01 STIF E7.7-10068. II

NASA-CR-149434

ERIM

FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN

(E77-10068) WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA Progress Report, 16 Aug. - 15 Nov. 1976 (Environmental Research Inst. of Michigan) 47 p HC A03/MF A01 CSCL 02E

N77-16403

Unclas G3/43 00068

WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA

TYPE II PROGRESS REPORT

16 August 1976 - 15 November 1976

114800-25-L

NASA Contract No. NAS5-22389

Prepared by

Richard F. Nalepka - Principal Investigator
John Colwell - Co-Principal Investigator
Daniel P. Rice

for

Mr. C. R. Stonesifer, NASA Technical Officer
Code 902
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt Road
Greenbelt, Maryland 20771

2062L

WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA
TYPE II PROGRESS REPORT
16 August 1976 - 15 November 1976

The following report serves as the sixth Type II Progress Report for Landsat Follow-on Investigation #2062L which is entitled "Wheat Productivity Estimates Using Landsat Data".

This investigation has several objectives, including the following:

1. to develop techniques and procedures for using Landsat data to estimate characteristics of wheat canopies which are correlated with potential wheat grain yield
2. to demonstrate the usefulness of Landsat data for estimation of wheat yield
    a. for irrigated and for non-irrigated LACIE (Large Area Crop Inventory Experiment) intensive test sites
    b. for two different years with varying weather conditions.

1.0 PROBLEMS

No significant problems were met during this reporting period.

2.0 ACCOMPLISHMENTS AND RESULTS

In this section, we discuss the technical accomplishments during this reporting period. Included are subsections discussing:

1. relative utility of Landsat, meteorological, and ancillary data
2. analysis of newly processed Landsat data
3. indicators of wheat green development
4. extension of yield prediction over time and space
5. normalization of soil reflectance variability
6. utilization of leaf area duration information.

2.1 RELATIVE UTILITY OF LANDSAT, METEOROLOGICAL, AND ANCILLARY DATA

One of the goals of this project is to assess the relative utility of Landsat data, meteorological data, and some combination of Landsat, meteorological, and ancillary data for predicting wheat yield. This section discusses analyses carried out to address the above goal and some resulting conclusions.

2.1.1 Landsat Data vs Meteorological Data

Our analysis tends to substantiate our earlier hypothesis that there is important wheat estimation information contained in Landsat data that is not provided by standard meteorological data. The differences in meteorological conditions (particularly temperature and precipitation) over a 30 sqaure mile site are generally not substantial. For example, for the 30 rain gauge stations on a test site in Finney County, Kansas in 1975, precipitation was relatively constant over the entire site. During the important growing months of May and June the coefficient of variation (σ/m) in precipitation between rain gauge stations was only about 0.10 (see Table 1).

TABLE 1. RAINFALL DATA FOR 30 RAIN GAUGE STATIONS AT FINNEY SITE, 1975

Month	Mean Value Of Rainfall	Standard Deviation	Coefficient Of Variation (σ/m)
May	3.76	0.43	0.11
June	2.85	0.27	0.09

Despite the relative constancy of important meteorological conditions, the yield on the 1975 Finney site varied substantially (21.0 bu/acre to 74.0 bu/acre) from field to field. The reasons for this variation in yield are apparently largely non-weather related. These differences in yield may more likely be related to such factors as differences in topography, soil type, planting density, fertilization, cropping practices in a field, and irrigation, none of which are accounted for by most meteorological yield models. The resulting differences in crop condition and eventual yield found in the 1975 Finney site are, to a substantial degree, manifested in Landsat data, as indicated in the results reported in this and other quarterly reports. Thus, it appears that Landsat data can better account for local variations in yield than can meteorological data.

Since yield models require as inputs certain measures of field condition, Landsat data may be a reasonable source for such inputs. For example, the Kanemasu-ET model depends on periodic estimates of leaf area index (LAI) [5], while a proposed ERIM yield prediction method [4] requires leaf area duration or percent vegetation cover duration as input variables.

To assess the utility of Landsat data several questions can be raised with regard to these variables:

1. How accurately can these variables be estimated by field personnel?
2. How well are these field estimates of variables related to yield?
3. How well can the variables be estimated using Landsat data?
4. How well are the Landsat estimates of the variables related to yield?

If we assume that the carefully made ERIM objective field measurements of percent cover are correct, we can assess how well other field personnel can make estimates of such a parameter, relative to how well Landsat data can be used to make such estimates. For the 1975 Finney Site, the May 21 ASCS (Agricultural Stabilization and Conservation Service) subjective estimates of percent cover and the ERIM field measurements of percent cover have a correlation of 0.52. Statistically this correlation is not significant at the 5% level. On the other hand, the correlation between Landsat data (Band 7/Band 5) and ERIM objective field measurements for the same fields at the same time is 0.93, which is statistically significant. Therefore, preliminary indications are that for yield models that require estimates of degree of crop vegetative development, Landsat data may furnish a better estimate than some subjective estimates made by field personnel using traditional approaches.*

It should be noted that the amount of crop vegetative development also might be estimated by use of meteorologically-based growth models. However, since meteorological conditions did not vary drastically over the Finney site, it seems highly unlikely that any meteorologically-based growth model would have predicted the large variation in vegetative percent cover that was found to exist between fields on 21 May 1975 (from 28.9% to 93.6% cover on fields which were sampled by ERIM). Therefore, using Landsat data one apparently can estimate amount of vegetative development (a possible variable in some yield models) better than growth models that are based on meteorological data.

The correlations between various estimates of field vegetative condition and <u>actual yield</u> are shown in Table 2. The ERIM objective measurements of percent green wheat cover on May 21 were significantly correlated with yield, as were measurements of green LAI and Landsat data.** However, ASCS estimates of percent cover and height were not significantly correlated with yield. It appears, therefore, that for yield models that require periodic estimates of vegetation condition that are correlated with potential yield, Landsat estimates of these inputs are as good as or better than the traditional subjective field estimates.*

---

* Traditional methods using trained field personnel can certainly be more precise than Landsat data, but the traditional methods are sufficiently time-consuming so that they cannot routinely be made on enough samples to characterize large, variable fields. The advantage of using Landsat data is that it samples the whole field.

** Various Landsat measures of green vegetative development were analyzed and are discussed in Section 2.3. The particular value used here is the square of the ratio of Landsat Bands 7 and 5.

TABLE 2. CORRELATIONS BETWEEN VARIOUS INDICATORS OF CROP CONDITION AND YIELD, 21 MAY 1975, FINNEY DATA (N=6)

Variable	Correlation
Percent Cover (ASCS)	0.601
Height (ASCS)	0.795
Green Cover (ERIM)	0.912*
Green LAI (ERIM)	0.826*
SQRAT2. ($\sqrt{7/5}$)	0.916*

* Significance at 5% level = 0.811

Other yield models require actual (subjective or objective estimates of probable yield. For example, the USDA/SRS pre-harvest yield forecasts are based on weather variables such as actual and predicted precipitation, plus field condition or probable yield as reported by farmers or other field personnel [6].

We now address the question of whether Landsat data could improve on some traditional estimates of probable yield. One way of making this comparison is to examine the correlations with yield and Landsat data and alternative methods of estimating probable yield. Such alternative methods might include stand quality ratings (made by ASCS personnel), and objective estimates of yield made by ASCS from field sampling just prior to harvest (FCIC). The available comparisons for three sites for which we have processed Landsat data are indicated in Table 3.

On the basis of the results shown in Table 3, our preliminary conclusion is that Landsat estimates** of probable yield are as good as or better than the traditional field alternatives which we examined, even when the Landsat estimates are made as much as two months before the estimates using alternative methods.

** Other selected Landsat estimates also discussed in Section 2.3.

TABLE 3. CORRELATIONS WITH YIELD OF INDIVIDUAL FIELDS
ASCS ESTIMATES AND LANDSAT DATA

	DATE, SITE						
	FINNEY, 1975		ELLIS, 1975		FINNEY, 1976		
	N=11 Date	N=11 Correlation	N=18 Date	N=18 Correlation	N=11 Date	N=11 Correlation	Average Correlation
FCIC Estimate	Pre-harvest	0.95	Pre-harvest	0.74	Pre-harvest	0.45	0.71
Stand Quality Rating	Pre-harvest	0.47	Pre-harvest	0.89		---	0.68
Landsat Predicted Yield (4 bands)	April 15	0.94	May 21	0.79	May 6	0.87	0.87
Landsat Predicted Yield (TVI)	April 15	0.93	May 21	0.64	May 6	0.77	0.78

We make the following preliminary conclusions as a result of the material presented in this section.

1. Landsat data can provide at least as good an indicator of field condition (percent cover, LAI) as can subjective field estimates for use in existing yield models.
2. Landsat indicators of probable wheat yield are as good an indicator as are subjective, and some objective, field estimates of probable yield, for use in existing yield models.
3. Therefore, Landsat data can be used as a substitute for field estimates of field condition or probable yield in wheat yield models that require such inputs.

2.1.2 Landsat data vs Ancillary Data

Many meteorological yield models do not include potentially important environmental/cultural factors which are not routinely available from local weather stations. The relative importance of some of these environmental/cultural factors and the degree to which they can be accounted for by Landsat data, is discussed in this section.

The fact that non-meteorological factors can be important determining factors of wheat yield is indicated by the following example. Precipitation data, frequently the most important input to meteorological yield models, was recorded at four fields for which yield was also determined on the 1975 Finney site. Although the yield from field to field varied greatly, from 21.0 bu/acre to 60 bu/acre, there was not a clear association between yield and precipitation for the months of May and June. Since temperature and solar irradiance were probably quite similar over the entire site, they can not explain these large differences in yield, either.

The relationships between several other environmental/cultural factors and wheat yield have been investigated for the 1975 Finney test site. The specific factors investigated were:

1. wheat variety
2. irrigation/no irrigation
3. fertilization/no fertilization
4. amount of irrigation.

