Pakistan Space and Upper Atmosphere Research Commission

Pakistan: *How SUPARCO Makes Crop Forecasts and Estimates*

based on integral use of RS data

A joint FAO, UN & SUPARCO publication

2012

PREFACE

Timely crop forecasts & estimates are crucial to both policy makers and other stake holders involved in taking decisions, initiating policies, planning, investment and marketing. In today's fast paced information age, the statistical methods described in this manual provide a pragmatic, rapid and cost effective solution for the provision of timely and verifiable information on crop growth, forecasts and estimations. The technique would subsequently assist in establishment of benchmark against which other data sources of crop data could be compared. SUPARCO has a solid record of objectivity and ability to meet the target date for reporting. Objectivity provides crop estimates credibility because it is assured that the estimates are based on scientific principles of probability sampling and not subjective opinions. SUPARCO also provides forecasts and crop estimates in an unbiased, credible and timely manner.

This manual has been developed in consultation with FAO, UN and with the funding support of USDA for use as training manual by the Crop Reporting Services (CRS) personnel. It also highlights the work carried out by SUPARCO in collaboration with FAO, UN in improving forecasting and estimation methodologies in Pakistan.

John S. Latham
Senior Land and Water Officer (Geospatial)
NRL - UN/FAO

ISBN : 978-969-9102-08-0

TABLE OF CONTENTS

1. Introduction

Monitoring crop growth and production of early forecasts of planted crops is of immense importance for food security, planners, policy makers and other stakeholders in Pakistan. Reliable, timely and credible information enables planners and decision makers to handle deficits or surpluses of food crops in a given year in an optimum manner. SUPARCO, during each cropping season, publishes crop forecasts and estimates of the major crops in Pakistan. These estimates are used as benchmarks in the system because of their objectivity and timeliness. The estimates also feed into the agricultural system and assist in decision making process of planners, policy experts as well as food security professionals. Farming and business communities are also now using this data.

2. SUPARCO Crop Production Forecasts and Estimates

There are two components of crop forecasts and estimates:

a. Estimation of area planted and area to be harvested
b. Estimation of yield per hectare or acre.

These components are multiplied to obtain forecasts and estimates of total production. They are obtained differently and documented in separate sections of this publication.

3. Estimation of Area Planted and to be Harvested

Timely and reliable national statistics of a country's agricultural sector can be obtained through the establishment of an adequate, periodic, national agricultural survey based on probability sampling methods, image classification and adhering to well defined and reproducible techniques.

SUPARCO acquires wall to wall satellite imagery and conducts field surveys each year to establish the area planted and to be harvested during Rabi (Spring) and Kharif (Autumn) cropping seasons. The imagery pertaining to the planted area is acquired twice through the SPOT satellite constellation, while the field survey is conducted once during the season. It takes five to six weeks to collect and summarize the data. Thirty SUPARCO professionals work on satellite image processing, data collection, summarization and estimate of planted areas for each crop.

This report explains how the adopted methodology is both objective and credible; moreover the estimates produced with the small field sample are also timely and efficient.

Image Map of Pakistan

The following two methods are being used by SUPARCO for area estimation:

a. Area Frame Sampling Technique

b. Satellite Image Classification

3.1 Area Frame Sampling Technique

- An agricultural mask of the area of interests is prepared based on the satellite imagery of 5m resolution during the peak crop growth seasons of February/March for Rabi crops and in September for Kharif crops.
- SUPARCO utilizes satellite-based area frame sampling developed in collaboration with Land & Water Geospatial Unit of FAO-UN, which is part of a fully operational system for the estimation of crop areas.
- The Satellite based area frame technique involves a three stage stratification process to group districts and homogenous areas in order to use statistical inference to estimate crop areas.
- Stratification makes it possible to produce more accurate estimates by reducing variability between samples. Each stratum becomes a separate population.

These statistical procedures are all part of a scientific probability survey system, implemented by SUPARCO. The following types of stratification are made on the satellite imagery with the help of GIS techniques:

a. Administrative stratification by grouping districts with similar cropping practices.

b. Agriculture land and non-agriculture land.

c. Agriculture land based on cropping pattern, irrigation systems and cropping intensity.

