WSS TALK 10/27/93

History and Procedures of Objective Yield Surveys in the United States

Eric Waldhaus, Eddie Oaks, Mike Steiner, National Agricultural Statistics Service
Eric Waldhaus, NASS, USDA, Room 4162, South Building, Washington, D.C. 20250

KEY WORDS: Crop cutting, Enumerators, Objective Yield, Yield Surveys

This paper reviews the efforts made by the National Agricultural Statistics Service (NASS) in the United States Department of Agricultural (USDA) to measure crop production by direct measurement of plant characteristics. NASS implies the current agency and all of its predecessors. These efforts are collectively known as Objective Yield Surveys (OYS). This program is similar to Crop Cutting surveys done in many parts of the world. The major difference is that OYS includes non-destructive field counts prior to harvest to facilitate yield forecasts during the growing season. Yield is defined as the weight of targeted crop, at standard moisture, produced per unit of harvested area. Sampling for NASS Objective Yield Surveys have been on a probability basis from the inception. This is, however, not the implication of the word 'objective' in the title. Objective refers to the direct collection of plant characteristic measurements, instead of subjective estimates of yield reported by an observer.

NASS is a recognized world leader in the use of objective yield technology. Objective yield surveys produce the primary indications for yield forecasts and estimates for the major feed and food grains in the United States. Additionally, NASS has made long term commitments to make this technology available internationally. Through cooperative arrangements NASS has demonstrated or helped implement objective yield programs in many countries of Asia, Africa, and Central and South America.

Three aspects of NASS' objective yield program for major field crops are considered: The history and evolution of the program, current sampling procedures, and general concepts of objective yield survey field procedures. Specifically not included is a discussion of the use of survey data in preparing yield forecasts. This major topic is covered in another paper presented at this conference.

HISTORY

Yield and production of major field crops in the United States have been forecasted and estimated by USDA since President Abraham Lincoln's administration in the 1860's. Crop condition surveys were prepared monthly by the Statistics Division, USDA as early as 1863, the year following the creation of the Department. Until 1884 pre-harvest reports were in terms of condition as compared to an 'average' crop. In 1884 the reporting concept changed. Condition began being asked as a percent of a 'normal' crop, given no adverse effects of weather, disease or pests.

Although crop area changes from year to year, some of the largest variations in crop production are caused by fluctuations in production per unit area or yield. For more than a century, yield forecasts were based solely on voluntary producer appraisals of expected yield. It was recognized early that actual changes in yield were not fully reflected in subjective grower appraisals. By 1898 traveling agents supplemented farmer-crop reporters' information with on site observations of crop conditions. By 1903 more than 100,000 agriculture related business operators, including cotton ginners, millers, elevator operators, and transportation agents were paneled to gain insight into the agricultural situation.

In 1910 a shift began in the practice of reporting crop condition to forecasting actual production during the growing season. By 1915 cotton production forecasts became available during the growing season. The transition from condition to yield forecasts required regression modeling. This was almost entirely done by visual interpretation of charts prior to the use of computers in the late 1960's.

Objective measurements for forecasting yield started with cotton in 1928. These early efforts involved statisticians driving along the perimeter of cotton fields, making boll counts at predetermined locations in fields. There appears to have been no effort made to relate the field counts to yield. Thus, it may be more appropriate to think of this early effort as 'Objective Condition' surveys. Later corn and wheat were added to this program, but this early effort in objective methods was discontinued at the start of the World War II. Research into objective measurements of wheat, corn, and cotton resumed in 1954.

The 'birth' of probability sampling for agricultural statistics and objective yield methods came in 1957 when the United States Congress funded an initiative titled "A Program for the Development of the

Agricultural Estimating Service". The project provided for an annual enumeration of a large area frame probability sample for crop area estimates. This area frame survey evolved into the current June Agricultural Survey (JAS). Target crop fields identified during the JAS provide the sampling universe for the OYS (except winter wheat).

Cotton and corn objective yield programs became operational in 1961. Wheat came on line a year later. Soybeans joined the national program in 1967 and potatoes in the early 1970's. Grain sorghum, sunflowers and rice were added in the 1980's, but due to budget constraints grain sorghum and sunflowers were dropped in 1988. The rice program was reduced then and finally discontinued in 1993.

