certain basic types. For example, pure maize, maize with beans, maize with other and so on.

Estimation of crop area in sub-Saharan Africa has always been a problem because of the rotational farming, availability of seeds, rainfall, pests, diseases and weather calamities. The choice of procedure depends on the focus and objective of the survey. The presentation of crop area by certain basic types, that is area of pure stand and area of mixed crop recorded with separate estimate of yields (see Table 7.13 and Takwimu 1988, p 5), is important to agricultural economists and/or farm management to not only know what crops are grown, but also how they are grown. However most organizations interested in agriculture, including FAO, send out requests for return as if cropping were always in pure stands. Moreover a breakdown of area into certain basic types, as recommended by Poate and Casley (1985), may make some crops in mixed farms of more than three crops at the same location insignificant and one might have to create a category "miscellaneous" which would include everything else recorded. This section will attempt to present crop areas as pure stands and the problems involved.

Table 6.16: Percentage of plots containing mixed cropping

Country	Mixed crop (%)
Benin	39
CAR	74
Kenya	47

Source: Longacre, Verma *et al.* (1988).

Verma *et al.* (1988) in their comparative study of obtaining production estimates from five sub-Saharan countries in Africa, from samples of between 100 and 120 in each country, came across approximately two thirds of the plots containing other crops beside the main crop. Table 6.16 shows the percentage of plots by countries which were containing mixed cropping.

To determine the proportion of a particular crop area in the fields or plots in a household demands rather complex perception on the part of the enumerator. What one can assess on the ground is likely to be either the number of plants of the respective crops per unit area, or that part of the ground surface covered by the plants of respective crops. The exercise is highly subjective. However, with the help of the farmer's experience on seed input, number of paces made while planting, say in a row, number of days needed to plough (whether human or oxen driven) and together with a fairly regular and systematic intercropping procedure, it is likely to produce an acceptable proportion. Otherwise it is quite evident, as Idaikkadar (1979) put it, that judgement of areas occupied by each crop is beyond the capacity of an average field officer.

Petricevic (1988) conducted a pilot survey in area measurements in Botswana on some of the mentioned variables. The results do not seem to be reliable when only one variable is considered as can be seen in sketches of Figure 6.1.

It is therefore anticipated that by combining enumerator's field experience and farmer's practical experience an area proportion for the crop under study in the fields for each household can be acceptably calculated, such that we can proceed to estimate a representative proportion for the entire region.

A simple and reliable form to specify a proportional procedure in the estimation of crop area in mixed cropping farms will be envisaged in this section. Data from Agricultural Sample Survey (Agsasu) 1987/88 by the Bureau of Statistics, Dar es salaam, Tanzania are used for the calculation of proportions of maize from a combination of maize and beans. The intensive cultivation in some fields has rendered it very difficult to eliminate other crops from this combination apart from major crops like sorghum, millet and paddy. Table 6.17 presents area of maize with combinations of different other crops.

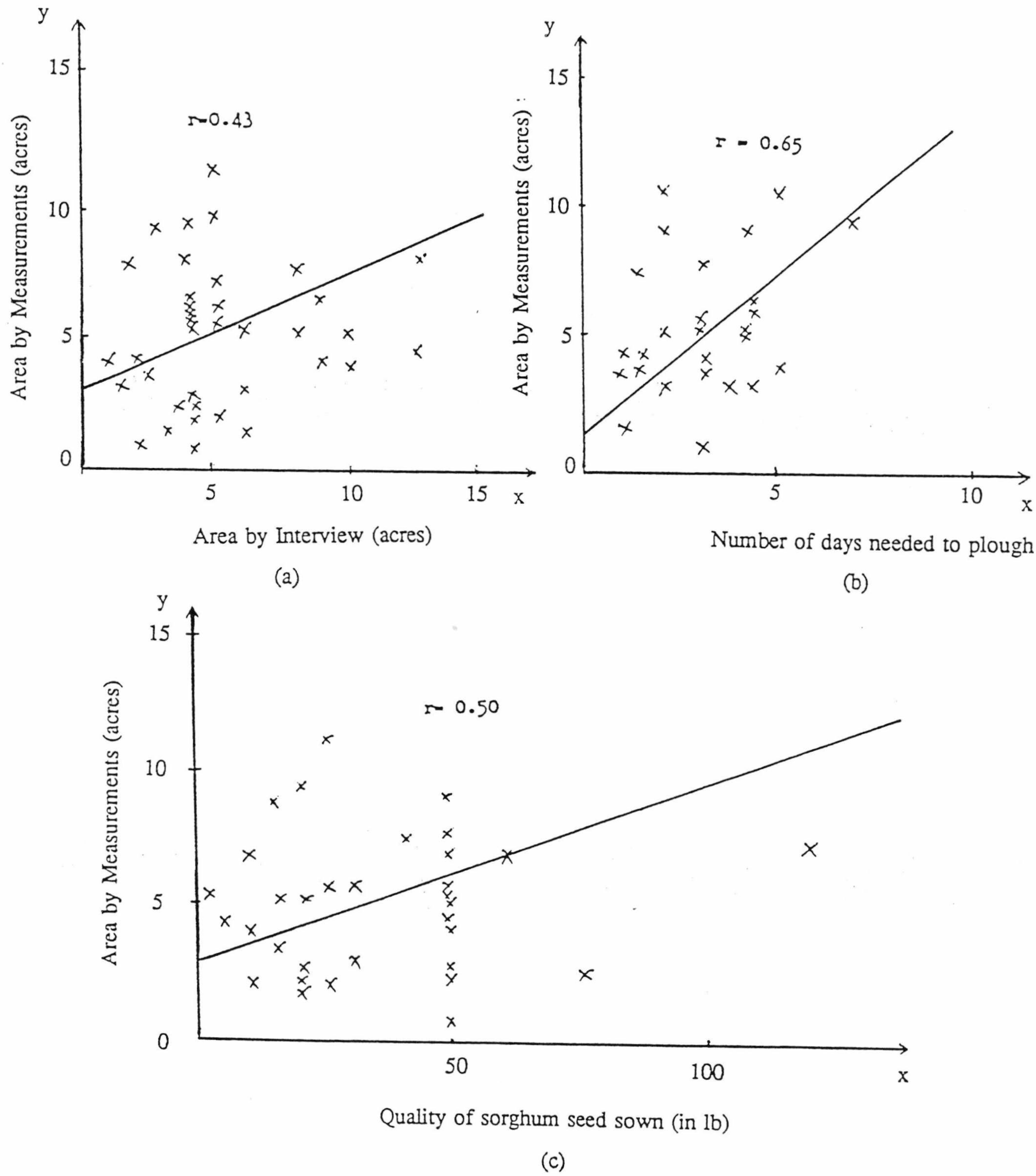


Figure 6.1: Correlation sketches between the area obtained by measurements and by using some other variables

Table 6.17: Planted area for maize combination (mixtures) by season 1986/87 ('000 *ha*)

Crop combination with maize	Season	
	Short rains	Long rains
Pure Stand	35	421
Bean	124	530
Groundnut	17	251
Cassava	90	124
Sorghum	12	146
B. Millet	0	101
P. Peas	4	71
Other	1	49
Paddy	0	45
F. Millet	0	44
Cowpea	15	29
Greengarm	0	24
S. Potatoes	2	12
Sunflower	2	5

Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

6.5.2 Estimation of proportions

A simple random sample of n households is considered in order to estimate the population proportion P of area in the maize and bean combination that is attributed to maize. Let x_i be the total area for a number of fields in a household that are occupied by maize and bean combination. Also let p_i be the proportion of maize in a household. The proportion can be expressed in the form

$$p_i = \frac{y_i}{x_i}$$

where y_i is the area covered by maize ($y_i \leq x_i$), and so the area of maize can be explicitly represented as

$$y_i = p_i x_i \tag{6.6}$$

The relation between y_i and x_i is linear (see Figure 6.2). This would suggest using an estimate based on linear regression of y on x. However, on testing the null hypothesis for zero intercept all but one village gave data consistent with this null hypothesis as shown in the table below.

Table 6.18: Intercept coefficients estimates by villages

Village	Sample size	Intercept	Standard deviation	t-calc. ratio	t-table ratio
Himiti	8	-0.01278	0.01677	-0.76	2.447
G-Lumbo	9	-0.01792	0.09448	-0.19	2.365
Ngoha*	14	0.00000	0.00000	-	-
Matwiga	14	0.07660	0.06173	1.24	2.179
Kagunga	14	-0.19110	0.16320	-1.17	2.179
Irunda**	10	0.13970	0.04358	3.21	2.306
Kanyelele	14	0.08405	0.07987	1.05	2.179

* There was no mixed cropping sampled in Ngoha

** The null hypothesis of zero intercept is rejected.

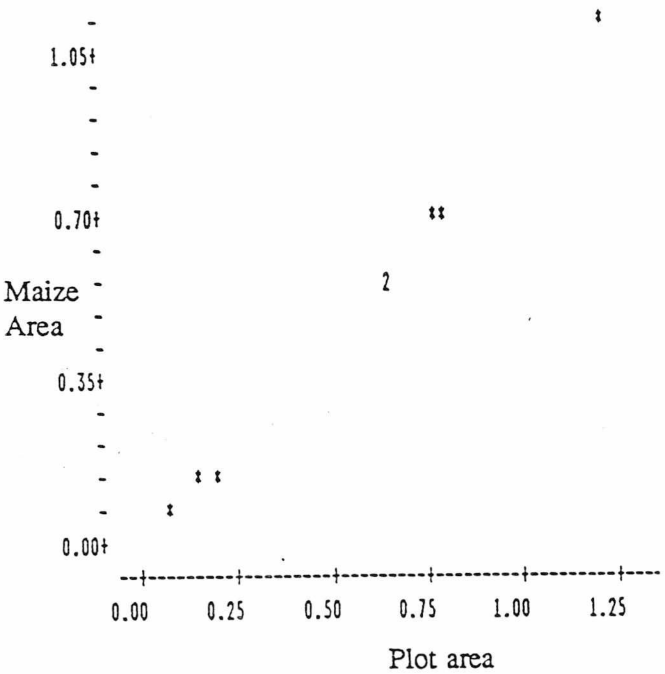
Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

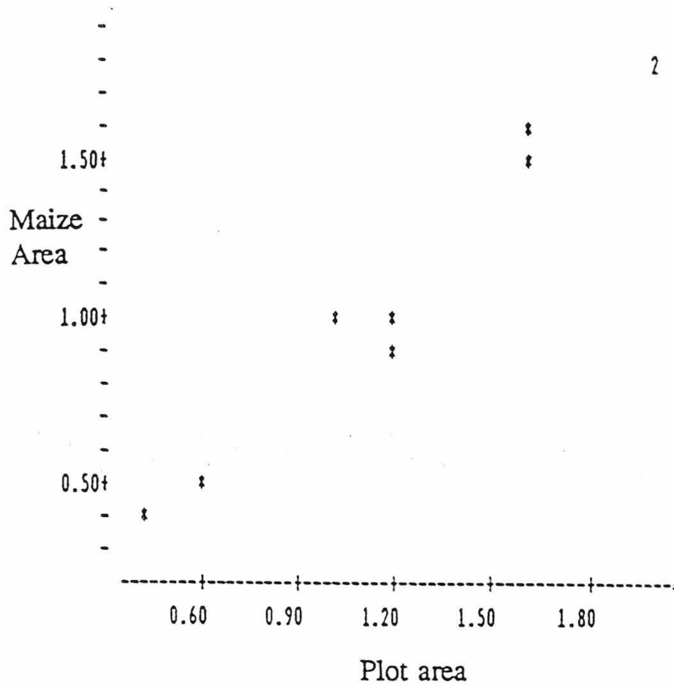
The relation is also presented on the basis of data pooled from the three regions. Pooling of the data across regions has the obvious advantage of increasing the available sample size which helps one to discern the overall pattern of results more clearly. The combined regression also gave data consistent with the null hypothesis, with the intercept coefficient of -0.02298 and its t-calculated ratio as -0.59. The linear relation for the pooled data is shown in Figure 6.3. Moreover the regression residuals indicate the normality of y_i's as shown in Figure 6.4. In such situations a ratio estimate is better than a regression estimate.

The proportion of area for maize in the entire sample can be estimated as

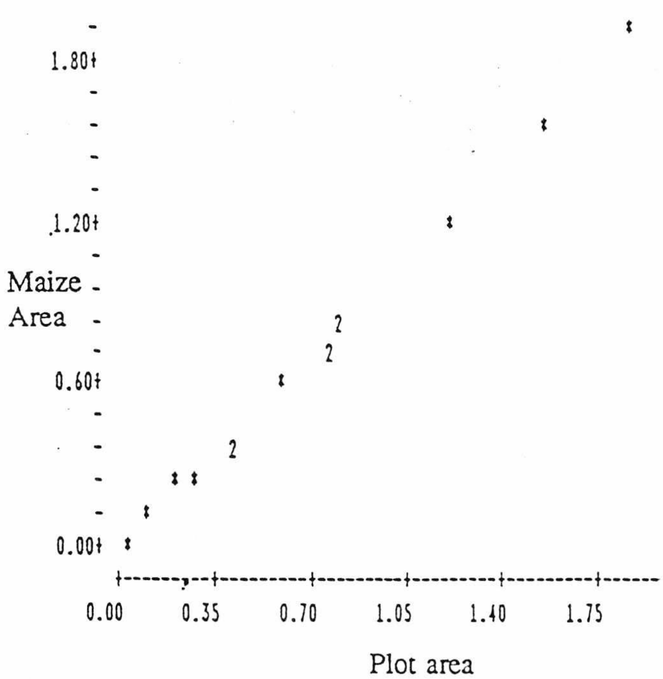
$$\hat{p} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$



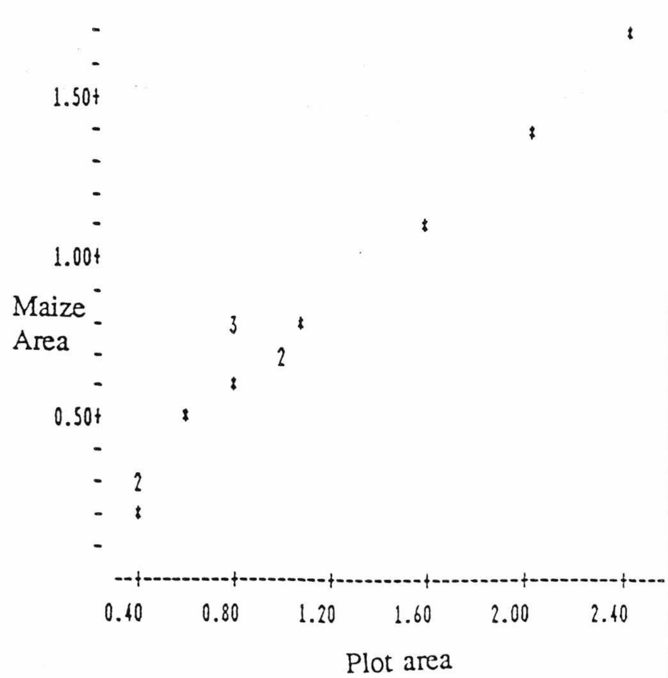
(a) Himiti;Arusha



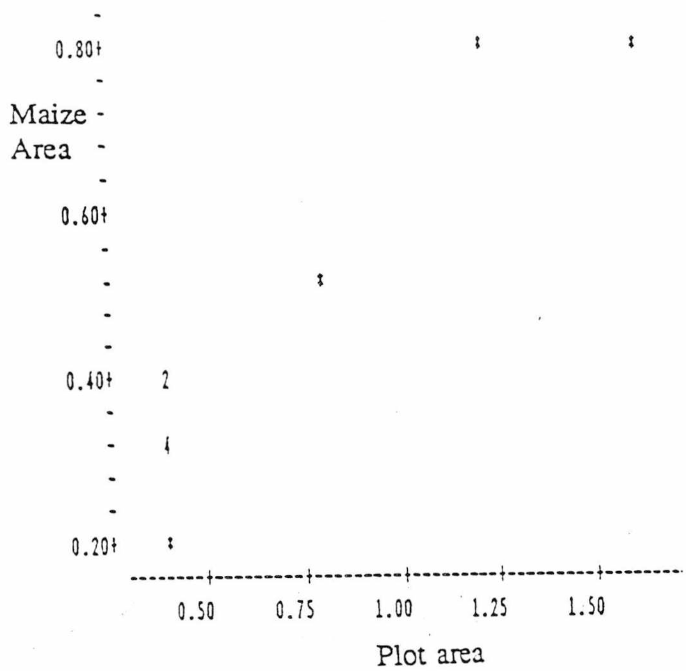
(b) Gekrum-Lambo;Arusha



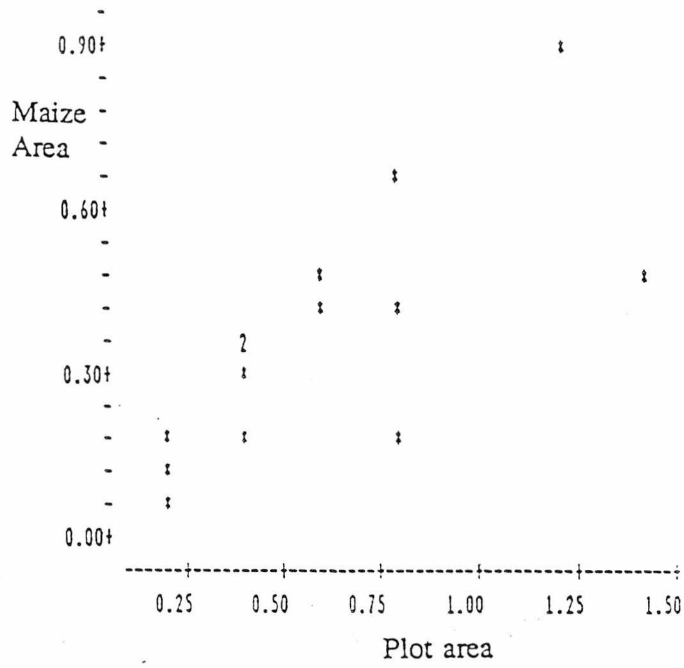
(c) Ngoha;Mbeya



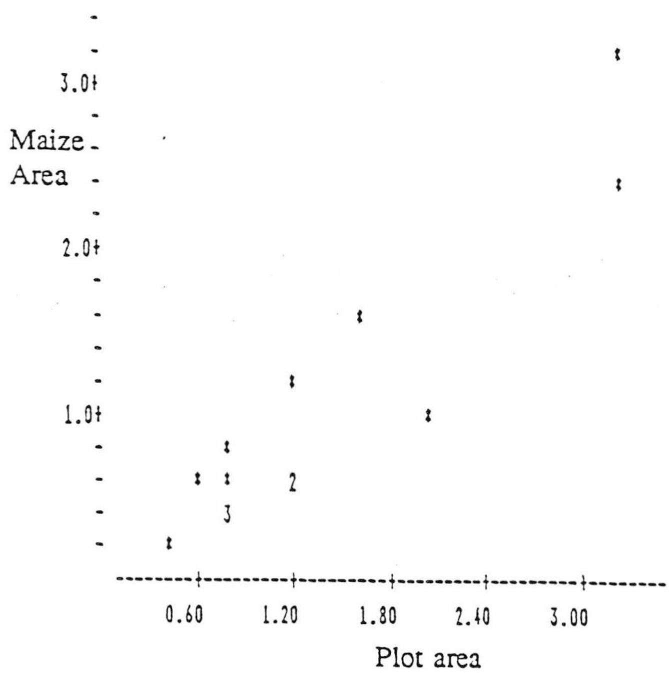
(d) Matwiga;Mbeya



(e) Irunda;Mwanza



(f) Kanyelele;Mwanza



(g) Kagunga;Mwanza

Figure 6.2: The regression of maize area on plot area for different villages;regions in Tanzania
(MINITAB plots)

