R0-10010

Lockheed Electronics Company, Inc.

A SUBSIDIARY OF LOCKHEED CORPORATION
1830 NASA Road 1, Houston, Texas 77058
Tel. 713-333-5411

NASA CR-160315
JSC-16020
AUG 21 1979

Ref: 642-7714
Contract NAS 9-15800
Job Order 73-705

TECHNICAL MEMORANDUM

EXPERIMENT PLAN FOR EVALUATION OF LANDSAT AGRONOMIC VARIABLES USING WHEAT INTENSIVE TEST SITES

By

D. E. Phinney

(E80-10010) LARGE AREA CROP INVENTORY EXPERIMENT (LACIE). EXPERIMENT PLAN FOR EVALUATION OF LANDSAT AGRONOMIC VARIABLES USING WHEAT INTENSIVE TEST SITES (Lockheed Electronics Co.) 27 p HC A03/MF A01 N80-13589
Unclas
G3/43 00010

Approved By: J. E. Wainwright, Manager
EO Development and
Evaluation Department

August 1979

LEC-13632

CONTENTS

Section	Page

PRECEDING PAGE BLANK NOT FILMED

TABLES

FIGURE

1. INTRODUCTION

The development of spectral crop growth models and research into spectral indications of agronomic variables governing plant yield require knowledge of the fundamental informational content of spectral data. An extensive set of coincident spectral and ground-truth observations was obtained through reduction of data collected at intensive test sites (ITSs) of the Large Area Crop Inventory Experiment (LACIE). Although the ITS data base is specific for wheat, it seems likely that many valid generalizations to other crop types may be made from an analysis of this data set.

The purpose of this experiment plan is to outline a systematic analysis which will demonstrate the nature of the variations in spectral response for possible use in interpreting and predicting significant variation in crop development and yield. Of particular interest are indications of separability of development stages and agronomically significant yield parameters. The influence of soil background, varying agronomic practices, and recent meteorological events will be evaluated.

The general approach will include several steps. The inital phase will be aimed at making a choice between alternative preprocessing options and data transforms. The main part of the study will be an evaluation of separability of specific ground condition classes using spectral information. During the final phase, attempts will be made to construct trial predictive models and to evaluate the models. During all phases of the investigation, individual case studies will be conducted to evaluate anomalous situations.

The following sections will discuss the characteristics of the data, spectral data processing, and the available analysis tools. The final section proposes an analysis methodology for the study and provides a schedule.

2. CHARACTERISTICS OF THE DATA

The source of data for this study is the ITS spectral and meteorological data base compiled by Trenchard et al. (ref. 1) from ground and Land Satellite (Landsat) data collected during the 1975 to 1977 crop years as part of LACIE. For this study, the raw data set was reduced by selecting only those cases with same-day ground and satellite observations. The resulting data set contains 1539 observations. Data were available for 25 different segments. Observations during the 1976 crop year dominated (1024), with crop years 1975 (246) and 1977 (259) accounting for the remaining third. Table 2-1 gives a breakdown of the data set by segment, location, number of wheat fields (spring and winter), the number of acquisitions, and harvest year.

Sections 2.1, 2.2, and 2.3 discuss the characteristics of the ground truth, spectral data, and meteorological data, respectively.

2.1 GROUND-TRUTH DATA

Two types of ground observations for each designated field were recorded. The first type of data, obtained from inventories at the beginning and the end of the growing season, was recorded on the ITS inventory form. Included in these observations is information on acreage, land use, cultural practices, planting date, harvest date, and yield. These observations are available in machine-processable form. Additional information in the form of written comments includes seeding rates, row orientation, and fertilizer amounts. The second type of data, obtained from periodic observations throughout the growing season scheduled to be coincident with the overpass of Landsat, was recorded on the ground-truth periodic observation form. These data include observations on growth stage, ground cover, surface moisture conditions, weed growth, field operations, growth and yield detractants, and stand quality. These observations are available in machine-processable form with miscellaneous comments in written form.

Sections 2.1.1 through 2.1.12 detail the characteristics of the parameters selected for analysis in this study.

