United States
Department of
Agriculture

Statistical
Reporting
Service

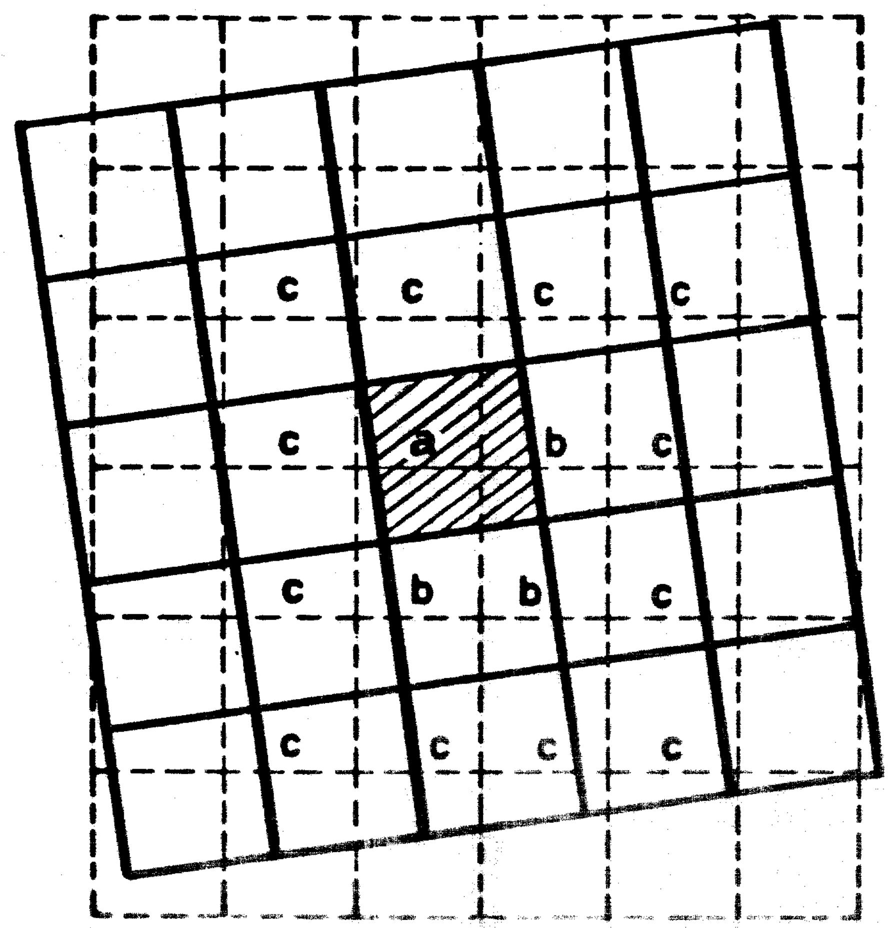
**Statistical
Research
Division**

Washington, D.C.

December 1983

The Effect of Cubic Convolution Resampling on Crop Area Estimates

J. Paul Zuttermeister
Richard Sigman



The Effect of Cubic Convolution Resampling on Crop Area Estimates

by

J. Paul Zuttermeister
Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C.

and

Richard Sigman
Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C.

December 1983

The Effect of Cubic Convolution Resampling on Crop Area Estimates. J. Paul Zuttermeister and Richard Sigman, Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C. 20250, SRS Staff Report No. AGES 831209, December 1983.

ABSTRACT

The purpose of this report is to investigate the effect of pixel resampling on the crop area regression estimates of USDA. Classifiers developed for non-resampled and resampled data are evaluated and show similar consistancy patterns. The authors recommend continued use of the resampled data for this specific application.

TABLE OF CONTENTS

	Page
Introduction	1
Summary	2
Landsat Analysis Procedures	2
Data Sets	2
Registration	2
Segment Shifting	2
Packing/Clustering	3
Estimation	3
Comparative Analysis	4
Conclusions	5
List of References	7
Regression Equations	A_1
Differences in Absolute Value of Residuals	A_2

THE EFFECT OF CUBIC CONVOLUTION RESAMPLING ON CROP AREA ESTIMATES

Introduction.

