PN-AEE-492
59206

Report #18

Report of GARD Consultancy

Agricultural Statistics
Short-Term Technical Assistance

Aaron C. Johnson, Jr. and John S. Rowe

University of Wisconsin

February - March, 1987

AGRICULTURAL STATISTICS SHORT-TERM TECHNICAL ASSISTANCE

February 12 - March 28, 1987

Aaron C. Johnson, Jr. and John S. Rowe

Prepared for the

Gambia Agricultural Research and Diversification Project

Banjul, The Gambia

February-March, 1987

AGRICULTURAL STATISTICS SHORT-TERM TECHNICAL ASSISTANCE

February 12 - March 28, 1987

Aaron C. Johnson, Jr. and John S. Rowe

BACKGROUND

Input to the Agricultural Statistics segment of the GARD project began with the review of The Gambia's agricultural data system in January - February, 1986 by Aaron C. Johnson, Jr. and John S. Rowe from the University of Wisconsin. As a result of the review, a long term program of cooperation between Johnson and Rowe and the Statisticians from PPMU was proposed. Rowe returned to The Gambia in April, 1986 to assist with the questionnaire design and the training of field enumerators for the 1986-87 National Agricultural Sample Survey (NASS). A third trip in October, 1986, allowed Rowe to review the NASS data collection procedures and advise on specific technical issues at the mid-point of the survey. These previous STTA are summarized in GARD Reports 2, 7, and 14.

PRIMARY TERMS OF REFERENCE

The primary objectives of this trip were to continue working with the statisticians at PPMU and to strengthen the NASS system. The Terms of Reference are listed in Appendix A. Upon our arrival in country, Chief of Party Elon Gilbert gave us several other items to review during our visit. In this report we comment on the results of our efforts in addressing the primary Terms of Reference first, and then comment on the other items studied. Finally, in the last section, we make recommendations on the training, commodity, and long term technical assistance needs of PPMU.

I. Review of NASS 86/87 Data and Publication of Results

By the time we arrived, the Farm Economics and Statistics Section of PPMU had produced a summary of the 1986 NASS data and had submitted a preliminary report to the Permanent Secretary. We reviewed the report in light of 1986-87 crop conditions and compared the estimates to the previous PPMU times series data. We found several major changes from previous estimates that merited more detailed analyses.

Despite the commendable effort by the clerical and data processing staff in preparing the data for summary, our review of the summary process revealed several procedural errors. These errors should have been discovered if a serious review of the process had been done by the professional staff.

After correcting the errors, we prepared several new Lotus worksheets to facilitate the estimating procedure. The data review and estimating procedures used are described in detail in Appendix B of this report. The actual review of the indications and the setting of the area, yield, and production estimates was done in conjunction with the four senior people in the Farm Economics and Statistics section. The estimates were put on Lotus worksheets that can readily reproduce tables of estimates for a publication. A narrative is now being written for publication and we were assured that the report would be available for the ARAB meetings in mid-April.

We recommend a number of specific steps for PPMU to follow that will help eliminate errors in the review and estimating process in the future:

1. Include questions on the NASS area measurement form to determine the final status of each sample field: harvested, crop failure, or enumerator failure. This distinction must be made so that valid zeros and missing data are handled properly in the summary system.

2. Review the listing procedures to make sure that enumerators have done a complete job of listing and have included only dabadas in the sample villages.

3. Maintain time-series charts at the division and national levels on:

   --Average number of plots per dabada, by crop
   --Average size of plots, by crop
   --Average crop area per dabada, by crop
   --Yield by crop

   Plotting the current year's estimates on the appropriate time-series charts is the best first check on the quality of the new indications. Any questionable deviations from recent years should be verified immediately, returning to the original questionnaires when necessary.

4. The data review and estimating procedure would be significantly enhanced if the <u>objective</u> NASS data were supplemented with <u>subjective</u> information on crop conditions from the enumerators, supervisors, extension workers, farmers, and other knowledgeable individuals. We recommend that these subjective inquiries be established and summarized monthly so a record is kept throughout the year. There are several current procedures that can be coordinated in the efforts:

--Early Warning/Crop Situation questionnaires collected biweekly on the NASS
--Monthly Aghyramet summaries
--Crop Situation Report ("Rep 6") of the DOA
--Monthly summary of field conditions by NASS supervisors

We recommend that these reports be combined with notes of the agricultural statistician's staff accumulated from personal contacts and travels in the field during the growing season. In this regard, what was a promising beginning of summarizing and publishing early warning data by NASS was abandoned in November, 1986. We strongly recommend that this report be re-established and published monthly.

II. Data Analysis Workshop

We conducted a two-week training course on Agricultural Data Analysis from February 23 to March 7, 1987. This training course was organized by PPMU in collaboration with the GARD Project. The purpose of the workshop was to provide mid-level economists, technicians, and statisticians from various government organizations an opportunity to discuss and solve mutual problems in data collection and analysis. The workshop was organized around the following major topics:

--Review of statistical theory
--Data collection methods
--Editing and error detection
--Data analysis techniques
--Publishing results

Sixteen participants attended the course representing the following government agencies:

Agency	Number of Participants
MOA--Extension	4
--Animal Health and Production	1
--Forestry	1
PPMU	4
MEPID	1
ADP II	1
Water Resources--Meteorology	1
--Fisheries	1
Central Statistics	1
Jahally Pacharr	1

Based on attendance, participation in discussions, comments to instructors, and written evaluations, the course was well-received and felt to be helpful by the participants. A graduation ceremony was held on the last class day during which participants received a

certificate of completion. A government official gave a brief address in which he stated the need for and importance of good statistics in development planning. The ceremony was covered by Radio Gambia, which gave a five-minute report about the course on the evening news.

III. Preparing for NASS 87/88

We met with PPMU staff and CILSS representatives to formulate a sample design for the 1987-88 NASS. The meeting was productive, with agreement on a two-stage cluster design. While many of the survey features are similar to previous years, there are three important statistical design changes: method of sample selection, estimation procedures, and computing variances of the estimates. Because these involve new concepts and require different ways of collecting and handling data, we spent considerable time working with PPMU on developing detailed procedures. One thing that emerged from that effort was a document that lays out the steps that must be followed. (See Appendix B.) The reader interested in the details can turn to the appendix. Here we make a few general comments about the sample design that we think all readers should be aware of.

For the NASS survey, The Gambia's existing geo-political areas are used to define strata boundaries. The Divisions form the first level of stratification and are then further subdivided into Districts. The number of Districts within a Division ranges from five to nine. Major advantages of this stratification are: the boundaries are well-defined, the sample is spread throughout the country, and it provides the basis for making agricultural estimates for each division.

For selecting the sample, each District is further stratified into small villages, medium villages, and large villages, where the expected number of dabadas in a village, (as reported by the Central Statistics Department), is used as the measure of size. This stratification is important because it guards against the possibility of structural differences between dabadas of the same size located in villages of different sizes. Possible structural differences might be that dabadas of the same size produce different quantities of a crop or a different crop mix, depending on the size of their village. Also, producers in larger villages might have easier access to credit or to Extension services, in which case they might be more efficient than similar producers in smaller villages. Without this stratification, the sample could be biased toward atypical producers.