An analysis of variance was performed for the above factors by regression with wheat yield for the 16 fields for which such data was available. From this analysis it was possible to determine the percent of the variance in yield accounted for separately by each of the factors, and also the percent of yield variance accounted for by Landsat data for the three dates analyzed. The site used for this analysis is a predominantly irrigated site, and the environmental/cultural factors are not entirely independent. In fact, all fields which were irrigated were also fertilized, and the converse, so these two variables were combined into a single irrigation-fertilization variable. Other factors had lesser, but non-zero correlations. The results of the analysis are presented in Table 4.

The analysis shows that there was not a large amount of yield variance accounted for by wheat variety. This is not surprising since farmers in a given location might be expected to use wheat varieties that are "best" for that location and cultural practices, and therefore the varieties should not differ appreciably from each other. The three principal wheat varieties represented in this analysis are Eagle, Scout, and Satanta. Although we do not have information on yielding ability of Satanta variety, Eagle and Scout varieties of wheat are known to have virtually identical "yielding ability" [7].

Irrigation-fertilization accounts for somewhat more of the variance in yield than wheat variety but surprisingly not a significant amount. Such cultural practices would not be economically justifiable if there were no effect on yield. The amount of variance in yield accounted for in this analysis by the irrigation-fertilization variable would have been higher (a value of 53.6%) if a field which had been deleteriously affected

TABLE 4. PERCENT OF VARIANCE IN YIELD ACCOUNTED FOR SEPARATELY BY SEVERAL ENVIRONMENTAL, CULTURAL, AND LANDSAT VARIABLES (1975 Finney data, 16 Fields)

Variable	Variance in Yield Accounted for by Variable (%)
1. Wheat Variety	18.5
2. Irrigation-Fertilization	24.9
3. Amount of Irrigation	79.9
4. Landsat Data (22 Nov 1974) [TVI]	59.8
5. Landsat Data (15 Apr 1975) [TVI]	85.9
6. Landsat Data (21 May 1975) [TVI]	75.5

by a treatment of herbicide was not included in the analysis. The total amount of irrigation (inches) applied to individual fields during the growth of the crop accounted for nearly 80% of the variance in yield. Again, this value would have been higher (85.6%) if the herbicide-treated field had not been included.

Landsat data yield indicator transforms for the three available dates were analyzed individually for their utility in predicting yield on the fields for which ancillary data was available. Yield indicator transformations (in this case TVI) for all three dates of Landsat data account for a high proportion of variance in yield.

In addition to the above analysis a coarse evaluation was made of the relative utility of ancillary variables and Landsat variables for predicting yield by determining the percent of variance in yield accounted for by several combinations of variables. The results are presented in Table 5.

In this analysis wheat variety and knowledge of whether the fields were irrigated and fertilized (variables 1 and 2) accounts for 31% of the variance in yield, but addition of information on amount of irrigation (variable 3) raises the variance accounted for to 93.5%. Since precipitation, temperature, and solar irradiance were in all likelihood essentially constant over the entire site, it is not surprising that

Table 5. PERCENT OF VARIANCE IN YIELD ACCOUNTED FOR BY SEVERAL COMBINATIONS OF ENVIRONMENTAL, CULTURAL, AND LANDSAT VARIABLES (1975 Finney data, 16 Fields)

Variables (From Table 4)	Variance in Yield Accounted for by Variables (%)
1, 2	31.0%
1, 2, 3	93.5%
1, 2, 3, 5	95.1%
4, 5, 6	85.5%
4, 5, 6, 1, 2	90.0%
4, 5, 6, 3	90.0%
1, 2, 3, 4, 5, 6	95.3%

Variable Key

1 -- Variety

2 -- Irrigation/Fertilization (Yes or No)

3 -- Irrigation Amount

4 -- TVI (22 November 1974)

5 -- TVI (15 April 1975)

6 -- TVI (21 May 1975)

three factors (variables 1-3) which do vary over the site account for as much of the variance in yield. However, the amount of irrigation, the most important of these three ancillary variables, is not a variable that is likely to be routinely available information, and hence is not a likely candidate variable for a wheat yield model.*

Without utilizing any ancillary information, the yield indicator transformations for the three Landsat data sets (variables 4-6) account for 87.5% of the variance in yield. When the two ancillary variables most likely to be available (variety and irrigation-fertilization) are included with the three Landsat data transforms, the yield variance accounted for is 90%. A similar result occurs if amount of irrigation is added to the Landsat variables. Apparently much of the variability accounted for by amount of irrigation is also accounted for by the Landsat data. If the best single date of available Landsat data (April 15) is included with the three ancillary variables, 95.1% of the variance is accounted for, while inclusion of all three Landsat data sets raises the value to 95.3%.

The foregoing discussion furnishes the basis for some preliminary conclusions regarding the relative utility of Landsat data and ancillary data for predictions of wheat yield on a predominantly irrigated site in southwestern Kansas. If data on important ancillary variables (especially amount of irrigation) is available, such data is a good indicator of wheat yield on an individual field basis, perhaps somewhat better than several dates of Landsat data (ancillary variables 1-3: 93.5%; Landsat variables 4-6: 87.5%). If both Landsat and ancillary data are simultaneously available, wheat yield prediction performance is improved only

* There is also a significant correlation between amount of irrigation and both percent cover and LAI, thus indicating that amount of irrigation is a factor that should be considered in growth models, as well as yield models.

slightly over either type alone (all variables: 95.3%). Therefore, using Landsat data alone may be an acceptable procedure. In situations where Landsat data, meteorological data, and ancillary data are available, use of some combination of this data will probably improve yield prediction performance. In such situations, the appropriate approach is probably a function of the marginal costs in increasing the complexity of a yield prediction model compared to the marginal benefits.

2.2 ANALYSIS OF NEWLY PROCESSED LANDSAT DATA

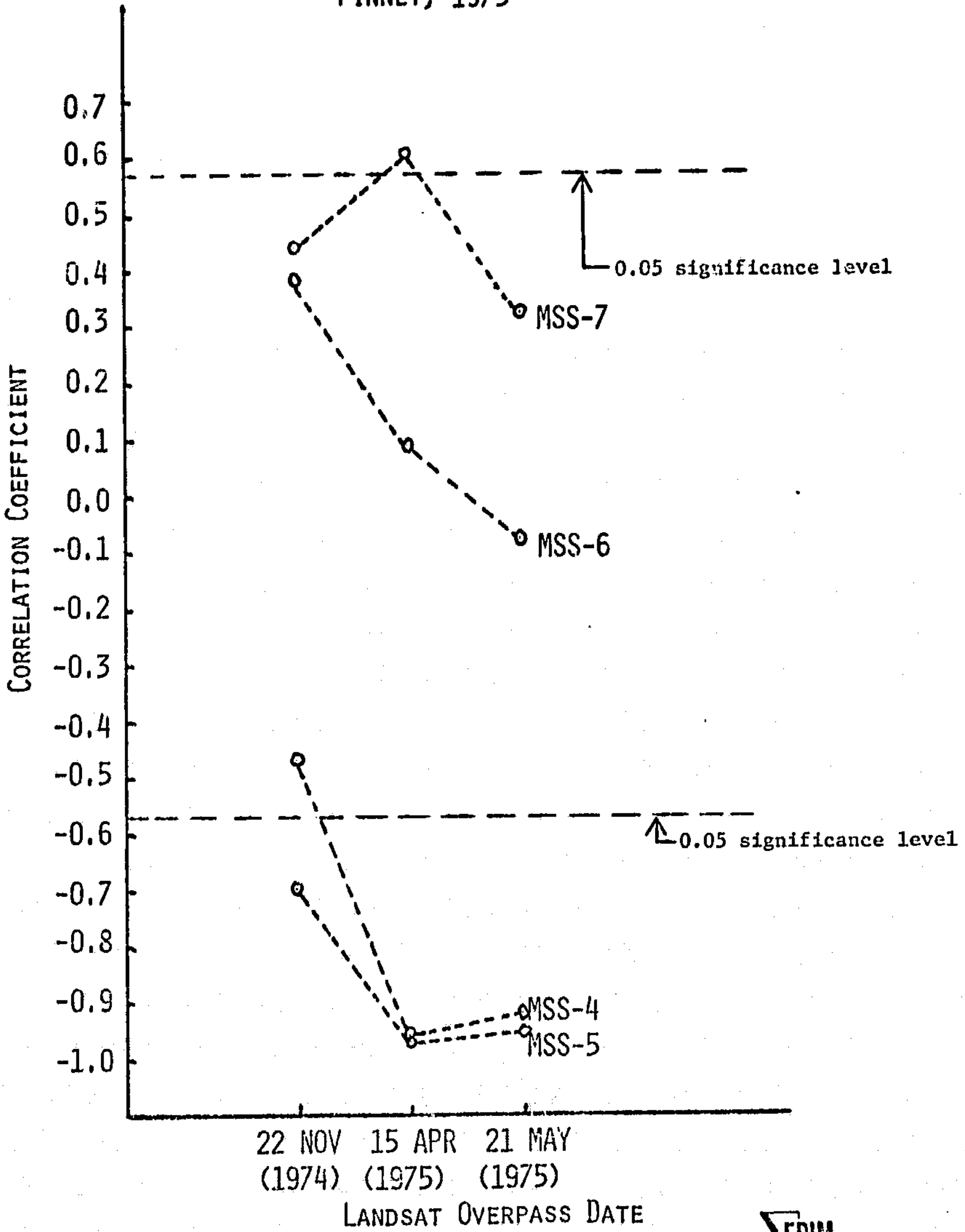
During this reporting period, Landsat data in addition to that previously available has been processed. In this section more details of the data processing is provided.

2.2.1 The 1974-75 Finney Test Site

With the addition of two data sets covering the original, largely irrigated, Finney county intensive test site, 12 Landsat spectral-temporal bands were available for analysis. The three data sets analyzed included Landsat passes on 22 November 1974, 15 April 1975, and 21 May 1975.