Digital data coming from the Landsat multispectral scanner are available in two different tape formats designated "A" and "P". The A data are processed for radiometric corrections only, while the P data are fully processed to include geometric corrections and resampling. The P pixel is generated by combining with various weights the values of sixteen A pixels. [6]* This process is called cubic convolution resampling and produces P pixels with higher spatial resolution. For example, a two by two pixel block of A data becomes approximately a two by three pixel block of P data.

This resampling procedure produces two potentially competing effects as far as crop estimation is concerned. First, the higher resolution P data may increase the proportion of pure pixels, that is those that represent a single ground cover. This may increase the precision of crop estimation.

On the other hand, the cubic convolution process by definition smears the signatures of neighboring A pixels to generate the P. This adulteration especially along crop field boundaries may make crop signature development and classification more difficult and may decrease precision of crop estimation.

Etheridge and Nelson (1980) found that cubic convolution produced brightness values outside the range of the original data but had no significant effect on maximum likelihood classification results. [1] Verdin (1983) on the other hand, found that cubic convolution produced anomalous values which had a significant effect in this water quality study. [9]

The question addressed in this paper is the following: Does the format of the Landsat data have any statistical effect on crop area estimation?

---

* Numbers in brackets refer to the numbered List of References at the end of the paper.

Summary.

This analysis showed no statistically significant differences in sample level regression estimates obtained from the two different data formats.

Landsat Analysis Procedures.

Data Sets.

Data were obtained in A format for five Landsat 3 scenes corresponding to P data analyzed earlier for the 1982 DCLC (Domestic Crop and Land Cover) project. [3] The three scenes in Iowa and the two in Missouri constituted one analysis district in each state.

The Iowa scenes contained 41 segments and the Missouri scenes 45 segments. A classifier for each of the two analysis districts was trained with the associated ground data--first for the P data and then with the A data.

Registration.

The current application of Landsat data employs the P format. Image products are available for this format and are used to register the satellite data to map coordinates. To register the A data, for which no such image products are avaliable, the control points used to register the P tape were transformed to A tape control points by the procedure described by Wolfe, et. al. (1980). [10] These transformed control points were then used to calculate a scene-to-map transformation for the A tape.

Segment Shifting.

Even though the registration process does a good job of locating satellite data in terms of ground coordinates, some local segment shifting is required to true up segment/field boundaries of the ground and satellite data. The magnitude of these shifts is generally one to three pixels in one or both of the coordinates. ASMA, the automatic segment shifting algorithm [2,5] does most of the shifting and is fairly reliable within its range of 5 pixels plus or minus. Table 1 shows the Root Mean Square (RMS) errors of registration in terms of segment shifts for the 5 scenes used. It compares the A and P formats for both row and column pixel shift RMS and direct meter distance RMS.

TABLE 1.

SCENE ID	SEGMENTS	P FORMAT(57x57m)			A FORMAT(79x57m)		
		PIX RMS		MTR RMS	PIX RMS		MTR RMS
		ROW	COL		ROW	COL	
I-AD29E	7	2.3	2.3	185	1.8	4.5	293
I-AD29F	26	2.3	2.3	185	2.7	5.3	370
I-AD29G	8	1.3	3.2	197	3.1	3.5	316
M-AD25I	29	1.5	2.8	181	3.7	8.9	585
M-AD25J	16	1.3	1.9	131	3.8	10.3	659

The larger RMS for A format seems to indicate a problem with the modified A registration procedure mentioned above. It also appears to be a scene dependent problem since no consistent pattern between the P and A RMS emerges. Segments shifted within the initial range of ASMA accounted for only 37 percent of the A segments compared to 100 percent for the P segments.

Packing/Clustering.

Packing is a method of selecting pixels to be used for signature development by CLASSY, an adaptive maximum likelihood clustering algorithm. All crops having at least 100 useable pixels were packed from the P format and those crops having at least 65 useable pixels were packed from the A format. This differential is roughly equivalent to the difference between the two pixel sizes. Useable pixels are those not falling on crop field boundaries and those not in crop fields predetermined to be inadequate for classifier training because of ambiguous boundaries or excessive waste.

