LARS Technical Memorandum T-10 060773

Calibration of the University of Michigan Aircraft Multispectral Scanner Data Using Smoothed Calibration Coefficients

by

David F. Strahorn & Paul E. Anuta

Abstract

A preprocessor (SMCAL) is described which produces a data tape which is calibrated in any desired calibration code with smoothed calibration coefficients. The C_1-C_0 and C_2-C_0 values used to gain calibrate each line of data are linearly smoothed over nine lines before and after each line.

Calibration of the University of Michigan Aircraft Multispectral Scanner Data Using Smoothed Calibration Coefficients

David F. Strahorn & Paul E. Anuta

Calibration of the University of Michigan aircraft multispectral scanner data has been found to degrade the data quality in many cases. Quite often, the classification accuracy of this calibrated data is lower than for the same uncalibrated data. This degradation of the data quality is not believed to be due to the calibration technique, but to noise in the calibration pulses associated with each line of data. When these noisy calibration pulses are used to make gain changes in a line of data, as in the LARSYS (GADLIN) program, each sample of data in the line has this noise superimposed on it. Minimization of the noise in the calibration pulses used for gain calibration should then decrease the degradation of the data which results from gain calibration.

Sources of Calibration Pulse Noise

The noise enters the calibration pulses during data collection, analog data storage, and data digitization. Although errors are also introduced during these stages which cause the values stored of the LARS data tape to be poor estimates of the actual situation that they were intended to monitor, this study is primarily interested in the variation from line to line of these pulses.

Since the scanner is analog, the calibration pulses are distorted slightly by photomultipliers, preamplifiers, amplifiers, and the analog recorder. The signal stored on the analog tape then has a noise signal superimposed. Three calibration pulses per line are digitized:

C_0; black reference level, C_1; calibrated lamp illumination, and C_2; solar illumination. These pulses are digitized by averaging a number of samples taken from each pulse to get a value for each. Unfortunately, the C_1 and C_2 pulses are so narrow that only three samples can be taken from each while fifty samples are taken for C_0. Figure 1 illustrates a normal line of analog data with the calibration pulses magnified showing variation along the pulse plateaus. Because of this variation, the same pulse digitized several times would yield several values clustered about the actual pulse mean. Since the average of three samples does not accurately approximate the pulse mean, it would be expected that the digitized values of a string (number of lines) of pulses of equal mean would vary about the mean. Variations in the actual calibration pulses are expected to be slow drift changes with little change from line to line. However, significant line to line changes have been noted in much of the Michigan scanner data. Since these line to line variations don't reflect the actual system variations, they may be considered noise and basically undesireable.

Magnitude of Calibration Pulse Noise

The magnitude of the line to line variations should be about constant for all channels and pulse means (each channel has a different circuit in the scanner and A-to-D system so that the noise would be a function of channel). It should therefore be expected that the channels with lowest pulse mean would also have the lowest signal to noise ratio. Figure 2 contains a $IDPRINT for a typical LARS data set (run number 71045700; segment 225, mission 42 of the Corn Blight Watch Experiment) collected with the "new" (July 1971) University of Michigan multispectral scanner.

The values for channels 9, 10, and 11 of $\bar{C}_1 - \bar{C}_0$ (where: $\bar{C}_0$ - mean black level pulse from ID record, $\bar{C}_1$ - mean calibrated lamp pulse from ID record, and $\bar{C}_2$ - mean solar illumination pulse from ID record) and for channels 8 and 9 of $\bar{C}_2 - \bar{C}_0$ are relatively low. These mean calibration pulses are the averages of the first 200 lines of individual calibration pulses. Since gain calibration is accomplished by multiplying each sample of data by the pulse mean divided by the pulse for that line ($(\bar{C}_1 - \bar{C}_0)/(C_1 - C_0)$ or $(\bar{C}_2 - \bar{C}_0)/(C_2 - C_0)$), the gain factors will vary greatly from line to line when the pulse means are low. This causes the data to become "streaky". Figure 3 contains a column graph of $C_1 - C_0$ for lines 100 - 150, channels 3, 6, and 9, and figure 4 contains a similar graph of $C_2 - C_0$. The values of $\bar{C}_1 - \bar{C}_0$ for channel 9 is 12.85 while the range of the individual $C_1 - C_0$ values is 11 over the fifty lines of the graph, almost 90% of the mean. These noisy gain factors are used to gain calibrate, serious degradation of the data results. Figure 5a contains a digital display photograph of channel 9 which has been gain calibrated using $C_1 - C_0$. This calibrated data is of little use for spectral identification of the ground scene.