OBJECTIVE YIELD SURVEYS OVERVIEW

NASS is organized into 45 State Statistical Offices (SSO). There is one in each state except in New England where six states are combined. There is a centralized Headquarters in Washington, D.C. Sample design and selection, planning and coordination between states, centralized data processing, and quality assurance are the major roles of Headquarters in the OYS. Headquarters prepares and distributes two major OYS manuals. The Supervising and Editing manual is focused on the tasks completed in the SSO. The Interviewers' Manual is a training and reference manual for enumerators in the field. Each SSO coordinates field work and other data collection activities independently within established guidelines.

Qualified, adequately trained field personnel, including SSO staff and field enumerators, are essential for a quality job. States send a survey statistician, designated the State Survey Statistician, to a National training workshop to learn and reinforce correct procedures. State Survey Statisticians return to their states to train field supervisors and enumerators.

Objective Yield Surveys begins with intensive training for field enumerators. Training is more intensive for OYS than many other NASS data collection operations. The need for rigorous training stems from the fact that data collection usually is accomplished in remote locations in the field where supervision is minimal and there is not the opportunity to clarify procedures. It is also recognized that data collection is often a very time sensitive process so it may be impossible to reconstruct an 'interview' when errors are discovered after the field work is complete. The cost of training is very high, but the need is critical. NASS has consistently recognized this need and continues to make a substantial resource commitment to training.

NASS field enumerators are part time employees of the National Association of State Departments of Agriculture (NASDA). NASDA contracts their services to NASS. There are approximately 600 NASDA enumerators who work on OYS. OYS enumerators are almost exclusively rural people, and most are from farm families, typically retired or part-time farmers or farm spouses. Understanding agricultural practices is a prerequisite for a successful OYS enumerator. Enumerators also have to demonstrate literacy and computational skills about equivalent to a high school graduate.

In addition to training, the State Survey Statistician is charged with monitoring survey progress, and is the resource person for enumerators. Responsibly also extends to oversight of all SSO processing of survey data, and supervising laboratory processes. The State Survey Statistician and assistants review all edit and summary output. In most states the final yield recommendations (proposed estimates) submitted to Headquarters are not prepared by the Survey Statistician, but a Commodity Specialist.

Field Quality Control is conducted by supervisory enumerators and statisticians from the State office. A random sample of each enumerator's field work is selected for personal inspection. The sample selected for quality control is unknown to the enumerator and the supervisor in advance to insure an accurate assessment of quality. The sample pattern is such that at least one quality check for each enumerator is insured, and multiple checks throughout the survey cycle are possible. Supervisors may inspect additional work of the enumerators in their charge on an 'as needed' basis.

Occasionally, deficiencies in field procedures are discovered by the quality control process. When this occurs remedial action is taken, both to correct errors in a particular sample and to re-train the errant personnel. Discovery of deliberately falsified survey results is another potential benefit of the Quality program. The authors, with about twenty years of objective yield experience each, have no personal knowledge of this ever occurring.

Objective yield surveys are timed for making crop production estimates which are released to the public in the monthly *Crop Production* report. *Crop Production*

is published during the second week of the month, between the 8th and the 12th. To complete field work, process all data, and remain timely, the OYS adheres to a very rigid schedule. Data collection starts on the 22nd of the month prior to the survey reference date, and must be completed by the first of the reference month. Laboratory work, data processing, and summary review are completed, and recommendation submitted to the NASS Agricultural Statistics Board in Headquarters by the second day before the *Crop Production* release.

Concepts and methodology used in the OYS for forecasting and estimating yields are similar for all field crops. Two components of yield -- weight of the fruit and number of fruit -- are used to forecast a yield. Various plant characteristics are used to predict these components during the growing season. Harvest losses, estimated by gleaning small plots in the sample fields after harvest, are deducted to obtain a net yield.

During the early growing season, crop maturity varies considerably by region. As the season and plant maturity progresses the plant characteristics and measurements made to forecast yield change . The enumerator determines the maturity stage of the crop in the sample field during each visit and makes the appropriate counts and measurements for the growth stage.

