Reprinted from

Tenth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Thematic Mapper Data and

Geographic Information Systems

June 12 - 14, 1984

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

ALGORITHMS FOR THE ESTIMATION OF FAILED DETECTOR DATA

B. GUINDON

Canada Centre for Remote Sensing
Ottawa, Ontario, Canada

I ABSTRACT

Imagery acquired by the LANDSAT-D Thematic Mapper exhibits partial data loss due to failed detectors. A recent study has been undertaken at the Canada Centre for Remote Sensing (CCRS) to identify a suitable algorithm for missing data replacement.

Two classes of algorithms have been considered. The first type involves grey level interpolation based on good data from adjacent detectors of the same band. The principal disadvantage of this approach is the potential distortion of the image spatial frequency spectrum introduced by the interpolation process. In order to overcome this problem, a second class of estimation has been studied which utilizes the observed grey level modulation in adjacent bands to aid in the estimation process. The success of the modulation techniques depend on the level of intraband correlation.

Extensive evaluations of candidate algorithms have been carried out using both real TM imagery and simulated imagery derived from airborne multispectral scanner coverage. It has been found that inclusion of adjacent band modulation improves estimation accuracy. In the cases of TM bands 1,2 and 3 intraband correlations are observed to range between 0.89 to 0.99 for a variety of surface classes. Estimation accuracies of 0.5 to 1.5 grey levels have been achieved in such cases.

Based on these evaluations optimum replacement algorithms are proposed for each TM band.

II INTRODUCTION

The thematic mappers of LANDSAT-4 and LANDSAT-5 employ arrays of detectors in order to acquire either 16 or 4 lines of video per swath. Partial data loss has already been experienced with the LANDSAT-4 sensor due to failed detectors (e.g. band 2 detector 4 and band 5 detector 3). Since multiple detector arrays can be expected to be a feature of many future sensors (e.g. the MLA and PLA of SPOT), the development of algorithms to accurately estimate failed detector data is needed.

A study has been carried out at the Canada Centre for Remote Sensing (CCRS) to compare the performances of a number of computationally simple replacement algorithms. Most of these algorithms utilize only data from neighbouring detectors of the band containing the failed detector. Recently, Bernstein and Lotspiech (1983) have suggested that information from adjacent bands might be usefully employed if adjacent band correlation is high. An adjacent band modulation technique has therefore been developed and evaluated in this paper.

In section III of the paper the candidate algorithms are listed while section IV contains a description of the test imagery used during the evaluation experiments.

Section V includes a discussion of the statistical parameters used to rank replacement algorithms and a summary of the test results.

Section VI concludes with a discussion of other implications of the algorithms in particular problems related to their operational implementation and finally, with an algorithm recommendation for TM data estimation.

III REPLACEMENT ALGORITHMS

A total of 4 algorithms have been evaluated, three of which involve grey level interpolation using only good data from a single band (i.e. the band containing the failed detector). These single band algorithms are listed below.

1) <u>Adjacent Line Substitution (ALS)</u>

$$B_{ij} = B_{ij-1} \quad (1)$$

2) Linear Interpolation (LI)

$$B_{ij} = (B_{ij-1} + B_{ij+1}) / 2 \qquad (2)$$

3) 4-Point Cubic Spline Interpolation (CS)

$$B_{ij} = \frac{11}{16} \bullet (B_{ij-1} + B_{ij+1})$$

$$- \frac{3}{16} (B_{ij-2} + B_{ij+2}) \qquad (3)$$

In the above equations B_{ij} is the grey level of line j pixel i in some band B which has to be estimated due to a failed detector. It is also assumed that good data is available from either the two adjacent or nearest 4 detectors. It should be noted that the coefficients of the cubic spline algorithm have been derived based on the assumption that a gap of two lines exists between samples B_{ij-1} and B_{ij+1}.

