USE OF AN AREA SAMPLING FRAME TO IDENTIFY THE SPATIAL DISTRIBUTION OF LIVESTOCK IN THE GAUTENG PROVINCE

by

CRAIG VON HAGEN

DISSERTATION

submitted in the fulfilment
of the requirements for the degree

MASTER OF SCIENCE

in

GEOGRAPHY

in the

FACULTY OF SCIENCE

at the

RAND AFRIKAANS UNIVERSITY

SUPERVISOR: DR. L. G. C. SCHEEPERS

AUGUST 2001

PREFACE

Gauteng Veterinary Services contacted the Agricultural Research Council's Institute for Soil, Climate and Water about the problem of unreliable livestock statistics. These talks resulted in a project for Gauteng Veterinary Services that concentrated on producing a methodology for producing reliable animal statistics for the province.

The author saw the opportunity to combine this research with formal study towards a Masters degree, on which all concerned parties agreed. The writer was the person who did the work with help from internal and external experts. The project was financially supported by Gauteng Veterinary Services and the Agricultural Research Council's Onderstepoort Veterinary Institute, regarding the extensive sampling undertaken. Gauteng Veterinary Services also supplied the manpower for the fieldwork.

This research was firstly written up as a formal Masters manuscript, then from there it was shortened into a much more condensed report for Gauteng Veterinary Services. The work in this manuscript is solely the work of the author.

ABSTRACT

In South Africa, there are no reliable statistics regarding animal numbers and distribution. The goal, therefore, of this research is to provide the framework and procedure for obtaining these statistics efficiently and accurately.

Available sampling methods and sampling frames were investigated and it was decided to carry out a sample survey because the Gauteng Province consists of a large number of holdings (land parcels). In the Gauteng Province, where a complete list of farmers or land owners is not available, it was decided to use an area sampling frame. Once the choice of sample design was made, the survey objectives were defined according to the clients' needs.

The sampling frame was constructed using various land parcel layers. These land parcels were merged, using GIS software, into one continuous layer of land parcels. They were then stratified to reduce the variance of the variable (animals) under study over the entire area, using area of land parcel and land-cover. The sample size was then calculated and the land parcels were selected randomly for survey purposes.

The survey was conducted between September and December 1999 and the questionnaires were input into a database for the estimation procedures. The closed estimation procedure was used because it is the only possible option if the data surveyed are referenced to the land parcel (and not to a farm that includes several land parcels).

The area frame sampling methodology worked well for cattle, sheep, horses, pigs and dogs/cats and to a lesser extent for goats, donkeys and game. The area frame method did not work well for poultry (because of extremely high values in a few land parcels), ostriches or mules (these are rare in the province).

Spatial distributions and density distributions were then interpolated from the animal counts taken in the survey and they give a general idea of the location of animals. The distributions of cattle, sheep, horses, pigs and dogs/cats are reliable. The distributions of the rest are distorted due to extreme counts in a few land parcels but a general idea of concentrations can still be inferred.

Considering that no historical data exists and that the overall goal of this research was to get an idea of animal numbers and the distribution of animals in Gauteng province, it can be considered successful, in that decision-makers now have a reliable source of information from which good decisions can be made.

SAMEVATTING

Daar bestaan geen betroubare statistiek rakende die aantal diere en hul verspreiding in Suid-Afrika nie. Die doel van hierdie studie is om ‘n raamwerk en prosedure daar te stel waarin hierdie statistiek effektief en akkuraat versamel kan word.

Beskikbare steekproefmetodes en steekproefraamwerke is ondersoek en daar is besluit om ‘n steekproef-opname uit te voer, omdat die Gauteng-provinsie uit ‘n groot aantal persele bestaan. In die Gauteng-provinsie, waar daar nie ‘n volledige lys van boere of grondeienaars beskikbaar is nie, is besluit om gebruik te maak van ‘n gebied steekproefraamwerk. Daarna is die opname doelwitte gedefinieer na aanleiding van die kliënt se behoeftes.

Die steekproefraamwerk is saamgestel deur gebruik te maak van verskeie bodembenuttingslae. Die bodemeenhede is deur middel van GIS sagteware in een aaneenlopende laag van bodemeenhede saamgevoeg. Hierdie laag is gestratifiseer om die variansie van die veranderlike (diere) oor die totale studiegebied te verminder, deur gebruik te maak van die oppervlakte van die bodemeenhede en die benuttingstipe. Die steekproefgrootte is bereken en die bodemeenhede is ewekansig gekies vir opname-doeleindes.

Die opname is tussen September en Desember 1999 gedoen en die vraelyste is in ‘n databasis ingesleutel vir beramingsprosedures. Die geslote beramingsprosedure is gebruik, omdat dit die enigste moontlike keuse is as die waargenome data gekoppel word aan die bodemeenheid (en nie aan ‘n plaas wat ‘n hele aantal persele kan insluit nie).

Die oppervlak-raamwerk steekproef metodologie het goed gewerk vir beeste, skape, perde, varke en honde/katte en in ‘n mindere mate vir bokke, donkies en wild. Die oppervlak-raamwerk metode het egter nie goed gewerk vir pluimvee (as gevolg van uitsonderlike hoë waardes in min persele), volstruise of muile (skaars in die provinsie) nie.

Ruimtelike-en digtheidsverbreidings is geïnterpoleer vanaf dieregetalle wat verkry is uit die opname en dit gee ‘n algemene idee van die ligging van die diere. Die verbreiding van beeste, skape, perde, varke en honde/katte is betroubaar, maar die verbreiding van die ander diere is verwronge as gevolg van uiterste tellings op ‘n klein aantal persele. ‘n Algemene idee van konsentrasies kan egter hiervan afgelei word.

Ten spyte daarvan dat geen historiese data bestaan nie, en met in agneming van die hoofdoel van die navorsing (naamlik om ‘n algemene idee te kry van dieregetalle en hul verspreiding in die Gauteng-provinsie), kan die navorsing as suksesvol beskou word, omdat besluitnemers nou ‘n betroubare bron van inligting het waarvan goeie besluite geneem kan word.

ACKNOWLEDGEMENTS

The Gauteng Veterinary Services and ARC–Onderstepoort Veterinary Institute for funding.

Mr G. Narciso for the introduction to area frame technology.

Mr A. B. Potgieter for general guidance and assistance.

Gauteng Veterinary Services for providing manpower to carry out the survey.

Dr F. van der Vyfer, Dr F. van der Meijden and Mr G. van der Zel from Gauteng Veterinary Services for their critical inputs.

Mr F. Luccarrini from Aquater (Italy) for assisting with the estimation procedure.

TABLE OF CONTENTS

CHAPTER ONE – BACKGROUND AND PROBLEM

CHAPTER 2 – DEVELOPMENT OF SAMPLE DESIGN

CHAPTER 3 – APPLICATIO N OF PROCEDURES TO THE STUDY AREA

LIST OF FIGURES

LIST OF TABLES

LIST OF MAPS

All maps are to be found at the end of the manuscript (pg 111), following in a sequential order.

CHAPTER ONE – BACKGROUND AND PROBLEM

1.1. INTRODUCTION

Agricultural statistics contribute to a stable economic atmosphere and reduce uncertainty for production, marketing and distribution operations. Uses of agricultural statistical information are extensive and varied. A main user is the producer, although the unique impact of crop and livestock data on a given producer depends on the type and size of operation. Information on production and stocks helps non-perishable crop producers decide whether to store their production to find their best market opportunity. Producers of perishable crops are interested in the timing of plantings and area planted as an indication of market flow during the harvest period. Other major users of agricultural statistics are farm organisations, agri-business, policy makers and agricultural economists (analysts). Farmers, ranchers and agri-business firms need information about current and future supplies of agricultural commodities for marketing, planning and decision-making. This information is also necessary for policy decisions concerning Government programs affecting the agricultural economy in specific ways (Fecso *et al.*, 1986). Uses of data by farm groups may range from maintaining a basic data series to preparing an important marketing campaign. Government agencies at various levels are important users of statistics. Agricultural statistics are used to plan and administer programs in such areas as consumer protection, conservation, foreign trade and education. The analyst transforms statistics into projections of current trends, interpretations of their economic implications, and evaluations of alternative courses of action in terms of prospective outcomes. These projections multiply the usefulness of statistics (Garibay *et al.,* 1996).

Most national agricultural statistics are obtained by collecting data at farm level. This can be done through a complete census of all farm operators, usually a costly and time-consuming process. It is also important to note that censuses are error prone because of omissions of farms and the difficulty in quality control on a large mass of data. The alternative to using a complete enumeration of a population is to conduct a sample survey. Both cost efficiencies and timely results favour the sample survey. Data are of most use when gathered and disseminated quickly, thus providing information for timely decision making. A well-designed sample survey is much more cost-efficient than a complete census, and may even be more accurate (Garibay *et al.,* 1996).

In South Africa, there are no reliable statistics regarding animal numbers and distribution. Provision for collecting such information is not made in the National census and consequently this information is not collected on a regular basis. The last agricultural census in 1993 (CSS Report No. 11-02-08) also did not collect information on animal numbers (only income and expenditure information was collected). Currently, according to Gauteng Veterinary Services (GVS), animal counts are done in an irregular way by extension officers going about their daily work. This results in duplication of animal counts and therefore inaccuracies. Some areas are visited more often than others and consequently, some areas are ignored. Experts need reliable, up-to-date, information regarding animal numbers and distributions in order to make informed decisions regarding disease control and high risk areas.

Statistics on agriculture are generally gathered by questioning farmers in exhaustive censuses or surveys relating to some farmers only. In order to apply such a method, an exhaustive list of farmers is necessary. Lists of farmers are not always available. This happens when regulations forbid agricultural statisticians to use them, and also when rapid changes in the structure of agriculture make it too difficult to prepare and update such lists (Deneufchatel & Porchier, 1993).

The South African Government has set out three major goals for policy reform in agriculture as laid out in the Ministry for Agriculture and Land Affairs (1998) document:

- to build an efficient and internationally competitive agricultural sector,
- to support the emergence of a more diverse structure of production with a large increase in the numbers of successful smallholder farming enterprises, and
- to conserve our agricultural natural resources and put in place policies and institutions for sustainable resource use.

All three points are relevant to this research, but the first point is the most relevant. The way to build an efficient and competitive agricultural sector, which can compete internationally, is by using scientific methods that have a proven track record in the international agricultural community. This is what this research hopes to achieve i.e. to put in place a system that can provide up-to-date and reliable statistics with regard to animals in South Africa. South Africa, however, is too large an extent for masters level research and a smaller area had to be selected.

GVS had talks with the Agricultural Research Council's Institute for Soil, Climate and Water (ARC-ISCW) about the problem of unreliable livestock statistics. These talks resulted in a project for GVS that concentrated on the Gauteng Province. The writer was the principle researcher for this project, which was financially supported by GVS regarding the extensive sampling undertaken. GVS also supplied manpower for the fieldwork. The project was also registered as a masters research project, on which all concerned parties agreed. This research thus resulted into this masters manuscript from which a condensed report was extracted for GVS.

Market information is crucial to the proper functioning of any market. It promotes efficient arbitrage between markets, which is to the benefit of both consumers and producers, and the efficient allocation of productive resources. It improves the bargaining power of producers when dealing with traders and processors, and it reduces transaction costs by reducing risks (Ministry for Agriculture and Land Affairs, 1998).

A brief overview of agriculture in South Africa will now be given to give a better understanding of the current structure of agriculture in South Africa. It also provides an idea of how changing legislation defines the priorities of government institutions such as GVS and how this research can help GVS to adapt and comply with new legislation and policies.

1.2. BACKGROUND - THE SOUTH AFRICAN AGRICULTURAL SITUATION

The South African government has issued the following mission statement for agriculture.

Mission Statement for Agricultural Policy

Ensure equitable access to agriculture and promote the contribution of agriculture to the development of all communities, society at large and the national economy, in order to enhance income, food security, employment and quality of life in a sustainable manner (Department of Agriculture, 1995).

The above mission statement has an indirect impact on this research as reliable animal statistics are needed as part of an encompassing plan to improve the development of communities and to reduce food insecurity. A brief overview is given to impart some idea of the agricultural situation in the country with special reference to livestock.

Agriculture in South Africa has emerged from a history of protection and subsidisation, which affected the efficiency, and competitiveness of the sector. The strategy for achieving the set objectives of making agriculture more efficient, creating jobs and opportunities and using resources sustainably, is based on an outward looking approach. In this approach the global village is seen not only as a market for output, but as a tool for effecting efficiency by exposing producers to international competition (Ministry for Agriculture and Land Affairs, 1998). Previously such efficiency was not needed as farmers were heavily subsidised.

Agricultural and economic policies encouraged commercial farmers to increase the farm size and to substitute labour with capital. Single-channel fixed price marketing schemes and the cross-subsidisation of agricultural product prices and transport costs caused a distortion of price relationships, which resulted in the incorrect allocation of agricultural resources (Department of Agriculture, 1995). This is no longer the case. By early 1998, the control boards dealing with maize, sorghum, oilseeds, wool, meat, wheat, cotton, mohair, lucerne, citrus, deciduous fruit, dried fruit, milk and canned fruit had all been closed (Ministry for Agriculture and Land Affairs, 1998) and subsidies removed.

Research is needed to improve the efficiency of the agricultural sector so as to be able to compete in the international market. This research aims to provide some assistance in this regard. A brief breakdown of the research structure with regards to agriculture in South Africa will now be given.

The government funds some 70 to 80% of all agricultural research in South Africa. Until 1990, most research was done within the National Department of Agriculture's (NDA) 11 institutes to the benefit of white farmers. These well-funded and well-maintained research facilities were then handed over to the Agricultural Research Council (ARC). The ARC receives about 60% of the public funds allocated to agricultural research and development. The balance of public funding goes to universities, the NDA, the Foundation for Research and

Development, the Council for Scientific and Industrial Research (CSIR), the Protein Research Trust and the Water Research Commission (Ministry for Agriculture and Land Affairs, 1998).

The ARC was established in terms of the Agricultural Research Act (Act No. 86 of 1990) and reports to Parliament through the Minister of Agriculture. The ARC has focused almost entirely on managing its own capacity, organised into some 16 research institutes, most of which do research on commodities. Some 70% of the ARC's funds comes from a Parliamentary grant, a further 20% comes from National or Provincial Agricultural Department grants, and the remaining 10% is provided by the private sector through research contracts (Ministry for Agriculture and Land Affairs, 1998).

The research activities of the above-mentioned institutions need to be brought in line with the government's research priorities, which are discussed below.

Numerous reviews of South Africa's agricultural research systems have identified the need to set research priorities to ensure that public expenditure in research helps the Government reach its objectives. Farmers can choose between crops, livestock and technologies that differ regarding their susceptibility to drought, disease and infestation, and regarding the costs of preventative measures such as dipping and spraying. Research will be aimed at supporting the development of more robust technologies as well as preventative measures to reduce risk. While priorities set for research in agriculture in the past were supportive of the Apartheid policy objectives, it has not been easy to reorientate these priorities to fit in with the new dispensation. Such reorientation requires major institutional reform (Ministry for Agriculture and Land Affairs, 1998). One such research priority is livestock management systems.

1.2.1. **Livestock management systems** (National Department of Agriculture, 1998)

Although the former homelands occupy only 13% of the agricultural land in South Africa, they carry some 30% of the livestock. Livestock in South Africa represents a valuable asset to large numbers of poor people, yet currently this asset does not deliver the material benefit to its owners that it is capable of. It is government policy to address the factors that currently constrain increases in productive output from animal ownership in small scale and traditional livestock areas, and design policy that will:

- lead to more efficient and sustainable management of land under the control of livestock owners,
- improve the effectiveness of support services, particularly those pertaining to animal health, animal nutrition and the marketing of cattle, small stock and poultry, and
- enhance the overall productive potential of animals.

This research will have a direct bearing on the improvement of support services. By providing a system for collecting reliable animal statistics, it will improve the decision-maker's ability to make informed choices regarding animal disease. It may also identify areas where animal numbers and densities exceed the carrying capacity of a certain type of land. Decisions can then be made regarding sustainable land management.

Livestock farming in the small-scale sector is technically very inefficient. Reproduction rates are low, the control of important parasites is patchy, and supplementary feeding is virtually unknown. Livestock and crop production are practised in isolation from each other, despite the fact that many households have their own stocks and produce crops. Manure accumulates to nuisance levels in the kraals, and crop residues are used in an unsystematic way in the fields as part of the common property resource. There is thus considerable scope for raising rural incomes through improvements in livestock productivity. Efforts to enhance household and social welfare through livestock improvement will provide multiple benefits if the livestock systems are dealt with as part of the whole farm system, with cropping and livestock activities being integrated.

Traditional livestock farming systems differ across South Africa. In the eastern and south-eastern regions, cattle herds are firmly locked in systems of use by the household, with very few animals being available for disposal on the formal market. In the north and west of the country, cattle are actively traded, up to 12% being disposed of on the formal market. While cattle are not traded actively in the east and south-east, small stock and their fibre products are. This variability means that, within broad policy, local approaches are likely to result in greater benefits than the blanket programmes of the past.

1.2.1.1. Cattle

Cattle account for between 80 and 90% of the asset value of livestock in the small-scale sector. The highest priority with regard to cattle is the improvement of animal health. This includes the control of ticks to prevent tick-borne diseases, prophylactic inoculations, and the treatment of illness and injury. Of these, tick control is generally more efficiently provided for by collective action, whether by the group management of dip tanks or the group purchase of material for individual treatment with hand sprays or pour-on treatments. All other treatments are more efficiently performed by the individual farmer.

Chronic subclinical malnutrition is, in the absence of disease, the prime cause of low productivity in cattle. This in turn is attributable to the lack of any organising institution. In contrast to the collective action required to improve nutrition from grazing, each owner has the option of improving supplements for his or her own cattle. This is not widely practised. Options include feeding from preserved crop residues, growing improved pastures on the owner's arable land, and the purchase of supplements.

The reduction of Government involvement in the areas of animal health and marketing cannot be achieved in too short a time without jeopardising the agricultural sector. Setting up competent cattle-owner associations to take over the collective interests previously administered by the Government is a prerequisite for the long-term sustainability of livestock improvement. They would take responsibility for the organisation of tick control, obtaining acaricides, remedies and prophylactic injections, liaison with the private sector suppliers of these commodities, marketing; and management of the common grazing resources.

1.2.1.2. Small stock

Small stock are kept in all parts of the country but achieve importance in terms of formal marketing only in the Eastern Cape. In this province, wool and mohair are significant contributors to the rural economy and are growing in importance. Small stock do, however, play a valuable role in all rural communities because of:

- low cost, which makes accumulating herds easier for the poorer groups,
- their small size, which means that they are easily disposed of when slaughtered, and
- the ease with which they can be marketed to meet minor cash demands.

The particular importance of small stock is that they offer the marginalised members of the community a means of support, and security against total destitution.

The priorities with regard to small stock are disease and pest control, extension and marketing. Small stock are affected acutely by an array of diseases and parasites, which significantly reduce productivity if not controlled. Indigenous flocks enjoy a degree of resistance in that, when affected, they do not necessarily die, and could even show spectacular improvements if treated. In the case of sheep scab, however, there is no resistance and the disease does present a real threat to the wool industry where it is left unchecked. The point of departure for improving the small stock sector therefore is the control of key diseases, particularly blue tongue, the reduction of internal parasites; and in the case of wool sheep, the elimination of scab.

The health and welfare of small stock, particularly in the subtropical areas, depends on a routine of dosing and dipping. Doing this effectively requires, firstly, that the owner must be aware of the effect of diseases and parasites and, secondly, that the control methods must be understood. With important exceptions, the general experience in the small-scale farming sector is that diseases and pests are viewed as unavoidable hazards of small stock production. With the exception of the Eastern Cape province, there appears to be a lack of trained extension staff dealing with small stock. A prerequisite for improvement in the sector is therefore that specialist advisory should be introduced or strengthened.

1.2.1.3. Game ranching

The game-ranching industry has shown substantial growth over the past few decades, to the point where it is currently estimated that game is ranched, either alone or in combination with livestock, on approximately 8 000 ranches covering about 16 million ha. The industry has grown largely in response to new levels of demand for game products ranging from hunting trophies, skins and horns to meat, especially venison, and live sales. In addition, value has been added by the provision of tourism facilities, such as lodges and bed-and-breakfasts.

The Government recognises the contribution made to rural economies by this form of land use, and will ensure that the same set of research, extension and veterinary considerations applicable to other sectors of the livestock industry will be ap plied to game ranching. It is also expected that the industry itself will devote more of its growing income to research and conservation.

1.2.1.4. Poultry

Poultry are kept throughout rural areas and constitute a significant portion of the animals kept by poorer households for consumption and marketing. They have the singular advantage of being able to provide producers with regular significant cash incomes, and therefore fit in well with the increasing need for alternative sources of income for rural people. Even more so than the small stock sector, poultry production has been market driven with limited support from the Government. Production systems are far ranging, from household flocks, small-scale broiler or egg production to sophisticated production in specialised housing. The demand for poultry products in the rural areas is substantial – quite adequate to support local marketing enterprises.

.

1.2.2. Animal disease control (Ministry for Agriculture and Land Affairs, 1998)

Considerable uncertainty and risk is attached to farming. Agriculture in South Africa is inherently more risky than in many other countries because of low average rainfall, and the wide variability in rainfall both between and within seasons in most parts of the country. In addition to the risks associated with drought, farmers are also confronted with a range of other hazards, including hail, fire, pests and diseases.

The Constitution provides a framework for the Government's livestock and animal health services. Animal health and disease control are listed as a concurrent national and provincial competency. A number of veterinary related spheres of the Government have also been listed as provincial and local competencies. They include veterinary services (excluding regulation of the profession); facilities for the accommodation, care and burial of animals; the licensing and control of undertakings that sell food to the public, municipal abattoirs and pounds.

The Animal Disease Act of 1984 (Act No. 35 of 1984) emphasises the threat that infectious animal diseases and parasites pose to the agricultural sector in South Africa, and the Southern African region as a whole. The Act now needs revision to bring it in line with the Constitution and to clarify provincial and national responsibilities.

It is proposed that under the Animal Health Bill, the NDA be made responsible for the co-ordination of all aspects of animal disease control and eradication throughout the country. This would involve setting standards for the control of notifiable diseases in animals (including game), that are applicable to all the provinces.

The legislation will authorise the Government to:

- co-ordinate and maintain a competent epidemiological database and information system of notifiable disease surveillance, based upon disease incidence reporting and supported by field and laboratory testing,

- develop programmes in consultation with provincial governments and private agricultural stakeholders to contain and eradicate diseases that may pose a threat to the national economy,

- set standards for routine control measures for those notifiable diseases and parasites, which are the agreed responsibility of the provincial Governments, and institute effective monitoring procedures to ensure compliance with those measures, and

- adopt quality control measures for the accreditation of all laboratories offering veterinary testing services.

1.3. RESEARCH PROBLEM STATEMENT

In order to adapt and comply with the goals and regulations of the Government, GVS have identified the lack of spatially linked livestock information as a major problem, especially with regard to the outbreak, and subsequent control of diseases. Veterinary Services have to answer the following set of important questions:

- What is the geographical extent of the disease?

- How many heads of livestock might be affected?

- How much vaccine has to be sent to the government veterinarian in the affected area?

- Will other animal species in the area be affected?

– Which farms will be put in quarantine?

– What short and long term plans should be put in place to deal with the situation?

Presently, answers and decisions flowing from these questions are made based on how well the local government veterinarian knows the area. Consequently a lot of guesswork is involved. To prevent the outbreak of an epidemic, GVS have realised that accurate, spatially linked livestock information has to be available in order to take the most appropriate and cost effective decisions.

The goal, therefore, of this research is to provide the framework and procedure for obtaining spatially linked livestock statistics efficiently and accurately. From the results provided by the framework and procedure, estimations will be made and incorporated into a Geographic Information System (GIS) for the Gauteng Province.

Statistics based on an area-sampling framework are an internationally recognised tool to satisfy the above requirement. Possible trends and areas of high occurrence can then be identified, investigated and monitored at a later stage**.**

As it's main objectives, this research undertakes to answer a number of questions:

(i) will the area frame sampling method provide accurate statistics for use by GVS,

(ii) can area frame sampling be implemented successfully in South Africa, and

(iii) can the area frame sampling method be implemented for use in other sample surveys e.g. estimating crop statistics?

In an attempt to answer these questions, several methodologies had to be researched and evaluated, from which an appropriate set of procedures, each containing specific techniques, were selected for application. These procedures rendered results, which were assessed using statistical analysis and indicated varying success rates for the different types of selected animals. The research framework outlined below explains the procedure followed to select the appropriate sampling method and subsequent determination of animal distributions.

1.4. RESEARCH FRAMEWORK

To solve the problem of the lack of reliable, up-to-date, spatially linked livestock information, a certain procedure was followed. This section outlines this procedure and the structure of the research.

In the initial stage of the research, sampling as a means of establishing total animal numbers versus complete census statistics had to be evaluated. If the existing literature reports favourably towards the use of sampling techniques to derive animal statistics, then a specific applicable sampling technique will have to be selected for implementation. The results of this applied sampling technique will then have to be extrapolated to supply total estimates and these estimates then assessed in terms of the real situation using statistical techniques. Distribution patterns must then be derived from the sampling point information using some form of interpolation procedure. For each of these steps, in depth literature research must be done to select the most appropriate technique amongst the multitude of available choices. Specific characteristics of the study area should be taken into consideration during such a selection procedure. After careful consideration of all these aspects, a very specific order of research procedures were adopted which forms a logical stepwise development, as follows:

- Firstly, an overview of the physical study area was needed. This will give an indication towards the particular techniques to be selected, as in most sampling techniques the type and size of land units play an important role in its selection or not (see section 1.5).

- Secondly, the advantages and disadvantages of complete agricultural censuses and sample survey as a means of obtaining animal statistics had to be investigated, mainly though existing literature. With this knowledge and knowledge about the situation in the study area, a choice of one or the other must be made (see section 2.1).

- Thirdly, if sampling is to be selected as a technique, it must be thoroughly investigated in existing literature. The different techniques and their specific applicational advantages under certain circumstances must then be evaluated against the situation in the study area. After careful assessment a specific method to be used must then be chosen (see section 2.2).

- Fourthly, the selected area frame sampling method must be developed for the area under study, applied and the data thus collected processed into total estimates of animal numbers. This involves a whole set of subsequent steps building onto each other:

  – Preparation of the area framework (section 3.1),
  – stratification of the area framework (section 3.2),
  – determination of the sampling size (section 3.3),
  – cartographic preparation of maps to identify sampling units (section 3.4),
  – preparation of an adequate survey questionnaire (section 3.5),
  – tabulation and database development to contain the derived data in such a way that linkages with topological map data would be possible (section 3.6),
  – organisation of the actual field work to gather data using the developed questionnaire at the selected sampling points (section 3.7),
  – checking the quality of collected data and field work (section 3.8),
  – processing of the gathered data into the built database and editing of the database (section 3.9), and
  – deriving estimates for the total animal numbers after selection of a specific estimation technique (section 3.10).

