METHODS OF COLLECTING AGRICULTURAL STATISTICS IN THE PEOPLE'S REPUBLIC OF CHINA

Ron Fecso, USDA

INTRODUCTION The author was a member of a technical delegation which visited the People's Republic of China (PRC) from November 8, 1982 to December 7, 1982, as part of the Agricultural Science and Technology Exchange Agreement with the PRC. The goals of this visit were to help improve U.S. understanding of the PRC's agricultural forecasting procedures and to demonstrate how statistical methods used in the U.S. can provide accurate agricultural statistics in lieu of a census.

The programs presented by the PRC during the visit included an overview of the historic development and policies of the PRC, government organization, planning policies, and the methodology used for censuses and sample surveys. Lectures and meetings were held in Beijing as well as in regional and local offices. Field trips were used to further explain the material presented. A synopsis of the PRC's organizational and economic presentations during this trip is included in (1). (Underscored numbers in parenthesis refer to literature cited in the References.) In return, seminars on the construction and use of area frames and on the agricultural forecasting methods used in the U.S. were presented by the U.S. delegation.

This paper describes the methodology and terminology of agricultural statistics used in the PRC and recommends areas for continued cooperation in the Exchange. For information concerning agricultural planning, coordination between agencies in the PRC, and an excellent reading list, the reader is referred to the report by Tuan and Crook which is based on an earlier visit to the PRC as part of the exchange program (2). Taun and Crook also present: (a) the organization of the agricultural statistical agencies; (b) the relationship of these agencies to the collection and use of the statistics; (c) the agricultural statistics reporting schedule; (d) the 1981 PRC statistical reporting tables (a translation); and (e) their assessment of the reliability of the agricultural statistics.

AGRICULTURAL STATISTICS The PRC did not inherit a strong statistical system when it came to power in 1949. With the aim to create an efficient statistical system, the State Statistical Bureau (SSB) was established in 1952. The statistical system grew through mid-decade, but was disrupted substantially during the frantic activity of the Great Leap Forward (1958-1960) when statistical reports were grossly exaggerated to match production goals. An attempt was made to restore objectivity to the system in the early 1960's.

The statistical system was again severely disrupted during the Cultural Revolution (1966-1976). Statistical organizations were abolished and workers were assigned to other occupations. In 1972, some government statistical institutions were reestablished, but by 1980 staffing at the national level was less than half that of 1956. It should be noted that considerable emphasis is now placed on keeping the statistical reports from being influenced by official production goals.

The PRC's main agricultural reporting system is based on the collection of census data. Each of the PRC's more than 5 million agricultural production teams (small collective farms) enter about 500 different categories of data in 13 quarterly and annual statistical reporting tables each year. For details see (2). Brief descriptions of the types of information and reports follow.

There are three categories of agricultural statistics:

1. Basic Information - construction of communal property and arable land, commune and team prices, and scientific information similar to that provided by our Agricultural Extension Service.

2. Agricultural Production - crops, livestock, and forestry.

3. Miscellaneous - mechanization, fertilizer, distribution of commodities produced, and income.

There are three types of reports:

1. Seasonal - planted area statistics, yield forecasts and estimates, and livestock inventory and use.

2. Midyear - livestock, forestry, cocoon, tea, fruit.

3. Annual - End of year production inventory and use of the commodities listed above as well as some economic data such as income and expenditures.

AGRICULTURAL FORECASTING AND ESTIMATION Three methods are used to collect data for regularly scheduled agricultural forecasts and estimates: censuses, typical area investigations, and sample surveys.

Censuses Two types of agricultural censuses are conducted: (1) the annual (sometimes referred to as "fixed time") census which is completed by the end of February and (2) a seasonal census which is done in each growing season. The SSB provides the leadership and reporting format for the collection of census data which they sometimes call "all around" or "overall" surveys. Forms are adjusted for local conditions and filled out under the combined leadership of the local Government agencies so that the various data requirements can be met. Census data are provided by the basic production teams and summarized upward through governmental levels to the State level. Although a census

requires a considerable amount of labor , the methodology is simple and administration is relatively easy due to the size and structure of the Government bureaucracy. The information is used for planning at all levels. The main shortcomings of censuses in the PRC are the amount of labor and time required, the possibilities for data handling errors, and as explained later, some subjective measures used. At this time, census is the only method used for animal and farm equipment information. The general feeling of the U.S. delegation members was that the PRC's officials are much more confident about their current census data than those collected in past years.