Since the above interpolation algorithms utilize at most 4 neighbouring samples and the input data for TM is undersampled (i.e. sampled at less than the Nyquist rate), one can expect distortions in the spatial frequency content of the reconstructed image particularly at the highest spatial frequencies. Bernstein and Lotspiech (1983) have suggested that recovery of high spatial frequencies may be enhanced by using the observed grey level modulation in an adjacent spectral band as an additional information source. An adjacent band modulation (ABM) algorithm has therefore been developed and evaluated in this paper. The form of the algorithm is based on the assumption that the correlation between adjacent band grey levels can be expressed in a simple multiplicative sense and that this correlation is stationary over distances of the scale of the interpolation window (i.e. 3 to 5 lines). The algorithm has the form:

$$B_{ij} = A_{ij} \left[W1 \bullet \left(\frac{(B_{ij-1} - B_0)}{A_{ij-1}} + \frac{(B_{ij+1} - B_0)}{A_{ij+1}} \right) + W2 \bullet \left(\frac{(B_{ij-2} - B_0)}{A_{ij-2}} + \frac{(B_{ij+2} - B_0)}{A_{ij+2}} \right) \right] / 2(W1+W2) + B_0 \qquad (4)$$

where

A_{ke} ≡ grey level of pixel k line e in a selected adjacent band A

W1, W2 ≡ weighting factors

B_0 ≡ grey level in band B corresponding to a grey level of zero in band A.

The variable B_0 has been included to allow for the possibility that a grey level offset exists between bands A and B. For a given band pair it can be estimated from a linear least squares fit to corresponding grey level measurements from bands A and B.

In some of the experiments described below two adjacent bands were available for modulation. The adjacent band exhibiting the highest correlation with band B was selected for the estimation process.

The final consideration in utilizing the ABM algorithm is the selection of weighting factors. Two cases have been studied here, one involving only adjacent lines (W1=1, W2=Ø) and the other the nearest 4 lines all equally weighted (W1=W2=1). These cases will be abreviated for conveninece to ABM(1,0) and ABM(1,1) respectively.

IV TEST DATA

The characteristics of the test scenes used in the evaluation are summarized in Table 1. Both LANDSAT-4 TM and simulated TM imagery were considered. The test data set was selected to reflect both a range in surface classes and also terrain exhibiting a high spatial frequency information content.

The simulated images (test subscenes 1 and 2) were acquired with a Daedalus multispectral scanner (MSS) operated by CCRS and included both agricultural and forest areas near Ottawa, Canada. The high resolution data (15 meter resolution) was subsequently degraded to 30 meter resolution by means of truncated sinx/x convolution (Shlien, 1979).

The 3 LANDSAT-4 subscenes were extracted from a scene covering the coastal regions of California (WRS path 42 row 36). The scene was acquired at the CCRS Prince Albert receiving station on November 15, 1982. Preprocessing of the image was limited to detector alignment and swath reversal (Butlin and Murphy, 1983).

The California frame was deemed useful for the present study because of its range of surface classes (ocean, mountains, agriculture and urban coverage). Of the three subscenes studied one contained all major classes, a second only agricultural fields and a third only mountains. Hereafter these subscenes will be referred to as test subscenes 3, 4 and 5 respectively.

V STATISTICAL PARAMETERS AND TEST RESULTS

A. STATISTICAL PARAMETERS

The performance of a replacement algorithm has been evaluated based on its ability to accurately predict the grey levels of subsets of pixels in each of the test image set channels. The observed error distribution for a given algorithm/channel combination (i.e. the distribution of (known-predict) grey levels) was characterized by three performance indicators, namely, mean error (m_e), standard deviation of the error distribution (σ) and largest observed error (l_e). Rankings of algorithms were then made by comparing the indicator values observed from application of all algorithms to the same test channel. The robustness of an algorithm could also be evaluated by comparing the resulting indicator values observed from its application to the same TM band of different subscenes.

