81-5

POTENTIAL UTILITY OF THEMATIC MAPPER DATA IN ESTIMATING CROP AREAS*

by

Richard Sigman and Michael Craig

Economics and Statistics Service
United States Department of Agriculture
Washington, D.C., USA

ABSTRACT

This paper predicts potential improvements offered by thematic mapper data over multispectral scanner data when utilized in regression estimation of crop areas. A study comparing LANDSAT data and simulated thematic mapper data is described. Quantitative measures of potential improvements in crop-area estimates of corn, soybeans, and dense woodlands are calculated, and the sensor characteristics causing these improvements are determined.

1981

1. BACKGROUND

Since the launch of LANDSAT I in 1972, the Economics and Statistics Service of the U.S. Department of Agriculture has investigated the utility of multispectral sensor (MSS) data in estimating crop areas. ESS's approach to utilizing MSS data has been to use it to supplement enumerator-collected ground observations available from ESS's operational surveys. These ground surveys consist of interviews with farm operators residing in randomly selected areas of land, referred to by ESS as segments.

The findings of ESS's MSS studies have been that such supplementary use of remote sensing data does yield statistical improvements in crop-area estimates but possess numerous difficulties for successful operational implementation. More specifically, ESS has found that MSS-based crop-area estimates are measurably more precise (i.e., have smaller variance) than estimates based only on ground data. However, operational implementation is hampered by MSS data problems of uneven quality, untimely delivery, and lost acquisitions because of clouds.

This paper describes an ESS investigation to predict to what degree, if any, that the supplementary use of thematic mapper (TM) data in crop-area estimation will provide additional statistical improvements compared to the supplementary use of MSS data. The research methodology employed was to analyze simulated TM (S-TM) data acquired by the aircraft-borne NS001 sensor.

2. CROP-AREA ESTIMATES FROM REMOTE-SENSING DATA

*Presented at the Fifteenth International Symposium on Remote Sensing of Environment, Ann Arbor, MI, May 1981.

Though requiring minor processing changes to accomodate TM's increased number of bands and higher resolution, ESS's basic approach for calculating MSS-based crop-area estimates can be applied without modification to TM data. Briefly, this approach consists of the following procedure sequence:

- Use segment ground-truth and corresponding pixels for supervised training of a pixel classifier,

- Classify segment pixels and develop regression relationship between segment classification results (independent variable) and segment ground truth (dependent variable),

- Classify all pixels in the area of interest, and finally

- Calculate crop-area estimates by applying the regression relationship to the all-pixel classification results.

The resulting MSS-based or TM-based crop-area estimates are evaluated in terms of their relative efficiency (RE), which is the improvement in precision (i.e., reduction in variance) compared to crop-area estimates based only on ground data. For example, an MSS/TM-based estimate having a relative efficiency of 3.0 means that the estimate is three times more precise (i.e., has a variance one third as large) as the corresponding estimate based only on ground data.

The ground-truth segments which are used in calculating MSS/TM-based crop-area estimates are selected by ESS's operational survey program using stratified random sampling. The utilized strata are areas of common general land use that have been photo interpreted from non-current aerial photography and LANDSAT image products. When an area of interest is completely contained within a single stratum, however, there is a very simple relationship between classifier performance and the relative efficiency of a corresponding MSS/TM-based crop-area estimate. This relationship is that the relative efficiency for a particular cover type is inversely proportional to the quantity $1 - r^2$, where r^2 is the coefficient of determination (correlation coefficient squared) between the ground-truth and classification results for the sample segments.

Thus, a measure for comparing two sources of remotely sensed data is

$$I(RE) = RE_2/RE_1,$$

where RE_1 and RE_2 are the relative efficiencies of source 1 and source 2, respectively. If I(RE) exceeds 1.0 then source 2 is an improvement over source 1.

If the area of interest is completely contained within a single stratum, then

$$I(RE) = (1 - r_1^2) / (1 - r_2^2).$$

As can be seen from Table I, when r_1^2 is large, even very small improvements in r^2 can produce major improvements in relative efficiency.

3. DATA SOURCES

The S-TM study area consisted of the two counties of Knox and Lewis, located in northern Missouri. Ground truth was collected by ESS's June Enumerative Survey (JES) during the period from late May to early June, 1979, in eleven segments randomly located within the two-county area. The eleven segments were each approximately one half square mile in size and all eleven segments were selected from the same ESS stratum--the stratum for intensively cultivated land.

At the time of the JES, some soybean fields had not yet been planted. Consequently, the JES recorded intentions to plant soybeans and these were verified by a follow-up survey conducted later in the 1979 growing season.

During the JES interviews, field boundaries were drawn on non-current photography, resulting in only approximate delineation of fields which had changed in shape after the date of the JES photography. Exact field boundaries within the eleven segments were determined from post-JES aerial infrared photography. This photography was flown during the first week of July, 1979, by a commercial flying service.