Observations for each sample are made on two randomly selected plots (units) in each of the selected fields. Each plot consists of a specified number of parallel rows of predetermined length, or a rectangular unit drawn to specification if crop rows are indistinguishable.

SAMPLING

OYS samples are selected from acreage reported of the target crop in the March Agricultural Survey (MAS) or the June Agricultural Surveys (JAS). Spring and durum wheat, corn, cotton, potatoes, and soybean samples are selected from the JAS. The winter wheat sample comes from the MAS.

Winter wheat samples are unique as they are selected from the March Agricultural Survey using a multiple frame (combined list and area survey) design. Also, winter wheat varies in that samples are drawn from 'fields to be harvested for grain', while other crops are sampled from fields 'planted and to be planted' on the parent survey.

The objective yield sample for each crop is allocated to the most important production states such that 80 percent or more of the nations crop is included. Allocations are made to minimize production estimate coefficient of variation (CV). Until about 1990 allocations were made to maintain minimum harvest level CV's. As estimation models have improved, an effort has been made to allocate samples to maintain a minimum CV across the growing season.

The JAS, which is the parent survey for OYS, is the major, once a year, multiple frame survey conducted by NASS. Nationally, the area frame component includes approximately 15,500 segments, each about 1 mile square, representing about 52,500 farms which are enumerated in early June to identify land use. The area of target crop planted is expanded by the associated expansion factor for the area frame sample. OYS samples are then selected proportional to the expanded acreage. Proportional sampling insures that the distribution of the OYS sample will approximate the distribution of the crop as discovered in the JAS. Sampling procedures are similar for winter wheat except MAS is the base survey.

Survey States, sample size, and sample distribution are reviewed annual, but NASS has attempted to maintain consistent State involvement and sample sizes to maintain year to year comparability. In 1993 1,670 winter wheat samples were selected in 13 States. Spring wheat samples totaled 380 in four States, and 150 durum samples in one State were selected. Corn samples equaled 2,010 spread over 10 States, while 1,360 Cotton samples were drawn in six States. Soybeans samples totaled 1,330 in eight States, and 2,080 Potatoes samples were distributed over 11 States.

FIELD PROCEDURES

Enumerators are provided aerial photograph with the area frame segment containing the selected sample field outlined in red. Operators of land in these segments were interviewed during the JAS. Within the segment there may be more than one tract (farm). The enumerator locates and interviews the operator of the tract which contains the selected target crop field for OYS.

Six reporting forms are used through the growing season to collect information from the farm operator or to record counts and measurements. The reporting forms are identified by letter initials, which reflect the chronological order of use of the forms during the growing season. The data collected on each form are

similar for all crops in the OYS program.

A convenient way to describe the field procedure for implementing the OYS is to describe each reporting form, and explain its use.

Form A - is an interview form, used to update the crop acreage intended for harvest and to identify the sample field. It shows which field (area frame) or how to select a field (list frame) that will be used for making actual field counts and measurements. The Form A is completed on the first visit to the selected farm. It is also used to gain permission from the farmer to enter the field to set out OYS sample units, and to query the farmer about pesticide usage so the enumerator can take appropriate personal safety precautions.

Pesticide usage has expanded over the years both in the crops treated and the variety of chemicals available. Consequently pesticide safety training and enumerator exposure monitoring has become an integral part of the OYS program. This is especially true for Cotton OY, where the use of organophosphorus pesticides is nearly universal.

Form H - also an interview form, is used to collect data on seed, fertilizer, and pesticide application rates and tillage practices. These data are used for further economic analysis, and are not part of the yield estimation program directly. It is completed at the same time as the Form A.

Form B - is a field observation recording form. It is used to record counts and measurements of the plants and fruits. This form also reiterates instructions for locating, constructing, and processing the sample units.

The following two sections: Locating the Sample, and Counts and Measurements are presented here because these activities are associated with completion of Form B. A separate Form B is completed each survey month until harvest time, when a final Form B is completed.

LOCATING THE UNIT:

After completing Forms A and H, the units are constructed in the sample field by the enumerator. Two units are laid out for each sample. Unit 1 and Unit 2 are located independently of each other (except in wheat where unit locations are dependent). The random number of rows and paces for locating Units 1 and 2 are computer generated and preprinted on a label on the Form B.