The brigade provides the basis for the seasonal census. In some sense the brigade is like our farm as a reporting unit in that the brigade keeps planting and yield records, as reported by the teams, for each field in the brigade. For a census during the growing season, the brigade has the following three subjective methods to "forecast" brigade yield.

1. A group formed from the statistical cadre and informed peasants drives around with the brigade statistician and subjectively evaluates the yield of the fields.

2. In some fields, grains are counted on some plants (not selected in a mathematically random manner). Historic records of these counts are compared to the current count and used with historic yields to subjectively derive yield. This method reportedly works better than the first.

3. A "typical" field is selected and the group carefully evaluates the yield.

Fertilizer use, weather, and their historical relationships are examined in the subjective development of these yield forecasts. For spring crops, the data is reported to the county no later than June 10th, for autumn crops, no later than October 20th. Besides crop forecasting, the census system is used to report livestock numbers and economic data. The commune reviews and adds the data from brigade reports and passes them onto the county with written comments. The county reviews the data (usually from about 650 brigades), summarizes the commune reports to county level, and passes this information to the next higher level of government. The manual summarization process is repeated up to the State level. As will be explained later in this paper, the census information from each of these levels provides crucial data for the sampling plans used in the PRC.

Typical Area Investigations This method, sometimes referred to as the "spot" or "key area method" has been adopted at the province, prefecture, and county level. The data must be reported twice each year, for the summer by the end of April and for autumn plantings by the 20th of September. At the county level, technicians from the county offices of the SSB, the Agriculture Bureau, and the Food Group form a team to forecast yields. Originally, they chose "typical" brigades or pieces of land on which they did their survey work. Now the team goes by truck and stops every five minutes to evaluate a field. These are the "typical" fields in a nonmathematically random sense. Visual examination, counts of stalks or kernels, and past harvest data are used to make the estimate of yield for each field.

The prefecture and province also supply technicians to go with the county teams to the "typical" fields to do forecast work. This grouping provides agreement on the data used in the summary of average yield for the county.

Sample Surveys Although some counties have continuously used the sampling system initiated in the 1950's, regional scale sampling has only been reinstituted in recent years. More emphasis is now being placed on sample forecasting methods, especially since the 30th session of the 11th National Congress. At this session the PRC instituted a new system of agricultural production responsibility which made the production unit smaller and thus may affect existing forecasting methods.

SAMPLE SURVEYS FOR CROPS Crop forecasts are one of the most important statistical series. Although the PRC makes widespread use of census methods to generate statistical information, sample surveys are not uncommon. Sampling methods were initiated in the early 1950's. In the early 1960's, groups were set up to investigate crop yield estimation using sampling. During the Cultural Revolution most of this sampling work stopped. Therefore, until recent years, most information was collected via census. Although agricultural censuses are still vital to the collection of planning information, sample surveys are being reinstituted and interest in improved sampling techniques continues to grow.

Survey Design Sample surveys are conducted during the two main crop seasons. Each crop has an individual sample survey rather than a single sample survey used for the collection of multiple data items. On the surface this might seem inefficient, but the total survey design and the use of local level production forecasts for crop management must be considered. Depending on the province, surveys might be done for paddy rice, wheat, corn, sorghum, millet, and/or sweet potatoes. The sample surveys are conducted in some of the agricultural provinces using a five stage sampling design which utilized prior census data to develop the sampling frame. As other provinces establish survey teams, sample surveys are expected to become more widespread until the goal of nationwide sample surveys is reached.

In selecting the sample for use in the first year, the province uses the prior three years of census data for the counties to draw the first stage sample. The three-year average (TYA) of yield per mu (a unit of area equal to

one-fifteenth of a hectare or 666.7 square meters) for each county is computed using end of season census data for the crop being surveyed. (On the average, provinces consist of about 80 counties.) Counties are arranged in ascending order by this average yield; then the cumulative county area of the crop is computed. A sampling interval is determined such that approximately 20 percent of the counties (with a minimum of 7) are selected using a systematic sample of the cumulative area values. The TYA yield per mu is computed for these selected counties and compared to the province TYA yield per mu. If the sample value is not within ± 2 percent of the province value, the sampling interval is reduced and a new and larger sample is drawn.