On September 9, 1979, the NS001 scanner [1] was flown at 20,000 feet over the entire two-county area in an aircraft operated by NASA's Johnson Space Center (JSC). Seven of the eight spectral bands of the NS001 scanner are the same as the TM bands. At the altitude flown, the ground resolution of the NS001 scanner was fifteen meters. In the post-flight processing, JSC degraded the ground resolution to thirty meters, the resolution of the TM. The NS001 data werre quantized into 256 levels (eight bits), the same quantization as the TM.

LANDSAT MSS data over the study area was obtained for September 7, 1979, two days previous to the NS001 overflight. As is well known, the MSS data characteristics are the following: four spectral bands, 60 meter resolution, and 64 quantization levels (six bits).

4. DATA ANALYSIS AND RESULTS

By calibrating the aerial photographs of the eleven JES segments to a cartographic coordinate system and then digitizing the segment field boundaries, cartographically-referenced ground truth was obtained. The remote sensing data was then also cartographically referenced. For the LANDSAT MSS data this was done in several steps. First an entire-image map-to-scene transformation was obtained by fitting a cubic, bivariate polynomial to ground control points identifiable on both a USGS quadrangle map and the 1:250,000 scale B&W LANDSAT image. The entire-image map-to-scene transformation was then used to scale and locate a plot of the segment field boundaries with respect to the coordinate system of a line-printer greyscale for a window of LANDSAT data containing the segment. The final step of LANDSAT data registration was to overlay the segment plot on the greyscale and by comparing the field boundaries on the plot with the lightness and darkness patterns in the greyscale determine local line and column shifts for each segment.

The registration of the NS001 data was done somewhat differently. Approximate segment locations in the NS001 data were determined by measuring the segment displacements in the JSC-supplied Visicorder strip charts. An affine map-to-scene transformation for each segment was then computed using ground control points selected for each segment from line-printer greyscales and cartographically-referenced aerial photographs. The rest of NS001 registration was then the same as that for LANDSAT. The registration procedures for both LANDSAT and NS001 were strictly mathematical in nature producing as end products map-to-scene transformations. No resampling of data to achieve geometric correction was performed.

Initial analysis was performed on the following types of remote sensing data:

-MSS data from LANDSAT,

- Simulated-TM (S-TM) data; i.e., the NS001 data as supplied by JSC, and

- Simulated-MSS (S-MSS) data created by degrading the NS001 data to MSS characteristics.

The S-MSS data set was created from NS001 data by using the four NS001 bands closest to the LANDSAT bands, by averaging 2x2 windows of NS001 pixels to achieve 60 meter resolution, and by integer dividing the NS001 data values by 4. For neither the S-TM nor the S-MSS data sets were degradations performed to model atmospheric effects.

For each type of remote sensing data the following steps were performed:

- By means of the cartographically-referenced ground truth and the developed map-to-scene transformation, extraction of ground-truth-labelled field-interior pixels for supervised training of a maximum-likelihood pixel classifier;

- Classification of all segments-interior pixels with the developed classifier; and finally

- Calculation of the correlation and coefficient of determination between segment-level JES-reported crop area and segment-level crop classification results.

Table II compares the coefficients of determination for MSS and S-MSS. The observed S-MSS coefficients are larger than the corresponding observed MSS values as would be expected because of fewer atmospheric effects on the aircraft-borne S-MSS. The observed differences between MSS and S-MSS correlations are not, however, statistically significant ($p > .05$). The observed differences in correlation were tested with Hotelling's (1940) test for comparing dependent correlation coefficients [2]. The use of Williams (1959) modification of Hotelling's test [3] yielded identical conclusions.

Table III compares the coefficents of determination for S-TM and S-MSS. The observed S-TM coefficients of determination are larger than the corresponding S-MSS coefficients of determination. For corn the observed difference in correlation statistically significant ($p = .003$), but for soybeans and dense woodlands the differences are not significant ($p > .05$).

Table IV uses the results in Tables II and III to compute a predicted-TM (P-TM) coefficient of determination to compare with the coefficient of determination for MSS. The P-TM correlation, r(P-TM), is calculated as follows:

$$r(P\text{-}TM) = r(S\text{-}TM) - [r(S\text{-}MSS) - r(MSS)].$$

Table IV also lists comparison relative efficiencies and P-TM confidence intervals. The latter are based on the Hotelling-test-statistic generated confidence interval for R(S-TM) - R(S-MSS) (R for population correlation) translated to the right by the amount r(MSS). The predicted relative efficiency improvement factors (I(RE)) of TM over MSS are 3.0, 1.4, and 1.3 for corn, soybeans, and dense woodlands, respectively.