The point of entry into the field, or starting corner, is the first corner reached when approaching the field that allows the units to have a chance of falling anywhere within the field boundaries. The shape of the field must be considered to insure that the entire field has a chance of selection. Research has indicated that there is no statistical differences related to starting corners. Therefore, any field corner which does not exclude some part of the field is acceptable.

The following steps are followed when locating and laying out units:

Step 1: The enumerator marks the staring corner with a piece of plastic flagging ribbon so it will be clearly visible on later visits.

Step 2: The enumerator then walks along the end of the crop rows the number of rows (or paces for wheat and broadcast seeded fields) indicated for Unit 1. A piece of flagging ribbon is tied onto the first plant in Row 1. This helps locate the same row on later visits. The next row in the direction of travel will be Row 2 of Unit 1.

NOTE: The enumerator walks his or her normal paces when locating the units within the field. It is not necessary to measure the distance traveled as it is not necessary to locate a precise point in the field, only one determined by a random process.

Step 3: The enumerator then walks the required number of paces into the field between Row 1 and Row 2, starting the first pace 1.5 feet outside the plowed end of Row 1. This makes it possible for a unit to fall anywhere in the field including the very edge.

Step 4: After the last of the required paces is taken, a dowel stick is laid down so that it touches the end of the enumerator's shoe. The dowel is placed across Row 1 and Row 2, at a right angle to the rows. The unit is laid out in the direction of travel of the last pace.

Step 5: The zero end of a 50 ft. tape is anchored at the dowel stick directly beside the plants in Row 1. The sample number is written on a florist stake and inserted at the anchor point.

Florist stakes are colored lath about 6 to 8 inches long. They are highly visible markers commonly used in nursery and greenhouse operations to mark seed beds. Florist stakes deteriorate quickly so no hazard will be created if lost or abandoned in the field after the

survey.

Step 6: In row 1 a starting florist stake is placed exactly 5 feet from the anchor point. It is marked "U1-R1". This measured 'buffer zone', helps insure that the unit location is not subjectively biased in its location by the enumerator. The florist stake should be placed beside the row about 2 inches from the base of the plants. The marker is placed outside the plant row to avoid any damage to the developing crop.

Step 7: Working outside the unit, the enumerator carefully measures the unit length and places a florist stake at the designated point. Corn, cotton and potatoes have larger unit lengths which are measured with a tape. For example, the corn count area is 15 feet long. A rigid metal frame is used for marking wheat and soybeans where the unit size is smaller. The wheat unit is 21.6 inches.

Not all fields are square or rectangle and other special situations may arise when locating and laying out a unit. The Interviewers' Manual gives details on how to handle most of these situations. Some of the problems that more commonly occur include: blank areas in the field that were known or unknown during the mid-year survey; the field is not large enough to accommodate the number of rows or paces specified; row direction changes; odd shaped fields are encountered as circular fields under pivot irrigation; fields planted on contours; or crop rows that are not distinguishable due to sowing practices. These situations are covered with precise instructions.

The Form B is the recording form for counts and measurements that are made at the units. Visits to these sample units will take place monthly during the growing season except for potatoes, when only one visit is made within 3 days of harvest or when vines are dead.

Because the same sample unit must be revisited monthly it is important the enumerator precisely mark the location of the unit. Plastic flagging ribbon is used. This is highly visible, but like the florist stakes, quickly disintegrates so it may be abandoned after the survey.

COUNTS AND MEASUREMENTS

Step 1: Measure 1-row space and then 4-row spaces. Measurements are made from the plants in row 1 to row 2 and then from row 1 to row 5. These measurements are used to calculate area of the unit.

Step 2: Count the number of plants in each row in the designated unit.

Step 3: Classify the unit by maturity category. Descriptive four page handouts with color picture examples are helpful in determining maturity.

Step 4: Make the specific counts and measurements of plant characteristics required. Different counts are made depending on the maturity level category. The crop and type of counts are as follows:

Soybeans: 1) plants; 2) nodes; 3) lateral branches with blooms, dried flowers, or pods; 4) blooms, dried flowers and pods; and 5) pods with beans.