Once the sample frame has been stratified as described above, two villages are randomly selected from each stratum in the District. Three sample dabadas are then selected from each of the sample villages. The data collected from these sample dabadas provide the basis for the area, yield, and production estimates for the major food and cash crops in The Gambia.

We think this new sample design represents a substantial improvement over previous years. It is solidly grounded in statistical theory and practice, and of great importance it permits PPMU to compute for the first time legitimate variances of their estimates. These variances are useful because they provide an indication of the statistical precision of the estimates.

Some modifications of the new survey design will require adjustments by PPMU. One major change from previous surveys is listing six villages in each District rather than three villages as in the past. This will require more time and money, but CILSS has foreseen this and has increased its budget allowance to cover the increased listing costs. PPMU must, however, get into the field earlier than usual to get the listing completed on schedule.

Other aspects of the 1987 survey design requiring attention are the implications for questionnaire design, enumerator training, and the training manual. These items were explicitly identified on our terms of reference, but there was not enough time to give them more than cursory treatment. In this respect, we feel we have not satisfactorily fulfilled our mission, but there is little we could do about it, other things simply took more time than expected.[1]

IV. GADS, Prices and Horticultural Marketing

GARD and the Gambia Agricultural Data System (GADS)

There are two aspect of GADS: software and data collection. Each is discussed below.

Software: We have heard innumerable (and often unmentionable) comments about the GADS summary system. This is the classic case of computer programs written by an individual for a specific purpose becoming a source of frustration when (a) the programs are used for slightly different purposes and (b) the person who wrote the programs is no longer available.

There is, nevertheless, a need by various researchers for software that will generate enterprise budgets, partial budgets, full-farm budgets, and the like. After discussing the GADS software with several people and reviewing the types of problems faced when using it, we do not think it advisable, even if it is possible, to try to salvage the GADS software. Rather, we think it best to explore the potential of taking a data base approach, specifically using Paradox.

---

1 After returning to Madison, we reviewed last year's NASS questionnaires and the enumerator's manual and offered some suggestions for revisions. These were sent to PPMU on April 9. In addition, we designed a Lotus worksheet, which works off three macros to summarize 1987-88 NASS data to the District-level for each crop. A disc containing the worksheet, two sample data sets, and a set of instructions was sent to PPMU on April 23.

This has several advantages. It eliminates the special-purpose, written-by-one-person program, the current problem. A data base allows more flexibility, has more editing power, and permits greater control over the entire operation. Important in this regard is that GARD has Paradox and many personnel are already using it.

However, the proposed switch would not be easy and could not be done quickly. A major concern is identifying who has the ability and time to work on it. The computer staff currently working at GARD has the ability, but they would need substantial startup time to understand the complexities of budget analysis. This, coupled with the many other and perhaps more important responsibilities, eliminates them as candidates for rewriting the system. All we can do for now is discuss the issue with a number of people on our return to Madison, especially expert budget analysts, to see what suggestions they might have.[2]

Data Collection: A meeting was held at the Cape to discuss coordination between DOA and GARD. Attending were Gilbert, Ceesay, Jones, Gai, Rowe, and Johnson. The group discussed the possible duplication of collecting GADS-type data by the two agencies. It was agreed that some duplication now exists but that it could be avoided by making some adjustments and by more closely coordinating activities. Further conversation with Director Kinteh confirmed PPMU's desire to make the optimum use of their resources by cooperating with DOA in the FSR/E cluster areas. A meeting will be held following Gilbert's return to The Gambia, at which time the details will be worked out.

Horticultural Marketing and Price Reporting

We attended a meeting at which new terms of reference for Lisa Daniels were established. In addition, we discussed the research plan at some length with John Caldwell and with G.O. Gaye. We will discuss these meetings with Campbell and Daniels upon our return to Madison.

The price reporting system established by MFP is continuing, but in a highly modified way. Because it is difficult to get country prices to Banjul in time for the radio broadcast on Monday, only Banjul-area prices are broadcast. While the broadcast is still of importance, its value is lessened by the restricted geographic scope of the reported prices.

Initially, prices were collected weekly for livestock, grains, and selected vegetables in 14 markets across the country. Price reporting stepped up to twice-a-week collection beginning in February,

---

2 Later conversations with Paul Jakus, the original programmer of GADS, and with Rick Klemme, a UW farm management specialist, confirmed the need to rewrite the system rather than try to patch the existing GAD programs. Similar systems have been programmed on Lotus 1-2-3. They also thought a data base system like PARADOX would be ideal for the needs of a farm budget program. We will forward more specific recommendations to GARD in the near future.

reporting stepped up to twice-a-week collection beginning in February, 1987. A major deficit of this program is the lack of any price analysis. A solid price data base seems to be available, with some series available for as long as three years in seven markets, yet no summary has been done, much less any research on market price relations. We were unable to determine precisely why the price data are languishing, and we have no specific recommendations regarding GARD involvement.

We were asked to conduct a seminar for current and potential users of price data. Time did not permit us to do that, but we urged the marketing specialist at PPMU to conduct such a seminar in the near future. We recommend that GARD in conjunction with the Director of PPMU provide encouragement and support for the seminar. We were told that various agencies request price data and we know that more use would be made of the data if others knew what was available.

SECONDARY TERMS OF REFERENCE

I. GARD and Computers

We spent considerable time discussing GARD's current and future computer needs. The draft of a detailed memo by Patty O'Neil served as the basis of this discussion. Since the memo will be revised and circulated to all interested parties following Patty's return to The Gambia, we present here only a brief summary of the memo, along with our recommendations.

Currently there are three computers at the Mile 7 office and two at the Cape office (including a portable that belongs to Madison). Two additional computers are on order. When all the equipment is received and installed, the computer situation for GARD will be

Cape	- 1 XT, word processing
	- 1 XT, administrative
	- 1 XT, research
Yundum	- 1 XT, research
Sapu	- 1 XT, research
TDY house	- 1 laptop for STTA and backup.

The current demand at Sapu is heavy and much of the computer work must be done at Mile 7. Because of uncertainty about the arrival of the computers on order, the current plan for Sapu is:

-- create a temporary facility in an available room
-- transfer an XT from Mile 7 to Sapu about July 1
-- this corresponds to the time that PCV (Brad Mills) will be trained and ready for posting at Sapu

Several problems are currently delaying any computer installation at Yundum:

-- no room suitable for locating computer
-- no personnel trained in computer use and maintenance
-- no computer available for Yundum until new one arrives
-- a computer should not be located there until a professional is assigned responsibility for machine and personnel are trained to use it.

In summary, once the needed renovations/construction are completed, responsibilities assigned, and new machines installed, GARD should have adequate computer facilities at all locations.

USAID regulations require that equipment be purchased from U.S. firms. There is little GARD can do about this restriction other than request waivers when absolutely necessary and be patient in the meantime. With this understanding, GARD personnel should strive to take a longer view when ordering equipment. The need to purchase on a bid basis and the time required for receiving and reviewing bids from the U.S. can take a long time. GARD staff and researchers must make contingency plans for those times when equipment does not arrive on schedule.