In the case of the simulated imagery, each algorithm has been applied to predict the grey levels of all pixels within the pixel, line coordinate limits (1,3) to (400,398). This is reasonable since the airborne MSS acquires only 1 line of video per swath.

The LANDSAT-4 TM, on the other hand, acquires 16 lines of video per swath, each line from a different detector. In order to better simulate the operational application of a replacement algorithm and to exclude known failed detector data from the study, 5 test runs were applied per channel per algorithm. Each test run corresponded to the replacement of a different detector. The estimated performance indicators for a given test channel was then taken to be the average of the 5 runs.

B. PRINCIPAL RESULTS

The following results are derived from analysis of the data presented in Tables 2 to 6. These results are grouped according to performance indicator.

Mean Error

1. Although non zero mean errors were observed during individual trial experiments, only linear interpolation exhibited a consistent bias in mean error. The LI offset, of approximate magnitude -0.25 grey level, is a quantization effect which results from the rounding up of non integer interpolated grey levels.

2. In only 3 of the 150 cases studied did the mean error (me) to standard deviation (σ) ratio exceed 0.2. This suggests that standard deviation is a better measure of algorithm performance than mean error.

Standard Deviation

Table 7 lists the preferred algorithm for each test subscene channel based on minimum observed standard deviation.

3. Overall, the ABM algorithms are consistently preferred for TM bands 1,2,3 and 7.

4. There appears to be no preferred weighting factor set for the ABM algorithm.

5. Single band interpolation is favoured for band 4. The poor performance of the ABM algorithms for this band can be attributed to the fact that band 4 is poorly correlated with bands 3 and 5. In the 5 cases studied, the full subscene correlations between band 4 and its best adjacent band were characterized by correlation coefficients ranging from 0.08 to 0.90. This is in contrast with bands 1,2,3 and 7 where best adjacent band correlation coefficients always exceeded 0.89.

6. The discrepant selection of CS over ABM for band 5 of test subscene 2 can also be explained from correlation arguments. In subscenes 3,4 and 5 it was observed that bands 5 and 7 were highly correlated while bands 4 and 5 were poorly correlated. In subscene 2 band 7 was unavailable, hence, the ABM algorithm was run on band 5 using the observed modulation of band 4.

7. Although the test subscenes contain quite diverse scene contents, the preferred algorithm selections based on the trial runs are, in general, consistent for all subscenes. This indicates that the ABM algorithm is robust and could be utilized for a wide range of data scene types.

Largest Observed Error

Table 8 lists the preferred replacement algorithm selections based on minimization of the largest observed error.

8. Comparing Tables 7 and 8 it can be seen that both indicators reflect the strong performance of the ABM algorithms for bands 1,2,3,5 and 7, and a consistent preference for single band interpolation for band 4. As discussed in result 6, the discrepant subscene 2 band 5 preference can be attributed to the absence of band 7 data.

VI IMPLEMENTATION CONSIDERATIONS AND CONCLUSIONS

A. IMPLEMENTATION CONSIDERATIONS

Besides algorithm accuracy in predicting grey levels, there are other factors which should be considered in algorithm selection. The pros and cons of each interpolation candidate algorithm are listed below.

Linear Interpolation

pros
- Computationally simple.
- Can be applied prior to radiometric calibration.

cons
- Spatial smoothing effects.
- Cannot be applied readily to replace either the 1st or the 16th detector unless scan gap widths are incorporated into the algorithm.

Cubic Spline

pros
- Can be applied prior to radiometric calibration.
- Relatively simple to apply from a computation point of view.

cons
- Not readily applicable for the replacement of the 1st, 2nd, 15th or 16th detectors.
- Spatial smoothing.

Adjacent Band Modulation

pros
- Geared to maintaining spatial resolution.
- Readily applicable to the replacement of any failed detector.
- Could be applied using information from only 1 adjacent line.

cons
- Requires adjacent band as well as adjacent line information.
- Radiometric calibration as a desired prerequisite process.
- Computationally more complex than single band interpolation algorithms.