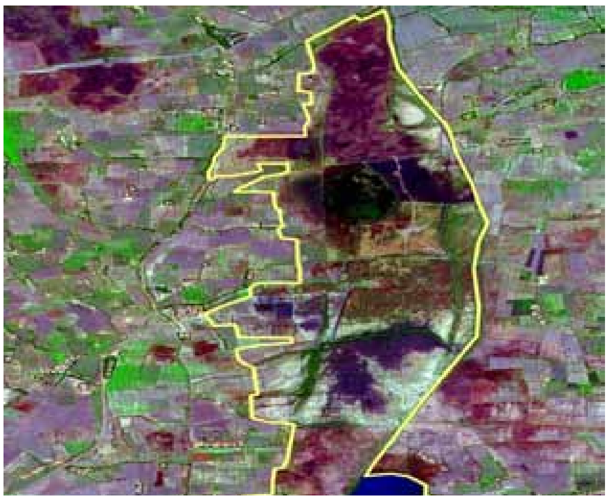
Each agricultural zone is divided into Primary Sampling Units (PSU) of approximately 1000 ha.

These PSUs are then assigned strata IDs based on the cropping intensity as in the next table:

S #	Strata ID	Visible Area Planted
1	11	Greater than 75%
2	12	50% to 75%
3	21	25% to 50 %
4	42	0-25%

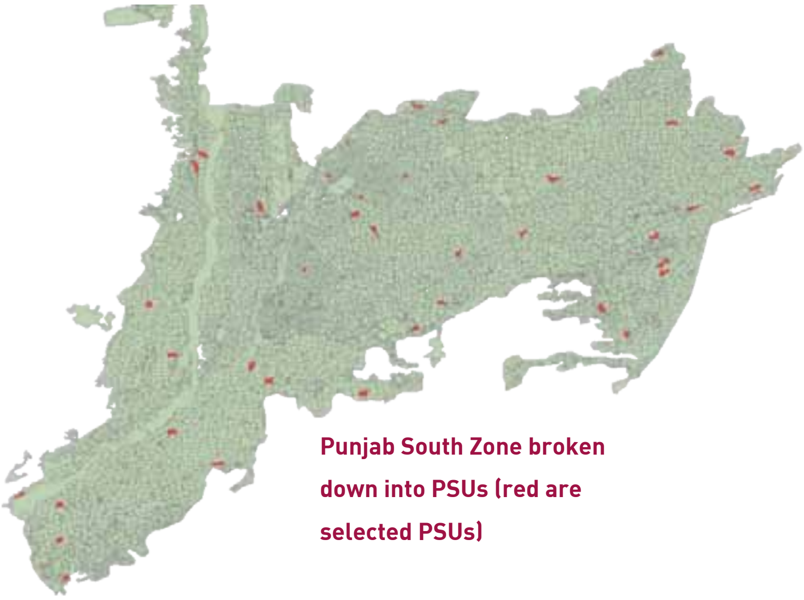
Agricultural land with strata 11

Agricultural land with strata 12



Agricultural land with strata 42

- A systematic random selection is made to select a sample of Primary Sampling Units (PSUs).
- Selected PSUs are sub-divided into Secondary Sampling Units (SSUs) of nearly equal size (300-400 Ha) keeping in view the sanctity of strata.
- The selected SSUs are further divided into Terminal Sampling Units (TSU) having a size of approximately 30 hectares.
- Finally, a sample of TSUs is selected and prepared for data collection in the field.



Punjab South Zone broken down into PSUs (red are selected PSUs)

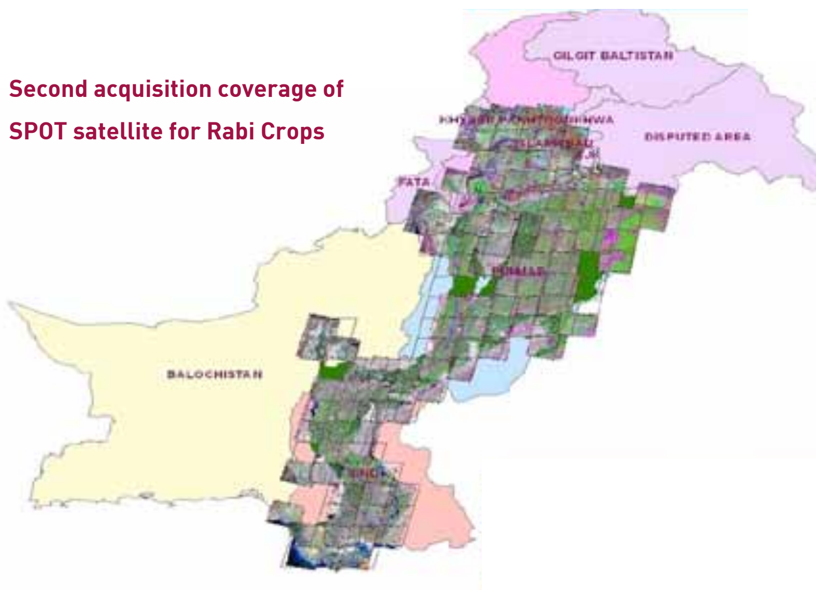
The picture above shows PSUs selected in southern zone of the Punjab province.