Corn: 1) plants; 2) average length of kernel rows; 3) diameter of ear; 4) stalks with ears or silked ear shoots; 5) number of ears; 6) ears with kernel formation; and 7) cob length; and 8) field weight of corn.

Cotton: 1) plants; 2) burrs, open and partially opened bolls; 3) large unopened bolls; 4) small bolls and blooms; and 5) squares.

Wheat: 1) stalks; 2) heads in late boot; 3) emerged heads on all stalks; and 4) detached heads.

Potatoes: 1) hills; 2) tubers; and 3) field weight of tubers in the unit.

After completing Unit 1 counts and measurements go back to the beginning of the Row 1, and walk to the designated row, or number of paces, for Unit 2. Continue in the original direction of travel as when locating Unit 1 if Unit 2 count exceeds the Unit 1 count. After locating the Row 1 of Unit 2, walk the required paces into the field to set up Unit 2, and make the counts and measurements required.

A Form B is done for each month until very near harvest. Close contact is made with the operator so a sample field will not be harvested before a final Form B (just before harvest) can be completed. During this last visit before the farmer harvests, a sample of mature crop is sent to the laboratory. This sample is the basis for at harvest yield estimates.

FORM C-1 and C-2 - These forms record laboratory observations, and are not seen by the field enumerator. Form C-1 records data from pre-harvest field visits, while the C-2 is generated from the last field visit made at, or just before the farmer harvest.

FORM D - is used to record the actual number of acres harvested at the end of the year and the operator estimated yield of the field.

FORM E - is a field observation form used to collect data for determining field harvest loss so a net yield estimate can be made. The field visit to collect data must be within 3 days after harvest to determine harvest loss accurately as loose grain deteriorates quickly or is lost when left in the open. Harvest losses are subtracted from gross yield to arrive at a net yield. Finding the location of this post-harvest unit is similar to the original unit location. A measured rectangle is staked out and fruit from the crop is collected, and sent to the lab. There it is counted, weighed, and moisture tested to determine the field loss.

NON SAMPLING ERROR

Controlling non-sampling error is a major concern of the OYS program as in any large scale sampling survey project. Cause for OYS non-sampling error can be divided into two major categories: faulty procedures, and faulty procedure implementation. Additionally, as OYS use sub-samples from other surveys, non-sampling error present in the parent survey is passed on or magnified. This is out of the control of the OYS personnel except to monitor the larger survey for consistency. This source of error will not be considered further herein.

Non-sampling error which are the result of faulty procedures can be dealt with in a straight forward manner. The NASS research unit continuously reviews various aspects of the OYS program to insure survey validity. Validation surveys are conducted for each crop on a rotational basis. These surveys explore many aspects of OYS, such as the independence of the starting corner as noted earlier.

The survey quality program is also useful in discovering faulty procedures. Most often procedural difficulties that are discovered in the quality control program relate to some 'special case' which was not adequately considered when preparing manuals. Instruction changes that clarified selecting starting corners that do not exclude some part of the field developed largely through this route.

Insuring that procedures are consistently and accurately followed across the country is the greatest challenge in controlling non-sampling error. The most important control for non-sampling error is training. OYS training is continuous. The training cycle starts with training for State Survey Statisticians at National workshops. Usually there are three held in a year, one for Wheat, another for Corn, Cotton and Soybeans, and the third for Potatoes. Corn, Cotton, and Soybeans are combined for training because procedures, growing seasons, and States involved largely overlap.

Training continues with workshops for field enumerators, conducted by the State Survey Statistician. Assistance from the Headquarters OYS unit is available to the SSO's in conducting local training. This can be an important resource for a new State Survey Statistician, and gives Headquarter personnel the opportunity to observe local operations.

The formal quality control program in which individual enumerators have work inspected at random is an important part of the NASS non sampling error control program. While the potential is in place to discover an enumerator who is intentionally falsifying reports or 'table topping', this is not a major concern. The real value of the quality control program is to assess the level and effectiveness of training. Another important benefit of the program is its moral boosting effect on enumerators. The normal out come of quality control is that the enumerator is 'caught doing it right'. When fed back to the enumerator in a positive way this can be excellent reinforcement for continued quality field work.