II. GARD and Computer Software

Considerable discussion centered on what software should be available to GARD-supported researchers. MSTAT4 is currently used for analyzing trial data and seems to be sufficient to meet most needs. SAS is also available, but is not at the moment being used by anyone. We were told that no one has yet learned how to use it, although some are trying.

To date, there has been little social science research, especially by agricultural economists, so there has been little demand for software, especially a regression package. MSTAT4 does have some regression options, but they are quite specialized to trial data analysis. This would not be an acceptable package for any serious social science analysis. SAS is, as mentioned, available and is a powerful program, but not all researchers know how to use it and many may not be willing to make the time commitment necessary to master it. GARD and PPMU computer personnel are experimenting with MINITAB, a program designed primarily for teaching, that, nevertheless, is an excellent statistical program if used properly. We recommend that GARD acquire the latest version of MINITAB. Johnson will contribute a copy of MINITAB to PPMU from his research budget.

III. Data Base Library

Our broad mandate here ranges from checking with Paul Jakus regarding a "clean" GADS data set to establishing a Gambian Agricultural Data Base.

We discussed the notion of a "data base" with many people who had differing ideas of what should be done. Yet there seemed to be a common thread: "It is good and necessary that all data on The Gambia be immediately entered into a computer for easy retrieval by the push of a button." This is, of course, an overstatement, but it does reflect the gist of the conversations we had with several people.

We attach considerable merit to the ideal of a centrally-located data library where policy makers, planners, and researchers can easily obtain the type of data they need in the form they want it. But this ideal is not achieved by simply entering all available data into a computer on a catch-as-catch-can basis. We have witnessed a number of ill-planned data bases that have floundered because of too many data too soon. We recommend a cautious approach with some outside expert assistance if a serious effort is to be made at computerizing The Gambia's agricultural data. Many major questions must be asked and problems resolved before a viable data library can be put together. The following terse questions and comments are suggestive of some of the issues involved:

1. Precisely why is a data library needed? Simply saying that "we want data" is insufficient.

2. Who will use the data? How, when, and in what form are the data likely to be needed? It should be possible to define the specific outputs that data users want.

3. Do all data relevant to The Gambia need to be included? What about Gambian data published by such organizations as FAO and the World Bank? What about researchers' data sets, such as field trial data?

4. Must the data be on computer disks? Is having statistical publications filed on a shelf of value?

5. If much of the data are to be stored on disks, additional points must be raised:

   a. Where should the data base reside? Who is to be responsible for its long-term maintenance? A librarian with computer expertise is needed. Without someone primarily responsible for adding to and maintaining the library, the quality of the holdings will decay faster than new acquisitions are added.

b. All data files must be completely documented in detail if they are to be of any value to subsequent users. We know from personal experience that otherwise useful data sets are discarded when their format or content cannot be verified because of inadequate documentation.

c. A committee should be created to screen data sets before wasting resources for entering data of such poor quality that they are utterly useless. An example. Rowe talked with Eckert regarding our interest in the "base line" data generated by MFP. The response was that due to editing, loss of documentation, and so on, the CSU people can no longer make any sense out of the raw data.

Many other issues could be raised but we think the point is made: if GARD is serious about a long-term, viable data library, it must be prepared to spend considerable time and effort in planning and organizing the system before seriously accumulating data.

If, on the other hand, GARD's interest is far more modest, perhaps accumulating PPMU data, rainfall data, prices, and the like, then we think existing computer personnel could put together a Gambian data set, using Paradox as the operating software. But we remain somewhat skeptical about how useful it will be and about how long it will remain current, and hence useful.

IV. Meeting of Socioeconomists

A meeting of socioeconomists was held at PPMU. Representatives from DOA, PPMU, Jahally Pachaar, MOA, plus a researcher completing field research and returning to Oxford University, attended the meeting.

The meeting began with each person discussing current work, problems and constraints, and preliminary plans for 1987. This revealed several areas of possible duplication of efforts. These included the fertilizer-use surveys of PPMU and ADP II; GADS data collected by PPMU, DOA, and Jahally Pachaar; and some socioeconomic survey work by PPMU and DOA.

Five areas of concern to socioeconomists were discussed by the group.

1. All agreed that duplication of effort was wasting resources and that cooperation and coordination are badly needed. How this is to be achieved was not clear. Trying to hold a quarterly meeting of socioeconomists on a regular basis was suggested as one way to begin working toward greater coordination. PPMU will be invited to the ARAB meeting to summarize the 1986 crop estimates and to review its GADS program. This is a start toward establishing a broader base

of cooperation and coordination among all researchers in The Gambia.

2. A problem faced by several groups is the apparent lack of appreciation by MOA of the logistics needs of socio-economists. This is especially troublesome for PPMU. The possible termination of the CILSS and FAO projects in the next year will eliminate funds currently used for data collection. If offsetting funds are not provided, the work of PPMU will be severely curtailed. If only minimal funds are made available, resulting in low night allowances and little fuel, it will be impossible to sustain collection of good quality data.

3. All participants agreed that high priority must be placed on better training and supervision of enumerators. All too often data collection efforts fail due to poor enumerator perform- ance (we heard of data sets that could not be summarized, much less analyzed, because of low quality resulting from poor supervision and enumeration). Several ideas for improving enumeration were discussed. One possibility is for PPMU and Jahaly Pachaar to cooperate in training since they collect similar types of data.

   Another proposal to conduct a nationwide data collection training session during the coming year. The idea would be to give enumerators from all agencies and projects rigorous, general training in enumeration methods and techniques. This would be followed by additional training by agencies and projects that would focus on their specific requirements.

   Apparently the Central Statistics Department (CSD) is developing a three-month enumerator training program. Discussions should be held with CSD to see if agricultural enumerators could attend this training school.

4. The need to share common data collection methods among different organizations was discussed. Currently different organizations use different methods for measuring such things as area planted to crops, crop yields, and the like. Different methods produce different results, and this leads to considerable confusion among data users. It is mandatory that responsible agencies work together to establish common measurement procedures. This will eliminate much of the confusion that currently results in a substantial waste of time and resources, and it would give more credibility to the data collected.

5. The group recognized the need for a data library. This has been presented in the previous section (III.) in a broader context.

The meeting closed on the positive note that it had been worthwhile and that such meetings should be held on a regular basis. We strongly support this proposition and recommend that GARD do all it can to promote a regular meeting of all socioeconomists to facilitate further cooperation.

RECOMMENDATIONS FOR ADDITIONAL GARD SUPPORT

Training Needs

In his October, 1986 report, Rowe made several recommendations for training at PPMU. Having received only indirect feedback on the GARD training committee's action, we will reiterate the training needs.

Yaya Jallow: We understand that Yaya Jallow has been approved for the five and a half month program (training periods 3 and 4) at the International Statistics Program Center (ISPC) in Washington, D.C., beginning in January 1988. By attending the second semester courses, he will build upon his previous training as well as limit his absence from the unit.

Momodou Jallow: Momadou Jallow is the senior supervisor in the Farm Economics and Statistics Section of PPMU. His primary responsibilities are supervising the NASS field operations and overseeing the in-office data summary procedures. He has been working in agricultural statistics since his completion of the agriculture program at Gambia College in 1979.