B. CONCLUSIONS

A study has been carried out to evaluate algorithms which could be used to estimate failed detector data. Of the four algorithms studied, three involved single band data (adjacent line replacement, linear interpolation and cubic spline interpolation). A fourth algorithm also utilized the observed modulation in an adjacent band to aid in missing data estimation (adjacent band modulation (ABM)).

Results from tests carried out on five LANDSAT-4 and simulated TM subscenes have shown that adjacent band modulation in general, outperforms all other methods. The exception to this result is in band 4, a band which is not strongly correlated with either of its adjacent bands. Full subscene correlation for this band and its best adjacent band were found to exhibit coefficients in the range 0.08 to 0.90. Since results with weighting factors (1,0) and (1,1) are not significantly different, it concluded that accurate estimations (i.e. scene rms errors of typically 1 to 3 grey levels) can be achieved using adjacent detector data only. Accuracies of this level were achieved when strong adjacent band correlation was observed (i.e. in excess of 0.89) and were consistently met for bands 1,2,3,5 and 7 over a wide range of surface classes.

The ABM algorithm is more complex to implement since it requires data from two bands and the a priori estimation of a band to band grey level offset; however, unlike cubic spline or linear interpolation the ABM algorithm is readily applicable to all detectors since implicitly it does not require knowledge of pixel spacing.

Finally, based on the trial runs the selected adjacent bands for ABM processing of TM 2,3,4 and 5 are observed to be 3,2,3 and 7 respectively. ABM experiments with TM 1 and 7 have been limited to adjacent band selections of 2 and 5 respectively.

REFERENCES

Bernstein, R. and J. Lotspiech, 1983, 'Analysis and Evaluation of the LANDSAT-4 MSS and TM Sensors and Ground Data Processing System', Early Results Symposium.

Butlin, T. and J. Murphy, 1983, 'Early Processing of Thematic Mapper Spectral Data by the Canada Centre for Remote Sensing', Proceedings of the Seventeenth International Symposium on Remote Sensing of Environment, in press.

Shlien, S., 1979, 'Geometric Correction, Registration and Resampling of LANDSAT Imagery', Canadian Journal of Remote Sensing, 5, pp.74-89.

AUTHOR BIOGRAPHICAL DATA

B. Guindon. Dr. Bert Guindon is a Research Scientist with the Canada Centre for Remote Sensing. He received a Doctorate in Physics from Queen's University in 1976. Since joining CCRS in 1978, he has been actively involved in a variety of image processing projects. His primary research interests include geometric correction and digital SAR image analysis.

TABLE 1: Summary of the characteristics of the test subscenes used in the replacement algorithm evaluations

Data Characteristic	Test Subscene 1	2	3	4	5
Data Source	Airborne MSS (Simulated TM)	Airborne MSS (Simulated TM)	LANDSAT-4 TM	LANDSAT-4 TM	LANDSAT TM
TM Bands	1-4	1-5	1-5,7	1-5,7	1-5,7
Subscene Size					
No. of Lines	400	400	400	400	400
No. of Pixels	400	400	1024	512	512
Surface Classes	Agriculture	Agriculture, Forests	Ocean, Agriculture, Mountains, Urban	Agriculture	Mountains

TABLE 2: Summary of the performances of the adjacent line replacement (AL), linear interpolation (LI), cubic spline (CS) and the two adjacent band modulation (ABM (1,0) and ABM (1,1)) algorithms on test subscene 1. The performance indicators listed include mean error (m_e), standard deviation of the error distribution (σ) and largest observed error (l_e)