- Fifthly, the results of this sampling and estimation procedure are shown in tables for each of the selected groups of animals. Each table also shows the reliability of these sampled results using the coefficient of variation (see section 4.1).

- Sixthly, the sampling point data are used to interpolate animal distribution patterns for all selected groups of animals (see section 4.2).

- Lastly, the results of the previous two steps are considered to give an answer towards the initially stated problem and sub-problems.

As the characteristics of the study area plays an important role in decisions to be made on the selection of specific sampling, estimation and interpolation techniques to be used; the study area will now be discussed in term of the characteristics to be considered during the rest of the procedures.

1.5. THE STUDY AREA

The study area is the Gauteng Province of South Africa. Gauteng is the commercial centre of South Africa as about 35-40 % of the country's Gross Domestic Product (GDP) is generated here. Gauteng is the most densely populated province in South Africa. It is the smallest province but it houses more than 7 million of the country's people. The level of urbanisation is at 97% (Burger, 1999). Gauteng's agricultural sector is geared to provide the cities and towns of the province with fresh products daily. These include dairy products, fruit, meat, eggs and flowers. A large area of the province falls within the maize production area of the country. Groundnuts, sunflower, sorghum and cotton are also produced. The province is characterised by convective rainfall during summer with warm summer temperatures, followed by a cold winter with frost (Burger, 1999). Snowfall is a rare event.

Map 1 shows the major roads at 1:500 000 scale and the magisterial districts in the Gauteng Province (all maps are included as an appendix at the end of the document so as to enable quick comparisons and easy integration with the synthesis). Map 2 shows the location of some major towns, while map 3 shows the 1:50 000-map sheet grid that covers the province. These maps help to give some orientation information about the province within the South African context.

No comprehensive, up-to-date list of farmers, livestock owners or land operators exist for the Gauteng Province, so a sampling method that does not require such a list is needed. No reliable statistics regarding animal numbers and distribution exist. The province consists of urban and rural areas and as a consequence the area is made up of large and small land parcels and land-cover varies considerably throughout the province.

Cattle, sheep and pigs are found on the fringes of the urban areas. Chickens are mainly confined to large commercial chicken batteries. Game is found mainly in the north-east of the province, while ostriches and donkeys are relatively rare.

With the necessary characteristics of the study are highlighted, the advantages and disadvantages of complete enumeration via censuses and estimation of total numbers via sampling can now be addressed.

CHAPTER 2 – DEVELOPMENT OF SAMPLE DESIGN

To establish data with acceptable levels of accuracy for planning purposes, complete census information can be gathered or a sample survey could be done. Both these main methods have certain advantages and disadvantages, which should be carefully considered before selecting a particular method for gathering the required information. In this chapter, the inherent characteristics of censuses and sample surveys will be compared, focussing on survey designs that are more applicable for agricultural purposes. These characteristics and advantages will then be taken into consideration for the selection of a survey design to establish animal numbers for the Gauteng Province by also taking into account the study area's nature and particular aspects related to the implementation of such a survey design.

2.1. AGRICULTURAL CENSUSES AND SAMPLE SURVEYS

An agricultural census has been defined by the Food and Agricultural Organisation (FAO) as a large-scale, periodic, statistical operation for the collection of quantitative information on the structure of the agriculture of the country. The basic objectives of an agricultural census are (Food and Agricultural Organisation Statistical Development Series, 1996):

- to provide aggregate totals for fundamental agricultural data for use as the benchmark for inter-censal estimates,
- to provide a framework for other agricultural sample surveys, and
- to provide data for small administrative units.

Since an agricultural census is not a frequent data collection activity, it is natural to associate it with those aspects of agricultural structure, which change relatively slowly. Census data are also useful in the design of annual or seasonal agricultural sample surveys. Stratification criteria, clustering procedures, etc. are often based on agricultural census data (Food and Agricultural Organisation Statistical Development Series, 1996).

A census or complete enumeration is a major undertaking requiring enormous amounts of people, training, forms, travel, summary equipment and money. The tabulation of the millions of questionnaires takes months and sometimes even years. Census results provide a wealth of information for government planners, businessmen, analysts of various disciplines, educators and others. It is also important to note that census statistics are usually the only source of data for smaller geographic subdivisions. Unfortunately due to their time and resource requirements censuses are normally taken only once every ten years. Most sectors of the economy need more frequent statistics in order to adjust operation or change policies. Sample surveys meet these statistical needs (International Training, 1987).

Beside the census, a well-designed national agricultural statistical system includes surveys aimed at collecting more specific and timely information, for example crop yields, that is not feasible to collect through an agricultural census (Food and Agricultural Organisation Statistical Development Series, 1996).

2.1.1. Complete enumeration censuses vs. censuses carried out on a sample basis

Agricultural censuses can be classified into two categories: censuses conducted by complete enumeration (survey) of all holdings or by a sample enumeration. The main characteristics of these two categories are the following (Food and Agricultural Organisation Statistical Development Series, 1996).

Complete enumeration

For agricultural statistics by complete enumeration, the enumerators complete a questionnaire for each holding. The result for each characteristic is obtained from the values of the characteristics in all holdings. The census will only include non-sampling errors.

Sample enumeration

Agricultural censuses based on sample enumeration are probability sample surveys, that is, surveys for which a probability sample is selected, and for which the methods of estimation for each census characteristic allows for the establishment of their statistical precision. This requires defining the sampling units and their probability of selection, from a known universe (frame).

2.1.2. Advantages of carrying out a census on the basis of complete enumeration (Food and Agricultural Organisation Statistical Development Series, 1996)

Censuses carried out on a basis of complete enumeration have several advantages when data derived from it is used for planning purposes. The census results can be obtained for small administrative and other area units. Such information is sometimes required by law, for local planning or for practical purposes. Some crops, although cultivated only to a limited extent, may be of great economic importance. Information on such crops can only be reliably obtained from a complete census. The situation is similar with regard to rare species of animals, or any other rare variable.

Current agricultural statistics in most countries have to be collected through annual sample surveys. These sample surveys can be planned much more efficiently if census results are available for small area units. The census listings can be used as a frame for the selection of the sample. This may prove an important saving because the preparation of the frame itself generally consumes a significant part of the budget allocation for the sample survey. Census data can be used for improved sampling design and better estimation procedures, which may lead to more reliable estimates of the characteristics under study.

The planning and implementation of an agricultural census conducted on a complete enumeration basis requires fewer highly qualified statistical personnel than a census conducted on a sample basis. This constitutes an important advantage in countries with limited expertise. Processing data from a complete enumeration is straightforward and does not involve the calculation of sampling errors or expansion factors and therefore requires fewer skills.

2.1.3. Disadvantages of carrying out a census on the basis of complete enumeration (Food and Agricultural Organisation Statistical Development Series, 1996)

Although data derived from complete enumeration censuses is preferable for planning purposes, these types of censuses can have significant disadvantages. In a country with a large number of holdings, to conduct a census by complete enumeration is, in practice, more expensive and time consuming than a census conducted on a sample basis. This is a particularly important consideration in areas with poor communication and transport infrastructure. For a complete enumeration census, a very large number of enumerators and supervisors are required. Quite often candidates with the desired qualifications are not available in the required number so the standard has to be lowered with a consequent effect on the quality of data. The quantity of data to be processed is very large for a complete census and therefore the cost of data processing will be higher due to the large volume to be processed. The results may also be considerably delayed if insufficient data processing equipment is available.

2.1.4. Factors for consideration in choosing between a complete or a sample census

Several important aspects should be considered when making a choice between a complete enumeration and a sample census. The sample enumeration is a very attractive proposition where there is a severe limitation of funds and personnel and the aim is confined to securing data with reasonable accuracy for major administrative units and the country as a whole. A decision whether to carry out a complete enumeration or to plan a sample survey of holdings will depend on the level at which the results are required. That is whether the results will be tabulated for the entire country, for individual provinces, for individual districts, or for smaller administrative subdivisions (Food and Agricultural Organisation Statistical Development Series, 1996).

Even those countries that lack resources should consider the possibility of undertaking at least a minimum part of the census on a complete enumeration basis. This is to ensure a good base for preparing an efficient sampling design for the collection of detailed data on important items, as well as planning future agricultural surveys to collect current agricultural statistics, and to be able to produce at least some data for small administrative units (Food and Agricultural Organisation Statistical Development Series, 1996).

In practice, control of non-sampling errors, including those of non-response, may be possible only on the basis of a sample. Even though a census may have been planned on the basis of a complete count, sampling techniques will have to be used for controlling the census operations, the response errors and those of data processing. However, in order to carry out these tasks effectively there is a need for highly qualified persons to design and execute the sampling plans (Food and Agricultural Organisation Statistical Development Series, 1996).

It was decided to carry out a sample census for this research because there was a limitation regarding funds and personnel and because the results only needed to be tabulated to the provincial level (i.e. Gauteng Province). Once the decision of choosing between a complete census and a sample survey has been made the selection of sample and survey design must be considered.

2.2. SURVEY DESIGN AND SAMPLE DESIGN

The **survey design** refers to the definitions and established methods and procedures concerning all phases needed for conducting the operation. These include: sample design, the selection and training, the organisation of the logistics involved in the distribution and receipt of questionnaires, the data collection and data processing procedures, and the analysis of data needed for the release of the final results. An agricultural census may have a number of different survey designs. For instance, it can be carried out using a number of different enumeration procedures (e.g. by personal field interview, by expert observation, self–administered questionnaires, by mail or telephone), by complete or sample enumeration, different types of sampling methods, sampling selection methods and estimation procedures, etc. In agricultural censuses, most data are obtained by enumerators through personal interviews with the holders, by using a questionnaire for each holding (Food and Agricultural Organisation Statistical Development Series, 1996).

The **sample design** of a survey or an agricultural census conducted by sample enumeration refers to the techniques for selecting a probability sample and the methods to obtain the estimates of the desired characteristics from the selected sample (Food and Agricultural Organisation Statistical Development Series, 1996).

As it was previously decided to make use of a sample survey for this study, the rest of this chapter will focus on possible sample designs that can be implemented and then selecting an appropriate design for this particular research.

2.2.1. Main types of sample designs for agricultural censuses

For agricultural censuses based on a sample enumeration, there are two basic types of sample methods commonly used concerning the final stage sampling unit and their probability of selection, namely list sample methods and area sample methods. In addition multi-frame sample methods are used, these being survey designs that combine an area sample with a list sample to obtain the census estimates (Food and Agricultural Organisation Statistical Development Series, 1996).

2.2.1.1. List sample designs (Food and Agricultural Organisation Statistical Development Series, 1996)

A list frame is a frame used for the last selection stage of a list sample design. Lists of holdings or holders' addresses form list frames. List sample designs are the most commonly used sampling procedures. In this case, the last stage sampling units are generally the holdings or the holders' addresses.

List sample censuses often include some strata of special holdings that are completely enumerated, or have a high sampling-selection fraction. Such strata consist of holdings that either corresponds to a significant proportion of the total estimated value of important census characteristics, or, if selected in the sample, whose characteristics may distort the results. For example, the strata of special holdings may consist of large holdings, holdings with the largest area for a given crop, with the largest livestock herds, highly specialised holdings or those corresponding to a localised production.

Cluster sampling is commonly used to account for geographic contiguity in the first stages of the sampling method. A sample of holders can be selected indirectly by first selecting a sample of villages (PSUs) with probability proportional to their total population (or housing units), since such information is usually available in most countries and approximates to the number of holders. Some additional information about the villages, such as farm population and primary agricultural activity, might also be available for stratification purposes. The total population of each selected village would then be screened for agricultural holders and a sample of holders will be selected in a second stage of sampling. During data collection, the enumerator usually completes a questionnaire for each selected holding by conducting an interview with the holder. In addition, in some cases the enumerator measures the fields and gathers whatever other data is needed to complete the census questionnaire.

2.2.1.2. Area sample designs

Since 1954, the National Agricultural Statistical Service (NASS) of the US Department of Agriculture (USDA) has been developing, using and analysing area sample frames as a vehicle for conducting surveys to gather information regarding crop acreage, cost of production, farm expenditures, grain yield and production, livestock inventories and other agricultural items (Cotter & Nealon, 1987). Since 1967, the NASS has been using area frame sampling in all 48 conterminous states in a system of surveys for obtaining this information (U SDA, 1983).

Iowa State University began construction of area frames for use in agricultural surveys in 1938. NASS began research into the use of area sampling frames in the mid – 1950's to provide the foundation for conducting probability surveys based on complete coverage of the farm sector. In 1954, area frame surveys were started on a research basis in ten states - 100 counties with 703 ultimate sampling units or segments. These surveys were then expanded over the years and made operational in 1965 in all states.

Changes made to the area frame methodology during the sixties and early seventies were mainly associated with sampling methods such as land use stratification and replicated sampling. Technological changes were incorporated during the seventies and eighties in the form of increased computerisation, use of satellite imagery, use of analytical software and development of an area frame sample management system among others. The area frame program has grown over the years and is conducted in 48 states with approximately 16 000 segments being visited by data collection personnel for the major agricultural survey conducted during June of each year (Cotter & Nealon, 1987).

An area frame for a land area consists of a collection or listing of all parcels of land for the area of interest. These land parcels can be delineated based on factors such as ownership, or based simply on easily identifiable boundaries as is done by the NASS. An area frame is critical to producing quality estimates, as it provides complete coverage with all land areas being represented in a probability survey with a known (not necessarily equal) chance of selection (Cotter & Tomczak, 1994). In circumstances where a complete list of farmers or land owners is not available, one may, instead of drawing from a list, draw a sample of areas within the territory in question. It can be used to obtain area estimates directly, but it can also become a basis for other studies on variables associated with the areas contained in the sample (such as yields), or the farms to which these areas belong. In this way a representative sample of farms is drawn indirectly without making use of a list (Deneufchatel & Porchier, 1993).

The concepts of area frame sampling are very simple: divide the total area to be surveyed into *N* blocks, without any overlap or omission; select a random sample of *n* blocks; obtain the desired data for reporting units of the population that are in the sample blocks through survey methods; and estimate population totals by multiplying the sample totals by *N/n*. The simplicity of the idea is in striking contrast to the complexity of successful application of the concepts (International Training, 1987). The technique also produces a variance and an estimation of error. Most area frame sample survey designs for agricultural censuses consist of a stratified probability sample of segments. The strata are defined by intensity of cultivated land, predominance of certain crops or other land-use characteristics (Food and Agricultural Organisation Statistical Development Series, 1996).

Two types of segments have been used for the area sample design of censuses in developing countries, namely segments that have recognisable physical boundaries and segments that coincide with the land of agricultural holdings (Food and Agricultural Organisation Statistical Development Series, 1996).

The boundaries of **segments that have recognisable physical boundaries** are physical terrain features (roads, rivers) that are readily found and provide an unambiguous identification of the segment. In this case, for each stratum (a stratum is a grouping of areas that have similar characteristics e.g. cultivated lands), the segments are defined with approximately equal area and a constant sampling expansion factor is used to obtain the estimates derived from data in the segments. The sample usually consists of a number of selected independent sample replicates in each stratum, that facilitates the rotation of the sample to reduce respondent burden if the sample is to be used periodically over time. These designs generally involve an annual (or seasonal) field data collection carried out by enumerators who complete a questionnaire for each tract included in each selected sample segment but can also be used in taking censuses. The data collection may involve objective measurement of agricultural areas utilising aerial photos. For each tract of a given sample segment, the enumerators delineate on the photo, the boundaries of the tract included inside the segment and the boundaries of all fields included in the tract. They also verify the crops planted and other uses of land for each field, as well as the information provided by the holder. Such identified agricultural areas in each sample segment are later measured in the office, providing the basis, through sample expansion procedures, for an objective estimation of agricultural areas (Food and Agricultural Organisation Statistical Development Series, 1996).

For **segments that coincide with the land of the agricultural holdings (point sampling)**, a grid is overlaid on the strata (areas with similar characteristics e.g. cultivated lands or grazing lands) and a sample of points is selected. Then, the points are identified on the ground and the corresponding holdings form the area sample. Thus, the design can be considered a stratified sample of holdings selected with the probability proportional to their sizes.

For the objective measurement of agricultural areas, the measurement on aerial photos is a very important advantage particularly if the interviewed holder does not know or does not want to report the area of land operated (Food and Agricultural Organisation Statistical Development Series, 1996).

Satellite imagery, and especially images from high-resolution satellites, is particularly well suited to use in combination with area frame survey techniques. The additional information they supply also concerns areas that can be related to the territory within which the sample is drawn, or to the sample itself (Deneufchatel & Porchier, 1993). However, there is a limit, while satellite imagery can complement surveys by supplying exhaustive information on the area under study, it cannot be interpreted in its entirety because a degree of uncertainty accompanies the operation of assigning the pixels to a particular type of land use. This uncertainty accounts for the need to make use of ground data to initialise an interpretation process, validate the satellite image or assess its reliability (Deneufchatel & Porchier, 1993).

Ground surveys are of great importance and their role varies depending on the techniques adopted. Their contribution can be considerably reduced if the information they supply is replaced by that which can be deduced from the use of images of the same area taken on successive dates (multi-temporal methods) (Deneufchatel & Porchier, 1993). But multi-temporal methods demand a certain level of prior knowledge of the territory under study: changes in the radiometric response of each crop during one season, nature of the principal species grown, farming practices and schedules, etc, as well as information on local weather conditions. This information must be kept in databases, or handled by expert systems (Deneufchatel & Porchier, 1993).

In short, there are many possible combinations of data collection between ground surveys and satellite imagery for implementation with area sample designs, which may be summed up as follows:

- Ground surveys on samples drawn in the traditional manner, without recourse to remote sensing;
- use of remote sensing for simple stratification of the territory as a preliminary to ground surveys;
- a combined system with a continuous mix of ground data and satellite images; and
- a pure remote sensing system, thus described because it is based on multi-temporal analyses. In such a system, ground surveys are infrequent and used only to supply the database with information (Deneufchatel & Porchier, 1993).

2.2.1.3. Multiple frame sample designs (Food and Agricultural Organisation Statistical Development Series, 1996)

A survey design that combines an area sample design with a list sample design is called a multiple frame sample design. Multiple frame estimates combine area sample with list frame estimates for each census characteristic. The results obtained from the list sample are expanded and added to the area sample estimate.

The list sample component of a multiple frame design is called the list of special holdings, and is usually formed by those holdings with the largest total area, the largest area for a given characteristic or highly specialised holdings. The list of special holdings should all be enumerated if possible. The results obtained from the list are added to the area sample estimate with no contribution to the overall variance.

A list sample is a necessary addition to an area sample in order to provide adequate estimates for important agricultural statistics. In fact, for list samples, many important agricultural characteristics have a skew distribution, concentrating a significant proportion of the total estimate in a small proportion of the holdings. For some of those characteristics, the list sample component should account for the skewness of their distribution so that the multiple frame estimates will be significantly more precise than the area sample estimates.

For multiple frame designs, any duplication between the list of special holdings and holdings included in the selected area sample segments must be eliminated from the selected segments. This operation of removing duplications of holdings requires special attention and resources. For this reason, it is particularly important in developing countries to utilise a relatively short list of special holdings, which would be feasible to inspect. Every sample survey requires the availability of a sampling frame. The sampling frame defines the target population and identifies the members that are available to be sampled.

2.2.2. The sampling frame (Garibay *et al.*, 1996)

The basic requirements of an effective sampling frame are that its sampling units, when aggregated, contain the entire population of interest and the individual sampling units are not duplicated. In other words, an effective sampling frame must be complete and its units unique. Knowledge of the characteristics of the different frame types are essential to determine their applicational advantages for a specific task, such as for this research, before a most suitable frame type can be selected. Therefore, **list and area frame types** will be investigated more closely in this section and **some practical considerations in selecting a particular frame type** considered, before a **selection of a specific frame type** can be made.

2.2.2.1. List sampling frames

One sampling method for agricultural surveys is to first compile the best list possible of farm operators, then select a sample of names from that list. This list of farm operator names becomes a list frame. A limitation of this type of sampling frame is that the target population and the survey population may be different. Compiling a complete list of every farm operator in a very large area is nearly impossible. This is especially true in developing countries as well as countries with rapidly changing economies. As soon as one farm operator dies or sells his farm to someone not on the list, the list becomes incomplete. The fact that lists become out-of-date quickly is a major disadvantage of this method.

Several types of list sampling frames can be considered as basis for the list sample survey, such as:

- A list of every animal in the area of interest. Of course, if such a list were available a survey would not be necessary (International Training, 1987).

- A list of all livestock owners and their correct address. This list would have to be constructed so that every animal would have a chance of being included in the sample. The survey process would have to ensure that each animal be associated with only one owner to avoid duplicate reporting and to ensure that none were missed (International Training, 1987).

- A list of farms or farm operators and their correct addresses. Again, it would have to be necessary that every animal have an equal chance of being included in the sample, which means that the list of farms must be complete. Special rules and procedures would be required to associate each animal with a unique farm operator so as to ensure that no animals are duplicated and none are omitted (International Training, 1987).

These three frames are not applicable to the Gauteng province, as these lists are not available.

One method used to overcome the problem of incomplete or unavailable lists is to base the survey sample on land. In this method, instead of composing a list of farm operators, a list of blocks of land is compiled. If these blocks of land completely cover the area of interest and do not overlap, then true probability sampling can be accomplished. This is because every piece of land has a known and non-zero probability of selection. Since any agricultural activity requires space, or land, all farming operations can be associated with the land they occupy.

2.2.2.2. Area sampling frames

In area frames, the target population is still all farming operations and the reporting unit is still the farm operator. However, the sampling unit becomes a piece of land. Herein lies the strength of the area frame. Since the sampling unit is a piece of land, changes in farm ownership are not a problem. Thus, the frame does not become incomplete, or out-of-date. Also, an area frame can be used to survey any item that can be associated with land.

An area frame is then a list of every section of land in the area of interest and its location. In South Africa, this list is available from the Surveyor General in the form of a land parcels layer. The land could be divided into segments with identifiable boundaries on maps or aerial photos. Every segment of land would be given a number on a listing. The coverage of the frame would be complete if all of the land areas were included in the listing. A random sample of land areas could be selected from the listing and the animals associated with each selected land area counted. The requirements of a probability survey would be satisfied, provided all animals were associated with a unique segment and not counted in more than one segment (International Training, 1987).

There are some disadvantages to area frames compared to list frames. Generally, an area sample is not as efficient as a list sample in terms of cost and precision. List samples can be selected with a narrow focus if enough prior information is already known about the farm. Usually area samples are more general. In addition, the initial cost of building an area frame is usually higher. However, this is offset because the cost of maintaining an area frame is lower.

2.2.2.3. Practical considerations for frame type selection

For an agricultural census to be conducted by sample enumeration, the following preliminary considerations can be used to compare the different types of sample designs (Food and Agricultural Organisation Statistical Development Series, 1996):

- **Level of bias.** An area sample can generate unbiased estimates since it is based on a sample from a frame that provides complete coverage of the area of interest. This cannot be said for a census based on a list sample since, in practice, a complete and perfectly updated list of holdings, valid during the data collection period, cannot be established: list frames of holdings are often incomplete and outdated. Coverage errors are a major problem in list sampling, but not in area sampling provided the rules of association linking holdings with selected segments are performed correctly.

- **Precision of the estimates.** The multiple frame method obtains more precise estimates of agricultural areas than list sample. In fact, by definition, in area sampling the probabilities of selection and the sampling expansion factors utilised are precisely proportional to agricultural areas.

- **Non-sampling errors and objective measurement of areas.** In area sample designs in which segments have recognisable physical boundaries, non-sampling areas associated with area measurements are reduced by using aerial photographs of the selected segments that clearly indicate the boundaries of the holdings and the fields. The photographs are used to check the reported area of fields and the total area of the holding it self. The holder is more inclined to be truthful when confronted with questions about specific portions of his holding that are also being observed by the enumerator at that moment. Such area sample designs provide more precise estimates than list sample designs. Many list sample designs do not require measurement of areas but rely on respondent information, which can be unreliable.

- **Durability.** If an annual agricultural survey is to be implemented using the census sample design, it is worth noting that the area frame is generally far more durable than a list frame of holdings. An area frame can be used over a period of years (5 – 10) without updating the sampling units in areas where agriculture is stable. An area frame does not become outdated unless the population extends into areas not covered by the frame. Changes in land-use, or in the number and location of holdings, may reduce the precision of the estimates but they do not introduce bias.

- **Complexity of implementation.** The implementation of an area sample design that has recognisable physical boundaries requires more technical expertise than the implementation of a list sample design.

- **Cartographic requirements.** The selection of an area-sampling frame requires accurate cartography on which to identify and measure areas. It requires the availability of suitable topographic charts, and preferably up-to-date satellite images, as well as scale-transfer and area measurement instruments. Aerial photos of the selected segments are a great advantage if objective measurement of area is required.

- **Costs.** There are high costs involved in the selection of an area sample of segments and these costs may be higher than those needed for the selection of a list sample. However such a high investment may easily be justifi ed if the samples are to be used on a regular survey basis. Lists have a short term of implementation and therefore would need to be updated all the time. If the survey must be repeated frequently, the cost of keeping the lists up-to-date would eventually be costlier than an area sample.

- **Lack of permanent boundaries.** For an area sample design, the lack of permanent boundaries in the maps, satellite image and aerial photos constitutes a serious problem. In tropical areas, because of the climatic conditions and shifting cultivation systems, boundaries change more frequently or get covered by bush and are not visible on the cartographic materials.

- **Distinguishing characteristics.** A distinguishing characteristic of multiple frame sample designs is that they have incorporated important technological advances in computer data processing to a larger extent than list sample methods. In fact, area sample methods can utilise satellite imagery or even digital satellite data as part of Geographic Information Systems (GIS), hand-held Global Positioning Systems (GPS) and generally a variety of automated procedures and techniques for sample selection and data analysis.

Considering the specific characteristics of the list and area frames as basis for sampling, together with practical considerations that should be acknowledged, it was **decided to select the area frame as a basis for this research**. This choice will affect the results obtained and need to be given careful attention before continuing.

2.2.2.4. Advantages and disadvantages of selecting the area frame sampling method

Taking into consideration the characteristics of the different sampling frame types and the practical aspects of their implementation, it is clear that, in choosing the area frame sampling method to conduct the estimation of animal numbers in Gauteng, there are several associated advantages and disadvantages, namely:

- **Advantages**

Versatility – an area frame can be used to generate statistics for any characteristics that are associated with the land. This would include houses, persons, farms, crops, animals, storage facilities, etc (International Training, 1987). Data collected from area frame surveys can also be used to verify and classify digital satellite data (Cotter, & Nealon, 1987).