The second stage sampling unit is the commune. (On the average there are 23 communes per county.) The TYA yield per mu from census data reported by the communes in each selected county is computed. Communes are arranged in ascending order by the TYA yield, then the cumulative commune area of the crop is computed. A sampling interval is set such that about 30 percent of the communes in each county (with a minimum of 7) are selected using a systematic sample of the cumulative area values. Again, the TYA yield for the communes in the sample is compared with county TYA yield. The process is repeated with a larger sample size if the ± 2 percent tolerance is not met.

The third stage sampling unit is the production team. A production team has the responsibility for the agricultural activities of, on the average, about 1 or 2 percent of the commune's land. At this stage, the TYA yield per mu from census data reported by the production team in the selected communes is computed. Teams are arranged in ascending order of the TYA yield, then the cumulative team area of the crop is computed. A sampling interval is set such that about 5 percent of the production teams in each commune (with a minimum of 3) are selected using systematic sampling of the cumulative area values. As in the previous stages, the ± 2 percent tolerance between the sample and commune TYA yield is required.

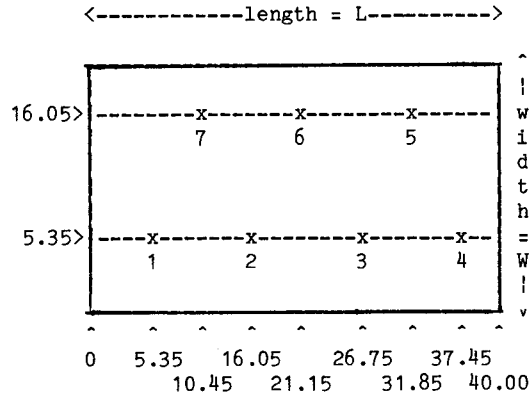
At the fourth stage of sampling, the sample units are the fields. Sampled units vary each year in areas where crop rotation is practiced. A group of technicians visits each field. Each production team has a list of fields by name, e.g. round or crescent, and a historic record of production for each field. The group uses the historic information and the observed condition of the growing crop to assign the field an expected yield. When this work is complete, the team arranges the fields in ascending order based on expected yield. In some areas of the country, a three-year average production rather than field visits is used to arrange the fields. Next, the cumulative production is computed. At this stage, using a predetermined sample size, a sampling interval is set to obtain a systematic sample of fields. In the North where fields are larger, at least 7 fields are chosen. In the South where fields are smaller, at least 15 fields are chosen. The number of fields chosen above the minimum depends upon the ± 2 percent criterion used at this stage.

At the fifth stage of sampling, sample plots are laid out in the selected fields. A systematic point sampling scheme is generally used to determine the centers of the plots. The points are laid out based on "the method of square roots" which is illustrated in Figure 1. The first step in defining the plots is to use tape measures to get an accurate measurement of the field's dimensions in order to compute the area. The number of sample points assigned to the field is based on the size of the field as follows:

Size of Field	Number of Points
under 5 mu	7
5 to under 10 mu	9
10 or more mu	11 to 20

To illustrate this method, assume that L = 40 (measurements in meters) and W = 20, so L x W = 800 square meters or 1.2 mu. Then the number of points based on the field size is m = 7. So L x W / 7 = 114.3, then d = the square root of 114.3 = 10.7, and the half distance is 10.7 / 2 = 5.35.

Figure 1 -- Illustration of the method of square roots.



Considering the field as a rectangular coordinate system, the points are laid out, as indicated in Figure 1 by the "x's" labeled 1 through 7, at the following places:

$$x_1 = (d/2,\ d/2) \qquad x_2 = (d/2 + d,\ d/2)$$

$$x_3 = (d/2 + 2d,\ d/2)$$

$$x_i = \begin{cases} (d/2 + (i-1)d - (j-1)L,\ d/2) & \text{if } j \text{ is odd} \\ (jL - d/2 - (i-1)d,\ d/2 + (j-1)d) & \text{if } j \text{ is even} \end{cases}$$

where j = integer portion of $((d/2 + (i-1)d)/L) + 1$.

Even in nonrectangular fields this procedure is used by modifying the x's to "zig zag" along parallel lines which go through the field at a distance of d apart.

In the special case of a narrow field, parallel lines cannot be formed within the field at the required distance, so an alternative procedure is used. In this instance, the field is divided into m (the desired number of sample points) equal rectangles with the center of each rectangle as the sample point.