Replacement Algorithm

TM BAND	AL m_e	AL σ	AL l_e	LI m_e	LI σ	LI l_e	CS m_e	CS σ	CS l_e	ABM(1,0) m_e	ABM(1,0) σ	ABM(1,0) l_e	ABM(1,1) m_e	ABM(1,1) σ	ABM(1,1) l_e
1	-0.02	4.10	121	-0.25	2.58	100	-0.04	2.54	92	0.03	1.05	22	0.02	1.29	26
2	-0.02	3.16	78	-0.25	2.01	66	-0.04	1.91	59	0.01	0.77	24	-0.01	0.94	26
3	-0.02	4.44	94	-0.25	2.84	78	-0.04	2.67	71	0.04	1.03	23	0.03	1.21	31
4	-0.02	5.53	87	-0.25	3.72	62	-0.03	3.29	56	0.01	3.79	71	0.02	5.04	86

TABLE 3: Summary of the performances of the adjacent line replacement (AL), linear interpolation (LI), cubic spline (CS) and the two adjacent band modulation (ABM(1,0) and ABM(1.1)) algorithms on test subscene 2. The performance indicators listed include mean error (m_e), standard deviation of the error distribution (σ) and largest observed error (l_e)

Replacement Algorithm

TM Band	AL			LI			CS			ABM(1,0)			(ABM(1.1)		
	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e
1	0.01	6.70	112	-0.25	4.11	86	-0.03	3.99	78	0.03	1.91	63	0.03	2.39	67
2	0.01	5.51	139	-0.25	3.38	109	-0.04	3.23	97	0.06	1.34	41	0.07	1.71	42
3	0.02	7.67	136	-0.25	4.82	97	-0.03	4.52	94	-0.07	1.85	55	-0.12	2.38	56
4	0.06	8.81	112	-0.25	5.82	69	-0.03	5.17	69	-0.05	6.04	87	-0.05	8.14	147
5	0.01	3.06	95	-0.24	1.87	79	-0.03	1.77	63	-0.04	2.33	83	-0.05	3.20	103

TABLE 4: Summary of the performances of the adjacent line replacemnt (AL), linear interpolation (LI), cubic spline (CS) and the two adjacent band 3 modulation (ABM(1,0) and ABM(1,1)) algorithms on test subscene 3. The performance indicators listed include mean error (m_e), standard deviation of the error distribution (σ) and largest observed error (l_e). Values quoted are the average of 5 trials each involving the replacement of a different TM detector.

Replacement Algorithm

TM Band	AL			LI			CS			ABM(1,0)			ABM(1,1)		
	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e
1	0.01	7.10	145.0	-0.32	2.25	60.0	0.07	2.46	57.4	-0.08	1.46	55.6	-0.02	1.31	77.0
2	0.02	4.53	83.0	-0.23	1.36	34.6	-0.11	1.43	33.2	0.10	0.68	12.4	0.04	0.62	11.6
3	-0.21	4.61	79.0	-0.22	1.35	31.8	0.01	1.44	26.2	0.03	1.01	53.4	-0.02	0.90	72.6
4	-0.01	8.64	92.6	-0.20	2.39	42.0	0.11	2.38	39.8	-0.02	2.44	45.4	0.04	2.95	50.6
5	0.18	16.95	136.6	-0.29	4.54	69.4	-0.04	4.39	67.6	0.11	2.39	26.6	0.14	2.37	29.8
7	0.04	9.54	71.2	-0.19	2.52	37.0	-0.04	2.54	35.8	0.03	1.32	11.6	0.01	1.33	14.2

TABLE 5: Summary of the performances of the adjacent line replacement (AL), linear interpolation (LI), cubic spline (CS) and the two adjacent band modulation (ABM(1,0) and ABM(1,1)) algorithms on test subscene 4. The performance indicators listed include mean error (m_e), standard deviation of the error distribution (σ) and largest observed error (l_e). Values quoted are the average of 5 trials each involving the replacement of a different TM detector.

