



**Space and Upper Atmosphere
Research Commission (SUPARCO)
and
Food and Agriculture Organization
Of the United Nations (FAO)**

Project: *“Agricultural Information Systems – Building Provincial Capacity in Pakistan for Crop Estimation, Forecasting, and Reporting based on the integral use of Remotely Sensed Data. GCP/PAK/125/USA.”*

**BTO Report
William H Wigton**

March 20 to April 7, 2012

This report is an effort to improve and document crop statistics methodology at SUPARCO.

Table of Contents

TOR 3
Summary 4
Main Report 5
Additional Research 5
Identify where the wheat went 8
Observations and Recommendations 11

Acknowledgements

I had a great deal of help this trip. SUPARCO management was extremely supportive providing me with everything I asked for to make my visit productive and safe. I was fortunate to have been in Islamabad when Imran Iqbal, Director General was in Islamabad and he provided excellent direction. Zuhair Bukhari, Director, provided generous time and support. I worked on a day to day basis with Abdul Ghafoor, Head of the Applications Division and his Application Division. Financing for this phase 2 of the project and my consultancy is coming from USDA and Pakistan through FAO/UN.

For technical support, Bhatti Iftikhar was with me most of the time and many of the Applications Division staffers were also present. I had amazing support from Ibrahim and Ibrar. Ibrahim did the analysis of most all area frame studies and he did it very efficiently. Bhatti took a few pages for the Manual and made a masterpiece in the draft out of it. Ibrar added to our discussion and added pages and understanding about yields in the manual. Dr Hanif was busy working on trying to make sense out of the Crop estimates.

I got to the markets two times, once under the security of FAO and once under the security of SUPARCO. Both times were welcome reliefs. I had delicious food and may have gained a kilogram in weight.



FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Terms of Reference for Consultant/PSA

Name: W.H.Wigton	
Job Title: Consultant (Improvement in Area Frame /Crop estimation system)	
Division/Department: NRL	
Programme/Project Number: GCP/PAK/125/USA	
Location: Islamabad,Pakistan	
Expected Start Date of Assignment: 10-March_ 2012	Duration: 3 weeks – comprising 2 weeks Pakistan 1 week USA - Washington
Reports to: Name: John Latham	Title: Senior Land and Water Officer (Gesopatial)

GENERAL DESCRIPTION OF TASK(S) AND OBJECTIVES TO BE ACHIEVED

Agricultural Information Systems – Building Provincial Capacity in Pakistan for Crop Estimation, Forecasting, and Reporting based on the integral use of Remotely Sensed Data, GCP/PAK/125/USA: The overall project goal is to help the Government of Pakistan enhancing and improving current systems for the integral use of remotely sensed data into existing data collection, analysis and dissemination systems as well as the development of complementary systems to enable the integration of satellite remotely sensed data.

In particular the consultant will :

- 1- review the current approach methodologies on agriculture crop acreage estimations at the federal level and review the potential improvement for implementation at provincial level in Punjab and Sindh including changes , if required in sample intensity, stratification , regression estimation procedures;
- 2- undertake a needs assessment for the design of a capacity building programme with crop reporting services in the field of agricultural statistics and consulting in Punjab and Sindh;
- 3- undertake preliminary training of representatives of the 2 crop reporting services of Punjab and Sindh, and the University of Faisalabad and SUPARCO;
- 4- Make recommendations and develop a work plan for productive improvements in crop acreage estimations and the capacities in the CRS for the 2 provinces and the improvement and finalization of crop bulletins.

The consultant will undertake a two-day backstopping mission in Rome on his way back for topic debriefing.

The consultant will finalise the report and develop the training programme based on findings whiel in USA after the field visit.

Finally the consultant will produce a report to FAO on the mission and activities undertaken.

Summary

The timing of this trip, March 20 to April 7, 2012 was fortunate because SUPARCO was in the process of making the final forecasts for the Rabi (winter) Season. Moreover, it has been an unusual season in that there has been a reduction of hectares according to the scientific probability survey system used by SUPARCO.