In his October report, Rowe proposed that Momadou enroll in the Introduction to Agriculture Statistics offered by the ISPC from August to December, 1987. The mix of courses is tailored to individuals like Momadou, who have a wealth of practical experience but are in need of further technical and theoretical training.

During the summer of 1986, when both Gai and Jallow were in the U.S. on training, Momadou had full responsibility for the 1986 NASS survey. His dedication in carrying out this task re-enforces our belief that he has the qualities necessary to be a valuable professional statistician in PPMU. But, as recommended by Rowe, Momadou must have the type of basic training offered at ISPC to achieve that status. Consequently, we again recommend as strongly as possible that Momadou Jallow attend the ISPC program from August to December, 1987.

Lamin Jabang: Rowe recommended that Lamin Jabang, a computer specialist at PPMU, be sent to the U.S. for degree training. Because of the important role he plays in the operation of PPMU, we reassert the recommendation that he be sent for training. Lamin has been diligently studying, using the services of a tutor, preparing for the O-level exams. He has registered and is scheduled to take the exam in

June of this year. On the assumption that he does well on the exam, he should begin degree training in January, 1988.

Computer-Related Needs

PPMU badly needs a printer in their computer facility. We recommend the purchase of a ProPrinter XL, as this will best serve their needs. Also during this trip it was agreed that PPMU was to receive two Uninterrupted Power Supplies (UPS), and a small generator to keep them running during power outages.

As mentioned earlier, we think both PPMU and GARD would benefit from the availability of MINITAB. We also recommend that GARD provide PPMU with copies of HAL and PARADOX. These software packages will provide better statistical analysis capability, more efficient use of Lotus 1-2-3, and data base capability for developing new NASS summary procedures.

We also recommend the following books to supplement their computer and statistics library:

Anderson and Sclove, The Statistical Analysis of Data, Second Edition, The Scientific Press, 540 University Ave., Palo Alto, CA 94301, 1986.

Anderson and Sclove, Minitab Guide to: The Statistical Analysis of Data, The Scientific Press, 540 University Ave., Palo Alto, CA 94301, 1986.

Kish, Leslie, Statistical Design For Research, John Wiley and Sons, New York, 1987.

Long-Term Technical Assistance

At our debriefing with Sompo-Ceesay, we were asked to make recommendations concerning the need for Long-Term Technical Assistance (LTTA) at PPMU. In this regard, we think our evaluation from a year ago is still valid (see GARD Report #2). At that time we outlined three options for technical assistance in agricultural statistics. They were:

1. Resident professional statistician
2. Resident research assistant
3. Short-term technical assistance

We also commented on the expected improvement of professional and technical skills given the three levels of input. Participant training and our STTA trips to date have improved the technical skills of the staff, but have had little impact on professionalism. We are still of the opinion that the only way PPMU will make substantial gains toward a more professional handling of their data is through Long-Term Technical Assistance.

APPENDIX A: TERMS OF REFERENCE

AGRICULTURAL STATISTICS SHORT-TERM TECHNICAL ASSISTANCE

February 16 - March 28, 1987

Aaron C. Johnson, Jr. and John S. Rowe

I. Assist PPMU statisticians with the review and publication of the 1986/87 National Agricultural Sample Survey (NASS).

II. Conduct a two-week statistics workshop for approximately 15 mid-level managers that have responsibility for the analysis of data in the Ministry of Agriculture and the Ministry of Water Resources. The curriculum will include:

   A. Sample design
   B. Data Collection Techniques
   C. Data Analysis Techniques
   D. Publishing and Report Writing

III. Assist PPMU statisticians in preparations for the 1987-88 NASS in the following areas:

   A Sample Design
   B. Questionnaire Design
   C. Enumerator Training Manual and Enumerator Training School

IV. Review the progress of:

   A. Price Reporting System
   B. Gambia Agricultural Data System (GADS)
   C. Horticultural Marketing Study

and assist in coordinating the future development of these studies.

APPENDIX B: THE NASS 87/88 SURVEY DESIGN[3]

GENERAL COMMENTS

The survey design for NASS 87/88 is slightly different from the design used for recent NASS surveys. The differences for NASS 87/88 are:

1. The three strata in a district are based on equal number of dabadas in each stratum, rather than on equal number of villages in each stratum.

2. A two-stage sample design with paired-selections of villages at the second stage is used. In the first stage, two villages (psu's) are selected by PPS from each stratum (rather than one village as in NASS 86/87). In the second stage, three dabadas are randomly selected from each of the two sample villages. This yields six dabadas in each of the three strata for a total of 18 dabadas in each district. This compares to 15 dabadas per district NASS 86/87.

3. Statistically legitimate variances for the estimates can be computed by the paired-selection method advocated by Kish.

4. Another modification for NASS 87/88 is the switch from the plot (or field) to the dabada as the unit for summarizing the survey data and preparing estimates of area, yield, and production at various levels of aggregation. This modification is, we believe, well-advised for reasons discussed below.

The following section of this report discusses sample frame construction. The next section outlines the procedure for selecting 18 dabadas from each District. The third section presents the estimation procedure, and the final section presents a suggested identification scheme.

---

3 This report was prepared by Cobe Johnson and John Rowe, University of Wisconsin-Madison, in consultation with Baboucar Gai and Yaya Jallow, PPMU Issa Ndiaye, CILSS, Ouagadougou; and Mbye Kebbeh, Statistical Consultant to CILSS. Statistical procedures are based on a consultancy report to PPMU by Kebbeh and Ndiaye, a meeting of the people named above in Banjul, March 1987, and by frequent references to Kish, Survey Sampling, John Wiley & Sons, 1965.

FRAME CONSTRUCTION

Sample frame construction begins with a listing of all villages and their 1983 Census population provided by the Central Statistics Department (CSD). This list was updated in March 1987 for NASS 87/88 using the latest CSD population data. The steps in constructing the frame are as follows:

1. Compute the expected number of dabadas in each village by dividing the village population by the average dabada size for the District. (District-level estimates of dabada size obtained from NASS 86/87 listing sheets.) Let Mos_{α} represent the expected number of dabadas in the αth village in the District.

2. Create three strata with an (approximately) equal number of dabadas, where:

   Stratum A: small villages
   Stratum B: medium villages
   Stratum C: large villages

3. To accomplish this, sort the villages by increasing size (number of dabadas) and list the accumulated sum of dabadas over all villages in the District. Let Mos_t represent the total number of dabadas in the District.

4. Create the strata boundaries as follows:

   a. The desired stratum size is obtained by dividing the total number of dabadas in the District by 3 (number of strata), rounding to an integer value if necessary. Let DSS represent the desired stratum size.

   b. In principle, the boundary between Stratum A and Stratum B occurs at the village having an accumulated total equal to DSS, and the boundary between Stratum B and Stratum C occurs at the village having an accumulated total equal to 2*DSS. This assures an equal number of dabadas in each stratum.

   c. In practice, the cumulated total for the villages will not be an integer multiple of the desired stratum size. The Stratum A/Stratum B boundary in this case falls after the first village with an accumulated total equal to or greater than the DSS. The Stratum B/Stratum C boundary occurs after the next village on the list with an accumulated total equal to or greater than 2*DSS. Because of the difference between desired stratum size and the actual sample size, the created strata will have unequal numbers of dabadas. Table 1 illustrates a sample frame.