Complete coverage – this is achieved provided there are no omissions of land area. This is a tremendous advantage since it provides the vehicle to generate unbiased survey estimates. Complete coverage is also useful in multiple frame (area and list) surveys where the area frameis used to measure the degree of incompleteness of the list frame (Cotter, & Nealon, 1987).

Statistically sound – an area frame provides complete coverage of the farm sector. If the segments for the surveys are selected at random, the surveys yield theoretically unbiased statistics for which precision levels can be measured (International Training, 1987).

Non-sampling errors reduced – face to face interviews generally result in better quality data being gathered. The interviewers have aerial photos showing the location of the sample segment to facilitate with the collection of data such as crop acreage within the segment. Also, if the respondent refuses to participate in the survey, the interviewer has the chance to observe some of the data and to make notes which are helpful when making nonresponse adjustments (Cotter, & Nealon, 1987).

Longevity – the area frame can be used over a period of years without having to update the sampling units (Cotter, & Nealon, 1987).

- **Disadvantages**

Less efficient than a list frame – an area frame would be less efficient for sampling purposes than an up-to-date list, stratified by type and size of farms. However, a complete list at the regional and national levels seldom exists (International Training, 1987).

Cost – an area frame can be very expensive to build and sample. Face to face interviews conducted by trained staff are also very costly (Cotter, & Nealon, 1987).

Inadequate for rare items – an area frame is not efficient for estimating rare items in a general-purpose survey. However, the area frame can be used in conjunction with a list frame for these rare items to combine the advantages of both frames (International Training, 1987).

Lack of sufficient boundaries or materials – in some regions there may be a lack of acceptable boundaries or a lack of current photography or material for developing an area frame. The quality of the area frame is reduced by these factors (International Training, 1987).

Sensitive to outliers – area frame surveys are sometimes plagued by a few extremely large operations that are in sample segments. These operations can greatly distort the survey estimates. A solution to this problem is to identify all very large operations prior to the survey (special list frame) and sample them with certainty (Cotter, & Nealon, 1987).

Although several disadvantages are linked to area frame sampling, knowledge of these can be utilised to prevent or limit the impacts of these limitations. Thus there can be much more emphasis placed on the advantages mentioned. Therefore some applicational aspects are necessary to consider before designing such an area frame for this specific research theme.

2.2.3. Applicational considerations using the Area Frame sampling design

The statistical survey method on which to base an agricultural census or survey in a given country should carefully consider the local conditions, resources and requirements. The choice of survey design should be made taking into account the trained personnel, the resources available and the desired accuracy of estimates of the principle characteristics. The sample design should also be simple enough to operate in the field with the help of available personnel. Past experience indicates that it is difficult to make adjustments for any significant deviations occurring from the sample design (missing questionnaires, errors in selection of the sample, errors in the frame, etc.) if the sample design is not simple. Also, the size of the sample has to be fixed at an adequate level so that it can be enumerated within the time limit prescribed for the statistical operations. The total cost and requirements of personnel (the number and the period), the construction of the frames and the other equipment required should be clearly assessed before any expenditure takes place (Food and Agricultural Organisation Statistical Development Series, 1996).

The survey objectives must be known before the area frame can be designed optimally. These objectives are needed to make decisions on the most effective sampling unit and reporting unit. Once the sample size, number of sampling units and survey data is known, the values are substituted in a formula known as an estimator. The **closed segment estimator** is more suited for crop estimates, while the **open or weighted estimator** is more suited for economic data. This information will also be useful in determining types of boundaries required and segment sizes (International Training, 1987).

It must be determined whether estimates are needed at national, regional or provincial levels. This determination should be made so that area frame sampling can be compared to other survey alternatives. A much larger sample size will be needed if precise regional or provincial level estimates are required. The question must be asked whether the desired precision can be accomplished within the budget and area frame constraints (International Training, 1987).

For multi-purpose surveys, the commodities that are of highest priority must be determined. This can be used to determine stratum requirements (International Training, 1987). Table 1 gives the animals of interest to the GVS and their priorities. The priorities were given by GVS experts according to their current needs regarding disease control.

Because of different distributions within the province and the amount of each commodity produced, the levels of precision will be different for different commodities. Some commodities may not be accurately estimated using only an area frame sample. These may require specialised lists and a multiple frame approach. It is very beneficial, at this point, to establish a level of desired precision for major survey items. This will have a considerable impact on the sample allocation and sample size (International Training, 1987).

Table 1. Animals with priorities and precision requirements.

Priority	Item	Description	*a Priori* data		Error (d)	
			Population Total*	Confidence (a)	%	±
1	1	Cattle	284 285	0.10	4	11 371
2	2	Sheep/Goats	230 164	0.10	5	11 508
6	3	Horses/Donkeys/Mules	19 991	0.30	11	2 199
5	4	Pigs	174 190	0.20	5	8 710
4	5	Poultry	2 475 680	0.20	8	198 054
3	6	Dogs/Cats	3 500 550	0.10	3	105 017
7	7	Ostriches	1 106	0.20	40	442
8	8	Game	7 332	0.20	50	3 666

(* Numbers supplied by Gauteng Vet erinary Services in 1997).

2.2.4. Segment size

By definition, the optimum size of the final sampling unit (or segment) has traditionally been the size that provides the most precision in the survey estimates for a given cost. The optimum size depends on a multitude of often interrelated factors such as the survey objectives, estimation method, data collection costs, data variability among segments, interview length, population density, concentration of cropland, the reporting unit, and the availability of identifiable boundaries for the segments (Cotter, & Nealon, 1987).

The optimum size for segments is very difficult to determine in practice, especially for multipurpose and multi-temporal surveys where information for a wide variety of agricultural characteristics is obtained. This difficulty is compounded when multiple estimation methods are used, not only within the same survey, but often for the same survey item (Cotter & Nealon, 1987).

Non-sampling errors also need to be considered when determining the target size of the segments. Two types of non-sampling errors are of concern (Cotter & Nealon, 1987):

– As the size of the segment decreases, the availability of suitable boundaries for the segments also decreases. This decline in good boundaries results in more reporting errors during the data collection phase.

– As the size of the segment decreases, so does the ability to delineate segments manually that are homogeneous with respect to the amount of cultivated land. Therefore, the sampling variability among segments increases for a given sample size.

According to Cotter & Nealon (1987), there are two rules of thumb that guide the decisions on what the target segment size will be. These are:

- Use the smallest size that is practical in terms of providing easily identifiable boundaries.

- Select a size for which data collection can be completed within one day under normal conditions.

In this research sampling units were not constructed from scratch. Instead, the latest (1999) existing land parcel boundaries were used. These were obtained from the Department of Land Affairs' Surveyor General office. This choice was made because funds and manpower were a limiting factor. Satellite images and/or aerial photographs needed for this exercise are too expensive and the amount of people needed to do this within a specific time frame are just not available. Another reason is that because Gauteng is very dynamic, changes in 'recognisable' boundaries often occur. Land parcel boundaries may not be recognisable on aerial photographs but the existence of a polygon theme with land parcel boundaries eases the plight. Even these parcel boundaries are soon out of date as farms are subdivided for new settlements or for new commercial zoning. The land parcels and therefore the area frame will have to be updated at least every five years to remain accurate.

2.2.5. Conclusion

Although a complete enumerate census always delivers the best data, it is not a viable option to acquire data on animal numbers and distribution via this means on a regular short term basis because cost and labour are always practical limitations.

Therefore, institutions such as GVS will have to make use of sampling techniques to obtain such data on a regular basis for proper management of animal numbers, distributions and related diseases, vaccination programs, production estimates, etc. The results delivered by using sampling techniques can have varying accuracy. The accuracy of these estimates is highly dependent on the specific design of the sample survey. Accuracy has a cost however and therefore the type of sample design must also consider factors such as the size of the area, the specialised skills of enumerators required and the particular accuracy levels required by the end-user.

Taking into consideration the above-mentioned, it was decided to adopt the area frame sampling design in combination with an area-sampling frame to define the sampling units. This design then makes use of existing land parcel boundaries as sampling units to investigate several animal objects, as selected by GVS as important study items.

In the following chapter this chosen sampling design will be developed further in detail to be applicable for this specific research problem and then combined with other steps in survey design to acquire animal estimates for the Gauteng Province.

CHAPTER 3 – APPLICATION OF PROCEDURES TO THE STUDY AREA

In the previous chapter it was decided to use the area frame sampling methodology to obtain animal estimates in the Gauteng Province. This chapter will focus on the construction of the area sampling frame, using existing land parcel boundaries and the preparation of the frame for survey purposes, which details the stratification procedure and the calculation of the sample size. The chapter then focuses on existing procedures in survey design and their application to the study area, such as the preparation of survey maps and questionnaires, the tabulation plan and the organisation of the fieldwork. The data processing of the survey questionnaires is also discussed and the estimation procedures used in area sample frames are introduced.

3.1. PREPARATION OF THE FRAMES

The sampling frame for a given sampling selection stage is the set of sampling units from which the sample is selected, together with their probabilities of selection (Food and Agricultural Organisation Statistical Development Series, 1996).

A sampling frame most often does not exist in practice. For example, a complete list of all holdings in the country is never available before conducting a complete enumeration census. A frame for each sampling stage of an agricultural survey design is a set of physical material (maps, census statistics, lists, directories, records) that cover all holdings and allow for providing suitable sampling units (Food and Agricultural Organisation Statistical Development Series, 1996).

A prerequisite for the organisation of an agricultural census, whether conducted by complete or sample enumeration, is the preparation of suitable frames. The preparation of a frame for an agricultural census conducted by complete or sample enumeration, requires a large proportion of the total effort, time and resources invested in a census program (Food and Agricultural Organisation Statistical Development Series, 1996).

In developing countries it is advantageous to use the opportunity of conducting an agricultural census to prepare an efficient sampling frame which can be utilised not only for the census but also for subsequent annual or seasonal agricultural surveys, thus improving the agricultural statistical system (Food and Agricultural Organisation Statistical Development Series, 1996).

According to Deneufchatel & Porchier (1993) satellite images can be used during the preparation of the **ground data**, but then the ground data must have the following characteristics:

- It must be possible to identify them on a satellite image; this implies their collection in the form of geographically referenced plots or points.

- They must have been obtained by means of a statistical method to make it possible to extract from the image a representative sample of pixels of the various types of land use.

The ground data are gathered by taking area samples, which may be areas in the strict sense (plots, for the most part grouped in “segments”), or dimensionless elements (lines or points), from which it is possible to construct areas, or estimate variables related to the area under study. These collection methods may be combined: lines or points may be drawn in a segment, points in a transact, and so on (Deneufchatel & Porchier, 1993).

A **segment** is a portion of territory whose form and size are selected in the light of the aims of the survey. Segments fall into two main types (Deneufchatel & Porchier, 1993):

- A segment with a form that is not defined in advance, whose limits are topographical elements that are chosen after detailed stratification.

- A segment of geometric form chosen without detailed stratification.

Within the sampling zone, the segment may itself be the ultimate sampling unit; in other words, the variable will be measured directly onto it. The survey then involves simplified mapping of each segment by drawing the outline of each plot and assigning a land use code to each plot. Measuring the area of the plots on the drawing provides an estimate for each of the categories of land use. This plot segment method is the most common, but it is not the only way of making use of segments (Deneufchatel & Porchier, 1993).

Each segment may be regarded not as the ultimate sampling unit on which the variable under study is directly measured, but as a primary sampling unit within which to draw a new sample on which the variable will be measured. This is therefore a two-stage exercise. The most convenient approach is to draw a sample of points in each segment (Deneufchatel & Porchier, 1993).

The ideal size of the segment is that which provides the greatest accuracy at a given cost. Minimising the size of the segment reduces the cost of the survey, but the following must be kept in mind (Deneufchatel & Porchier, 1993):

- Scope for position–finding in terms of fixed physical boundaries diminishes with the size of the segment, i.e. the probability that the segment will contain a stretch of a road or watercourse that is represented on a topographical map;

- the smaller the segment, the more time is wasted by the investigator in moving form one segment to another; and

- the segment must be big enough to constitute in itself a sample of the principal crops grown in the region; if it is too small, it is no longer representative, and the variance between segments will be high.

Before conducting the survey, it is thus necessary to define (Deneufchatel & Porchier, 1993):

- the variables to be studied; and
- the level of representativeness at which the results are expected (national, regional, sub-regional, etc.) on the basis of administrative divisions.

These points will be decisive in making technical decisions, particularly regarding nomenclature and sampling rate. It is important that the results obtained in the project should be drawn up, or should be able to be aggregated, at a level which enables them to be compared with other surveys or data obtained from other sources (Deneufchatel & Porchier, 1993).

The frame for this study was constructed using the following layers of information: farm[1] subdivisions[2], smallholdings[3] and erf[4] data. The Surveyor General[5] supplies this land parcel data in South Africa. These land parcels were merged, using ArcView GIS software, into one continuous layer of land parcels. The frame data was then checked using Arc/Info and ArcView GIS software so as to ensure that no land was omitted or duplicated in the merge process (Map 4, Map 5, and Map 6). Once this frame has been constructed, the stratification of the frame can be done.

3.2. STRATIFICATION

According to existing literature, stratification within a sample frame needs careful consideration of several important aspects, which will now be highlighted, before describing the actual process of stratification for this research. Stratification is done to achieve homogeneity within a frame. Stratification is a major portion of the frame development cost and the main control of the population variance (Fecso & Johnson, 1981). The stratification of a survey consists of dividing the population surveyed into homogeneous groups in terms of the variable under study, known as strata, so as to draw a sample in each group independently (Deneufchatel & Porchier, 1993).

The purpose of stratification is to reduce the variance of the variable under study in each stratum so as to obtain a lower variance over the entire area under study, or to obtain the same variance as that of a non-stratified sample

[1] Farm - A farm is a piece of rural land, falling mainly outside the urban local authority areas, although there are farms and farm portions within these areas, because all townships are laid out on either a farm or farm portion. All land at one stage or another, was farmland (Anderson, Surveyor General, pers. comm., 1999).
[2] Subdivision - a division of land, urban or rural, for which there is a diagram and in most cases, a survey lodged with the Surveyor-General's Office. It is always a portion of an erf or a farm area (Anderson, Surveyor General, pers. comm., 1999).
[3] Smallholdings - a collection of land parcels, no bigger than 25 morgen (1 morg = 0.856532 ha), represented on a general plan, which were laid out for agricultural purposes of small farmers. They are also subject to certain conditions to prevent subdivision or multiple residence. (Anderson, Surveyor General, pers. comm., 1999).
[4] Erf – a piece of urban land. It is also the collective name for lots and stands within a township or allotment area (Anderson, Surveyor General, pers. comm., 1999).
[5] The Surveyor-General's Office main function is determined by the Land Survey Act of South Africa, which states that they must examine and approve all surveys of land which are done for registration purposes. They must also archive the records, make the information available and maintain the records (Anderson, Surveyor General, pers. comm., 1999).

using a smaller sample. The variance ratio between the non-stratified sample and the stratified sample is a measure of the effectiveness of stratification (Deneufchatel & Porchier, 1993).

If the sole purpose of the survey is to estimate areas under crops, the territory must be stratified on the basis of cropping intensity. If, on the other hand, the aim is to obtain detailed knowledge of all the types of cover, agricultural and non-agricultural, very fine stratification may be less effective or even impossible if divergent stratification criteria are used. However, it can in many cases be useful to divide the territory into major types of cover: predominant arable land, predominant pasture, predominant forest land, and predominant non-agricultural uses (Deneufchatel & Porchier, 1993).

Delimiting the strata involves representing in summary form on a map the knowledge concerning land-use already acquired in previous surveys, topographical maps or aerial photography campaigns. Hence this presupposes prior knowledge of the territory, which is not always available. This is further complicated as aerial photography campaigns are only periodic, as is the updating of topographical maps and agricultural censuses, when such exist. Satellite images can be very useful in the absence of other records. As satellites are in continuous operation, satellite images can be obtained more or less on the data required, and can therefore give the most recent state of ground cover. They also offer the advantage, in comparison to aerial photography, of covering a larger area in a homogeneous manner, and with less geometric distortion. As a result of the digital nature of satellite images, a variety of processing options can be carried out which highlight landscape features of interest for a particular application (Deneufchatel & Porchier, 1993).

Although interpretation of satellite images allows very detailed stratification into ecological or agricultural zones, dividing up the territory into too many very small units must be avoided. The ultimate purpose of stratification is not to map agro-ecological zones but to allow more effective sampling. Except for the case of specific studies on small areas (hillside vineyards, valley bottoms used for horticulture or rice growing, etc.), the strata should not be smaller than around 2000 km^2. Strata that are too small require increases in the sampling rate that cancel out the cost-effectiveness of stratification, are difficult to handle, and are hard to use with satellite images. In the context of a combination with remote sensing, very detailed stratification is of little use. The territory has already been divided into small areas resulting from the intersection between administrative boundaries and those of satellite images. Additional stratification on the basis of type of land would produce still smaller zones of analysis, which could not be used (Deneufchatel & Porchier, 1993).

The number of sampling units in each stratum depends on the budget of the study and the constraints of remote sensing. The number of sampling units per stratum should permit reliable calculation of the coefficient of correlation, and this excludes tiny strata in which the sampling rate would have to be increased, which would be contrary to the purpose in view (Deneufchatel & Porchier, 1993).

Sampling rates generally adopted, range between 0,5 % and 2 %. In theory they should be modified in the light of the variance within each stratum in order to achieve the best possible allocation. This presupposes use of the results of an earlier similar survey, which is not always possible (Deneufchatel & Porchier, 1993).

Land-use stratification is the division of a land area into broad land-use categories. Stratification in this way makes for more efficient sampling and estimates for livestock, crop and farm counts (Dunleavy, 1988). Although certain parts of the process are subjective, precision work is required of the personnel stratifying the land to ensure that overlaps and omissions of land area do not occur and that land is correctly assigned to land-use categories (Cotter & Tomczak, 1994). Stratum boundaries should not, under any circumstances, split a field or sampling unit with a boundary (Dunleavy, 1988). By stratifying the land according to its use, homogeneous groups are created, that can be sampled at lower rates, thus reducing field and office workloads. Lower sampling errors are also achieved, thereby giving more confidence that the sample provides reliable estimates of the population totals (International Training, 1987).

Before a logical procedure is developed to arrive at the optimum stratification, it should be kept in mind what types of surveys the frame will be used for and what kind of statistics must be produced. If wheat were of interest, for example, the frame would be constructed to maximise its efficiency for wheat, with little regard for other items, such as livestock. In this case only two strata, wheat and non-wheat, may be needed, thereby greatly simplifying the design. It would be rare, however, to find such a singular interest in the agricultural sector (International Training, 1987).

Usually data is needed for many components of agricultural production and there are usually limitations on funds and resources. This means that the design must support multi–use or general-purpose surveys. While this permits greater utilisation of the frame, it also requires more consideration in the design, to assure that it meets various sampling needs (International Training, 1987). It is normally impossible to place all land areas into the most appropriate stratum. In most cases, small areas of differing land will have to be merged into a single stratum. (Dunleavy, 1988). Stratification for more than a few specific commodities is difficult. A sample allocation, which is optimal for one commodity, could reduce the efficiency for other commodities (Fecso *et al.*, 1986).

Considering these opinions in the literature and taking into account the limitations of existing data on these aspects in South Africa (no previous research data available, outdated maps and aerial photographs), **it was thought best to make use of the South African land-cover data to derive strata for this research**.

3.2.1. South African land-cover data

The South African land-cover data represents a standard hierarchical framework for the classification of remotely sensed data and it has been specifically designed to fit the South African environment. The framework is based on known land-cover classes that can be derived from high resolution, remotely sensed data such as SPOT or LANDSAT TM imagery, and links to existing classification systems or codes that have been used within various organisations e.g. the FAO and USGS. The South African land-cover data, derived from LANDSAT TM, covers the entire country and is provided in various GIS formats at 1:250 000 scale (Thompson M, 1996).

The land-cover classification does not contain a mix of both land-cover and land-use categories. Land–cover and land-use are closely related parameters, but they are not the same and it is important to distinguish clearly between the two in any given classification design. Land-cover refers to all the natural and man-made features that cover the earth's immediate material surface. Land-use refers to the human activity that is associated with a specific land unit, in terms of utilisation, impacts or management practices. As such, there can only be one land-cover type associated with a point on the ground, but this may be associated with several land-uses (for example, a grassland may be used for communal grazing within a conservancy area). However, the interdependence of the two components has often resulted in land-cover being used as a major diagnostic tool in the identification of land-use, leading to a common mapping association within many remote sensing classification schemes, for example, within the USGS land-cover/use model. Because mixed land-cover/use classification schemes can result in confusing or inappropriate classification structures, there is now a universal move towards the development of separate, but parallel land-cover and land-use classification schemes and data sets (Thompson M, 1996).

Map 7 shows land-cover types occurring in Gauteng. Figure 1 shows the land-cover coverage in the province, while Table 2 shows the percentage of coverage and the total hectares for each land-cover class.

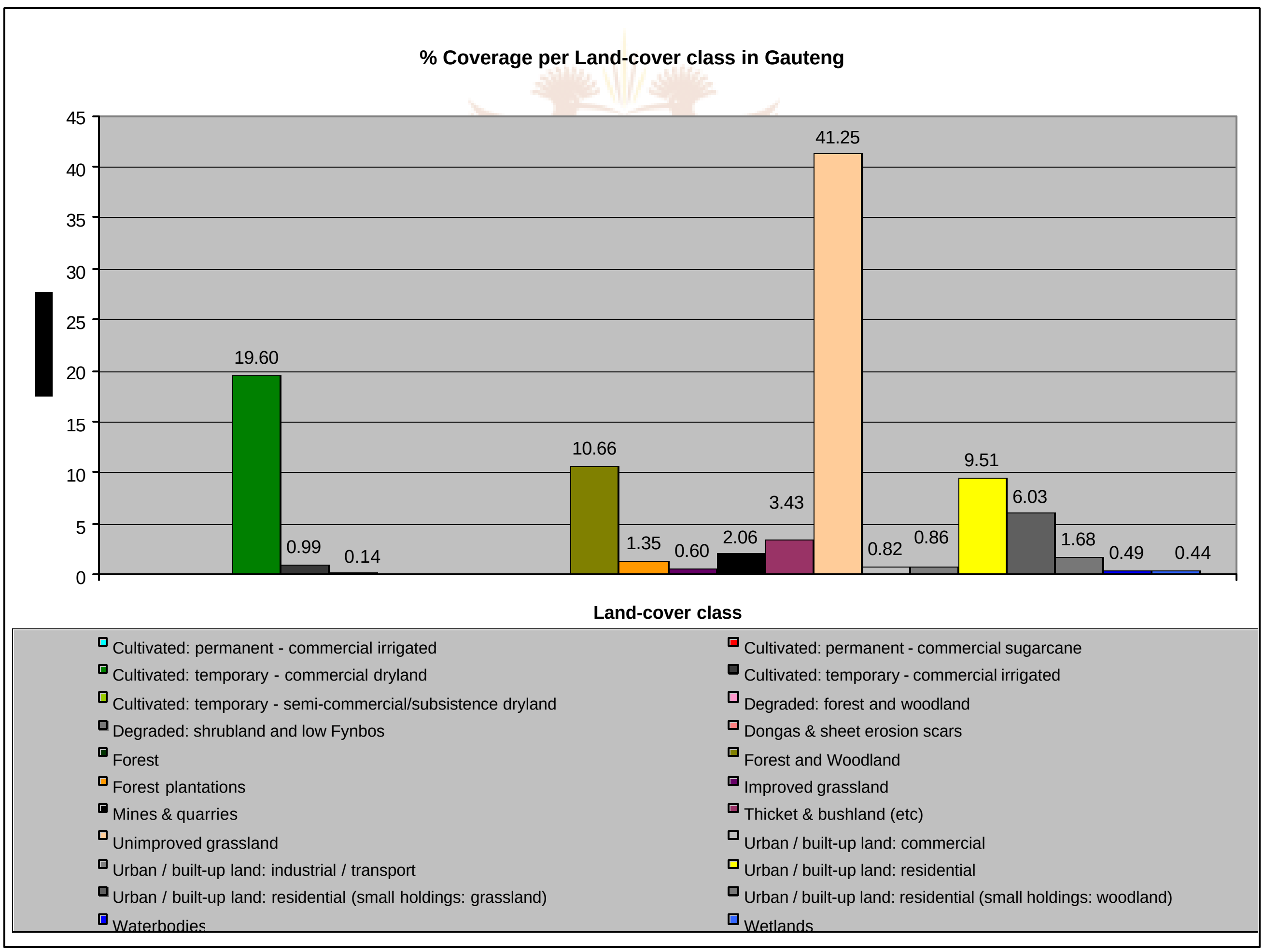

Figure 1. Land-cover classes in Gauteng province(derived from 1995, 1996 LANDSAT TM imagery)**.**

Table 2. Percentage hectares per total land-cover area in Gauteng.

AREA Hectares	LAND-COVER	% HECTARES
179 572.82	**Forest and woodland**	10.66%
0.39	**Forest**	0.00%
57 847.65	**Thicket & bushland (etc)**	3.43%
694 894.73	**Unimproved grassland**	41.25%
10 122.25	**Improved grassland**	0.60%
22 728.05	**Forest plantations**	1.35%
8 270.68	**Waterbodies**	0.49%
7 449.74	**Wetlands**	0.44%
538.45	**Dongas & sheet erosion scars**	0.03%
578.74	**Degraded: forest and woodland**	0.03%
20.59	**Degraded: shrubland and low Fynbos**	0.00%
18.13	**Cultivated: permanent-commercial irrigated**	0.00%
15.55	**Cultivated: permanent-commercial sugarcane**	0.00%
16 604.70	**Cultivated: temporary-commercial irrigated**	0.99%
330 143.44	**Cultivated: temporary-commercial dryland**	19.60%
2 412.91	**Cultivated: temporary-semi-commercial/subsistence dryland**	0.14%
160 260.91	**Urban / built-up land: residential**	9.51%
28 360.66	**Urban / built-up land: residential (small holdings: woodland)**	1.68%
101 561.95	**Urban / built-up land: residential (small holdings: grassland)**	6.03%
13 866.53	**Urban / built-up land: commercial**	0.82%
14 421.19	**Urban / built-up land: industrial / transport**	0.86%
34 770.80	**Mines & quarries**	2.06%
1 684 460.83	**ALL LAND-COVER CLASSES TOTAL**	**100.00%**

Only land-cover classes that cover a significant percentage of Gauteng's area will briefly be mentioned here. A detailed description of the land-cover classes is given in Appendix A (descriptions from Thompson M, 1996).