After mean signal values in each band were computed for each sufficiently large wheat field, the mean values were correlated with the farmer estimates of wheat grain yield in order to assess relative information content. The resulting correlations with yield as a function of time are shown in Figure 1 where the horizontal dotted lines represent the 5% level of significance (correlation values which fall between the dotted lines are not considered significant at the 5% level). It is clear that the single best spectral-temporal band for predicting yield is the 15 April red band (0.6-0.7 μm, Band 5), with the 15 April green band (0.5-0.6 μm, Band 4) a close second. April 15 is also the only one of the three dates on which Band 7 (0.8-1.1 μm) is significantly correlated with yield, although for all dates, Band 7 is more highly cor-

FIGURE 1. CORRELATION OF INDIVIDUAL LANDSAT BAND DIGITAL COUNT VALUES vs YIELD AS A FUNCTION OF DATE
FINNEY, 1975



ERIM

related with yield than Band 6 (0.7-0.8 μm), and Band 5 is more highly correlated with yield than Band 4.

The optimum spectral-temporal bands for predicting yield were then determined by stepwise regression. The result of the regression indicated that the four optimum spectral-temporal bands* came from the April 15 and November 22 Landsat data, and accounted for over 85% of the variation in yield as measured by the coefficient of determination, R^2.

In order to determine the best single date for predicting wheat grain yield using all four bands on a given date, a regression was performed between yield and the set of four bands for each date. The results are presented in Figure 2 which shows that the single best date is 15 April, where 82% of the variance in yield is accounted for. This "optimum date" differs from the result obtained for the same test on Ellis County 1975 Landsat data, where May 21 (approximately the time of heading) was the best single date. This result may be related to the fact that the wheat fields in the 1974-75 Finney site are primarily irrigated while the Ellis site wheat fields are primarily not irrigated. Irrigated fields achieve high vegetation cover earlier in the growing season than non-irrigated fields, and the ability of Landsat data to indicate differences in vegetation cover tends to decrease at higher values of vegetation cover (see February - May quarterly progress report). Therefore, the "optimum" Landsat date may be a compromise between the degree of correlation between vegetation cover and yield, and the capability of Landsat data to indicate differences in vegetation cover on a given date. For the 1974-75 Finney data set, the optimal combination of these two factors appears to occur prior to heading. However, additional Landsat data need to be investigated before we can have confidence in this hypothesis.

---

* (April 15 Bands 5, 6, 7, and November 22 Band 6)

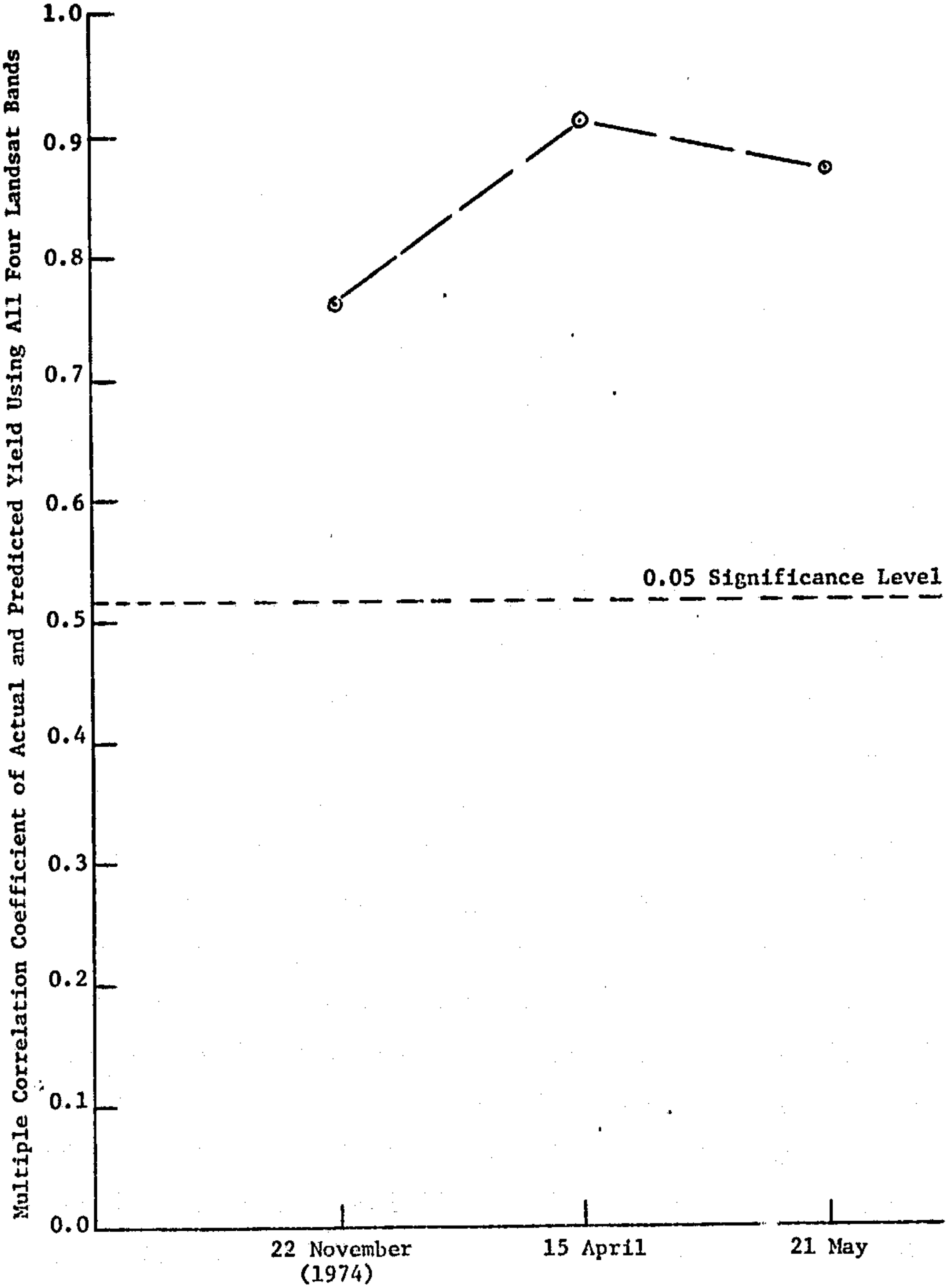


FIGURE 2. MULTIPLE CORRELATION COEFFICIENTS BETWEEN PREDICTED AND ACTUAL YIELD USING THE SET OF FOUR LANDSAT BANDS FOR EACH OF THREE DATES. (1975 Finney Site, Kansas)

2.2.2 The 1975-76 Finney Test Site

For the 1975-76 growing season, the Finney site was moved to a largely non-irrigated area and some of the yield data was received during this reporting period. Landsat data from May 6 was processed to obtain field mean signature values and, as shown in Figure 3, the data was found to be highly correlated with yield.

The Finney 1976 data provided a new situation not encountered in other data sets. A few fields or portions of fields contained stands of wheat sufficiently poor that the farmer decided to plow up all or portions of the field. A result of this situation is that an opportunity was available to determine if the Landsat data could indicate which fields are not likely to be harvested, an important consideration for early forecasts of total production. It was necessary for this analysis, to decide whether to predict yield/planted acre or yield/harvested acre.

In the long term, our interest is in contributing to a wheat <u>production</u> forecasting system, in which wheat area estimates and wheat yield estimates are aggregated to form production estimates. In such a system, the yield estimates must be in units of yield per planted acre if the area estimates are planted acres, or must be in units of yield per harvested acre if area estimates are harvested acres. In the short run, due to the difficulty of continually updating unplowed (or harvestable) acreage and redefining field boundaries for determination of yield, it was decided to work with yield/planted acre. Since the initial yield data were reported as yield/harvested acre, we obtained actual production data (bu) for the individual fields, and recomputed yield on a planted acre basis for the affected fields.

The definition of poor stands of wheat which are not likely to be harvested is associated with cultural/economic factors in addition to the remotely sensible factors. A farmer decides to plow and replant a marginal wheat field based on the expected market value of wheat and possible replacement crops, crop insurance payments, etc., as well as the