I attended a meeting where the Rabi (winter) Season crop estimates were openly discussed. Several pages in this report have been written as a result of that crop report. It seems some progress has been made accepting the scientific area frame estimates at SUPARCO since my last visit. However there are still areas where progress is needed.

According to the probability Area Frame (AF) Survey, there has been a reduction of wheat planted this year. One of the questions brought up during the discussion of the Rabi Season estimates was, “How is the land that is normally used as wheat crop being used this year?” The Punjab Central Zone (PCZ) was where the largest reduction of wheat has been found using the area frame estimates. To answer this question, we studied the 56 segments selected in PCZ and summarized the data for all fields. There were huge increases in Fallow and Fodder in the 2012 Rabi season.

In addition, the sugar cane estimates seem high when compared to the sugar reported by the sugar factories. My first question is, was there a mistake in the data collection? If not, is there a mistake in the sugar reported from the factories? SUPARCO must start questioning other data sources in addition to questioning their own AF probability survey results.

USDA has stipulated in the contract with FAO that SUPARCO would train the provincial Crop Reporting Services (CRS) from Sindh and Punjab Provinces. While I was consulting, SUPARCO organized a training program and implemented training for CRS reporters. While this course was mostly done by the SUPARCO staff I did participate on all four days with lectures and exercises.

I see a need for some research that should be carried out by SUPARCO to improve their crop estimating program. That research activities are identified in Annex I - Additional Research. It is discussed briefly in the main body of the report as well.

In addition, I was asked to develop a manual. A draft of a manual can be found in Annex 4 “*How SUPARCO Makes Crop Estimates, Bringing science to traditional Crop Estimating in Pakistan.*” The manual was developed by myself and Bhatti Iftikhar with substantial contributions by Ibrahim and Ibrar. It is my hope that this manual will make an important contribution to SUPARCO. It was written to explain how SUPARCO generates crop estimates using scientific probability sampling methods. The target readers are the professionals who will use these estimates and need to understand how they are made as well as CRS professionals who want to understand and be trained in Area Frame Surveys.

3.0 Main Report

In the main report of this document, I cover the four main topics: Research Required, Training Course - Flyer, 3) Punjab Central Zone wheat and sugar and a manual that was drafted with Applications Division personnel to explain how SUPARCO makes crop estimates.

This was a full amount of work for my short consultancy of 17 days in country. Some progress has been made from my last trip. It seems the AF estimates are established. There are some areas where progress needs to be improved. The Applications Division is extremely productive implementing a full national project of crop estimation as well as other projects such as Monitoring flood damage, earthquake damage, and other tasks that can be uniquely monitored using satellite remote sensing data and professional staffers.

Because of the importance of these jobs, there is no time for research. However, SUPARCO needs to undertake some research but can't with limited resources and time. I recommend setting up a research section inside SUPARCO that can concentrate on completing research projects without having responsibility of project work.

3.1 Recommended Research

The first item in recommended research is improving digital image processing. A tremendous improvement can be achieved and will be needed if this technology is to be fully exploited.

3.1.1 Digital Image Processing

The digital image processing research should be done with outside support from an expert who has done this for sometime because using SPOT imagery is difficult. I recommend Mike Craig for this task because he has done digital image processing for USDA/NASS, he is retired and he has the software USDA uses. Since he can't travel because of his health, he will have to SKYPE to communicate and there has been some preliminary discussions with Ibrahim and Bhatti. If FAO can't hire him, perhaps the University of Maryland can hire him and provide some assistance. I recommend Mike Craig because he does it digital image processing for a living, is still involved with USDA even though he is retired and has done research using SPOT imagery in the US. (SPOT is extremely difficult to use with this advanced technology because the spectral values are weak and the scenes are small.)

In 2011, Mike conducted research using SPOT to develop regression estimates and the cropland data layer (CDL). A CDL is a map of classified imagery showing each pixel as the crop that was classified. Currently USDA NASS gets approximately 95 percent correct classification. The last time I checked in Pakistan, the classification results were too low to be useful for either the CDL or regression estimation.