At each of these sample points a mechanism is used to define the sample plot for crops not planted in rows or those with narrow rows. In some parts of the country, the plot is determined using a two-thirds meter by one meter rectangular frame. When crop rows are present, these frames are laid out at a 45 degree angle to the crop rows. In some areas an instrument similar to a protractor is used which measures a circular sample of about 1.1 square meters. In wide row crops, a fixed length of row (3.3 meters to 6.7 meters depending on the crop and area) is used. The systematic procedure (illustrated in Figure 1) is measured across the rows with the selected points determining the chosen rows. The sample length is centered at each point. A measurement is taken from the sample row across eleven rows to obtain the average space between rows. These methods were presented during field trips near the end of the visit. The lack of highly trained statisticians in the remote areas made it difficult to ascertain why this method was chosen or whether any bias problems have been investigated. Future exchanges might request information on these topics.

Data Collection Data collection for the sample plots consists of two types: forecast and harvest data. The collection of this data is done by survey teams in the local areas. These enumerators are of two types: full-time employees of a State Statistical Bureau's Sample Survey Team and part-time workers hired for the survey. The same plots are used for both the forecast and harvest estimate of production surveys.

Currently no prefruiting surveys are conducted. The only forecast survey is conducted after the grains develop. In some areas, early season visits are made to these plots to estimate the number of seedlings per unit area and the tillers. These statistics are used to improve and promote field management such as the allocation of fertilizer and irrigation to the areas in most need. Whether the treatments bias the estimates made later from the plots has not been determined.

During the forecast survey, a team of enumerators counts the number of ears or heads of grain within the sample plots. From a subsample (first ten ears in a plot), the number of grains are counted. Historic information about the average weight per grain by variety is used to forecast the yield, as will be shown in the following sections.

The peasant family which manages the sampled field is instructed to give the survey team one or two days notice before harvesting the field. The survey team harvests the sample plots before the field is harvested. The samples are threshed or dehusked, dried and weighed. After taking local harvest loss into consideration, the estimated yield for the field can be computed.

Estimation The estimation process conforms to the general structure of government, with the SSB cadre of each level of government computing the estimates and passing the information up to the next higher level in the SSB.

At the first level of estimation, the yield per mu is calculated for the field using the sample plot data. A jin, one-half Kg, is the commonly used weight measure. The field level calculation for the forecast survey is of the form:

jin per mu = (the number of sample plot areas in one mu of area)

x (the average number of heads or ears of grain in the sample plots)

x (the average number of grains per head from the sample)

x (adjustment for harvest lost)

x (historic relationship between variety and jin per grain).

For the harvest survey, the calculation is:

jin per mu = (the number of sample plot areas in one mu of area)

x (average threshed and dryed weight per plot)

x (adjustment for harvest loss).

The field level calculation is used to estimate production for the unit responsible for the planting and care of the land which includes the sampled field. Usually this unit is the peasant household, but in some cases, the production team is the smallest unit responsible for the care of the land. The area for which the unit is responsible is known from prior census data as is the area of the sampled field (values used in the fourth stage of sample selection). The measurement of field size taken during the fifth stage of sample selection is used to adjust for regional variation in the area defined as a "mu" and also as a control for reporting error in the fourth sampling stage.

The calculation here is:

responsible unit production=
(the sample estimate of jin per mu)

x (the area planted by the responsible unit to the crop being estimated)

x (the measured area in the sample field)

÷ (The sample field size which was used in the fourth stage selection)

The estimate of the crop's production for the production team is computed as a ratio which uses the responsible unit estimates made in the production team as follows:

the production team total=
(total estimated production for the responsible units having samples)

x (total mu in the team)

÷ (total mu, from the sample selection listings, in the sampled responsible units).

The commune and county estimates are a direct analogue to the ratio used for the team estimate. The total for the sampled teams is multiplied by the ratio of the total commune mu divided by the mu from the sampled teams. Similarly, the total for the sampled communes is multiplied by the ratio of the total county mu divided by the mu from the sampled communes.

The province total is estimated using a ratio similar to the previous estimates for the team, commune, and county. The State estimate of total production is the sum of the province estimates.