TABLE 2-1.— SUMMARY OF DATA DISTRIBUTION IN THE ITS DATA BASE

Segment	Location	Number of wheat fields		Number of acquisitions	Harvest year
		Spring	Winter		
1687	Hand (1), S. Dak.	9	22	4	1976
1687	Hand, S. Dak.	10	6	2	1977
1960	Finney, Kans.	0	13	1	1975
1961	Morton, Kans.	0	10	1	1975
1961	Morton, Kans.	0	5	3	1976
1962	Saline, Kans.	0	15	1	1975
1962	Saline, Kans.	0	32	6	1976
1962	Saline, Kans.	0	19	4	1977
1963	Rice, Kans.	0	9	1	1975
1963	Rice, Kans.	0	7	3	1976
1963	Rice, Kans.	0	8	2	1977
1964	Ellis, Kans.	0	12	1	1975
1965	Burke, N. Dak.	8	0	2	1975
1965	Burke, N. Dak.	8	0	2	1976
1966	Williams, N. Dak.	32	0	2	1976
1967	Divide, N. Dak.	5	0	6	1976
1970	Liberty, Mont.	4	2	4	1976
1971	Hill, Mont.	0	17	5	1976
1972	Whitman (1), Wash.	1	8	1	1975
1972	Whitman (1), Wash.	6	14	1	1976
1973	Whitman (2), Wash.		9	1	1975
1974	Whitman (3), Wash.	3	16	3	1976

TABLE 2-1.– Concluded.

Segment	Location	Number of wheat fields		Number of acquisitions	Harvest year
		Spring	Winter		
1975	Oneida, Idaho	7	7	4	1975
1975	Oneida, Idaho	8	7	6	1976
1977	Bannock, Idaho	4	10	2	1975
1977	Bannock, Idaho	8	7	4	1976
1978	Randall, Tex.	0	14	1	1975
1978	Randall, Tex.	0	11	4	1976
1978	Randall, Tex.	0	4	2	1977
1979	Deaf Smith, Tex.	0	2	1	1975
1979	Deaf Smith, Tex.	0	7	2	1976
1980	Oldham, Tex.	0	4	2	1977
1982	Madison, Ind.	0	4	1	1975
1982	Madison, Ind.	0	2	3	1976
1983	Boone, Ind.	0	2	1	1977
1986	Hand (2), S. Dak.	7	0	1	1975
1986	Hand (2), S. Dak.	4	1	4	1977
1987	Polk, Minn.	45	0	1	1975
1987	Polk, Minn.	12	0	3	1976
1987	Polk, Minn.	11	0	3	1977
1988	Finney, Kans.	0	34	3	1976
1988	Finney, Kans.	0	16	4	1977

2.1.1 LAND-USE CODE

The land-use code gives the type of wheat planted (spring, including durum, or winter) and in some cases the variety. In the basic data set, there are 1045 observations on winter wheat, with 586 indicating variety. There were 475 spring wheat observations with 143 indications of variety. There were 9 missing land-use codes.

2.1.2 GROWTH STAGE

The periodic observations of growth stage were made using a generalized 11-point scale. There were 401 missing observations. Table 2-2 gives the stage-by-stage distribution of the remaining 1128 cases along with a short description of each stage. A general conversion of the growth stage data to the Robertson biometeorological time scale (BMTS) was adopted for use during LACIE. Because the conversion to the BMTS was an approximation, the growth stage observations will be left in the original form for this study.

TABLE 2-2.— DISTRIBUTION OF GROWTH STAGE OBSERVATIONS IN THE ITS DATA BASE

Growth stage	Code	Number
Not planted	1	6
Preemergence	2	3
Emerged	3	48
Tillering	4	190
Booted	5	69
Beginning to head	6	84
Headed	7	123
Beginning to ripen	8	164
Mature	9	114
Harvested	10	210
Does not apply	11	117

2.1.3 GROUND COVER

The periodic observations of the percentage of ground cover were made on a 5-point scale. There were 407 missing observations. The distribution of available cases is given in table 2-3.

TABLE 2-3.— DISTRIBUTION OF GROUND COVER OBSERVATIONS IN THE ITS DATA BASE

Range of ground cover, %	Code	Number
0 – 19	1	287
20 – 39	2	111
40 – 59	3	136
60 – 79	4	167
80 – 100	5	421

2.1.4 SURFACE MOISTURE CONDITIONS

The observations of surface moisture conditions were made on a 4-point scale. There were 400 missing observations. The distribution of available cases is given in table 2-4.