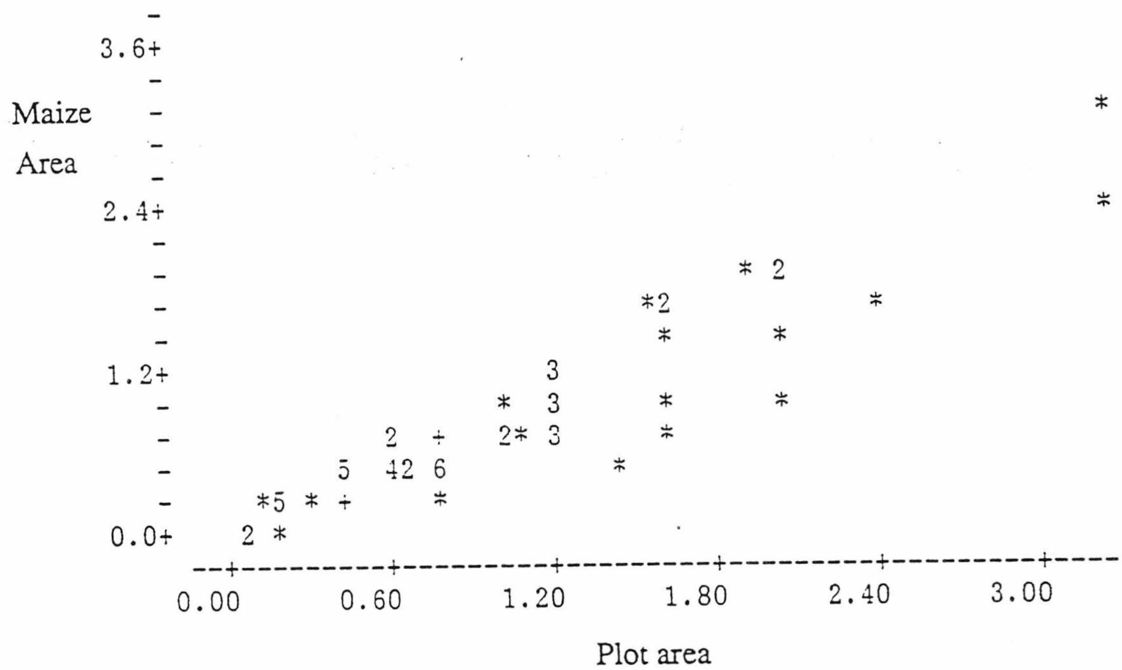


Figure 6.3: The regression of maize area on plot area for the pooled data (MINITAB plots)

```
Histogram          N = 83

Midpoint  Count
   -3.0       2  **
   -2.5       1  *
   -2.0       1  *
   -1.5       3  ***
   -1.0       8  ********
   -0.5      13  *************
    0.0      21  *********************
    0.5      15  ***************
    1.0      14  **************
    1.5       3  ***
    2.0       1  *
    2.5       0
    3.0       0
    3.5       1  *
```

```
Residuals -
          -                                                      *
          -
       2.5+
          -                                                   *
          -                                              ** *
          -                                     22 6 2 2
          -                               *335*2
       0.0+                         3225333
          -                    **433*
          -                *4 2*
          -              2*
          -            *
      -2.5+          *
          -     *  *
          -
          -
            --------+---------+---------+---------+---------+--------Nscores
                 -2.0      -1.0       0.0       1.0       2.0
```

Figure 6.4: A Histogram and a Normal probability plot of residuals from pooled data (MINITAB plots)

which is a typical ratio estimator. From equation (6.6), P can be estimated as

$$\hat{p} = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} x_i}.$$

The appropriate variance of $\hat{p}$ (see Cochran 1977, p66) is given by

$$V(\hat{p}) = \frac{1-f}{n} \sum_{i=1}^{N} \left[\frac{x_i}{\bar{X}}\right]^2 \frac{(p_i - P)^2}{N-1}.$$

This form shows that the approximate variance involves a weighted sum of squares of deviations of the p_i's from the population value P. For estimated variance we have

$$v(\hat{p}) = \frac{1-f}{n\bar{x}^2} \frac{\sum_{i=1}^{n} (p_i x_i)^2 - 2\hat{p} \sum_{i=1}^{n} p_i x_i^2 + \hat{p}^2 \sum_{i=1}^{n} x_i^2}{n-1} \tag{6.7}$$

where $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$ is the average area per household in the sample.

The estimated proportions of maize for the seven villages in three regions, their lower and upper bounds in the 95% confidence interval and their corresponding estimates of variance are presented in the following table.

Table 6.19: Maize proportion estimates by villages

Region	Village	p	Lower bound	Upper bound	$v(p)$
Arusha	Himiti	0.889881	0.875138	0.904624	0.000327
	G-Lumbo	0.891481	0.870736	0.912226	0.000757
Mbeya	Ngoha	1.000000	1.000000	1.000000	0.000000
	Matwiga	0.750149	0.730490	0.769808	0.001176
Mwanza	Kagunga	0.734657	0.702795	0.766519	0.003089
	Irunda	0.628645	0.581063	0.676227	0.004561
	Kanyelele	0.577163	0.524117	0.630209	0.008562

Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

There was no mixed cropping sampled in Ngoha. The practice is not common, hence the proportion appeared as 1.00 indicating pure maize. In fact the region is the major maize producer with very little livestock farming.

In Arusha region the two villages seem to agree on a proportion while there are differences for villages in Mwanza. As explained in chapter four (section 4.8), Mwanza region is the chief grower of cotton, which is one of the major cash crops in Tanzania and is also among the leading regions in livestock farming. In addition to that, as mentioned earlier, the region is actively engaged in gold mining. Thus, with all those potential resources, food crop becomes only a general farming activity for mere subsistence. The major part of the land is used either for cotton farming or for livestock grazing. The allocation is not systematic but depends on the number of heads of cattle and the capacity for cotton farming. An interesting observation is that although the population density for this region is high (see Table 6.20), households acquire relatively bigger lands which aggravate disproportionate farming systems. This is because the land is comprised of patchy grassland and scattered cultivation with tress and bushland. Some rocky structures make the land less homogeneous. It is unlike Arusha region, as pointed out earlier, that it is a major coffee grower and also engages actively in livestock farming, but because the greater portion of land is dry grassland and a national reserve for wildlife there is a high concentration of people in villages despite the low overall population density. Land is intensively used, cultivation extends up to the slopes of mountains and grazing is either zero or on specially combined allocated areas. Therefore a proportionate farming system is necessary.

Table 6.20: Population density for the regions

Region	Population	Density (*per sq. km.*)
Arusha	1351675	17
Mbeya	1476199	25
Mwanza	1878271	96

Source: 1988 Population Census

Bureau of Statistics, Dar-es-salaam, Tanzania.

6.5.3 Weighted averages

It is certain that the composition of a mixed crop field will vary from time to time. Also not all the crops in the mixture may be planted at the same time nor harvested at the same time. The situation therefore needs a carefully balanced but simply formulated proportion of area on yearly basis.

Despite the above concerns, proportion estimates for villages in each region seem to show a similar pattern from one to another. This to some extent also applies to their variances (Table 6.19). It is therefore quite appropriate to pool the region results for a combined representative proportion estimate.

The proposed proportion estimate for a region is to be obtained by the weighted average method. The land use structure for the regions as explained earlier favours number of households in a village as weights. This is because households make overall use of the land in a village, and that number of households in a village is a readily available statistic in each region.

Given N_j, number of households in sampled village j, then the weight $w_j = \dfrac{N_j}{\sum_{j=1}^{n_k} N_j}$

where n_k is the number of sampled villages in the region k. The representative proportion for the region, $\hat{p}_k$, is given by

$$\hat{p}_k = \sum_{j=1}^{n_k} w_j \hat{p}_j$$

where $\hat{p}_j$ is the estimated proportion in the j^{th} village. The estimated variance of $\hat{p}_k$ is given by

$$v(\hat{p}_k) = \sum_{j=1}^{n_k} w_j^2 v(\hat{p}_j)$$

where $v(\hat{p}_j)'$s as calculated in equation (6.7) are shown in Table 6.19 for each village. The estimated representative proportion for each region is shown in Table 6.21 below with its corresponding variance.

Table 6.21: Estimated proportions and their variances for the regions

Region	$\hat{p}_k$	$v(\hat{p}_k)$
Arusha	0.890733	0.000286
Mbeya	0.858887	0.000375
Mwanza	0.639060	0.002014

Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

Area for pure stand crops

The values of $\hat{p}_k$ for respective regions in Table 6.21 can be used to calculate areas for maize only from maize+bean areas given in Table 6.14. The area is calculated as

$$Area(Maize) = \hat{p}_k \times Area(Maize + bean)$$

with respective variance as

$$V[Area(Maize)] = \hat{p}_k^2 \times V[Area(Maize + bean)].$$

The total maize area is achieved by summing the above calculated area with the area under maize only given in Table 6.14 with the sum of their respective variances to get the overall variance of the pure stand. Bean area and its variance are calculated the same way, either by applying $(1-\hat{p}_k)$ on the above equations, or subtracting the result of the above equation from the Maize+bean area. Therefore, Table 6.14 with maize, maize+bean and bean areas with their respective standard errors can be re-written as Table 6.22 with maize and bean areas only with their respective standard errors in brackets.

Table 6.22: Area for the pure stand crops by village

Village	Year	Area (*ha*) Maize	Bean
Nyankumbu	1989	345	195
		(48.2)	(27.2)
	1990	375	212
		(36.9)	(20.8)
Igawilo	1989	431	55
		(80.1)	(29.2)
	1990	473	40
		(113.8)	(29.1)
Ngarash	1989	995	226
		(146.9)	(63.5)
	1990	992	210
		(158.5)	(63.9)
Sinoni	1989	554	47
		(56.0)	(4.8)
	1990	500	38
		(59.8)	(4.7)

Production is estimated as the product of area and yield rate. In the next chapter an attempt is made to arrive at an appropriate way of attaining yield rate in order to combine with areas calculated in this chapter to estimate production.

Chapter seven

Production: crop cutting versus farmer's statements

7.1 Introduction

Determination of crop yields coupled with data on areas at about the time of harvest is primarily for the estimation of crop production. Production is of vital importance in economic planning. Data on yields are also used for other purposes. Most attempts at improving agricultural production are aimed at yields. At a glance they show the differences in the yield rates between different households/fields as provided in the survey structure, but at greater length they represent a basis for the preparation and formulation of many economic measures.

When data on a yield survey are to be collected, there is essentially a choice between two main data collection methods:-

1. Crop-Cut Method - mainly by the use of crop cutting subplots of known sizes and shapes accompanied by measurement of area of the plot;
2. Farmer's Interviewing Method - by farmer's statements given in local units and subsequently translating them into standard units.

Other methods include: eye estimate method, based on visual inspection of the crop by knowledgable surveyor or external agent; crop reporting committees; extension officers estimation; reporting service based on agents; harvesting the whole field. Olsson (1990) argued that harvesting of whole fields is too cumbersome for practical survey work.

The traditional way of collecting yield statistics, according to Zarkovich (1966), was through some kind of reporting service in which farmers themselves were involved. In many developing countries, however, the author observed that such an approach was difficult to operate for many reasons. He gave these examples:-

1. Extension services were not effectively established over the country as a whole;
2. The extension workers were not experienced enough;
3. Farmers would not be able to evaluate the yields of various crops in units meaningful to the data collectors; etc.

For all these and similar reasons experienced at that particular time, it has been taken for granted that usable information on production can be obtained only on the basis of methods involving crop cutting. The method is said to give the necessary protection on incorrect reporting if there is reason to suspect the farmer's response (suspicion about the purpose of statistics, income taxes, obligatory deliveries in kind, *etc.* (Zarkovich 1966)). In addition it is also said to protect against the consequences of an apparently unavoidable tendency on the part of reporters either to be conservative in their judgement or to stick to some kind of pre-existing figures irrespective of their quality.

Under the above circumstances it was pointed out that data obtained from eye estimates are often considerably lower than the corresponding figures resulting from the use of the crop-cut method. As an illustration Zarkovich (1966, pp 332-34) extracted and displayed some tables from surveys performed in England and Sweden to show the differences between the crop-cut method, which he termed an "objective method," and others, which were known as "subjective methods". Scott *et al.* (1989) discarded this misleading dichotomy. They favoured distinguishing methods on the basis of their method of data collection, each with its own potential and limitation on particular circumstances. Indeed, they concluded that all methods suffer from measurement errors and each method has its own strengths and weaknesses in terms of accuracy, cost and feasibility depending on the objectives and conditions of data collection.

Despite some strong arguments in favour of the crop-cut method Zarkovich (1966) pointed out that the cost aspect was a big disadvantage. A major problem with any approach involving physical measurement has been its high cost and time required. Often enumerators have to be stationed in sampled areas for long duration, which in turn necessitates the use of a heavily clustered sample design and small size (Casley and Lury 1986). High cost has also meant that, in many developing countries, it has not been possible to establish and sustain broadly based systems for regular collection of agricultural statistics (Verma *et al.* 1988, Scott *et al.* 1989).

Given these problems, there has been cause of concern shared both by the governments in African countries and by such international development and funding agencies for the need to have reliable and prompt estimates. It is desirable therefore to look into ways and possibilities of using less costly and relatively quicker methods.

This chapter will deal with the question of the crop-cut method and the various measurement errors that creep into the results and also explore the method of obtaining estimates of production directly by interviewing farmers.

7.2 Crop-cut method

A sample consists of several subplots which are harvested and the produce weighed. They are of prescribed dimensions located and marked according to some defined procedures. From the yield rates of these subplots, larger units like household's/village's yield rates could be determined with their standard errors. The extrapolation can be up to national level.

Crop-cut method is in widespread use as a technique. It was originally used by research workers to compare yields on an objective basis for selected treatments applied to accurately measured plots. Randomisation was used to justify or validate statements made about the "effects" that treatments had by the response of crops to them. Rutherford (1989, personal communication) pointed out that most research workers achieved much higher yields than farmers, because, if for no other reason, research workers worked on *net* areas (excluding crop failures they could explain), while farmers reported on grown areas (including crop failures within the gross area). However, up to about five years ago, relatively little written guidance was available for field survey units, as observed by Poate and Casley (1985). But its use raised practical discrepancies long before, when Zarkovich (1966) was concerned about the various creeping biases emanating from selection of fields, border of subplots, location, shape and size of subplot. Recent surveys conducted in Nigeria, Zimbabwe and Bangladesh have indicated that the method tends to suffer from a systematic bias of over-estimation as well as from large variance due to heterogeneity of the the crop within plots and holding (Casley and Kumar 1988, Scott *et al.* 1989).

The report containing the results of studies in five African countries in 1987, known as Longacre report, (Verma *et al.* 1988), has prompted a critical review of the method and made a comparison with other methods. The study in each of those countries was carried out using a common design with the objective of comparing crop estimates based on the crop-cut method with estimates obtained by asking farmers directly to state their production. Both types of estimates were compared against a complete harvest and weighing of the sampled plots.

The experiment was carried out in Benin, Central African Republic (CAR), Kenya, Niger and Zimbabwe. This is a cross section through sub-Saharan Africa. The study was restricted to maize only (main crop) in all countries except Niger, where millet (which is the main crop) was studied instead.

In each country, two or three regions were selected purposively and, within these, a two-stage sample of districts and plots was selected using random procedure. Each country had a total sample of about 100 or 120 plots. In all selected plots, the enumerators recorded length and bearing of the boundaries, and finally used a pre-programmed hand-held calculator to obtain area. The closing error of the plot measurements as a percentage of the perimeter was not to exceed 2%. Two crop cutting squares of $5m$ x $5m$ were placed by a random method within the area just measured. At the time of harvest the enumerators first harvested the two subplots. Farmers' estimates were given in local units and were obtained twice before, and once after, harvest.

The main results were that the crop-cut method gave over-estimates of 14% to 38% in each country. Farmer's statements produced relatively insignificant errors of −8% to +7% in each country. It was concluded that farmer's estimates were indeed more reliable than the crop-cut method.

On the other hand Olsson (1990) argued that the method of interviewing farmers has so many pitfalls that it would be unwise to abandon the crop-cut method based on Longacre report. He stressed the condition on the farmer's willingness to report the correct information. He also cautioned the need to examine carefully the issue of the standard measurements and questionnaire design to accommodate traditional units. The author is sceptical on the question of static enumerators, and doubtful of their understanding of the standard units of measurements.

This study intends to review the discussion mentioned above in the following two sections through the various sources of error in the crop-cut and farmer's estimates. Agsasu data for 1987/88 are also used to compare results from the crop-cut method with that from interviewing farmers. Finally in section 7.6 and 7.7, yield rate and production will be respectively estimated from the Mussa Field Survey's data.

7.3 Sources of errors: (1) Crop-cut method

In this section the procedure used to select a subplot, whether a circle or square as the case may be, will be outlined. Various sources of errors will be discussed in the subsequent sections and conclusion given in section 7.3.7 regarding the use of crop-cut method.

7.3.1 Procedure for subplot location

One method for randomly placing the subplot in the plot is based on coordinates. The maximum length X and the maximum width Y of the plot are determined as sketched in Figure 7.1. Random numbers x and y between 0 and X, and between 0 and

Y are selected respectively using, for example, random number tables. The point (x,y) on the plot forms the basis for allocating the subplot. It can be the starting point of a diagonal for a square subplot or a centre point for a circular subplot.

Another method also using two random numbers (x,y) is based on walking along the perimeter of the plot and then turning inside the plot for allocating point (x,y). The two random numbers are often restricted between one and half the length of the perimeter respectively. From any starting point, at the boundary of the plot one measures a distance x along the perimeter, and then turns inwards, roughly perpendicular to the boundary line, and measures a distance y into the plot. The diagonal of the square subplot will fall on the perpendicular line through the point y. If the subplot is circular, then the centre of the circle is taken at point y.

With minor variations these are the standard procedures mostly used in surveys. However they have been recognized as imperfect and always lead to bias. The following is a review of various sources of errors from the crop-cut method as experienced through various surveys conducted in several African countries.

7.3.2 Allocation bias

One of the important sources of bias according to Olsson (1990) is the tendency of the enumerators to "improve" the random allocation of subplots. African crops are often patchy in their pattern of planting. As pointed out in section 4.5 the within field/plot growing pattern is often very irregular, where crops grow correctly only where there is good soil. Enumerators may try to apply their judgement and prevent point (x,y) from falling into infertile areas or other sterile ground. This will therefore lead to upward bias.

7.3.3 Border bias

Yield along the border is often different from that of the remaining parts of the plot. If (x,y) is so close to the border of the plot that part of the crop cutting subplot would fall outside the plot, different actions could be taken. The first possibility is to disregard

(x,y) and select a new point. The consequence of such rejection is obvious: the points close to the border will have less chance of being included in the sample. This procedure will lead to under-representation of near border areas. Secondly it is possible to pull back the subplot according to some rule. For a square subplot, it normally happens that its diagonal is nearly perpendicular to the boundary line of the plot because of the way the subplot is laid, it happens that some part of the square (often corner edge) lies outside the plot. As the square subplot is pulled inside the plot along the perpendicular line through the point y, it therefore occupies smaller portion of the area along the border of the plot than before and creates under-representation close to the border. Scott *et al.* (1989) observed that this will lead to a lower sampling probability for a border band just inside the perimeter. For a circular subplot, Olsson (1990) commented that it leads to an over-representation of borderline subplots in the sample. A common approach, the author suggested, is to move the circle inside if the centre falls inside the plot and to select a new point if the centre falls outside. He however suggested that a theoretically preferable approach is to include all subplots irrespective of their coverage on the borderline, and to base the estimate of yield per hectare on the part of the area which is inside the plot. This approach is inapplicable in African surveys for a simple reason that it is difficult to calculate that portion of area of the circle inside the plot. These cases are illustrated in Figure 7.2.