Variance Computations Variance computations are made for the estimated yield per mu. The average yield per mu in the province is computed as from a three stage sample design where county within the province is the first stage, commune within the county is the second stage, and the team within the commune is the third stage. The commune yield per mu is computed as the simple average of the yields per mu computed for each team sampled in the commune. The county yield per mu is computed as the simple average of the commune yields per mu, while the province yield per mu is computed as the simple average of these county yields per mu. The variance is computed by using the familiar estimate of the variance of the average from a three stage sample.

Quality Control Quality control is accomplished through reinspection. Whenever the forecast and the estimate or the estimate and the harvest data differ markedly, someone from the brigade goes to the field to reevaluate the yield and to determine what caused the difference. Although "markedly" was not strictly quantified during the visit, over 2 or 3 percent difference seemed to start quality control processes in much of the statistical work.

Updating the Sample At each stage a systematic sample is chosen from the list of units arranged by yield. As mentioned earlier, there is a criterion which the sample must meet in the initial year of sample selection. This same criterion applies each subsequent year to determine whether the sample should be updated. If the criterion is not met in subsequent years, the sample size is increased; when the difference is large, a new sample is drawn.

The criterion is used so extensively it deserves to be formally stated. Let,

y_i = the three year average or presurvey estimated yield of sample unit "i" in the population of size N, and

y_j = the same yield value for sample unit "j" in the selected sample of size n,

then the criterion ratio must satisfy

$$\left| \frac{\sum_{j=1}^{n} \frac{y_j}{n} - \sum_{i=1}^{N} \frac{y_i}{N}}{\frac{\sum_{i=1}^{N} y_i}{N}} \right| \leq .02 .$$

When the criterion ratio is much larger than .02 (how "much" was not clear), a new larger size sample is drawn. If there is only a slight difference, the sample is increased as follows. If the sample average is more than 2 percent below the population average (the average of the frame units at this stage), an additional sample is drawn systematically from the units in the population ordering which are above the population average yield. If the sample average was above the population average, the additional units are selected from population units below the average. The criterion is reapplied and additional samples drawn until the criterion is met.

RECOMMENDATIONS This section outlines areas for continued cooperation in the Agricultural Science and Technology Exchange. The increased understanding of the agricultural statistics methodology used in the PRC gained from this portion of the Exchange allows USDA statisticians to proceed with the goal of demonstrating how additional statistical methods can be utilized within the PRC's statistical system to improve the agricultural statistics. The topics presented here were raised as specific interests by the Chinese technicians or show promise for improvement based on statistical merit.

Training Possibly the most important long term contribution to improving the PRC's agricultural statistics would be assistance in statistical training. This training should not be geared toward the explanation of the U.S. methods, but should be academically oriented training in statistical analysis, sampling theory, and methods related to agricultural estimation.

Developing a Research Program A group of highly trained statisticians should be developed

with the specific purpose of conducting research in the area of agricultural forecasting and estimation. This group could supply high level technical consultants with experience in agricultural surveys to the various government agencies as well as develop a research program for survey improvements.

Survey Research Questions concerning multicrop estimation were raised by PRC technicians several times. The PRC does a specific sample survey for each crop. Information and/or research concerning the use of one survey to estimate several crops was desired.

PRC personnel also desired to exchange information on nonsampling errors. Items of particular interest include crop damage due to repeated visits to the same yield plot, harvest estimate bias associated with field treatment resulting from the early season forecasts, bias due to treatment effects, and nonsampling errors associated with the tremendous amount of hand summary.

A venture which should be developed is in the area of sample design. Although the systematic nature of the PRC's multistage sampling serves much of the purpose of stratification, there is a considerable amount of auxiliary data, now used only in sample selection, which almost surely could be utilized in the form of improved estimators. Other areas of survey research could include improving the variance estimators which are now simple random sample approximations, exploring the optimal allocation of samples to the various sampling stages, yield forecasting methodology, and multicrop survey designs. The use of minicomputers to improve estimation and reduce survey time and cost should be evaluated.

REFERENCES Fecso, Ron and George Hill, Agricultural Statistics in the People's Republic of China , Statistical Reporting Service, USDA, Washington, D.C., Staff Report No. AGES831208, December 1983.

Tuan, Francis C., and Frederick W. Crook. Planning and Statistical Systems in China's Agriculture , International Economics Division, Economic Research Service, USDA, Washington, D.C., FAS No. 181, April 1983.