Table 1

Creating Three Strata in a District
(District villages sorted by number of dabadas)

	Village	Number of Dabadas	Cumulate Number of Dabadas
Stratum A	1	8	8
	2	9	17
	3	9	26
	4	11	37
	5	12	49
Stratum B	6	15	64
	7	15	79
	8	18	97
Stratum C	9	20	117
	10	25	142 Mos_{α}

DSS = desired stratum size = 142/3 = 47.3 = 47
Boundary for Stratum A and Stratum B:
Between #5 and #6 -- Mos_5 of 49 greater than DSS of 47
Boundary for Stratum B and Stratum C:
Between #8 and #9 -- Mos_8 of 97 greater than 2*DSS of 94

Stratum	Desired Size	Actual Size
A	47	49
B	47	48
C	47	45

SAMPLE SELECTION

The NASS 87/88 survey uses a two-stage sample selection procedure. In the first stage two villages (primary sampling units) are selected in each stratum. At the second stage, three dabadas (the units of analysis) are selected from each of the two sample villages selected in the first stage. These stages are discussed in the following two sections.

First-Stage Selection

A village is viewed as a group of dabadas; technically, a village is a cluster of elements. As suggested by Table 1, different villages will have different numbers of dabadas. The sampling "statistics" for unequal-sized clusters are complex, involving complicated sample selection techniques and sophisticated estimation procedures.

For this reason, applied sampling statisticians have devised a number of practical methods that reduce the complexity of sampling unequal clusters yet yield acceptable estimators of population parameters. One widely used method, and the one used for NASS 87/88, is selecting the sample clusters (villages) with probability proportional to size (PPS).

Two villages are systematically selected from each stratum.[4] The sampling interval (SI) is computed and then a random number, called a random start (RS), is selected from the interval 1 to SI.[5] The first village selected is that village with a cumulated total, Mos_{α} equal to or greater than the random start. The selection number for the second village is obtained by adding the sampling interval to the random start and selecting the village with an Mos_{α} equal to or greater than this sum.

This selection procedure is illustrated in Table 2. The population total in Stratum A is 49, the Mos_{α} for village #5. Two of the five villages are selected in Stratum A, so the sampling interval for Stratum A is 49/2 = 24.5 = 24. Generate a random number for the interval 1 to 24. In the illustration, a random start of 14 is obtained. Consequently, the first sample village is #2, since its cumulated population (17) is the first one that is equal to or greater than the random start. The selection number for the second village is 14 + 24 = 38. This identifies village #4 as the second sample village in Stratum A. The same procedure is followed to obtain the sample villages in the other two strata.

---

4 In theory this requires that the villages be randomized within the stratum prior to selecting the two sample villages to avoid the selection bias than can occur when systematically selecting a sample from an "ordered" population. Because randomization usually requires more effort that can be justified and because there is no reason to believe that the list of villages involves any "ordering," the villages are selected systematically. This is the practical equivalent of selecting a single selection from a random half-stratum (Kish, p. 230).

5 This random start can be generated by a computer program (Lotus 1-2-3, for example), by a hand calculator with statistical functions, or obtained from a random number table or by drawing number slips from a hat. The important point is that the random start is obtained by a mechanical method and not in a haphazard, uncontrolled fashion that is open to individual discretion.

Table 2

Selecting Two Villages From Each Stratum

	Original List		Randomized List			
	Village	Number Dabadas	Village	Number Dabadas	Cumulated Dabadas	
Stratum A	1	8	2	9	9	
	2	9	5	12	21	village #1
	3	9	1	8	29	
	4	11	4	11	40	village #2
	5	12	3	9	49	
Stratum B	6	15	8	18	18	
	7	15	6	15	33	village #3
	8	18	7	15	48	village #4
Stratum C	9	20	9	20	20	village #5
	10	25	10	25	45	village #6

Stratum A: Sampling Interval = 49/2 = 24.5 = 24
Random Start = 14 ==========> village #1
14 + 24 = 38 ==========> village #2

Stratum B: Sampling Interval = 48/2 = 24
Random Start = 19 ==========> village #3
19 + 24 = 43 ==========> village #4

Stratum C: Only two villages, so both are selected.

As the illustration shows, there may be only two villages in a stratum. In this case, both villages are selected into the sample. It is possible, especially if there is considerable variation in village size. to end up with only one village in Stratum C. Several methods are available for dealing with "oversized" units. To keep the NASS 87/88 procedures easy to follow, yet maintain statistical integrity of the results, we will select this one village. The method for selecting the six dabadas from this one village is discussed in the following section.

Second-Stage Selection

In this stage, three dabadas are selected from each village selected in the first sampling stage. The first step at the second stage is preparing the listing sheet of all dabadas for each of the sample villages. The listing sheet contains certain basic data for the dabada, such as dabada head, various crops planted, whether the dabada owns livestock, and the like. The listing sheet for a sample village might look like the following:

		Crops Planted (Enter a "1" if Yes)				Livestock Owned	
Dabada	Head	Maize	Sorghum	Millet	Groundnuts	Cattle	Sheep
1	Jallow	1	1	-	-	-	1
2	Johnson	1	1	1	1	1	1
3	Gai	-	-	-	1	1	-
4	Kebbeh	1	1	-	1	1	1
5	Njai	1	-	1	1	1	-
6	Rowe	-	-	1	1	-	-
7	Jabang	-	-	1	1	-	1
8	Momodou	1	1	1	-	1	1
9	...						

Three dabadas will be selected from this list using systematic random sampling. If the sampling interval = (number of dabadas)/3 is a whole number (an integer value), standard systematic sampling will be used. For example, if the number of dabadas is 30, the sampling interval = 30/3 = 10. Select a random start on the interval 1 to 10. Select that dabada and every 10th one following until three have been selected.

If, on the other hand, the sampling interval is not at integer value (and this is the typical case), then the listing is treated in a circular fashion (see Kish, p.116) in the following way. Suppose there are 43 dabadas, so the computed sampling interval is 14.1. Use 14 as the sampling interval, but select a random start on the interval 1 to 43 (this is N, the population size). Select that dabada, say dabada #22, then take every 14th dabada until three are selected. Depending on the random start, selections following the first may come from the beginning of the list. For example, if there are 43 dabadas and a random start of 40 is obtained, select dabadas #40, #11, and #26. This procedure gives every dabada the same chance of being in the sample and retains all the desirable characteristics of systematic selection: easy to use, easy to check errors in sample selection, and spreads the sample through the listing sheet.

The case of only one village in a stratum mentioned above, which can occur because of unequal numbers of dabadas in the villages, poses a problem for obtaining paired selections (six dabadas) from the stratum. The problem of oversized units (see Kish, 7.5E) can be resolved by one of several procedures, each of which would cause a serious distortion of the overall NASS 87/88 survey design, especially the use of the paired-selection formula for computing variances. For NASS 87/88, the following procedure is used for a single village in a stratum:

1. Select the one village with certainty.

2. List all dabadas in the village.

3. Draw a line on the listing sheet to create two zones with the same numbers of dabadas in each zone. If there is an odd number of dabadas on the listing sheet, draw the line following dabada with number = (total dabadas/2) rounded up to next integer.