One of the reasons classification results are better in the US is that they use large areas in one scene. For example, one Landsat scene is 32,000 sq kilometers while a SPOT scene is about one ninth of that size. One can overcome the small area problem if multiple SPOT scenes can be downloaded in one path. Five to seven SPOT scenes are taken at one time and used as one scene. The Punjab will be taken in strips rather than scenes.

USDA uses the SEE5 Decision Tree classifier and indicates that the results are as good as when he uses the maximum likelihood classifiers.

3.1.2 Yield Crop Cutting Surveys

One issue that has not been resolved is to improve yield forecasting and estimation. CRS is doing objective yield surveys using crop cutting methods that I think are inefficient. I recommend some research to prove or disprove better procedures leading to efficient objective yield surveys. One way to carry out this research is to: 1) Identify large areas (several hectares) of wheat but small enough to be harvested soon after the research. 2) Lay out 40 to 50 samples plots of 3 by 4 ft and collect crop cut data in the fields under study. 3) Harvest the total areas to get yields for the area. 4) Compare the results of the samples with the actual production harvested. This entire process may need to be repeated in several large areas to properly calibrate the crop cuts.

If SUPARCO can complete this research with the CRS, then CRS should also have multiple plots laying out their 15 by 20 plots and harvesting them. The research would determine which is more accurate and which takes less time.

Dr. Hanif showed me a manual that has been translated into Urdu for field work where Eric Waldhaus's suggestions were put into a manual. If SUPARCO wants a consultant, consider an objective yield survey (OYS) expert such as Eric Waldhaus or someone of his experience to support the effort.

I have carried out this same type of research for the soybean OYS. The key is to have more enumerators involved so you can observe any problems with the instructions and then have a small enough area to be able to harvest the fields so that you can compare the objective yield crop cut estimates against the real economic yields obtained from the harvested fields. This research would go a long way to move SUPARCO forward in the field of yield estimation. Moreover, it will help any research being suggested at the University of Maryland. It is difficult to model yields without proper calibration data.

3.1.3 Stratification in Area Frame by regrouping districts

Ibrahim and Bhatti showed me some interesting observations regarding stratification of districts into homogeneous agriculture zones. Punjab Province is currently divided into four Agricultural Zones and Sindh has been divided two agricultural zones. These Agricultural Zones have served

SUPARCO well and moreover, no one has questioned these zones. However, they have observed that some districts are not fitting into the average agriculture zone and may be more similar to an agriculture zone that borders. For example, there are districts in Left Sindh that are more closely aligned with Right Sindh and in Punjab there are some districts that don't produce crops similar to other districts in their agriculture zone. They are looking at setting up additional agriculture zones and in some cases it would be beneficial to regroup. In this first analysis, we suggest some additional subdivisions that seemed to improve the estimates for the Punjab Central Zone. We did analysis and it didn't seem to make much difference however, when estimating a crop area and there are segments in the zone with zero acres in the crop, the CVs are raised substantially. This is the reason wheat estimates are so good in the Rabi Season and why some crops are not estimated well in the Kharif Season.

When the area frame was constructed, SUPARCO was estimating wheat in the Rabi Season. Now they are estimating all the summer crops and finding that some reshuffling would better serve SUPARCO. This regrouping might be observed with an unsupervised clustering algorithm or a mathematical solution exists with a procedure called simulated annealing. SUPARCO will need outside assistance with simulated annealing but not with unsupervised clustering.

I emailed Dr. Sohaib Khan at LUMS. He seems familiar with simulated annealing but not for this application and this is a very specific application. I have a colleague in the US who is familiar with this specific application but it would be better to have Dr. Sohaib provide this support because he will be closer to SUPARCO. I have used this technique in Arkansas and I was able to reduce the sample by 25 % and at the same time reduce the CV by 10%. Perhaps the clustering algorithm readily available at SUPARCO can serve the purpose.

3.2 Training Crop reporting Service Professionals

Annex 1. Training Course – Flyer is made for the CRS professionals in Punjab and Sindh. This is a continuous task and SUPARCO is up to it. A four-day training course was implemented specifically for CRS professionals who collect and summarize data.