Correction of the Problem

It is clear that calibration of LARS data collected with the University of Michigan multispectral scanner, with calibration pulses having line to line variations which are large compared to their means, results is severe degradation of the data quality. Minimization of this problem may be accomplished in one of three ways: 1) raise the mean level of all calibration pulses to about 150 - 200 where the variations will be small enough to ignore, 2) decrease the noise level of the pulses, or 3) smooth out the variations in the pulses.

The first two alternatives require modifications to the data collection, storage, and digitization hardware, therefore, the third was selected. The data reformatting program was modified on August 5, 1971 so that the calibration coefficients stored with each line are the averages of the calibration pulses for that line and the nineteen preceding lines. This still leaves two problems, much data was reformatted before the modification of the program and even though the calibration coefficients have been smoothed, their restorage on the data tape results in a quantization error of up to 10% which can still cause serious degradation of the data. The only way to eliminate the effect of quantization error is to do the smoothing and calibration at the same time. For this purpose, the smoothing calibrator (SMCAL, LARSYS 1039) was written.

Smoothing Algorithm

The black level coefficient, used to calibrate out bias changes, is the average of fifty samples per line on the analog tape, therefore its variations are considered significant and are not smoothed. The values used for gain calibration $C_1 - C_0$ and $C_2 - C_0$ are smoothed with the following algorithm:

$$S_{i,j,k} = C_{o,j,k} + [20C_{i,j,k} + \sum_{L=1}^{9} (20-L)((C_{i,j,k+L} - C_{o,j,k+L}) + (C_{i,j,k-L} - C_{o,j,k-L}))]/290$$

Where:

$C_{i,j,k}$ - original calibration coefficient

$S_{i,j,k}$ - smoothed calibration coefficient

k - line number

j - channel number

i - calibration coefficient code

0 - black level

1 - calibrated lamp illumination

2 - solar illumination

Test of Algorithm

The Smoothing Calibrator was tested on six sets of data from the 1971 Corn Blight Watch Experiment (LARS run numbers: 71029100, 71040100, 71045700, 71078000, 71078400, and 71078600). The ability to distinguish corn from noncorn ground cover was tested for SMCAL using calibration code 4 against original data using calibration codes 1 and 4. The results of the classification test are shown in tables 1a through 1f. The percent improvement in overall test performance was statistically compared (t test) for three sets of paired data at a 90% confidence level and the results displayed in table 2. The performance of the original data in calibration code 1 was found to be significantly higher than either gain calibrated data set. The performance of the smoothed data was on the average about 3/4% higher than the unsmoothed gain calibrated data, but this was not a significant improvement at the 90% confidence level. None of the effects were found significant at the 99% confidence level.

Discussion of Results

The test classifications were all performed with the sets of channels selected during the Corn Blight Watch Experiment for the original data in calibration code 1. It is possible that this is not the best set of channels for the gain calibrated data and that if the classifications had been performed with the proper set of channels. The test performance of the gain calibrated data would have been higher.

Even though the original data was significantly better at the 90% confidence level, the mean difference was only 1% over the original data in code 4 and only 0.28% over the SMCAL data. These differences are small and of relatively no importance. Although it can not be stated that the SMCAL calibration grealy improves classification accuracy, it doesn't decrease the test performance much over the bias calibrated data and even improves the test performance over the gain calibrated original data.