- Professional enumerators travel to the selected segments and complete a total census on the 30-hectare sample areas.
- Every crop and land cover inside the segment is recorded.
- This total enumeration of a 30-hectare segment takes approximately 3 hours. Before the enumerator leaves the segment, it is assured that the entire segment is covered and no field and area is left without survey (even areas of mosques and cemeteries etc are brought into account).
- In most cases, the enumerator does not measure any field because measurements have been accomplished on the satellite image.
- This ground information is fed into software and registered to identify different crops on the satellite imagery during the digital image processing and classification processes.

3.2 Satellite Digital Image Processing and Classification

The SPOT satellites orbit the earth in sun-synchronous near polar orbit about 830 kilometres above sea level. As the satellite passes overhead, it records solar electromagnetic radiation in the visible light and near infrared ranges in 5 spectral bands. The spatial resolution of the imagery is approximately 5 metres. The satellite has the capability to tilt its imaging cameras to collect imagery 450 km left and right of its track, providing the opportunity to cover the same area within 3 to 5 days depending on location. SUPARCO downloads this satellite imagery from SPOT satellite at the receiving station in Islamabad on a daily basis.

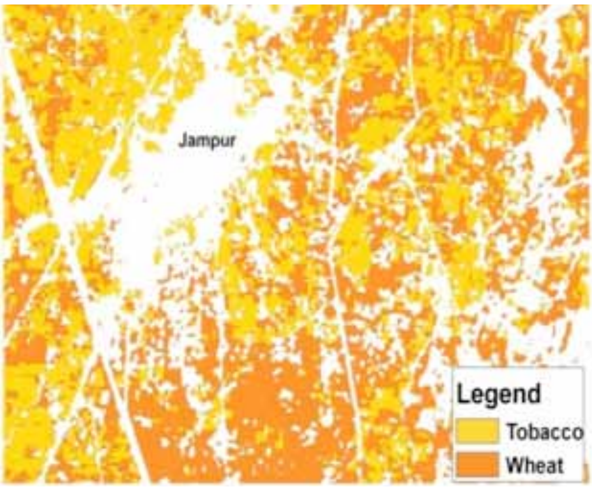
The current SPOT imagery becomes another source of crop information that employs digital



Second acquisition coverage of SPOT satellite for Rabi Crops image classification process. Different crops absorb, reflect and scatter solar electromagnetic radiation differently in different wavelength. When calibrated properly with ground data, computer software having a registry of temporal/ spectral signatures is used to assist identification of different crops. In the Rabi season, crop identification is easier because there is a great deal of wheat and less sugarcane and fodder.

The sugarcane has a unique spectral reflectance (at least during period of its growth) so, it is easy to separate it from wheat especially when multi-date imagery is used. The fodder has a bright tone as compared to wheat which is easily distinguishable.

Satellite Imagery - District Rajanpur



Crop Identification - District Rajanpur

The satellite analyst takes the current imagery of the crop season at a time when the crops are mature enough to provide unique reflectance and stores reflectance of each crop in the system. Then the software takes one picture element (pixel) at a time and assigns it as to the crop with the most similar signature.

At the end, there may be some misclassification errors but in general, the classification is consistent and the errors can be removed with mathematical techniques. SUPARCO professionals are in the process of improving the methodology through the use of the latest software and classification techniques. No direct crop acreage estimation is made based solely on pixel counting.

4. Crop Yield Forecasting/ Estimation System

Statistical models are being used by SUPARCO for the estimation of the yield. The following parameters are being incorporated on decadal basis in the model:

1. Data base consisting of 15 years temporal data on satellite VGT, crops, agro met, fertilizers, irrigation water supplies has been developed.