SUPARCO has trained Pakistani professional many times for Punjab and Sindh. I suggested that SUPARCO not emphasize area frame construction but how data must be collected to make scientific probability estimates. My reasoning is as follows. When I was Section Head of Area Frame at USDA, I gave talks to our state office professionals. I didn't expect them to construct area frames but I did expect them to collect data and make estimates. Therefore I stressed the data collection and estimation. Still I had to give them enough information about construction so that they understood the principles. I think SUPARCO accomplished this in this training provided to the Punjab and Sindh professionals. Moreover, I think the area frame construction will always remain at SUPARCO in the near future while data collection could well be transferred to CRS institutions in the provinces.

Part of my terms of reference (TOR) for FAO was to discuss training and capacity building inside the CRS institutions. This technology is not taught by formal training as well as on the job (OTJ) training. CRS is ready to have more on the job training programs and actually take over data collection in some districts in Punjab. They have young staff who want to move Pakistan forward. Some have requested letting them help so SUPARCO management should discuss partial involvement in data collection with Punjab. I know this is already being done because CRS is collecting data for SUPARCO. I am suggesting that SUPARCO give the CRS those responsibilities to do the job correctly without compensation as their mandate for estimating crops. I may be behind on this recommendation but it seems feasible to have CRS start to accept the responsibility of the operational aspects to relieve SUPARCO so SUPARCO can perfect the digital image processing and the yield research.

The professionals in the training course are requesting the transfer. This transfer of responsibility can be accomplished one district at a time. If CRS personnel in a district are ready, then transfer the responsibilities and monitor activities. SUPARCO should provide support as if the a piece of hardware is being purchased.

3.3 Rabi Wheat Estimates and What Happened to All the Wheat Hectares? Changing Estimates Based on Subjective Evaluations

SUPARCO has spent money and years of research setting up a scientific method to estimate crops. I have been involved since 2007 but SUPARCO has been involved much longer than this. It is my understanding that the Government of Pakistan requested SUPARCO to use science and best practices to help the government improve crop estimates.

The area frame estimates are based on probability sampling and statistical inference and are scientific. It is easy to make mistakes but when they have been made correctly, then SUPARCO must defend them as scientific. These estimates can be changed based on theoretical considerations or identified current problems with implementation. There are always improvements to methodology that can be made.

However, no person should be able to make big changes to the estimates based on subjective evaluations because SUPARCO estimates must be based on science. **In the United States, the Secretary of Agriculture cannot change estimates that National Agriculture Statistics Service (NASS) makes.** Before every crop release, the Secretary is briefed and he can ask questions about the estimates but he is not permitted to change estimates even when he is sure they are incorrect.

In Pakistan I suggest that identifying problems with stratification or data collection is always up for discussion but not changing estimates except as I present later in this report. If someone wants to make changes in the estimates, then find something wrong with the sample, data collection, edit and data summary.

How should the estimates for Pakistan be made? SUPARCO should set the national estimate first because it is based on the largest sample of 366 segments. Then Punjab Province should be set second

because it is second best estimate. As the area becomes smaller, the estimates become less accurate. It turns out that Punjab Central Zone looks like it has an estimate of wheat that is short. What should be done?

As I mentioned, the national estimate is set first. After that, the province estimates and then the lower estimates are set. If one of the lower level estimates seems out of place and needs to be changed, then add or subtract some from someplace else but you must remember to balance what you take from an estimate by adding to another estimate so the national estimate remains what was originally set.

Table 1 shows the estimates of Punjab Central Zone

Agriculture Zone	sample number n	Wheat Rabi		% change	Estimate	
		2011	2012			
PNE	46	1162.60	1178.20	1.32%	1050.00	-10.72%
PCZ	56	2341.40	1998.50	-17.16%	2200.00	-6.43%
PSZ	78	2739.90	2644.00	-3.63%	2578.00	-6.28%
PPZ	23	396.70	387.80	-2.29%	380.00	-4.39%
Total Punjab		6640.60	6208.50	-6.96%	6208.00	-6.97%
SLS	52	1098.90	944.30	-16.37%		
SRS	42	496.50	574.60	13.59%		
Total Sindh		1595.40	1518.90	-5.04%		
KP-N	20	173.10	183.00	5.41%		
KP-S	24	450.00	466.10	3.45%		
Total KP		623.10	649.10	4.01%		
Baluchistan	25	304.50	345.10	11.76%		
Total Baluchistan		304.50	345.10	11.76%		
Total Pakistan	366	9163.60	8721.60	-5.07%		