The packed files were subjected to 7 iterations of CLASSY. Clusters identified with fewer than 1 percent of the total pixels for all crops and less than 10 percent of that particular crop were eliminated from the statistics files.

Estimation.

The standard Gaussian Maximum Likelihood pixel classifier produced the Segment pixel crop counts for the statistical analysis. The Missouri analysis produced sample level estimates of regression parameters for corn, soybeans, sorghum, rice, cotton, dense woodland, and winter wheat in land use strata 50 using 28 segments. In Iowa estimates for corn, soybeans, alfalfa, oats, wasteland, and permanent pasture were produced in strata 17 using 19 segments. The segments used for training but not included in the estimates fell among different land use strata with insufficient frequency to justify additional strata estimation.

Comparative Analysis.

The precision of Landsat-based techniques used by SRS for crop-area estimation is determined not by the accuracy of the Landsat classification but by the consistency of the classification results.

Consequently, the question posed earlier, "Does the format of the data have any statistical effect on crop-area estimation?": can be restated as follows: Does one of these two formats produce more consistent classification results when similar analysis procedures are applied to both? Sample level estimation generally includes more than one crop and regresses the jth crop pixels classified in the segment (3). The theoretical regression model describing a segment observation i, for crop j, is given by $Y_{ij} = My_j.x_{ij} + E_{ij}$ where $My_j.x_{ij}$ is the mean value of the reported crop acres distribution associated with a particular observation x_{ij} of classified crop pixels for the jth crop. The E_{ij}'s are differences (residuals) between the expected and the actual measured y_{ij}. "Reported crop acres" comes from the ground data collection carried out during the June Enumerative Survey and subsequent follow up contacts with land operators. "Classified crop pixels" comes from the Landsat sample classification. The regression equations obtained are presented in Appendix 1.

The units of the residuals are hectares. The residuals for the jth crop are assumed to come from a normal distribution with zero mean and variance σ_j^2. The more consistent a Landsat pixel classifier is for the jth crop, the smaller σ_j^2 is.

The P and A data formats produce different classifiers. One way to determine if a classifier is more consistent than another is to examine the residuals of the regression. Because of the underlying assumptions about the error term, the absolute size of the residuals is a function of σ^2. Hogg and Craig [4] show that if a random variable is from a normal distribution with mean zero and variance σ^2 then the mean of its absolute value is $\sigma\sqrt{2/\pi}$.

The single sample of segments, the two data formats, and the several crops suggest the application of the multivariate paired t-test to evaluate the hypotheses.

The null hypothesis is that within each crop, the variance of the residuals from the A regression and P regression are the same. This hypothesis is tested indirectly by looking at the means of the differences between absolute value of the residuals. (These data are presented in Appendix 2). That is, under the null hypothesis, the distribution of these differences has a mean of zero.

The hypothesis can be tested by means of the Hotelling T^2 criterion [8] where the observation units are segments and the dimensionality of the data is equal to the number of crops.

The test was carried out in equivalent General Linear Model framework using Proc Reg procedure of SAS [7] and gave the following F values.

STATE	CROPS	SEGMENTS	F-VALUE	D. F.	Prob F
Iowa	6	19	1.0449	6/13	0.4413
Missouri	7	28	2.0737	7/21	0.0927

With an alpha level at .05 the analysis shows no significant differences between the consistency of P and A format classifiers in either Iowa or Missouri.

Conclusions

From this study, it does not appear that the cubic convolution resampling transformation has any effect on the precision of crop estimation. Since the P tape requires one less processing step to register, SRS use of the P tape is currently preferred. If in the future, however, the A tape becomes less expensive or can be delivered more quickly than the P tape, and the A Registration problem can be solved then SRS use of the A tape is recommended.

LIST OF REFERENCES

1. Etheridge, J. and C. Nelson. "Some Effects of Nearest Neighbor, Bilinear Interpolation, and Cubic Convolution Resampling on Landsat Data." Proceedings Fifth International Symposium Machine Processing of Remotely Sensed Data. LARS, Purdue University, West Lafayette, Indiana, June 1979, Page 84. (Also See U.S. Geological Professional Paper number 1175. 1980. Page 306).