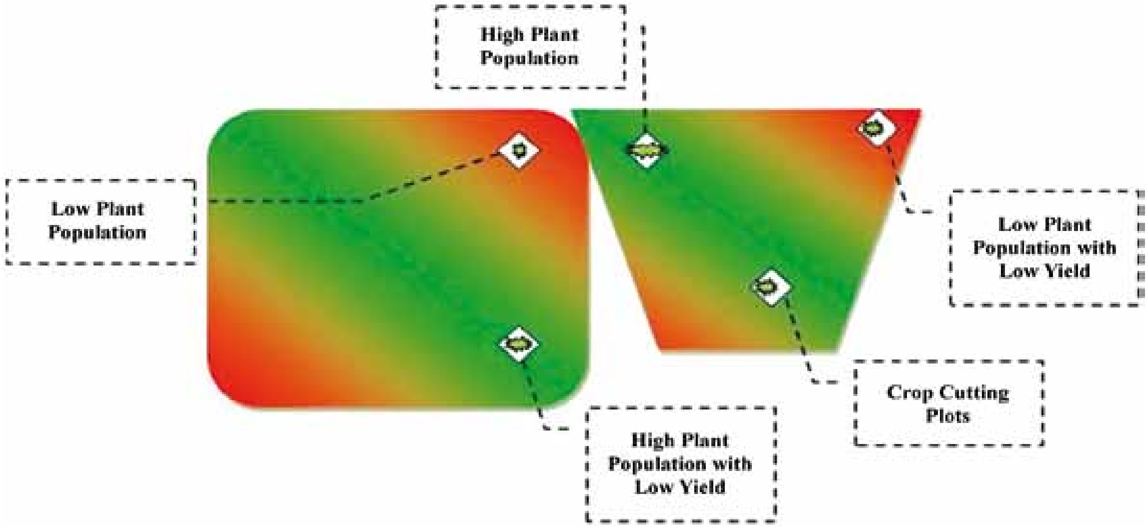
2. Current season information on:

    a. Agro-met parameter (Temperature regimes, rainfall, sunshine duration)
    b. Satellite VGT NDVI
    c. Fertilizer - Monthly off take by district
    d. Irrigation - 10 daily water withdrawals from head works

- Crop yield is a major component for the crop production forecast and estimation. SUPARCO starts with Crop Reporting Service' s (CRS) yield estimates and refines them based on meteorological conditions.

- Provincial CRSs (PCRSs) provide yields for major and minor crops. There is one crop yield forecast, which is opinion based, and a crop yield estimate, which is based on crop cuttings experiments at the harvest time.

- The sample villages of the provinces are used as the master sample for listing fields.



- PCRS enumerators visit selected villages in order to carry out the crop cuttings experiment based on random table techniques.
- They select three crop fields in each village and carry out crop cuts of two sample plots of 20x15 square feet (major crops) and 15 ft x 20 ft (minor crops) as in the figure above.
- These crop cut surveys act as reference information from fields for a given season and reflect the availability of different crop inputs like irrigation water, fertilizers usage, pesticides applications and prevailed weather impacts.
- This information is then compiled at district and provincial level to produce average crop yield estimates based on harvesting as well taking into consideration the adjustment in loss due to harvesting procedure etc.
- Crop cuts are only possible at harvest time and do not provide information about actual crop growth and its relation with final yields.
- Furthermore, CRS based system does not provide a possibility for crop yield forecasting for early warning related to crop yield assessments.
- SUPARCO starts with the CRS crop yield estimates and adjusts them based on remote sensing and GIS techniques.
- Crop yield research at SUPARCO is currently focused on crop yield forecast modelling for early assessments and warning through monitoring of vegetation dynamics, agricultural input situations like water, fertilizers etc.
- This yield modelling component is based on the modular FAO-UN approach.

Fields being prepared with traditional technology

Off season cultivation of high value crops

CROP CUT PROCEDURE FOR WHEAT

Step 1
Pegging to fix corners of 15ft X 20ft sample plot for crop yield Estimation

Step 2
Further explanation of the procedure

Step 3
In situ threshing floor

Step 4
In situ threshing demonstration

5. The Area Frame versus Village List Methodology

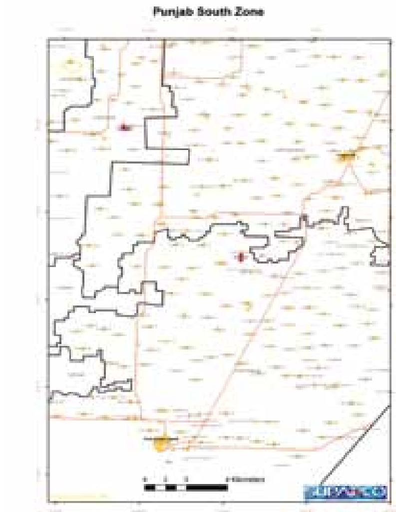
Both SUPARCO's segment area frame and the CRS village list frame require sampling frames, probability sample selection, data collection, raising factors and summary. Moreover, both sampling designs are estimating the same target population parameters, such as total wheat production or total rice production etc.