TABLE 2-4.— DISTRIBUTION OF SURFACE MOISTURE CONDITION OBSERVATIONS IN THE ITS DATA BASE

Moisture condition	Code	Number
Dry	1	778
Damp	2	203
Wet	3	129
Standing water	4	19

2.1.5 WEED GROWTH

The occurrence of weeds in the designated field was reported on a 4-point scale. There were 401 missing observations. The distribution of observations of weed growth is given in table 2-5.

TABLE 2-5.— DISTRIBUTION OF WEED GROWTH OBSERVATIONS IN THE ITS DATA BASE

Weed growth	Code	Number
Negligible	1	908
Slight	2	122
Moderate	3	80
Heavy	4	18

2.1.6 FIELD OPERATIONS

Periodic observations of field operations in progress were reported for 14 categories. The occurrence of such operations was extremely rare for the observations in the ITS data base. It appears that little practical use can be made of this parameter.

2.1.7 GROWTH AND YIELD DETRACTANTS

Growth and yield detractants were reported in 349 observations. A significant number exists for moisture stress and uneven stands. Although limited in number, those indications of disease, frost, and winterkill will be examined with interest. Table 2-6 gives the distribution of reported detractants. Note that additional information on the nature of the detractant is often given in the written comments included in the periodic observations.

2.1.8 YIELD

All fields observed in this data set have one or more estimates of field yield. Estimates are made quasi-independently by the farmer, the Agricultural Stabilization and Conservation Service, and the Federal Crop Insurance Corporation and by direct sampling. Preliminary evaluation of these data reveals no systematic bias attributable to any particular data source. That evaluation also indicated a probable variance due to sampling of 17 (bushels per acre)^2. For this study, an average of the individual estimates will be used.

TABLE 2-6.— DISTRIBUTION OF GROWTH AND YIELD DETRACTANTS REPORTED IN THE ITS DATA BASE

Detractant	Code	Number
Disease	3	9
Drought	4	97
Moisture	5	20
Wind	6	5
Frost	8	4
Pot holes	10	1
Uneven stand	11	46
Weeds	12	81
Winterkill	13	9
Other	15	77

2.1.9 STAND QUALITY

Stand quality at each observation was reported on a 5-point scale. Data were either missing or not applicable for 693 cases. Table 2-7 gives the distribution of usable observations.

TABLE 2-7.— DISTRIBUTION OF STAND QUALITY OBSERVATIONS IN THE ITS DATA BASE

Stand quality	Code	Number
Poor	1	25
Below average	2	128
Average	3	484
Above average	4	166
Excellent	5	33

2.1.10 ROW ORIENTATION

The orientation of rows has been identified as a potentially key parameter in the interpretation of spectral data in this study. Row orientation is available in the form of written comments for a number of fields, particularly in the Phase III data. The parameter will be placed in machine-processable form for inclusion in this study.

2.1.11 CULTURAL PRACTICES

The ground-truth inventories provide additional information on cultural practices which may explain variation in spectral response. These data include information on whether or not a given field was fertilized and/or irrigated. There are a total of 179 irrigated and 801 fertilized fields. The written comments on the form sometimes contain ancillary information indicating whether the previous year was fallow and what amount of fertilizer was applied.

2.1.12 PLANT HEIGHT

Data on the plant height in inches are available for most fields. The potential exists for calculating an effective canopy volume from plant height and percentage of ground cover.

2.2 SPECTRAL DATA

Spectral data are available for each designated field in two basic formats. The first format gives statistics for the entire field. Field-averaged data will be used for the major part of this study. Data are also available from the raw imagery data tapes on a pixel-by-pixel basis. The pixel-level data will be used primarily for special case studies in which the finer detail justifies the additional processing.

2.2.1 RAW SPECTRAL DATA

The spectral data consist of the mean digital counts and the standard deviations for each observation. In addition, the mean digital counts for the entire segment for each band for each acquisition are available. The solar elevation angle for each acquisition is also recorded. These parameters allow the calculation of the data preprocessing and transformation options discussed in section 3.

2.2.2 IMAGERY DATA

Imagery in the form of LACIE product 1 is available for each acquisition. These data will be used extensively in the case studies.

2.3 METEOROLOGICAL DATA

Daily meteorological data for the period January 1, 1975, to December 31, 1977, for each location are available. Maximum and minimum air temperatures and total precipitation from a nearby climatological station have been coded. Table 2-8 gives the location of each station.

