DISCUSSION OF SATELLITE AGRICULTURAL SURVEYS

Emil H. Jebe, Environmental Research Institute of Michigan

It is indeed a pleasure for me to have this opportunity to discuss this series of interesting papers. First, I wish to thank my colleague and one-time office-mate, Prof. Jessen for suggesting me as a discussant. It has seemed appropriate for me to set this session in a historical perspective.

2. HISTORICAL COMMENTS

We are meeting here in August 1982. In the fall of 1938, Jessen was busy planning the first stratified random sampling survey for collecting agricultural information in Iowa. These plans were being made through the cooperative support of the USDA-SRS and the Stat Lab, ISU, under the leadership of W.F. Callander, Charles F. Sarle, Arnold J. King, and Prof. G.W. Snedecor. Four students were research assistants under that beginning cooperation, E.E. Houseman, P.G. Homeyer, Emil H. Jebe and Raymond J. Jessen. Out of that research came the Iowa and Arkansas "Farm Employment Surveys" of 1940-42, and the Master Sample of Agriculture.

Then came WWII. There were increased demands for sampling in many areas on a variety of topics. Jebe was an officer and entered the AUS. Houseman went to the SRS at the USDA. Jessen continued his work at ISU and his research promoted activity in many other places.

Now, what happened at the USDA. Research continued by dribbles on various aspects of the SRS needs. In the 1950's through some fortuitous circumstances, Houseman was able to expand activities in several areas, but it was not until in the mid 60's before there was a full implementation of the fruits of the research of 1938-1941 at ISU. That is an incubation period of about 25-30 years.

In 1964-65, the WRL of the University of Michigan had a contract within the IBM Federal Systems Division to survey and discuss possible civilian applications of the "Remote Sensing Techniques", which had come out of WWII and the research after the war. At the WRL, I just happened to be available and was a person with some agricultural background plus agricultural experiment station experience. Under I. Sattinger's direction, my assignment was to survey and describe the potential agricultural applications of the remote sensing devices. Although I had little knowledge of the physics involved, the agricultural applications part was relatively easy for me. I visited Purdue University (the late Dr. Shea) and the USDA-SRS (Glen Simpson was Chairman of the Crop Reporting Board at that time). My part of the WRL report considered principally the agricultural censuses and crop reporting aspects of the work of the USDA, Bureau of the Census and state agencies [1]. In 1965 I discussed these potential applications of satellite and aerial remote sensing at a conference at Cornell University on "The Role of Aerial Photography for Resource Studies in the Developing Countries" [2]. The speaker who preceded me on the program was Prof. Robert Colwell, U of C., Berkeley, who was already well known for his pioneering work in use of infrared films for aerial photography, disease detection in forests, etc.

It is interesting for me to reflect now on some of my wild ideas about how to implement the satellites viewing from on high into real time (almost) identification of crops in the fields and summation of those field areas by crop. I shall not describe them but will merely note two thoughts of that time. I did recognize some of the statistical problems present and I said to myself - give us 20 years and maybe we shall be doing something. By doing something I meant routine utilization, not research on how to to do it. The prediction, of course, was based on the period of 25-30 years I noted above.

Yet things have moved quite rapidly. In 1972, the first Landsat was launched and operated as expected. Digital tapes of what the sensors saw soon became available. How excited we all were to be able to identify rice fields quite well in northern CA (Fred Thomson & Frank Sadowski) [3].

Through my suggestions, Work, et al. made a successful application of regression analysis for estimation of wild life habitat in North Dakota [4]. Perhaps, I also made one of the early suggestions for correction of satellite interpretation bias by use of some form of ground truth and regression analysis. Dr. Feivison, Prof. Hartley, our own staff at ERIM, and others, followed up and expanded on this approach greatly. Thus, the problems which I recognized but to which I have contributed so little are before us today in the papers which we have heard. To repeat, it is now August 1982 - my 20 years prediction ends in 1985, just 3 years away, but it appears that we are already making some utilization of satellite data for agricultural estimates although this usage may not yet be routine.

In giving these historical comments, I should note that they are essentially personal with my own biases and are not intended to be complete in any sense - just a piece of "the story". Another relevant comment is that I am delighted to see the interest and participation of ASA members in this session and their contributions. By interesting a wider community and tapping their talents we shall make greater and more rapid progress on the problems still before us. Also, papers are appearing in the journals. I noted, with some interest in the May 1982 issue of Technometrics, the paper by Grassia and Sundberg on calibration of classifiers [5].

3.