4. Treat these zones as villages.

5. Select three dabadas from each zone (pseudo-village) using the selection procedures outlined above.

6. Use these to represent paired selections from this stratum when doing the variance computations below.

MEASUREMENTS IN THE FIELD

For each sample village, all fields for the sample dabadas will be listed by type of crop. Measurement procedures are as follows:

1. AREA MEASURES: Measure all fields for each crop on this list.

2. YIELD MEASURES: Select a 1/5 random sample of fields for each crop and obtain yield measures. One randomly located 5X5 meter square will be cut on each of the sample fields. Sample fields are selected as follows:

number of fields on list	number of fields for crop cut
1 to 9	1
10 to 14	2
15 to 19	3
20 to 24	4
and so on	...

For 1 to 9: use 1-digit random number table, select one random number in interval in 1 to number of fields.

For 10 to 14: use 2-digit random number table, select two random numbers without replacement in interval from 1 to number of fields.

For 15 to 19: use 3-digit random number table, select three random numbers without replacement from interval from 1 to number of fields.

. . . and so on . . .

AREA ESTIMATES

The example below illustrates the procedure for estimating area at the District level. Yield and production estimates are discussed following this illustration, and Division and National estimates are discussed in the following section.

Example of Estimation Procedures

DISTRICT No._____ Mos_t = 128 dabadas

Stratum A: Mos_a = 39 Stratum B: Mos_b = 44 Stratum C: Mos_c = 45

Sequence of steps, or levels, for estimating area, yield, and production:

1. Village
2. Stratum
3. District

Steps 1 and 2 are illustrated using two sample villages in Stratum A. Assumed results will be used for the four sample villages in Strata B and C to illustrate estimating for step 3, the District level. Assume the sample data from Stratum A questionnaires are as follows:

Village #1			Village #2		
Db	Plot	Area	Db	Plot	Area
1	1	.68	1	1	.93
1	2	1.30	2	1	1.61
2	1	.75	2	2	1.11
3	1	.89	2	3	1.30
3	2	1.10	3	1	.89
Total Area		4.72			5.84

The total area for each village is computed first, but before proceeding an adjustment may be necessary. The sampling probabilities for selecting the villages were based on the expected number of dabadas in the village as computed from the Census data. However, it is possible that the listing sheet, which provides a complete count of dabadas in the village, may show a difference between the expected and the actual number. If this is the case, improper selection probabilities were used and an adjustment is needed. As an example, we have:

	Expected Dabadas	Actual Dabadas
Village #1	10	11
Village #2	8	6

Village #1 should have had a larger chance of being selected in order to be "statistically" representative, while village #2 should have had a smaller chance of being selected. To correct for this misrepresentation, the village total is adjusted by the ratio of the actual number to the expected number of dabadas:

Village #1: (11/10)*4.72 = 5.192 = adjusted total
Village #2: (6/8)*5.84 = 4.38 = adjusted total

These adjusted totals are used as the basis of estimation. (Notice that village #1's total is adjusted upward to compensate for the "under" probability of its selection and village #2's total is adjusted downward to compensate for the "over" probability of its selection.) Rather then first checking to see if there is a difference between the expected and actual number of dabadas, simply include the adjustment routine and run all village totals through it--if the expected number equals the actual number, the adjustment factor is "1" and nothing is changed. This will make programming or writing Lotus macros for computing estimates much easier.

For computing variances later, the area per dabada in each village is needed[6]:

Area Per Dabada

Village #1:	5.192/3 = 1.73067	(= village total area/3)
Village #2:	4.38/3 = 1.46000	(" " " ")

NOTE: the number of decimal places does not imply accuracy, it should be retained for subsequent computations. Once the final estimates are computed they should be rounded down to 100, i.e., 12,896.1519 to be presented as 12,900.

Stratum A Area Estimates:

Sample total in Stratum = Sum of total in villages = 5.192 + 4.38
= 9.572

mean = area/dabada = 9.572/6 = 1.595

Estimate of Stratum Mean = 1.595
Estimate of Stratum Total = (No. Dabadas in Stratum)*(Stratum Mean)
= 39*1.595 = 62.205

Estimates for Strata B and C within the District are prepared in the same way.

---

6 Remember for the special case of only one village in a stratum that the sample data from the two zones are to be treated as data from two villages.

District Estimates: Total Area and Area Per Dabada

To illustrate the area estimate for the District, assume we have:

	Area per Dabada		Stratum
	Village #1	Village #2	Mean
Stratum A:	1.73	1.46	1.595
Stratum B:	2.01	1.98	1.995
Stratum C:	2.96	2.84	2.900

	Expected No. Dabada	Mean	Estimate of Total Area
Stratum A:	39	1.595	62.205 = 39*1.595
Stratum B:	44	1.995	87.780 = 44*1.995
Stratum C:	45	2.900	130.500 = 45*2.900
	128		

Estimate of District Total = 280.485

Estimate Area per Dabada = 280.485/128 = 2.19 for the District

Estimating the Variance for the District Estimates

Since two villages are selected from each stratum within the District, the "paired-selection" formula from Kish provides an easy way to compute the variance for the District estimates. The basic formula is

$$\mathrm{var}(\bar{x}_D) = \frac{1}{2^2} \sum_{\kappa} (1 - F_{\kappa}) (\bar{x}_{\kappa 1} - \bar{x}_{\kappa 2})^2$$

where 2 is the number of clusters (villages) selected; κ = a,b,c for the three strata; 1,2 refer to villages 1 and 2; and f_{κ} is the sampling fraction in the κ th stratum.

In general, the sampling fraction = f_{κ} = $1/(Mos_{tk}/6)$. For the example, we have:

$f_a = 1/(39/6) = 1/6.5 = .15385$

$f_b = 1/(44/6) = 1/7.33 = .13633$

$f_c = 1/(45/6) = 1/7.5 = .13333$

Using these results, the variance of the estimate is obtained from:

$$\text{var(district mean)} = 1/4[(1-.15385)(1.73-1.46)^2 + (1-.13643)(2.01-1.98)^2 + (1-.13333)(2.96-2.84)^2]$$

$$= .01881$$

and

se(mean) = .13715
se(total) = .13715 * 128 = 17.55552

In summary the District area estimate is:

Estimate of total area = 280.485, se(total) = 17.55552, CV = 6.26%
Estimate area/dabada = 2.191, se(mean) = .13715, CV = 6.26%

Yield and Production Estimates

The simple average of the crop cuts on the sample fields will be used as the yield estimate. (see MEASUREMENTS IN THE FIELD above.) The production at the village level is computed as the area estimate times the yield estimate for the village. Using the example in Table 4:

For Village No. 1:

area * yield = production
5.192 * 1.2 = 6.23040
6.23040/3 = 2.07680 = production/dabada (for variance)

For Village No. 2:

area * yield = production
4.38 * 1.49 = 6.52620
6.52620/3 = 2.17540 = production/dabada (for variance)

where 1.2 and 1.49 are assumed yield measures.