The **unimproved grassland** land-cover is by far the majority land-cover in Gauteng (41,25 %). This class consists of essentially indigenous species growing under natural or semi-natural conditions. It comprises areas with less than 10 % tree cover and more than 0.1 % total vegetation cover. This class is expected to carry the most animals in the province. **Cultivated: temporary-commercial dryland** is the other majority land-cover (19,60 %). This class is characterised by large, uniform, well-managed field units of crops, which are harvested at the completion of the growing season and then remain idle until replanted. A low animal count is expected for this class. **Forest and woodland** (10,66 %) are wooded areas with more than 10 % tree cover growing under natural or semi-natural conditions. As expected the **urban / built-up** land also cover a significant part of Gauteng (18,91 % for all urban classes). The urban class is expected to have a low livestock count but a reasonably high dog and cat estimate.

As this general classification contains numerous classes not contained in Gauteng and several classes that could be merged, because they will have the same effect when considered in terms of animal distributions, it was decided to reclassify these classes to fit the purpose of this study better.

3.2.2. Reclassification of land-cover for stratification purposes

The land-cover was reclassified so as to minimise the number of strata in the end (Map 8). The reclassified classes were the result of meetings with Gauteng Veterinary Services. The classes were chosen by looking at each land-cover class and deciding if the livestock distribution would be significantly different from that of any other class. For example, cultivated land will have a different livestock distribution than open vegetation or residential areas. The decision was based on the knowledge of livestock distribution from GVS, since no previous statistics were available.

The **urban: residential** reclassified class is made up of the urban residential land-cover class since this class was viewed as being distinctive on its own. The **cultivated land** class is made up of all the cultivated land-cover classes except subsistence. It was felt that the subsistence class would fall better within the **urban: smallholding** class. The **animal deficient** class includes the mines & quarries, urban: commercial, urban: industrial/transport, waterbodies, wetlands and donga classes. Although it is acknowledged that this class contains animals (especially game in wetlands), the term “animal deficient” was used for want of a better description. This class was sampled but at a less accurate sampling rate because it was not considered as important, with regard to livestock, as the other land-cover classes. The **vegetation** class includes the leftover land-cover classes. The area in hectares and the percentage coverage of the new reclassified classes is shown in Table 3 and Figure 2.

The reclassification was done using ArcView GIS by assigning each land-cover class a new reclassified code and then dissolving (or grouping) using the new code.

Although the vegetation class seems extremely large when compared to the other classes, it is further subdivided during the next stage according to land parcel size (defining the strata).

Table 3. Area of reclassified land-cover and the original classes making up the new class.

RECLASSIFIED CLASS	ORIGINAL CLASS
Vegetation	Degraded: forest and woodland
966 303.66 ha	Degraded: shrubland and low Fynbos
57.37% of total area	Dongas & sheet erosion scars
	Forest
	Forest and Woodland
	Forest plantations
	Improved grassland
	Thicket & bushland (etc) Unimproved grassland
Cultivated Land	Cultivated: permanent - commercial irrigated
346 781.81 ha	Cultivated: permanent - commercial sugarcane
20.59% of total area	Cultivated: temporary - commercial dryland
	Cultivated: temporary - commercial irrigated
Urban: Residential	Urban / built-up land: residential
160 260.91 ha	
9.51% of total area	
Urban: Smallholdings/Subsistence	Cultivated: temporary - semi-commercial/subsistence dryland
132 335.52 ha	Urban / built-up land: residential (small holdings: grassland)
7.86% of total area	Urban / built-up land: residential (small holdings: woodland)
Animal deficient	Mines & quarries
78 778.94 ha	Urban / built-up land: commercial
4.68% of total area	Urban / built-up land: industrial / transport
	Waterbodies
	Wetlands

3.2.3. Defining the strata

The strata were defined using two criteria i.e. land parcel size and land-cover. Land parcels were divided into three size classes namely small, medium and large. These classes were also decided on in various meetings with GVS. It was decided that parcels smaller than 25 ha would be classed as small. Parcels between 25 ha and 400 ha would be medium, while parcels with an area greater than 400 ha would be classified as large (Map 9). Again it must be emphasised that this division was subjective as no previous statistics exist.

To define the final strata, the two criteria (reclassified land-cover and parcel sizes) were combined using ArcView GIS. At first it was necessary for each land parcel to be assigned to a reclassified land-cover class according to majority of area covered by a specific land-cover class. In other words, a specific land parcel might have two or more land-cover classes covering it. For example it might have both cultivated land and vegetation classes. If cultivated land took up 100 ha and vegetation took up 300 ha, it was assigned to the vegetation class because it took up the majority of the area. This was done to avoid having too many polygons in the area frame.

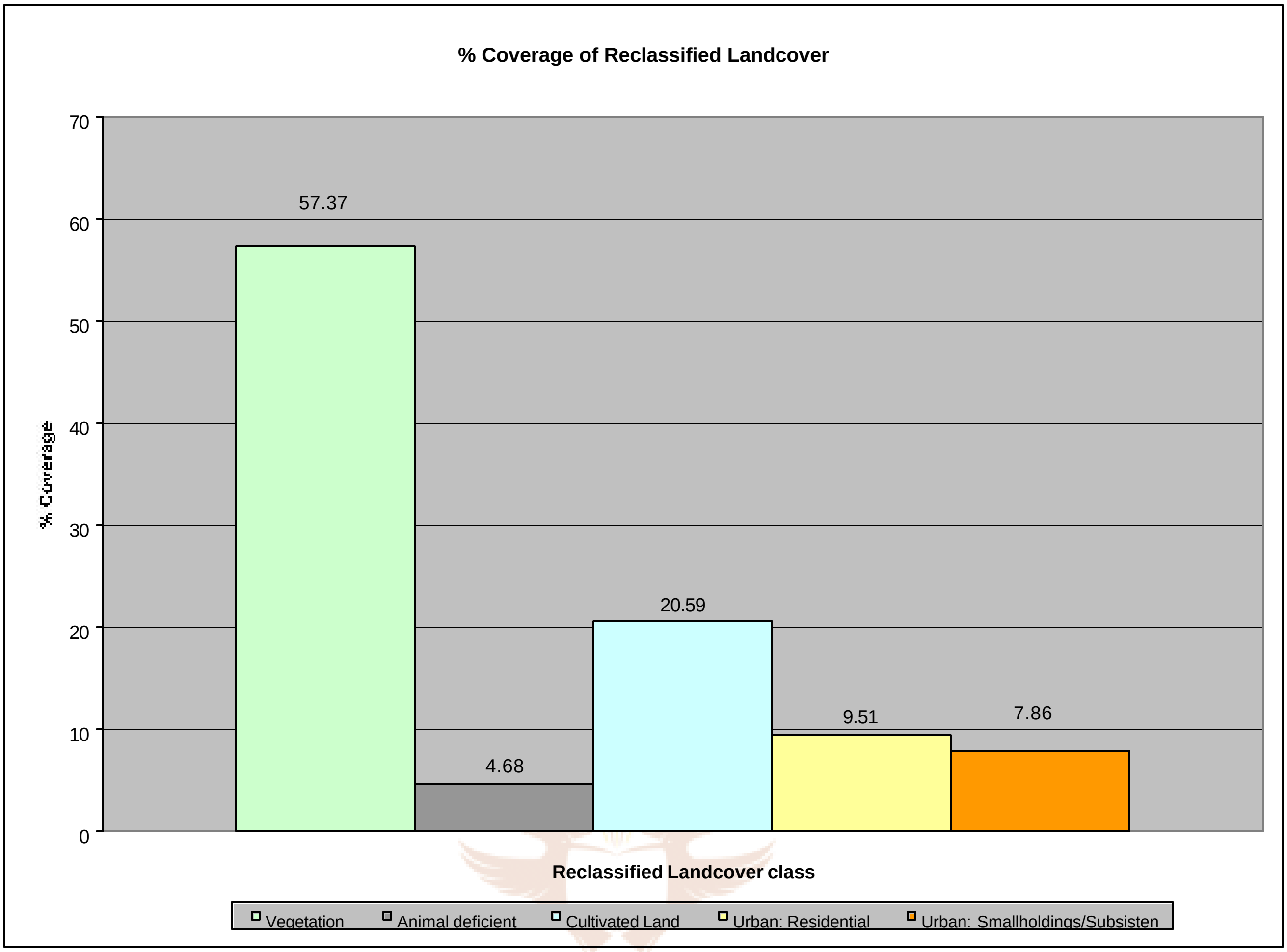

Figure 2. Reclassified land-cover in Gauteng province (reclassified from 1995, 1996 LANDSAT TM images).

Simple queries were then run in ArcView GIS to assign a stratum value to each land parcel. An example of such a query would be "select all the land parcels between 25 ha and 400 ha with a vegetation land-cover." The parcels selected in this query would be assigned to the "vegetation with parcels 25 – 400 ha" stratum (Table 4 and Map 10).

As can be seen from Table 4 and Figure 3, the vegetation strata with parcel size between 25 – 400 ha and greater than 400 ha were the largest strata. The smallest was the animal deficient strata, which includes the urban industrial and commercial areas. An attempt was made to keep the area coverage more or less equal between the different strata, to keep the accuracy of the estimate as reliable as possible between them.

The land-cover data used to create these strata is at 1:250 000 scale. This results in a rather crude stratification when taken down to a land parcel scale. This, however, was the only data available at the time that could have been used as a stratification tool. This stratification could be improved on at a later stage when the Urban Eye Project of the CSIR's Satellite Applications Centre (SAC) is completed. This maps the Gauteng land-cover at a 1:50 000 scale and would result in a much more accurate stratification.

With the strata defined, there can be progressed to the next stage of the sample design that is the determination of the size of the sample to be investigated in the survey.

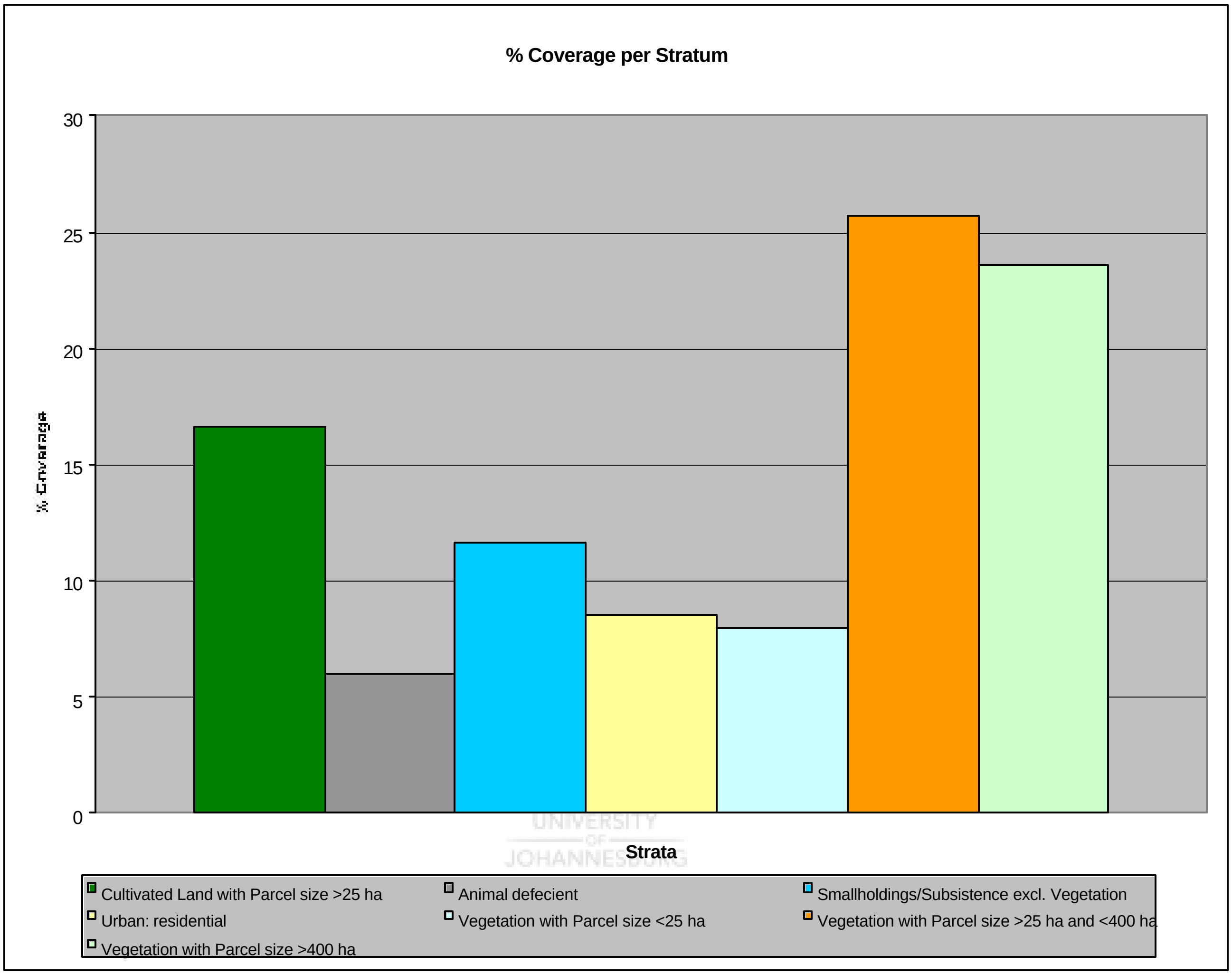

Figure 3. Area coverage of the strata.

Table 4. Final strata and area covered.

STRATUM	AREA HECTARES	% HECTARES
Cultivated Land with Parcel size >25 ha	284 843.23	16.63
Animal deficient	102 767.67	6.00
Smallholdings/Subsistence exc. Vegetation	199 221.32	11.63
Urban: residential	145 801.74	8.51
Vegetation with Parcel size <25 ha	136 331.71	7.96
Vegetation with Parcel size >25 ha and <400 ha	439 978.58	25.69
Vegetation with Parcel size >400 ha	403 723.63	23.57
TOTAL	**1 712 667.90**	**100.00**

3.3. SAMPLE SIZE

The sample size in a given situation is difficult to determine. The appropriate sample size depends on many factors, such as the efficiency of the sampling design adopted and the number of areas and classifications for which estimates are required. If the data is needed for planning at administrative subdivision level, a much larger sample will be needed; a sampling fraction of 2 – 10 % of the holdings may be adequate for providing reliable estimates for all the principle characteristics. However, if such information is needed at only the broad provincial or national level, a sampling fraction of 1 – 2 % of holdings or agricultural areas may be adequate. As a rule of thumb, no attempt should be made to make sampling estimates for areas or groups of holdings for which less than 200 – 300 sample questionnaires have been completed. A critical examination of the sample size should be made at the time of the pilot census. One of the objectives of the pilot census should be to study the variability of different characteristics and the time and cost involved in obtaining information on them (Food and Agricultural Organisation Statistical Development Series, 1996).

There are a variety of ways by which the sample may be selected. For each, a rough estimate of the sample size can be made from the degree of desired precision specified by the client. The relative costs and time involved for each sample strategy must be compared before taking a decision (Cochran, 1977).

More than one item or characteristic is usually measured in sample survey: sometimes the number is large. If a desired degree of precision is prescribed for each item, the calculations lead to a series of conflicting values for *n*, one for each item. Some method must be found to reconcile these values (Cochran, 1977). These conflicts were solved by using the animal priorities given by GVS.

To determine the sample size for the Gauteng area frame, two assumptions were considered. Firstly, if it was assumed that nothing was known about the population distribution and that money was no problem, over-sampling (sample size of 25 – 50%) could be decided on. Secondly, if it were assumed that the GVS has some idea of the population distribution and that a limited budget was available, then a much smaller sample size would be selected. In this study the latter assumption was the most appropriate and the sample extraction definition was built accordingly.

The following statistical procedure was used in calculating the sample size for each stratum (in Potgieter, Narciso & von Hagen, 2000 from Cochran, 1977). Potgieter and experts at Statistics South Africa were consulted to verify the validity of the procedure. According to Cochran (1977), if Pr is probability and $\overline{y}$ is the average of the observations from a simple random sample, it is wished to have

$$\mathrm{P}r\left(\left|\,\overline{y}-\overline{Y}\,\right|\geq d\right)=a \qquad \textbf{(1)}$$

Where d is the chosen margin of error and t is the abscissa of the normal curve that cuts off an area α at the tails. Solving for n, it is found that

$$s_{\bar{y}} = \sqrt{\frac{N-n}{N}} \times \frac{S}{\sqrt{n}} \qquad \textbf{(2)}$$

Hence,

$$d = t \times \sqrt{\frac{N-n}{N}} \times \frac{S}{\sqrt{n}} \qquad \textbf{(3)}$$

This gives,

$$n = \frac{\left(\frac{tS}{d}\right)^2}{1 + \frac{1}{N}\left(\frac{tS}{d}\right)^2} \qquad \textbf{(4)}$$

Also take as first approximation, where V is the variance,

$$n_0 = \left(\frac{tS}{d}\right)^2 = \frac{S^2}{V} \qquad \textbf{(5)}$$

This is true unless $\frac{n_0}{N} * 100$ is appreciable (in almost all examples given in research literature, the rule of thumb is if $n_0 > 5\%$ then calculate n), in which case compute n as

$$n = \frac{n_0}{1 + \left(\frac{n_0}{N}\right)} \qquad \textbf{(6)}$$

In the livestock inventory study it is necessary to estimate the population total with a margin of error d, therefore take as a first approximation

$$n_0 = \left(\frac{NtS}{d}\right)^2 = \frac{(NS)^2}{V} \qquad \textbf{(7)}$$

instead of **equ ation 5**. **Equation 6** remains unchanged.

Deming (1960) showed that some simple mathematical distributions could be used to estimate S^2 from a knowledge of the range (h) and a general idea of the shape of the distribution (Figure 4). The sample sizes are mostly large enough to approximate to the normal curve. Where this was not the case, Deming's (1960) estimate of the appropriate distribution was used.

According to Hale, (NASS USDA, pers. comm. 1999) an expert guess of h is better than no data. If no data of the distributions of the items within a region are known and one cannot guess either, an expensive over-sampling strategy should be followed. The GVS have a fair enough idea of h and the distribution of the items within the Gauteng Province.

Table 1 depicts the list of items that are to be surveyed. Eight items were decided on, with the highest priority going to cattle and the lowest to game. Other *a Priori* data like total population, confidence level (a) and error margin (d) are also depicted in Table 1. Consequently, these relations were effectively used in calculating h, t, S^2, n_0, hence n.

The final sample size to be extracted, for each stratum, is depicted in Table 5. Furthermore, N was calculated as the final total SU's within each stratum (Table 5).

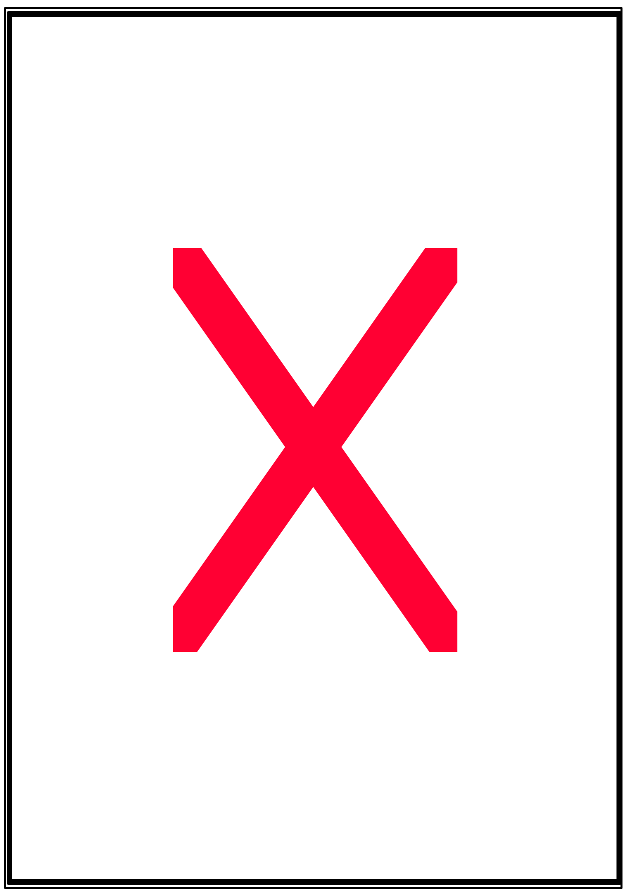

Figure 4. Some simple shapes of distributions with their given variances (Deming 1960).

Table 5. Final sample extraction size for each stratum.

Strata	Description	SU Total (N)	Sample Size (n)	Fraction (%)
1	**Vegetation with parcel size < 25 ha (exc. erfs, parks, urban areas)**	19 358	737	3.8077
2	**Vegetation with parcel 25-400 ha (exc. erfs, parks, urban areas)**	4 403	706	16.0296
3	**Vegetation with parcel size > 400 ha (exc. erfs, parks, urban areas)**	492	153	31.1497
4	**Cultivated land with parcel size > 25 ha**	1 873	158	8.4337
5	**Smallholdings/subsistence (exc. vegetation)**	39 883	208	0.5225
6	**Urban-residential**	1 528 857	204	0.0134
7	**Animal deficient (incl. waterbodies/parks/mines/urban-industrial)**	8 215	164	1.9949
	TOTAL:	1 603 081	2 331	0.1454

Map 11 shows the distribution of the selected samples. The land parcels in each stratum were selected by running a random selection script on each stratum individually. The script was run within ArcView GIS directly on the polygon land parcel theme. The script was written by ESRI (ArcView GIS developers) and it is included in ArcView as a sample script. It can be seen in Appendix B.

These selected sample points had to be clearly defined and labelled in order to direct enumerators to a particular location. For this purpose survey maps need to be compiled for usage as set out in the next phase.

3.4. CARTOGRAPHIC PREPARATION

Survey maps to be used by enumerators were prepared using ArcView GIS. After the random selection of land parcels to be sampled was made, the centroid co-ordinate of each parcel was calculated and included on the survey form. This co-ordinate was used to locate each land parcel. Each of the selected land parcels was then labelled, using a unique number, to keep track of the survey forms when they came back and were entered into the database. The 1:50 000 grid was then overlaid on the selected parcels. Size A3 maps were then printed per 1:50 000 map sheet, per stratum for easy reference. For orientation purposes, 1:50 000 road data were also included on each map sheet. A total of 280 maps of A3 format were printed for survey purposes. Map 12 shows an example of the survey maps.

These survey points must now be surveyed in a specific manner by using a specifically designed questionnaire.

3.5. **SURVEY QUESTIONNAIRE** (Food and Agricultural Organisation Statistical Development Series, 1996)

Once the decision on the scope and coverage of the census has been taken, a questionnaire can be developed in order to secure the relevant information in an orderly and co-ordinated manner. The census questionnaire is the most basic document in the census programme since it becomes the vehicle for collecting the desired information. Any deficiencies in the questionnaire design will lead to incomplete and inaccurate data being collected. Considerable thought was given to the design and formulation of the questionnaire and input was sought from available experts on this subject matter.

Generally three methods (self-enumeration, interview method, objective measurements) are considered for obtaining census data. Objective measurements of areas and yields are used in countries where holders are not familiar with standard units of measurement and where cadastre information does not exist.

The format of the questionnaire will depend on the method of inquiry, whether by interview, by self-enumeration, intended for objective measurement or a combination of these methods. The self-enumeration method was preferred for this survey because there is no link between landowner or land user and the actual land parcel. It would therefore be less time consuming for the enumerator to visit the land parcel using the given co-ordinates and A3 survey maps, than to find out who, or where, the land owner is located.

The size and form of the questionnaire are the first items to be considered. The questionnaire should not be too large and should be easy to handle in the field. Its size and shape should be such that the enumerator can easily handle it in the field while recording the respondents' answers.

The temptation to ask a great number of questions should be resisted. Even if it is argued that once the holder has been contacted, maximum advantage should be taken in collecting the necessary information, as it is more costly and time consuming to meet him/her than to obtain data through the mail. However, this argument is not valid for many reasons. Frequently the data requested is not readily known to the holder, who may need to consult records, which takes time. Furthermore, if the questionnaire is lengthy, the holder, who at the outset is prepared to reply to the questions, may become less co-operative after being questioned for a long time. It is therefore very important that the questionnaire is not too lengthy. As a general rule, an interview should not exceed 45 minutes. The questions should also be formulated in a clear simple language, using, wherever possible, the vocabulary familiar to the holder.

Attention should be given to the quality of the paper used. Thin paper should not be chosen, because during the fieldwork the questionnaire is often subjected to very unfavourable climatic conditions and to constant handling during the distribution and subsequent tabulation of the data.

The size of the print should be easily read even when light is not adequate. This occurs frequently in rural areas when the holders are interviewed in the evening, since in many rural areas there is minimum lighting. Often, in order to keep questionnaires to a reasonable size, a small print is used which is unacceptable. Effort should also be made to use a different type of print for the questions and for notes or instructions to the enumerators. The questions must be easily distinguished since most of them will have to be read aloud to the holder. However, very heavy print should not be used, as the questionnaire will look overloaded. The space for replies should be made large enough so that there is room for responses to be entered and the lines should not be printed too close together.

The questionnaire for the survey was developed following the above FAO guidelines in conjunction with GVS. The questionnaire was kept as simple as possible so as to ensure that the enumerator spends the least possible time on each land parcel. An example of the survey questionnaire is given in Figure 5. The information on the top three lines of the questionnaire was provided using the land parcel data and therefore did not have to be filled in again.

The information contained in the completed questionnaires then have to be made available in some easily accessible format to enable further processing of the data. This is normally done through some form of tabulation.

26/08/1999 **Gauteng Animal inventory** **PAGE 1**

PARCEL ID	**SG_CODE**	**FARMNAME**	**X**	**Y**
16698	T0JR00000000042000073	DOORNKRAAL 420 JR	28 31 55	-25 35 25

DO YOU HAVE ANY OF THE FOLLOWING ANIMALS ON THE LAND PARCEL?

	YES	**NO**	**IF YES, HOW MANY**
CATTLE	X		91
SHEEP	X		160
GOATS		X	
HORSES	X		2
DONKEYS		X	
MULES		X	
PIGS		X	
POULTRY		X	
DOGS?CATS	X		8 DOGS, 2 CATS
OSTRICHES	X		4
GAME		X	
OTHER		X	

Figure 5. Example of survey questionnaire.

3.6. TABULATION PLAN

The database for the collection and input of the survey questionnaires was done in Microsoft Access. The database consists of three tables namely, land parcels, inventory and animal. The tables with the field definitions are shown in Figure 6. The relationships between the tables in the database are given in Figure 7. This set of linked tables contains all the data about the selected land parcels within specific survey strata and will also contain the information on animal counts once the field work is done by the enumerators.