7.3.4 Bias on subplot size

The size of subplot will affect the precision of the estimates. There is some evidence that smaller subplots give, on average, a higher estimate of mean yield per hectare (Olsson 1990). In their study of measurement problems in the 1986/87 agricultural sample survey of Tanzania, Anderson and Holmberg (1988) made some measurements on the crop-cut method with circular subplots of $3m^2$ and $9m^2$ demarcated from a square subplot of $25m^2$ of maize plots. In their comparison of yield per hectare, from the various sizes of subplots used, they indicated that there was a significant statistical

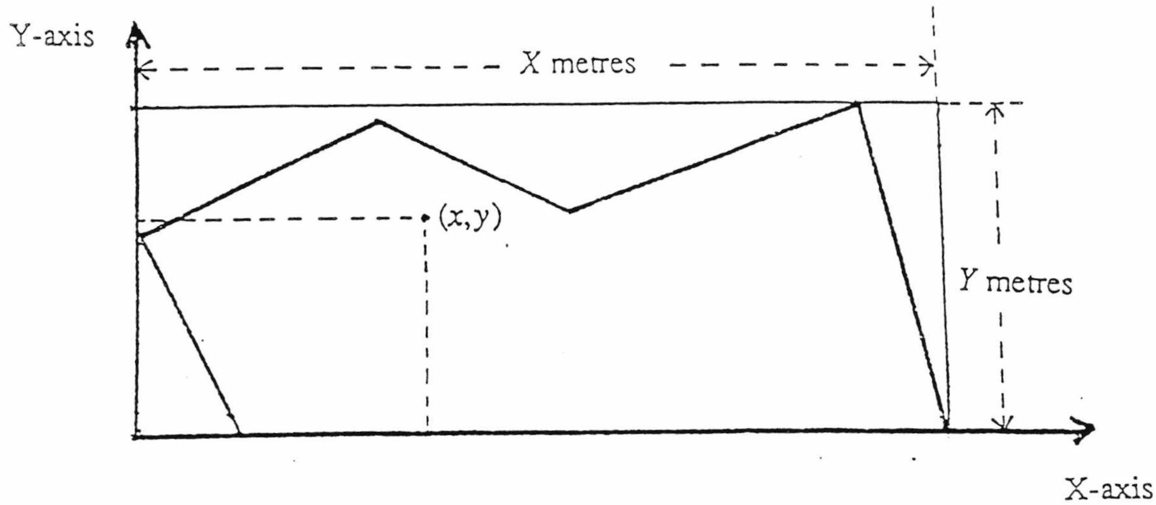


Figure 7.1: Procedure for subplot location

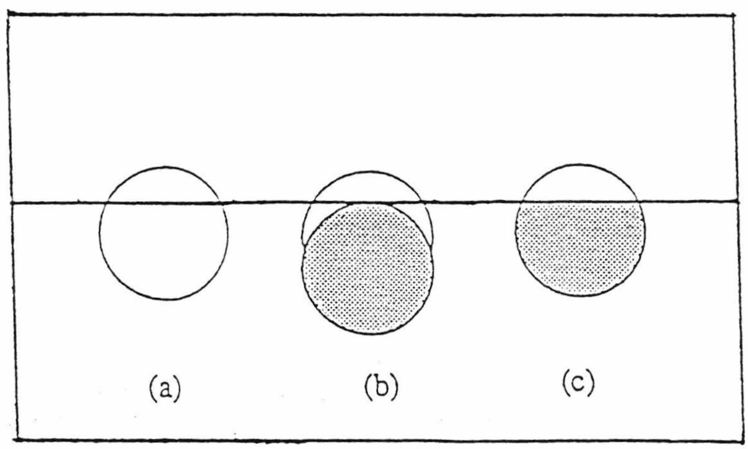


Figure 7.2: Border case strategies

Border case in: (a) disregarded

(b) moved into the field

(c) included with an area proportional to the area inside the field

(adapted from Olsson 1990)

difference between the estimates from a $3m^2$ subplot and $25m^2$ subplot. The authors acknowledged that a $3m^2$ subplot is too small, and even $9m^2$ subplot is not that big, because the fields/plots are so uneven. In the extent of mixed cropping the size is adequate around $100m^2$ (FAO 1982).

Over-estimation with small subplots had been reported a long time ago (see Sukhatme 1947a, 1947b, Mahalanobis and Sengupta 1951, Panse 1963 and Sengupta 1964) from the Indian experience. The basis of their over-estimate of small plots was to assume that their largest plot was unbiased. No whole field estimates were used, where gross area rather than crop area was used to estimate the yield per unit area. The seriousness of practical over-estimation is now confirmed in Africa.

7.3.5 Bias on subplot shape

Certain standard shapes of plots have been used for yield surveys based on crop cutting. The most common ones are circular, rectangular and triangular. Mahalanobis and Sengupta (1951) studied the yield problem relating to the shape of subplots. Triangular shape produces the most biased results followed by square shape in comparison with circular subplot.

The problem of shape of plots must be considered not only from the error point of view but also from that of practical convenience. Andersson and Holmberg (1988) reported that in most cases a square subplot of $25m^2$ varied between 15.9 and 26.5 m^2. Also they noted a problem in getting the angles perpendicular, so they concluded that circular subplots are easier to use and that work takes less time. However, because of the design structure of the circular instrument used, the authors could not exceed an area of $15m^2$ (radius $\cong 2.19m$) and therefore tentatively suggested a radius of between $2m$ and $2.5m$.

An ideal subplot is one which has minimum perimeter for a given area. A circle seems to be appropriate. However, larger radii proved to be inconvenient to manufacture as well as to carry. The situation is even more serious in the case of mixed cropping which is dominant in African traditional farming.

7.3.6 Bias on number of subplots

Where the crop under study is grown in evenly planted, dense stands and under carefully controlled conditions, the level of within-plot variation would be expected to be low compared with variations between plots. Analysis of Nigerian data revealed a high level of within-plot variation; this finding has been confirmed by an independent set of observations from Niger (Poate and Casley 1985).

Olsson (1990) stated that since the within-plot variances are often very large in traditional African farming, then the idea of using a design with at least two subplots is not uneconomical.

While it is true that precision increases with the number of subplots per plot, the question of enumerator work load should not be disregarded. There will be extra effort required to locate, lay and demarcate the subplot boundaries. Supervision of the enumeration quality of such surveys, which requires several visits, is difficult. Without close supervision, the enumerator's tendency to create fictitious data after harvesting just one subplot is to be expected. Because of the variation within the plot, this attitude will lead to such problems of enumerator bias. Poate and Casley (1985) cited two examples experienced in Africa. In a closely supervised experiment on seven villages, enumerators estimated plot yields of millet and sorghum by crop cutting two subplots per plot and comparing them with the total harvest of the plot. In four of the seven villages the coefficients of variation recorded from the subplot were only half the order of magnitude of the coefficients of variation from the harvest of the corresponding whole plot. That is to say subplots were much less variable than whole plots. The correlation coefficient of differences between subplot yields and the corresponding whole-plot yields was greater than 0.8 in eleven of the fourteen cases. If one subplot over-estimated the whole-plot yield by a large amount, the second supposedly independently located subplot usually over-estimated by the same order of magnitude. Records revealed enumerator-specific cases where pairs of subplot yields were similar in size and in dimension of bias. In a second example two subplots were laid as part of the procedure in a project survey on thirteen villages. The coefficients of variation were higher than in the previous example but were still less than 40% of the mean yield in eight villages. The level of correlation between subplots was greater than 0.7 in six villages.

Although these surveys were conducted under controlled conditions, the authors observed that even then many enumerators did not follow the rules necessary to ensure independence of the subplots. Therefore under such circumstances an upward bias is expected.

7.3.7 Conclusion

Nearly all the sources of error reviewed above predict a positive bias. While the magnitude of these biases is hard to foresee, Scott *et al.* (1989) could not attribute to any one of them an average anything near 30%. This is the overall over-estimate for the crop-cut method observed by the authors in their five nations experiment survey. This suggests that the observed bias is due to a combination of the above sources in addition to others such as edge effects (a tendency of inclusion of plants that lie fractionally outside the subplot), over-thorough harvesting (over conscientiously harvesting subplots leaving no waste), non sampling of areas impossible to fit a subplot (acute angle corners) and others.

It seems crop cutting, which is restricted to a few main crops only, is more suitable for a detailed study of crop response in fieldwork on small project evaluation, which is closely supervised, than for a wide current national survey on estimation of crop output.

7.4 Sources of error: (2) Farmer's statements

In this section the procedure used in getting the amount of harvests from the farmers will be outlined. Also various sources of errors will be discussed in the subsequent sections and conclusion given in section 7.4.6 on the method of obtaining information directly from the farmers.

7.4.1 Procedure used

A farmer reports to the enumerator how much he/she expects to harvest (pre-harvest) or if already harvested, how much the farmer got (post-harvest). Depending on the enumerator's question, farmers can give harvest estimates from a plot production to the entire holding. The statements are usually made in terms of traditional units of the farmer's own choice which are subsequently converted into kilograms. Like any other process in a survey, there are errors attributing to the methodology. The following are some of the main sources of error by farmer's statement.

7.4.2 Bias on prevailing circumstances

When a farmer starts to answer questions, like any other respondent in a survey, a complicated machinery of psychological process often starts working. In agricultural surveys, farmers are sensitive to taxation, prices of seeds and fertilizer and many other costs affecting agricultural inputs. Thus even if a farmer knows the accurate replies, chances of deviating from the truth are always there. It is the responsibility of the enumerator, as in any other survey, to explain to the farmer the objective of the survey. Once the enumerator has gained the confidence of the farmer, there is no indication to assess bias on this effect.

7.4.3 Bias on the size of unit

In different parts of the country farmers use different units for measuring their harvest. In Benin, for example, there are about seven such units ranging from about one to four kilogrammes. The cuvette (bowl) in Central African Republic is a standard container but is filled in a different manner in different regions (Scott *et al.* 1989). In Kenya, Tanzania and Zimbabwe a sack (gunny bag) of about 90 *kgs* is a standard container, and a metal box (*debe*) is a smaller unit of about 18 *kgs* for Kenya and

Tanzania. Data become more reliable where there are only a small number of local units. A bias in the conversion of these local units into the standard kilogramme unit is expected.

7.4.4 Bias on rounding up

Some units are inefficiently large. There is no doubt that over-estimation of production occurred for some smaller plots in Zimbabwe because of the only available unit, the bag (Scott *et al.* 1989). A common response of "not even a bag" might be reported as "one bag" for lack of anything precise. This will definitely lead the bias upward.

7.4.5 Bias on state of crop

A farmer's statement of production is usually given in units of volume, whereas what is required is their weight equivalent. The difficulty is that the weight of a volumetric unit changes over time. From their survey, Scott *et al.* (1989) showed that a given volume of maize cobs when weighed would have lost 17% - 20% of its original weight. When subsequently converted into grain equivalent, the samples showed a further loss of between 20% and 40%. This can lead to a substantial upward bias if inappropriate conversion factors are used when the enumerator is not aware of the state of crop during the survey reporting.

7.4.6 Conclusion

Farmer's statements on the production depend very much on the circumstances at the time of survey. A slight suspicion concerning taxation or any other unwanted purpose will create inaccurate data. Apart from the population census, there is little public notice and awareness of on-going surveys. Many developing countries just plunge into surveys without adequate public notice. Well organized, well informed ways and appropriate means of conducting a survey make the farmer cooperative, willing and eager to respond. Enumerators must be able to explain the objectives and aims of surveys in the way the farmers can understand them. Also enumerators must make farmers feel part of them, for a successful interview.

There is no doubt that traditional tools of units contribute a bias, but this can be reduced by applying appropriate conversions.

Farmers' statements are less reliable for harvest on a piecemeal basis. This is mainly common for tuber and root crops.

There has been a vigorous literacy campaign in many African countries which substantially decreases illiteracy of farmers. For example, the literacy rate in Tanzania is over 80%. In the pre-harvest estimate the farmer uses the knowledge of environment and gives a more reliable estimate than the extension field officer's eye estimate. When comparing with the crop-cut method Poate and Casley (1985) remarked that farmer's estimate of crop output, which can be obtained from a large sample, will be no more biased than crop cutting on a similar size and can be collected without great expenditure of resources and skills.

7.5 Data analysis

Data from a crop cutting square subplot of $25m^2$ is derived from Agsasu 1987/88. The crops for crop cutting were maize, paddy, sorghum and millet. A subplot is normally to be demarcated early in the season. However, when the location can not be traced, either because of poles disappearing or ropes decaying in the soil, a new subplot is allocated. The $25m^2$ subplot takes 10 - 15 minutes to demarcate (Andersson and Holmberg 1988), and then the harvest could start. In most circumstances the subplot harvest starts first and then the whole plot. Seldom does it happen the other way. Later when the harvest is over a farmer is asked to report the harvest on the whole measured area of the plot in his own units. The key elements for a plot i are:

1. The crop cutting estimate of production C_i kgs.
2. The farmer's statement on harvest F_i kgs.

7.5.1 The crop cutting weight

In the crop-cut method, the weighing process undergoes several steps. The following were the steps considered in estimating C_i.

1. Raw weight from the subplot - w_i.
2. Sample weight from the collected raw weight - s_i.
3. Final weight; weight after drying - f_i.

Sample weight is taken when the raw weight exceeds 5 *kgs*, otherwise raw weight is the same as sample weight. The weighings are necessary for adjustment to the yield estimate to allow for the moisture content of the grain. The concept of yield rate is that when multiplied by area harvested it would give the harvested production at farm-gate. The loss weight for sampled grains (*kgs*) in a subplot is $(s_i - f_i)$. Therefore the proportion of loss for a sample weight is $\frac{(s_i - f_i)}{s_i}$, and the loss weight for the subplot is

$$\frac{w_i(s_i - f_i)}{s_i}.$$

The final weight for the subplot is $w_i - \frac{w_i(s_i - f_i)}{s_i} = \frac{w_i f_i}{s_i}$.

Since the subplot size is $25m^2$ then the yield rate in kilograms per hectare (*kg/ha*) is

$$400\frac{w_i f_i}{s_i}.$$

Therefore crop cutting estimate of production for plot i is

$$C_i = 400\frac{w_i f_i}{s_i} A_i \tag{7.1}$$

where A_i is the area of the whole plot as measured by the enumerator.

The farmer's statement on harvest for plot i, F_i, is achieved after making standard conversion from the traditional units provided. The primary interest is to compare C_i with F_i.

A simple ratio of the means for C_i compared to F_i is achieved, considering the latter as base value. That is to say

$$\tau = \frac{\sum C_i}{\sum F_i}. \tag{7.2}$$

The comparison will be made on:-

- crop combinations of maize with other crops,
- regional level with crops breakdown,
- regional level, combined.

Tables 7.1, 7.2 and 7.3 below show the results of the above calculations.

Table 7.1: Ratios of crop-cut production to farmers' statement for different crop types in the regions

Region	Crop types	Ratio	Sample size
Dodoma	sorghum	3.002	5
Arusha	maize	3.096	3
	maize+bean	2.410	4
	maize+others	4.245	6
Kilimanjaro	maize+others	6.822	4
Tanga	maize	0.762	7
	maize+others	2.401	4
Morogoro	paddy	2.647	6
Ruvuma	maize	0.798	5
	maize+others	0.804	5
Iringa	maize	1.456	6
	maize+bean	3.605	8
Mbeya	maize	1.106	12
Singida	millet+sorghum	6.258	7
Tabora	maize+others	3.346	6
Shinyanga	maize	5.919	5
	maize+others	1.896	7
	sorghum	3.506	7
Mwanza	maize+others	10.284	8
	paddy	5.844	6

Table 7.2: Ratios of crop-cut production to farmers' statement for maize (pure and combined with other crops) for the country

Crop Combination	Ratio	Sample Size
Maize	1.535	42
Maize+bean	3.101	18
Maize+others	1.568	38

Table 7.3: Ratios of crop-cut production to farmers' statement: by regions

Region	Ratio	Sample size
Dodoma	3.473	6
Arusha	3.459	13
Kilimanjaro	6.822	4
Tanga	1.886	11
Morogoro	2.175	7
Coast	3.934	4
Ruvuma	0.799	11
Iringa	1.891	15
Mbeya	1.509	14
Singida	6.454	12
Tabora	3.061	9
Shinyanga	2.936	20
Mwanza	6.099	16

Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

Ruvuma region, which is situated in the southern part of the country, is the major producer of maize in the country. The region has always enjoyed high yields of maize, except in 1987/88, as can be seen from the figures provided by Regional Agricultural Development Officers in the table given below.

Table 7.4: Yield trend for Ruvuma region

1983/84	1984/85	1985/86	1986/87	1987/88
2400	1970	1960	2435	1800

Source: Ministry of Agriculture and Livestock

Planning and Marketing Division, Dar es salaam, Tanzania.

The drop in yield for the year 1987/88 is an indication that it was a bad year for that region. A look at the data on the area of sampled plots taken at Namatuhi village, Songea sub-urban in Table 7.5 shown below indicates that harvested areas were much greater than planted areas, though planted areas were somehow agreeable to the areas measured by enumerators.

Table 7.5: Areas as reported by farmers and enumerators: Namatuhi, Ruvuma region

Crop type	Area (*ha*)			Total field(s) area
	Planted	Measured	Harvested	
Maize	1.50	1.55	6.00	6.50
	2.50	1.91	3.00	3.50
	2.00	2.33	7.00	7.00
	2.50	2.79	5.00	6.00
	5.00	0.10	5.00	8.50
Maize+others	2.00	2.58	5.00	5.00
	1.00	0.59	4.50	4.50
	1.50	1.45	3.50	4.50
	1.50	1.80	4.00	5.50
	4.00	3.32	5.00	8.50
C.V.	50%	51%	23%	27%

Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

One reason for the values shown in Table 7.5 can possibly be that on realizing the bad situation, farmers extended their planted area to balance production. This must have been done in the early stages of planting and after enumerators had measured the fields.

The crop cutting production in equation (7.1) will then have to be low, since A_i, the measured area, is relatively smaller compared to harvested area. This in turn must have affected τ in equation (7.2). The coefficient of variation for harvested area is low, indicating that farmers were trying to make use of the whole farming land which does not vary much within a village.

On the other hand the high ratio in Mwanza region indicates over-estimation of maize combined with other crops. In this region because of the nature of land structure, that is, scattered cultivation separated by rocky areas and bushland, a common combination with maize apart from cow peas and greengarm is cassava. This root crop normally creates a wider spacing for maize plants which suggests that enumerators were somehow avoiding the less dense areas and created a notable upward bias. The same is noted in the neighbouring Shinyanga region which has similar features but to a lesser extent.

There also had been a substantial over-estimation in Arusha region and neighbouring Kilimanjaro region. These are intensive cultivated areas. Most settlements are along the numerous mountainous sites. Because of that there are a considerable number of fields in the hilly areas. The problem of area measurements as discussed in section 2.8 prevails. Also selection of subplots at steep slopes where crops are grown in small patches may contribute to higher ratios in these regions.

The same size of the subplot has been used throughout the survey for different sizes of plots. This is appropriate for estimating the national production and its errors, as the ratios shown in Tables 7.1, 7.2 and 7.3 give greater weight to larger plots. However since the interest is in measurement errors then it becomes less appropriate. In chapter four it has been experienced that there were variations of yields not only between fields/plots but also within, which affected production. One statistical approach to this problem is to perform some mathematical transformation to the observations. This means therefore, analysing not the original observations, but a transformation of them, such as their square roots or their logarithms, though the new scales are not so easy to appreciate (Clarke 1980). However such kind of transformations simplify, create less variabilities and make distributions nearly normal. The kind of transformation considered in this case is taking the logarithm of the ratios at the individual plot level. When these are averaged the variation in plot size is eliminated. Scott *et al.* (1989) noted that the transformation reduces the impact of cases with very large error which are given undue weight in the conventional estimate, and at the same time normalizes the observation and allows us to present meaningful sampling standard errors. Tables 7.6, 7.7 and 7.8 show the corresponding estimates and their standard errors using the variable

$$\log_{10}\left[\frac{C_i}{F_i}\right].$$

Table 7.6: Log of ratio of crop cutting production to farmer's statement: maize and combination for the country

Crop combination	Mean	Standard Error
Maize	0.1713	0.0860
Maize+bean	0.4884	0.0956
Maize+others	0.4877	0.0915

Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

Table 7.7: Log of ratios of crop cutting production to farmer's statement: crop types in the regions

Region	Crop types	Mean	Standard Error
Dodoma	sorghum	0.4144	0.1124
Arusha	maize	0.7728	0.2846
	maize+bean	0.4938	0.2029
	maize+others	0.6943	0.1412
Kilimanjaro	maize+others	0.8332	0.0519
Tanga	maize	0.1227	0.2645
	maize+others	0.4070	0.3647
Morogoro	paddy	0.4114	0.1136
Ruvuma	maize	-0.3012	0.2585
	maize+others	-0.2077	0.1526
Iringa	maize	0.2366	0.1506
	maize+bean	0.4747	0.1213
Mbeya	maize	-0.0361	0.0901
Singida	millet+sorghum	0.7608	0.1152
Tabora	maize+others	0.4817	0.0634
Shinyanga	maize	0.6215	0.2249
	maize+others	0.4642	0.1797
	sorghum	0.0864	0.2493
Mwanza	maize+others	0.8788	0.1348
	paddy	0.4753	0.2307

Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

Table 7.8: Log of ratio of crop cutting production to farmer's statement: by regions

Region	Mean	Standard Error
Dodoma	0.4935	0.1183
Arusha	0.6507	0.1156
Kilimanjaro	0.8332	0.0519
Tanga	0.2261	0.2182
Morogoro	0.3639	0.0860
Coast	0.5029	0.1734
Ruvuma	-0.2545	0.1508
Iringa	0.4117	0.1005
Mbeya	0.0087	0.0900
Singida	0.7905	0.0761
Tabora	0.3878	0.1125
Shinyanga	0.3637	0.1307
Mwanza	0.6747	0.1210

Source: Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

The farmers were asked to report genuinely their production. There was no material reward given to enhance cooperation nor was any special favour provided to the sampled farmers. At least efforts were made to inform the farming community at all levels including the village's. One may not take for granted the farmer's willingness to report crop production accurately, but one should also not undermine the farmer's integrity and honesty. The objective of Agsasu was not to compare crop-cut method with farmer's statement, and the result is therefore free from any prejudice which makes it valid and generalizable. It is an important practical aspect in our intention in the search for appropriate methodologies for generating food production statistics in developing countries, particularly in this case, sub-Saharan Africa.

