7.9-10.0

NASA CR.

15183

Analysis of the Effects of Interpolation and Enhancement of LANDSAT-1 Data on Classification and Area Estimation Accuracy

N. Chu

C. McGillem

P. Anuta

RECEIVED
DEC 1978
NASA STI FACILITY

The Laboratory for Applications of Remote Sensing
Purdue University   West Lafayette, Indiana

1977

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Analysis of the Effects of Interpolation and Enhancement of Landsat-1 Data on Classification and Area Estimation Accuracy		**5. Report Date** November 4, 1977
		6. Performing Organization Code
7. Author(s) N.Y. Chu, C.D. McGillem, P. Anuta		**8. Performing Organization Report No.**
9. Performing Organization Name and Address		**10. Work Unit No.**
		11. Contract or Grant No. NAS9-14970
		13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address Purdue University Laboratory for Applications of Remote Sensing 1220 Potter Drive, West Lafayette, In. 47906		**14. Sponsoring Agency Code**

15. Supplementary Notes

16. Abstract

The effect of the scanner instantaneous field of view and sampling on classification accuracy and area estimation is investigated for crop fields and water bodies. Comparison of classification results using conventional interpolation and an enhancement which compensates for the field of view is presented. It is shown that enhanced and interpolated data produces improved area estimation for water body classification.

17. Key Words (Suggested by Author(s)) Landsat MSS Data Interpolation Enhancement Classification	18. Distribution Statement

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price*

Analysis of the Effects of Interpolation
and Enhancement of LANDSAT-1 Data on
Classification and Area Estimation
Accuracy

N. Chu

C. McGillem

P. Anuta

I. Introduction

Numerical classification of digital multispectral scanner
data from aircraft and satellite sensors using computer techniques
is complicated by many factors. One of these is the effect of the
finite instantaneous field of view of the scanning sensor which
"blurs" or averages the signal from a finite area into a single
generated by the data system. For the Landsat-1 sensor the "blur"
area is approximately an 80 meter diameter circle and for a typical
aircraft scanner system the area may be a circle 10 meters or less
in diameter. This finite area sample may contain "pure" or homo-
geneous scene material or it may contain a mixture of two or more
materials whose boundaries pass through the pixel area. For pixels
covering homogeneous areas the finite pixel area causes little
trouble and in fact classification may be improved due to the smooth-
ing effect of gathering energy from the surrounding areas. For the
overlap case, however, a contamination of pure spectral signatures
results, causing difficulty in properly classifying the boundary
pixels. The work reported here is a preliminary evaluation of the
effects of a particular data enhancement approach aimed at improving
classification performance in such cases.

Some researchers have approached the boundary classification
problem by attempting to model mixture spectra as linear combi-
nations of pure spectra and in so doing attempt to determine the
fractional area in each pixel covered by each pure material.[1,2]
These approaches seek to analyze and model mixture phenomena of
each original pixel and have not proven particularly effective. The
work reported here takes a different approach by attempting to
improve the resolution of the image through use of special signal

processing techniques.

Since the size of the point spread function of the LANDSAT-1 MSS system is fixed it is not possible to directly alter the area encompassed by one pixel of the system output. It is possible, however, to carry out signal processing operations utilizing weighted sums of surrounding pixels to generate new data points or to modify existing data points in a manner so as to reduce the fraction of the data that falls into the boundary category.

One such method makes use of interpolation procedures that generate new points between the original points. This is accomplished by fitting a smooth surface to surrounding points and then computing intermediate points from the equation for the smooth surface.[3] This leads to a more gradual transition to the boundary and thus an increased likelihood that a portion of what was formerly the boundary will fall into one or the other of the classes on either side of the boundary.

A more powerful method of reducing the effects of boundaries is through use of an image restoration filter designed to reduce the effective instantaneous field of view of the scanner.[4,5,6] One such filter that has been developed for LANDSAT data preprocessing provides approximately a 65% reduction in the effective area of a single pixel of ERTS data while still controlling the noise and sidelobe levels in the resultant image.