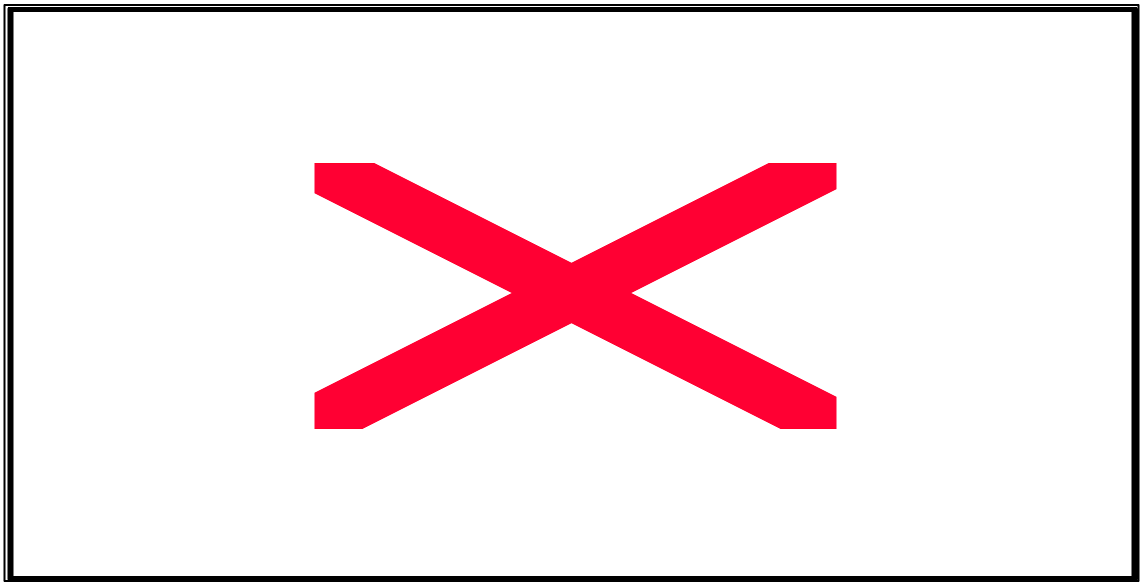

Figure 6. Tables and field definitions in the survey table.

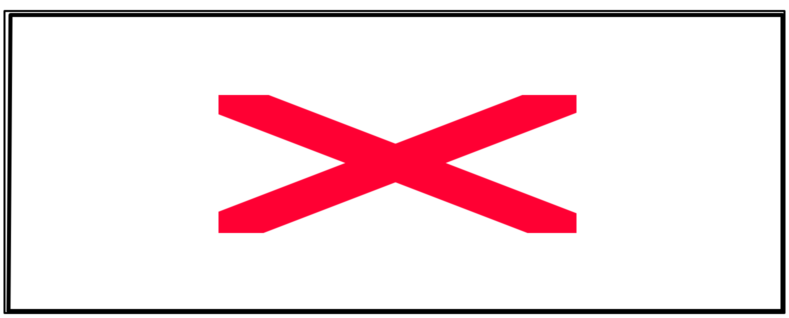

Figure 7. Relationships between database tables.

3.7. ORGANISATION OF FIELDWORK

Guidelines for the organisation of the fieldwork were taken from Deneufchatel & Porchier, 1993.

The size of a segment to be surveyed must be such that the segment can be surveyed in less than a day, including travel time. The ideal size is that which enables an investigator to complete work on two segments per day. This ideal size varies from one region to another, since the time spent depends on:

- Accessibility: existence and standard of the road network;
- the ease of finding plot boundaries on the ground;
- the number of plots; and
- the complexity and clarity of the nomenclature and the instructions provided to the investigators.

The survey calls for two categories of topographical records:

- Those which enable the location of segments to be found on the ground; and
- those that guide the investigator during the ground survey.

In order to enable the investigator to locate the segments on the ground, a topographical map or a road map on the scale of 1:200 000 is adequate. The segments must be identified on the map, without seeking an excessive degree of accuracy. It is useful to mark on the map any notable topographical features that are not represented, such as buildings, non-surfaced roads that the investigator can use, etc.

On approaching the segment, the investigator must possess a very accurate topographical document which will enable him, whether he wishes to locate points or outliner plots, to find his position to within a few metres.

In both the point and plot segment methods, the investigator is equipped with:

- The basic topographical document in which the outline of the segment is drawn;
- a topographical map on which the location of the segment is indicated;
- as appropriate, such instruments as a compass or a GPS; and
- a survey sheet for each segment.

In the case of point segments, the basic topographical document also bears the grid of points. The points take the form of dots or crosses. It is not necessary for the outline of the segment to be represented.

The sheets in the two methods differ in that the sheet used in the point survey is “closed”, with numbered lines corresponding to the number of points. The sheet for the plot survey is “open-ended”, with an unlimited number of lines, since the number of plots is not known in advance.

The plot survey involves covering the whole of the plot and noting all the categories of land use to be found in it, sketching their location, generally on a transparent sheet with a fine-point indelible felt-tipped pen, and transferring the number of each plot and its cover onto the sheet (Deneufchatel & Porchier, 1993).

For the point survey, the investigator must:

- Locate, or relocate, the exact position of each point, and
- observe the portion of territory denoted by each point and designate it in accordance with the land use nomenclature.

An essential component of a point survey is the stability of the points. From one year to another, the investigator must be certain that he is continuing to observe the same points, so as not to introduce changes which would in fact constitute errors of observation. Every stable element which can pin down the location of the point must be mentioned in the survey documents (distance from a noteworthy tree, a building, and so on).

Points located on plot boundaries give rise to many cases of ambiguity. The usual method for solving this problem is to select at random one or other of the two adjacent plots. The choice is made not by the investigator himself, but in the office, by the person in charge of the survey, in order to avoid the frequent bias in favour of agricultural uses of the land. The choice is made once and for all to cover the entire duration of the survey, so as to keep the sample stable.

Another contentious case involves types of land use, which are heterogeneous. These types of cover are found in certain farming areas, but especially in transitional zones between farming uses and non-farming uses or between fallow land and forest. Other cases are mixed orchards, crops or pasture under fruit trees and so on.

The survey maps were distributed to GVS who in turn distributed the maps to the veterinary technicians in each district. Each technician then visited the land parcels allocated to him. The technicians could locate the land parcel using the survey maps in conjunction with 1:50 000 topographical maps and the co-ordinate of the land parcel.

The actual survey was carried out between September and December 1999 by veterinary technicians with people from GVS helping out where possible. Future surveys should strive to collect data during the same general time period.

3.8. QUALITY CHECKS

Guidelines for quality checking were taken from Deneufchatel & Porchier, 1993.

Quality control must be exercised on:

- The quality of the drawing of the plots, where appropriate;
- the geographical location;
- identification of the cover; and
- interpretation of the instructions in contentious cases.

The risk of a poor quality drawing arises only for plot segments, and this is a drawback compared with the point survey. It is hard to recruit investigators who can both make a good drawing and correctly identify crops.

An error in geographical location occurs when the investigator makes a mistake concerning his location, or concerning the width of the plot. In this way a point, a plot or part of a plot may be poorly identified not because the investigator has incorrectly identified what he has seen, but because he has not seen what he should have seen. He may, for example, have located a point in a wheat field when it was in barley field. This type of error is difficult to identify as such, since the error is often accompanied by an identification error. What appears to be an error of identification is often an error of location.

Errors in interpretation of the instructions are much more frequent with plot segments. They arise from the simplifications made necessary by the drawing process. No two investigators will simplify or regroup the results in the same way.

By its very nature, a plot segment is more susceptible to error than a point segment. Since a grid comprises a specific number of points, it is possible to check that there is a cover code that corresponds to each point. In the case of a plot segment, if a plot has been overlooked, this means in practice that, if no number has been given to it, since the total number of plots is not known, the oversight can pass unnoticed.

The different types of error are difficult to identify, and it is possible to measure only an overall error, grouping together errors of location and errors of identification, possibly with degrees of seriousness attached to the errors of identification.

In a point survey consistency with the previous year's survey can be checked. Highly unlikely successions at one point – for example, wheat that becomes forest – may be noted and corrected, or, exceptionally, confirmed. This is not possible with plot data, since a plot is often divided or merged with neighbouring plots from one year to the next, so that a particular plot cannot be permanently given a number. Attempts to keep record of "parent plots" and "offspring plots" have run up against insurmountable difficulties simply because of the complexity of the task.

The simplest checking method is to have the survey carried out by two different investigators. However, in the event of a divergence, it is impossible to tell which one is in error, unless a third survey is carried out for the purpose of analysing the divergences. Consequently, the method is limited to counting the divergences between the two investigators. The number of errors is deemed equal to half the number of divergences, the assumption being that the two investigators are in error equally often. If the investigators are both in error (for example, if all the investigators have misunderstood an instruction), the error will pass unnoticed.

The rate of error must be evaluated in the light of the nomenclature. It will be higher if the nomenclature is unrealistic and if the investigator is asked to differentiate between types of cover, which is difficult or impossible to distinguish on the ground. For example, grain maize cannot unambiguously be distinguished from maize for silage if the investigator makes only one visit.

Care was taken to ensure that the co-ordinates of the selected land parcels were correct to ensure easy location. It seemed plausible that the land parcel encountered by the enumerator might have been developed or might now have a different land-cover type. This might happen because the land-cover data used is based on 1995 imagery and development or degradation might have occurred in certain areas. When this happened, it was decided to keep the land parcel in its assigned stratum, but noting on the questionnaire the change that occurred, so as to assign the land parcel to a different stratum in future surveys.

3.9. DATA PROCESSING

Questionnaires completed and returned were input into an Access database in a format that would be ready to put into the estimate procedure using the form in Figure 8.

LandParcel ID refers to the GIS code assigned to each land parcel as a unique identifier. Sg_code is the Surveyor-General 21 character code, while X and Y refer to the longitude and latitude. These three fields all help to identify the spatial locations and uniqueness of each land parcel. The contact field is for any contact person that is associated to the land parcel, while the comments field leaves space to report any changes in land-cover or subdivisions of the land parcel. These two fields are not critical for the running of the estimate. Provision is made, by means of drop-down lists, to report on the animal type and amount of each animal occurring on the land parcel. Provision is also made for comments about the animal i.e. specific type of species.

Once all the data has been processed and is in the database, the estimation procedure can be done.

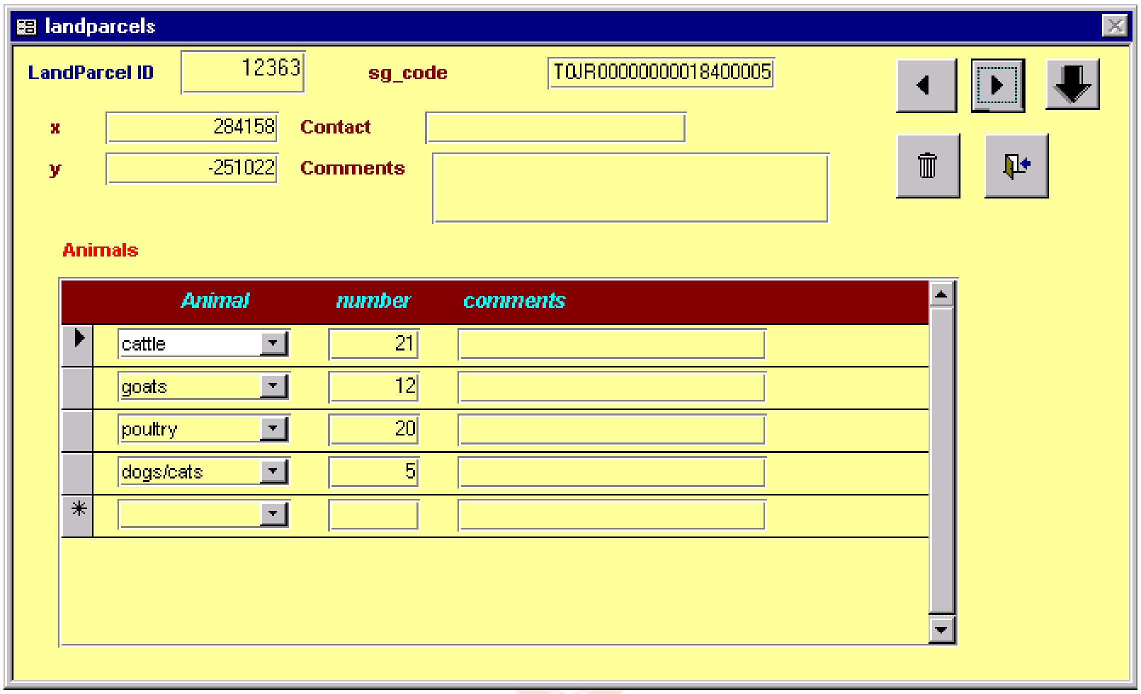

Figure 8. Example of the input form.

3.10. ESTIMATION PROCEDURES IN AREA SAMPLE DESIGNS

Since sampling units are land areas that may not coincide with the land of a holding, it is necessary to establish a criterion to associate each segment with a holding. This is so that a census characteristic value can be assigned to each segment as a function of its value in the associated holding. Then, a uniform expansion factor is applied in each stratum to obtain the census estimate for the characteristic (Food and Agricultural Organisation Statistical Development Series, 1996).

The choice of a reporting unit is important because it establishes a point of reference to associate crops and livestock with a unique segment. Three different reporting units can be used (International Training, 1987):

– Closed – land within segment boundaries associated with a unique operation. All livestock, regardless of ownership on the tract of land, are reported.

- Open – all land, both inside and outside the segment associated with a unique operation. The "farm" is associated with the segment containing the residence of the farm operator. For example, all livestock on the "farm" are reported regardless of ownership.

- Weighted – all land, both inside and outside the segment, associated with a unique operation. Data for the entire farm are prorated to the sample segment. The proration or weighting is usually done using the ratio of the segment area to the entire farm area. Reporting unit is the same as the open approach except data is collected for the entire farm even when the residence is not inside the segment.

The closed segment approach requires that the enumerator obtain data for only that part of the farm within the sampling unit, while the open and weighted segment approaches require that the enumerator obtain data on the entire farm (Ford *et al.*, 1986).

For agricultural censuses based on an area sample with segments that have recognisable physical boundaries, two types of direct expansion methods have been used to associate segments with holdings and to expand the results obtained in the holdings associated with the selected sample of segments (Food and Agricultural Organisation Statistical Development Series, 1996).

The direct expansion estimations refer to the way the holding data are distributed among the segments before they are expanded by the inverse of the sampling fraction. The direct expansion methods used are called the weighted segment method and the open segment method (Food and Agricultural Organisation Statistical Development Series, 1996).

The following example from Ford *et al* (1986) illustrates the difference between the three.

Suppose the following situation occurs for a specific farm: land parcel acres = 10, farm acres = 100, hogs on the land parcel = 20, and hogs on the farm = 40. The closed segment value of number of hogs would be 20; the weighted segment value would be 40 X (10/100) = 4; and the open segment value would be 40 (if the headquarters is in the land parcel) or 0 (if the head quarters is not in the land parcel).

3.10.1. The closed segment estimation method

The "closed" concept is easiest to execute because the interviewer and the respondent can observe the photograph and pinpoint the segment. The "closed" procedure is useful for estimating crop acreage and livestock inventories. However, it is difficult or often impossible to associate economic characteristics, such as value of sales, with a tract of land. To estimate economic characteristics it is necessary to identify an entire farming unit (International Training, 1987).

If stratification is good and parcel sizes are approximately equal then, in repeated sampling, the closed segment estimator will have good precision for estimating total area (Faulkenberry & Garoui, 1991).

3.10.2. The open segment estimation method

The use of the “open” concept is more difficult because it requires special rules to associate each farm with a unique segment. Sampling variability is generally greater for the “open” concept than the “closed” concept because the segment size puts an upper bound on the number of animals or crop acres planted in a tract. There are no such bounds when using the “open” concept. The optimum segment size for the “open” concept is that it must generally be large enough to contain about one farm headquarters. This may allow for a larger segment than when using the “closed” method (International Training, 1987).

The direct open estimation method involves the following steps (Food and Agricultural Organisation Statistical Development Series, 1996):

- The open segment method associates to the segment all holdings with headquarters inside the segment. For this purpose, clear rules have to be established to define a unique reference point for each holding, called the headquarters. There are several ways to do this but the most common procedure is to define the headquarters as the residence of the holder. In this case, a one-to-one correspondence between holding and headquarters can be established, with an additive rule in case of joint holders living in different dwellings.

- The value of the characteristic in the segment is defined as the sum of the characteristic in each of it’s associated holdings, i.e. the holdings with headquarters inside the segment.

- The census estimate of the characteristic in the stratum is the sum of the characteristic in all its segments multiplied by the expansion factor of the stratum.

- The census estimate of the characteristic is the sum of the characteristic in all strata.

Open segment estimators are not generally recommended due to the difficulty that arises in identifying and locating the residence of the holders in urban areas (Food and Agricultural Organisation Statistical Development Series, 1996).

3.10.3. The weighted segment estimation method

The “weighted” segment method is a compromise between the “open” and “closed” concepts. Entire farm data are collected and prorated to the tract. The primary disadvantage of this method is the increase in enumeration time required. The sampling variability is lower than the “open” method because it spreads the larger operations over many segments. However, non-sampling errors may be greater with the "weighted" concepts due to difficulties in obtaining accurate weights. Because of cost considerations, the optimum segment size will be smaller than for the “closed” and “open” procedures (International Training, 1987).

The direct weighted segment estimation method involves the following steps (Food and Agricultural Organisation Statistical Development Series, 1996):

- The weighted segment method uses all holdings with a tract in the segment, by associating to the segment all holdings with any land inside the segment.

- The value of a given segment characteristic in each tract (e.g. the number of cows) is defined as the value of the characteristic in the holding multiplied by a factor equal to the ratio between the area of the tract divided by the area of the holding. For instance, if 10% of a holding's area is in the tract, 10% of the total number of cows of the holding are assigned to the tract.

- The value of the characteristic in the segment is defined as the sum of the characteristic in each of its tracts, as defined above.

- The census estimate of the characteristic in the stratum is the sum of the characteristic in all its segments multiplied by the expansion factor of the stratum.

- Finally, the census estimate of the characteristic is the sum of the characteristic in all strata.

The weighted segment estimator does not require establishing precisely who is the holder, or the holdings headquarters, but it is necessary to identify the holding, calculate it's area inside the selected segments, and find an informant who can provide data for the total holding (Food and Agricultural Organisation Statistical Development Series, 1996).

After considering the estimators described above and consulting with necessary experts, **the closed estimator is chosen for the Gauteng animal estimate**, since this is the only possible choice if the data surveyed are referred to the land parcel (and not, for example, to a farm that includes several land parcels).

3.10.4. Sampling efficiency

Sampling efficiency refers to the precision and cost of the estimators. The precision of an estimate can be measured by: 1) the variance of the estimate, 2) the standard error (se) which is the square root of the variance, or 3) the coefficient of variation (CV) which is the standard error divided by the estimate. An estimate becomes less precise as any of these measures increases (Ford *et al.*, 1986).

Given the same number of land parcels to make each estimate, weighted segment estimates are usually more precise than closed segment estimates, and closed segment estimates are usually more precise than open segment estimates (Ford *et al.*, 1986).

The CV is one major tool for evaluating the quality of the estimates. It measures the precision of an estimate, but not the accuracy. A low CV shows that the estimator has very little variation relative to the point estimate and is precise. Conversely, a high CV means that the estimator has a wide confidence interval and that the estimated value could change greatly given a different sample (Garibay *et al.*, 1996).

According to GUSS/NASS Project (1995), estimates should have CV's less than 10 % for items of major importance. Budget limitations on the survey and therefore the sampling rate, may produce higher than desired CV's. The fact that no historical data exists will also lead to higher CV's.

3.11. CONCLUSION

In this chapter, the survey design and sampling frame were further refined by stratifying the sampling frame. The strata were then defined using land-cover and land parcel area as stratification criteria. This was done so as to reduce the variance of the items under study. Once this step was completed, the survey was planned by determining the sample size, preparing the survey maps, preparing the survey questionnaire and designing the tabulation plan for the input of the data.

After careful implementation of the previously mentioned procedures and by adhering to the specifications as defined by the GVS, data was gathered by GVS personnel between September and December 1999 and input into the database for estimation. The methods for the estimation procedures were also reviewed, from which the closed estimator was chosen for the next phase to derive the animal statistics for Gauteng.

The following chapter will concentrate on the acquired results, taking into consideration that these results are only a set of sample data, which must also be evaluated in terms of their accuracy of estimation. This data can actually be regarded as points of information that can be related to a larger area. These points can be assigned to the larger area by applying an interpolation technique that converts the points into a continuous surface. Only then would a more complete picture of distribution be derived.

CHAPTER 4 – ESTIMATION RESULTS AND SPATIAL INTERPOLATION

The previous chapter highlighted the steps taken for the preparation of the survey and the conducting of the survey itself. Choices were made regarding stratification criteria and sample size. A review of the three types of estimation techniques was also given and a choice of which estimator to use was made. In this chapter, the chosen closed estimator is applied to each survey item (animal type) and to each stratum (land-cover and land parcel size) to come up with a global estimate for each item for the province. Each item is then given an accuracy assessment using the coefficient of variance as an indicator of accuracy.

Later in the chapter, an overview of the main interpolation techniques is given and then one is selected to derive spatial distribution maps of each item for the Gauteng province.

4.1. THE GAUTENG ANIMAL ESTIMATE PER STRATUM

For calculation of statistics derived from sample data (expansion to the population level), it is absolutely necessary to know the probability of selection of a sampled item (Garibay *et al.,* 1996). In this case, the probability of selection was equal for all land parcels regardless of area. This is because the random sampling script used is based on the number of polygons, not the size.

As was mentioned previously, the closed segment estimator was chosen, since it is the only possible choice if the data surveyed are referred to the land parcel (and not, for example, to a farm that includes several land parcels), as is the case in this research study.

Francesco Luccarrini, a statistician from Aquater in Italy, who specialises in Area frame statistics, developed a computer programme to run the estimate.

The following is the estimator for stratified random sampling, the variance, the standard error of estimation and the coefficient of variance:

Notation:

h = stratum indicator h=1,…,H
H = total number of strata

n_h = number of units sampled in stratum h
N_h = total number of units in stratum h

i = sample unit (land parcel) indicator
y_{ih} = number of individuals surveyed in sample unit i stratum h

[1] estimate $$\hat{Y} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{ih}$$ **(8)**

[2] variance $$v\left(\hat{Y}\right) = \sum_{h=1}^{H} N_h^2 \frac{1 - n_h / N_h}{n_h} \cdot s_h^2(y)$$ **(9)**

where $s_h^2(y) = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(y_{ih} - \bar{y}_h\right)^2$ is the sampling variance in stratum *h* **(10)**

[3] standard error $$s.e.\left(\hat{Y}\right) = \sqrt{v(y)}$$ **(11)**

[4] Coefficient of variation (%) $$CV\left(\hat{Y}\right) = \frac{s.e.\left(\hat{Y}\right)}{\hat{Y}} 100$$ **(12)**

Land parcel area is not directly included in the estimator because the selection of the sample was completely random (same probability for each land parcel) and not proportional to land parcel area. According to Luccarrini (pers. comm. 2000), land parcel area could be included *a-posteriori* in a weighted estimator but the correlation between number of animals and land parcel area are not that strong, so that the results could be misleading.

On the advice of Luccarrini (pers. comm. 2000), two distinct estimates have been produced for each stratum. One including outliers and one excluding outliers. This is because extremely high outliers (e.g. poultry) lead to a distorted estimate and a high CV. The outliers were calculated by selecting values exceeding the mean by five times the standard deviation. The usual limit is three times the standard deviation (Luccarrini, pers. comm., 2000). The difference in the two estimates comes from only a few land parcels with extremely high values and sometimes this difference is very high as in the case of ostriches and poultry.

If properly used, stratification nearly always results in a smaller variance for the estimated mean or total than is given by a comparable simple random sample (Cochran, 1977). The difference in the estimate's accuracy between strata comes from both different sampling rates for each stratum and a different data structure (i.e. higher variability of the counts). The CV acts as an indicator of the accuracy of the estimator.

Table 6 shows the original sample size and the sample size used in the estimator. The reason for the discrepancy in values is nonresponse i.e. the enumerator was unable to gain access to the land parcel and the form was not filled in.

It must be noted that the order of animals from table 7 onwards is according to the priority set by GVS for each animal.

Table 6. Original sample sizes and sample sizes used in estimator.

Stratum	Total sampling units	Original sample	Sample size used
1	19 358	737	722
2	4 403	706	694
3	492	153	153
4	1 873	158	158
5	39 883	208	189
6	1 528 857	204	199
7	8 215	164	164

4.1.1. Stratum 1: vegetation with land parcel size < 25 ha

Table 7. Estimates for stratum 1: vegetation with parcel size < 25 ha (incl. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV)
cattle	290 825	133 413	45.87
sheep	91 937	18 194	19.79
goats	10 081	2 732	27.11
horses	10 617	2 359	22.22
donkeys	1 903	639	33.59
mules	831	601	72.34
pigs	38 421	8 030	20.90
poultry	6 382 241	5 294 315	82.95
dogs/cats	54 829	2 409	4.40
ostriches	102 527	97 345	94.95
game	54400	32 098	59.00

Table 8. Estimates for stratum 1: vegetation with parcel size < 25 ha (exc. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV)
cattle	156 767	23 535	15.01
sheep	80 676	14 539	18.02
goats	10 081	2 732	27.11
horses	9 196	1 912	20.80
donkeys	1 903	639	33.59
mules	831	601	72.34
pigs	38 421	8 030	20.90
poultry	1 019 914	601 912	59.02
dogs/cats	54 829	2 409	4.40
ostriches	3 324	1 657	49.86
game	54 400	32 098	59.00

In this stratum, no outliers occur for goats, donkeys, mules, pigs, dogs/cats and game so the estimate remains the same. The CV for dogs/cats indicates that the estimate for this stratum is reasonably accurate (Table 7). The CV's for pigs, goats, donkeys are between 20 % and 40 % giving an acceptable accuracy for this stratum. The CV's for game and mules are unacceptably high. This could be explained by the fact that these items are extremely rare in the stratum and the sampling rate is not high enough to get a reliable estimate.

The estimates for sheep and horses do not change much when run with or without the outliers. This is because the outlier values are not extremely large and lie close to the cut-off for outliers. The CV's for these two items are acceptable.

The cattle, poultry and ostrich estimate for this stratum change drastically when done without the outliers (Table 8). The cattle estimate becomes more reliable with a CV of 15 %. The change occurs because there are a few land parcels that have very high cattle counts.

The CV's for poultry and ostriches are still unacceptably high when outliers are ignored. This is due to extremely high values occurring on a few land parcels in the case of poultry. In the case of ostriches, it is probably due to the fact that ostriches occur rarely in the stratum and the sampling rate is not large enough for a reliable estimate.