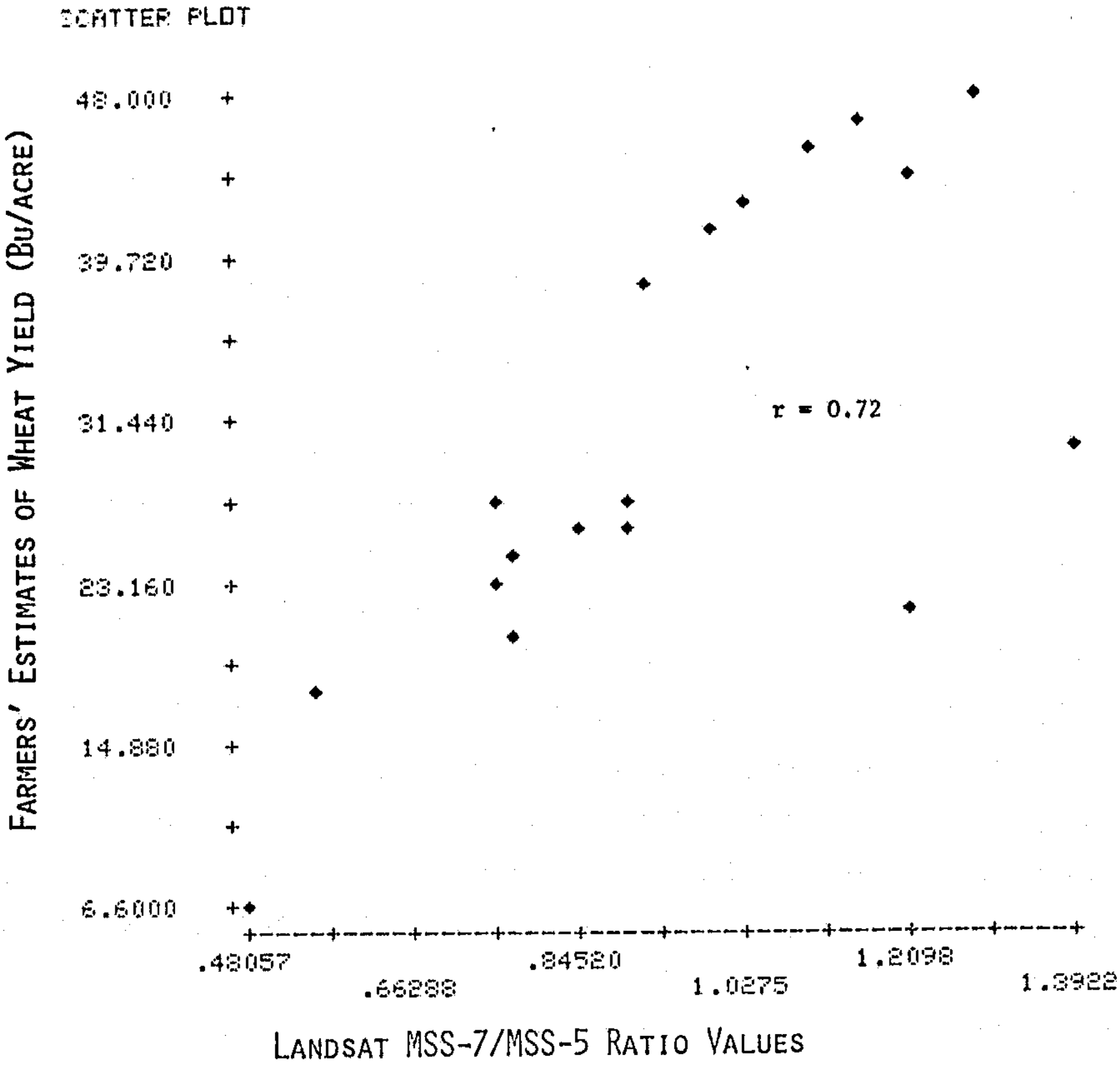


FIGURE 3. SCATTERPLOT OF WHEAT YIELD vs LANDSAT MSS-7/MSS-5 RATIO. (6 MAY 1976, FINNEY)

ERIM

farmer's expectation of the wheat's productivity. Thus, the critical level of stand quality below which a farmer would plow up his stand of wheat may vary from year to year or even from farmer to farmer. Our experience to date suggests that Landsat data can be used to differentiate between fields of varying stand quality, at low values of stand quality. But in order to estimate which fields are not likely to be harvested, the critical level of stand quality or probable yield must be determined from a source other than remote sensing.

2.3 INDICATORS OF GREEN WHEAT DEVELOPMENT AND YIELD

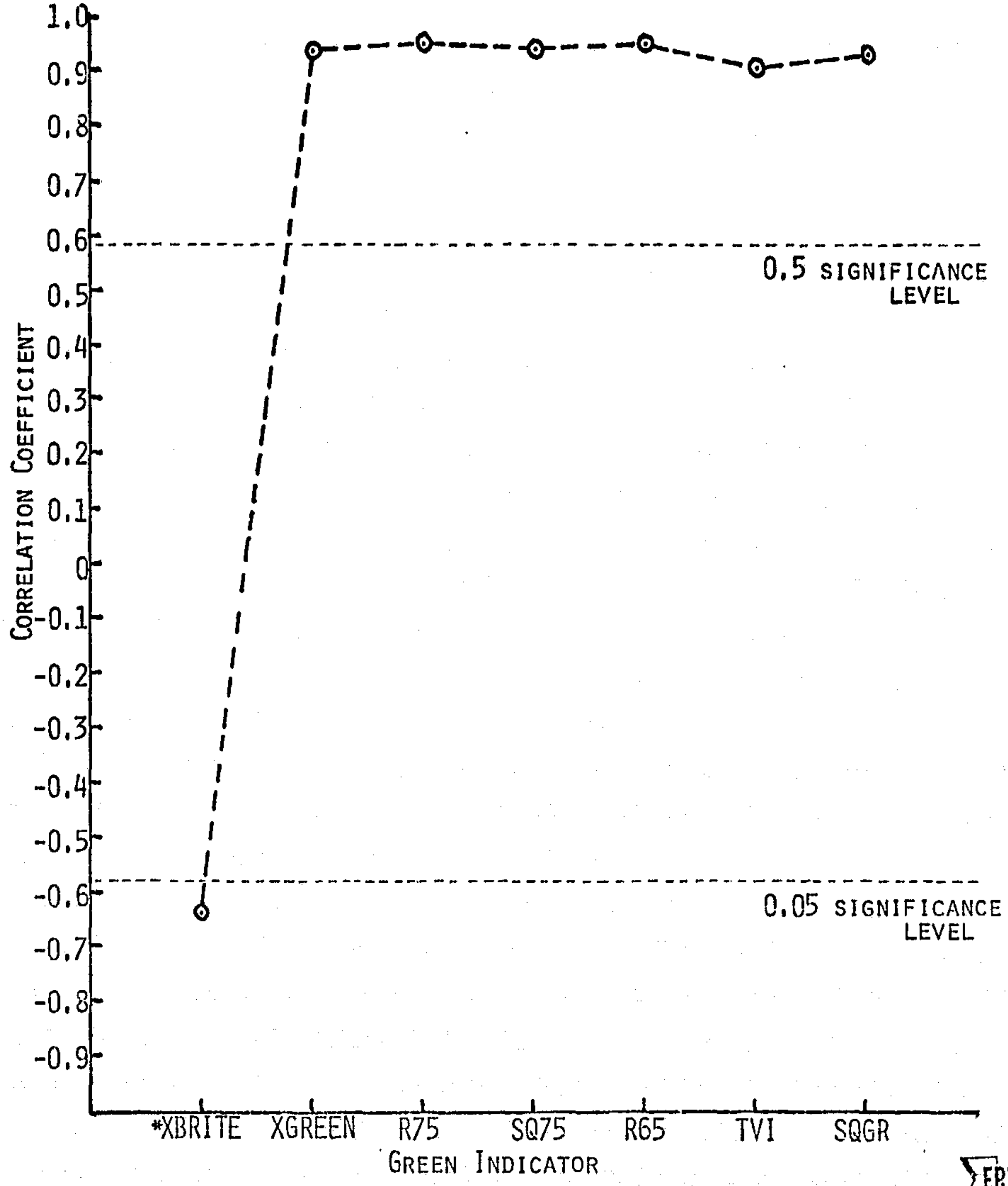
In Section 2.1 reference was made to various Landsat indicators of green wheat development. These indicators are discussed in this section. The usefulness of various green feature indicators was investigated using Finney, 21 May 1975 Landsat data for which field measurements of vegetation condition were available. The green feature indicators examined include:

1. the brightness EXTEC3* channel (XBRITE)
2. the green stuff EXTEC3 channel (XGREEN)
3. the ratio of original Landsat Band 7 and 5 (R75)
4. the square root of the Band 7/Band 5 ratio (SQ75)
5. the ratio of original Band 6/Band 5 (R65)
6. TVI ($\sqrt{\frac{7-5}{7+5} + 0.5}$)**
7. The square root of the green stuff EXTEC3 channel (SQGR).

The correlation between various green feature indicators and measurements of green leaf area index is shown in Figure 4. Features 2-7 are all highly positively correlated with green LAI. The brightness feature (#1) is negatively correlated with green LAI, the reasons for which were hypothesized in the previous quarterly report. Although it is not expected

* EXTEC3 related features are described in the previous quarterly report.
** Previously used by Texas A&M University (TAMU) [2].

FIGURE 4. CORRELATION OF LANDSAT "GREEN INDICATORS" WITH FIELD MEASUREMENTS OF GREEN LAI. (FINNEY SITE, 21 MAY 1975)



* For definition of "Green Indicators" see first paragraph of Section 2.3.

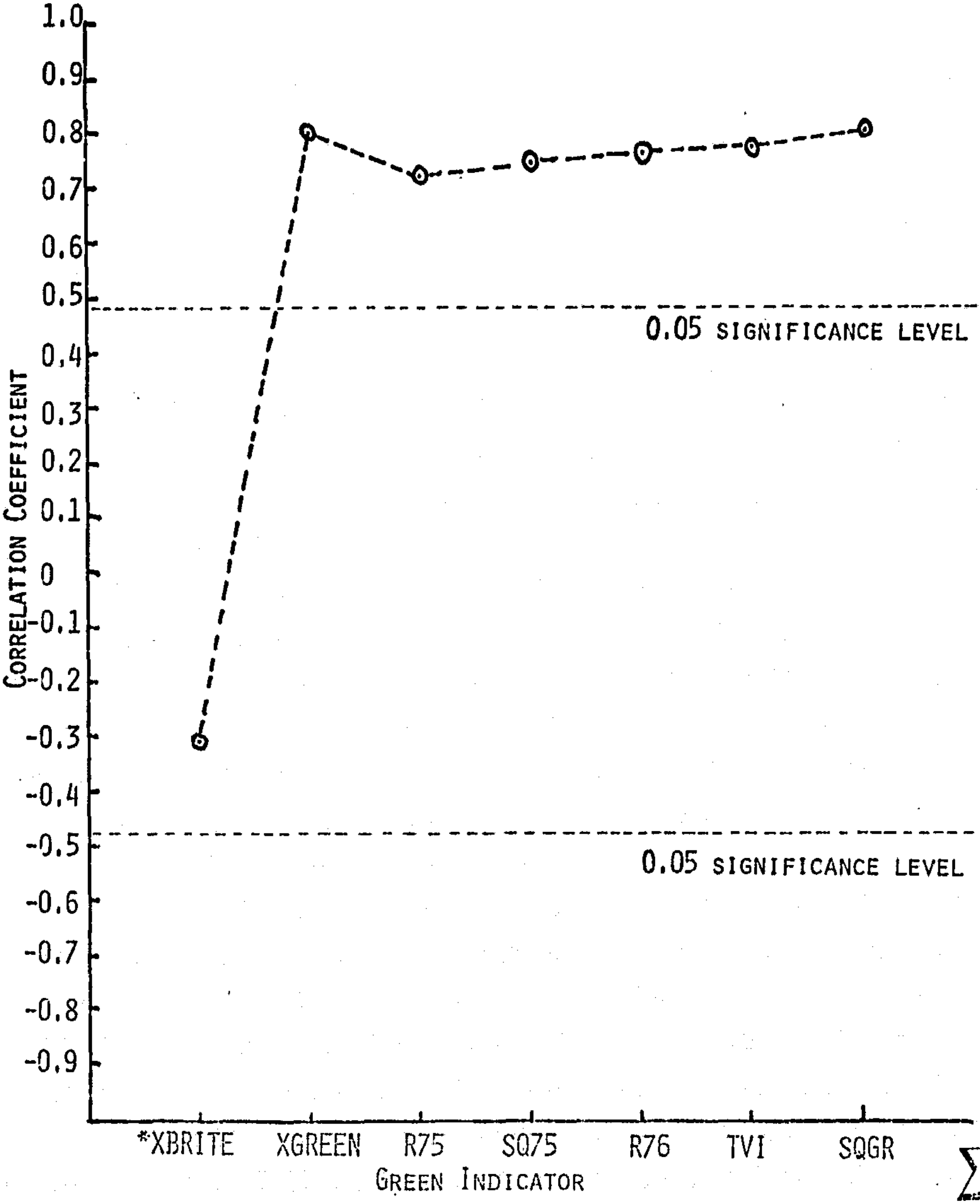
that the brightness channel (which is indicative of relative soil reflectance) will always be as useful as it is in this data set, Johannsen and Barney [1] state that in the Great Plains a very dark soil is indicative of a more fertile soil and, therefore, higher yield potential.