3. SPECTRAL DATA PROCESSING

A large number of preprocessing and data transformation algorithms has been developed by a variety of investigators. It has been noted that a some of the proposed transforms are highly correlated and thus contain the same information. An attempt has been made to select for evaluation a set of transforms and preprocessing options which embody fundamentally different approaches to the analysis of spectral data.

3.1 SPECTRAL PREPROCESSING OPTIONS

Three basic data preprocessing steps will be considered in this study. All these algorithms are designed to remove one or more major sources of scene-to-scene variation. During the initial phases of the projected study, an analysis will be conducted to select a reduced set of preprocessing steps. The options presented in the following sections represent steps either required or suggested by the authors of the data transforms to be evaluated.

3.1.1 SUN-ANGLE CORRECTION

The Kauth-Thomas transform (ref. 2) and the transform vegetation index (TVI) (ref. 3) require the use of a sun-angle correction. All other data transforms inherently correct for sun angle by ratioing. For this study, a simple cos correction will be used with all digital data corrected to a reference solar elevation angle of 51°.

TABLE 2-8.— SOURCE OF METEOROLOGICAL OBSERVATIONS FOR SEGMENTS INCLUDED IN THE ITS DATA BASE

Segment number	County	State	Station name	Station number	Coordinates		Elevation	
					Latitude	Longitude	Meters	Feet
1687	Hand	S. Dak.	Miller	395561	44°52' N.	98°98' W.	483.718	1587
1960	Finney	Kans.	Garden City	142980	37°98' N.	100°82' W.	865.632	2840
1961	Morton	Kans.	Elkhart	142432	37°00' N.	101°91' W.	1103.376	3620
1962	Saline	Kans.	Salina	147160	38°80' N.	97°63' W.	383.134	1257
1963	Rice	Kans.	Sterling	147796	38°22' N.	98°20' W.	498.653	1636
1964	Ellis	Kans.	Russell	147046	38°87' N.	98°82' W.	568.147	1864
1965	Burke	N. Dak.	Bowbells	320961	48°80' N.	102°25' W.	596.798	1958
1966	Williams	N. Dak.	Wildrose	329400	48°63' N.	103°17' W.	691.896	2270
1967	Divide	N. Dak.	Crosby	321871	48°90' N.	103°30' W.	594.970	1952
1970	Liberty	Mont.	Joplin	244512	48°58' N.	110°76' W.	1024.128	3360
1971	Hill	Mont.	Havre	243996	48°55' N.	109°71' W.	787.603	2584
1972	Whitman	Wash.	Rosalia	457180	47°23' N.	117°37' W.	731.520	2400
1973	Whitman	Wash.	LaCrosse	454338	46°82' N.	117°88' W.	451.104	1480
1974	Whitman	Wash.	Colfax	451586	46°88' N.	117°38' W.	595.844	1955
1975	Oneida	Idaho	Malad	105544	42°20' N.	112°27' W.	1432.560	4700
1977	Bannock	Idaho	Fort Hall Indian Agency	103297	43°03' N.	112°43' W.	1359.408	4460
1978	Randall	Tex.	Canyon	411430	34°98' N.	101°93' W.	1094.232	3590
1979	Deaf Smith	Tex.	Hereford	414098	34°80' N.	102°47' W.	1170.432	3840
1980	Oldham	Tex.	Vega	419330	35°25' N.	102°42' W.	1222.248	4010
1982	Madison	Ind.	Anderson Sewage Plant	120177	40°10' N.	85°72' W.	258.166	847
1983	Boone	Ind.	Whitestown	129557	40°00' N.	86°33' W.	249.631	819
1986	Hand	S. Dak.	Miller	395561	44°52' N.	98°98' W.	483.718	1587
1987	Polk	Minn.	Crookston NW Experiment Station	211891	47°80' N.	96°62' W.	269.138	883
1988	Finney	Kans.	Garden City	142980	37°98' N.	100°82' W.	865.632	2840

Let X_i represent the Landsat signal in channel i. The sun-angle correction, X_i', is calculated as follows:

$$X_i' = \frac{\cos \Theta_0}{\cos \Theta} X_i,$$

where Θ is the solar elevation angle and Θ_0 is the reference solar elevation angle ($\Theta_0 = 51°$). The resulting data will appear to have been aacquired at the reference solar elevation angle.