The restoration filter permits generation of new data points between original points which have a smaller instantaneous field of view as well as reducing the instantaneous field of view of the original points. Thus, the restoration filter method has the potential for increasing the effective resolution of the data

and thereby reducing the percentage of overlap pixels occurring at boundaries relative to the total number in the scene. The problem of the overlap pixel is therefore attacked here by reducing the effective size of the pixel rather than trying to analyze the fractional components of the original pixels.

The results of classifying LANDSAT-1 MSS data after preprocessing both by interpolation and by restoration filtering are described. In Section II, results are presented for the straightforward application of interpolation to typical farm land for the purpose of estimating crop acreages. No general improvement in accuracy is found to result from this procedure. In fact, although the results are mixed, there may be a slight reduction in average accuracy using this technique. These results are inconclusive due to the lack of training statistics, and clear knowledge of the placement of boundaries. Restoration filter preprocessing was not carried out for the crop classification experiment due to resource limitations. This is suggested for further work.

Section III describes the application of interpolation and enhancement techniques to estimation of the areas of lakes. Again, it is found the conventional processing of interpolated data using a single set of training areas gives no improvement in accuracy over uninterpolated data. However, by selecting special training areas from the lakes, it is found that a significant improvement in accuracy is obtained. When the enhancement preprocessing technique is employed a very marked improvement in accuracy is obtained and the results become very consistent. With this procedure the estimation error is reduced by a factor of two over that obtained with the unpreprocessed data.

Section IV discusses certain peculiarities of the analysis procedure used and suggests how further improvements might be made with both the interpolation and enhancement techniques.

II. Crop Acreage Estimation.

The area selected for analysis lies in DeKalb, Ogle and Lee Counties in northern Illinois.[7] These areas are primarily farmland and considerable ground truth is available for this region. The Landsat-1 MSS data for the area was collected on August 9, 1972 (Scene No. 1017-16093).

An area of slightly more than 18,000 acres (128 x 128 pixels) was interpolated with a cubic polynomial (POLYINT)[3] to provide a 4 x 4 enlargement (512 x 512 pixels) of the original data set. The interpolated data set was then classified using standard procedures[8] and compared with classification of the non-interpolated data. The results are shown in Table 1. The classes considered are corn, soybeans, and "other", consisting of all other materials found in the area such as alfalfa, oats, pasture, trees, water, bare soil, etc. The training and test fields were selected from the imagery using ground truth and the boundaries were set so that the IFOV did not include border mixture pixels. Classification was carried out using the statistics of training sets taken from the interpolated data and also using the statistics of training sets taken from the original (uninterpolated) data. It is seen that interpolation does not significantly change the classification accuracy. There is a slight increase in the average class accuracy and a slight decrease in the overall accuracy.

Data	Training field performance		Test field performance	
	overall	average by class	overall	average by class
ORIGINAL DATA	71.3%	67.9%	75.1%	64.5%
INTERPOLATED DATA (All New Points Used Used in Training)	71.3%	68.6%	73.5%	65.9%
INTERPOLATED DATA (Training Fields Same As Original Data)	69.9%	67.0%	73.9%	63.8%

TABLE 1.   Crop Classification Results (line 701-828, col. 1073 - 1328, run 72032803).

The original data set had spectral components with amplitudes in the range 16-40 out of the maximum possible dynamic range of 0-127. In order to see whether this limited dynamic range had adversely affected the interpolation process the dynamic range of the original data was doubled by multiplying all amplitudes by a factor of two. The interpolation was then carried out on this new data set and classification carried out in the same manner as before. The results are shown in Table 2 and are essentially the same as those obtained with the data having a more restricted dynamic range.

Results for a different area are shown in Table 3. Again no appreciable changes in classification accuracy were obtained.

From the above results it appears that there is no improvement in training and test field performance using interpolated data and that there may in fact be a slight loss (1-2%) in accuracy. One possible explanation for this result is as follows. The interpolation procedure produces new points near a boundary that are different from the original boundary pixels and also different from the class within the boundary. However, the training and test areas are chosen completely from within the boundaries and therefore do not include any of these "intermediate" points. Thus the classifier rejects these points as being part of the class corresponding to the training class. As discussed in Section 3 it is likely that by expanding the training areas to include interpolated points near the boundary it may be possible to obtain significant improvement in performance.