Sample Total = 6.23040 + 6.52620 = 12.75640
Production/dabada = 12.75640/6 = 2.12607

Stratum A production estimate: (number of dabadas)*(production/dabada)
= (39) * (2.12607) = 82.9166

Repeat these computations to obtain production estimates for Strata B and C. The sum of these three estimates yields the production estimate for the District. The important point to note in all this is

that the same formulas are used for estimating area and production. This carries through to computing the variance of the area and production estimates.

DIVISION AND NATIONAL ESTIMATES

NASS uses a stratified design: the country is first stratified into Divisions and each Division is stratified into Districts. Consequently, the District estimates are expanded up to the Division level by using the standard stratified formulas (see Kish, p. 79 for a discussion of these formulas.

Estimate Division Mean = $\Sigma N_k/N$ * (District Mean_k)

where: N_k = number of dabadas in kth District k
N = number of dabadas in the Division

Estimate Division Total = N*(Estimate Division Mean)

Estimate Variance Mean = $\Sigma\ N_k^2/N^2$*(Variance Division Mean_k)

where Variance Division Mean_k is computed above by the paired selection procedure

se(Division Mean) = sqrt(Division Variance Mean)

se(Division Total) = N*se(Division Mean)

To obtain National estimates, use the same formulas, replacing "Division" with "Nation" and "District" with "Division."

The above is done for both area and production estimates. To get a yield estimate at any level, divide the production at that level by the area at that level. This gives a properly weighted average yield.

Finally, for any estimate at any level, the coefficient of variation, the CV, is obtained by se(estimate)/estimate. This expresses the plus/minus one standard deviation range as a percent of the estimate. We say that the true value falls in the interval of the estimate plus or minus the CV.

SUGGESTED CODING SCHEME

It will be very important for NASS 87/88 to keep track where data are located by village, by dabada, and so on. This will be particularly important for NASS 88/89 when half of the NASS 87/88 villages will be rotated out of the sample. Any number of schemes can be used, we suggest using sequential numbering. For example,

Sample Villages

Stratum A:	Village #1
	Village #2
Stratum B:	Village #3
	Village #4
Stratum C:	Village #5
	Village #6

In this way, villages can be accessed in any stratum with have to indicate a stratum identification. We also suggest numbering the sample dabadas in the stratum 1 to 18. Starting, say, with numbers 1, 2, 3 in Stratum A, Village #1 and ending with 16, 17, 18 in Stratum C, Village #2. Again individual dabadas can be located directly without need for a stratum and village designation.

APPENDIX C: NASS ESTIMATING PROCEDURES - 1987

Below are the steps to follow in using the indications summarized from the NASS to estimate agricultural items.

1. Raw Data Review: PPMU statisticians should review the raw data to be certain all sample dabadas have a complete complement of data. If a plot is missing a field measurement, the average size of the other dabadas in the district should be substituted. For yields, the statistician must distinguish between missing data and valid zeroes. Valid zeroes represent crop failure; the field was planted, but no harvest was attempted. Missing data occur when the field was harvested but the enumerator failed to obtain a yield sample. For missing yields, a "999" should be entered as a missing data code. Other data items (such as equipment inventory and livestock numbers) should also be reviewed and a district average entered for missing data for these items.

2. Data Expansions: After accounting for all the raw data, the direct expansion summary is run (EAEXPAND.BAS). Output (NATION.crp)[7] should be checked and transferred to worksheets (NATcrp.WK1) and then combined with all direct expansions on one worksheet (DIREXP.WK1). The resulting indications should be entered on the national and divisional time-series graphs on file at PPMU.

3. Additional Area Indications: To aid the estimating process, we recommend calculating two additional indications: averages per dabada, and averages per plot. These will assist in compensating for the problem of missing data by allowing the use of the estimating formula $Y = N\bar{y}$, where N is the estimated total number of dabadas and $\bar{y}$ is the average per dabada. As explained by Kish, this estimating procedure not only is more reliable, but is also easier to explain to the data user. Refer to item 6 below for notes on the relative precision of the indications.

Additional calculations necessary for both 1985-86 and 1986-87:

1) From the listing sheets for each district:

   a) Count the number of dabadas reporting the item of interest. Enter the numbers on PROPOR85.WK1 and PROPOR86.WK1.
   b) Calculate proportion of total dabadas reporting item on the worksheet.

---

[7]See listing of Lotus worksheets and explanation of naming conventions at the end of this appendix.

c) Sum districts to obtain divisional and national totals and proportions.
d) Calculate average number of plots per dabada for each crop for each district. Enter these data on Lotus worksheets (crpINDIC.WK1).

2) From the raw data files:

a) Calculate the average area per plot for each crop for each district. Enter the data on the crpINDIC.WK1.
b) Calculate the average area (or other desired number) per dabada for each item of interest. Store on crpDBAVE.WK1.

3) Develop and maintain a time-series indications for each item of interest:

a) Dabadas per district from the listing stage of sampling.
b) Total number of dabadas in the sample, number of dabadas reporting the item of interest, and proportion of dabadas reporting the item.
c) Average area per plot.
d) Averages per dabada (include all items of interest: population, area, number of cattle, and so on).

4) Maintain worksheets outlined in the attached list of computer files and Lotus worksheets.

NOTE: The files used for making agricultural estimates should be treated as "Internal Use" working documents. The contents would be difficult for a layman to understand and should be disclosed to researchers or statisticians only at the direction of the Head of the Farm Economics and Statistics Section. The procedures discussed below, however, are open for discussion and should be reviewed and improved whenever possible.

4. Division and National Level Estimating Procedures: Using the national and divisional Lotus 1-2-3 graphs and the estimating worksheets, estimate area, yield, and production for each crop.

1) Plot the three indicators of planted area for each crop. Identify the three indicators on the graphs with a "P" for average per plot, a "D" for average per dabada, and an "E" for direct expansion.

2) Compare the indicators and choose what appears to be reasonable, remembering that there is approximately a 10% error on either side of each division indicator plus an unknown amount of non-sampling error. When indicators are extremely divergent, the reviewer should search for errors in the raw data and in the summarization process.

3) Enter the divisional area estimates on the Lotus worksheets (crpEST.WK1). Using the F9 "recalculate" function, the worksheets will calculate national totals and calculate divisional data as percent of total and percent of previous year.

4) Using the ratio of harvested area to planted area from the DIREXP.WK1, make estimates of harvested area for each division. Enter the numbers on crpEST.WK1.

5) Plot the division yield indicators from DIREXP.WK1 on the yield graphs. Remembering that these indicators tend to have an upward bias (refer to precision comments below) and an amount of sampling and non-sampling error, select a tentative yield, usually at the nearest 50 or 100 kilograms. In selecting a yield, some consideration should be given to trend, but with an emphasis on growing conditions in the division. As always, widely divergent indicators should be verified with a recheck of the raw data and the summarization process.

6) Enter the tentative yield estimates on the Lotus worksheet (crpEST.WK1). Use the recalculate key to calculate the production and national yield. Review the percent of total and percent of previous columns and "fine-tune" the estimate. Adjust area and yields to nearest 50's if necessary, but <u>avoid</u> estimates that give a false sense of precision (refer to section 8 below).