3.1.2 XSTAR HAZE CORRECTION

The XSTAR haze correction algorithm (ref. 2) developed at the Environmental Research Institute of Michigan has been suggested as a preprocessing step valid for all data transforms. The XSTAR haze correction is applied after a sun-angle correction. Let $\hat{X}_i$ be a scene diagnostic signal for the *ith* Landsat channel. Let α_i and X_i^* be coefficients. The change in optical thickness, γ, from the reference condition. Let Y^* be a reference yellow value for the scene.

$$\alpha = \begin{pmatrix} 1.2680 \\ 1.0445 \\ .9142 \\ .7734 \end{pmatrix} \quad X^* = \begin{pmatrix} 61.9 \\ 66.2 \\ 83.2 \\ 33.9 \end{pmatrix}$$

$$Y^* = -11.2082$$

Calculate the following:

$$a = \sum_{i=1}^{4} \alpha_i^2 (\hat{X}_i - X_i^*) R_i$$

$$b = \sum_{i=1}^{4} \alpha_i (\hat{X}_i - X_i^*) R_i$$

$$c = \left(\sum_{i=1}^{4} \hat{X}_i R_i \right) - Y^*$$

$$\gamma = \frac{-b}{a}\left\{1 - \sqrt{1 - \frac{2ac}{b^2}}\right\}$$

where

$$R_i = \begin{pmatrix} -0.89952 \\ .42830 \\ .07592 \\ -.04080 \end{pmatrix}$$

The corrected value, X_i', is calculated from the sun-angle-corrected channel value, X_i, as follows.

$$X_i' = e^{\alpha_i \gamma}(X_i - X_i^*) + X_i^*$$

The utility of this algorithm will be evaluated during the initial phases of this study.

3.1.3 CATE SCENE STANDARDIZATION

The use of the scene's mean as a standardization procedure is required for the evaluation of the Cate transform. It has been suggested that this procedure will enhance the information content of channel ratios and transformations based on channel ratios. Let X_i be the Landsat signal for the field (or pixel) of interest and $\bar{X}_i$ be the scene mean for the *ith* channel. The standardized value, X_i', is calculated as follows.

$$X_i' = 5\,\frac{X_i}{\bar{X}_i}$$

The standardized data are rescaled to have a 0 to 10 range. The desirability of using this preprocessing option with channel ratios will be investigated in the initial phases of this study.

3.2 DATA TRANSFORMS

Three basic data transforms and a set of channel ratios will be examined during this study. The specific transforms are discussed in the following sections.

3.2.1 TRANSFORM VEGETATION INDEX

The TVI developed by Rouse et al. (ref. 3) has been used successfully as an indicator of plant biomass. The TVI may be calculated using either Landsat channels 4 and 2 (TVI7) or channels 3 and 2 (TVI6). Let X_i represent the Landsat signal in the *ith* channel after sun-angle correction.

$$TVI6 = \sqrt{\frac{X_3 - X_2}{X_3 + X_2} + 0.5}$$

$$TVI7 = \sqrt{\frac{X_4 - X_2}{X_4 + X_2} + 0.5}$$

Both formulations will be examined during this study.

3.2.2 KAUTH-THOMAS TRANSFORM

The Kauth-Thomas transform converts four-channel Landsat data into an orthogonal two-dimensional feature space. The "brightness" dimension (SBI) lies along a line of soils. The greenness axis (GVI) is perpendicular to the brightness and seems to be an indication of plant biomass. This transform maintains that variations in spectral response may be adequately characterized by two dimensions.

The Kauth-Thomas transform requires sun-angle correction. Let X_i be the Landsat signal in the *ith* channel after sun-angle correction.

$$SBI = 0.433X_1 + 0.632X_2 + 0.586X_3 + 0.642X_4$$

$$GVI = -0.290X_1 - 0.562X_2 + 0.600X_3 + 0.491X_4$$

It has been suggested that the XSTAR haze improves the performance of this transform. The desirability of the XSTAR correction will be determined during the initial phase of this study.