Data	Training class performance		Test class performance	
	overall	average by class	overall	average by class
ORIGINAL DATA	71.3%	67.9%	75.1%	64.5%
INTERPOLATED DATA	71.3%	68.5%	73.6%	66.1%
INTERPOLATED DATA (Training Field Same As Original Data)	70.1%	67.1%	74.3%	63.8%

TABLE 2. Results of Crop Classification. Area is Line 701-828, col. 1073-1328, run 72032803. The original data is dynamically doubled. Overflows are less than 0.01%.

Data	Training class performance		Test class performance	
	overall	average by class	overall	average by class
ORIGINAL DATA	73.5%	75.4%	60.9%	56.9%
INTERPOLATED DATA	73.2%	74.8%	58.4%	54.4%
INTERPOLATED DATA (Training Field Same As (1))	71.3%	73.8%	62.0%	57.3%

TABLE 3. Crop Classification Results (line 573-828, col. 1301 - 1328, run 72032803).

III. Estimation of Water Acreage

The accuracy of estimating water area by classification of ERTS MSS data has been studied previously by Bartolucci.[9] Seven of the lakes used in this previous study were selected for analysis. The areas of the lakes range from 15 to more than 1,800 acres and their "true" areas are taken from USGS data. Surveys during the years 1969 to 1971 provide a reliable source of the actual average water area of these lakes in the month of May and these areas were taken to be the true values. The Landsat-1 data was gathered on May 4, 1973 (Scene No. 128515595).

Three types of data sets were analyzed: original data; 4 x 4 interpolated data (POLYINT); and 4 x 3 interpolated and enhanced data. For each of the chosen lakes, the surrounding land area was classified against the class water. A clustering routine was used as a guide to provide the training samples required by the classifier.[10] In general there are several classes existing between the lake water and the surrounding land; e.g., water-land boundary, water-vegetative boundary, and shallow or muddy water. These classes can be investigated by studying their spectral signatures as required. This is discussed in detail by Bartolucci.[9] There are two processes affecting results here. One is the existence of the several boundary classes (which in fact may be a continuous gradation from deep water to land cover) and the other is the effect of the instantaneous field of view. Thus, the situation is more complex than that for the crop field case where the boundary between fields is sharp relative to the instantaneous field of view of the scanner.

It is found experimentally that selection of training areas strongly affects the classification accuracies obtained. As an attempt to reduce the variability produced by this subjective aspect of classification, it was decided to classify all of the chosen lakes using the same set of training areas. The training set was selected from several lakes judged to have typical spectral characteristics. The results of this analysis using interpolated data are shown in Columns 1, 2 and 3 of Table 4. In Columns 1 and 2 the training set was selected from the original data while in Column 3, the training set was selected from the interpolated data.

Since a single training set was used for all classifications it follows that if a particular lake has spectral characteristics that deviate significantly from the norm, then the results may prove less accurate than what is possible when the training sets are selected for each lake individually. Column 4 of Table 4 shows the results obtained when the training areas were selected for each lake individually.

Comparing the results for the original data (Column 1) with those for the corresponding interpolated data (Columns 2 and 3) shows a slight reduction (1-3%) in accuracy of the estimates of area. Note the errors are always on the low side and always are greater percentage-wise for smaller lakes than for larger lakes. This supports the assumption that the error is coming from the inability of the classifier to properly allocate the boundary points to the adjacent classes. In Column 4 where individual training sets for each lake are used and where all points interior