Different regions pose different obstacles for effective use of the crop-cut method which ultimately when performed without careful supervision and experienced and careful enumerators, leads to over-estimation. It is therefore possible under these circumstances to concur with the conclusion of Scott *et al.* (1989), that estimation of production on the basis of farmer's post-harvest reports are likely to be a more accurate, economical and practical approach which may be the approach which can be sustained in most difficult cases.

If the crop-cut method is to be used as the method of estimating yield rate, then with the big number of village establishments, the number of subplots will have to be increased greatly (Panse *et al.* 1966). The authors warned that the organisational and financial implication of such huge programmes will impose an unmanageable burden on the governments and urged for alternative techniques.

Tanzania, like most of the countries in sub-Saharan Africa, is a large country requiring large scale crop cutting for her varieties of crops. The above results suggest that enquiries based on farmer's statements can employ larger and more efficient samples, quickly and cheaply, and can be more easily sustained due to the limited resources and statistical capability.

7.6 Estimation of yield rate

In the Mussa Field Survey, farmers were asked to report production after harvest in units which they understood. On a few occasions the problem of conversion was experienced, for example when some farmers had already stored their produce in storage huts (*vihenge*) of different sizes and could not be able to recall the quantity. In such situations, comparison was made with similar huts of known production encountered earlier, also by enquiring on consumption rate, *etc.*, to make a comparative conversion.

From the sampled households in a village, crop yield rate can be achieved from the production that the farmer stated over the measured cropped land area. The summary data on area and production from the sampled villages are given in Appendix IX. Two types of yield rates have been calculated. One type is the crop category yield rate, this is shown in Table 7.9. The other type is particular crop yield rate, after adjusting mixed cropped land area to its respective crops. The latter makes use of proportions in Table 6.21, and is shown in Table 7.10.

Table 7.9: Average crop category yield for the villages

		Yield rate (kg/ha)			
			Maize+bean		
Village	Year	Maize	Maize	Bean	Bean
Nyankumbu	1989	na	341	75	na
	1990	na	238	47	na
Igawilo	1989	2371	na	na	390
	1990	2236	na	na	855
Ngarash	1989	881	1151	470	670
	1990	1641	712	87	1325
Sinoni	1989	1946	1021	263	na
	1990	1414	644	211	na

na = not available or not analysed because of the insignificant number of households.

Table 7.10: Average crop yield for the villages after mixed cropping adjustment

		Yield rate (kg/ha)	
Village	Year	Maize	Bean
Nyankumbu	1989	533	209
	1990	373	131
Igawilo	1989	2371	390
	1990	2236	855
Ngarash	1989	1027	1282
	1990	1333	1224
Sinoni	1989	1376	2403
	1990	952	1931

The yield rates of beans for Ngarash and Sinoni are high because of the small land apportioned to beans from the mixed cropping land area of maize+beans. From the Agsasu data analysis, the proportion shown in Table 6.21 is the smallest among the three regions $(1 - 0.890733 = 0.109267)$.

The overall yield rates from the selected villages for particular crop and year are the averages calculated from Table 7.10 and are shown in Table 7.11. The averages of crop yields as obtained from surveys performed by Bureau of Statistics and Ministry of Agriculture and Livestock are shown in Table 7.12.

Table 7.11: Average crop yield based on the sampled villages

	Yield rate (kg/ha)	
Crop	1989	1990
Maize	1327	1224
Beans	1071	1035

Table 7.12: Average crop yield from other surveys

	Yield rate (kg/ha)			
	1986/87		1987/88	
Survey	Maize	Bean	Maize	Bean
Agsasu/Crop Cutting	1560	na	1082	na
Agsasu/Farmer's Interview	1430	1330	1304	461
MALD	1714	na	1450	na

Key:

Agsasu ═ Agricultural Sample Surveys.
MALD ═ Ministry of Agriculture and Livestock.
na ═ not available

Source: (i) Mussa Field Survey

(ii) Agricultural Sample Survey 1987/88

Bureau of Statistics, Dar es salaam, Tanzania.

(iii) Ministry of Agriculture and Livestock

Planning and Marketing Division, Dar es salaam, Tanzania.

While at the moment it may be difficult to state which one is a better value for crop yield because of the difference of years between the other surveys and the Mussa Field

Survey, it does seem from Tables 7.11 and 7.12 that farmer's interview trend on crop yield tends to agree with the average based on sampled villages. The value of maize yield in Agsasu crop cutting for 1987/88 is very low. It is about 31% lower than that of 1986/87, while that of Agsasu farmer's interview is about 9% lower and the MALD 16%. The value seems to suffer errors of the kind explained in chapter two (section 2.1.1), because it contradicts the ratios obtained in Table 7.1 from the same set of data.

7.7 Estimation of Production

Production from the sampled villages can be potrayed in various ways depending on the objective and the interest of the users. One way is to show as discussed in chapter six (section 6.5.1), by considering the areas and production of the sampled households as extracted from Appendix IX.

Table 7.13: Area and production for crop/crop mixture by village for the sampled households

Village	Year	crop/crop mixture	Area (*ha*)	Production	
				Maize	Beans
Nyankumbu	1989	Maize+bean	42.28	14405	3192
	1990	Maize+bean	46.17	10991	2175
Igawilo	1989	Maize	14.43	34208	na
		Beans	1.55	na	605
	1990	Maize	14.69	32850	na
		Beans	1.62	na	1385
Ngarash	1989	Maize	42.62	37545	na
		Maize+bean	26.29	30270	12360
		Beans	14.17	na	9490
	1990	Maize	42.20	69270	na
		Maize+bean	27.35	19480	2380
		Beans	12.75	na	16890
Sinoni	1989	Maize	6.67	12980	na
		Maize+bean	18.61	19010	4887
	1990	Maize	7.57	10710	na
		Maize+bean	17.18	11070	3625

na = not available or not analysed because of insignificant number of households.

Although Table 7.13 is very informative on how crops or farming is carried out, however when all crops/crop mixtures are involved it can extend much further and become large. Another way to show production from the sampled villages is to give areas for single crops as obtained from the estimation of proportion in chapter six (section 6.5.3). This kind of way is said to give a wrong impression of how crops are grown but the situation can be cleared by providing detailed explanation.

Production for maize and beans as pure stands can therefore be estimated by multiplying the yields from Table 7.11 with the areas obtained through the remote sensing process in Table 6.22. The estimated production is therefore shown in Table 7.14 with standard errors in brackets. The standard errors for the yield have been calculated from the average yields in Table 7.10.

Table 7.14: Production for the pure stand crops by village

Village	Year	Production (*kg*) Maize	Bean
Nyankumbu	1989	183885	40755
		(37466)	(27317)
	1990	139875	27772
		(28852)	(15604)
Igawilo	1989	1021901	21450
		(62262)	(29326)
	1990	1057628	34200
		(88980)	(21831)
Ngarash	1989	1199536	357678
		(114185)	(63773)
	1990	1551612	314568
		(123931)	(47938)
Sinoni	1989	1104928	156195
		(43529)	(4821)
	1990	621656	84964
		(46758)	(3526)

These production values can be extrapolated by appropriate weights to regional and national (mainland) estimates.

Chapter eight

General conclusion

8.1 Findings and conclusions

This study realized that there is a problem of coordination between various data collection sources in most countries of sub-Saharan Africa. The data collected seemed to lack reliability because of statistical flaws in the collection process.

It is recommended that a single body be nominated in a country for coordinating and justifying national statistics before they are put forward to the users.

In chapter four, the application of remote sensing for providing information on cropped land was made. Indeed remote sensing has been applied in several other projects, but a major disadvantage in most of the applications was noted. The projects were mostly carried out by foreign experts or firms. Adeniyi (1989) observed that the knowledge of how remote sensing has been applied is not known by nationals and the advantages of using remote sensing are usually lost when experts leave at the end of assignments.

Remote sensing data had played a major role in the area estimation of an entire segment of interest. However, it was found out that subdivision of land use into smaller units cannot be carried through computer analysis or classification of digital data, since such a division is based on per pixel which easily leads to misclassification. Some farms which constitute the smaller units are hardly the size of a pixel. Visual analysis was able to provide quick and reliable results concerning the cultivated area, distinguishable on the image village areas (the segment of our interest).

Therefore, for the present circumstances, the study maintains the use of remote sensing data by resorting to simple and affordable measures to encourage its continuation and the involvement of their own nationals in the sub-Sahara African developing countries, which must rely on easily transferable technology. However, further research is recommended in the use of digital image classification by devising adaptable programs to be used in ordinary personal computers (PCs), since at the moment digital image processing is performed only in mainframe or micro-computer with graphics or a purpose built digital processor (*eg*. GEMS, IDP 3000 and Dipix Aries II). Digital image processing is the numerical manipulation of digital images and includes preprocessing, enhancement and classification. By devising techniques in which classification can be managed by computer user friendly packages, it is expected that subdivision of smaller units can possibly be made based on land use/cover or varying spectral properties of the ground.

The role of remote sensing was not to replace other systems completely, but rather to be used in conjuction with other techniques, including ground sampling, in an overall multistage sampling procedure. The remote sensing data, which dealt with the overall area of the entire segment, was used as auxiliary information. In chapter five, ratio-type estimators were found to be suitable in area estimation because of their effective use of auxiliary information. Thus, a combination with the ground sampling data enabled ratio-type estimators to be appropriately applied. The modified ratio estimator $\overline{y}_{MOD}$, proposed in that chapter, proved to be better when compared with $\overline{y}_r$, $\overline{y}_{C1}$ and $\overline{y}_{C2}$ because it is more efficient and has smaller bias as calculated from the Field Survey - Mussa 1990 data. Also it is preferable because of its simplicity and suitability in its application to African farms' structures.

As the farm areas vary within and between households annually because of farming pattern, fallowing, *etc.,* chapter six managed to show that crop area measurements must be done by enumerators, as farmers seemed not to apprehend consciously the changing areas. Also in the chapter, data obtained from satellites was integrated with ground sampling data on crop areas at household level. The method managed to eliminate the discrepancy of misrepresentation caused by others not incorporating satellite data from about -10% up to +90%.

In chapter seven an alternative method to the crop-cut method was adopted as an annual measure for estimating yield rate. The crop-cut method was found to be highly biased, difficult to handle in certain situations and expensive in terms of both equipment, supervision and enumerator's time. On the other hand, the alternative method, the farmer's statements, was found to be capable of being handled by the limited available enumerators with marginal additions to provide for requisite supervision of the field work and data processing. The method is much simpler and more accurate than the crop-cut method. In the comparison of 1987/88 Bureau of Statistics' agricultural sample survey data, it was shown that in all but one of the 13 regions, the crop-cut method led to over-estimation.

Finally production statistics, which are a product of areas and yield rates were found to be calculated with much reliability and confidence from the remote sensing data combined with ground sampling area measurements and farmer's statements on yields respectively.

In a real situation, problems exist which emphasize the recommendation for simple analyses, not that they are second best but rather all that possibility can support. As in most developing countries, the resource constraints in government institutions in sub-Saharan Africa place a limit on the amount of funds and staff time spent, so an estimation procedure without resorting to complex and time consuming measures is preferable.

References

Abiodun, A.A., Oesberg, R.P., 1988; Basic principles of remote sensing, *Report of the Eleventh International Training Course on Applications of Remote Sensing to Agricultural Statistics, (5-20 May, 1986), Rome and Montpellier.* FAO, Rome.

Adeniyi, P.O., 1989; Remote sensing in Nigeria, *Remote Sensing for Development: Proceedings of the International Conference, September, 1986, Berlin.* part II, Ed. Abertz, J., Tauch, R.

Ahlcrona, E., 1988; The impact of climate and man on land transformation in Central Sudan - Application of remote sensing. *Meddelanden Fran Lunds Universitets Geografsika Institutioner.* Avhandlingar **103:**

Alonzo, D.C., 1988; An overview of food and agricultural statistics, *Report of the Eleventh International Training Course on Applications of Remote Sensing to Agricultural Statistics, (5-20 May, 1986), Rome and Montpellier.* FAO, Rome.

Anderson, P., Holmberg, M., 1988; Some measurement problems in the 1986/87 agricultural sample survey in Tanzania, *Working Paper.* **58:** Uppsala.

Battese, G.E., Fuller, W.A., 1981; Prediction of county crop areas using survey and satellite data, *American Statistical Association: Proceedings of the Section on Survey Research Methods.* pp 500-505.

Barrett, E.C., Curtis, L.F., 1982; *Introduction to Environmental Remote Sensing.* (2nd edn), Chapman and Hall, London.

Bartholome, E., 1986; Remote sensing as an aid for area estimation of rainfed crops in Sahelian countries, *presented at 20th International Symposium on Remote Sensing of Environment, (4-10 Dec.), Nairobi.*

Bird, A.C., 1991; Principles of remote sensing: Electromagnetic radiation, reflectance and emissivity, *In* Belward, A.S., Valenzuela, C.R., (eds) *Remote Sensing and Geographical Information Systems for Resources Management in Developing Countries.* Vol.1, Kluwer

Academic Publisher, London.

Brooks, S.H., 1955; The estimation of an optimum subsampling number, *Journal of the American Statistical Association.* **50:** pp 398-415.

Casley, D.J., Kumar, K., 1988; *The Collection, Analysis and Use of Monitoring and Evaluation Data.* Johns Hopkins, A World Bank Publication, London.

Casley, D.J., Lury, D.A., 1985; *Data Collection in Developing Countries.* Clarendon Press.

Chakrabarty, R.P., 1973; A note on the small sample theory of the ratio estimator in certain specified populations, *Journal of the Indian Society of Agricultural Statistics.* Vol.XXV, No.2, pp 49-57.

Chakrabarty, R.P., 1979; Some ratio-type estimators, *Journal of the Indian Society of Agricultural Statistics. Vol.XXXI, No.1, pp 49-62.*

Chaudhary, S.M., 1973; *Introduction to Statistical Theory – part II.* Markazi Kutub Khana, Lahore.

Chhikara, R.S., 1984; Effect of mixed (boundary) pixels on crop proportion estimation, *Remote Sensing of Environment.* **14:** pp 207-218.

Chhikara, R.S., 1986; Error analysis of crop acreage estimation using satellite data, *Technometrics.* Vol.28, No.1, pp 73-80.

Clarke, G.M., 1980; *Statistics and Experimental Design.* (2nd edn), Edward Arnold, London.

Cochran, W.G., 1977; *Sampling Techniques.* (3rd edn), John Wiley, New York.

Colwell, J.E., 1977; *Uses of Remote Sensing for Agricultural Statistics in Developing Countries.* ERIM Report No. 111800-18-T, Ann Arbor, Michigan.

Curran, P.J., 1980; Multispectral remote sensing of vegetation amount, *Progress in Physical Geography.* **4:** pp 315-341.

Curran, P.J., 1985; *Principles of Remote Sensing.* Longman group, New York.

Daily News, 1989; Report on the Royal Agricultural Society of the Commonwealth, *Tanzanian*

Newspaper. Saturday, February 25, p 3.

David, I.P., Sukhatme, B.V., 1974; On the bias and mean square error of the ratio estimator, *Journal of the American Statistical Association*. **69:** pp 464-466.

Des Raj, 1964; A note on the variance of the ratio estimate, *Journal of the American Statistical Association*. **59:** pp 895-898.

Dunford, C., Mouat, D.A., Norton-Griffiths, M., Slaymaker, D.M., 1983; Remote sensing for rural development planning in Africa, *I.T.C. Journal*. **2:** pp 99-108.

Durbin, J., 1959; A note on the application of Quenouille's method of bias reduction to estimation of ratios, *Biometrika*. **46:** pp 477-480.

Eklof, Jan A., 1984; Institutional cooperation between Sweden and developing countries in the field of Statistics, *Statistick Tidskrift Statistical Review*. **4:** pp 331-338.

FAO, 1982; *Estimation of crop areas and yields in agricultural statistics*. **22:** Food and Agriculture Organization, Rome.

FAO, 1986; *Programme for the 1990 World Census of Agriculture*. Food and Agriculture Organization, Rome.

Fischer, W.A., 1975; *History of Remote Sensing, Manual of Remote Sensing*. (1st edn), Vol.1.

Fisher, R.A., 1958; *Statistical Methods for Research Workers*. (13th edn), Oliver and Boyd, Edinburgh.

Goodman, L.A., Hartley, H.O., 1958; The precision of unbiased ratio-type estimators, *American Statistical Association Journal*. pp 491-508.

Hall-Konyves, K., 1988; Remote sensing of cultivated lands in south Sweden. *Meddelanden Fran Lunds Universitets Geografsika Institutioner*. Avhandlingar **102:**

Hartley, H.O., Ross, A., 1954; Unbiased ratio estimators, *Nature*. Vol.174, pp 270-271.

Helden, U., 1981; Landsat data as an aid for obtaining Agricultural Statistics, *TCP Project NER/8904*. FAO.

Helden, U., 1984; Drought impact monitoring - A remote sensing study of desertification in Kordofan, Sudan, *Lunds Universitets Naturgeografiska Institution.* **61:**

Helden, U., 1987; *An assessment of woody biomass, community forests, land use and soil erosion in Ethiopia - A feasibility study on the use of remote sensing and GIS-analysis for planning purposes in developing countries.* Lund Univ. Press.

Helden, U., Eklundh, L., 1988; *National drought impact monitoring - A NOAA NDVI and precipitation data study of Ethiopia.*

Howard, J.A., 1988; Outline of remote sensing, *Report of the Eleventh International Training Course on Applications of Remote Sensing to Agricultural Statistics, (5-20 May, 1986), Rome and Montpellier.* FAO, Rome.

Idaikkadar, N.M., 1979; *Agricultural Statistics - A handbook for Developing Countries.* Pergamon Press, Oxford.

Idso, S.B., Jackson, R.D., Reginato, R.J., 1977; Remote sensing crop yield, *Science.* p 19.

Kendall, M.G., Stuart, A., 1963; *The Advanced Theory of Statistics.* Vol.1, (2nd edn), Charles Griffin, London.

Kershaw, C.D., 1987; Discrimination problems for satellite images, *International Journal of Remote Sensing.* Vol.8, No.9, pp 1377-1383.

Kikula, I.S., 1987; Surveillance of agriculture and its impacts on environment, *Research Proposal Paper.* Institute of Resources Assessment, Dar es salaam.

King, R.B., 1981; An evaluation of Landsat-3 RBV imagery for obtaining environmental information in Tanzania, *In* Allan, J.A., (ed.) *Matching Remote Sensing Technologies and their Applications.* Remote Sensing Society, Reading, pp 85-95.

King, R.B., 1982; Rapid rural appraisal with Landsat imagery: A Tanzanian experience, *Z. Geomorph. N.F. Suppl.-Bd.* **44:** pp 5-20, Berlin-Stuttgart.

King, R.B., 1984; *Remote Sensing Manual of Tanzania.* Land Resources Development Centre, UK.

Lahiri, D.B., 1951; A method for sample selection providing unbiased ratio estimates, ***Bulletin of** the International Statistical Institute.* **33:** 2, pp 133-140.

Lamprey, R., 1985; Impact on Kenya savanna vegetation: A remote sensing approach, ***Ph.D.** Thesis.* Univ. of Aston, UK.

Latham, J., Som, H., 1991; Remote sensing and agricultural statistics in developing countries, *Euro Courses Manuscripts on Remote Sensing Applied to Agricultural Statistics: **Regional** Inventories (20-24 May), Ispra.*

Lechi, G., 1988; Satellite sensors and satellite image characteristics, *Report of the Eleventh International Training Course on Applications of Remote Sensing to Agricultural Statistics, (5-20 May, 1986), Rome and Montpellier.* FAO, Rome.