Should it be necessary to gain calibrate data because there are known to be gain changes during the data run, or it is desired to standardize the gain of the data, SMCAL calibration provides a method of calibration which does not severly degrade the data.

Recommendations

It is recommended that data not be calibrated unless there are scanner gain changes (use code 4), illumination changes (use code 5), or if it is desired to standardize the gain of the data (either code 4 or 5), and generate a calibrated file with SMCAL.

There are two areas which warrant further study: 1) using the solar illumination gain calibration (codes and 2) using the "best" set of channels when classifying calibrated data.

Table 1 Results of Classification Tests

a) Run Number - 71029100 SEG. 208 41M

	ORIGINAL DATA CODE 1	ORIGINAL DATA CODE 4	SMCAL DATA CODE 4
TRAINING PERFORMANCE-%			
Corn	98.9	98.4	99.3
Other	99.8	98.7	99.9
Overall	99.7	98.7	99.8
TEST PERFORMANCE-%			
Corn	86.3	84.0	86.6
Other	99.8	98.7	99.9
Overall	88.9	86.7	89.0

b) Run Number - 71040100 SEG. 207 42M

	ORIGINAL DATA CODE 1	ORIGINAL DATA CODE 4	SMCAL DATA CODE 4
TRAINING PERFORMANCE-%			
Corn	97.1	96.8	97.2
Other	99.1	98.9	99.0
Overall	98.2	98.2	98.2
TEST PERFORMANCE-%			
Corn	81.1	83.0	80.4
Other	88.7	87.1	88.8
Overall	84.0	84.6	83.6

Table 1 Results of Classification Tests (continued)

c) Run Number - 71045700 SEG. 225 42M

	ORIGINAL DATA CODE 1	ORIGINAL DATA CODE 4	SMCAL DATA CODE 4
TRAINING PERFORMANCE-%			
Corn	100.0	100.0	100.0
Other	99.7	98.1	99.7
Overall	99.7	98.5	99.8
TEST PERFORMANCE-%			
Corn	86.1	95.7	83.3
Other	89.3	77.4	89.7
Overall	87.9	84.8	87.3

d) Run Number - 7107800 SEG. 209 46M

	ORIGINAL DATA CODE 1	ORIGINAL DATA CODE 4	SMCAL DATA CODE 4
TRAINING PERFORMANCE-%			
Corn	99.0	99.0	99.0
Other	99.7	99.7	99.8
Overall	99.6	99.6	99.6
TEST PERFORMANCE-%			
Corn	94.5	94.9	94.8
Other	98.4	98.4	98.4
Overall	96.9	97.1	97.1

Table 1 Results of Classification Tests (continued)

e) Run Number 71078400 SEG. 217 46M

	ORIGINAL DATA CODE 1	ORIGINAL DATA CODE 4	SMCAL DATA CODE 4
TRAINING PERFORMANCE-%			
Corn	100.0	100.0	100.0
Other	100.0	99.8	100.0
Overall	100.0	99.9	100.0
TEST PERFORMANCE-%			
Corn	91.4	91.7	91.3
Other	87.8	86.3	87.3
Overall	88.7	87.6	88.2

f) Run Number - 71078600 SEG. 219 46M

	ORIGINAL DATA CODE 1	ORIGINAL DATA CODE 4	SMCAL DATA CODE 4
TRAINING PERFORMANCE-%			
Corn	99.6	99.5	99.7
Other	99.8	99.8	99.6
Overall	99.7	99.6	99.7
TEST PERFORMANCE-%			
Corn	87.1	86.9	86.9
Other	93.8	93.1	93.3
Overall	93.5	92.8	93.0

Table 2 Statistical Information from the Classification Tests

COMPARISON		MEAN DIFFERENCE	STANDARD DEVIATION	+ VALUE (5 D.F.)
		%	%	%
ORIGINAL CODE 4	ORIGINAL CODE 1	-1.05	1.41	-1.82*
SMCAL CODE 4	ORIGINAL CODE 1	-0.28	0.34	-2.02*
SMCAL CODE 4	ORIGINAL CODE 4	0.77	1.37	1.37