Similar relationships have been found for the correlation of the Landsat green feature indicators discussed above and wheat grain yield (Figure 5). Once again, all of the green feature indicators are highly positively correlated with yield, and the soil brightness feature is negatively correlated with yield, although not significantly so.

The justification for the square roots in some of the features is related to evidence to date which indicates that the relationship between percent cover or LAI and yield is not linear. This effect is intuitively reasonable since there must be an upper bound on yield that is approached more or less asymptotically as percent cover and LAI increases. Therefore, the best possible indicator of percent cover or LAI would not necessarily be the best possible indicator of yield over a broad range of yield values. In practice, we have found that the square root transformation of most Landsat green indicators is more highly correlated with yield than the green indicator is otherwise.

Because no manipulation of the original Landsat 4-channel data can create information, data normalization techniques such as green feature transformations can do no better than maintain the total amount of information that was originally present. All the techniques employed to date have led to some degree of reduction of information. However, a majority of information lost is unrelated to crop vegetative condition and potential yield. Therefore, our initial results suggest that a modest reduction in the amount of yield-predicting capability is justified, since substantial improvements in the capability to extend yield prediction from one set of measurement conditions to another can be achieved in the process.

FIGURE 5. CORRELATION OF LANDSAT "GREEN INDICATORS" WITH FIELD YIELD. (FINNEY SITE, 21 MAY 1975)



ERIM

*For definition of "Green Indicators" see first paragraph of Section 2.3.

2.4 YIELD PREDICTION EXTENSION

In order for Landsat data to be used most effectively as part of a wheat yield forecasting system, a relationship between Landsat data and wheat yield developed under one set of conditions (environmental conditions, cultural practices) should be extendable to Landsat data collected under different conditions at a different place and/or time. In any event, the limitations to the extendability of a relationship between Landsat data and the wheat yield should be known, in order to minimize the possibility of large errors in yield forecasting. There are at least three possible sources of variability that could potentially cause a deviation in a Landsat-wheat yield relationship:

1. changes in environmental conditions (e.g., atmospheric haze and soil reflectance)
2. changes in cultural practices (e.g., irrigation, fertilization, wheat variety)
3. changes in previous crop history (e.g., planting date, previous cropping practice, and previous weather conditions insofar as they affect plant development and potential yield).

The following sections discuss the importance of the effects of some of the above sources of variability, with respect to extension of a yield prediction relationship, and an investigation of possible ways of minimizing the effects of such variability.

Tests of the feasibility of extending a Landsat-wheat yield relationship over time and/or space were performed for three sets of conditions. In order of expected increasing complexity and difficulty, the three types of conditions tested were:

1. local (adjacent day) yield prediction
2. extension from a predominantly non-irrigated site to another predominantly non-irrigated site.
3. extension from a non-irrigated site to a predominantly irrigated site.

In the initial testing of yield prediction extension, three normalization/extension techniques were examined. The techniques examined were:

1. EXTEC3 *
2. SQ75 ($\sqrt{7/5}$)
3. TVI ($\sqrt{\frac{7-5}{7+5}} + 0.5$)

2.4.1 Same Site, Adjacent Day Prediction

We examined the ability to predict yield on a Landsat data set using a yield relation developed on Landsat data gathered over the same site on an adjacent day. Landsat data were used from May 20 and 21, 1975 on the Ellis site for 33 fields. (Pixels from these fields were taken using a 1.0-pixel inset. A similar test using a 1.5-pixel inset for 18 fields was reported in a previous quarterly report.) A mean square error (MSE) for the regression using the 33 fields was calculated by

$$MSE = \frac{1}{n-m-1} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

where n = number of cases (fields)

m = number of variables (channels used in regression)

Y_i = yield for field i

$\hat{Y}_i$ = Landsat predicted yield for field i.

A regression was performed on the May 21 Landsat data. The resulting regression equation was subsequently applied unchanged to the May 20 Landsat data to predict yield, and the MSE was again calculated.

In order to statistically quantify the degree to which performance was degraded in extending a yield predicting regression equation from one

* EXTEC3 is an algorithm developed at ERIM to account and correct for variable external effects such as atmospheric condition.

data set to another, an "F-statistic" was computed as the ratio of the MSE of the extended equation to the base equation. The larger the F-ratio, the worse the prediction of individual field yields was compared to the base prediction of yield.

Another statistical test performed was to determine how well the <u>average</u> yield for all fields was predicted. This test, a "t-test", was then computed as

$$t = \frac{\hat{\bar{Y}} - \bar{Y}}{s/\sqrt{n}}$$

$\bar{Y}$ = average value of yield

$\hat{\bar{Y}}$ = average predicted value of yield

$s^2 = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2/n-1$

The null hypothesis is $\hat{\bar{Y}} - \bar{Y} = 0$, or that the mean values of actual and Landsat-predicted yield are the same. The larger the t-value, the less likely the hypothesis is to be true.

F and t tests were computed for data that was not normalized in any way, in order to determine the severity of the problem of using unnormalized Landsat data. F and t tests were subsequently computed for the three normalization techniques mentioned previously, namely EXTEC3, SQ75, and TVI. The results are presented in Table 6.

The F and t values that are statistically significant are indicated by asterisks. Note that if the data is not normalized at all, both the F and t tests are significant. In other words, neither individual field yields nor mean value of yield for all fields is predicted accurately without any normalization of the data. All three of the normalization procedures, however, result in no significant differences (F or t tests) in yield prediction performance by the extension, indicating that the normalization procedures have been useful in extending yield prediction capabilities.

TABLE 6. PREDICTION OF YIELD FOR ELLIS SITE, 20 MAY 1975, USING (A) RELATION DEVELOPED LOCALLY, AND (B) RELATION DEVELOPED FOR ELLIS SITE, 21 MAY 1975

METHOD	A (LOCAL) MSE[1]	B (NON-LOCAL) MSE[1]	MEAN DIFFERENCE[2]	F - STATISTIC	T - STATISTIC
ORIGINAL LANDSAT BANDS	19.4	45.0	- 5.0	2.3*	4.2*
EXTEC3-TRANSFORMED BANDS	19.6	20.9	0.4	1.1	0.5
SQUARE ROOT OF BAND 7/ BAND 5 RATIO	27.5	25.9	-.002	0.94	.002
TVI	25.4	23.9	.02	0.94	.02

1. MEAN SQUARE DIFFERENCE BETWEEN ACTUAL AND PREDICTED YIELD
2. DIFFERENCE BETWEEN MEAN OF PREDICTED YIELD AND MEAN OF ACTUAL YIELD

* SIGNIFICANT AT 0.05 LEVEL

2.4.2 Dryland Site/Dryland Site Prediction

The second and more difficult test of yield prediction extension performance was made using Ellis 11 May 1975 data and Finney 6 May 1976 data. Both sites are predominantly not irrigated, but the fact that the data is for different locations and different years implies that the weather conditions may have been different during the growing season. Crop phenological development was also somewhat different.

Degree days from March 1 (using 40^{o} F as the threshold level) were computed for both sites, and on this basis the Ellis site was slightly ahead of the Finney site phenological development on May 6. For this reason, we assume that the Ellis 3 May 1975 Landsat data would have been more analogous to the Finney 6 May 1976 data in terms of crop phenology.

However, Ellis 3 May 1975 data were collected by Landsat 1, whereas the Ellis 11 May 1975 and Finney 6 May 1976 data were collected by Landsat 2. Because of the differences in calibration between the two satellites, we chose to use the Ellis 11 May 75 site, rather than 3 May 76, for extension from the Finney 6 May 76 site.

The May 11 data was used as the base data set, and yield prediction was attempted using a relationship developed on May 6. The results are presented in Table 7.

Again, data that had not been normalized failed both the F and t-tests. In other words, neither individual field values nor average yield for all fields were predicted accurately.