NAME OF LAKE	AREA (ACRES)	1	2	3	4	5
		TRAINING FROM ORIGINAL DATA		TRAINING SET FROM INTER-POLATED DATA	TRAINING SET FROM INDIVIDUAL LAKES	
		ORIGINAL DATA	INTERPOLATED DATA	INTERPOLATED DATA	INTERPOLATED DATA	ENHANCED DATA
MAXINKUKEE	1864	1750.0 (93.8%)	1737.0 (93.1%)	1625.0 (87.1%)	1693.0 (90.8%)	1806.0 (96.8%)
BASS	1400	1310.0 (93.5%)	1305.0 (92.2%)	1313.8 (94.2%)	1290.0 (92.0%)	1409.0 (100.6%)
YELLOW	151	124.4 (82.3%)	121.0 (80.1%)	127.0 (84.0%)	137.9 (91.0%)	127.9 (84.0%)
SILVER	102	81.7 (80.0%)	78.7 (77.1%)	80.4 (78.8%)	100.9 (98.9%)	84.7 (83.0%)
ROCK	56	31.3 (53.8%)	30.0 (53.5%)	27.8 (53.2%)	42.0 (75.0%)	44.7 (79.0%)
LOON	40	29.1 (72.8%)	27.1 (67.7%)	31.7 (79.3%)	33.9 (84.0%)	31.4 (77.0%)
FISH	15	7.8 (52.0%)	7.0 (46.6%)	7.8 (52.0%)	9.0 (60.0%)	11.0 (73.0%)

TABLE 4. Results of Water Area Estimation. Table entries show acreage implied from number of pixels classified as water followed by the percent of correct acreage this represents.

to the points classified as "boundary" are included there is a significant improvement in accuracy. This is very evident for the smaller lakes where the results are substantially better than for the original data.

The analysis of data that was preprocessed with the enhancement algorithm[5] is shown in Column 5 of Table 4. In this case only those points classified as "water" in the training set are includ l in the area estimate. It is seen that improvement in the accuracy of the area estimate is present in every case.

A comparison of the accuracy of the estimates of area as a function of size is given in Figure 1. In this figure data is shown for the original data (Column 1, Table 4), and the enhanced data (Column 5 of Table 4). The ordinate in the figure is the percent of the estimate that must be added to it to give the correct value. The most significant features evident in this figure are the smooth behavior of the estimates obtained from the enhanced data and the erratic behavior for small lakes of the estimates based on the original data. There is clearly a significant improvement in the estimation procedure that results from using the enhanced data. If the results for the interpolated data (Column 4, Table 4) were plotted in Figure 1 they would fall between the curves for the original and enhanced data. However, the points would not fall on a smooth curve but would be somewhat oscillatory.

IV. Discussion and Conclusions

As discussed by Bartolucci[8] there are two basic approaches to water acreage estimation. The first approach is to classify