In addition to the above-described two-treatment comparisons (MSS versus S-MSS, S-MSS versus S-TM, and P-TM versus MSS), called 2^1 designs in design-of-experiments terminology [4], an eight-treatment comparison (2^3 design) was also performed. These eight treatments resulted from analyzing eight different remote sensing data sets arising from the NS001 data through various combinations of MSS versus TM for the following characteristics:

- -Number of spectral bands (MSS = 4, TM = 7),
- -Spatial resolution (MSS = 60 m, TM = 30 m), and
- -Quantization (MSS = 64 levels, TM = 256 levels).

Correlation coefficients between crop reported area and crop classification results were calculated for each of the eight data sets (See Table V.), and the response equation

$$\text{correlation} = \text{average correlation} + A*D_A + B*D_B + C*D_C$$
$$+ AB*D_A*D_B + AC*D_A*D_C + BC*{}_B*D_C + ABC*D_A*D_B*D_C$$

was computed. The quantities D_A, D_B, D_C indicate the levels for the three factors of number of spectral bands (factor A), spatial resolution (factor B), and quantization (factor C) and assume the value -1 for the MSS level of the factor and +1 at the TM level for the factor. The coefficients A, B, and C indicate the main-effect contributions for factors A, B, and C, respectively, to the observed variability in correlation. Coefficients AB, AC, and BC indicate the two-way interaction effects, and coefficient ABC indicates the three-way interaction effect. The coefficients A, B, C, AB, AC, BC, and ABC are called half-effect coefficients because their value is one half the swing in the response variable between when in the response equation the coefficient multiplies -1 and when it multiplies +1. The sign and magnitude of a half-effect coefficient indicate the direction and amount of influence that the corresponding factor or combination of factors has upon the response variable--in this case, correlation.

Table VII presents the analysis of variance for the 2^3 design. Significant effects were determined by means of a forward-selection partial-F test [5] on the half-effect coefficients ranked by decreasing magnitude. For corn the significant effects were A (bands), B (resolution), and AB (bands x resolution). Figure 1 plots the corn response equation for these factors. Note that when $D_A = -1$ (MSS bands) increased resolution causes correlation to decrease from .77 to .76. For soybeans the only significant effect was B (resolution), which since the B half-effect coefficient is positive implies increasing resolution increases correlation for soybeans. For dense woodlands there were no effects which were significantly different from zero.

5. CONCLUSIONS

For the conditions of northern-Missouri location and early-September date, the predicted relative efficiency improvement factors of unitemporal TM over unitemporal MSS data are 3.0, 1.4, and 1.3 for corn, soybeans, and dense woodlands, respectively. The improvement in performance for corn is due primarily to a complex interaction of the effects of increased number of spectral bands and increased resolution for the TM compared to the MSS. The improvements in performance for soybeans is primarily due to the correlation-increasing effect of the increased resolution of the TM.

REFERENCES

[1] R. R. Richard, R. F. Merkel, G. R. Meeks. "NS001MS--LANDSAT-D Thematic Mapper Band Aircraft Scanner", Twelfth International Symposium on Remote Sensing of Environment, Manilla, Philippines, 1978, pp 719-727.

[2] H. Hotelling. "The Selection of Variates for Use in Prediction with Some Comments on the General Problem of Nuisance Parameters," Annals of Mathematical Statistics, 11, 1940, pp 271-283.

[3] E. J. Williams. Regression Analysis, Wiley, 1959, pp 72-80.

[4] C. Daniel. Applications of Statistics to Industrial Experiments. Wiley. 1976.

[5] N. R. Draper and H. Smith. Applied Regression Analysis. Wiley. 1966, pp 169-171.

Table I. Relative Efficiency Improvement Factor, I(RE), as a Function of Baseline, R_1^2, and Increase, $D(R^2)$, of Coefficient of Determination

	I(RE)		
R_1^2	$D(R^2)=.02$	$D(R^2)=.04$	$D(R^2)=.06$
0.5	1.0	1.1	1.1
0.7	1.1	1.1	1.4
0.8	1.1	1.3	1.7
0.9	1.6	1.7	2.5
0.95	1.7	5.0	na

Table II. Comparisons of Simulated-MSS and Actual-MSS

	sample coefficients of determination between segment reported area and classified pixels		
cover	S-MSS	MSS	p-value* for difference
corn	0.55	0.51	.83
soybeans	0.97	0.93	.24
dense woodlands	0.83	0.71	.10

*pvalue = probability of occurrence of the observed difference between sample coefficients of determination under the hypothesis that the population coefficients of determination are equal (two-tailed Hotelling test)

Table III. Comparisons of Simulated-MSS and Simulated-TM

cover	sample coefficients of determination between segment reported area and classified pixels		p-value* for difference
	S-MSS	S-TM	
corn	0.55	0.89	.003
soybeans	0.97	0.99	.22
dense woodlands	0.83	0.90	.10