Lennington, R.K., Sorensen, C.T., Heydorn, R.P., 1984; A mixture model approach for estimating crop areas from Landsat data, *Remote Sensing of Environment.* **14:** pp 197-206.

Lillesand, T.M., Kiefer, R.W., 1987; *Remote Sensing and Image Interpretation.* (2nd edn), Wiley, New York.

MacDonald, R.B., Hall, F.G., 1980; Global crop forecasting, *Science.* **208:** pp 670-679.

Mahalanobis, P.C., Sengupta, J.M., 1951; On the size of sample cuts in crop-cutting experiments in the Indian Statistical Institute: 1939-1950, *Bullettin of the International Statistical Institute.* **33:** 2, pp 349-404.

Makusi, R.R., 1990; The need of coordination of agricultural statistics in MALD, *Unpublished Manuscript.*

MALD, 1989; *Current Agriculture Survey 1987/88.* Ministry of Agriculture and Livestock, Dar es salaam, pp 11-20.

Malingreau, J-P.,1981; Remote sensing and technology transfer in a developing society, *I.T.C. Journal.* Vol.2, pp 125-138.

McCloy, K.R., Smith, F.R., Robinson, M.R., 1987; Monitoring rice areas using Landsat MSS data, *International Journal of Remote Sensing.* **8:** pp 741-749.

Mickey, M.R., 1959; Some finite population unbiased ratio and regression estimators, *Journal of the American Statistical Association.* **54:** pp 594-612.

Midzuno, H., 1951; On the sampling system with probability proportionate to sum of sizes, *Ann. Inst. Stat. Math.* **2:** pp 99-108.

Moreira, M.A., Chen, S.C., Batista, G.T., 1986; Wheat area estimation using digital Landsat MSS data and aerial photographs, *International Journal of Remote Sensing.* **7:** pp 1109-1120.

Olsson, K., 1985a; Remote sensing for fuelwood resources and land degradation studies in Kordofan, the Sudan, *Meddelanden Fran Lunds Universitets Geografsika Institutioner.* Avhandlingar **C:**

Olsson, L., 1985b; An Integrated study of Desertification, *Meddelanden Fran Lunds Universitets Geografsika Institutioner.* Avhandlingar **XCVIII:**

Olsson, U., 1990; *Approaches to Agricultural Statistics in Developing Countries – an appraisal of ICO Experiences.* International Consulting Office, Reports 12, Statistics Sweden.

Ottichilo, W.K., Peden, D.G., Agatsiva, J.L., Mwendwa, H., 1985; Maize harvest forecast for 1984 in Bungoma, Elgeyo Marakwet, Kakamega, Nakuru, Nandi, Trans Nzoia, Uasin Gishu and West Pokot districts, *Technical Report No. 115.* KREMU, Nairobi.

Ouedraogo, H., 1980; The African remote sensing programme, *Proceedings of the Fourteenth International Symposium on Remote Sensing of Environment, (23-30 April), Addis Ababa.* Vol.1, UN. Economic Commission for Africa, pp 103-111.

Pacheco, R.A., 1980; Applications of remote sensing to agriculture development in tropical countries, *In* Balkema, A.A., (ed.) *Remote Sensing Application in Agriculture and Hydrology.* Rotterdam, pp 299-308.

Panse, V.G., 1963; Plot size again, *Journal of the Indian Society of Agricultural Statistics.* Vol.15, pp 151-159.

Panse, V.G., 1964; *Estimation of crop yields.* FAO, Rome.

Panse, V.G., Rajagopalan, M., Pillai, S.S., 1966; Estimation of crop yields for small areas, *Biometrics.* **22:** pp 374-384.

Pearce, S.C., Clarke, G.M., Dyke, G.V., Kempson, R.E., 1988; *A Manual of Crop Experimentation.* Charles Griffin & Co., London.

Peden, D.G., Ottichilo, W.K., Mwendwa, H., 1985; Estimating pre-harvest production of maize in Kenya using large-scale aerial photography and radiometry, *Remote Sensing Yearbook.* pp 129-138.

Petricevic, P.I., 1988; Objective measurement of area and yields, *Report of the Eleventh International Training Course on Applications of Remote Sensing to Agricultural Statistics, (5-20 May, 1986), Rome and Montpellier.* FAO, Rome.

Poate, C.D., Casley, D.J., 1985; *Estimating Crop Production in Development Projects: Methods and Their Limitations - A Technical Supplement to Monitoring and Evaluation of Agriculture and Rural Development Projects.* World Bank, Washington.

Quenouille, M.H., 1956; Notes on bias estimation, *Biometrika.* **43:** pp 353-360.

Rao, J.N.K., Webster, J.T., 1966; On two methods of bias reduction in estimation of ratios, *Biometrika.* **53:** pp 571-577.

RCO, 1981; *Arusha Region Today: 1981.* Regional Commissioner's Office and Development Alternatives, Inc, Washington D.C.

Reddy, V.N., 1978; A study on the use of prior knowledge on certain population parameters in estimation, *Sankhya.* **4:** Series C, pp 29-37.

Rothfield, R., Kumar, B., 1980; Report on the census of agriculture 1978, *Parliamentary Paper.* No.28, Fiji.

Rutherford, A.A., 1988; Critique of 1986/1987 agricultural sample survey, *Unpublished Manuscript.*

Sabins, F.F., 1986; *Remote Sensing: Principles and Interpretation.* (2nd edn), W.H. Freeman & Co., New York.

Sandbrook, R., 1985; *The Politics of Africa's Economic Stagnation.* Univ. Press, Cambridge.

Schowengerdt, R.A., 1983; *Techniques for Image Processing and Classification in Remote Sensing.* Academic Press.

Scott, C., 1989; Notes on sampling theory, *Unpublished Manuscript.*

Scott, C., Merchant, T., Verma, V., 1989; Experiments on the estimation of agricultural production in Africa, *I.S.I.* pp 305-327, 47 Session, August 29-September 6, Paris.

Sengupta, J.M., 1964; On perimeter bias in sample cuts of small size, *Sankhya.* Series B, Vol.26, pp 53-68.

SERISS, 1987; *Sudan Emergency and Recovery Information and Surveillance System.* Remote Sensing Project, Final report, Ministry of Agriculture, Sudan.

Sisodia, B.V.S., Dwivedi, V.K., 1981; A modified ratio estimator using coefficient of variation of auxiliary variable, *Journal of the Indian Society of Agricultural Statistics.* **33:** pp 13-18.

Smith, J.H., Ramsey, D.B., 1982; A crop area estimator based on changes in the temporal profile of a vegetative index, *American Statistical Association.* pp 495-498.

Srivastava, S.K., 1967; An estimator using auxiliary information in sample surveys, *Calcutta Statistical Association Bulletin.* Vol.16, pp 121-132.

Srivastava, S.K., 1980; A class of estimators using auxiliary information in sample survey, *Canadian Journal of Statistics.* **8:** pp 252-254.

Srivastava, S.K., Jhajj, H.S., Sharma, M.K., 1986; Comparison of some estimators of K and B in finite populations, *Journal of the Indian Society of Agricultural Statistics.* Vol.XXXVIII, No.2, pp 230-236.

Standard, 1990; FAO says it has a solution to Africa's hunger, malnutrition, *Kenya Newspaper.* Friday, July 13, p 6.

Stern, M., 1985; Census from heaven ?, *Meddelanden Fran Lunds Universitets Geografsika Institutioner.* Avhandlingar **XCIX:**

Stromquist, L., Larsson, R.A., Bystrom, M., 1988; *An Evaluation of the SPOT Imagery Potential for Land Resources Inventories and Planning – A Lesotho Case Study.* UNGI Rapport Nr.68, Uppsala.

Sukhatme, P.V., 1947a; The problem of plot size in large-scale yield surveys, *Journal of the American Statistical Association.* Vol.42, pp 297-310.

Sukhatme, P.V., 1947b; Use of small-size plots in yield surveys, *Nature.* Vol.160, p 542

Sukhatme, P.V., Sukhatme, B.V., 1974; *Sampling Theory of Surveys with Applications.* (2nd edn), Asia Publishing House, London.

Taherkia, H., 1985, Remote sensing applied to slope stability in mountainous roads in Iran, *Ph.D. Thesis,* Univ. of Aston, UK.

Takwimu 1988 *Agriculture Sample Survey of Tanzania Mainland 1986/87,* Bureau of Statistics, Daressalaam, Tanzania.

Takwimu, 1989. *Agriculture Sample Survey of Tanzania Mainland 1987/88,* Bureau of Statistics, Daressalaam, Tanzania.

Taylor, J., 1985; Agricultural applications of remote sensing, *Proceedings of the International Symposium on Farm Electronics and Computing, Royal Agricultural Society of England, (20-25, October), Stratford.* pp 319-331.

Taylor, J., 1991; Principles of remote sensing and available data, *Euro Courses Manuscripts on Remote Sensing Applied to Agricultural Statistics: Regional Inventories (20-24 May), Ispra.*

Tin, M., 1965; Comparison of some ratio estimates, *Journal of the American Statistical Association.* **60:** pp 294-307.

Verma, V., Merchant, T., Scott, C., 1988; Evaluation of crop-cut methods and farmers report for estimating crop production, *Longacre.* London.

Wigton, W.H., Huddleston, H.F., 1978; A land use information system based on statistical inference, *Proceedings of the Twelfth International Symposium on Remote Sensing of Environment, (20-26, April), Ann Arbor.* Vol.I, pp 429-441.

Williams, W.H., 1978; How bad can "good" data really be ?, *The American Statistician.* Vol.32, No.2, pp 61-65.

Wu, C.F., 1982; Estimation of variance of the ratio estimator, *Biometrika.* **69:** pp 183-189.

Zarkovich, S.S., 1966; *Quality of Statistical data.* FAO, Rome.

APPENDIX I

CALCULATORS FOR AREA MEASUREMENT

CASIO FX-795P AND FX-790

The instruction and programme for the calculator fx-795P and instruction for calculator fx-790P

(chapter 2, section 2.2.1.4.3)

How to use the area measurement program

1. Power on
2. Press P4
3. Screen will show " Angle No.1 "
4. Input first angle in degrees
5. Screen will show " Side No.1 "
6. Input first side distance in meters
7. Operations 3-6 will repeat for second angle and second side respectively
8. After last side, execute P1 for closing error
9. Execute P2 for area in hectares
10. Execute P3 for area in acres

AREA MEASUREMENT PROGRAM CODED IN
BASIC PROGRAMMING LANGUAGE
APPLICATION SYSTEM - CASIO FX-795P
PROGRAM P0

This is the program which will start after pressing P4

```
5  SET F0: CLEAR 10 N=N+1
20 PRINT " ANGLE No. " ; N;
30 INPUT "        ", A
40 IF A > 360 THEN GOTO 170
50 PRINT " SIDE No. " ; N;
60 INPUT "        ", S
70 I=I+S
80 H=H+S*S
90 T=S*COS(A)
100 U=S*SIN(A)
110 B=B+T
120 D=D+U
130 F=F+U*B-T*D
140 C=C+B
150 E=E+D
160 GOTO 10
170 BEEP 1: BEEP 1
180 PRINT " Too Large Re-enter! " ;
190 GOTO 30
```

PROGRAM P1

This program gives the closing error

```
5 SET F2
10 V=SQR(B*B+D*D)/I*100
20 PRINT CSR1; " Closing error = "; CSR16;V;CSR22; " % "
```

PROGRAM P2

This program gives the area in hectares

```
5 SET F2
10 X=((E*B-C*D)/(N-1)+F/2)/10000
20 PRINT CSR1; " Area = "; CSR12;X;CSR21; " ha."
```

PROGRAM P3

This program gives the area in acres

```
5 SET F2
10 Y=2.47*((E*B-C*D)/(N-1)+F/2)/10000
20 PRINT CSR1; " Area = "; CSR12;Y;CSR21; " ac."
```

PROGRAM P4

This program initiates PROGRAM P0

```
5 COSUB #0
```

CASIO FX-790P

How to use the area measurement program

	Key	Display
1	Power on	Ready P0
2	Shift 0	Angle no. 1?
3A	{Angle in degrees}	
	EXE	Side no. 1?
3B	{Side in meters}	
	EXE	Angle no. 2?

Repeat steps 3A and 3B for all angles and sides of the field. After the last side press

4. Shift 1 Closing error=

5. Shift 2 Area=...ha

6. Shift 3 Area=...ac

Repeat steps 2-5 for each new field. Note: The order of steps 4 - 6 is irrelevant.

EXAMPLE OF AREA CALCULATION

	Angle in degrees	Side in meters
A - B	84	266
B - C	200	228
C - D	305	181
D - A	332	98

Results:

Closing error:
Shift 1 1.12 %

Area in ha
Shift 2 3.16 ha

Area in acres
Shift 3 7.82 ac.

APPENDIX II

LANDSATS 1, 2 and 3

Launch:
Landsat 1: 23 July 1972
operation ended: 6 January 1978
Landsat 2: 22 January 1975
operation ended: 25 February 1982
Landsat 3: 5 March 1978
standby mode: 31 March 1983

Orbital parameters:
Orbit: near polar
sun synchronous
Altitude: 919 km
Inclination: 99.09°.
Coverage: 82°N to 82°S.
Period: 103 minutes, crossing the equator at 9.30hrs local time.
Repeat cycle: 18 days.

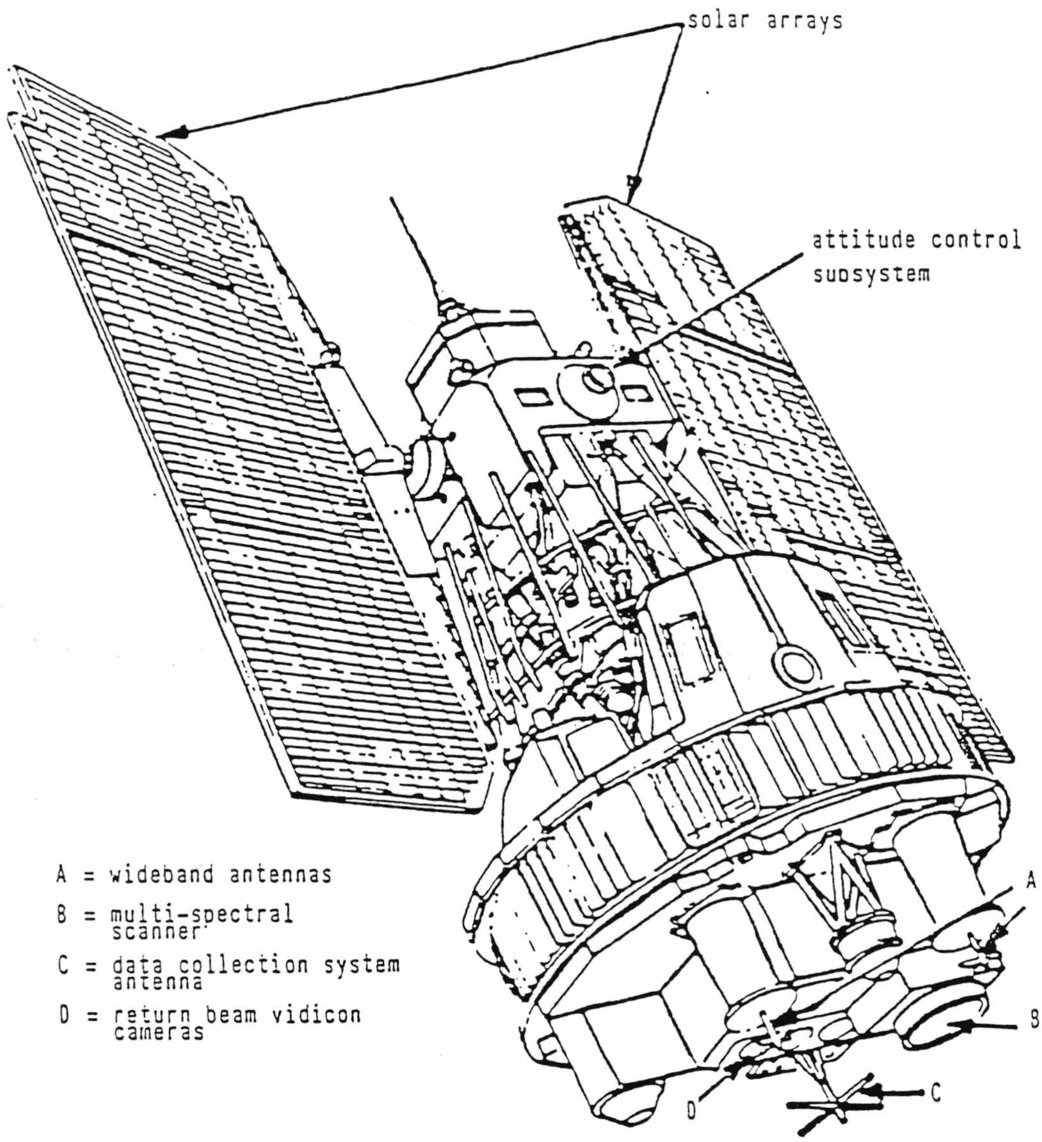
Satellites and sensors

Landsat-1 was the first satellite designed to collect data about the Earth's surface and resources. Three satellites, initially called Earth Resources Technology Satellites (ERTS), launched by NASA between 1972 and 1978 comprised the first generation of the Landsat series.

The satellites' payload consisted of two sensors:

- return beam vidicon (RBV) cameras
- a four-band multi-spectral scanner system (MSS)

The data were either transmitted directly to specially equipped ground stations or stored on tape recorders for later transmission to data centres in the U.S.A. The Landsat receiving stations covering the U.K. and Europe are Fucino in Italy and Kiruna in Sweden, both are part of the Earthnet programme.



Return Beam Vidicon Camera (RBV)

	Wavelength (μm)	Resolution	Image format and comments
Landsats 1 and 2 – three RBVs:			Simultaneous view from three cameras of a scene
Band 1	0.475-0.575 (blue–green)	80m	185 x 185km with 14% side
Band 2	0.580-0.680 (yellow-red)	80m	overlap at the equator and
Band 3	0.690-0.830 (red-near IR)	80m	10% forward overlap.
Landsat 3 – two RBVs, single band:			Two side-by-side images,
	0.505-0.750 (visible to near IR – panchromatic)	40m	98 x 98km, four RBV images approximately coinciding with scene.

Multi-spectral Scanner (MSS)

	Wavelength (μm)	Resolution	Image format and comments
Landsats 1, 2 and 3:			
Band 4	0.50-0.60 (green)	80m	185 x 185km images with 10%
Band 5	0.60-0.70 (red)	80m	forward overlap and 14% side
Band 6	0.70-0.80 (red-near IR)	80m	overlap at the equator,
Band 7	0.80-1.10 (near IR)	80m	increasing towards the poles.
Landsat 3 only:			Range of thermal sensitivity:
Band 8	10.40-12.50 (thermal IR)	120m	Only a few scenes available of limited area.

LANDSATS 4 and 5

Launch:
Landsat 4: 16 July 1982
Landsat 5: 1 March 1984

Orbital parameters:
Orbit: near polar sun synchronous
Altitude: 705 km
Inclination: 98.2°.
Coverage: 81°N to 81°S.
Period: 99 minutes, crossing the equator at 9.45hrs local time.
Repeat cycle: 16 days.