FIG. 1 PERCENT CORRECTION TO BE ADDED TO ESTIMATE TO OBTAIN CORRECT AREA

ORIGINAL DATA

ENHANCED DATA

% OF ESTIMATE TO BE ADDED

ESTIMATED AREA OF LAKE (ACRES)
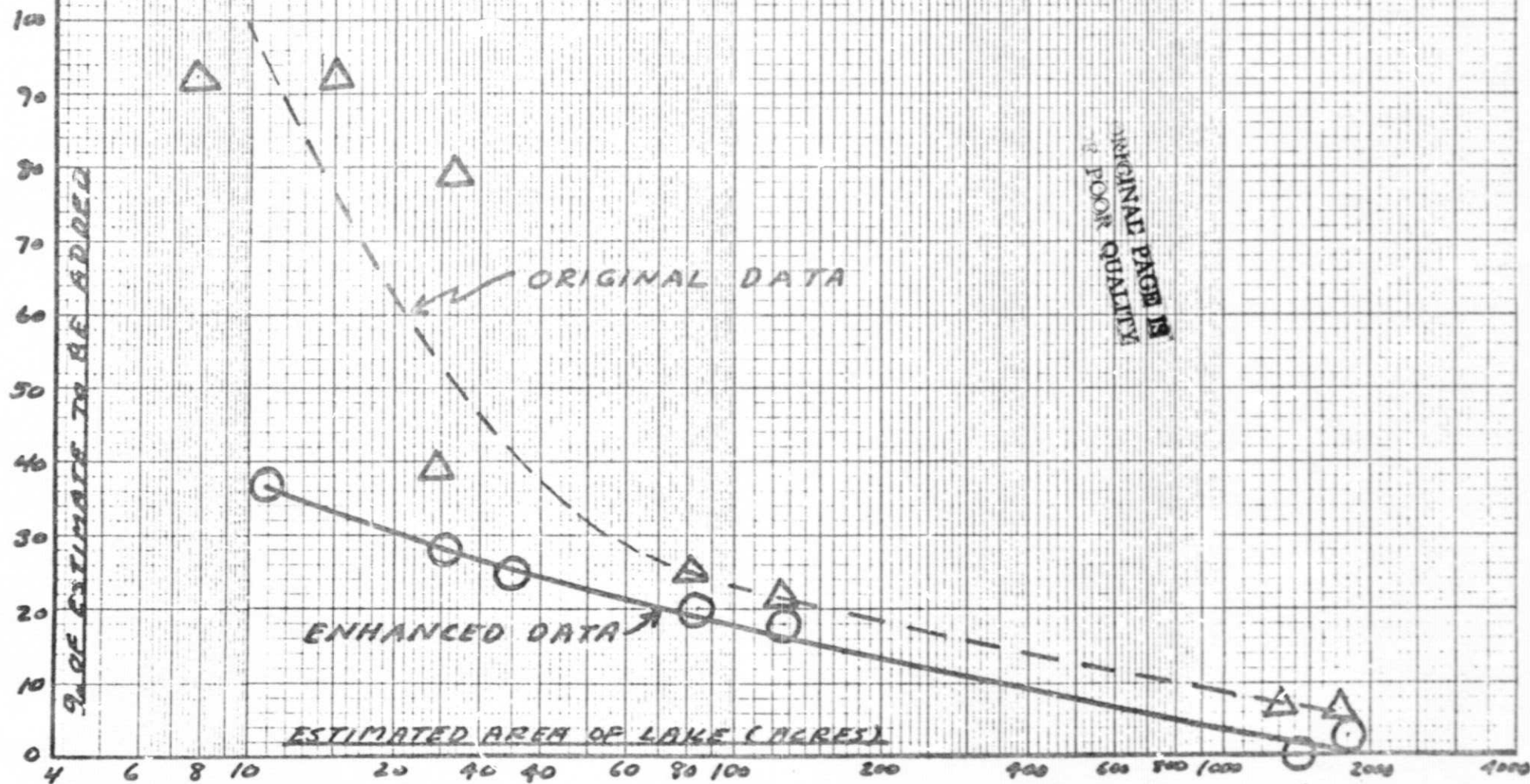
all the water against all other classes present. The number of
points in the class water found by this procedure is then mul-
tiplied by an appropriate scale factor to obtain the final acreage
estimate. This is the method used in Columns 1, 2, 3 and 5 of
Table 4. For this procedure interpolation provides no improve-
ment while enhancement provides a significant improvement.

The second approach is to estimate a boundary and subwater
classes near the boundary. Which particular points fall in the
subwater class is determined from the spectral characteristics
of the clustered data. The subwater class points inside the
boundary are then added to the water class points to give the
total used in making the estimate. A typical set of cluster re-
sults for interpolated data is shown in Table 5 and Figure 2 which
corresponds to data for Rock Lake. It is seen that between the
class water (symbol W) and land (symbol F) there are two distinct
intermediate classes. These are designated the boundary (symbol B)
and the subwater (symbol O). If the basis of employing interpo-
lation is that it reveals more details near the boundary then
these classes correspond to that information and should be used
to improve the estimation. It is this procedure that was used to
produce the data of Column 4 in Table 4. Clearly the error of
the estimate was reduced below that of the original data. How-
ever, it is believed that further improvements can be made by more
careful determination of the proper subwater class charecteristics
and the number of such classes to utilize in the processing opera-
tion.