3.2.3 CATE TRANSFORM

The Cate transform has a required preprocessing step and two forms of data presentation. The preprocessing step is one of normalization based on the scene's mean and transformation into a relative energy form (section 3.1.3). The standardized Landsat signal, X_i, for the *ith* channel may be utilized in two ways. The first presentation ranks the channel values in desending order. The channel rankings of the data constitute a signature classification which have been used to characterize crop stage and stand heterogeneity. The values have also been used to define a three-dimensional space (channels 1, 2, and 4). The location in this three-dimensional space is given in terms of its false-color value, chroma, and hue.

$$\text{Value} = \frac{1}{3}(X_1 + X_2 + X_4)$$

$$\text{Chroma} = a \sin B$$

$$\text{Hue} = \cos^{-1}\left[\frac{2X_4 - X_1 - X_2}{\sqrt{2}\sqrt{(X_1 - X_2)^2 + (X_4 - X_1)^2 + (X_2 - X_4)^2}}\right]$$

where

$$a = \sqrt{X_1^2 + X_2^2 + X_4^2}$$

$$b = \sqrt{(10 - X_1)^2 + (10 - X_2)^2 + (10 - X_4)^2}$$

$$c = 10\sqrt{3}$$

$$B = \cos^{-1}\left(\frac{c^2 + a^2 - b^2}{2\,ac}\right)$$

The channel ranking and the value-chroma-hue triplet will be examined during this study.

4. ANALYSIS TOOLS

A number of basic analytical and statistical procedures will be utilized throughout this study. The following sections identify some, but not necessarily all, of these procedures.

4.1 PLOTS

Graphical analysis in the form of computer-generated plots will be used extensively. It is anticipated that plots will be used to examine the variability of spectral response as a function of time given a similar set of surface conditions and, at a given time, given varying surface conditions. Time, in the sense used here, may indicate either calendar or biological time.

4.2 SIMPLE STATISTICS

Standard statistical software will be used to generate simple statistics such as means, variances, and correlations between spectral and ground-truth variables. The software selected has the capability to delete observations from the analysis on the basis of Boolean logic applied to the data set. For example, it would be possible to examine the correlation between growth stage for those cases at heading, given that the percentage of ground cover was between 20 and 39 percent without extensive editing of the basic data set. The software has been modified to generate all of the preprocessing options and data transforms discussed in section 3.

4.3 MODEL DEVELOPMENT

Two different statistical procedures will be used in the development of predictive models during this study. When the dependent variable is continuous (e.g., yield or plant height), the law-of-the-minimum algorithm developed by H. O. Hartley will be employed. When the dependent variable is not continuous (e.g., surface moisture conditions), a linear discriminant analysis program will be used for model development.

5. ANALYSIS METHODOLOGY

The analysis program proposed for this study will consist of three basic phases. Prior to the first phase will be a short data verification step in which simple consistency checks will be applied. The first phase will consist of selecting preprocessing options and identifying redundant data transforms. The second phase will consist of conducting a stratified correlation study in which selected data transforms will be evaluated for use in separating and predicting specific agronomic variables. The third phase will consist of developing and evaluating baseline predictive models for selected agronomic variables. Concurrent with each phase will be individual case studies for anomalous situations.

The following sections provide details on the projected analysis. Figure 5-1 is the schedule for this study. A review will be held at the end of each phase of the study.

5.1 VERIFICATION OF DATA AND SELECTION OF TEST DATA

The data will be examined for obvious errors. Coding errors and possible cloud contamination in the spectral data will be screened out of the data set. The screening procedure will use simple histograms, visual scanning of the data, and an examination of imagery products.

When data verification is completed, a set of test data will be selected. The test data will consist of approximately 500 observations. The test cases will be selected from those observations with growth stage data. Approximately one-half of the observations in each stage will be designated test data. Histograms of the development and test data sets will be compared for other important ground-truth variables to ensure that reasonable distributions of the agronomic variables have been retained in each data set.

5.2 SELECTION OF PREPROCESSING OPTIONS

The first main activity of this study will be the evaluation of preprocessing options. Two primary questions will be examined in this phase. First, a