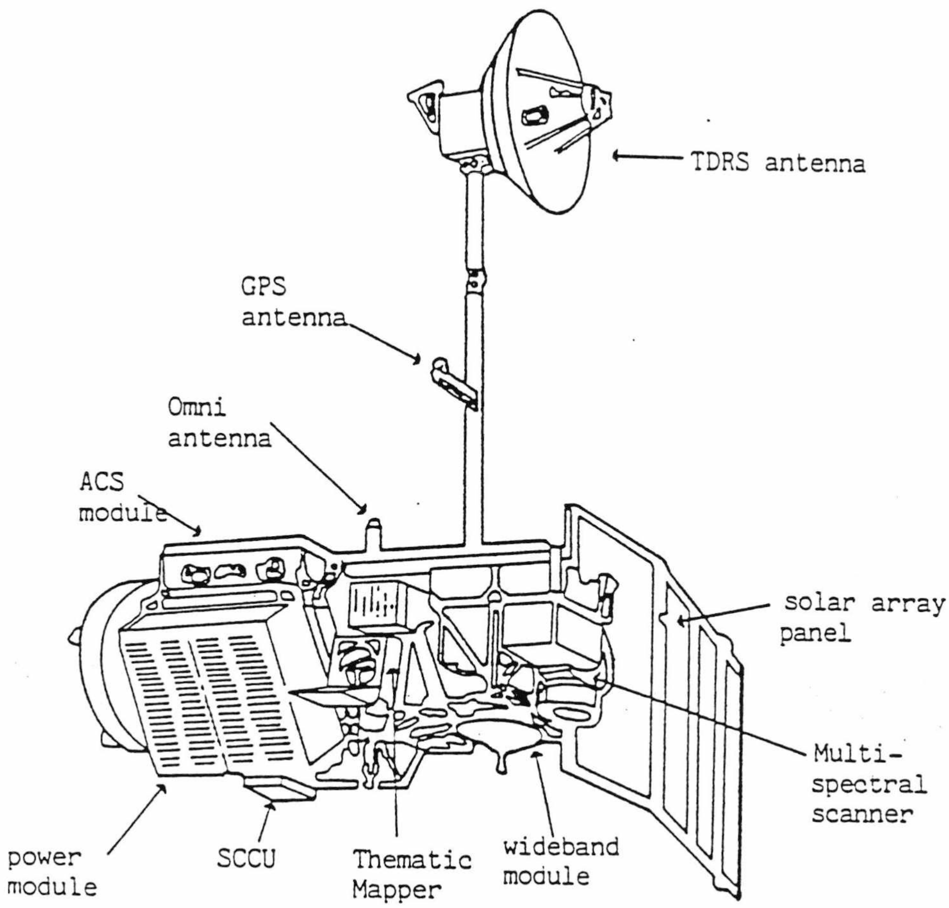
Satellites and sensors

The second generation of NASA's Landsat series was initiated in July 1982 with the launch of Landsat 4. In addition to the conventional multi-spectral scanner system (MSS – four wavebands and 80m ground resolution) familiar to all users of the earlier Landsats, the payload included the Thematic Mapper (TM), which recorded data in seven wavebands with a resolution of 30m in the visible, near and middle infra-red wavebands and 120m in thermal infra-red. Unfortunately power supply problems terminated the reception of TM data in February 1983, although MSS acquisition continues. However, Landsat 5 was launched on 1 March 1984 and both MSS and TM sensors, are operational.

Multi-Spectral Scanner

The MSS on Landsats 4 and 5 are similar to those flown on earlier Landsat missions. The ground pixel resolution is 80 by 80m. However, the four spectral bands are identified by a new numbering system – although the spectral coverage remains unchanged.

Landsats 1, 2 and 3	Landsats 4 and 5
Band 4 →	Band 1
Band 5 →	Band 2
Band 6 →	Band 3
Band 7 →	Band 4



Multi-Spectral Scanner (MSS)

	Wavelength (μm)	Resolution
Band 1	0.50-0.60 (green)	80m
Band 2	0.60-0.70 (red)	80m
Band 3	0.70-0.80 (red-near IR)	80m
Band 4	0.80-1.10 (near IR)	80m

Thematic Mapper

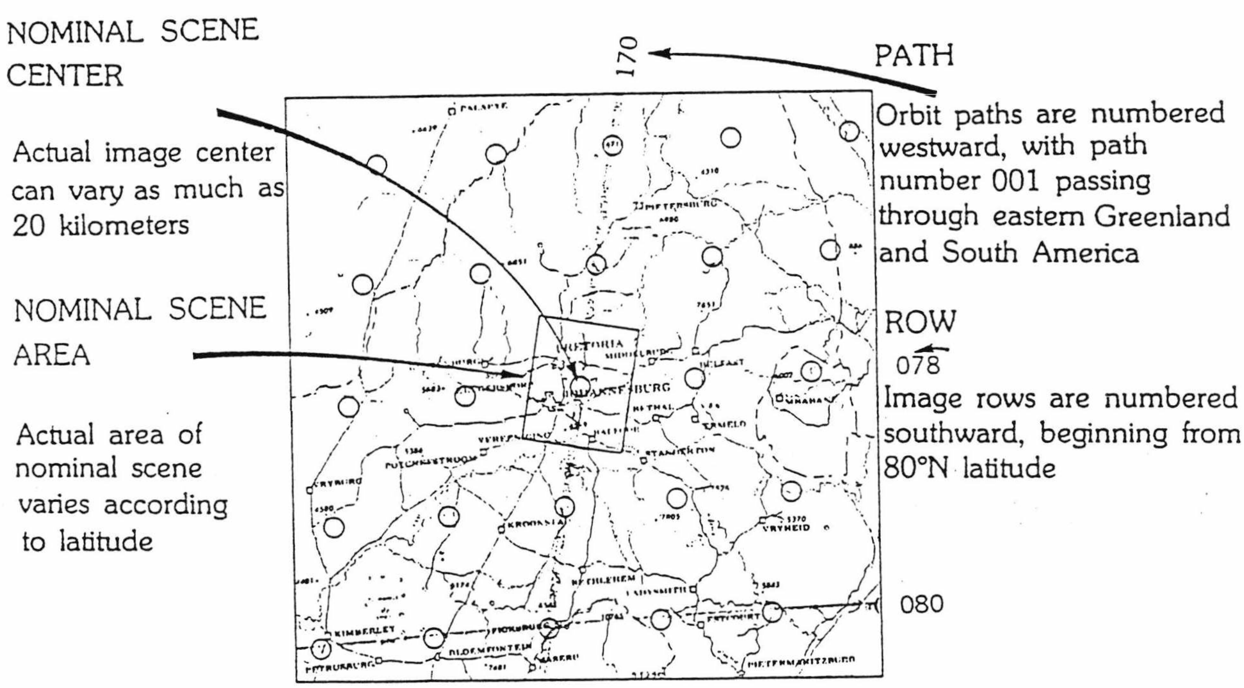
	Wavelength (μm)	Resolution
Band 1	0.45 - 0.52	30m
Band 2	0.52 - 0.60	30m
Band 3	0.63 - 0.69	30m
Band 4	0.76 - 0.90	30m
Band 5	1.55 - 1.75	30m
Band 6	10.40 - 12.50	120m
Band 7	2.08 - 2.35	30m

Both sensors provide image data with 185 x 185km coverage, with 5.4% forward overlap and 7.3% side overlap at the equator increasing at higher latitudes.

WORLDWIDE REFERENCE SYSTEM (WRS)

The Worldwide Reference System (WRS) is an extension of a reference system developed by the Canadian Centre for Remote Sensing for catalogue Landsat imagery. The WRS consists of a global network of 251 paths and 119 rows for Landsat 1, 2 and 3; a network of 233 paths for Landsat 4 and 5. The path and row intersections correspond to geographic locations over which Landsat scenes are generally centred. These locations are identified by three-digit path and row numbers and, when combined, identify a nominal scene centre.

Care should be exercised when ordering Landsat data to ensure that the appropriate WRS paths/rows are used; that is, WRS Landsat 1, 2 and 3 or WRS Landsat 4 and 5 must be designated.



Example: WRS Landsat 4, 170–078 (path 170, row 078) identifies the nominal scene center covering Pretoria, Republic of South Africa.

APPENDIX III

SPOT

The SPOT (Systeme Probatoire d'Observation de la Terre) satellite has been designed by the French Centre National d'Etudes Spatiales (CNES) and built by French, Belgian and Swedish partners.

The payload of SPOT-1 consists of two identical high resolution visible (HRV) imaging instruments and a package comprising two magnetic tape data recorders and a telemetry transmeter.

The HRV sensors are designed to operate in either of two modes - panchromatic (black and white) or multispectral (colour) - in the visible and near infrared portions of the spectrum.

The capabilities include:-

nadir viewing (vertical) providing a swath width of 117 km
off-nadir viewing (oblique) up to 27 degrees from the vertical
revisit possibilities to increase the frequency of imaging individual sites
stereoscopic imaging

SPOT is the first satellite to use pushbroom sensors. These consist essentially of CCD (charged-coupled device) linear arrays. This approach avoids the problems associated with mechanical means of moving a scanning mirror. With the pushbroom device, the scanning of an individual line of a scene is performed electronically by successively measuring the current generated by each detector within the linear array. Each spectral band uses four linear arrays, each consisting of 1728 elementary detectors (CCD devices).

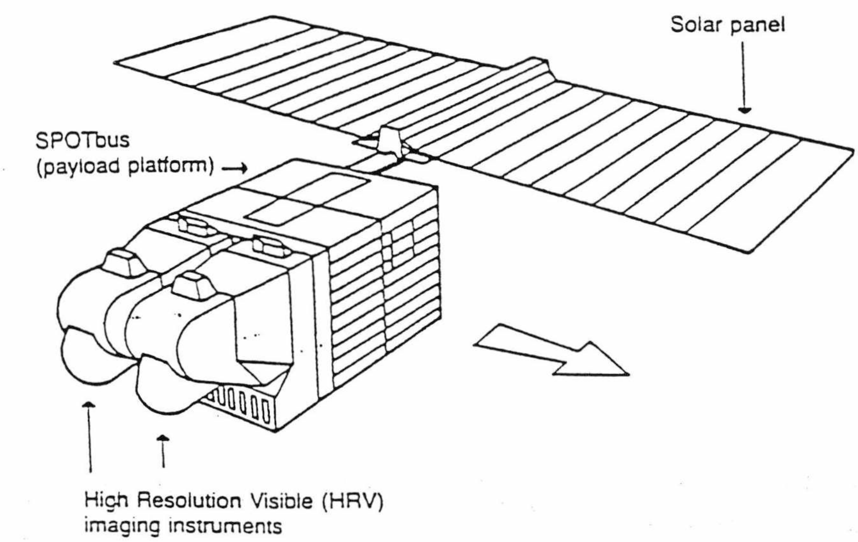
The sun-synchronous orbit ensures the satellite always passes overhead at the same (solar) time. In its normal orbit, SPOT crosses the equator, going south, at about 10.30 a.m. and at that time precisely on 15 June of each year.

Each HRV instrument can be pointed so that two adjacent strips of the Earth's surface are covered, giving a total swath width of 117 km (nadir) and an overlap of 3 km. Since the distance between adjacent ground tracks at the equator is about 108 km, complete Earth coverage can be obtained with this fixed setting of instrument fields.

It will be possible to steer the morrors of the HRV instruments to view obliquely (off-nadir) up to 27 degrees either side of vertical to cover an area of interest within a 950 km wide strip centred on the satellite ground track.

The width of the observed swath varies between 60 km for nadir viewing and 80 km for extreme off-nadir viewing.

The programme of planned observations is controlled by the satellite's onboard computer. A sequence of recorded images may include both modes of operation (panchromatic and multispectral) and changes in each instrument's viewing directions.



Launch: 22 February 1986

Orbital parameters:

orbit: near polar, sun-synchronous
altitude: 832km
inclination: 98.7°
timing: crossing equator at 10.30 local sun time (from N to S)
repeat cycle: 26 days

Sensors: High Resolution Visible (HRV)

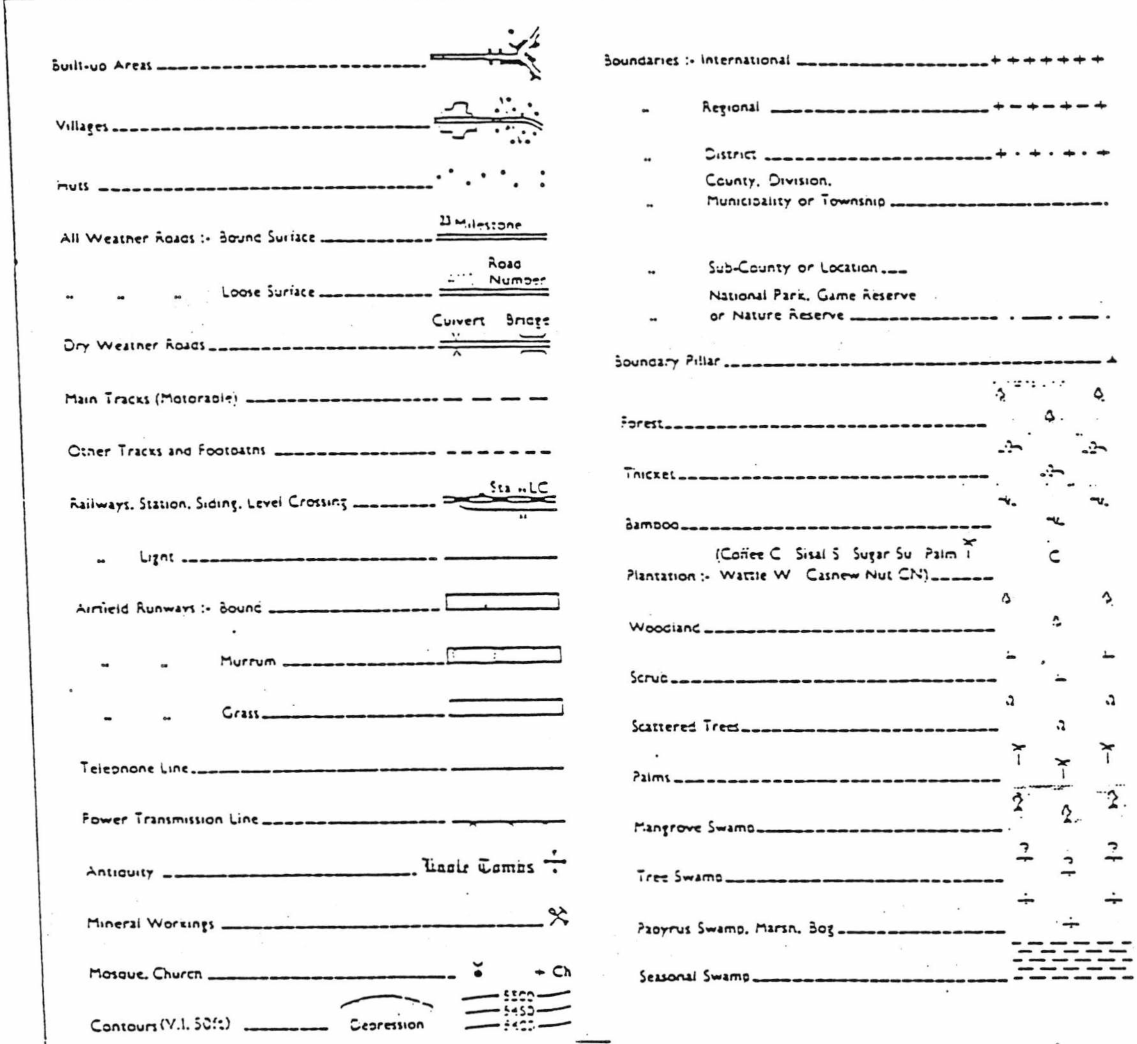
	wavelength (μm)	resolution (m)
Multispectral mode:		
Band 1	0.50-0.59	20
Band 2	0.61-0.68	20
Band 3	0.79-0.89	20
pancromatic mode:	0.51-0.73	10

APPENDIX IV

KEY FOR TOPOGRAPHICAL MAPS

Key for general features for all topographical maps

(chapter 4, section 4.8.1)



ABBREVIATIONS

AC	Area Commissioner	FR	Forest Reserve	PWD	Public Works Division
		Hosp	Hospital	RC	Regional Commissioner
Ch	Church	LGHQ	Local Government Headquarters	RH	Rest House
CHQ	County Headquarters	LHQ	Location Headquarters	Sch	School
Ct Ho	Court House	Mkt	Market	SCHQ	Sub-County Headquarters
DC	District Commissioner	MoW	Ministry of Works	T	Telephone
DHQ	Division Headquarters	PO	Post Office	Tg	Telegraph
Disp	Dispensary	PP	Police Post	TC	Trading Centre

APPENDIX V

THE ASYMPTOTIC BIAS OF t_{MOD}

A detailed derivation of the asymptotic bias of t_{MOD} (chapter 5, section 5.3).

From the following expression of t_{MOD}

$$t_{MOD} = r\left\{1 + \eta(C_{xy} - C_x^2)\right\}$$

which can also be expressed as

$$= r\left\{1 + \eta\frac{S_{xy}}{\bar{X}\bar{Y}} - \eta\frac{S_x^2}{\bar{X}^2}\right\}.$$

We have, from (5.7)

$$r = R\left\{1 + \delta_{\bar{y}} - \delta_{\bar{x}} + \delta_{\bar{x}}^2 - \delta_{\bar{x}}\delta_{\bar{y}} + \cdots\right\}.$$

Hence we can re-write the expression of t_{MOD} as

$$\begin{aligned} t_{MOD} &= R\left\{1 + \delta_{\bar{y}} - \delta_{\bar{x}} + \delta_{\bar{x}}^2 - \delta_{\bar{x}}\delta_{\bar{y}} + \cdots\right\}\left\{1 + \eta\frac{S_{xy}}{\bar{X}\bar{Y}} - \eta\frac{S_x^2}{\bar{X}^2}\right\} \\ &= R\{1 + \eta\frac{S_{xy}}{\bar{X}\bar{Y}} - \eta\frac{S_x^2}{\bar{X}^2} + \delta_{\bar{y}} + \eta\frac{S_{xy}}{\bar{X}\bar{Y}}\delta_{\bar{y}} - \eta\frac{S_x^2}{\bar{X}^2}\delta_{\bar{y}} \\ &\quad - \delta_{\bar{x}} - \eta\frac{S_{xy}}{\bar{X}\bar{Y}}\delta_{\bar{x}} + \eta\frac{S_x^2}{\bar{X}^2}\delta_{\bar{x}} + \delta_{\bar{x}}^2 + \eta\frac{S_{xy}}{\bar{X}\bar{Y}}\delta_{\bar{x}}^2 - \eta\frac{S_x^2}{\bar{X}^2}\delta_{\bar{x}}^2 \\ &\quad - \delta_{\bar{x}}\delta_{\bar{y}} - \eta\frac{S_{xy}}{\bar{X}\bar{Y}}\delta_{\bar{x}}\delta_{\bar{y}} + \eta\frac{S_x^2}{\bar{X}^2}\delta_{\bar{x}}\delta_{\bar{y}} + \textit{terms of higher power}\}. \end{aligned}$$

Taking expectation on both sides will make some of the terms in the above equation to have orders of n^{-2}, and by ignoring them we will have the following expression

$$\begin{aligned} E(t_{MOD}) &= R\left\{1 + \eta\frac{S_{xy}}{\bar{X}\bar{Y}} - \eta\frac{S_x^2}{\bar{X}^2} + \frac{S_x^2}{n\bar{X}^2} - \frac{S_{xy}}{n\bar{X}\bar{Y}} + O(n^{-2})\right\} \\ &= R\left\{1 + \eta C_{xy} - \eta C_x^2 + \frac{C_x^2}{n} - \frac{C_{xy}}{n} + O(n^{-2})\right\} \\ &= R\left\{1 + \frac{C_x^2}{n} - \frac{C_{xy}}{n} + \eta(C_{xy} - C_x^2) + O(n^{-2})\right\} \end{aligned}$$

$$= R\left\{1 + \frac{1}{n}(C_x^2 - C_{xy}) - \eta(C_x^2 - C_{xy}) + O(n^{-2})\right\}$$

$$= R\left\{1 + (\frac{1}{n} - \eta)(C_x^2 - C_{xy}) + O(n^{-2})\right\},$$

since $\eta = \left[\frac{1}{n} - \frac{1}{N}\right]$, the expectation of t_{MOD} is therefore

$$E(t_{MOD}) = R\left\{1 + \frac{1}{N}(C_x^2 - C_{xy}) + O(n^{-2})\right\}.$$

Finally the bias of t_{MOD} is

$$Bias(t_{MOD}) = R\left\{\frac{1}{N}(C_x^2 - C_{xy}) + O(n^{-2})\right\}.$$

APPENDIX VI

THE ASYMPTOTIC VARIANCE OF $\bar{y}_{MOD}$

The derivation of the asymptotic variance of $\bar{y}_{MOD}$ (chapter 5, section 5.4).

$$\begin{aligned}
\bar{y}_{MOD} &= (1-W)\bar{y} + W\, t_{MOD}\bar{X} \\
&= (1-W)r\bar{x} + W\bar{X}r\{1+\eta(C_{xy}-C_x^2)\} \\
&= r\bar{x} - Wr\bar{x} + W\bar{X}r + W\bar{X}\eta rC_{xy} - W\bar{X}\eta rC_x^2 \\
&= r\bar{x} - Wr(\bar{x}-\bar{X}) + W\bar{X}\eta rC_{xy} - W\bar{X}\eta rC_x^2 \\
&= \bar{y} - \frac{W\bar{y}(\bar{x}-\bar{X})}{\bar{x}} + \frac{W\eta\bar{y}S_{xy}}{\bar{x}\bar{Y}} - \frac{W\eta\bar{y}S_x^2}{\bar{x}\bar{X}} \\
&= \bar{Y}(1+\delta_{\bar{y}}) - W\bar{Y}(1+\delta_{\bar{y}})\delta_{\bar{x}}(1+\delta_{\bar{x}})^{-1} + \frac{W\eta\bar{Y}(1+\delta_{\bar{y}})(1+\delta_{\bar{x}})^{-1}S_{xy}}{\bar{X}\bar{Y}} \\
&\quad - \frac{W\eta\bar{Y}(1+\delta_{\bar{y}})(1+\delta_{\bar{x}})^{-1}S_x^2}{\bar{X}^2}.
\end{aligned}$$