4.1.2. Stratum 2: vegetation with land parcel sizes 25 – 400 ha

Table 9. Estimates for stratum 2: vegetation with parcel size 25 – 400 ha (incl. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV)
cattle	163 862	15 355	9.37
sheep	40 286	4 970	12.34
goats	4 599	1 420	30.88
horses	3 375	549	16.27
donkeys	748	142	19.00
mules	12	11	91.78
pigs	21 964	3 705	16.87
poultry	3 228 661	1 693 173	52.44
dogs/cats	16 527	1 074	6.50
ostriches	1 021	363	35.56
game	13 735	4 337	31.58

Sheep, donkeys, mules, ostriches and game have no outliers in this stratum and the estimates therefore stay the same. The sheep estimate gives a low CV and the estimator for sheep in this stratum is accurate (Table 9). Donkeys, ostriches and game have an acceptable CV and the estimate for these items is therefore reasonably reliable. Mules have an extremely high CV indicating an unreliable estimate. This is because mules are a rare item in this stratum.

Table 10. Estimates for stratum 2: vegetation with parcel size 25 – 400 ha (exc. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV)
cattle	155 170	13 245	8.54
sheep	40 286	4 970	12.34
goats	3 330	819	24.61
horses	3 051	464	15.22
donkeys	748	142	19.00
mules	12	11	91.78
pigs	18 347	2 893	15.77
poultry	849 518	426 259	50.18
dogs/cats	14 852	582	3.92
ostriches	1 021	363	35.56
game	13 735	4 337	31.58

The estimates of cattle, goats, horses, pigs and dogs/cats differ very slightly including and excluding outliers (Table 10). This is because the outliers lie very close to the cut-off of five standard deviations above the mean. The CV's for dogs/cats and cattle are below 10 % indicating a reliable estimate. The pig, horse and goat estimate for this stratum can also be considered reasonably reliable.

Poultry has an unacceptably high CV for both estimates and therefore cannot be considered reliable. This is because of the occurrence of extremely high counts in one or two land parcels.

Most CV's in this stratum range from good to reasonable indicating a reliable overall estimate for this stratum. This is due to quite a high sampling rate of 16 %.

4.1.3. Stratum 3: vegetation with land parcel size > 400 ha

Table 11. Estimates for stratum 3: vegetation with land parcel size > 400 ha (incl. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	38 990	7 781	19.96
sheep	8 261	1 618	19.59
goats	466	173	37.18
horses	492	171	34.84
donkeys	48	21	43.55
mules	16	13	83.01
pigs	1 656	619	37.42
poultry	199 919	135 904	67.98
dogs/cats	1 575	172	10.97
ostriches	38	23	61.71
game	10 277	4 838	47.08

The estimator for sheep, goats, donkeys, mules, pigs, poultry, ostriches and game have no outliers, so the estimates are the same in both tables. The low CV for sheep indicates a reliable estimate for this stratum

(Table 11). The CV's for goats, pigs and to a lesser extent donkeys and game can also be considered reliable. The CV's for mules, poultry and ostriches are extremely high and therefore unreliable. This is because these items are rare in this stratum.

Table 12. Estimates for stratum 3: vegetation with land parcel size > 400 ha (exc. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	26 789	3 526	13.17
sheep	8 261	1 618	19.59
goats	466	173	37.18
horses	321	99	30.88
donkeys	48	21	43.55
mules	16	13	83.01
pigs	1 656	619	37.42
poultry	199 919	135 904	67.98
dogs/cats	1 414	118	8.41
ostriches	38	23	61.71
game	10 277	4 838	47.08

The estimators for cattle, horses, dogs/cats show slight differences without outliers (Table 12). In both estimators cattle and dogs/cats show a low CV indicating an accurate estimate. Horses show a higher CV but it can still be considered relatively reliable.

Due to the high sampling rate of 31 % (the highest of all the strata), it was expected that the estimate would be fairly accurate for all items in the stratum. This is not the case except for cattle, sheep and dogs/cats. This could indicate that all the items, except for the three mentioned, are relatively rare in this stratum.

4.1.4. Stratum 4: cultivated land with land parcel size > 25 ha

Table 13. Estimates for stratum 4: cultivated land with land parcel size > 25 ha (incl. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	117 382	15 109	12.87
sheep	49 198	10 272	20.88
goats	1 931	878	45.49
horses	604	199	32.95
donkeys	325	120	37.07
pigs	7 398	3 442	46.53
poultry	22 964	5 265	22.93
dogs/cats	8 341	712	8.54
ostriches	430	247	57.41
game	290	267	91.84

All items (except sheep and pigs) have no outliers therefore giving the same number in both tables. Cattle and dogs/cats produced a low CV indicating an accurate estimate (Table 13). Poultry numbers are estimated reliably in this stratum when compared with previous strata. Goats, horses and donkeys give a higher, but still reliable, CV. Ostriches and game give very high CV's indicating that they occur rarely in this stratum.

Table 14. Estimates for stratum 4: cultivated land with land parcel size > 25 ha (exc. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	117 382	15 109	12.87
sheep	42 450	8 196	19.31
goats	1931	878	45.49
horses	604	199	32.95
donkeys	325	120	37.07
pigs	3 908	934	23.91
poultry	22 964	5 265	22.93
dogs/cats	8 341	712	8.54
ostriches	430	247	57.41
game	290	267	91.84

Sheep shows only a slight difference between the two estimates and gives a reliable estimate in both cases (Table 14). Pigs show an acceptable CV in the estimate excluding outliers, indicating that a few land parcels returned a high pig count.

No mules were counted in this stratum.

4.1.5. Stratum 5: smallholdings/subsistence exc. vegetation

Table 15. Estimates for stratum 5: smallholdings/subsistence exc. vegetation (incl. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	302 604	101 116	33.42
sheep	191 185	57 236	29.94
goats	92 638	56 451	60.94
horses	13 083	3 762	28.75
donkeys	422	296	70.36
pigs	63 306	20 250	31.99
poultry	10 165 311	9 473 421	93.19
dogs/cats	133 998	11 409	8.51
ostriches	8 440	4 893	57.98
game	13 716	10 973	80.00

All the items produce the same estimate, except for goats which has a few large outliers (Table 15). The difference between the two estimates for goats is high indicating that one or two land parcels have an extremely high count.

Table 16. Estimates for stratum 5: smallholdings/subsistence exc. vegetation (excl. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	302 604	101 116	33.42
sheep	191 185	57 236	29.94
goats	37 139	11 963	32.21
horses	13 083	3 762	28.75
donkeys	422	296	70.36
pigs	63 306	20 250	31.99
poultry	10 165 311	9 473 421	93.19
dogs/cats	133 998	11 409	8.51
ostriches	8 440	4 893	57.98
game	13 716	10 973	80.00

Only dogs/cats give a reliable estimate for this stratum (Table 16). Cattle, sheep, horses and pigs have acceptable CV's so that the estimate given is not too inadequate. Poultry, ostriches and game exhibit very high CV's. The reason for this is two-fold: these items occur rarely in the stratum and the sampling rate for this stratum was only 0,5 %. The low sampling rate also accounts for the higher CV's in the other items.

No mules were counted in stratum 5.

4.1.6. Stratum 6: urban – residential

Table 17. Estimates for stratum 6: urban – residential.

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	522 423	242 417	46.40
sheep	245 846	245 830	99.99
goats	676 077	311 802	46.12
horses	122 923	108 565	88.32
poultry	1 044 846	541 692	51.84
dogs/cats	2 235 665	177 214	7.93

No donkeys, pigs, mules, ostriches or game were counted in this stratum. There are no outliers for stratum 6, so the estimate will be the same for each item (Table 17).

All items except dogs/cats have high CV's. Cattle and goats can be considered acceptable when compared to the extremely high CV's of sheep and horses. A very low sampling rate of only 0.01 % accounts for these unreliable estimates. It seems unlikely to find cattle, goats, sheep, horses and poultry in the urban-residential stratum. Even though the estimates indicate an unreliable estimate for all these items, it is still significant to find these animals here at all. The occurrence of these animals in this stratum could be explained by the numerous informal settlements in the province. Horses are used to haul firewood, while the others are used for food or as an indication of wealth. Often these animals are found together on communal lands within informal settlements.

The reliable estimate for dogs/cats occurs because of the low range for each land parcel. It is reasonable to assume that each land parcel in the stratum has about two dogs/cats. This minimal variation leads to a low expansion factor that does not distort the estimator.

4.1.7. Stratum 7: animal deficient

Table 18. Estimates for stratum 7: animal deficient (incl. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	8 215	3 007	36.61
sheep	14 326	7 853	54.82
goats	2 254	1 659	73.61
horses	3 005	1 807	60.13
mules	250	247	99.00
pigs	951	848	89.12
poultry	37 518	13 164	35.09
dogs/cats	10 769	2 684	24.93
ostriches	450	331	73.61
game	5 910	5 851	99.00

Table 19. Estimates for stratum 7: animal deficient (excl. outliers).

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	8 215	3 007	36.61
sheep	14 326	7 853	54.82
goats	2 254	1 659	73.61
horses	3 005	1 807	60.13
mules	250	247	99.00
pigs	951	848	89.12
poultry	37 518	13 164	35.09
dogs/cats	8 265	1 144	13.84
ostriches	450	331	73.61
game	5 910	5 851	99.00

Dogs/cats is the only item to have outliers occurring in this stratum (Table 18). The estimator becomes more reliable when these outliers are removed (Table 19).

Poultry and cattle have an acceptable CV when compared to the other items in the stratum. All the rest exhibit extremely high CV's and the estimates for these items are unreliable.

No donkeys were counted in stratum 6.

The high CV's occur because all items are rare in this stratum because it includes mining, industrial and commercial areas.

These animal deficient areas contain livestock because they can sometimes be found grazing on abandoned mining and /or industrial land. Communal land is also known to occur on mining ground. Cats and dogs are often found on commercial, industrial and mining property.

4.1.8. All strata

The total of all animals (for all strata) actually counted is given in Table 20.

Table 20. Animals surveyed for all strata in selected land parcels.

Animal ID	Animal	Total surveyed
1	cattle	60 954
2	sheep	17 801
3	goats	1 984
4	horses	1 288
5	donkeys	234
6	mules	43
7	pigs	6 393
8	poultry	860 163
9	dogs/cats	7 008
10	ostriches	4 083
11	game	7 698

Table 21. Global estimate for all strata (excl. stratum 6) with outliers.

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	921 880	168 989	18.33
sheep	395 194	61 656	15.60
goats	111 970	56 567	50.52
horses	31 178	4 833	15.50
donkeys	3 448	729	21.16
mules	1 110	650	58.59
pigs	133 698	22 388	16.75
poultry	20 036 616	10 984 579	54.82
dogs/cats	226 041	12 036	5.32
ostriches	112 909	97 469	86.33
game	98 331	35 032	35.63

Table 22. Global estimate for all strata (excl. stratum 6) without outliers.

Animal	Estimate (y)	Standard error (se)	Coeff. of variance (CV) %
cattle	766 930	105 847	13.80
sheep	377 186	60 361	16.00
goats	55 203	12 442	22.54
horses	29 263	4 619	15.79
donkeys	3 448	729	21.16
mules	1 110	650	58.59
pigs	126 592	22 020	17.40
poultry	12 295 147	9 503 072	77.29
dogs/cats	221 701	11 753	5.30
ostriches	13 706	5 196	37.91
game	98 331	35 032	35.63

Stratum 6 was excluded form the global estimate because of the extremely low sampling rate which leads to unreliable expansions. Table 21 shows the estimate with outliers and Table 22 shows the estimate without outliers.

Cattle were counted in all strata. The cattle estimate with (921 880) and without (766 930) outliers differs slightly, but the difference is not too drastic. The CV for both estimators is relatively low given a reliable estimate for cattle. The *a Priori* figure given to calculate the sample size is well below the estimates (Table 1, pg. 37).

Sheep were counted in all strata. The sheep estimate (395 194 with outliers and 377 186 without outliers) is very close for both estimates. The CV for both estimates is low indicating a reliable estimate for sheep. The *a Priori* information given for the combination of sheep and goats is below the estimate (Table 1).

Goats were counted in all strata. The estimate with (111 970) and without (55 203) outliers is significantly different. The difference occurs because a few land parcels have a high goat count. The estimate without outliers has an acceptable CV.

Horses were counted in all strata. The estimate with (31 798) and without (29 263) outliers is almost the same. This is because the outliers lie very close to the cut-off of five standard deviations above the mean. The CV for both estimates is low indicating a reliable estimate for horses. The *a priori* value given for a combination of horses, mules and donkeys, is not too far off, but the difference is still a significant one (Table 1).

Donkeys were not counted in strata 6 or 7. No outliers occurred for this item so the estimate stays the same (3 448). The CV of around 20 % is acceptable for this item.

Mules were not counted in strata 4, 5 and 6. There were no outliers occurring for mules so the estimate remains unchanged (1 110). The CV is too high to provide a reliable estimate. This is because mules are a rare item in the province. A much higher sampling rate over the whole province would be needed to provide a reliable estimate for this item.

Pigs were not counted in stratum 6. The pig estimate is relatively close with (133 698) and without (126 692) outliers. The low CV for both estimates indicates a reliable estimate. It is interesting to note that the *a Priori* figure given is slightly higher than both estimates (Table 1).

Poultry were counted in all strata. There is an extremely large variation between the estimate with (approx. 20 million) and the estimate excluding outliers (approximately 12 million). The CV's for both estimates are extremely large and the estimate is unreliable. The huge difference between estimates is due to a few land parcels having very high counts of 10 000 or greater. To get a more reliable estimate, all chicken farms and batteries will have to be identified and included in a stratum on their own. The rest of the province will have to be sampled more intensively as well. The estimates are far higher than the *a Priori* information given to calculate sample size (Table 1).

The combination of dogs and cats was counted in all strata. The estimate with (226 041) and without (221 701) outliers is very similar indicating the outliers lie very close to the cut-off of five standard deviations. The CV for both estimates is very low indicating quite an accurate estimate. This is the only item for which the estimates for stratum 6 and the rest of the stratum can be combined because both have a relatively low CV and the expansion for stratum 6 was not distorted. The two can be combined using the formula:

[Global estimate without stratum 6] + [stratum 6 estimate]

The result including outliers is 2 461 706 and excluding outliers is 2 457 366. The global CV for dogs and cats for the whole province is 7,22 % indicating a reliable estimate for the whole province. The estimate is significantly lower than the *a Priori* value given (Table1).

Ostriches were not counted in stratum 6. The difference between the estimate with (112 909) and without (13 706) is extremely large due to high ostrich counts occurring in a few land parcels. Ostriches are also rare in the province. A much higher sampling rate or a list frame would have to be implemented to get reliable results for ostriches. The CV of the estimate without outliers of 37,91 % is however acceptable given the priority of this item as set by GVS.

No game was counted in stratum 6. No outliers occur for this item so the estimate remains unchanged (98 331). The CV for game is acceptable given the priority set by Veterinary Services for this item (Table 1). The estimate differs greatly from the *a Priori* information given in Table 1.

Overall it can be said that the estimate for dogs/cats, cattle, sheep, horses and pigs is reliable and accurate. The estimate for goats, donkeys, ostriches and game is acceptable while the estimate for mules and especially poultry is unreliable.

4.2. SPATIAL DISTRIBUTION OF ANIMALS IN GAUTENG

4.2.1. Methodology

In the previous section, estimations of population totals were made and accuracy for each of the items was given. These are only population totals and they give no indication of the distribution of items. Points of observation at specific locations are however available from the survey and these can be used to estimate the values of the unknown areas from the measured points in the survey. This is known as the interpolation process.

4.2.1.1. The Interpolation Process

There are many advantages in taking spatial data beyond a purely descriptive display method, such as thematic mapping of points using colours or proportionally sized symbols (Figure 9 and Figure 10). Modelling and interpolation software, such as ArcView Spatial Analyst, provide the means necessary to process and display data in a new derivative form (Wyatt, 2000).

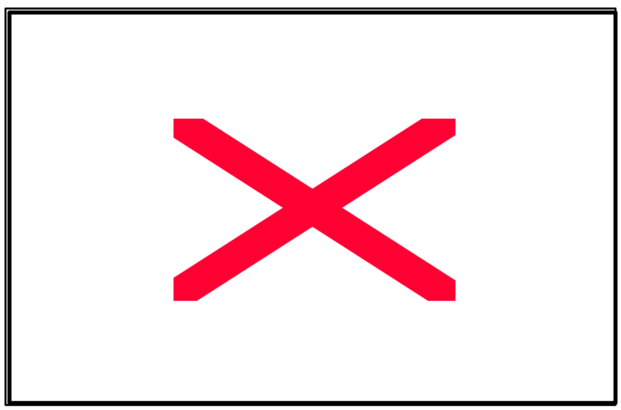

Figure 9. Example of sample points (Wyatt, 2000).

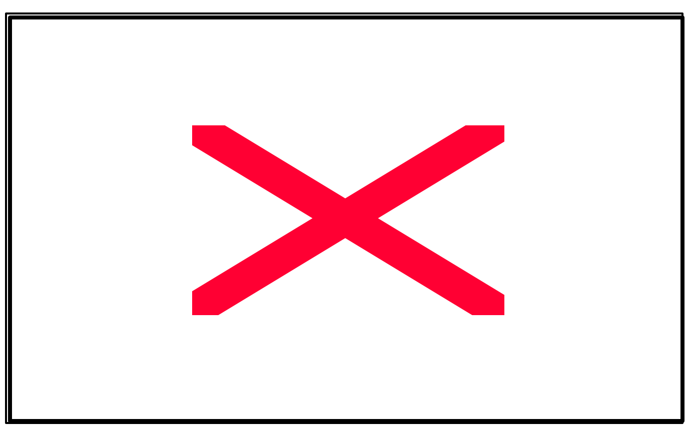

Figure 10. Point values represented in 3D space (Wyatt, 2000).

Point data usually arise as measurements or observations yielding numerical results at specific locations. The attribute can be anything. As long as it is numeric it can be represented in 3D form. Interpolation is the process of estimating a value at a location (x, y) from assumed or measured values at other locations (x1, y1), ..., (xN, yN) (Figure 11). Interpolation can be used to create a grid simply by estimating values at the centre of each grid cell (http://www.quantdec.com /).

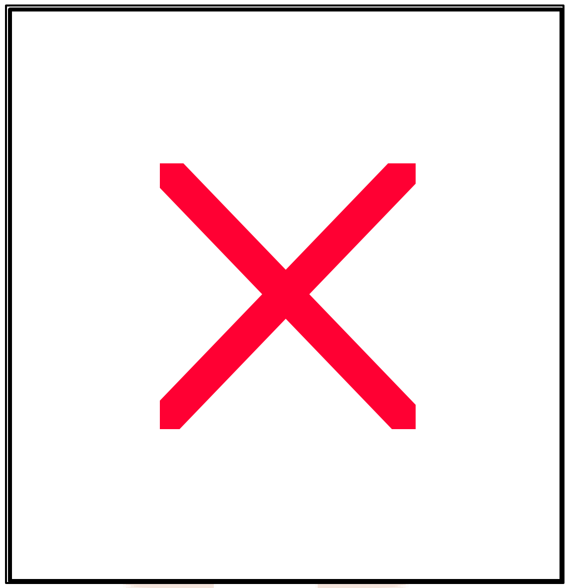

Figure 11. The interpolation process (http://www.quantdec.com).

The main reason for using grids is to build a continuous surface that connects all the data points in space, effectively removing gaps in the representation of data and facilitating comparison of datasets (Wyatt, 2000).

One way to interpolate is to average all the measured values. This makes sense when the measurements are really independent of location. Normally, however, a variable defined continuously over part of the earth's surface does not change wildly from one point to another. It can be expected that the values at unmeasured points be related to the values at nearby measured points. Local interpolators exploit this idea by using only the points near (x, y) for estimates. These interpolat ors vary according to how the nearby points (the "neighbourhood") are selected (http://www.quantdec.com/SYSEN597/GTKAV /section9/chapter_29b.htm).

4.2.1.2. Neighbourhood search methods (http://www.quantdec.com /SYSEN597/GTKAV/ section9/ chapter_29b.htm)

Two common approaches are 1) select all points within a fixed distance of the estimated point or 2) select the nearest K points, where K is predetermined. In Figure 12, the fixed-distance technique (using a distance equal to the radius of the circles) would interpolate the value at point A using four neighbours but would use only one neighbour to interpolate the value at point B.

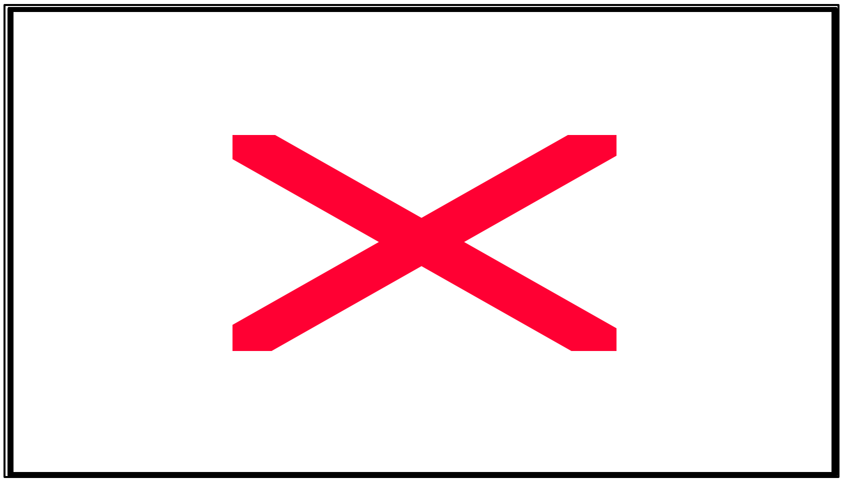

Figure 12. Fixed-distance interpolation method (http://www.quantdec.com/SYSEN597/GTKAV).

If, instead, a fixed number of neighbours were used, then the search around point B would have to expand until four neighbours are found (Figure 13).

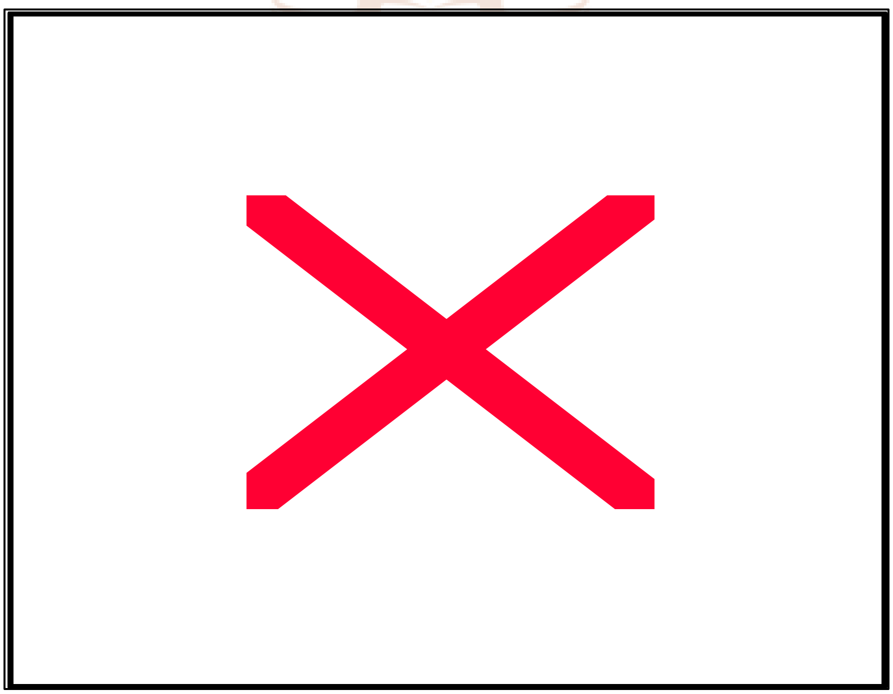

Figure 13. Nearest neighbour method (http://www.quantdec.com/SYSEN597/GTKAV/section9).

Each method has its advantages and drawbacks:

– The fixed-distance approach may fail to find any neighbours. It will be unable to interpolate into large spatial gaps in the data. In some circumstances this is good, because it helps reveal areas of undesirably low sampling; in other circumstances, this is bad because the holes in the grid may be undesirable.

– The K-nearest neighbour approach will always find neighbours, but they may be so far away from some points as to be essentially useless.

Which method to use will depend on the data, the method of collection, the desired grid characteristics and the nature of the decisions or analyses that will be performed with the resulting grid (http://www.quantdec.com/SYSEN597/GTKAV/section9/chapter_29b.htm).

Once a set of neighbours is found, the interpolator must combine their values to produce the estimate. The simplest combinations, such as the average or median are often reasonable, but they have several undesirable properties:

– Any statistical combinations of the neighbouring data depend only on the neighbours, which implies the estimated values will only change when the collection of neighbours changes.

– The values of the statistical combinations do not change except when the neighbourhood changes.

This leads to discontinuous step -like surfaces (Figure 14), which can be improved upon by using distance weighting to smooth out the distribution surfaces.

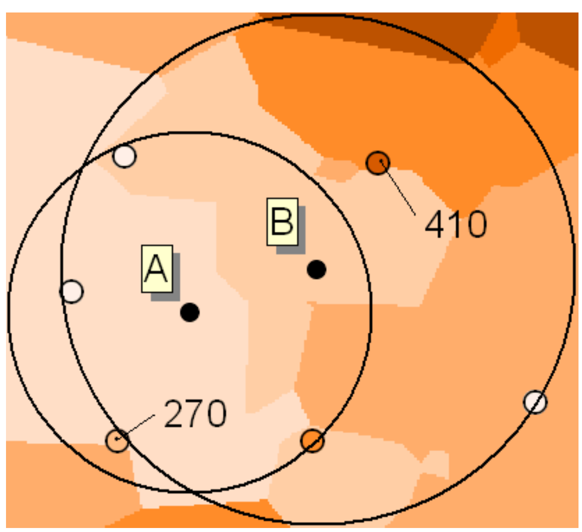

Figure 14. Average method and resulting step-like surfaces (http://www.quantdec.com).

4.2.1.3. Inverse Distance Weighting (IDW) Method

Because of these step-like surfaces, it is preferable to weight the averages according to the relationships among the points. The simplest form of weighting is IDW (Figure 15). The IDW interpolator assumes that each input point has a local influence that diminishes with distance. It creates weights according to the distances between the interpolated location (x, y) and each of its neighbours. The weight functions should be largest at zero distance and decrease as the distance increases.

Power functions provide an easy-to-compute set of weight functions. It controls the significance of the surrounding points upon the interpolated value. A higher power results in less influence from distant points (Watson & Philip, 1985).

IDW, therefore, works in four steps (http://www.quantdec.com/SYSEN597/GTKAV/section9/chapter_29b.htm):

- Find a neighbourhood {(x1, y1), ..., (xN, yN)} of points around (x, y), the location where the value is to be estimated.
- Compute the relative weight w_K of each point (xK, yK) in the neighbourhood as
- $wk = distance((x, y), (xK, yK))^{-p} = [(x-xK)^2 + (y-yK)^2]^{(-p/2)}$
- Divide each relative weight by the sum of all the weights: this causes the weights to sum to unity.
- Form the weighted average of the known or assumed values at each point (xK, yK). This is simply

  $z = \text{Sum of } w_k * z_k$

  where z_k is the value at point (xK, yK).