2. Graham, M.H. An Algorithm for Automating the Registration of USDA Segment Ground Data to Landsat Data. AgRISTARS Report DC-Y1-04211 (NSTL/ERL-201), December 1981.

3. Hanuschak, G., et al. Obtaining Timely Crop Area Estimates Using Ground-Gathered and LANDSAT Data. USDA Technical Bulletin No. 1609 U.S. Government Printing Office: 1980 0-310-945/ESCS-189

4. Hogg, R. V. and A. T. Craig. Introduction to Mathematical Statistics. Forth Edition, MacMillian Publishing Co. New York: Page 115.

5. Kalcic, M. T. Automatic Segment Matching Algorithm Theory, Test and Evaluation. AgRISTARS Report DC-YZ-04325 (NSTL/ERL-209), June 1982.

6. Manual on Characteristics of Landsat Computer-compatable Tapes Produced by the EROS Data Center DIGITAL IMAGE PROCESSING SYSTEM. Compiled by P. F. Holkenbrink, USGS 1978, Stock Number 024-001-03166-7 Page 6.

7. SAS Institute Inc. SAS User's Guide: Statistics, 1982 Edition. Cary, NC: SAS Institute Inc., 1982. Page 11.

8. Timm, N. H. Multivariate Analysis with Applications in Education and Psychology, Brooks/Cole Publishing Co. Monterey, CA: Page 149.

9. Verdin, J. Corrected vs. Uncorrected Landsat 4 MSS Data. Landsat Data Users Notes, Issue No. 27, NOAA, Sioux Falls, S.D. June 1983, Page 4.

10. Wolfe, R.H. Jr., et al., "Landsat Image Registration For Agricultural Applications." Proceedings Eighth International Symposium Machine Processing of Remotely Sensed Data. LARS, Purdue University, West Lafayette, Indiana, July 1982, Page 236.

APPENDIX 1

STATE: **IOWA** Strata 17 (19 Segments)

CROP	Regression Equation	R^2
A-Corn =	-12.252 + .486 * PIX	.704
A-Soybeans =	-.542 + .459 * PIX	.821
A-Alfalfa =	2.663 + .345 * PIX	.149
A-Oats =	-2.226 + .208 * PIX	.248
A-Waste =	11.233 + .757 * PIX	.536
A-Perm Past =	-11.192 + .425 * PIX	.599
P-Corn =	-25.253 + .381 * PIX	.691
P-Soybeans =	-2.280 + .344 * PIX	.807
P-Alfalfa =	10.776 + .081 * PIX	.011
P-Oats =	-3.380 + .642 * PIX	.152
P-Waste =	10.625 + .507 * PIX	.527
P-Perm Past =	-2.952 + .400 * PIX	.623

STATE: **MISSOURI** Strata 50 (28 Segments)

CROP	Regression Equation	R^2
A-Corn =	-.678 + 1.988 * PIX	.793
A-Soybeans =	21.697 + .467 * PIX	.394
A-Sorghum =	2.552 + .500 * PIX	.478
A-Woods =	.036 + .266 * PIX	.585
A-Rice =	2.637 + .636 * PIX	.764
A-Cotton =	.933 + .397 * PIX	.518
A-Winter Wheat =	-6.547 + 1.696 * PIX	.450
P-Corn =	-1.120 + 1.203 * PIX	.854
P-Soybeans =	2.942 + .293 * PIX	.381
P-Sorghum =	2.471 + .456 * PIX	.546
P-Woods =	-2.034 + .092 * PIX	.347
P-Rice =	-.638 + .481 * PIX	.824
P-Cotton =	-.321 + .267 * PIX	.589
P-Winter Wheat =	-5.935 + .723 * PIX	.463