Now

$$\begin{aligned}
(1+\delta_{\bar{y}})\,\delta_{\bar{x}}\,(1+\delta_{\bar{x}})^{-1} &= (\delta_{\bar{x}} + \delta_{\bar{x}}\,\delta_{\bar{y}})(1-\delta_{\bar{x}}+\delta_{\bar{x}}^2-\delta_{\bar{x}}^3+\ldots..) \\
&= \delta_{\bar{x}} + \delta_{\bar{x}}\,\delta_{\bar{y}} - \delta_{\bar{x}}^2 + \ldots...
\end{aligned}$$

Also

$$\begin{aligned}
(1+\delta_{\bar{y}})(1+\delta_{\bar{x}})^{-1} &= (1+\delta_{\bar{y}})(1-\delta_{\bar{x}}+\delta_{\bar{x}}^2-\delta_{\bar{x}}^3+\ldots..) \\
&= 1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2+\ldots...
\end{aligned}$$

Therefore $\bar{y}_{MOD}$ can be rewritten as

$$\begin{aligned}
\bar{y}_{MOD} &= \bar{Y}(1+\delta_{\bar{y}}) - W\bar{Y}(\delta_{\bar{x}}+\delta_{\bar{x}}\delta_{\bar{y}}-\delta_{\bar{x}}^2) + \frac{W\eta\bar{Y}S_{xy}}{\bar{X}\bar{Y}}(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2) \\
&\quad - \frac{W\eta\bar{Y}S_x^2}{\bar{X}^2}(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2) + \textit{terms of higher power.}
\end{aligned}$$

Squaring the above equation on both sides gives us

$$\begin{aligned}
\bar{y}_{MOD}^2 &= \bar{Y}^2(1+\delta_{\bar{y}})^2 + \\
&\quad + \left[\frac{W\eta S_{xy}}{\bar{X}}\right]^2(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2)^2
\end{aligned}$$

$$+\left[\frac{W\eta\bar{Y}S_x^2}{\bar{X}^2}\right]^2(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2)^2$$

$$-2W\bar{Y}^2(1+\delta_{\bar{y}})(\delta_{\bar{x}}+\delta_{\bar{x}}\delta_{\bar{y}}-\delta_{\bar{x}}^2)$$

$$+\frac{2W\eta\bar{Y}S_{xy}}{\bar{X}}(1+\delta_{\bar{y}})(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2)$$

$$-\frac{2W\eta\bar{Y}^2S_x^2}{\bar{X}^2}(1+\delta_{\bar{y}})(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2)$$

$$-\frac{2W^2\eta\,\bar{Y}S_{xy}}{\bar{X}}(\delta_{\bar{x}}+\delta_{\bar{x}}\,\delta_{\bar{y}}-\delta_{\bar{x}}^2)(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\,\delta_{\bar{y}}+\delta_{\bar{x}}^2)$$

$$+\frac{2W^2\eta\,\bar{Y}^2S_x^2}{\bar{X}^2}(\delta_{\bar{x}}+\delta_{\bar{x}}\,\delta_{\bar{y}}-\delta_{\bar{x}}^2)(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\,\delta_{\bar{y}}+\delta_{\bar{x}}^2)$$

$$-\frac{2W^2\eta^2\,\bar{Y}S_{xy}\,S_x^2}{\bar{X}^3}(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\,\delta_{\bar{y}}+\delta_{\bar{x}}^2)^2+\textit{terms of higher power.}$$

Now

$$\begin{aligned}(\delta_{\bar{x}}+\delta_{\bar{x}}\,\delta_{\bar{y}}-\delta_{\bar{x}}^2)^2&=\delta_{\bar{x}}^2+(\delta_{\bar{x}}\,\delta_{\bar{y}})^2+(\delta_{\bar{x}}^2)^2+2\delta_{\bar{x}}(\delta_{\bar{x}}\,\delta_{\bar{y}})-2\delta_{\bar{x}}(\delta_{\bar{x}}^2)-2\delta_{\bar{x}}\,\delta_{\bar{y}}(\delta_{\bar{x}}^2)\\&=\delta_{\bar{x}}^2+\ldots\ldots,\end{aligned}$$

$$\begin{aligned}(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\,\delta_{\bar{y}}+\delta_{\bar{x}}^2)^2&=1+\delta_{\bar{x}}^2+\delta_{\bar{y}}^2-(\delta_{\bar{x}}\,\delta_{\bar{y}})^2+(\delta_{\bar{x}}^2)^2-2\delta_{\bar{x}}+2\delta_{\bar{y}}-2\delta_{\bar{x}}\,\delta_{\bar{y}}\\&\quad+2\delta_{\bar{x}}^2-2\delta_{\bar{x}}\,\delta_{\bar{y}}+2\delta_{\bar{x}}(\delta_{\bar{x}}\,\delta_{\bar{y}})-2\delta_{\bar{x}}(\delta_{\bar{x}}^2)-2\delta_{\bar{y}}(\delta_{\bar{x}}\,\delta_{\bar{y}})+2\delta_{\bar{y}}(\delta_{\bar{x}}^2)-2\delta_{\bar{x}}\,\delta_{\bar{y}}(\delta_{\bar{x}}^2)\\&=1-2\delta_{\bar{x}}+2\delta_{\bar{y}}-4\delta_{\bar{x}}\,\delta_{\bar{y}}+3\delta_{\bar{x}}^2+\delta_{\bar{y}}^2+\ldots..,\end{aligned}$$

$$\begin{aligned}(1+\delta_{\bar{y}})(\delta_{\bar{x}}+\delta_{\bar{x}}\delta_{\bar{y}}-\delta_{\bar{x}}^2)&=\delta_{\bar{x}}+\delta_{\bar{x}}\delta_{\bar{y}}-\delta_{\bar{x}}^2+\delta_{\bar{x}}\delta_{\bar{y}}+\ldots..\\&=\delta_{\bar{x}}+2\delta_{\bar{x}}\delta_{\bar{y}}-\delta_{\bar{x}}^2+\ldots..,\end{aligned}$$

$$\begin{aligned}(1+\delta_{\bar{y}})(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2)&=1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{y}}^2+\ldots..\\&=1-\delta_{\bar{x}}+2\delta_{\bar{y}}-2\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2+\delta_{\bar{y}}^2+\ldots..\end{aligned}$$

and

$$\begin{aligned}(\delta_{\bar{x}}+\delta_{\bar{x}}\,\delta_{\bar{y}}-\delta_{\bar{x}}^2)(1-\delta_{\bar{x}}+\delta_{\bar{y}}-\delta_{\bar{x}}\delta_{\bar{y}}+\delta_{\bar{x}}^2)&=\delta_{\bar{x}}+\delta_{\bar{x}}\delta_{\bar{y}}-\delta_{\bar{x}}^2-\delta_{\bar{x}}^2+\delta_{\bar{x}}\,\delta_{\bar{y}}+\ldots..\\&=\delta_{\bar{x}}+2\delta_{\bar{x}}\,\delta_{\bar{y}}-2\delta_{\bar{x}}^2+\ldots...\end{aligned}$$

Therefore

$$\begin{aligned}\bar{y}_{MOD}^2&=\bar{Y}^2(1+2\delta_{\bar{y}}+\delta_{\bar{y}}^2)-W^2\,\bar{Y}^2(\delta_{\bar{x}}^2)\\&\quad+\left[\frac{W\eta S_{xy}}{\bar{X}}\right]^2(1-2\delta_{\bar{x}}+2\delta_{\bar{y}}-4\delta_{\bar{x}}\delta_{\bar{y}}+3\delta_{\bar{x}}^2+\delta_{\bar{y}}^2)\end{aligned}$$

$$- \left[\frac{W\eta \bar{Y} S_x^2}{\bar{X}^2} \right]^2 (1 - 2\delta_{\bar{x}} + 2\delta_{\bar{y}} - 4\delta_{\bar{x}}\delta_{\bar{y}} + 3\delta_{\bar{x}}^2 + \delta_{\bar{y}}^2)$$

$$- 2W\bar{Y}^2 (\delta_{\bar{x}} + 2\delta_{\bar{x}}\delta_{\bar{y}} - \delta_{\bar{x}}^2)$$

$$+ \frac{2W\eta \bar{Y} S_{xy}}{\bar{X}} (1 - \delta_{\bar{x}} + 2\delta_{\bar{y}} - 2\delta_{\bar{x}}\delta_{\bar{y}} + \delta_{\bar{x}}^2 + \delta_{\bar{y}}^2)$$

$$- \frac{2W\eta \bar{Y}^2 S_x^2}{\bar{X}^2} (1 - \delta_{\bar{x}} + 2\delta_{\bar{y}} - 2\delta_{\bar{x}}\delta_{\bar{y}} + \delta_{\bar{x}}^2 + \delta_{\bar{y}}^2)$$

$$- \frac{2W^2\eta\, Y S_{xy}}{\bar{X}} (\delta_{\bar{x}} + 2\delta_{\bar{x}}\, \delta_{\bar{y}} - 2\delta_{\bar{x}}^2)$$

$$+ \frac{2W^2\eta\, \bar{Y}^2 S_x^2}{\bar{X}^2} (\delta_{\bar{x}} + 2\delta_{\bar{x}}\, \delta_{\bar{y}} - 2\delta_{\bar{x}}^2)$$

$$- \frac{2W^2\eta^2\, \bar{Y} S_{xy}\, S_x^2}{\bar{X}^3} + \textit{terms of higher power.}$$

Taking expectation on both sides will make some of the terms in the above equation to have orders of n^{-2}, and by ignoring them we will have the following expression

$$E(\bar{y}_{MOD}^2) = \bar{Y}^2 + \bar{Y}^2 \frac{S_y^2}{n\bar{Y}^2} - W^2 \bar{Y}^2 \frac{S_x^2}{n\bar{X}^2} - 4W\bar{Y}^2 \frac{S_{xy}}{n\bar{X}\,\bar{Y}}$$
$$+ 2W\bar{Y}^2 \frac{S_x^2}{n\bar{X}^2} + 2W\eta\bar{Y}^2 \frac{S_{xy}}{\bar{X}\,\bar{Y}} - 2W\eta\bar{Y}^2 \frac{S_x^2}{\bar{X}^2} + O(n^{-2})$$

which can be rewritten as

$$E(\bar{y}_{MOD}^2) = \bar{Y}^2 \{1 + \frac{C_y^2}{n} - \frac{W^2\, C_x^2}{n} - \frac{4W\, C_{xy}}{n} + 2W\,\eta\, C_{xy}$$
$$+ \frac{2W\, C_x^2}{n} - 2W\,\eta\, C_x^2 + O(n^{-2})\}.$$

Therefore the variance of $\bar{Y}_{MOD}$ is

$$V(\bar{y}_{MOD}) = E(\bar{y}_{MOD}^2) - \bar{Y}^2$$
$$= \bar{Y}^2 \{ \frac{C_y^2}{n} + \frac{W^2\, C_x^2}{n} - \frac{4W\, C_{xy}}{n}$$
$$+ 2W\,\eta\, C_{xy} + \frac{2W\, C_x^2}{n} - 2W\,\eta C_x^2 \}$$
$$= \frac{\bar{Y}^2\, C_y^2}{n} \{1 + W^2\, K^2 - 4W\, K\, \rho_{yx}$$
$$+ 2W\, Kn\, \eta\, \rho_{yx} + 2W\, K^2 - 2W\, K^2 n\, \eta\}$$

$$= \frac{S_y^2}{n}\{1 + WK[WK - 4\rho_{yx}$$

$$+ 2(n\,\eta\,\rho_{yx} + K - Kn\,\eta)]\}.$$

Since $n\eta = n(1/n - 1/N) = (1 - n/N)$ then

$$V(\bar{y}_{MOD}) = \frac{S_y^2}{n}\{1 + WK[WK - 4\rho_{yx} + 2(\rho_{yx} - \frac{n}{N}\rho_{yx} + \frac{n}{N}K)]\}$$

$$= \frac{S_y^2}{n}\{1 - WK[WK - 4\rho_{yx} + 2\rho_{yx} + \frac{2n}{N}(K - \rho_{yx})]\}$$

$$= \frac{S_y^2}{n}\{1 + WK[WK - 2\rho_{yx} + \frac{2n}{N}(K - \rho_{yx})]\}.$$

APPENDIX VII

THE BIAS OF $\bar{y}_{MOD}$ (EXACT THEORY)

The derivation of the bias of $\bar{y}_{MOD}$ from the exact theory (chapter 5, section 5.5.1).

$$\bar{y}_{MOD} = (1-W)\bar{y} + W\bar{X}t_{MOD}$$

$$= (1-W)\bar{y} + W\bar{X}r\left\{1 + \frac{1}{n}\left(\frac{s_{xy}}{\bar{x}\bar{y}} - \frac{s_x^2}{\bar{x}^2}\right)\right\}$$

$$= (1-W)(\alpha + \beta\bar{x} + \bar{u}) + W\bar{X}\left(\frac{\alpha}{\bar{x}} + \beta + \frac{\bar{u}}{\bar{x}}\right)$$

$$+ \frac{W\bar{X}s_{xy}}{n\bar{x}^2} - \frac{W\bar{X}s_x^2}{n\bar{x}^2}\left(\frac{\alpha}{\bar{x}} + \beta + \frac{\bar{u}}{\bar{x}}\right)$$

$$= (1-W)(\alpha + \beta\bar{x} + \bar{u}) + W\bar{X}\left(\frac{\alpha}{\bar{x}} + \beta + \frac{\bar{u}}{\bar{x}}\right) + \frac{W\bar{X}s_{xy}}{n\bar{x}^2}$$

$$- \frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3} - \frac{W\bar{X}\beta s_x^2}{n\bar{x}^2} - \frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}.$$

But $s_{xy} = \beta s_x^2 + s_{ux}$, therefore

$$\bar{y}_{MOD} = (1-W)(\alpha + \beta\bar{x} + \bar{u}) + W\bar{X}\left(\frac{\alpha}{\bar{x}} + \beta + \frac{\bar{u}}{\bar{x}}\right)$$

$$+ \frac{W\bar{X}s_{ux}}{n\bar{x}^2} - \frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3} - \frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}$$

$$E(\bar{y}_{MOD}) = (1-W)\alpha + (1-W)\beta m + \frac{Wm\alpha}{m-1} + W\beta m$$

$$- \frac{Wm\alpha}{n}E\left(\frac{s_x^2}{\bar{x}^3}\right)$$

$$= \alpha - W\alpha + \beta m + \frac{Wm\alpha}{m-1} - \frac{Wm\alpha}{n}E\left(\frac{s_x^2}{\bar{x}^3}\right)$$

and so

$$Bias(\bar{y}_{MOD}) = -W\alpha + \frac{Wm\alpha}{m-1} - \frac{Wm\alpha}{n}E\left(\frac{s_x^2}{\bar{x}^3}\right)$$

$$= W\alpha\left\{\frac{m}{m-1} - 1 - \frac{m}{n}E\left(\frac{s_x^2}{\bar{x}^3}\right)\right\}.$$

From Rao and Webster (1966), $E(\frac{s_x^2}{\bar{x}^3}) = \frac{n}{(m-1)(m+1)}$. Therefore

$$
\begin{aligned}
Bias(\bar{y}_{MOD}) &= W\,\alpha\{\frac{m}{m-1} - 1 - \frac{mn}{n(m-1)(m+1)}\} \\
&= W\,\alpha\{\frac{m^2 + m - m^2 + 1 - m}{(m-1)(m+1)}\} \\
&= \frac{W\,\alpha}{(m-1)(m+1)}.
\end{aligned}
$$

APPENDIX VIII

THE VARIANCE OF $\bar{y}_{MOD}$ (EXACT THEORY)

The derivation of the variance of $\bar{y}_{MOD}$ from the exact theory (chapter 5, section 5.5.2).

$$\begin{aligned}
\bar{y}_{MOD} &= (1-W)\bar{y} + W\bar{X}t_{MOD} \\
&= (1-W)\bar{y} + W\bar{X}r\left\{1+\frac{1}{n}(C_{xy}-C_x^2)\right\} \\
&= (1-W)(\alpha+\beta\bar{x}+\bar{u}) + W\bar{X}\left[\frac{\alpha}{\bar{x}}+\beta+\frac{\bar{u}}{\bar{x}}\right]\left\{1+\frac{1}{n}(C_{xy}-C_x^2)\right\} \\
&= (1-W)\alpha+(1-W)\beta\bar{x}+(1-W)\bar{u}+\frac{W\bar{X}\alpha}{\bar{x}}+W\bar{X}\beta+\frac{W\bar{X}\bar{u}}{\bar{x}} \\
&\quad +\frac{W\bar{X}C_{yx}}{n}\left\{\frac{\alpha}{\bar{x}}+\beta+\frac{\bar{u}}{\bar{x}}\right\}-\frac{W\bar{X}C_x^2}{n}\left\{\frac{\alpha}{\bar{x}}+\beta+\frac{\bar{u}}{\bar{x}}\right\} \\
&= (1-W)\alpha+(1-W)\beta\bar{x}+(1-W)\bar{u}+\frac{W\bar{X}\alpha}{\bar{x}}+W\bar{X}\beta+\frac{W\bar{X}\bar{u}}{\bar{x}} \\
&\quad +\frac{W\bar{X}s_{xy}}{n\bar{x}^2}-\frac{W\bar{X}s_x^2}{n\bar{x}^2}\left\{\frac{\alpha}{\bar{x}}+\beta+\frac{\bar{u}}{\bar{x}}\right\}
\end{aligned}$$

Now since $s_{xy}=\beta s_x^2+s_{ux}$, then

$$\begin{aligned}
\bar{y}_{MOD} &= (1-W)\alpha+(1-W)\beta\bar{x}+(1-W)\bar{u}+\frac{W\bar{X}\alpha}{\bar{x}}+W\bar{X}\beta+\frac{W\bar{X}\bar{u}}{\bar{x}} \\
&\quad +\frac{W\bar{X}\beta s_x^2}{n\bar{x}^2}+\frac{W\bar{X}s_{ux}}{n\bar{x}^2}-\frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3}-\frac{W\bar{X}\beta s_x^2}{n\bar{x}^2}-\frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}.
\end{aligned}$$

Let

$$\bar{y}_{MOD'} = \bar{y}_{MOD}-(1-W)\alpha-W\bar{X}\beta.$$

Therefore

$$\begin{aligned}
\bar{y}_{MOD'} &= (1-W)\beta\bar{x}+(1-W)\bar{u}+\frac{W\bar{X}\alpha}{\bar{x}}+\frac{W\bar{X}\bar{u}}{\bar{x}} \\
&\quad +\frac{W\bar{X}s_{ux}}{n\bar{x}^2}-\frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3}-\frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}. \qquad (8.1)
\end{aligned}$$

Hence

$$E(\bar{y}_{MOD'}) = (1-W)\,\beta m + \frac{W m \alpha}{(m-1)} - \frac{W m \alpha}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right]$$

and

$$\left\{E(\bar{y}_{MOD'})\right\}^2 = ((1-W)\,\beta m)^2 + \frac{(W m \alpha)^2}{(m-1)^2} + \frac{(W m \alpha)^2}{n^2}\left\{E\left[\frac{s_x^2}{\bar{x}^3}\right]\right\}^2$$
$$+ 2(1-W)\,\beta m \frac{W m \alpha}{(m-1)} - 2(1-W)\beta m \frac{W m \alpha}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right]$$
$$- 2\frac{(W m \alpha)}{(m-1)}\frac{(W m \alpha)}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right].$$