The village list methodology is called Village Master Sample (VMS) Frame.

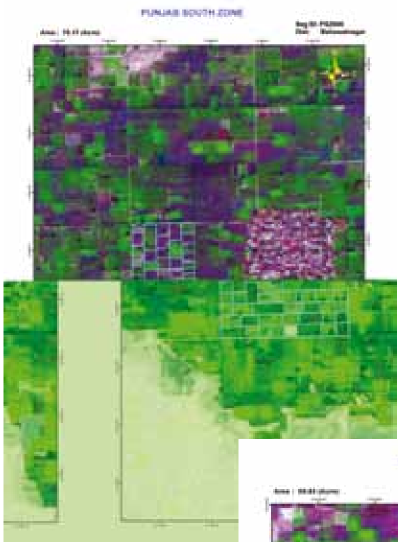
- This system was initially designed during 1978-84 in Punjab and revised during 2005-06.
- The system is based on land revenue records. 26109 Punjab villages have been stratified into small, medium and large based on areas of wheat, cotton, rice and sugarcane.
- Sample villages (1240) have been selected on the basis of probability proportion representing 5% of the total cropped area of Punjab. Total sampled area of Pakistan is 1.2 million ha.

The segment area frame of SUPARCO.

- It is based on satellite imagery of 5.0 m resolution acquired during peak growing seasons of September and February.
- The stratification is carried out on the basis of land cover features and cropping density. The principle of probability proportion is applied and 366 segments of 30 ha each were drawn. The total sampled area is about 11000 ha that is 1% of the sampled area of Crop Reporting Services.
- Satellite imagery and GIS programs allow the identification of agriculture land in each district and province. The land that is cropped can be separated into one of four cropland intensity strata's. Large towns, barren lands, lakes and non-agriculture land can also be separated.
- The area frame cropland is mostly cropland with other categories of land excluded (see section 3 of this document).
- GIS programs subdivides the land into 30-hectare parcels of land called segments. This process is carried out in several steps in a way that doesn't necessitate dividing all agriculture land into 30-hectare sampling units. However, at the end of area frame construction, the total number of 30-hectare sampling units (N) is known. Capital N is important because it provides the raising factor in each stratum.



Guide Map to locate a segment



Segment P1-06

Segment P1-07

- A random sample of 366, 30-hectare sampling units (called segments) is selected to represent total cropped area of Pakistan. This is a very small sample but even a small sample that is representative can provide accurate timely estimates.

- After the 366 segments are selected, they are prepared for data collection in the field. Imagery is identified for each selected segment of 30-hectares; maps are prepared for enumerators who visit the field and collect data.

- Enumerators are trained to visit each segment in the field and collect data for every field and land cover inside the segment. Even cemeteries, mosques, factories and housing units must be identified and recorded. For the fields, the crop or land cover is identified.

- Data for each segment are summarized and prepared for crop estimation.

- SUPARCO has chosen to implement the most direct estimator, the closed segment estimator. This means that data are collected for each segment and summarized by adding all the crops of like type together. When segment totals, they are multiplied by the raising factor, N/n each stratum at a time. This estimator works for crops, livestock, households and farm economic data.

- A simplified version of the direct expansion estimator is:

$$\hat{y}_c = \sum_{h \varepsilon C} N_h [\bar{y}_h]$$

Where $\hat{y}_C$ is the estimate of the variable y in population C.

$N_h [\bar{y}_h]$ is total number of SSUs in stratum h multiplied by the average segment value of the variable in stratum h.

$\sum_{h \varepsilon C}$ is the sum of all h strata in population C.

- Within each stratum, the average crop acrege of the sample is calculated and this average is multiplied by N, & the total number of segments in the stratum.

- An alternative way to think about the raising factor is to calculate the probability of selection for a segment. The probability of selection is n/N where n is the sample number and N is the population number. These numbers are calculated within each stratum.