In this case the EXTEC3 transformed data yield extension attempt also failed both the F and t tests, and was not much better than the unnormalized data extension attempt. While the parameters of EXTEC3 were derived for Landsat 1 data, we expected improvement in yield extension between Landsat 2 data sets, as long as both data sets had the same calibration and the Landsat 2 calibration differs not too greatly from the Landsat 1 calibration. But, in fact, GSFC changed the calibration procedure for Landsat 2 in July 1975. We will need to determine

TABLE 7. PREDICTION OF YIELD ON ELLIS SITE, 11 MAY 1975, USING (A) RELATION DEVELOPED LOCALLY, AND (B) RELATION DEVELOPED FOR FINNEY SITE, 6 MAY 1976

METHOD	A (LOCAL) MSE[1]	B (NON-LOCAL) MSE[1]	MEAN DIFFERENCE[2]	F - STATISTIC	T - STATISTIC
ORIGINAL LANDSAT BANDS	26.6	673.	-24.7	25.3*	5.4*
EXTEC3-TRANSFORMED BANDS	26.9	467.	-20.2	17.4*	5.4*
SQUARE ROOT OF B AND 7/ BAND 5 RATIO	39.9	91.5	- 2.1	2.4	4.1*
TVI	35.6	77.9	- 3.4	2.2	4.2*

1. MEAN SQUARE DIFFERENCE BETWEEN ACTUAL AND PREDICTED YIELD
2. DIFFERENCE BETWEEN MEAN OF PREDICTED YIELD AND MEAN OF ACTUAL YIELD

* SIGNIFICANT AT 0.05 LEVEL

ERIM

more details about calibration differences before we can fully interpret the EXTEC3 results.

Both SQ75 and TVI yield extensions "passed" the F test at the 5% level, but only barely so. In other words, prediction of individual fields is not statistically significantly degraded by the extension procedure. However, predicted average value of yield for all fields is significantly different. Apparently, the reason that individual field yields were predicted accurately (F-test), while the average value of field yields was not (t-test), is due to a small but consistent bias in individual field yield prediction.

The F-statistic compared the mean squared value of the individual field yield deviations and none of the individual field yield predictions were very far in error. However, they all tended to be in error in the same direction. Therefore, the cumulative effect on the average value of predicted yields showed up in a significant t-test.

2.4.3 Dryland Site/Irrigated Site Prediction

The third and also difficult test of yield prediction extension performance was made using 21 May 1975 Finney data and 21 May 1975 Ellis data. The Finney site is predominantly irrigated and fertilized, whereas the Ellis site is predominantly non-irrigated and non-fertilized. The phenological state of the two sites was assumed similar on May 21, based on both ASCS field observations and on the fact that both sites experienced nearly the same number of degreee days from March 1 to May 21.

The Finney data was used as the base data set, and yield prediction was attempted using a relationship developed on the Ellis data. The results are presented in Table 8.

Once again, the Landsat data that had not been normalized failed both the F and t-tests. Neither individual field yield values nor mean yield for all fields was predicted accurately. None of the three normalization techniques passed the F and t-tests, either. In other words, none of the normalization techniques tested were able to extend a yield prediction relationship from one site to the other.

TABLE 8. PREDICTION OF YIELD FOR FINNEY SITE, 21 MAY 1975, USING (A) RELATION DEVELOPED LOCALLY, AND (B) RELATION DEVELOPED FOR ELLIS SITE, 21 MAY 1976

Method	A (Local) MSE[1]	B (Nonlocal) MSE[1]	B (Nonlocal) Mean Difference[2]	F-Statistic	T-Statistic
Original Landsat Bands	60.2	460.	-19.3	7.7*	3.7*
EXTEC3 Transformed Bands	60.2	412.	-18.6	6.8*	3.8*
Square Root of Band 7/Band 5 Ratio	71.6	305.	-15.3	4.3*	3.6*
TVI	56.1	342.	-16.7	6.1*	3.7*

1. Mean Square Difference Between Actual and Predicted Yield

2. Difference Between Mean of Predicted Yield and Mean of Actual Yield

* Significant at 0.05 level

One of the probable reasons for this poor yield prediction extension is that most of the fields on the Ellis site were low to medium in yield values while most of the fields on the Finney site were medium to high in yield values. The average value of yield for the Ellis fields was 32.4 bu/acre and the average value of yield for Finney was 52.9 bu/acre. The non-linearity in the relationship between Landsat data and yield may, therefore, have caused some of the problems in extending predictive relationships from one site to another. It is also possible that the irrigated and fertilized fields on the Finney site have different structural and radiometric (spectral) properties than non-irrigated, non-fertilized fields on the Ellis site. Since no field data were collected at the Ellis site, we cannot confirm this.

On the basis of the preceding discussion we conclude that data normalization is still a significant problem for extension of yield prediction relationships.

2.5 SOIL REFLECTANCE VARIATION

The previous quarterly report discussed the importance of identifying transformations which adequately monitor features of interest (such as green development), and which simultaneously minimize the effect of variable features (such as soil) that affect unambiguous assessment of the features of interest. Since soil reflectance variations tend to interfere with unambiguous assessment of green development, we are interested in how much soil reflectance varies, as well as in how to reduce the effect of soil reflectance variations.

The fact that soil reflectance does vary considerably on the 1975-76 Finney site is indicated by ground-based measurements of soil reflectance made by Texas A&M University field personnel using an Exotech ERTS radiometer (Table 9). Additional evidence based on ERIM laboratory measurements of hemispherical soil reflectance for Finney soils was tabulated in the previous quarterly report, and is documented here in graphical form (Figure 6).

TABLE 9. AVERAGES OF BROAD-BAND GROUND SPECTRAL REFLECTANCE MEASUREMENTS MADE BY THE LACIE FIELD MEASUREMENTS TEAM USING AN EXOTECH ERTS RADIOMETER
[From Reference 3]

Soil Reflectance	Value In Landsat Band: 4	5	6	7
Mean, m	0.130	0.157	0.216	0.263
Standard Deviation, σ	0.060	0.049	0.057	0.068
Coefficient of Variation, (σ/m)	0.46	0.31	0.27	0.26

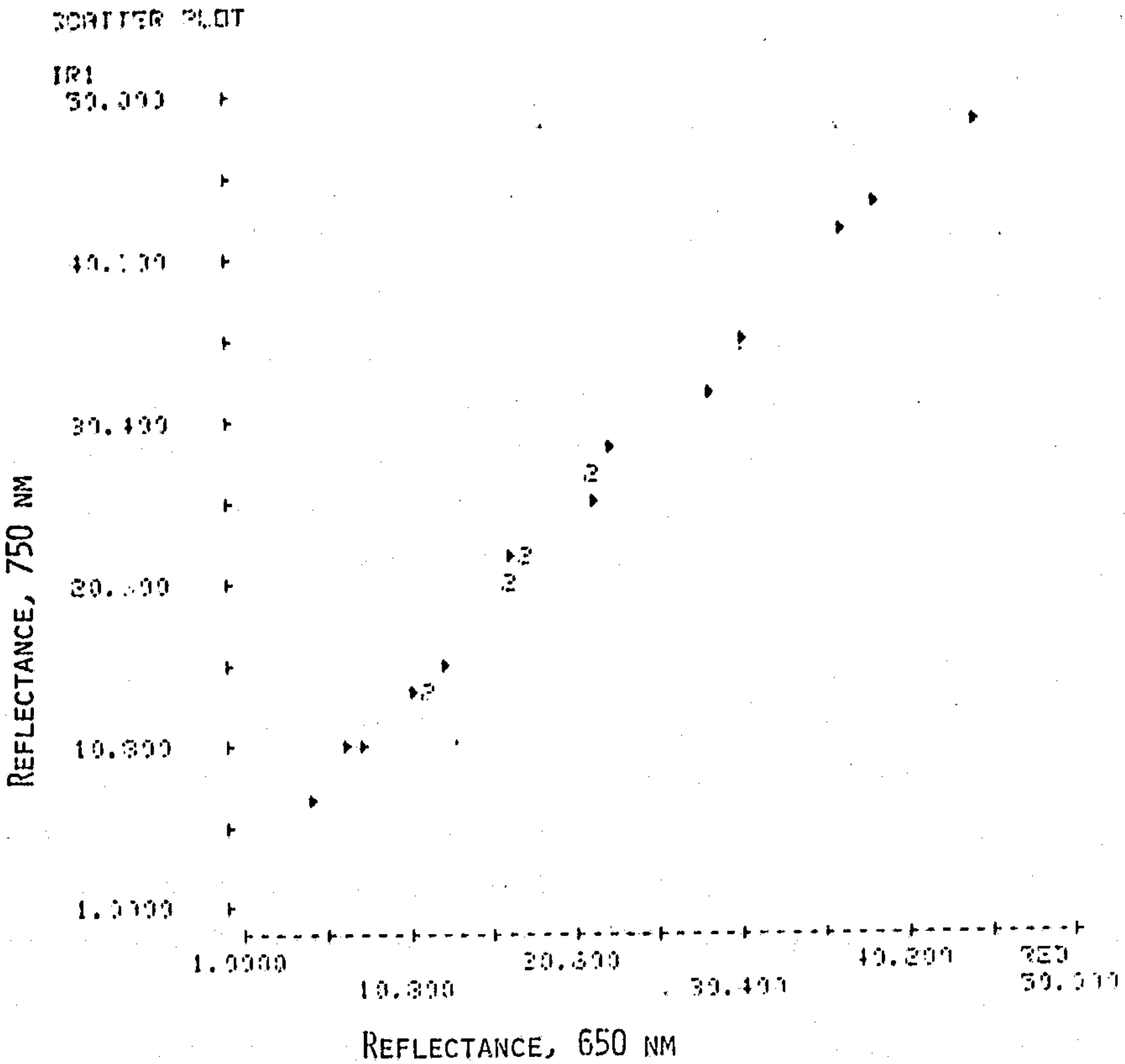


FIGURE 6. SCATTER PLOT OF 21 SAMPLES OF REFLECTANCE (%) FROM FINNEY SOILS AT 750 NM AND 650 NM (FROM ERIM FIELD MEASUREMENTS)

One method of soil reflectance normalizing makes use of various ratios of individual Landsat bands, such as Band 7/Band 5 and Band 6/Band 5, or other green feature indicators such as described in Section 2.3. May 6 Landsat data for the 1976 Finney site on three wheat fields that were plowed up prior to harvest shows a substantial variation in soil reflectance. The effect of several green feature/soil variation transforms on the Landsat data for the three fields is shown in Table 10. Note that the transforms have much less variability than the individual bands.