Also from (8.1) we have

$$(\bar{y}_{MOD'})^2 = ((1-W)\,\beta\bar{x})^2 + ((1-W)\,\bar{u})^2 + \frac{(W\bar{X}\alpha)^2}{\bar{x}^2} + \frac{(W\bar{X}\bar{u})^2}{\bar{x}^2}$$
$$+ \frac{(W\bar{X}s_{ux})^2}{n^2\bar{x}^4} + \frac{(W\bar{X}\alpha)^2 s_x^4}{n^2\bar{x}^6} + \frac{(W\bar{X}\bar{u})^2 s_x^4}{n^2\bar{x}^6}$$
$$+ 2(1-W)\,\beta\bar{x}(1-W)\,\bar{u} + 2(1-W)\,\beta\bar{x}\frac{W\bar{X}\alpha}{\bar{x}}$$
$$+ 2(1-W)\,\beta\bar{x}\frac{W\bar{X}\bar{u}}{\bar{x}} + 2(1-W)\,\beta\bar{x}\frac{W\bar{X}s_{ux}}{n\bar{x}^2}$$
$$- 2(1-W)\,\beta\bar{x}\frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3} - 2(1-W)\,\beta\bar{x}\frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}$$
$$+ 2(1-W)\,\bar{u}\frac{W\bar{X}\alpha}{\bar{x}} + 2(1-W)\,\bar{u}\frac{W\bar{X}\bar{u}}{\bar{x}} + 2(1-W)\,\bar{u}\frac{W\bar{X}s_{ux}}{n\bar{x}^2}$$
$$- 2(1-W)\,\bar{u}\frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3} - 2(1-W)\,\bar{u}\frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3} + 2\frac{W\bar{X}\alpha}{\bar{x}}\frac{W\bar{X}\bar{u}}{\bar{x}}$$
$$+ 2\frac{W\bar{X}\alpha}{\bar{x}}\frac{W\bar{X}s_{ux}}{n\bar{x}^2} - 2\frac{W\bar{X}\alpha}{\bar{x}}\frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3} - 2\frac{W\bar{X}\alpha}{\bar{x}}\frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}$$
$$+ 2\frac{W\bar{X}\bar{u}}{\bar{x}}\frac{W\bar{X}s_{ux}}{n\bar{x}^2} - 2\frac{W\bar{X}\bar{u}}{\bar{x}}\frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3} - 2\frac{W\bar{X}\bar{u}}{\bar{x}}\frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}$$
$$- 2\frac{W\bar{X}s_{ux}}{n\bar{x}^2}\frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3} - 2\frac{W\bar{X}s_{ux}}{n\bar{x}^2}\frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}$$
$$+ 2\frac{W\bar{X}\alpha s_x^2}{n\bar{x}^3}\frac{W\bar{X}\bar{u}s_x^2}{n\bar{x}^3}.$$

So

$$E(\bar{y}_{MOD'})^2 = (1-W)^2\,\beta^2 m\,(m+1) + (1-W)^2\,\delta + \frac{(W\,m\,\alpha)^2}{(m-1)(m-2)}$$
$$+ \frac{(W\,m)^2\delta}{(m-1)(m-2)}$$
$$+ \left[\frac{W\,m\,\alpha}{n}\right]^2 E\left[\frac{s_x^4}{\bar{x}^6}\right] + \frac{(W\,m)^2\delta}{n^2} E\left[\frac{s_x^4}{\bar{x}^6}\right] + 2(1-W)\,W\,m\,\alpha\,\beta$$
$$- \frac{2(1-W)\,W\,m\,\alpha\,\beta}{n} E\left[\frac{s_x^2}{\bar{x}^2}\right] + \frac{2(1-W)\,W\,m\,\delta}{(m-1)}$$
$$- \frac{2(1-W)\,W\,m\,\delta}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right] - \frac{2(W\,m\,\alpha)^2}{n} E\left[\frac{s_x^2}{\bar{x}^4}\right]$$
$$- \frac{2(W\,m)^2\delta}{n} E\left[\frac{s_x^2}{\bar{x}^4}\right]$$
$$= ((1-W)\,\beta\,m)^2 + (1-W)^2\,\beta^2 m + \{(1-W)^2 + \frac{(W\,m)^2}{(m-1)(m-2)}$$
$$+ \left[\frac{W\,m}{n}\right]^2 E\left[\frac{s_x^4}{\bar{x}^6}\right] + \frac{2(1-W)\,W\,m}{(m-1)} - \frac{2(1-W)\,W\,m}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right]$$
$$- \frac{2(W\,m)^2}{n} E\left[\frac{s_x^2}{\bar{x}^4}\right]\}\,\delta$$
$$+ \left\{\frac{(W\,m)^2}{(m-1)(m-2)} + \left[\frac{W\,m}{n}\right]^2 E\left[\frac{s_x^4}{\bar{x}^6}\right] - \frac{2(W\,m)^2}{n} E\left[\frac{s_x^2}{\bar{x}^4}\right]\right\}\alpha^2$$
$$+ \left\{2(1-W)\,W\,m - \frac{2(1-W)\,W\,m}{n} E\left[\frac{s_x^2}{\bar{x}^2}\right]\right\}\alpha\,\beta.$$

Therefore

$$v(\bar{y}_{MOD'}) = v(\bar{y}_{MOD}) = E\left[\bar{y}_{MOD'}\right]^2 - \left\{E\left[\bar{y}_{MOD'}\right]\right\}^2$$
$$= (1-W)^2\,\beta^2\,m^2 + (1-W)^2\,\beta^2\,m + \{(1-W)^2 + \frac{(W\,m)^2}{(m-1)(m-2)}$$
$$+ \left[\frac{W\,m}{n}\right]^2 E\left[\frac{s_x^4}{\bar{x}^6}\right] + 2\frac{(1-W)\,W\,m}{(m-1)} - 2\frac{(1-W)\,W\,m}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right]$$
$$- 2\frac{(W\,m)^2}{n} E\left[\frac{s_x^2}{\bar{x}^4}\right]\}\,\delta + \{\frac{(W\,m)^2}{(m-1)(m-2)} + \left[\frac{W\,m}{n}\right]^2 E\left[\frac{s_x^4}{\bar{x}^6}\right]$$
$$- 2\frac{(W\,m)^2}{n} E\left[\frac{s_x^2}{\bar{x}^4}\right]\}\,\alpha^2 + \left\{2(1-W)\,W\,m - 2\frac{(1-W)\,W\,m}{n} E\left[\frac{s_x^2}{\bar{x}^2}\right]\right\}\alpha$$

$$- ((1-W)\beta m)^2 - \{\frac{(W m)^2}{(m-1)^2} + \left[\frac{W m}{n}\right]^2 \left\{E\left[\frac{s_x^2}{\bar{x}^3}\right]\right\}^2$$

$$- 2\frac{(W m)^2}{n(m-1)} E\left[\frac{s_x^2}{\bar{x}^3}\right]\} \alpha^2$$

$$- \left\{2\frac{(1-W) W m^2}{(m-1)} - 2\frac{(1-W) W m^2}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right]\right\} \alpha\beta$$

$$= (1-W)^2 \beta^2 m + \{\frac{(W m)^2}{(m-1)(m-2)} + \left[\frac{W m}{n}\right]^2 E\left[\frac{s_x^4}{\bar{x}^6}\right]$$

$$- 2\frac{(W m)^2}{n} E\left[\frac{s_x^2}{\bar{x}^4}\right]$$

$$- \frac{(W m)^2}{(m-1)^2} - \left[\frac{W m}{n}\right]^2 \left\{E\left[\frac{s_x^2}{\bar{x}^3}\right]\right\}^2 + 2\frac{(W m)^2}{n(m-1)} E\left[\frac{s_x^2}{\bar{x}^3}\right]\} \alpha^2$$

$$+ \{2(1-W) W m - 2\frac{(1-W) W m}{n} E\left[\frac{s_x^2}{\bar{x}^2}\right] - 2\frac{(1-W) W m^2}{(m-1)}$$

$$+ 2\frac{(1-W) W m^2}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right]\} \alpha\beta$$

$$+ \{(1-W)^2 + \frac{(W m)^2}{(m-1)(m-2)} + \left[\frac{W m}{n}\right]^2 E\left[\frac{s_x^4}{\bar{x}^6}\right] + 2\frac{(1-W) W m}{(m-1)}$$

$$- 2\frac{(1-W) W m}{n} E\left[\frac{s_x^2}{\bar{x}^3}\right] - 2\frac{(W m)^2}{n} E\left[\frac{s_x^2}{\bar{x}^4}\right]\} \delta.$$

Simplifying the coefficient for α^2

$$\{\frac{W^2 m^2}{(m-1)(m-2)} + \frac{W^2 m^2 n^2 \left\{6(n-1) + m(n+1)\right\}}{n^2(n-1)(m-1)(m-2)(m+1)(m+2)(m+3)}$$

$$- 2\frac{W^2 m^2}{n} \frac{n}{(m-1)(m-2)(m+1)}$$

$$- \frac{W^2 m^2}{(m-1)^2} - \frac{W^2 m^2}{n^2} \frac{n^2}{(m-1)^2(m+1)^2} + 2\frac{W^2 m^2}{n(m-1)} \frac{n}{(m-1)(m+1)}\}.$$

Leaving the second factor in the above expression, we have

$$W^2 m^2 \left\{\frac{(m-1)(m+1)^2 - 2(m-1)(m+1) - (m-2)(m+1)^2 - (m-2) + 2(m-2)(m+1)}{(m-1)^2(m-2)(m+1)^2}\right\}$$

$$= \frac{W^2 m^2}{(m-1)^2(m-2)(m+1)^2} \{(m-1)(m+1)^2 - 2(m^2-1)$$

$$- (m^2 + 2m + 1)(m-2) - (m-2) + 2(m^2 - m - 2)\}$$

$$= \frac{W^2 m^2}{(m-1)^2(m-2)(m+1)^2}\{m^3 + m^2 - m - 1$$
$$- 2m^2 + 2 - m^3 + 3m + 2 - m + 2 + 2m^2 - 2m - 4\}$$
$$= \frac{W^2 m^2}{(m-1)^2(m-2)(m+1)^2}\left\{m^2 - m + 1\right\}.$$

Combining this result with the second factor left out gives

$$\frac{W^2 m^2}{(m-1)(m-2)(m+1)}\left\{\frac{(m^2-m+1)}{(m-1)(m+1)} + \frac{6(n-1)+m(n+1)}{(n-1)(m+2)(m+3)}\right\}$$
$$= \frac{W^2 m^2}{(m-1)(m-2)(m+1)}\{\frac{(m^2-m+1)(m+2)(m+3)+6(m-1)(m+1)}{(m-1)(m+1)(m+2)(m+3)}$$
$$+ \frac{m(n+1)}{(n-1)(m+2)(m+3)}\}$$
$$= \frac{W^2 m^2}{(m-1)(m-2)(m+1)}\{\frac{(m^2-m+1)(m^2+5m+6)+6m^2-6}{(m-1)(m+1)(m+2)(m+3)}$$
$$+ \frac{m(n+1)}{(n-1)(m+2)(m+3)}\}$$
$$= \frac{W^2 m^2}{(m-1)(m-2)(m+1)}\{\frac{m^4-m^3+m^2+5m^3-5m^2+5m+6m^2-6m+6+6m^2-6}{(m-1)(m+1)(m+2)(m+3)}$$
$$+ \frac{m(n+1)}{(n-1)(m+2)(m+3)}\}$$
$$= \frac{W^2 m^2}{(m-1)(m-2)(m+1)}\left\{\frac{m(m^3+4m^2+8m-1)}{(m-1)(m+1)(m+2)(m+3)} + \frac{m(n+1)}{(n-1)(m+2)(m+3)}\right\}$$
$$= \frac{W^2 m^3}{(m-1)(m-2)(m+1)(m+2)(m+3)}\left\{\frac{(m^3+4m^2+8m-1)}{(m-1)(m+1)} + \frac{(n+1)}{(n-1)}\right\}.$$

Simplifying the coefficient for δ

$$\{(1-W)^2 + \frac{W^2 m^2}{(m-1)(m-2)} + \frac{W^2 m^2\{6(n-1)+m(n+1)\}}{(n-1)(m-1)(m-2)(m+1)(m+2)(m+3)}$$
$$+ 2\frac{(1-W)Wm}{(m-1)} - 2\frac{(1-W)Wm}{(m-1)(m+1)} - 2\frac{W^2 m^2}{(m-1)(m-2)(m+1)}\}.$$

Again leaving the third factor, we have

$$\{(1-W)^2 + \frac{W^2 m^2}{(m-1)(m-2)} + 2\frac{(1-W)Wm}{(m-1)} - 2\frac{(1-W)Wm}{(m-1)(m+1)}$$

$$-2\frac{W^2 m^2}{(m-1)(m-2)(m+1)}\}$$

$$= \left\{(1-W)^2 \frac{W\,m}{(m-1)}\left[\frac{W\,m}{(m-2)} + 2(1-W) - 2\frac{(1-W)}{(m+1)} - 2\frac{W\,m}{(m-2)(m+1)}\right]\right\}$$

$$= \{(1-W)^2 + \frac{W\,m}{(m-1)}$$

$$= \left\{(1-W)^2 + \frac{W\,m}{(m-1)}\left[\frac{W\,m\,(m-1)}{(m-2)(m+1)} + 2\frac{(1-W)\,m}{(m+1)}\right]\right\}$$

$$= \left\{(1-W)^2 + \frac{W^2 m^2}{(m-2)(m+1)} + 2\frac{(1-W)\,W\,m^2}{(m-1)(m+)}\right\}.$$

Combining with the third factor left out, we have

$$\{(1-W)^2 + \frac{W^2 m^2}{(m-2)(m+1)} + 2\frac{(1-W)\,W\,m^2}{(m-1)(m+1)}$$

$$+ \frac{W^2 m^2 \{6(n-1) + m(n+1)\}}{(n-1)(m-1)(m-2)(m+1)(m+2)(m+3)}\}$$

$$= (1-W)^2 + 2\frac{(1-W)\,W\,m^2}{(m-1)(m+1)} + \frac{W^2 m^2}{(m-2)(m+1)}\left[1 + \frac{6(n-1) + m(n+1)}{(n-1)(m-1)(m+2)(m+3)}\right]$$

$$= \{(1-W)^2 + 2\frac{(1-W)\,W\,m^2}{(m-1)(m+1)}$$

$$+ \frac{W^2 m^2}{(m-2)(m+1)}\left\{\frac{(m-1)(m+1)(m+3) + 6}{(m-1)(m+2)(m+3)} + \frac{m(n+1)}{(n-1)(m-1)(m+2)(m+3)}\right]\}$$

$$= \{(1-W)^2 + 2\frac{(1-W)\,W\,m^2}{(m-1)(m+1)}$$

$$+ \frac{W^2 m^2}{(m-1)(m+1)}\left\{\frac{m^3 + 5m^2 + 6m - m^2 - 5m - 6 + 6}{(m-1)(m+2)(m+3)} + \frac{m(n+1)}{(n-1)(m-1)(m+2)(m+3)}\right\}\}$$

$$= \{(1-W)^2 + 2\frac{(1-W)\,W\,m^2}{(m-1)(m+1)}$$

$$+ \frac{W^2 m^2}{(m-1)(m+1)}\left\{\frac{m^3 + 4m^2 + m}{(m-1)(m+2)(m+3)} + \frac{m(n+1)}{(n-1)(m-1)(m+2)(m+3)}\right\}$$

$$= \{(1-W)^2 + 2\frac{(1-W)\,W\,m^2}{(m-1)(m+1)} + \frac{W^2 m^3}{(m-1)^2(m+1)(m+2)(m+3)}\left\{(m^2 + 4m + 1) + \frac{(n+1)}{(n-1)}\right\}\}$$

Simplifying the coefficient for $\alpha\beta$

$$\left\{2(1-W)\,W\,m - 2\frac{(1-W)\,W\,m}{n}E\left[\frac{s_x^2}{\bar{x}^2}\right] - 2\frac{(1-W)\,W\,m^2}{(m-1)} + 2\frac{(1-W)\,W\,m^2}{n}E\left[\frac{s_x^2}{\bar{x}^3}\right]\right\},$$

using Theorem 3 of Rao and Webster (1966), we have

$$E\left[\frac{s_x^2}{\bar{x}^2}\right] = \frac{1}{(n-1)}\left\{n^3 E\left[\frac{z_i^2}{(\sum z_i)^2}\right] - n\right\} = \frac{n}{(m+1)}$$

and so the coefficient becomes

$$\left\{2(1-W)\,W\,m - 2\frac{(1-W)\,W\,m}{n}\frac{n}{(m+1)} - 2\frac{(1-W)\,W\,m^2}{(m-1)} + 2\frac{(1-W)\,W\,m^2}{n}\frac{n}{(m-1)(m+1)}\right\}$$

$$= 2(1-W)\,W\,m\left\{1 - \frac{1}{(m+1)} - \frac{m}{(m-1)} + \frac{m}{(m-1)(m+1)}\right\}$$

$$= 2(1-W)\,W\,m\left\{\frac{(m-1)(m+1) - (m-1) - m(m+1) + m}{(m-1)(m+1)}\right\}$$

$$= 2(1-W)\,W\,m\left\{\frac{m^2 - 1 - m + 1 - m^2 - m + m}{(m-1)(m+1)}\right\}$$

$$= 2(1-W)\,W\,m\left\{\frac{-m}{(m-1)(m+1)}\right\}$$

$$= -2\frac{(1-W)\,W\,m^2}{(m-1)(m+1)}.$$

Therefore

$$v(\bar{y}_{MOD}) = (1-W)^2\,\beta^2\,m$$

$$+ \frac{W^2 m^3}{(m-1)(m-2)(m+1)(m+2)(m+3)}\left\{\frac{(m^3+4m^2+8m-1)}{(m-1)(m+1)} + \frac{(n+1)}{(n-1)}\right\}\alpha^2$$

$$+ \{(1-W)^2 + 2\frac{(1-W)\,W\,m^2}{(m-1)(m+1)}$$

$$+ \frac{W^2 m^3}{(m-1)^2(m+1)(m+2)(m+3)}\left[(m^2+4m+1) + \frac{(n+1)}{(n-1)}\right]\}\,\delta$$

$$-2\frac{(1-W)\,W\,m^2\,\alpha\beta}{(m-1)(m+1)}.$$

APPENDIX IX

SUMMARY DATA OF MUSSA FIELD SURVEY

The summary data on area and production from the sampled villages (chapter 7, section 7.6)

Key:

MzAr	Maize area for maize only
BnAr	Bean area for bean only
MzbnAr	Area for mixed maize and beans
MzPr	Maize production from MzPr
MzCpr	Maize production from MzbnAr
BnPr	Bean production from BnAr
BnCpr	Bean production from MzbnAr

Village	Year	Location	MzbnAr (*ha*)	MzCPr (*kg*)	BnCPr (*kg*)
Nyankumbi	1989	Nyantorotoro	19.97	7525	1044
		Mbugani	22.31	6880	2148
		Sum	42.28	14405	3192
	1990	Nyantorotoro	23.86	5421	685
		Mbugani	22.31	5570	1490
		Sum	46.17	10991	2175

Village	Year	Location	MzAr (*ha*)	BnAr (*ha*)	MzPr (*kg*)	BnPr (*kg*)
Igawilo	1989	Iganjo	6.18	0.70	13240	80
		Igawilo	8.25	0.85	20968	525
		Sum	14.43	1.55	34208	605
Igawilo	1990	Iganjo	5.37	0.95	11000	350
		Igawilo	9.32	0.67	21850	1035
		Sum	14.69	1.62	32850	1385

Village	Year	Location	MzAr (*ha*)	BnAr (*ha*)	MzbnAr (*ha*)	MzPr (*kg*)	BnPr (*kg*)	MzCpr (*kg*)	BnCpr (*kg*)
Ngarash	1989	Juu	8.98	3.75	12.39	13530	3695	17380	8345
		Kati	33.64	10.42	13.90	24015	5795	12890	4015
		Sum	42.62	14.17	26.29	37545	9490	30270	12360
Ngarash	1990	Juu	10.02	3.54	13.37	24940	6215	11770	360
		Kati	32.18	9.21	13.98	44330	10675	7710	2020
		Sum	42.20	12.75	27.35	69270	16890	19480	2380

Village	Year	Location	MzAr (*ha*)	MzbnAr (*ha*)	MzPr (*kg*)	MzCpr (*kg*)	BnCpr (*kg*)
Sinoni	1989	Bondeni	3.44	9.97	3190	7480	2357
		Milimani	3.23	8.64	9790	11530	2530
		Sum	6.67	18.61	12980	19010	4887
Sinoni	1990	Bondeni	2.73	11.83	480	7240	2185
		Milimani	4.84	5.35	10230	3830	1440
		Sum	7.57	17.18	10710	11070	3625