DISCUSSION

Raj S. Chhikara, Old Dominion University

Two of the three papers related to remote sensing, one by Professors Battese and Fuller and another by Amis, et al, deal with the problem of predicting county crop acreages using the regular USDA survey data in conjunction with satellite (Landsat) data. Currently in USDA, the regression estimator obtained by regressing the survey reported crop acreages onto those determined from Landsat data for sample segments in the area of interest is used for estimation of crop acreages. Each of these two papers discusses methods by which county crop acreages can be predicted more precisely with the utilization of satellite data in addition to the regular survey data. Professors Battese and Fuller have proposed a predictor which has the desirable property of minimum mean square error; whereas Amis, et al, have investigated empirically methods of improving classification of Landsat data.

First I discuss the paper by Battese and Fuller. It generalizes the classical regression estimator and gives a class of linear predictors assuming a nested-error regression model. This is achieved by weighting the classical estimator where weights depend upon the sample size, the county and error variance components. The minimum mean square error predictor is derived under the assumption of known variances. Consideration is given to the problem of bias and a modified predictor is suggested when the mean squared bias is desired not to exceed a threshold value. The method is applied to predict corn and soybean acreages for 12 counties in North-Central Iowa. The best predictor, standard regression predictor and two other predictors are computed. Numerical results show that the estimated mean square error is minimum for the best predictor.

This is an excellent paper showing how the present USDA crop acreage estimates can be improved upon at the county level. This is of course to meeting certain underlying assumptions. I have a few concerns regarding the assumptions made in the paper.

First, which is of minor importance, is the assumption of known variances. In general, variances are unknown and only their estimates can be available for use in prediction. So the optimal property of their predictor holds only conditionally.

My major concern is the assumption of known Landsat crop pixels in a segment. These crop pixels are estimated, sometimes with gross errors, and hence, the auxiliary variable is subject to measurement error. In the present context of classification of Landsat data with many limitations in adequately training a classifier, the measurement error is not necessarily uniform over an entire Landsat scene, and thus, it may introduce bias in the predictor. If this bias is considerable and dominates other errors, it should be investigated and, if necessary, a predictor which is at least approximately unbiased be constructed.

Lastly, the paper gives a predictor for the overall mean crop acreage for a collection of counties. This predictor is not necessarily the optimum one despite of its being the linear combination of the best predictors for individual counties. The investigators may consider constructing a predictor which has the minimum mean square error. The adjustment proposed in the paper seems artificial and does not necessarily improve the overall mean predictor except matching the large area crop acreage to the aggregated county crop acreages.

The paper by Amis, et al, focuses on the problem of classification of Landsat data to estimate the number of pixels for different crops and the extent to which the errors in classification affect the crop acreage estimation error. The classification system presently used--called EDITOR--may be iterated several times in selecting data for training of a classifier, if necessary, to achieve maximum value for the square of correlation coefficient, r^2, for the sample segments for which both survey reported crop acreages and Landsat data are available. The precision of a crop acreage estimate is based on the variance estimated by the residual mean square error times $(1-r^2)$. Since the value of r^2 for the sample segments which are used in training the classifier is expected to be higher than that for the segments which are not used in such training, a smaller value of r^2 is expected for the entire scene and hence, the gain in precision is likely to be overestimated.

The empirical study is conducted to evaluate overestimation of the gain in precision using 33 segments available from Missouri. Alternative clustering and classification techniques are also considered to seek improvement in the classification performance. The set of segments are treated in three different ways: (1) All 33 segments are used in training the classifier and in obtaining the regression equation, (2) 25 segments for training the classifier and 8 segments for an independent test set for the classification and regression, (3) jacknifing with 30 segments for training the classifier and 3 segments for the test repeated 11 times. Based on the test results it was concluded that:

(i) The classification error rates were higher for the test segments as compared to those for the segments used in training of the classifier.

(ii) The value of r^2 were smaller for the test segments as compared to those for the segments used in training of the classifier, implying that the gain in precision of a crop acreage estimate is overestimated.

(iii) Use of the alternative clustering method--called CLASSY--resulted in a smaller mean square error as well as in lower classification rates. The use of CLASSY was recommended to improve upon the present method of crop acreage estimation.

Most of the statistical analysis was based on comparative tests using the Hotelling's T^2 statistic. No specific statistical inference was made on the overestimation of precision and on the determination of bias in a crop acreage estimate resulting from the classification errors. Since each of these issues is equally important in the evaluation of the present approach to crop acreage estimation, different analyses of data should be constructed. Because of a limited data set these analyses may not be conclusive, yet the available data can be used to plan another empirical study.

It seems the present study is very limited and needs to be extended to a larger region with a larger set of test segments. It will be beneficial to include in any new empirical study the testing of the linear predictor proposed by Battese and Fuller to investigate how well it performs in the presence of classification errors.

The paper by Ron Fesco addresses the practical problem of stratification and area frame development for a large-scale crop survey using satellite data. He has presented a procedure for obtaining a land-use stratification using information derived mostly from the Landsat data. An automated area sampling frame, given in a digited form and which is flexible enough to permit changes in sampling unit size, is proposed and discussed. I did not get a chance to study his griding system in details. I hope that the stratification and area frame resulting from the proposed method is tested for its efficiency for a region in the U.S. before its implementation on a large-scale.