It is more difficult at the moment to indicate empirically what the usefulness of the soil normalizing transforms is in a vegetation canopy using actual Landsat data because insufficient ground data is available. The usefulness of the transforms in a vegetation canopy with variable reflectance can be investigated, however, using a vegetation canopy reflectance model. Malila, et al [3], calculated the canopy reflectance under a variety of conditions using structural and radiometric data collected on the 1975-76 Finney site as part of this project. The reflectance measurements were converted to simulated Landsat radiance values, and some of the results are shown in Table 11. Note that the variation in individual band simulated Landsat radiances is large for low vegetation cover canopies, but decreases as the vegetation cover increases. The ratio values are nearly constant for a given value of vegetation cover.

An additional indicator of the usefulness of a transformation which will result in normalization of variable soil reflectance is given by 7 June 1975 Ellis Landsat data which was processed during this reporting period. On the day before the Ellis Landsat overpass nearly 4 cm. of rain fell. As a result, the individual Landsat band correlations with yield were very anomalous due to the low reflectances of the wet soil. Because of the low soil reflectances, and the corresponding condition of anomalous correlations of Landsat data with yield, one might expect that extension of yield prediction would fail on this data set

TABLE 10. LANDSAT DIGITAL COUNT AVERAGE VALUES FOR INDIVIDUAL BANDS AND FOR THREE TRANSFORMS ON THREE PLOWED FIELDS (6 May 1976)

	Landsat Bands				Transforms		
Field	4	5	6	7	6/5	7/5	TVI
27	46.5	66.5	73.5	32.4	1.11	0.49	0.50
55	32.2	43.2	49.2	21.8	1.14	0.50	0.48
165	47.1	69.0	76.7	33.9	1.11	0.49	0.49

TABLE 11. MODELED VALUES OF LANDSAT RADIANCE AND RADIANCE RATIOS FOR CANOPIES WITH LOW TO INTERMEDIATE VEGETATION COVER AND HIGH SOIL VARIABILITY (After [3])

Stage	Green Cover (%)	Soil	Band 5	Band 6	Band 7	Band 6/ Band 5	Band 7/ Band 5
Emergent	3	1	0.326	0.369	0.477	1.13	1.46
		2	0.459	0.522	0.670	1.14	1.46
		3	0.591	0.676	0.865	1.14	1.46
Jointing	10	1	0.446	0.615	0.819	1.38	1.84
		2	0.606	0.830	1.100	1.37	1.82
		3	0.766	1.052	1.392	1.37	1.82
Pre-Heading (Booting)	38	1	0.393	0.828	1.196	2.11	3.04
		2	0.459	0.964	1.398	2.10	3.05
		3	0.526	1.109	1.619	2.11	3.08

using untransformed data. However, soil normalization transforms resulted in a less anomalous relationship with yield. We therefore expect that extension of yield prediction has a greater chance of success once methods of normalizing soil variations have been applied.

2.6 LEAF AREA DURATION

One of the hypotheses stated early in this investigation was that information about amount of photosynthetic material, integrated over time, would be more highly correlated with yield than information at a single point in time. For example, it was hypothesized that the integral of leaf area index (LAI) over time (leaf area duration - LAD) from heading to senescence would be more highly correlated with yield than LAI at a point in time. Similarly, the duration of percent green wheat cover might be more highly correlated with yield than the values at the time of heading*.

The relation between field condition at a point in time and grain yield was investigated by calculating the correlation between percent green wheat cover determined from ERIM field measurements and wheat grain yield. Four time periods were available for Finney 1975 data, namely May 21, (approximately the time of heading), May 30, June 9, and June 18. The results are presented in Figure 7. For this limited set of data on predominantly irrigated fields, the highest correlation occurred on May 21, and it decreased monotonically through June 18.

An approximation to percent cover duration was computed by successively adding percent cover information to the May 21 data to get a total. Successive summations were then correlated with yield to determine

---

* The relationships discovered for LAI and for percent cover as a part of this project have thus far been quite similar. Therefore, the two parameters will be used more or less interchangeably as indicators of field condition.

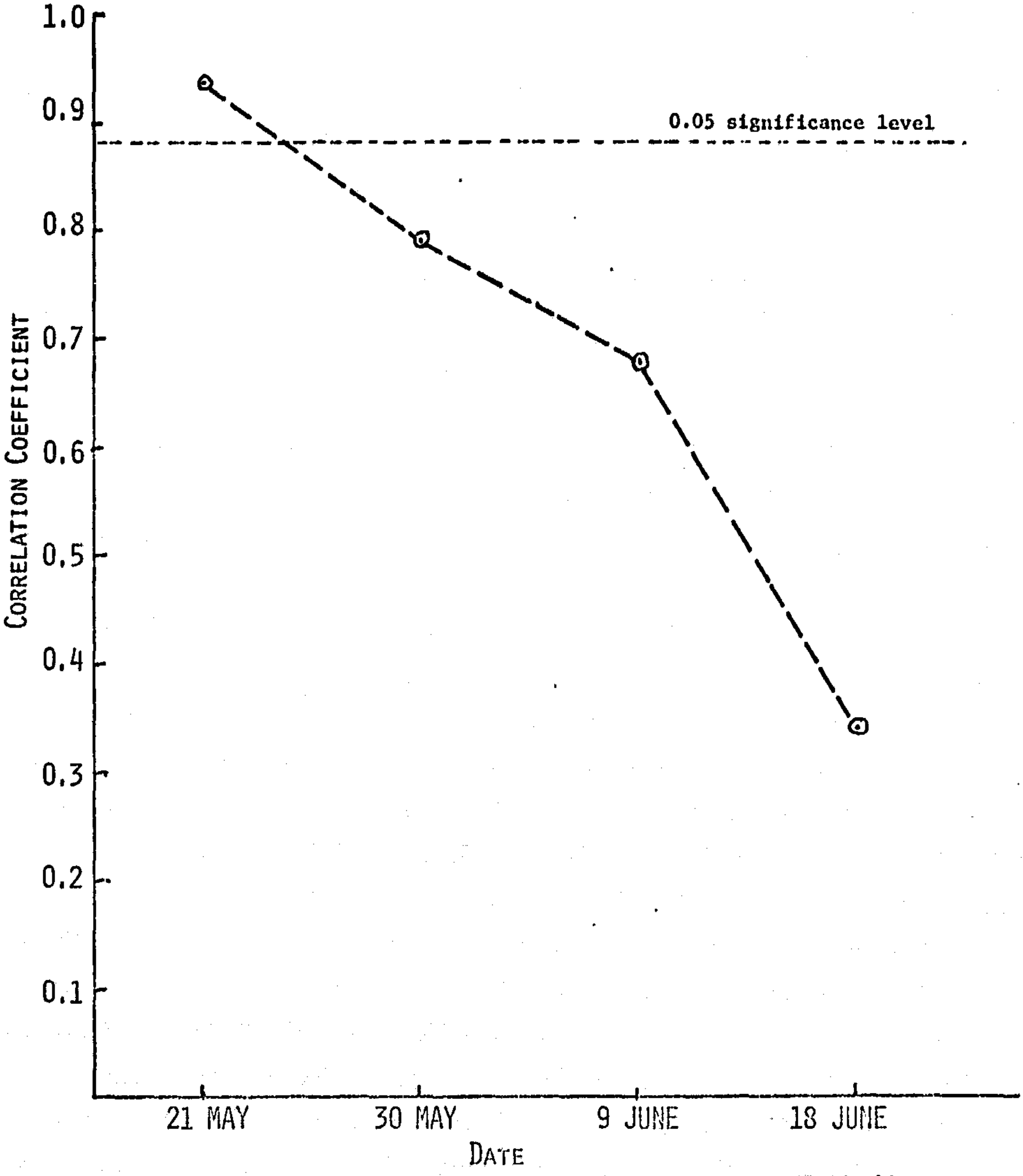


FIGURE 7. CORRELATION OF % GREEN WHEAT COVER WITH YIELD AS A FUNCTION OF DATE, FINNEY, 1975.

if correlations were improved. The results are shown in Figure 8 where it can be seen that none of the summations of percent cover over time improve the correlation with yield obtainable by the values of percent cover on May 21. Similar results were found for LAI and LAD.

As indicated by the above discussions, it is felt that the hypothesis (that leaf area duration or percent cover duration features improve yield estimation) has not been verified for this data set. The lack of verification of the hypothesis may be due to the fact that, for the highly irrigated fields in this data set, the amount of grain yield may be more closely related to the amount and timing of irrigation than to the amount and duration of green photosynthetic material.

The same kind of analysis can be made for a larger number of fields if Landsat data is used as a surrogate for amount of green vegetation present. However, we have not had the opportunity to establish a green feature indicator that works well in largely senescent wheat canopies such as were present on June 9 and June 18.

We can, however, use Landsat data with some confidence as a surrogate for green vegetative cover when the wheat is predominantly green. This was done for Ellis Landsat data using the May 20 pass as the best single date and summing backwards to May 11 and May 3. The correlation with yield was greater for May 20 SQ75* data than was the correlation with the sum of SQ75 May 20 and SQ75 May 11, which in turn was greater than the correlation with yield for the sum of SQ75 May 20 and SQ75 May 11 and SQ75 May 3 (see Figure 9). No sum of dates was as highly correlated with yield as the best single date (May 20).

Based on the ERIM field data on actual vegetation condition and on Landsat indicators of vegetation condition discussed above, it appears that a summation of amount of photosynthetic material over time is not more highly correlated with yield than is information at a point in time

---

* As before, SQ75 is the square root of the Landsat Band 7 to Band 5 ratio.