It is also a privilege for me to participate in this session dedicated to Professor Hartley. My association with HOH began when he first came to the Stat. Lab. at Iowa State University. We became close personal friends of the family and we "helped look after" Grace and the children when Herman had to make a return trip to England. We visited Herman's brother and family in London in 1956. Herman helped me with a number of my problems and answered some of my questions, although I must say that I did not understand some of the answers. A special memory is Herman's first class at ISU on the application of computers to a variety of problems in numerical analysis and statistics. Dr. Gerald Barger, who contributed a great deal to the area we are considering today, was an active participant in that class. Dr. Barger's death earlier in the year before HOH died was a loss to this same field of endeavor. Another special memory of Herman was his assistance to me when I was writing the first program to compute the multiple regression residuals for our newly acquired IBM 650 at the Stat. Lab., ISU. There was a shift required to make my program work, but I had not detected the need for it.

4.

Now, it is time for me to give attention to some of the papers presented. I turn first to Sielken's interesting paper [6]. The title gave me a bit of problem. It appeared to me at first that Sielken means "Within Year Estimation," which can be applied year after year. Later in the paper, he brings in the use of multi-year data, but it was not clear to me just how it would be used.

The paper is somewhat mathematical; for publication, I believe an Appendix should be added to give more details on the mathematical results given in the body of the paper, which are not obvious by inspection. From the series of papers we have heard, the two broad problems of remote sensing are evident -- first, a classification has to be made, and then an estimation procedure must be applied to the classification results. Sielken takes the classification as given and then proceeds from there.

The model which Sielken has chosen seems reasonable although it is clear that the segment effect will change as the season advances so that an interaction term would be appropriate. Sielken's model just puts the interaction in the residual.

Sielken has considered three transformations of the classified proportion for a segment -- the identity, the logarithm, and the logit. One of my colleagues has remarked, "Why should a transformation be needed?", but Sielken has given some possible reasons. He did not make clear for me, however, why the identity estimate would not always be between 0 and 1. I agree that the observed p's in a stratum will have large variation, ranging from zero to near 1. Not even in Texas would the observed p = 1 for large segments. The variance components due to the segment effects and the error components would be reduced by 1/n for n segments in a stratum, and the delta (ℓ) term would be smoothed, too, so p should not get outside of the range 0 to 1.

For the log model, Sielken's example is illustrative, but I believe more evidence is needed for the multiplicative form of the model. Sielken does not note that the inverse estimate from the log model will be a median value for the stratum. For the log model, Sielken hints at the use of previous year's data but does not make explicit how the combination will produce an estimate for the current year.

The logit model is advanced for the multiplicative model with small p, but again I believe that further justification of this model choice is required. I do recognize that the logit approach has nice properties.

Sielken does bring out both the variance and bias estimation problems for his non-additive model cases. Further, he expects heterogeneous variation and therefore proposes a weighted least-squares estimation approach. How much this approach will complicate things in practice we do not know.

It occurred to me that perhaps an extension of Professor W. G. Cochran's work for situations involving binomial and extraneous variation might be useful for Sielken's study of the variance for his estimaters [7].

In closing on Sielken, I note that no references were included with my draft copy. The Abstract says that "This procedure and its empirical behavior are discussed" (underline by Jebe). He did not give us any such information on actual applications of the method. The time constraint for the paper may have required omission of results.

5.

Turning to the Smith & Ramey paper, we find a somewhat different approach, but still one that utilized the seasonal development of a crop [8]. It was of some interest to me that the study area used by Smith & Ramey covered some of the same areas studied by King & Jebe (1940) [9]. The utilization of the growing degree days makes the Smith & Ramey paper interesting.

In looking at the paper as a whole, however, I could not connect well the Introduction and the Section 8, Conclusions and Remarks. It is not made clear how points 1.1 (analyst time) and 1.2 (computer time requirements) have been obviated, eliminated or reduced by the Smith & Ramey method. The term "accurate" is used in 1.3 (USSR sample size requirements) and in point 2. Now, accuracy is a combination of bias and precision. Both of these components play a major role in remote sensing applications. The "current techniques" are not defined in point 2 of the Introduction. Paragraph 2 of the conclusions is not documented in the paper.

Figure 2 gives us a combined picture for segment 1924 in 1978 and 1979. We do not know whether #1924 gives one of the best pictures of the growing degrees relation or not. In their model, they use a cubic term, but we do not know how much the cubic term improves the fit for segment 1924 or for the whole set of ten segments. Figure 3, which shows the relation curving upward in the latter part of the season leaves one in some doubt about the cubic term.

Since the Y in the model is "percent red (green) pixels," a computer classification still has to be made. It appears that column 4 of Table 1 may be a difference derived from an analysis of two Landsat images for a segment (1979-1978). The column 5 would be an Estimated Change derived from another source, e.g.,USDA-SRS, but it could be the other way around for Columns 4 and 5. The text is not helpful on Table 1 so this part of the paper needs amplification. The "Error" column will become clear after we understand columns 4 and 5.

The summary statistics given below Table 1 in Table 2 do not address the real problem, which is an estimate for this North-South Dakota area of small grains acreage in the current year, i.e., 1979.