* - Significant at the 90% confidence level

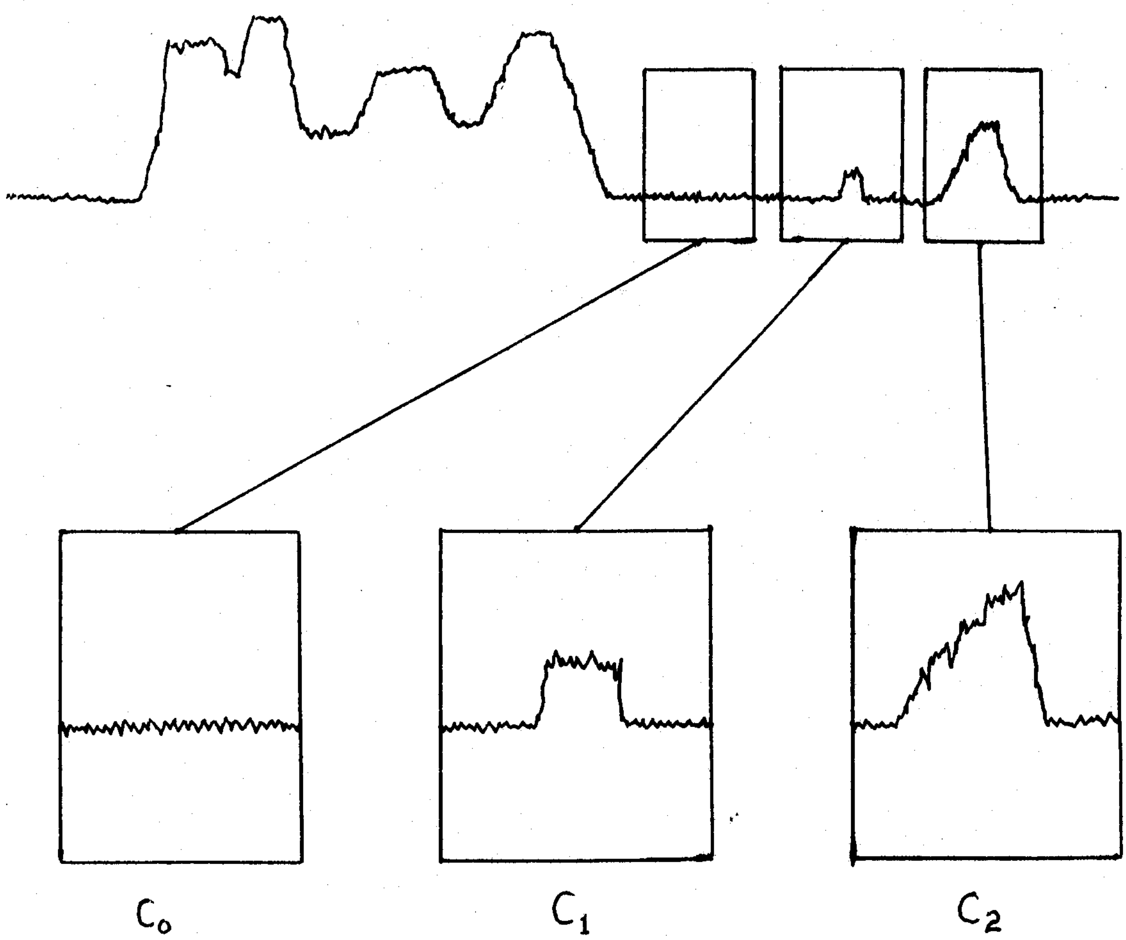


Figure 1. Sketch of typical line of data for one channel on an analog tape, showing the shape of the black level pulse (C_0), the calibrated lamp pulse (C_1), and the solar illumination pulse (C_2).