The crop reporting board in the USDA makes these smaller adjustments sometimes when they think they are required. The national estimate has not changed, the Punjab estimate hasn't changed but the distribution changed a bit. I think these changes should be possible if there is a good reason and the changes are small. The reasoning is that you are staying with the most accurate estimates you make and the ones that are less accurate you are redistributing some crops. However, the other crops seem to be well distributed, then all smoothing must be moderate.

Another problem came up that I want to discuss in this report. The national sugar cane estimates shows an increase in sugar cane for Pakistan. If this estimate is accepted, then there are consequences that don't

fit well. Let me explain. A hectare of sugar cane produces 70 to 55 tons of sugar juice and this equates to 5.4 mmt of sugar. When you go to the factories, they report about half the sugar you would expect.

What should SUPARCO do and what should it not do? It should definitely review all the data and expansion factors. My guess is if the raising factors are wrong, other crop estimates would show discrepancies. SUPARCO should check the field data to make sure fodder and maize are not being called sugar cane (remember the sugar estimate seems too high). After those issues are checked, then it is time to stick with the area frame estimates. If professionals from the Central Bank, the Planning Commission or the sugar industry try and ask for the estimates to be changed, the correct answer is, show us where we have errors.

If the data and the raising factors check out and are correct, the problem lies in the check data and that is out of your control. SUPARCO looks weak if it comes back and changes estimates to fit in line with check data that may be hiding the truth because of taxes or slippage. I realize I go home and management faces the difficult questions. However, the check data is not your concern or under your control. Establish the area frame estimate. When someone asks, “where did the estimate come from?” hand them the manual and say that data were collected from the samples and estimates were made using proper statistical inference and proper raising factors. The sample is scientifically selected and has provided estimates for three or four years. Please tell me where we have a mistake and we can change the values.

3.4 Documentation of “How SUPARCO Makes Crop Estimates”

Part of my time on this consultancy was spent documenting area frame and image processing methodology being implemented at SUPARCO. While it is a draft, I believe it will eventually make a good publication for SUPARCO to hand out to people who will use the data. I gave Bhatti a few pages and he added some tremendous pictures and written material. This document can be helpful to establish the area frame and image processing.

4.0 Observations and Recommendations

1. There are still questions about errors in data collection although those questions are much reduced. As I understand, SUPARCO professionals are asked to collect data from three or more segments in one day. That is too much for quality data collection. I recommend data be collected for two segments each day in order to improve the quality of data collection.
2. I realize SUPARCO will spend double if enumerators do half the segments in a day. However, probability sampling depends on small samples where data are collected without error.
3. SUPARCO is training Provincial Crop Reporting Service professionals all the time. Over 200 provincial professionals have been trained. The CRS provincial professionals should be ready take over some of the responsibility for field work. The participants in the 4-day workshop requested to implement the survey. Of course, the management must agree but it doesn't have to be done all at once. Punjab Central can wait until Punjab Northeast is trained so it can be turned over one district at a time. It will certainly be a big training job but as more of the data collection is turned over to CRS, more research can be completed by SUPARCO.
4. Field data collection can be improved if the imagery is improved. SUPARCO needs Google Earth quality spatial resolution imagery for the segments. **I recommend using better imagery for segment identification.** It is my understanding that in the near future SUPARCO will be able to download 50 cm data. This resolution will be perfect for segment identification.
5. There are several different types of research that SUPARCO needs to complete. There is research with PDA and GPS equipment, research for image processing and research for yield crop cutting. I have written these research efforts up and they are included in this report.
6. FAO has requested training programs and capacity building programs for the CRS personnel.

Annex I. Research Required

Annex II. Training Course - Flyer

Annex III. Punjab Central Zone Wheat

Annex IV. SUPARCO Crop Estimation Manual