7) Enter the Estimated number of dabadas (rounded to 50's or 100's) from crpINDIC.WK1 onto the lower half of crpEST.WK1. Also enter the percent of dabadas growing the crop as indicated on crpINDIC.wk1. The average area per dabada will be calculated by the worksheet. Review the numbers being careful to see that when there are a sizable number of dabadas reporting the crop, (say over 50), that there are estimates of area, yield, and production.

8) At this point, the entire worksheet should be reviewed. Harvested area and numbers of dabadas reporting will be rounded to the nearest 50 by the worksheet. The estimates are now ready for publication. Parts of the tables may be extracted to a word processing program for printing, or printed directly from the Lotus worksheet.

5. <u>District Level Estimating Procedures</u>: This review is done on the crpINDIC.WK1 worksheets.

1) Plot the three planted area indications on time-series graphs and make estimates. Try to smooth the estimates over time. Minimize abrupt changes unless there is evidence to

support them, such as abnormal climatic or growing conditions. Agricultural averages at an aggregated level (district, division, or national), such as average area planted to all crops, average number of people in dabadas, and average number of livestock owned per dabada, are relatively stable statistics that change rather slowly over time.

2) The district planted area estimates should be adjusted as necessary so they add to the previously established division totals.

3) The harvested area estimates should be set using time-series charts and the indication of area planted but not harvested from DIREXP.WK1. NOTE: At this time these procedures are rather weak. There is a need to develop better methods of producing reliable indications for the amount of a crop that is not harvested.

4) Review the yield indications and estimate a yield for each district. Again, heed the warnings not to follow widely divergent indications. When the yields are entered on the worksheet (crpINDIC.WK1), Lotus will recalculate the production for each district and the division total. The recalculated values will seldom agree with the original division estimates. However, do <u>not</u> change the division estimates. The districts should be adjusted to agree with the divisions. This process is sometimes difficult because it is necessary to adjust both harvested areas and yields, and then see if the result in the production agrees with the division estimate.

6. <u>Precision of Area Indications</u>:

<u>Average per Plot</u>: This indication is comprised of four elements: (Estimated number of dabadas) * (proportion of dabadas reporting the crop)*(plots per dabada) * (measured area per plot). It is probably the weakest of the three indications because of its reliance on the number of plots reported by the dabada at the time of listing. Since listing is done before planting, the numbers reflect the "intentions to plant", not actual plots planted. While the listing sheets offer a larger number of observations for a district than the sample of 15 dabadas, the quantitative information is a little suspect because the enumerator is usually in a hurry and little rapport is developed with the farmer. Most farmers under these conditions are reluctant to disclose information about their holdings. On the positive side, this indication is not sensitive to missing field measurements, or totally missing dabadas.

Average per Dabada: This indication is calculated as follows: (Estimated number of dabadas) * (proportion of dabadas reporting the crop) * (average area per dabada planted to the crop). If the estimated number of dabadas reporting the crop is reasonable, and the enumerator has measured all the field in the dabada, the indication should be sound. The indication is sensitive to fields not measured, so the raw data must be complete for those dabadas reporting. However, if some dabadas are not reported, the indication has the effect of substituting the average of the reporting dabadas.

Direct Expansion: Although this indication is in theory identical to the two mentioned above, it suffers most from missing data and errors in the listing process. Since it is calculated as: (Expansion Factor from the sampling process) * (accumulated total of area for all 15 sample dabadas in the district), a downward bias results if the enumerator is incomplete in either the listing process or the measurement process. There are no correction factors in either item, so the raw data must be complete, (or edited in), otherwise the indication will be less than it should be.

7. Precision of Yield Indications: Yield indications derived from small area crop cuts suffer from a number of biases. Unfortunately, most of the biases are usually in an upward direction. The enumerator's tendency to select non-random locations within the field usually results in a better yield Also, most enumerators tend to include too many plants in the crop cutting area whenever there is a question of whether the plant falls on the border. However, in situations where farmers do not have an idea of their yield per hectare, this process is the best available. If it is to be an useful indication, every effort must be made to train the enumerators to follow a totally random process in selecting the small crop cutting area and then to follow the instructions precisely. The procedure must also be well supervised because it is difficult work and many enumerators are tempted to take short cuts. The supervisor should be suspicious of data that show either too little, or too much variability. Researchers suggest that the C.V. for an enumerator's crop cutting samples should be in the range of 30-40%.

8. Rounding Published Data: Publishing estimates with several decimal places gives a false impression of high precision. We can assume that the expanded indications at the divisional level have a sampling error of about 10%. Thus, the indication of 2,642.93 hectares of maize for Western Division is really 2,642.93 plus or minus 528 hectares with 95% confidence, so the real value is anywhere between 2,015 and 3,171. We suggest adopting the practice of rounding all estimates to 100's, rounding to the nearest 50 only when necessary to make the estimates balance.

9. Publications: We recommend the formats shown in worksheets crpDIVPB.WK1 and crpDISPB.WK1. They allow the data user to view the number of dabadas growing the crop as well as the area, yield, and production of the crop. The district data should be published in a separate bulletin with a special explanation of the thinness of the data, possibly released only to bona fide agricultural technicians.

NOTE: Although the steps outlined above have included only the crop estimates, the same procedures and techniques should be applied for the other items queried in the NASS. The computer manager can adapt the Lotus worksheets to accept the equipment inventory and the livestock data.

10. Computer Files and Lotus 1-2-3 Worksheets:

The following computer files were used in the summary and review of the 1986-87 NASS:

name.PRN — Original raw data file for each crop. The "name" portion of the file title is the name of the crop.

NATION.crp — The expanded data file. Output from EAEXPAND.BAS. See below for an explanation of the "crp" term.

crpNAT.WK1 — Expanded raw data files. NATION.crp files read into Lotus.

DADABA.crp — Output from the average area per dabada calculation program.

crpDBAVE.WK1 — Contains the number of dabadas reporting the crop, plots, total area in the crop and the average per dabada. The 1985-86 averages are also included where available.

PROPOR85.WK1 and PROPOR86.WK1 — These worksheets contain listing sheet data on the number of dabadas reporting each crop and the proportion of the total dabadas is calculated. The proportions are used to estimate the number of dabadas growing the various crops.

DIREXP.WK1 — This worksheet contains the direct expansion indications for all crops at the district, division and national levels. Planted area, harvested area, yield, and production indications are included. It is too wide to print on one page.

crpINDIC.WK1 This worksheet contains the three indications used in estimating planted area. The indications are at the district level and are added to the division and national levels.

crpEST.WK1 This worksheet is used to enter the final estimates for fine tuning at the division level. The worksheet recalculates the national totals and percents of total and of the previous year. It also rounds production and number of dabadas to 50's.

crpDIVPB.WK1 This worksheet is a dummy table format suitable for publishing division level estimates. Estimates may be entered and the table extracted to a word processor or printed directly from Lotus.

crpDISPB.WK1 This is a dummy table format for the publication of the district level estimates.

The "crp" designation indicates the various crops, using the following codes:

maize - MZE	Early Millet - EML
Late Millet - LML	Sorghum - SRG
Upland Rice - UPL	Swamp Rice - SWP
Irrigated Rice - IRR	Groundnuts - GNT
Findo - FDO	