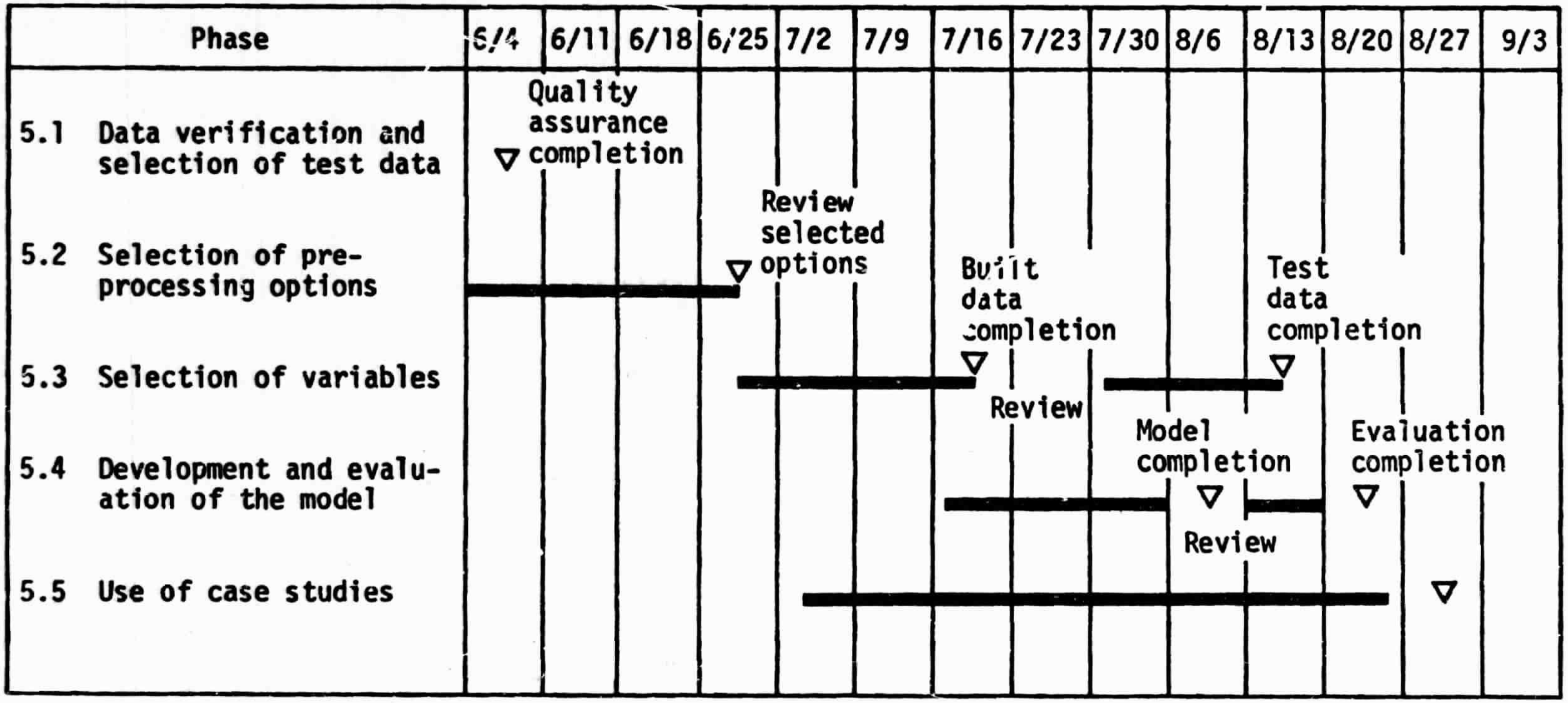

Figure 5-1.— Schedule for the evaluation of Landsat agronomic variables using wheat ITS's.

decision will be made on the use of the XSTAR haze correction as a general preprocessing step for all data transforms. Second, the use of the Cate scene standardization will be evaluated as a preprocessing step for channel ratios.

A limited set of data will be used to conduct these evaluations. The correlations between transformed spectral data for fields beginning to head (growth stage 6) and yield will constitute one decision criterion. A second criterion will be the separability between fields in growth stages 5 and 6 and/or 7 and 8 as indicated by differences in the means and the standard deviations of the transformed data.

The transforms described in section 3.2 will be calculated with and without the indicated preprocessing step for the development data set. Any conclusions drawn about the apparent relative performance of the techniques are to be documented and tested by applying the techniques to the independent test data set. The degree to which the initial conclusions are supported by test results are to be analyzed and documented.

The capability of the preprocessing and transformation techniques to normalize data from time to time and place to place and to enhance vegetation relative to the scene will be tested separately. The results will be evaluated for improvements in site-to-site variability for vegetation indicated by ground truth to be of comparable stage and quality and in ability to enhance strong vegetation signatures in sites with varying degrees of overall vegetation.