Another approach to the use of change detection for a revised stratification was advanced by Colwell (1979) in order to reduce overall cost [10]. In situations where the value of the information is proportional to the magnitude of the change detected, a first screening (visual or by computer) of the data might be used to detect change. Subsequent allocation of resources for the crop estimation could then be made in proportion to the indicated magnitude of change revealed by the screening. Thus, areas of apparent little or no change from a designated norm would receive no further sampling and considerable cost reductions might be achieved.

It is noted that the bias correction has not been addressed by Smith and Ramey. Achieving desirable precision by their approach may require a larger sample size. Further, their approach does not separate the "percent change" into the components of interest. Where is the change? Is it in wheat, oats, barley, flax or sunflowers?

The use of percent green pixels has been considered by a number of investigators. It may be more useful as an indicator of crop condition than for determination of crop area. Also, Δ (green pixels) may be highly correlated with Δ (yield/per unit area) in these ND-SD counties for 1978-1979. Since percent green pixels includes more than the crop of interest, there is a need for correction for background to remove the bias for the crops of interest. This correction may not be uniform from year to year.

As noted by Smith & Ramey, it is perfectly possible to process "wall to wall" (all pixels) in a county or a stratum such as a CRD. Such an approach will reduce sampling error to zero. But that still leaves us on the "horns of the great dilemma" of remote sensing. Moving in one direction leads us to apparent great precision but unknown bias that is not negligible. If the bias is to be reduced, it appears that some kind of auxiliary information is needed that must be collected either on the ground or much closer to the ground than what the satellite sees. Last year in a study of irrigated lands in Libya, we estimated biases of the complete Landsat coverage to be of the order of -40 and +30 percent for two different seasons within the same year. These biases were determined from a stratified random sampling ground survey for which the sample size was 115 segments that provided a sampling error (95 percent interval) of approximately 18 percent (Latham, et al., 1981 [11]).

It is necessary for me to make one comment on the JSC research program (last para. of paper). All sampling experience of the past 40 years has shown that smaller sampling units are more efficient because of higher precision and at little or no increase in cost. In remote sensing, however, we have an apparently new ball game. I think it will be some time before we learn how to optimize. When suitable procedures are not yet firm and routine, we cannot optimize although the directions to go may become clear. If we need to sample, larger s.u's are not the way to go. It was of interest to me that the Walker-Sigman study used quarter sections, which indicates that smaller sampling units than the current 5 x 6 mile units may be feasible [12].

In closing, I want to thank my colleagues, Dr. John Colwell and Brian Thelen for their assistance in reviewing these papers.

REFERENCES

1. Peaceful Uses of Earth-Observation Spacecraft, Vols. I and II, IR&OS Lab., Willow Run Labs., Inst. of Sci. and Tech., Univ. of Mich., 7219-1-P(I) & (II), February 1966.
2. E.H. Jebe, "Application of Satellites for Agricultural Surveys," (unpublished), presented at Cornell University Conference on "The Role of Aerial Photography for Resource Studies in the Developing Countries," 1965.
3. F.J. Thomson, et al., "Progress Report on Use of ERTS-1 Data," ERIM 193-300-16-P (for NASA), August 1973.
4. E. Work, et al., "Use of Landsat Data to Assess Waterfowl Habitat Quality," ERIM 120000-15-F (for NASA and USDI), Fish and Wildlife Service, Northern Prairie Res. Ctr., Contract No. 14-16-008-2018, January 1978.
5. A. Grassia and R. Sundberg, "Statistical Precision in the Calibration and Use of Sorting Machines and Other Classifiers," Technometrics 24, 117, 1982.
6. R.L. Sielken, Jr., "Multi-Year, Through-the-Season Crop Acreage Estimation Using Estimated Acreage in Sample Segements" (this ASA Session), 18 August 1982.
7. W.G. Cochran, "Analysis of Variance for Percentages Based on Unequal Numbers," JASA 38, 287, 1943.
8. J.H. Smith and D.B. Ramey, "A Crop Area Estimator Based on Changes in the Temporal Profile of a Vegetative Index" (this ASA Session), 18 August 1982.
9. A.J. King and E.H. Jebe, "An Experiment in the Pre-Harvest Sampling of Wheat Fields," IA Agri. Expt. Sta. Bull. 273, 1940.
10. J.E. Colwell, "Present Status and Future Directions in Crop Yield Forecasting," TM, ERIM (for NASA), NASA CR-ERIM 132400-6-R, July 1979.
11. J.S. Latham, et al., "Monitoring the Changing Extent of Irrigated Lands of the Gefara Plain, Libya," UN FAO, Contract No. TF/LIB/5-5 (LIB)/AGON, January 1982.
12. G. Walker and R. Sigman, "County Level Crop Area Estimates from Landsat: Evaluation of the Battese-Fuller Estimator," (First paper this session).