*pvalue = probability of occurrence of the observed difference between sample coefficients of determination under the hypothesis that the population coefficients of determination are equal (two-tailed Hotelling test)

Table IV. Comparison of Actual-MSS and Predicted-TM

cover	coefficient of determination between segment reported area and classified pixels:			Relative Efficiency		
	MSS	P-TM	P-TM 75% C.I.	MSS	P-TM	P-TM 75% C.I.
corn	0.51	0.84	(.74, .95)	2.0	6.0	(3.8, 21.3)
soybeans	0.93	0.95	(.94, .97)	14.3	20.0	(16.7, 35.5)
dense woodlands	0.71	0.77	(.73, .81)	3.4	4.4	(3.7, 5.3)

Table V. Response Values for 2^3 Design

cover	Correlation Between Segment Reported Area and Segment Classified Pixels							
treatment*:	(1)	a	b	ab	c	ac	bc	abc
corn	.739	.787	.758	.900	.801	.804	.754	.942
soybeans	.986	.992	.996	.998	.984	.990	.996	.994
dense woodlands	.912	.926	.938	.936	.932	.942	.941	.946

*The presence of a letter denotes the +1 (TM) level of a factor (a for factor A = bands, b for factor B =resolution, c for factor C =quantization). The absence of a letter indicates the -1 (MSS) level. (1) indicates all levels at -1.

Table VI. Analysis of Variance of 2^3 Design

cover	effect symbol	effect source	half effect (decreasing magnitude	% explanation of response variation (indiv)	(cumltv)
corn	A	bands	.047	48.8% *	48.8%
	AB	bands x rsltn	.034	25.6% *	74.4%
	B	resolution	.028	17.3% *	91.7%
	C	quantization	.015	5.0%	96.7%
	ABC	bands x rsltn x qntztn	.011	2.7%	99.4%
	BC	rsltn x qntztn	-.005	0.5%	99.9%
	BC	bands x qntztn	-.0001	0.1%	100.0%
soybeans	B	resolution	.00399	74.7% *	74.7%
	AB	bands x rsltn	-.00145	9.9%	84.6%
	A	bands	.00135	8.5%	93.1%
	C	quantization	-.00101	4.8%	97.9%
	ABC	bands x rsltn x qntztn	-.00051	1.2%	99.1%
	AC	bands x qntztn	-.00044	0.9%	100.0%
	BC	rsltn x qntztn	-.00004	0.0%	100.0%
dense woodlands	C	quantization	.0062	36.8%	36.8%
	B	resolution	.0061	35.6%	72.4%
	A	bands	.0035	11.7%	84.1%
	AB	bands x rsltn	-.0027	7.0%	91.1%
	BC	rsltn x qntztn	-.0026	6.5%	97.6%
	ABC	bands x rsltn x qntztn	-.0015	2.2%	99.8%
	AC	bands x qntztn	-.0004	0.2%	100.0%

*Significant at .05 level via forward-selection partial-F test.

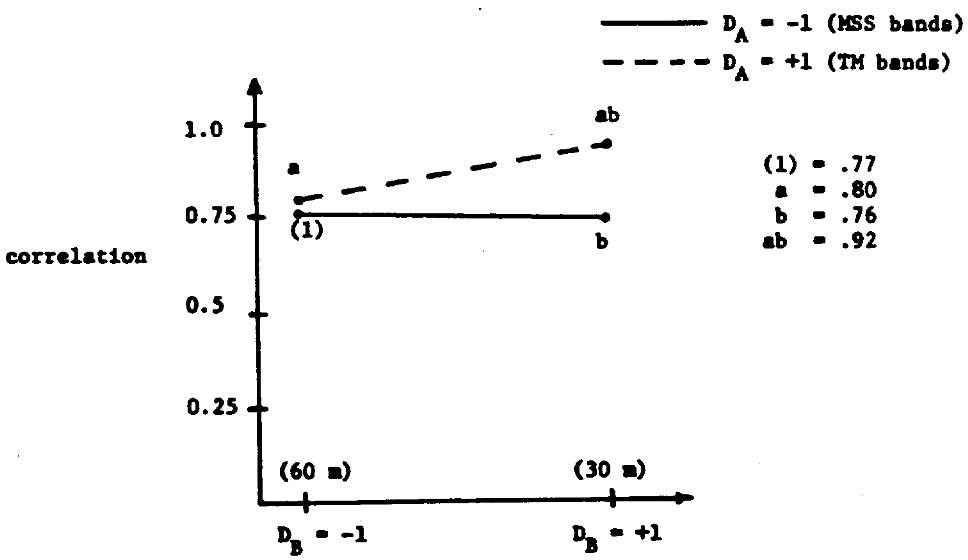

Figure 1. Average Effects of Spectral Bands and Spatial Resolution on Correlation between Corn Reported Area and Number of Pixels Classified as Corn