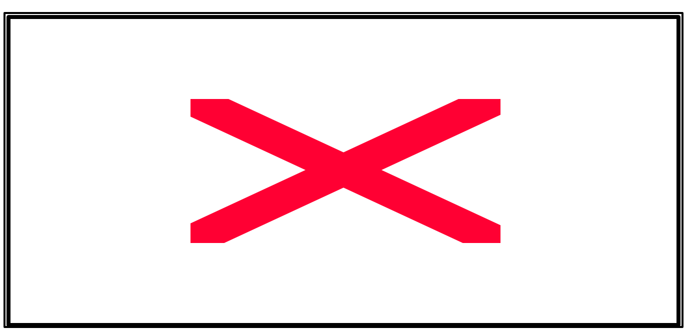

Figure 15. IDW with some smoothing (Wyatt, 2000).

Conclusions for the general situation of many data points (http://www.quantdec.com /SYSEN597/GTKAV/ section9/chapter_29b.htm):

- For powers > 1.0, IDW surfaces are smooth (have first derivatives) at the data locations.
- For powers < 1.0, IDW surfaces cusps (discontinuous or infinite first derivatives) at the data locations.
- As the power increases, the IDW surface becomes flatter near the data points and jumps more rapidly in value between the data points. The IDW estimator acts as if data have stronger spatial correlation as the power increases.
- At power = 0, the IDW estimator reduces to the usual unweighted average.

For these reasons powers of greater than 1.0 are often chosen.

The best results from IDW are obtained when sampling is sufficiently dense with regard to the local variation that is being simulated. If the sampling of input points is sparse or very uneven, the results may not sufficiently represent the desired surface (Watson & Philip, 1985). Therefore it is necessary to investigate more of these interpolation techniques.

4.2.1.4. Spline Method (Franke, 1982; Mitas & Mitasova, 1988)

The spline interpolator is a general-purpose interpolation method that fits a minimum-curvature surface through the input points. Conceptually it is like bending a sheet of rubber to pass through the points, while minimising the total curvature of the surface. It fits a mathematical function to a specified number of nearest input points, while passing through the sample points. This method is best for gently varying surfaces such as elevation. It is not appropriate if there are large changes in the surface within a short horizontal distance, because it can overshoot values. The regularise method yields a smooth surface. The tension method tunes the stiffness of the surface according to the character of the modelled phenomenon. The basic form of the minimum-curvature spline interpolation imposes the following two conditions:

- The surface must pass exactly through the data points.
- The surface must have minimum curvature.

4.2.1.5. Kriging Method (Oliver & Webster, 1990; Wendroth & Nielson, 1999)

Kriging (Figure 16) is an advanced interpolation procedure that generates an estimated surface from a scattered set of points with z-values. Kriging is based on the regionalised variable theory that assumes that the spatial variation in the phenomenon represented by the z-values is statistically homogeneous throughout the surface; that is, the same pattern of variation can be observed at all locations on the surface. This hypothesis is fundamental to the regionalised variable theory. Point sets known to have anomalous pits or spikes, or abrupt changes such as those that might be represented by breaklines in a tin, are not appropriate for the Kriging technique.

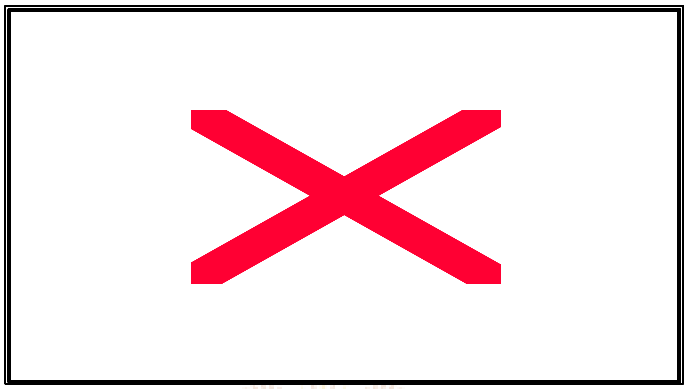

Figure 16. Kriging with heavy smoothing (Wyatt, 2000).

After careful consideration of the previously discussed interpolation methods and the sampling points used in the Gauteng study area, **it was decided to use the IDW technique for the following reasons**:

– irregularly spaced sample points were used in the study area,

– potentially large variation in livestock numbers and livestock densities were encountered during the estimation process, and

– surfaces can be interpolated through known points.

Density is calculated when the major concern is the relative geographical crowding or sparseness of discrete phenomena, such as the number of persons or cattle per area unit (Robinson *et al.*, 1984). The density is computed by

$D = N / A$

Where N is the total number of phenomena occurring in an enumeration unit and A is the area of the unit. Density is more closely related to the land than other averages and ratios, and the significant element in the relationship is area. Thus, for example, 5000 items in an area of 100 hectares is a density of 50 per hectare.

When working with density, the size of the land parcels limits the detail that can be presented. Generally, the larger the units, the less will be the differences among the values (Robinson *et al.*, 1984).

4.2.2. Results

The results of the interpolation process using the chosen IDW interpolation technique will now be discussed. Each survey item is discussed separately and possible reasons for the spatial variations in the distributions are given. Maps of the distribution and density distribution can be seen at the end of the document.

4.2.2.1. Cattle results

Cattle distribution (Map 13) mainly corresponds to the land parcels with an area greater than 25 ha and areas with a vegetation land-cover. The urban and residential areas of Johannesburg, Pretoria and the East and West Rand show little or no concentration of cattle. Relatively high concentrations occur in the districts of Nigel, Heidelberg, Krugersdorp, Bronkhorstspruit, Cullinan, Wonderboom and Vanderbijlpark. These districts occur on the periphery of the urban commercial and residential centres. Vanderbijlpark is dominated by cultivated land but still shows quite a high concentration of cattle. This might be an indication that farmers graze cattle on fallow land in this district. The interpolation shows a high concentration of cattle in the extreme south of the Johannesburg district. This could be associated with cattle grazing on communal land or on mining land. According to Van der Zel (pers. comm., 2000) the high values in Bronkhorstspruit and Heidelberg could be associated with feedlots. In Cullinan, state land is being given to developing farmers and this could account for the high cattle concentration there (Van der Zel, pers. comm., 2000). The high values in Nigel and Krugersdorp need to be investigated.

The cattle density map (Map 14) shows some slight differences. Krugersdorp, Nigel and Heidelberg still stand out as areas of high cattle concentration. As mentioned in the previous paragraph, the Krugersdorp and Nigel area need some investigation, while the Heidelberg values might be associated with feedlots. Even though the land parcels in these districts are mostly greater than 25 ha, there is a sufficient number of cattle to show a high cattle density. Districts such as Brakpan, Germiston, Wonderboom and Vereeniging have few cattle, but they still show high cattle density. These areas correspond with smallholdings, which means the land parcels in these areas are smaller and therefore indicate a high density. The high densities in Soshanguve and Johannesburg might be associated with communal grazing land, where all owners graze their animals on pieces of open land. The high density in the west of Johannesburg corresponds well with the location of Soweto so this reinforces the association to communal land.

4.2.2.2. Sheep results

Sheep distribution (Map 15) in Gauteng shows a relatively high concentration in Nigel, Heidelberg, Vanderbijlpark and Krugersdorp and to a lesser extent in Vereeniging, Cullinan and Bronkhorstspruit.

These areas mostly correspond to land parcels with an area greater than 25 ha and vegetation land-cover. Vanderbijlpark however has cultivated land as the majority land-cover and has a high concentration of sheep. According to Van der Zel (pers. comm., 2000), sheep farms are known to occur in Cullinan, Nigel, Heidelberg, Vanderbijlpark and Krugersdorp. A high concentration occurs in the extreme west of Kempton Park. This corresponds to the location of Modderfontein where there is mostly industrial activity. Surrounding these industries are large areas of open land where sheep can be found. The high values in Bronkhorstspruit and Vereeniging will have to be researched further.

The sheep density distribution (Map 16) shows a high density in Nigel and Heidelberg due to high sheep counts in these areas. High densities also occur in the south of Vanderbijlpark. As mentioned in the previous paragraph these are areas where sheep farms are known to occur. Wonderboom has a high density because these are areas with smallholdings that have a small area. A high density also occurs in the south-east of Johannesburg, which could be communal grazing land or open mining areas. A high density also occurs in the west of Johannesburg, which corresponds to the location of Soweto and therefore could be associated with communal grazing areas. Benoni's high density will have to be investigated because it falls mainly in cultivated land according to the land-cover data.

4.2.2.3. Goat results

High goat concentrations (Map 17) occur in Soshanguve, Wonderboom, Cullinan and Bronkhorstspruit. A small concentration also occurs in Heidelberg. The high concentration in Soshanguve, the north of Cullinan and the north of Wonderboom correspond to informal settlements and they could be explained by the reasons given in paragraph 1.2.1.2 regarding rural communities. The high concentrations in Nigel and Heidelberg need to be investigated.

High densities (Map 18) also occur in Soshanguve and the surrounding areas in the north-west of Wonderboom. The high number of goats and the small land parcel sizes that occur in this area can explain the high densities, along with the explanation that was given in the previous paragraph. The high density in Alberton is due to the small land parcel sizes in the area. The high density in the west of Johannesburg corresponds with Soweto and therefore communal grazing land. The high density in the south of Wonderboom corresponds to the location of smallholdings, which have a relatively small area.

4.2.2.4. Horse results

High horse concentrations (Map 19) occur in Nigel, Benoni, Krugersdorp, Pretoria, Wonderboom and Bronkhorstspruit. These concentrations could signal the occurrence of stables or riding schools on smallholdings in the rural-urban fringe. Stud farms in the Nigel district could also explain the high concentration in this area (Van der Zel, pers. comm., 2000)

High densities (Map 20) occur in Benoni and Wonderboom due to the small land parcel size in these areas. The high densities in Johannesburg and Randburg could correspond with communal areas. Here the locals use horses to haul firewood to their homes (Van der Zel, pers. comm., 2000).

4.2.2.5. Donkey results

High areas of donkey concentration (Map 21) occur in Bronkhorstspruit, Boksburg, Vanderbijlpark, Krugersdorp and Randfontein. The concentration in Bronkhorstspruit might be due to workers on farms that keep donkeys for hauling firewood, etc. The other areas will have to undergo further investigation, but there might be a correlation with the informal settlements.

Only Bronkhorstspruit stands out as having a relatively high donkey density (Map 22). This is because of the high donkey counts taken in this area which might be due to the reason mentioned above.

4.2.2.6. Pig results

High pig concentrations (Map 23) occur in Bronkhorstspruit, Nigel, Vanderbijlpark, Oberholzer, Randfontein, Heidelberg, Westonaria, Cullinan and Krugersdorp. These areas occur away from the urban commercial and residential centre of Gauteng and correspond mostly with the grassland land-cover and the land parcels with areas greater than 25 ha. Pig farms are known to occur in Bronkhorstspruit and Krugersdorp, possibly Nigel as well (pers. comm. Van der Zel, 2000). Pig farms sometimes utilise waste from hotels and restaurants as food and they therefore maybe situated in close proximity to areas with high concentrations of hotels and restaurants. The high values in Oberholzer, Vanderbijlpark, Westonaria and Randfontein will have to be investigat ed further.

Relatively low pig densities (Map 24) occur throughout Gauteng. The exception is in the south of Wonderboom where there is a high density. This area does not show up in the distribution map. The area corresponds with small land parcels and therefore even a small pig count will show a high density. According to Van der Zel (pers. comm., 2000) this high density might be associated with a pig abattoir in this area.

4.2.2.7. Poultry results

High poultry concentrations (Map 25) occur in Bronkhorstspruit, Randfontein, Krugersdorp, Soshanguve, Wonderboom and Cullinan. These areas might correspond with poultry farms or in some cases where counts of greater than 10 000 were found, it might indicate commercial chicken batteries.

High densities (Map 26) occur in Johannesburg, Krugersdorp, Randburg, Soshanguve, Wonderboom and Cullinan. The exceptionally high densities in Soshanguve, Wonderboom, Johannesburg and

Krugersdorp may correspond with commercial chicken batteries. The high density in the west of Johannesburg corresponds with the communal lands of Soweto.

4.2.2.8. Dog and cat results

Most of Gauteng land parcels have between three and ten dogs and cats (Map 27). The exceptions are the urban commercial and industrial areas as well as the mining areas. Higher concentrations occur in Nigel, Vanderbijlpark, Krugersdorp, Wonderboom and Springs. The very high concentrations in Nigel and Springs will need to be investigated.

High densities (Map 28) occur in most of the urban residential areas because of the small erf land parcel sizes. Relatively high densities also occur in Soshanguve indicating that communal areas also have a high dog and cat count.

4.2.2.9. Ostrich results

A high ostrich concentration (Map 29) occurs in Krugersdorp. This could correspond with an ostrich farm which is known to occur in the area (pers. comm., van der Zel, 2000). A national ostrich abattoir is also found on the ostrich farm.

A high density of ostriches (Map 30) occurs in Krugersdorp and Wonderboom as well as in the south-east of Johannesburg. The high density in Krugersdorp could be associated with an ostrich farm, while the high density in Wonderboom and Johannesburg will need some research. High densities and distributions might also be due to zoo farms. Zoo farms are places where sick or new animals meant for the zoo are kept for a time for monitoring.

4.2.2.10. Game results

High concentrations of game (Map 31) occur in Krugersdorp, Vereeniging, Johannesburg, and Wonderboom and in the north-east of Bronkhorstspruit. Game farming is known to occur in the north of Cullinan as well but the sample points have not picked this up. This could be due to the operator not being willing to co-operate and divulge his game numbers or the operator not knowing how much game he actually has. Another reason for non-response is the fact that poaching is known to occur in this area (pers. comm., van der Zel, 2000). Operators are not keen to divulge game numbers due to this fact. In Wonderboom and Bronkhorstspruit game farms are known to occur (pers. comm., van der Zel, 2000). Old army ground has been converted to game farms in the Wonderboom district (pers. comm., van der Zel, 2000). The high values in Krugersdorp might be associated with a game reserve in the area, while in Vereeniging the concentration will need further investigation.

High densities of game (Map 32) occur in Wonderboom because of the game farms in the area. The high density in Johannesburg will have to undergo some investigation. The possibility of zoo farms in the area, as mentioned in the previous section, might also be a cause of some of the high densities and distributions.

4.2.3. Conclusion

After application of the closed segment estimator on the sample data, estimated population totals were derived with varying degrees of acceptance, according to the coefficient of variation also calculated. The results obtained for all strata combined, showed that:

- Cattle were found in all strata and the CV for the estimates, with (921 880) or without (766 930) outliers, indicated reliable totals.
- Sheep were also counted in all strata, and reliable totals of 395 194 (with outliers) and 377 186 (without outliers) were estimated according to the CV of 16 %.
- Goats are also to be found in all strata, but the difference in reliability between CV's with inclusion of outliers (51 %) and excluding outliers (23 %) indicates great variance of occurrence between land parcels included in the sample. Therefore the estimate excluding outliers with an acceptable CV will be regarded as giving the more accurate total of 55 203 goats.
- Horses were counted in all strata and the CV for the estimates with (31 798) and without (29 263) both indicate a reliable total.
- Donkeys were not counted in strata 6 or 7. The estimate of 3 448 is the same for both estimates. The CV of 21 % indicates a reliable estimate for donkeys.
- Mules were not counted in strata 4, 5 and 6. The estimate of 1 110 mules is the same for both estimates. The CV of 59 % is too high to provide a reliable estimate. This occurred because mules are rare animals in the province.
- Pigs were not counted in stratum 6. The pig estimate is similar with (133 698) and without (126 692) outliers. The CV for both estimates indicates a reliable pig estimate.
- Poultry were counted in all strata. There is an extremely large variation between the estimate with (20 million) and without (12 million) outliers. The CV for both estimates is highly unreliable. The difference between the estimates and the high CV is due to the fact that a few land parcels had extremely high poultry counts.
- The combination of dogs and cats were counted in all strata. The estimate of two and a half million for both estimates are very reliable given the CV of 7 %.
- Ostriches were not counted in stratum 6. The estimate of 112 909 (with outliers) and 13 706 (without outliers) is vastly different. This is because of high counts encountered on a few land parcels, combined with the fact that ostriches occur rarely in the province. The estimate without outliers is acceptable, given the priority set by GVS, with a calculated CV of 38 %.
- No game was counted in stratum 6. The estimate of 98 331 is unreliable given a CV of 36 %, but acceptable given the priority set by GVS.

The interpolation process from the sample point data produced maps of distribution for each of the animal types within the Gauteng province. As expected, the distributions of cattle, sheep and pigs are concentrated away from the urban and urban: residential areas. These animals occur mainly on land parcels greater than 25 ha. Donkeys and goats are also distributed away from urban centres, but with high densities occurring in informal settlements. Dogs and cats are distributed almost evenly throughout the province with high densities occurring in the urban: residential areas. The distributions of cattle, sheep, horses, pigs and dogs and cats are reliable. The distributions of the other animals are distorted due to extreme counts in a few land parcels, although a rough idea of concentrations can still be inferred.

This research will now be summarised, in the form of a synthes is, according to the major tasks that were conducted. After the synthesis a brief conclusion regarding the success and precision of this research is given along with improvements and further research suggestions.

CHAPTER 5 - SYNTHESIS

5.1. PROBLEM RESEARCHED

In South Africa, there are no reliable statistics regarding animal numbers and distribution. Provision for collecting such information is not made in the National census and consequently this information is not collected on a regular basis. Currently, this kind of data is collected in an irregular way by extension officers going about their daily work. This results in duplication of animal counts and therefore inaccuracies. Some areas are visited more often than others and consequently, some areas are ignored. Experts need reliable, up-to-date, information regarding animal numbers and distributions in order to make informed decisions regarding disease control and high risk areas. In order to adapt and comply with the goals and regulations of the Government, Gauteng Veterinary Services (GVS) have identified the lack of spatially linked livestock information as a major problem, especially with regard to the outbreak, and subsequent control of diseases. GVS have to answer questions regarding the geographical extent of the disease, the amount of livestock affected, how much vaccine should be sent to an area, what other species might be affected, which farms should be put in quarantine and what plans to put in place to manage the situation.

The goal, therefore, of this research was to provide the framework and procedure for obtaining spatially linked livestock statistics efficiently and accurately. From the results provided by the framework and procedure, estimations will be made and incorporated into a Geographic Information System (GIS) for the Gauteng Province.

The research undertook to answer a number of questions:

(i) will the area frame sampling method provide accurate statistics for use by GVS;

(ii) can area frame sampling be implemented successfully in South Africa; and

(iii) can the area frame sampling method be implemented for use in other sample surveys e.g. estimating crop statistics?

With the help of literature and local experts, a complete set of procedures had to be developed for implementation of this specific problem in the Gauteng province. These procedures will now be summarised in short and evaluated in terms of their applicability and results rendered.

5.2. INFORMATION GATHERING

During this phase, available sampling methods and sampling frames were investigated. This took place in the form of a literature review and by consulting available experts on the subject. First it had to be decided whether to conduct a complete census or a sample survey.

Complete enumeration censuses vs. censuses carried out on a sample basis

Agricultural censuses can be classified into two categories: censuses conducted by complete enumeration (survey) of all holdings or by a sample enumeration. The main characteristics of these two categories are the following (Food and Agricultural Organisation Statistical Development Series, 1996).

Complete enumeration

For agricultural statistics by complete enumeration, the enumerators complete a questionnaire for each holding. The result for each characteristic is obtained from the values of the characteristics in all holdings. The census will only include non-sampling errors.

Sample enumeration

Agricultural censuses based on sample enumeration are probability sample surveys, that is, surveys for which a probability sample is selected, and for which the methods of estimation for each census characteristic allows for the establishment of their statistical precision. This requires defining the sampling units and their probability of selection, from a known universe (frame).

It was decided to carry out a sample survey because the Gauteng Province has a large number of holdings (land parcels). To conduct a complete census would have been more expensive and more time consuming than a sample survey. A large number of supervisors and enumerators would have been needed for a complete census. Manpower was a severe limitation, so again a sample survey was the better option. The quantity of data to be processed is very large for a complete census. The cost of data processing would have been higher due to the large volume to be processed. Data was needed on a major administrative unit (provincial) and not at the farm level. A sample survey is sufficient when the aim is securing data with reasonable accuracy for major administrative units. Given unlimited funding, skilled personnel and time, a complete census would have been conducted. This ensures a good base for preparing an efficient sampling design for the collection of detailed data on important items, as well as planning future agricultural surveys to collect current agricultural statistics, and to be able to produce at least some data for small administrative units.

Once the decision of choosing between a complete census and a sample survey has been made the selection of sample and survey design had to be considered.

5.3. DEVELOPMENT OF SAMPLE DESIGN

The survey design refers to the definitions and established methods and procedures concerning all phases needed for conducting the operation. These include: sample design, the selection and training, the organisation of the logistics involved in the distribution and receipt of questionnaires, the data collection and data processing procedures, and the analysis of data needed for the release of the final results (Food and Agricultural Organisation Statistical Development Series, 1996).

The sample design of a survey or an agricultural census conducted by sample enumeration refers to the techniques for selecting a probability sample and the methods to obtain the estimates of the desired characteristics from the selected sample (Food and Agricultural Organisation Statistical Development Series, 1996).

Main types of sample designs for agricultural censuses

For agricultural censuses based on a sample enumeration, there are two basic types of sample methods commonly used concerning the final stage sampling unit and their probability of selection, namely list sample methods and area sample methods (Food and Agricultural Organisation Statistical Development Series, 1996).

List sample designs (Food and Agricultural Organisation Statistical Development Series, 1996)

List sample designs are the most commonly used sampling procedures. In this case, the last stage sampling units are generally the holdings or the holders' addresses.

Area sample designs

An area frame for a land area consists of a collection or listing of all parcels of land for the area of interest. These land parcels can be delineated based on factors such as ownership, or based simply on easily identifiable boundaries as is done by the NASS. An area frame is critical to producing quality estimates, as it provides complete coverage with all land areas being represented in a probability survey with a known (not necessarily equal) chance of selection (Cotter & Tomczak, 1994). The concepts of area frame sampling are very simple: divide the total area to be surveyed into N blocks, without any overlap or omission; select a random sample of n blocks; obtain the desired data for reporting units of the population that are in the sample blocks through survey methods; and estimate population totals by multiplying the sample totals by N/n.

In the Gauteng Province, where a complete list of farmers or land owners is not available, it was decided to draw a sample of areas within the territory in question. It can be used to obtain area estimates directly, but it can also become a basis for other studies on variables associated with the areas contained in the sample, or the farms to which these areas belong. In this way a representative sample of land parcels was drawn indirectly without having to make use of a list.

Once the choice of sample design was made, the survey objectives had to be defined before the area frame could be designed optimally. The survey objectives and priorities were set in consultation with various experts from GVS. Table 1 gives the animals of interest to the GVS and their priorities. These priorities were given by GVS experts according to their current needs regarding disease control, listed from highest to lowest (cattle, sheep, goats, horses, donkeys, mules, pigs, poultry, dogs/cats, ostriches, game).

5.4. APPLICATION OF CHOSEN METHODS

Construction of the sampling frame

The frame was constructed using the following layers: farm subdivisions, smallholdings and erf data. These land parcels were merged, using ArcView GIS software, into one continuous layer of land parcels. The Surveyor General supplies this land parcel data in South Africa. The data was checked using Arc/Info and ArcView GIS software so as to ensure that no land was omitted or duplicated in the merge process (Map 4, Map 5, and Map 6). These land parcels showed great variation in terms of land-cover and size and therefore needed some form of stratification before sampling points could be selected.

Stratification

The purpose of stratification is to reduce the variance of the variable under study in each stratum so as to obtain a lower variance over the entire area under study. In the Gauteng Province, there were multiple animal items to sample. Therefore a general stratification was needed, as it could not be specific to one animal. The stratification was done in consultation with GVS, according to the best estimate of animal distribution that could be supplied by GVS. Stratification in Gauteng was done using two criteria: land-cover, which was reclassified into land-cover classes that would have approximately the same animal distribution. These new land-cover classes were then divided further on the basis of the area of the land parcel to produce strata, which were sampled separately.

By stratifying the land according to its use, homogeneous groups are created, that can be sampled at lower rates, thus reducing field and office workloads. Lower sampling errors are also achieved, thereby giving more confidence that the sample provides reliable estimates of the population totals (International Training, 1987).

In section 3.2, it was stated that a land parcel was assigned a land-cover class based on the majority area covered within that land parcel. In future surveys, the frame may be improved by using animal carrying capacity to stratify. This is due to the fact that emphasis should be placed on land-cover with a high carrying capacity. Some kind of weighting proportional to animal capacity could be considered, which might improve the reliability of the end results.

Once the stratification was done, the sample size for each stratum could be calculated.

Sample size

To determine the sample size for the Gauteng area frame, two assumptions were considered. Firstly, if it was assumed that nothing was known about the population distribution and that money was no problem, over-sampling (sample size of 25 – 50%) could be decided on. Secondly, if it were assumed that the GVS had some idea of the population distribution and that a limited budget was available, then a much smaller sample size would be selected. In this study the latter assumption was the most appropriate and the sample extraction definition was built accordingly. With a better knowledge about animal distributions through a study like this, future sampling sizes need not be large and sampling points can be allocated to strata more efficiently.

Survey and data input

The sample sizes were then used to randomly select land parcels to be surveyed. These land parcels were then plotted in ArcView GIS and survey maps were printed for the enumerators. This random selection can in future be improved by considering the animal distribution patterns derived from this study.

A survey questionnaire was also prepared for the enumerators to fill in when they visited the land parcels. The questionnaire was made as simple as possible to ensure that the enumerators spent the least possible time on each land parcel.

The survey was conducted by veterinary technicians between September and December 1999. It would have been ideal to have the survey completed in one month, but this was not feasible, as there were not enough skilled technicians available on demand. The survey had to be completed as veterinary technicians went about their daily tasks. As the survey questionnaires were received, they were inputted into a database that was specifically designed for the purpose of the estimation that followed.

The survey maps used could have been improved by printing them with the 1: 50 000 topographic map sheets as a backdrop. This would have made the technicians' task of locating the land parcels much easier.

Estimation procedures

There are three types of estimation procedures to be considered: closed, open and weighted. The following example from Ford *et al* (1986) illustrates the difference between the three.

Suppose the following situation occurs for a specific farm: land parcel acres = 10, farm acres = 100, hogs on the land parcel = 20, and hogs on the farm = 40. The closed segment value of number of hogs would be 20; the weighted segment value would be 40 X (10/100) = 4; and the open segment value would be 40 (if the headquarters is in the land parcel) or 0 (if the head quarters is not in the land parcel).