Whether the improved techniques using interpolated data will
exceed the performance with enhanced data and whether use of sub-
water classes with the enhanced data gives further improvement

Figure 2. Cluster Results for Interpolated LANDSAT Data
Rock Lake, 10 Clusters, 4 Channels.

```
CLUSTER   POINTS   MEANS
                        CH( 1)  CH( 2)  CH( 3)  CH( 4)
    1 (   )    113       28.52   20.18   59.42   38.05
    2 ( + )    155       37.33   40.46   43.99   21.95
    3 ( 1 )    284       32.15   28.93   43.98   24.15
    4 ( L )    271       28.49   22.04   50.04   30.65
    5 ( V )    325       26.96   20.47   43.72   25.28
    6 ( Y )    250       29.09   24.07   38.96   21.85
    7 ( F )    180       26.35   19.77   33.80   18.00
    8 ( B )    169       25.84   19.40   27.01   13.05
    9 ( C )    180       25.28   18.51   20.41    7.97
   10 ( W )    438       24.45   18.01   15.04    4.26
```

```
CLUSTER VARIANCES

        CH( 1)  CH( 2)  CH( 3)  CH( 4)
    1    2.02    4.18    6.85    5.60
    2    2.63   11.32    7.09    3.94
    3    2.27    7.08    5.96    3.92
    4    2.61    5.45    7.68    7.42
    5    3.06    4.06    4.56    2.73
    6    2.04    4.65    3.81    2.02
    7    2.63    6.41    5.04    3.36
    8    2.46    6.02    4.02    2.71
    9    1.56    3.65    3.28    2.24
   10    3.04    3.88    3.04    0.93
```

TABLE 5. Mean Vector and Covariance Matrix of the 10 Classes
          of Lake Rock.

are yet to be determined.

The crop classification experiment did not include boundary pixels in the tests and did not use resolution enhanced data. Both these elements should be included in future studies to explore the full value of the preprocessing techniques.

REFERENCES

1. Detchmendy, D. M., Pace, W. H., "A Model for Spectral
    Signature Variability for Mixtures," Proceedings of the
    Earth Resources and Information Analysis Systems Confer-
    ence, Tullahoma, Tennessee, March 13, 1972.

2. Salvato, P. Jr., "Iterative Techniques to Estimate Sig-
    nature Vectors for Mixture Processing of Multispectral
    Data," Proceedings of Conference on Machine Processing
    of Remotely Sensed Data," October 16-18, 1973, Purdue
    University, W. Lafayette, Indiana, pp. 3b-48 - 3b-59.

3. McGillem, C. D., Interpolation of ERTS-1 "Multispectral
    Scanner Data," LARS Information Note 022175, Purdue Univer-
    sity, W. Lafayette, Indiana, 1975.

4. Riemer, T. E., and McGillem, C. D., "Ghost Suppression
    in Image Registration Filters," Proc. IEEE Int'l Conf. on
    Communications, June 16-18, 1975.

5. McGillem, C. D., Riemer, T. E., and Mobasseri, G. B.,
    "Resolution Enhancement of ERTS Imagery," Proc. Symp. on
    Machine Processing of Remotely Sensed Data," Purdue Univer-
    sity, W. Lafayette. June 3-5, 1975.

6. Riemer, T. E. and McGillem, C. D., "Optimum Constrained
    Image Restoration Filters", LARS Information Note 091974,
    Purdue University, W. Lafayette, Indiana, 1974.

7. Landgrebe, D. A. and Staff, "A Study of the Utilization of
    ERTS-1 Data from the Wabash River Basin - Final Report,"
    LARS Information Note 052375, Laboratory for Applications
    of Remote Sensing, W. Lafayette, Indiana, May 1975.

8. Bauer, M. E. and Cipra, J. E., Identification of Agri-
   culture Crops by Computer Processing of ERTS MSS Data,
   LARS Information Note 030173, Purdue University, W.
   Lafayette, Indian 1973.

9. Bartolucci, L., Section in ERTS Wabash Valley Study Final
   Report "Water Resources Research", LARS Information Note
   052375, May 1975.

10. Tarnocai, C. and Kristof, S. J., "Computer-Aided Class-
    ification of Land-Water Boundaries Using ERTS Data -
    MacKenzie Delta Area, N.W.T., "LARS Information Note
    031875, Purdue University, W. Lafayette, 1975.