Replacement Algorithm

TM	AL			LI			CS			ABM(1,0)			ABM(1,1)		
Band	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e
1	-0.13	8.92	96.0	-0.41	2.55	55.8	0.15	2.81	55.2	-0.18	1.53	13.8	-0.30	1.40	13.2
2	-0.13	5.47	54.4	-0.28	1.46	30.6	-0.08	1.57	29.8	0.09	0.75	7.4	0.09	0.72	7.0
3	-0.23	8.67	75.6	-0.33	2.08	37.2	-0.01	2.23	37.2	-0.04	1.13	10.4	-0.10	1.08	9.8
4	-0.42	13.28	76.4	-0.31	2.75	33.4	-0.07	2.77	37.6	0.07	3.20	44.6	0.16	3.93	51.4
5	-0.68	19.20	112.4	-0.20	4.20	45.6	0.02	4.22	45.4	-0.09	2.60	34.2	-0.14	2.84	62.2
7	-0.38	14.23	73.4	-0.26	2.87	28.6	-0.08	2.88	27.8	0.17	1.66	22.4	0.19	1.66	31.8

TABLE 6: Summary of the performances of the adjacent line replacement (AL), linear interpolation (LI), cubic spline (CS) and the two adjacent band modulation (ABM(1,0) and ABM(1,1)) algorithms on test subscene 5. The performance indicators listed include mean error (m_e), standard deviation of the error distribution (σ) and largest observed error (l_e). Values quoted are the average of 5 trials each involving the replacement of a different TM detector.

Replacement Algorithm

TM	AL			LI			CS			ABM(1,0)			ABM(1,1)		
Band	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e	m_e	σ	l_e
1	0.11	7.33	138.6	-0.35	2.24	76.2	0.10	2.74	72.8	0.08	1.57	18.8	-0.21	1.64	16.0
2	-0.08	4.71	70.4	-0.17	1.37	38.6	-0.19	1.56	35.4	0.14	0.710	7.8	0.09	0.64	8.0
3	0.12	6.95	87.0	-0.26	1.97	45.6	-0.08	2.20	41.2	-0.33	0.82	8.6	-0.20	0.94	9.0
4	-0.17	11.98	69.0	-0.43	2.98	25.2	-0.05	3.14	25.6	0.01	2.26	32.2	-0.01	2.62	48.0
5	0.52	22.98	113.0	-0.17	5.79	50.6	0.01	6.04	53.0	-0.10	2.72	23.0	-0.33	2.66	20.6
7	0.33	10.73	55.2	-0.31	2.84	25.8	-0.02	3.00	27.8	0.13	1.27	12.4	0.21	1.18	10.8

TABLE 7: Summary of the preferred replacement algorithm for each test subscene channel based on minimum standard deviation

TM Band	Test Subscene 1	2	3	4	5
1	ABM(1,0)	ABM(1,0)	ABM(1,1)	ABM(1,1)	ABM(1,0)
2	ABM(1,0)	ABM(1.0)	ABM(1,1)	ABM(1,1)	ABM(1,1)
3	ABM(1,0)	ABM(1,0)	ABM(1,1)	ABM(1,1)	ABM(1,0)
4	CS	CS	CS	LI	ABM(1,0)
5		CS	ABM(1,1)	ABM(1,0)	ABM(1,1)
7			ABM(1,0)	ABM(1,0)/ ABM(1,1)	ABM(1,1)

TABLE 8: Summary of the preferred replacement algorithm for each test subscene channel based on largest grey level error

TM Band	Test Subscene 1	2	3	4	5
1	ABM(1,0)	ABM(1,0)	ABM(1,0)	ABM(1,1)	ABM(1,1)
2	ABM(1,0)	ABM(1,0)	ABM(1,1)	ABM(1,1)	ABM(1,0)
3	ABM(1,0)	ABM(1,0)	CS	ABM(1,1)	ABM(1,0)
4	CS	CS/LI	CS	LI	LI
5		CS	ABM(1,0)	ABM(1,0)	ABM(1,1)
7			ABM(1,0)	ABM(1,0)	ABM(1,0)