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

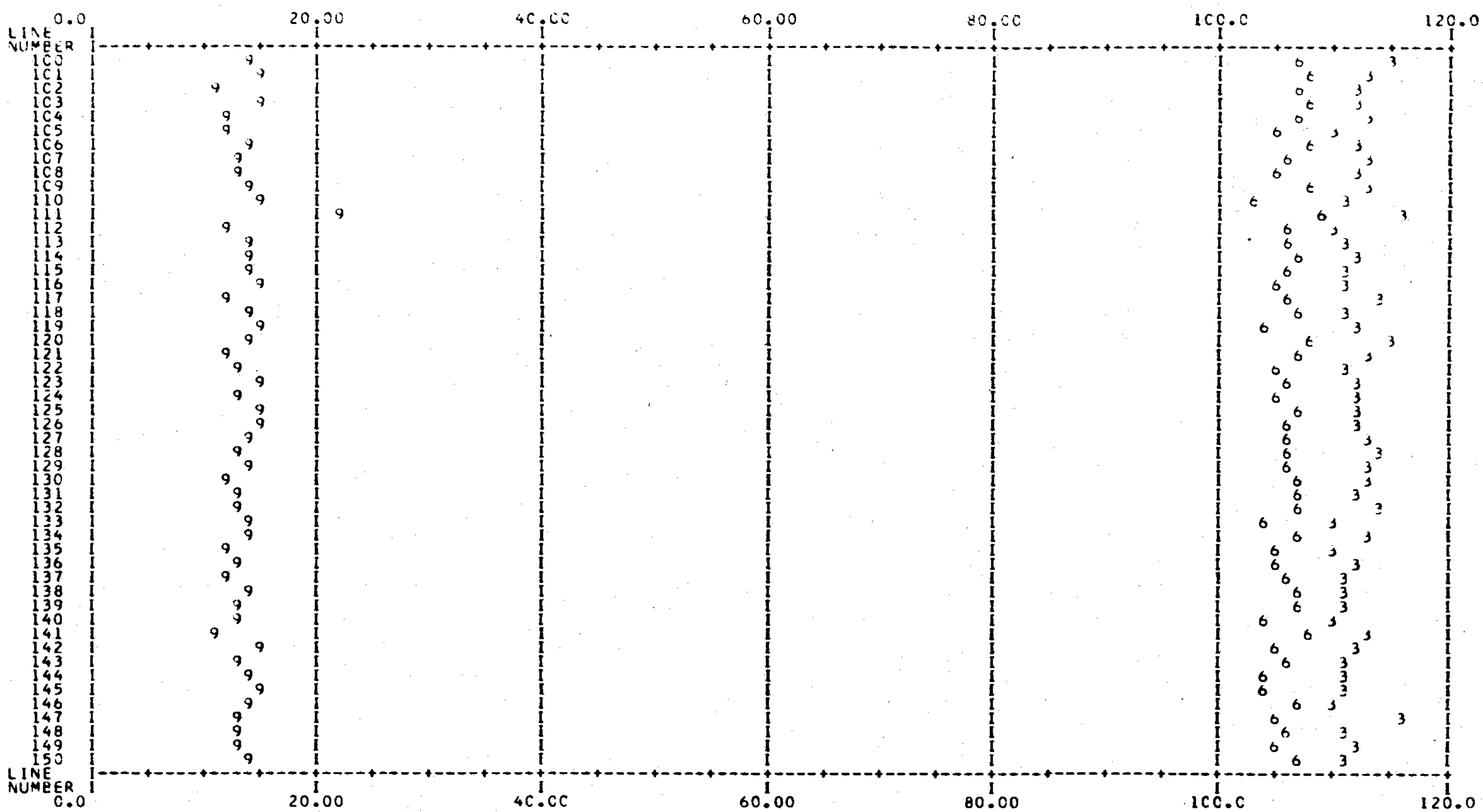
MAY 24,1972
12 11 45 PM

CORN BLIGHT SEGMENT 225 MISSION 42

TAPE NUMBER.............. 515	FILE NUMBER................ 3	RUN NUMBER........... 71045700
CONTINUATION CODE........... 0	NUMBER OF DATA CHANNELS.... 12	NUMBER OF DATA SAMPLES... 228
FLIGHT LINE.. CRN BLT LO FL225	DATE DATA TAKEN...... 7/27/71	TIME DATA TAKEN.... 1019 HOURS
AIRCRAFT ALTITUDE. 5000 FEET	GROUND HEADING.... 180 DEGREES	REFORMATTING DATE.JULY 29,1971