The results of this phase will be a set of recommended processing procedures for the stratified correlation and separability studies described in section 5.3.

5.3 SELECTION OF VARIABLES

The second phase of this study will be an in-depth evaluation of the potential information content of spectral data for specific agronomic variables. Growth stage, ground cover, surface moisture conditions, stand quality, and yield will be examined. The influence of weed growth, growth and yield detractants, and row orientation on the correlation and separability of the primary variables will be examined. The primary intent of this phase of the study is to gain an understanding of the sources of "noise" in the agronomic "signal."

Extensive use will be made of stratification of the data to isolate potential sources of variation. For example, in the evaluation of various data transforms as direct indicators of yield, stratification by growth stage is indicated. Further stratification within growth stage by percentage of ground cover would give an indication of the importance of varying ground cover on the spectral signal. By thoughtful use of data stratification, one should be able to identify the more desirable data transforms at the same time that major sources of data variability are being evaluated. Table 5-1 gives some of the primary and secondary stratifications being considered.

TABLE 5-1.— SUMMARY OF POTENTIAL DATA STRATIFICATIONS FOR THE ITS DATA BASE

[P = primary; S = secondary]

Stratify by	Variable of primary interest				
	Yield	Stage	Moisture	Cover	Quality
Yield					
Growth stage	P				P
Surface moisture					
Ground cover	S		P		
Stand quality	S				
Row orientation					
Weed growth					
Detractants					
Plant height					

The exact strategy used for each major variable will depend largely upon the results found as the study progresses. However, two broad groupings of the data exist as discussed in the following sections.

5.3.1 CORRELATION STUDIES

For continuous variables such as yield, simple correlation coefficients will be used as screening tools. For instance, the correlation between transform spectral data and yield would be calculated for all fields at the same growth stage. The change in correlation with time could then be plotted to show the stages of growth at which spectral data would be needed to predict yield. The influence of other variables could be evaluated by further stratifications.

5.3.2 SEPARABILITY STUDIES

For variables given in discrete classes such as growth stage, discriminant analysis will be used. The spectral data transforms will be selected using a stepwise discriminant analysis. The separability in a given case will be evaluated through the percentage of correct classification achieved using the selected spectral variables.

5.3.3 APPLICATION IN PRACTICE

Relations developed on the development set will be applied to the test set without stratification. The relations will be used in a predictive mode for yield or growth stage to determine whether the relations are meaningful and useful when stratification by ground truth is not possible. Evaluation will be based on how well the relations stand up, not on absolute predictive capability.

5.4 DEVELOPMENT AND EVALUATION OF THE MODEL

Relatively simple multivariable models will be constructed and evaluated during the final phase of the study. Modeling tools will include the law of the minimum and stepwise discriminant analysis. Models will be attempted for at least growth stage and yield. Models for other variables will be developed if the results of the second study phase are promising. All models will be tested on an independent set of data (see section 5.1).

Any relations resulting from the development data set which seem promising as models will be tested on the test data set with ground-truth information used to stratify data and without reference to the ground-truth information. The objective is to test the predictive capabilities of the models.

During the final phase, the correlation and the separability studies conducted during the second phase (section 5.3) will be repeated on the test data in order to replicate the results.

5.5 USE OF CASE STUDIES

Throughout all phases of this study, all significant anomalous situations will be examined by individual case studies. Special imagery product and reprocessing of selected fields at the pixel level will be required for most case studies. Use will be made of ancillary data sources such as detailed soil map as required by the individual anomaly.

6. REFERENCES

1. Trenchard, M. H; Sestak, M. L.; Kinsler, M. C.; and Phinney, D. E.: Composition and Assembly of a Spectral-met Data Base for Spring and Winter Wheat. LEC-13393 (JSC-14901), Vol. I, May 1979; Vol. II, June 1979.

2. Kauth, R.; Lambeck, P.; Richardson, W.; Thomas, G.; and Pentland, A.: Feature Extraction Applied to Agricultural Crops as Seen by Landsat. Paper presented at the LACIE Symposium (Houston, Texas), Oct. 1978.

3. Rouse, J. W., Jr.; Hess, R. H.; Schell, J. A.; and Deering, D. W.: Monitoring Vegetation Systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351, 1973.

END

DATE

FILMED

B 12 1980