The closed estimation procedure was chosen because this is the only possible choice if the data surveyed are referred to the land parcel (and not, for example, to a farm that includes several land parcels).

The results obtained using this methodology, and their reliability, will now be summarised and interpreted in terms of their utility as useful information for further decision-making.

5.5. RESULTS

Estimation results

Two estimates were calculated: one including outliers and one excluding outliers (an outlier being defined as animal counts that were above five standard deviations of the mean for each item). Overall estimates given are accurate and reliable with variations in accuracy occurring due to insufficient sampling and/ or the item sampled occurring rarely in the province.

Cattle numbers were estimated at 921 880 including outliers and 766 930 without outliers, with accuracy of 18 % and 14 % respectively according to their CV values. These low CV values indicate reliable estimates for both values obtained.

The **sheep** estimate (395 194 with outliers and 377 186 without outliers) is very close for both estimates. The CV for both estimates is low (16 % for both) indicating a reliable estimate for sheep.

The **goat** estimate (111 970 with outliers and 55 203 without outliers) is significantly different. The difference occurs because a few land parcels have a high goat count. The estimate without outliers has an acceptable CV (23 %).

The **horse** estimate (31 798 with outliers and 29 263 without outliers) is almost the same. The CV for both estimates is low (16 % for both) indicating a reliable estimate for horses.

No outliers occurred for **donkeys** so the estimate stays the same (3 448). The CV of around 20 % is acceptable for this item.

There were no outliers occurring for **mules** so the estimate remains unchanged (1 110). The CV is too high (59 %) to provide a reliable estimate. This is because mules are a rare item in the province and non-occurrence on most land parcels makes it impossible to accurately estimate their total numbers.

The **pig** estimate (133 698 with outliers and 126 692 without outliers) is relatively close. The low CV (17 %) for both estimates indicates a reliable estimate.

There is an extremely large variation between the **poultry** estimate with (approx. 20 million) and the estimate excluding (approx. 12 million) outliers. The CVs for both estimates are extremely large (55 % and 77 %) and the estimate is thus unreliable. The huge difference between estimates is due to a few land parcels having very high counts of 10 000 or greater. A number of chicken farms are located within Gauteng. These farms have extremely large numbers of chickens, which influence estimation when only area frame sampling is used. In such situations multiple frame sampling will provide more accurate estimates. It will allow for separate estimation of all chicken farms by utilising a list sample frame approach.

The result for **dogs and cats** including outliers is 2 461 706 and excluding outliers is 2 457 366. The global CV for dogs and cats for the whole province is 7,22 % indicating a reliable estimate for the whole province.

The estimate for **ostriches** with (112 909) and without (13 706) outliers differs significantly. One or two land parcels having extremely high ostrich counts cause this difference. The CV without the outliers (38 %) is acceptable for this item given the priority set by Veterinary Services.

No outliers occur for **game** so the estimate remains unchanged (98 331). The CV for game is also acceptable (36 %) given the priority set by Veterinary Services.

Three classes of accuracy can now be defined. Items surveyed with a CV below 15 % can be considered reliable. Those with a CV of between 15 % and 20 % can be considered acceptable. Items with a CV greater than 20 % start becoming unreliable and become totally unacceptable the higher the CV gets. According to GUSS/NASS Project (1995), estimates should have CV's less than 10 % for items of major importance. Budget limitations on the survey and therefore the sampling rate, may produce higher than desired CV's. The fact that no historical data exists will also lead to higher CV's. It is important to put the CV figures into descriptive classes because decision-makers sometimes do not have the necessary statistical background to be able to interpret numbers. The area frame methodology therefore produced **reliable results** for cattle and dogs and cats. It also produced **acceptable results** for sheep, donkeys, horses and pigs. The area frame method **did not work well for the other items** (goats, mules, poultry, ostriches and game) because of extremely high values in a few land parcels and due to the fact that some of these items are rare in the province. The accuracy of those items in the unacceptable category could be improved by adding a list frame of those land parcels that have the majority of these items.

Spatial results

Three **methods** of displaying the survey data spatially were considered: Inverse Distance Weighting, Spline and Kriging.

Inverse Distance Weighting (IDW) method

The simplest form of weighting is IDW. The IDW interpolator assumes that each input point has a local influence that diminishes with distance. It creates weights according to the distances between the interpolated location (x, y) and each of its neighbours.

Spline method

The Spline interpolator is a general-purpose interpolation method that fits a minimum-curvature surface through the input points. It fits a mathematical function to a specified number of nearest input points, while passing through the sample points.

Kriging method

Kriging is an advanced interpolation procedure that generates an estimated surface from a scattered set of points with z-values.

It was decided to use the IDW technique in the Gauteng Province because irregularly spaced sample points were used in the study area, large variation in livestock numbers and livestock densities were encountered during the estimation process, and surfaces can be interpolated through known points.

Spatial distributions were interpolated from the animal counts taken in the survey and they give a rough idea of the location of animals, but some animals gave a distorted view due to extreme counts occurring on some land parcels. The distributions of cattle, sheep, horses, pigs and dogs and cats are reliable. The distributions of the rest are distorted due to extreme counts in a few land parcels but a rough idea of concentrations can still be inferred. Maps 13 – 32 can be referred to for a more detailed look at the spatial distributions of the animals. These spatial distributions can be used by GVS to identify potential problem areas in the event of a disease outbreak. They can also be used in other disciplines by combining them with land-use and land-cover data to identify strains on the available resources due to high animal numbers

5.6. CONCLUSION

According to Cochran (1977), the more information we have initially about a population, the easier it is to devise a sample that will give accurate estimates. The precision of the sampling procedure for this survey can only really be judged by examining the frequency distribution generated for the estimate if the procedure is applied again and again to the same population. Any completed sample such as this is potentially a guide to improved future sampling, in the data that it supplies about the means, standard deviations and the nature of the variability of the principal measurements and the costs involved in getting the data.

The three questions set in the beginning can now be answered, considering the results as discussed.

Will the Area Frame sampling method provide accurate statistics for use by Gauteng Veterinary Services?

From the estimation procedure and the resulting CV values, it can be seen that the estimates are accurate enough for animals found throughout the province. It is however not accurate for rare animals or for animals that occur in high numbers on a few land parcels. The GVS now have a reliable, statistically sound animal estimate that includes an accuracy assessment. These estimates can also be used in future surveys carried out by GVS to improve on results through better sample allocation.

Can Area Frame Sampling be implemented successfully in South Africa?

From the results obtained in the Gauteng Province, it can be said that the area frame sampling methodology can be implemented successfully in other provinces. This will be the case only if the item to be surveyed is found throughout the area under survey and it is not a rare item. A list frame, or multiple frame survey, would be better for rare items. It is also advisable to have only one item under survey because the stratification can then be fine-tuned to suit the specific survey item.

Can the Area Frame be implemented for use in other sample surveys e.g. estimating crop statistics?

The area frame methodology can be used for any item that can be associated with an area (land parcel). The stratification of the frame would however have to be adjusted to suit the item of interest.

Considering that no historical data existed, and that the overall goals of this research were to get an idea of animal numbers and distribution in Gauteng province, the results and answers obtained can be considered successful, in that decision-makers now have a reliable source of information from which good decisions can be made. It also provides a procedure that can be improved upon for future repetition of surveying.

It must be stressed that the results presented in this research are from a sample survey. The figures and distributions should not be considered as absolute reality. Reality can and will differ from these results to a lesser or greater extent for the different items (animals), in accordance with each item's reflected accuracy as indicated in the results.

5.7. POSSIBLE IMPROVEMENTS AND FURTHER RESEARCH

This research and its results revealed and emphasised certain aspects, which can improve future similar applications. These aspects can be seen as confirmation of existing knowledge and new knowledge on specific local conditions, which will be highlighted in short.

Area sampling frame development is a major undertaking that must be considered a long-term investment of time, money and other resources. The efficiency of the frame over time depends on the strata definitions and quality of the stratification. When land-use does not change, restratification is not necessary and the work completed during the first year will be valid for many years (Garibay *et al.,* 1996). Unfortunately this is not the case in the dynamic Gauteng province and this could be overcome by using the "**Urban Eye**" data. Using the land-cover data from the "Urban Eye" project, which has recently been completed by the CSIR, could refine the stratification. This project has mapped the Gauteng land-cover at a 1:50 000 scale. This has resulted in a much more accurate (and more recent) land-cover assessment of the province. It has provided a more detailed breakdown of the urban land-cover classes and most likely will include classes such as communal land and/or informal settlements.

Using animal carrying capacity as a stratification tool could also improve this research. The interest is on land-cover with the ability to carry animals and some kind of weighting based on this could be beneficial.

Strata used for this study could also be combined in view of the results obtained. The vegetation with parcel size greater than 400 ha stratum and the vegetation with parcel size between 25 and 400 ha stratum, delivered the same results and need not be subdivided into two different strata. Further research and consultations with area frame experts would have to be undertaken to further refine and improve the stratification and estimation process.

More survey points should be added for rare items (e.g. ostriches and donkeys) so as to get a more accurate estimate and distribution. This is however limited by resource constraints (money and labour).

The area frame can also be broken down to district level to obtain a more detailed idea of the numbers and distribution of animals. This can be done in conjunction with the work that is done by the veterinary technicians in each district. Everytime a farm or land parcel is visited (for whatever reason), a complete count of animals will occur. It is imperative, in these cases, to record the date of the count so as to ensure that animals are not counted twice, due to animal movement.

Unexpected high concentrations in numbers and densities will need to be investigated, as mentioned before under section 4.2. This can be a verification exercise to deduce if the results of this research do in fact confirm the real situation.

GVS have disease occurrence points, where outbreaks of diseases have been recorded. A surface per disease can be created and then compared with the animal distribution and animal density surfaces. A relationship might exist between high numbers or densities and disease occurrence, which can improve decision-making on disease control considerably.

Another relationship might exist between soil type and disease occurrence and therefore, animal numbers and densities. This possibility also needs to be researched which could lead to Veterinary Services being able to make more informed decisions regarding disease control.

A way to incorporate the statistics from this survey and the statistics from the routine work of technicians needs to be found. This will create historical data that can be used in future to generate more accurate estimates and distribution surfaces, especially with regard to rarely occurring items.

It is also recommended that the Gauteng area frame and the resulting estimates be incorporated into the web-based Agricultural Geographic Information System (AGIS) as a Decision Support System for livestock estimating and distribution. AGIS is a collaborative effort between the National Department of Agriculture, the Provincial Departments of Agriculture and the Agricultural Research Council.

BIBLIOGRAPHY

Barnett, V. 1974: *Elements of Sampling Theory*. The English Universities Press. London.

Barnett, V. 1991: *Sample Survey: Principles and Methods.* Edward Arnold. London.

Bonar, B. 1989: *Guide for the LANDSAT Analyst.*

Burger, D (ed). 1999: South African Yearbook. South African Communication Service. Pretoria.

CSS Report No. 11-02-08. 1993: *Census of Agriculture 1993 – Provincial Statistics – Gauteng.* Central Statistical Service. Pretoria.

Chaudhuri, A. & H. Stenger. 1992: *Survey Sampling: Theory and Methods*. Marcel Dekker. New York.

Clark, I. 1982: *Practical Geostatistics*. Applied Science. Essex.

Cochran, W. G. 1977: *Sampling Techniques (3rd ed.)*. John Wiley & Sons. New York.

Cotter, C. & J. Nealon. 1987: *Area Frame Design for Agricultural Surveys.* National Agricultural Statistics Service – United States Department of Agriculture. Washington, D. C.

Cotter, J. J. & C. M. Tomczak. 1994: An Image Analysis System to Develop Area Sampling Frames for Agricultural Surveys. *Photogrammetric Engineering & Remote Sensing, Vol. 60, No. 3, pp. 299 – 306.* American Society for Photogrammetry and Remote Sensing. Washington, D. C.

Deming, W. E. 1960: *Some elementary theory for design*. John Wiley. New York.

Deneufchatel, D. & J. C. Porchier. 1993: *Area Frame Surveys and Remote Sensing: Guide to Surveying by Segments and Points.* Study Group on Food and Agricultural Statistics in Europe, 5 – 9 July 1993. FAO. Geneva.

Department of Agriculture 1995: *White Paper on Agriculture*. ISBN 0-621-16111-x. Pretoria.

Dobbins, R. *et al.* 1994: *Construction of a Statistical Area Sampling Frame using Remote Sensing and Geographic Information Systems: The Statistics Canada Experience.* Statistics Canada. Ottawa.

Dunleavy, K. P. 1988: *The Stratification Process*. John Wiley. New York.

Ebdon, D. 1985: *Statistics in Geography*. Basil Blackwell. New York.

Faulkenberry, G. D. and A. Garoui. 1991: Estimating a Population Total Using an Area Frame. *Journal of the American Statistical Association, Vol. 86, No. 414, pp. 445 – 449, Theory and Methods*. American Statistical Association. Washington, D. C.

Fecso, R. & V. Johnson. 1981: *The New California Area Frame: A Statistical Study.* Statistical Research division - United States Department of Agriculture. Washington, D. C.

Fecso, R. *et al*. 1986: Sampling Frames for Agriculture in the United States. *Journal of Official Statistics, Vol. 2, No. 3, pp. 279 – 292.* Statistics Sweden.

Food and Agricultural Organisation Statistical Development Series. 1996: *Conducting Agricultural Censuses and Surveys.* Food and Agriculture Organisation of the United Nations. Rome.

Ford, B. L. *et al.* 1986: Area Frame Estimators in Agricultural Surveys: Sampling versus Non-sampling Errors. *Agricultural Economics Research, Vol. 38, No. 2, pp. 1 – 9.* Statistical Reporting Service – United States Department of Agriculture. Washington, D. C.

Foreman, E. K. 1991: *Survey Sampling Principles.* Marcel Dekker. New York.

Franke, R. 1982: Smooth Interpolation of Scattered Data by Local Thin Plate Splines. *Comp. & Maths. With Applls. Vol. 8, No. 4, pp. 237 – 281.* London.

Gangkofner, U. *et al.* 1990: *Regional Inventory 1988/89 in Bavaria.* Conference on the Application of Remote Sensing to Agricultural Statistics, 10 – 11 Oct 1989. Joint Research Centre. Varese, Italy.

Garibay, R. *et al.* 1996: *Area Frame Point Sampling: An Exploratory Study to Measure Nicaragua Agriculture Production.* National Agricultural Statistics Service – United States Department of Agriculture. Washington DC.

Geuder, J. 1984: *Paper Stratification in SRS Area Sampling Frames.* Statistical Research Division – United States Department of Agriculture. Washington, D. C.

GUS/NASS Project. 1995: *Pilot Point Sample Area Frame – Poland.* Survey Specifications. Draft 3.

Gonzalez Alonso, F. *et al.* 1990: Comparison of JRC and USDA Segments. *Conference on the Application of Remote Sensing to Agricultural Statistics, 10 – 11 Oct 1989, pp. 111 – 117.* Joint Research Centre. Varese, Italy.

http://www.quantdec.com. 2000.

International Training. 1987: *Area Frame Development and Sampling.* Area Frame Section – NASS – USDA. Washington, D. C.

Isaaks, E. H. & R. M. Srivastava. 1989: *An Introduction to Applied Geostatistics.* Oxford University Press. New York.

Konijn, H. S. 1973: *Statistical Theory of Sample Survey Design and Analysis.* North – Holland. Amsterdam.

Kott, P. S. 1990: *Mathematical Formulæ for the 1989 Survey Processing System (SPS) Summary.* National Agricultural Statistics Service – United States Department of Agriculture. Washington, D. C.

Krige, D. 1990: *A Sampling Frame for the Greater Durban Area.* University of Natal. Durban.

Lehtonen, R. 1995: *Practical Methods for Design and Analysis of Complex Surveys.* John Wiley & Sons. New York.

Lepoutre, D. 1990: Crop Area and Production Evaluation for Agri-Business. *Conference on the Application of Remote Sensing to Agricultural Statistics, 10 – 11 Oct 1989, pp. 145 – 149.* Joint Research Centre. Varese, Italy.

Ministry for Agriculture and Land Affairs. 1998: *Agricultural Policy in South Africa – A Discussion Document.* National Department of Agriculture. Pretoria.

Mitas, L. & H. Mitasova, 1988: General Variational Approach to the Interpolation Problem. *Comput. Math. Applic. Vol. 16, No. 12, pp. 983 – 992*. London.

National Department of Agriculture, 1998: http://www.nda.agric.za/policies.

Narciso, G. 1996: *Area Frame Case Study: The Development of a Remote Sensing and Geographical Information System for the Estimation of Crop Area Statistics.* SAC/GDTA Training Course, 7 – 11 Oct 1996. Agricultural Research Council – Institute for Soil, Climate and Water. Pretoria.

Oliver, M. A. & R. Webster. 1990: Kriging: A Method of Interpolation for Geographical Information Systems. *Int. J. Geographical Information Systems, Vol. 4, No. 3, pp 313 – 332.* London.

Potgieter, A. B, G. Narciso & C. von Hagen. 2000. *A Geographical Information System Framework for Obtaining Livestock Statistics for the Gauteng Province in South Africa: Unpublished draft.* Pretoria.

Robinson, A. H., *et al.* 1984: *Elements of Cartography (5th ed.).* John Wiley. New York.

Sarndell, C. E. *et al.* 1992: *Model Assisted Survey Sampling.* Springer – Verlag. New York.

Siu-Ngan Lam, N. 1983: Spatial Interpolation Methods: A Review. *The American Cartographer, Vol. 10, No. 2, pp 129 – 149.* New York.

Stakenborg, J. 1990: *Data Treatment for Crop Statistics.* Conference on the Application of Remote Sensing to Agricultural Statistics, 10 – 11 Oct 1989, pp. 139 – 144. Joint Research Centre. Varese. Italy.

Stopher, P. R. & A. H. Meyburg. 1979: *Survey Sampling and Multivariate Analysis for Social Scientists and Engineers.* D. C. Heath and Company. Lexington. Massachusetts.

Thompson, M. 1996: A Standard land-cover classification scheme for remote-sensing applications in South Africa. *South African Journal of Science, Vol. 92, pp.34 – 43.* Pretoria.

Thompson, S. K. 1992: *Sampling.* John Wiley & Sons. New York.

USDA. 1983: *Scope and Methods of the Statistical Reporting Service.* Publication No. 1308. Washington D. C.

Watson, D. F. & G. M. Philip. 1985: A Refinement of Inverse Distance Weighting Interpolation. *Geoprocessing, 2, pp 315 – 327.* Elsevier. Amsterdam.

Wendroth, O. & D. R. Nielson. 1999: *Sampling Field Soils And Their Vegetation: Learning Spatial and Temporal Statistics By Example: Unpublished draft.* New York.

Wyatt, P. 2000: *The Interpolation Process.* Directions Magazine. http://www.directionsmag.com /article.asp?ArticleID=52.

APPENDIX A – DESCRIPTIONS OF THE SOUTH AFRICAN LAND-COVER CLASSES IN THE GAUTENG PROVINCE

These descriptions of the land-cover classes are taken from Thompson M, 1996. This land-cover data was used to stratify the area sample frame.

- Forest and Woodland - all wooded areas with more than 10% tree canopy cover, where the canopy is composed of mainly self-supporting, single stemmed, woody plants >5m in height. Essentially indigenous tree species, growing under natural or semi-natural conditions. Excludes planted forests and woodlots. Typically associated with the Forest and Savannah biomes in South Africa.

- Forest - Tree canopy cover > 70%. A multi-strata community, with interlocking canopies, composed of canopy, subcanopy, shrub and herb layers.

- Thicket and Bushland - communities typically composed of tall, woody, self-supporting, single and/or multi-stemmed plants, with, in most cases, no clearly definable structure. Total canopy cover > 10%, with canopy height between 2-5m.

- Grassland – all areas of grassland with less than 10% tree and/or shrub canopy cover, and greater than 0,1% total vegetation cover. Dominated by grass-like, non-woody, rooted herbaceous plants. Typically associated with the Grassland and Savannah biomes.

  - Unimproved grassland – essentially indigenous species, growing under natural or semi-natural conditions. Typically associated with the Grassland biome. 694 894,73 ha (41,25 %) of this class covers the province. It is by far the majority land-cover in Gauteng.

  - Improved grassland – planted grassland, containing either indigenous or exotic species, growing under man-managed conditions for grazing, hay or turf production, recreation (e.g. golf courses), etc.

- Forest plantations – all areas of systematically planted, man-managed tree resources, composed of primarily exotic species. Category includes both young and mature plantations that have been established for commercial timber production, seedling trials and woodlot/windbreaks of sufficient size to be identified on satellite imagery. Includes clear-felled stands within plantations. Excludes all non-timber based plantations such as tea and sisal, as well as orchards used in the production of citrus or nut crops.

- Waterbodies – areas of (generally permanent) open water. The category includes natural and man-made waterbodies, which are either static or flowing, and fresh, brackish and salt-water conditions. This category includes features such as rivers, dams (i.e. reservoirs), permanent pans, lakes, lagoons and coastal waters.

- Wetlands – natural or artificial areas where the water level is at (or very near) the land surface on a permanent or temporary basis, typically covered in either herbaceous or woody vegetation cover. The category includes fresh, brackish and saltwater conditions. Examples include salt marshes, pans (with non-permanent water cover), reed-marsh or papyrus-swamps and peat bogs.

- Barren lands – non-vegetated areas, or areas of very little vegetation cover (excluding agricultural fields with no crop cover, and opencast mines and quarries), where the substrate or soil exposure is clearly apparent. 1 137,78 ha (0.07 %) is considered barren in Gauteng. This is a combination of the two degraded vegetation classes and the dongas & erosion class.

  - Degraded land – permanent or seasonal, man-induced areas of very low vegetation cover (i.e. removal of tree, bush and/or herbaceous cover) in comparison with he surrounding natural vegetation cover. Category includes major erosion scars (i.e. sheet and gully erosion). Sub-divided by vegetation classes i.e. Degraded-Woodland, and Degraded-Grassland wherever possible to allow reconstruction of full class extent. Typically associated with subsistence level farming and rural population centres, where overgrazing of livestock and/or wood – resource removal has been excessive. Often associated with severe soil erosion problems.

  Characterised on satellite imagery by significantly higher overall reflectance levels (i.e. whiter appearance) and lower NDVI values (in comparison with the surrounding vegetation).

- Cultivated land – areas of land that are ploughed and/or prepared for raising crops (excluding timber production). The category includes areas currently under crop, fallow land and land being prepared for planting. 349 194,73 ha (20,73 %) is covered by cultivated land (permanent or temporary) in Gauteng. Most of this is land used in temporary-commercial dryland agriculture.

  Unless mapping scales allow otherwise, physical class boundaries are broadly defined to encompass the main areas of agricultural activity, and are not defined on exact field boundaries. As such the class may include small interfield cover types (i.e. hedges, grass strips, small windbreaks, etc.), as well as farm infrastructure.

  - Permanent crops – lands cultivated with crops that occupy the area for long periods and are not replanted after harvest. Examples would include tea plantations, vineyards, sugar cane and citrus orchards, hops and nuts.

  - Temporary crops – land under temporary crops (i.e. annuals) that are harvested at the completion of the growing season, which remains idle until replanted. Examples would be maize, wheat, legumes, potatoes, onions and lucerne.

  - Definitions applied to agricultural practices within the context of the remote-sensing classification:

(a) Subsistence/semi-commercial cultivation – characterised by numerous small field units in close proximity to rural population centres. Typically dryland crops produced for individual or local (i.e., village) markets. Low level of mechanisation.

(b) Commercial cultivation – characterised by large, uniform, well managed field units, with the aim of supplying both regional, national and export markets. Often highly mechanised.

(c) Irrigated/non-irrigated – major irrigation schemes (i.e. areas supplied with water for agricultural purposes by means of pipes, overhead sprinklers, ditches or streams) are characterised by numerous small farm-scale irrigation dams, close proximity to major water sources and/or centre pivot irrigation systems.

- Urban/built-up land – an area where there is a permanent concentration of people, buildings and other man-made structures and activities, from large villages to city scale. 318 471,24 ha (18,91 %) is covered by all of the urban classes.

  Small rural communities are often included within the surrounding land-cover category (i.e. subsistence/semi-commercial agriculture) if mapping scales do not permit identification of such settlements as individual features.

  Where mapping scales permit, the limits of the urban boundary are delineated to exclude open areas within the built-up region (i.e. vegetated or non-vegetated areas with few or no structures).

  - Residential – areas in which people reside on a permanent or near-permanent basis. The category includes both formal (i.e. permanent structures) and informal (i.e. no permanent structures) settlement areas, ranging from high to low building densities (including smallholdings on the urban fringe).

  - Commercial – non-residential areas used primarily for the conduct of commerce and other mercantile business, typically located in the central business district (CBD).

  - Industrial/transport – non-residential areas with major industrial (i.e. manufacture and/or processing of goods or products) or transport related infrastructure. Examples would include power stations, steel mills, dockyards and airports.

- Mines and quarries – areas in which mining activity have been done or are being done. Includes both opencast mines and quarries, as well as surface infrastructure, mine dumps, etc., associated with underground mining activities.

APPENDIX B – RANDOM SELECTION SCRIPT

The following is the Avenue (ArcView programming language) code for selecting features in a random manner. This script was used to randomly select the land parcels that were surveyed.

'Selects a percentage of a theme's features randomly'

```
theView = av.GetActiveDoc
theDisplay = theView.GetDisplay
ThemeList = av.GetActiveDoc.GetActiveThemes
if (ThemeList.Count <> 1) Then
 msgbox.Error("You must have one and only one active theme","Select Random Sample is aborting")
 exit
end
theTheme = ThemeList.Get(0)
theFtab = theTheme.GetFtab
theBitMap = theFtab.GetSelection
theBitMap.ClearAll
TotalRecs = theFtab.GetNumRecords
PercentSelect = MsgBox.Input("Enter the percentage of Total Population:","Select Random Sample","10")

if ((PercentSelect = NIL) or (PercentSelect.IsNumber.Not)) then exit end
if((PercentSelect.AsNumber > 99) or (PercentSelect.AsNumber < 1)) then
  Msgbox.Error("You must enter a number between 0 and 100","Select Random Sample is aborting")
  exit
end
selectnumber = (PercentSelect.AsNumber/100 * TotalRecs).Truncate
if ( Msgbox.YesNo("Do you require an exact sample percentage?","Select Random Sample",TRUE) ) then
 setbits = 0
 while (setbits < selectnumber )
  newnumber = Number.MakeRandom ( 0,  totalrecs)

  theBitMap.Set(newnumber)
  setbits = theBitmap.Count
 end
else
 for each n in 1..selectnumber
  newnumber = Number.MakeRandom ( 0,  totalrecs)
  theBitMap.Set(newnumber)
 end
end
theDisplay.Invalidate(True)
```