APPENDIX 2

DIFFERENCES IN THE ABSOLUTE VALUE OF RESIDUALS*

IOWA

SEGMENT	CORN	SOYBEANS	ALFALFA	OATS	WASTE	PASTURE
226	6.84	2.72	-5.60	-2.78	.10	3.04
227	-2.40	2.66	-4.07	-2.85	-1.45	-2.83
230	20.45	-9.95	1.64	1.22	-1.42	3.63
231	8.21	-1.25	-5.53	-.89	-.61	-3.16
232	-8.42	4.50	4.76	1.11	.16	-.62
1233	-8.00	-7.09	-13.39	1.22	.34	1.73
1234	-4.11	1.17	5.40	-2.20	.61	7.50
1236	-3.61	5.36	-2.79	-6.29	.62	-.84
1237	6.50	2.32	-5.64	-.74	.16	-.09
1238	-10.03	-5.25	5.59	1.63	-.66	-10.89
7240	-12.60	2.38	-1.33	-2.66	.47	-10.32
7241	2.04	-5.07	-2.18	-5.80	-1.09	-5.29
7244	10.00	4.16	3.07	-9.40	-1.42	.03
7245	-9.77	-10.68	-7.14	-.72	.61	1.27
7246	-7.88	-3.81	-1.32	5.04	1.35	8.54
8219	-13.95	-.78	-.46	-1.96	-.41	-.37
8223	+.55	-5.98	3.96	1.20	1.11	.16
8224	-2.68	-.95	-5.44	-3.97	.61	.24
Sum	-33.28	-19.02	-24.69	-27.73	0.00	-4.95
Average	-1.75	-1.00	-1.30	-1.46	0.00	-.26

*Absolute value of the residuals from the P regressions subtracted from those obtained in the A regression.

APPENDIX 2

DIFFERENCES IN THE ABSOLUTE VALUE OF RESIDUALS*

MISSOURI

SEG	CORN	SOY	SORGHUM	RICE	COTTON	WOODS	WHEAT
6316	8.44	14.99	1.82	10.33	.52	-3.11	5.18
6317	0.60	-16.46	-.54	.39	-1.52	-.06	1.01
6318	-0.65	6.71	.67	1.56	-.79	-2.76	3.00
6319	4.74	2.48	-1.96	2.06	-1.38	.57	3.00
6320	-0.44	8.72	-6.31	5.47	-.74	-.13	6.34
6326	-0.44	-16.70	2.26	-1.05	3.36	-.16	-16.62
6327	1.23	-21.72	-.12	1.49	1.97	-.16	-7.01
6328	-0.44	-24.90	.17	1.18	1.57	-.07	7.68
6329	0.19	2.64	2.30	1.18	2.41	-.30	-1.50
6330	-0.44	-.77	-6.07	-1.05	1.88	-.16	.45
6331	-0.44	11.72	5.47	-.13	-1.24	-.03	6.87
6332	-0.44	-6.97	1.02	2.48	4.46	-.16	1.55
6346	0.02	-9.60	-1.30	4.64	.61	-.49	6.87
6347	-4.79	37.31	-.79	2.00	4.70	.39	-5.55
6348	-0.44	9.82	1.08	-1.73	.85	-0.03	21.31
6349	-.44	-8.45	-.58	-.57	.98	-.13	-4.14
6350	-.44	-.31	1.76	5.97	2.50	-1.51	-5.15
6351	-.44	3.87	-2.37	1.35	1.77	-.49	-3.94
6356	-.90	-3.24	1.20	1.98	1.11	-.77	3.25
6357	5.20	17.52	.43	6.77	-2.10	-.35	.42
6358	5.21	6.08	5.48	3.10	-.08	1.84	5.42
6359	1.54	-10.11	.25	-.57	1.77	-4.04	2.75
6360	-1.18	-26.32	.08	3.65	.61	-1.52	-2.31
6361	.73	-.52	3.03	-2.50	-.73	-.22	.58
6362	-5.20	.59	-.74	2.46	.70	.60	1.86
6363	.39	-9.95	4.45	-.91	-7.46	-.31	.83
6364	-.44	14.60	.08	-2.55	.61	-.59	-4.48
6365	.60	-2.71	6.05	-9.72	8.81	-.16	1.55
Sum	2.45	-21.68	16.82	37.28	25.15	-14.31	29.22
Average	.09	-.77	.60	1.33	.90	-.51	1.04