- The raising factor is the reciprocal of the probability of selection. If the probability of selection is 5/1200 then the raising factor is 1200/5. We will use this principle in the next section when presenting the village list sample design.

In the booklet on Area Frame (AF) Sampling (AF), Earl Houseman wrote. "The simplicity of the idea (of AF sampling) is in striking contrast to the complexity of successful application of the concepts. A high proportion of the problems found in the application of AF sampling in agriculture are characteristic of the survey populations and therefore, common to all suvey methods, sampling or census. However, survey methods differ considerably with regard to effectiveness, or potential effectiveness in coping with practical problems that exist". Area frame sampling for agriculture has been developed to control both sampling and non sampling errors and is the best probability sampling system in the world. This cautionary quote by Earl Houseman above, is appropriate for any survey methodology.

The next part of this quote concerns selecting a random sample. In AF sampling, the sample is selected using sophisticated scientific sampling procedures. Systematic sampling is most efficient when the sample list is put in an ordered array. Population totals can be estimated accurately.

When surveys fail to give accurate results, they fail in data collection. The AF system allows control over the data collection process. Survey methods differ considerably with regard to effectiveness, or potential effectiveness in coping with practical problems that exist. Pakistan has its own set of practical problems in the field and adjustments are continuously being made to improve these crop estimates.

The village list frame design used by the CRSs requires a complete list of villages without omission or overlaps just as the area frame requires a list of all cropland without omission or overlaps. It is our understanding that the mouazas fulfil this requirement. Each mouza contains one or more villages and it has permanent boundaries. A sample is then selected from the mouza list using probabilities proportional to the size of the village where size is the number of hectares.

Villages vary in size, some are very large and some are very small. Assume for a moment that we have the perfect list of village mouzas in a province, the list is complete and there is no overlap and associated with each village is the size of the village in number of hectares.

The table below provides village information, its size in hectares, an accumulation of hectares and the probability of selecting the village.

S #	Village no.	Hectares	Accumulation	Sample	Probability	Probability
1	Village 1	254	1060	1-254	0.0230	254/11060
2	Village 2	572	1632	255-826	0.0517	572/11060
3	Village 3	45	1677	827-871	0.0041	45/11060
4	Village 4	2078	3755	872-2949	0.1879	2078/11060
5	Village 5	20	3775	2950-2969	0.0018	20/11060
6	Village 6	34	3809	2970-3003	0.0031	34/11060
7	Village 7	89	3898	3004-3092	0.0080	89/11060
8	Village 8	1680	5578	3093-4772	0.1519	1680/11060
9	Village 9	567	6145	4773-5339	0.0513	567/11060
10	Village 10	245	6390	5340-5584	0.0222	245/11060
11	Village 11	94	6484	5585-5678	0.0085	94/11060
12	Village 12	74	6558	5679-5752	0.0067	74/11060
13	village 13	689	7247	5753-6441	0.0623	689/11060
14	Village 14	2346	9593	6442-8787	0.2121	2346/11060
15	Village 15	58	9651	8788-8845	0.0052	58/11060
16	Village 16	842	10493	8846-9687	0.0761	842/11060
17	Village 17	28	10521	9688-9715	0.0025	28/11060
18	Village 18	333	10854	9716-10048	0.0301	333/11060
19	Village 19	1012	11866	10049-11060	0.0915	1012/11060

Assuming that two villages are selected (9 and 14), the following will be accomplished:

- Enumerators go to these two villages and collect data for wheat fields from each village.
- Then the wheat in village 9 is raised to the population total of all 19 villages by multiplying the wheat by the reciprocal of the probability of select or 11060/567 and the wheat in village 14 is collected and added and multiplied by the reciprocal of the probability of selection or 11060/2346.
- There are two estimates of wheat for the 19 villages and we average the two estimates.

What we have described is standard village list methodology. Any procedure must be close to this methodology described above. Notice that large villages have higher chance to be selected. Also notice that this is one stage sampling. Further, notice that when a large village is selected, there is a smaller raising factor and a smaller village has a smaller probability of selection and a larger raising factor.

The differences between these two methods are subtle. Both are probability samples and both in theory are reasonable.

The area frame based on land can be stratified better. Moreover, land can be listed without overlap or omission in a straight forward way.

The village list depends on the number of hectares. Households are less stable so the probabilities based on old census listing of households will be soon out of date. In addition, data collection from 30 hectare segments can be done accurately but data collection from a village with 2000 households will take months even then is not or may not be accurate.