CHANNEL	SPECTRAL BAND LOWER	SPECTRAL BAND UPPER	CALIBRATION PULSE VALUES C0	C1	C2
1	0.46	0.49	26.70	153.3	174.5
2	0.48	0.51	25.30	153.6	165.3
3	0.50	0.54	25.50	136.7	160.1
4	0.52	0.57	25.00	125.3	148.1
5	0.54	0.60	26.65	127.9	141.8
6	0.58	0.65	22.05	127.3	127.5
7	0.61	0.70	24.00	112.0	127.3
8	0.72	0.92	23.00	121.7	66.35
9	1.00	1.40	19.65	32.50	67.65
10	1.50	1.80	30.65	76.60	116.5
11	2.00	2.60	32.00	90.75	113.4
12	9.30	11.70	22.05	207.3	0.0

Figure 2. LARSPLAY $IDPRINT of run number 71045700 showing the averages over the first 200 lines of data of the calibration pulses.



$C_1 - C_0$

Figure 3. LARSPLAY $GCOL of the values of calibrated lamp level minus black level (C_1-C_0) for lines 100 to 150 of run number 71045700, channels 3, 6, and 9.

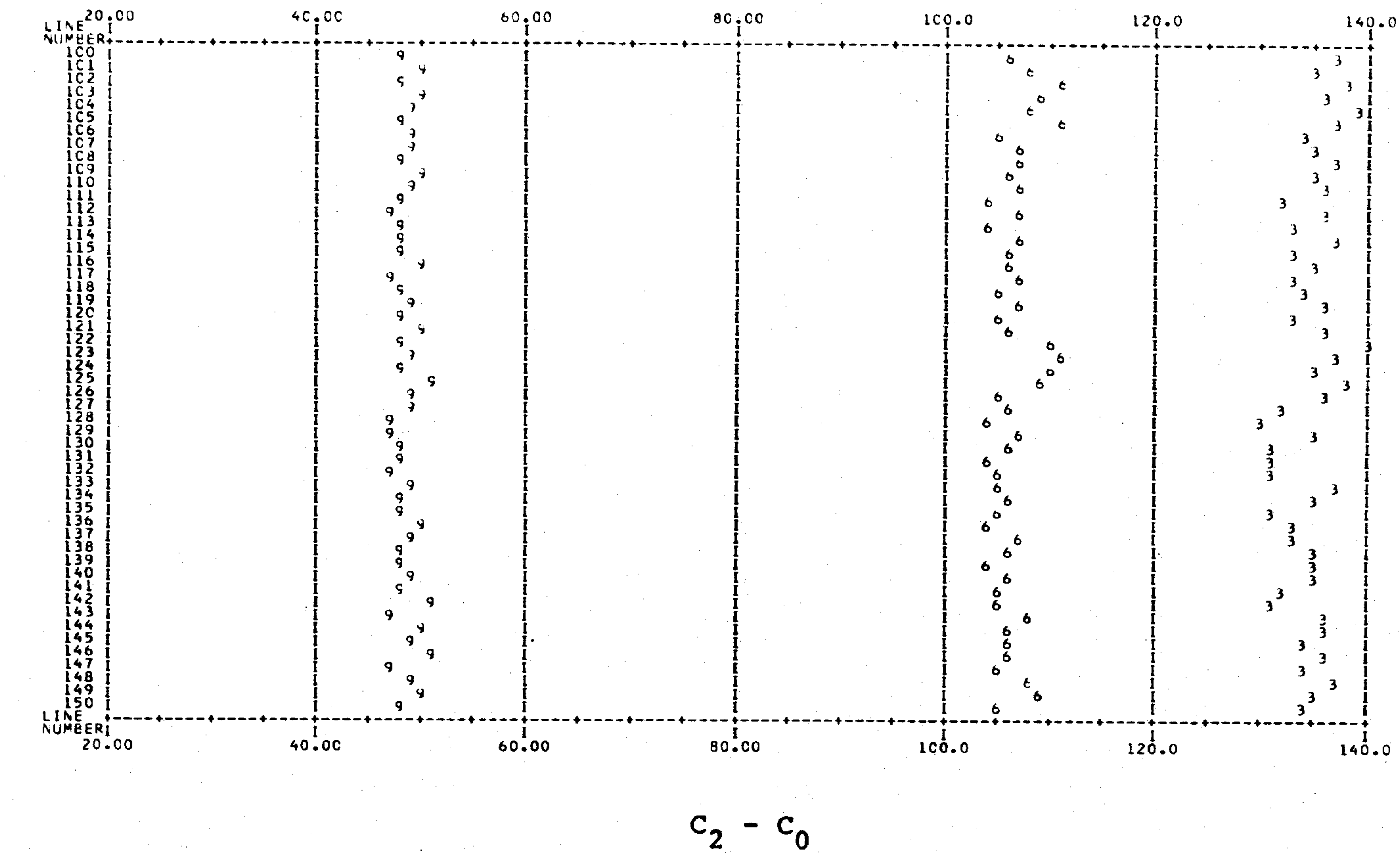
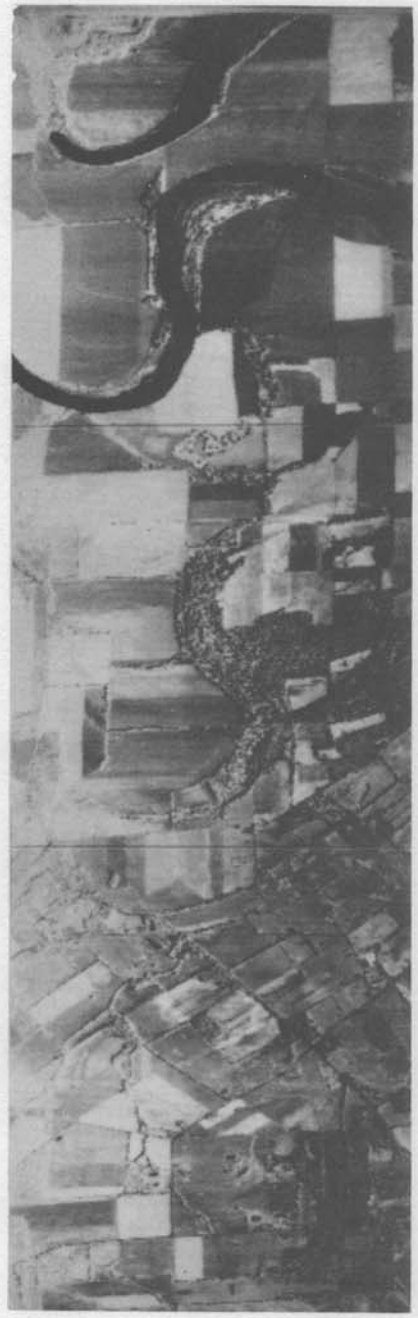


Figure 4. LARSPLAY $GCOL of the values of solar illumination level minus black level (C_2-C_0) for lines 100 to 150 of run number 71045700, channels 3, 6, and 9.

5a - LARSYS Code 4



5b - SMCAL Code 4

Figure 5. Digital display photographs of LARS run number 71045700, channel 9, lines 1 - 800. 5a is the result of using LARSYS code 4 (line by line C_1-C_0 calibration). 5b is the result of a SMCAL generated file using code 4 with smoothed coefficients.