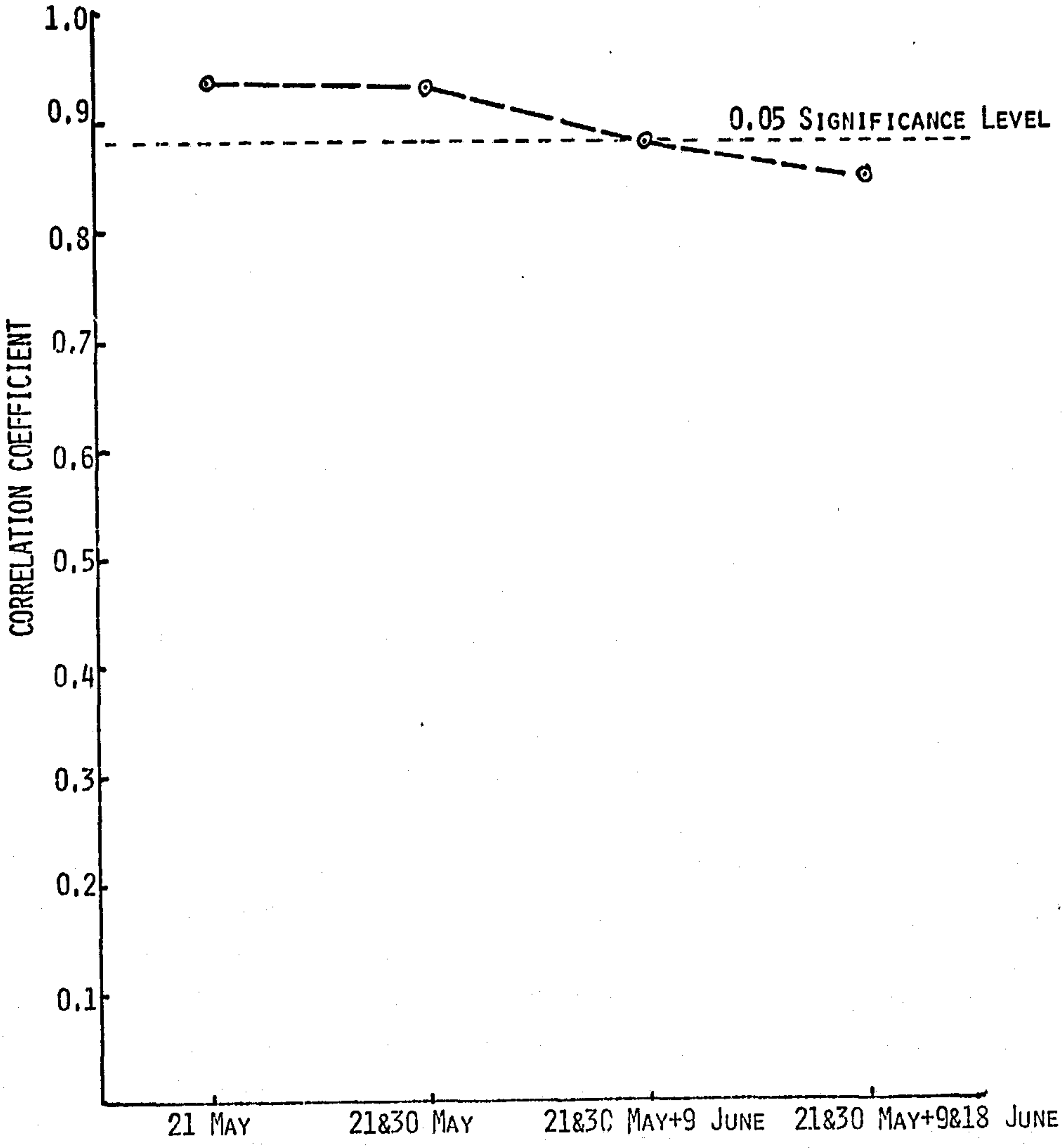


FIGURE 8. CORRELATION OF SUMMATION OF % GREEN WHEAT COVER WITH YIELD FOR VARIOUS DATES, FINNEY, 1975

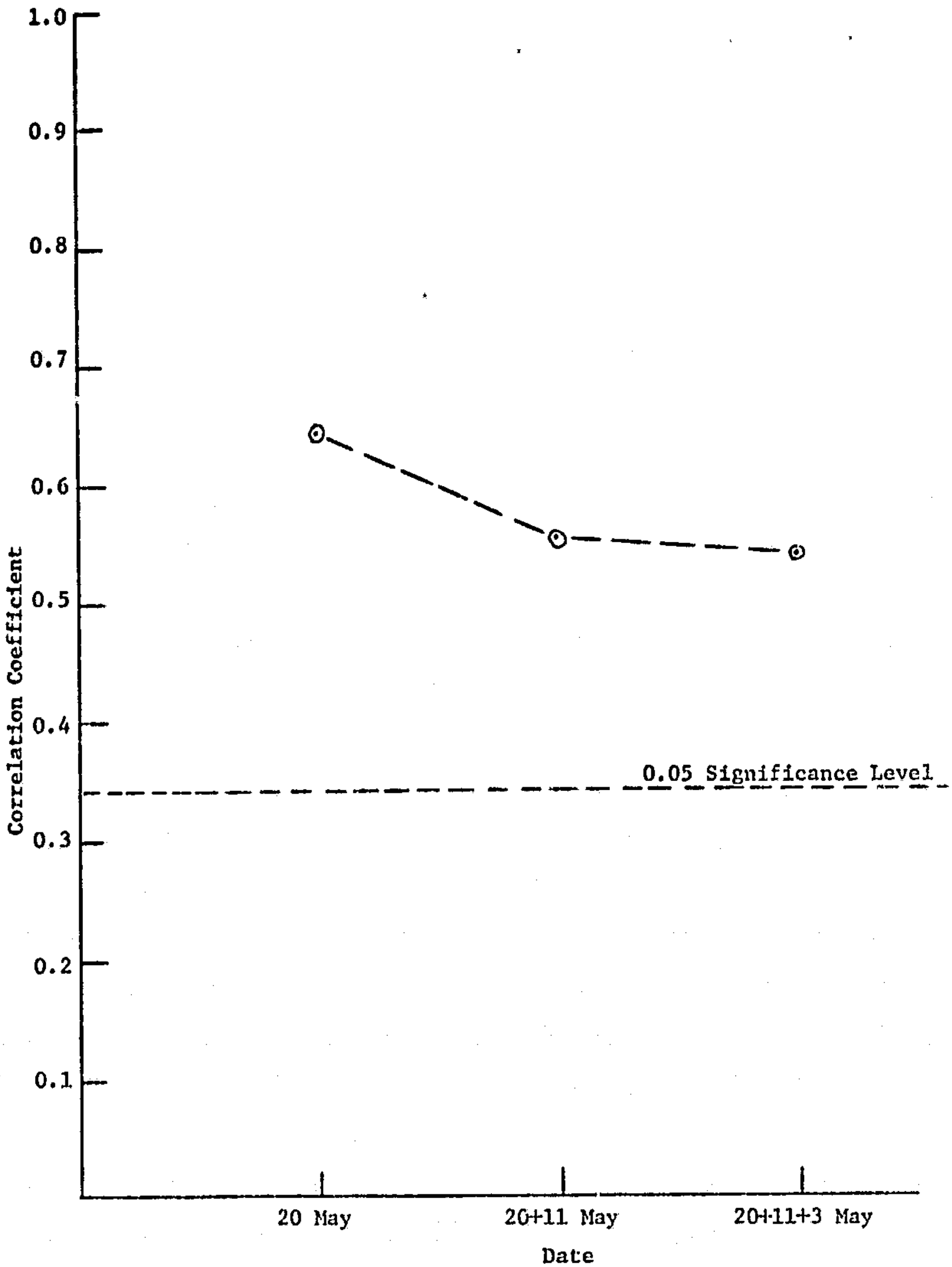


FIGURE 9. CORRELATION OF SUMMATION OF LANDSAT INDICATORS OF GREEN WHEAT COVER WITH YIELD, Ellis, 1975

for the cases investigated so far. In fact, initial results indicate that the opposite is true. Although it is disappointing that the hypothesis is not supported by the data analyzed thus far, it is encouraging to observe that it is also not necessary to have all of a certain sequence of dates to perform accurate yield prediction. In other words, the initially proposed yield prediction method (based on Landsat indicators of LAD) [Reference 4] may be more elaborate than is required. However, multitemporal data used independently (i.e., not summed) has proven to be useful in improving yield prediction, as indicated elsewhere in this report and also in previous quarterly reports.

3.0 TRAVEL/PRESENTATIONS/PUBLICATIONS

On 18 October 1976, Richard Nalepka and John Colwell participated in a Landsat Follow-on Program Review held at NASA/Goddard Space Flight Center. A summary of progress to date was presented to a panel of discipline specialists.

4.0 FUTURE PLANS

We will continue to investigate the relationship between Landsat data and yield as a function of time and site. The efforts will be concentrated on Landsat data from late fall (November), early spring (March), and near heading (mid-May). We will begin to analyze data from the 1975-76 Ellis site. The feasibility of extending yield prediction relationships over time and space will continue to be analyzed. We will also continue efforts to assess the relative utility of Landsat, meteorological, and ancillary data for prediction of winter wheat yield. We will pursue the matter of calibration procedure changes in Landsat, and their effects on year-to-year yield prediction efforts, and we will pursue methods of correcting for calibration differences and other external effects which interfere with reliable prediction of yield.

REFERENCES

1. Johannsen, J., and T. Barney, "A Review of Soil Color", University of Missouri, Columbia, Missouri, 1976.
2. Deering, D., et al, "Measuring 'Forage Production' of Grazing Units from Landsat Data", Proceedings of Tenth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, 1975.
3. Malila, W., R. Cicone, and J. Gleason, "Wheat Signature Modeling and Analysis for Improved Training Statistics", ERIM Report No. 109600-66-F, NASA Contract No. NAS9-14123, 1976.
4. Colwell, J., and G. Suits, "Yield Prediction by Analysis of Multispectral Scanner Data", ERIM Report No. 109600-17-F, NASA Contract No. NAS9-14123, 1975.
5. Feyerherm, A., "Prediction of Wheat Yield", Quarterly Progress Report, NASA Contract No. NAS9-14533, 1 May 1976.
6. U.S.D.A., Miscellaneous Report #967, Statistical Reporting Service of the U. S. Department of Agriculture.
7. Feyerherm, A., "Prediction of Wheat Yield", Quarterly Progress Report, NASA Contract No. NAS9-14533, 1 August 1976.