D40

N78-14570

WHEAT YIELD FORECASTS USING LANDSAT DATA*

John E. Colwell, Daniel P. Rice, and Richard F. Nalepka
Environmental Research Institute of Michigan
Ann Arbor, Michigan

ABSTRACT

Many of the considerations of winter wheat yield prediction using Landsat data are discussed. In addition, a simple technique which permits direct early season forecasts of wheat production is described.

---

1. INTRODUCTION

The accurate forecast of production of agricultural crops, particularly those subject to international trade, is becoming a more urgent requirement due to the growing world population and the resulting food supply problem. Past evidence shows that traditional approaches have sometimes proven inadequate. The purpose of the investigation described here is to determine the extent to which Landsat data can be used to improve winter wheat crop production forecasting capabilities.

The production of an agricultural crop can be thought of as the product of the yield (e.g., bushels/acre) and the area (e.g., acres). Remote sensing data, and Landsat data in particular, can potentially be used to assess both crop yield and crop acreage. In this study we consider first the problem of estimating wheat yield (per acre) using the data known to be from wheat fields. This problem is addressed by demonstrating the nature of yield prediction using Landsat data, by comparing such yield prediction to other methods, and by studying the consistency of Landsat yield relations from one site or acquisition to another. Second, we consider the possibility of estimating total wheat production without a determination of whether each portion of data is from a wheat field. An initial test of a technique designed to make such forecasts using early-season Landsat data is presented.

2. BASIS FOR LANDSAT WHEAT YIELD FORECASTING

The fundamental propositions on which Landsat forecasts of wheat yield are based are that: (1) a good early-season indicator of potential wheat grain yield is the degree of vegetative development; and (2) the degree of wheat vegetative development can be estimated using Landsat data.

Farmers and agronomists have long felt that there is a relationship between degree of vegetative development and yield. In fact, traditional ratings of "stand quality" are based on visual estimates of vegetative cover, measurements of stand height, or similar quantities. It has been recognized that such indicators are especially useful since they incorporate and integrate the effects of important environmental conditions, from meteorological factors such as precipitation and solar radiation, to cultural factors such as fertilization and irrigation. No growth model yet developed has been able to perfectly simulate the synergistic effects of all such variables, but the crop itself, by definition, does so. Until recently, it has been difficult to get precise and timely field observations of crop condition over large areas, so estimates of potential yield based on such observations have not generally been practical. However, the

---

*Funded by the NASA/Goddard Space Flight Center on Contract NAS5-22389.

advent of earth resource satellites such as Landsat has presented the possibility of monitoring actual crop condition over large areas in a timely fashion.

Returning to the fundamental propositions mentioned above, the question of whether field vegetative condition is a good indicator of yield was examined using ERIM field measurements of percent green wheat cover made so as to characterize entire fields. The measurements of green wheat cover for each field were then compared with the corresponding farmers' reports of actual wheat yield (bu/acre). Such comparisons using ERIM field measurements at a site in Kansas made during two successive years at equivalent phenological stages are indicated in Figure 1. For these data the correlation between green wheat cover and yield is 0.82. This is a statistically significant correlation and tends to support the proposition that vegetative condition is a good indicator of yield.

The hypothesis that Landsat data can be a good indicator of field condition was also investigated. A variety of transformations of Landsat data chosen to be good measures of green development* were compared to ERIM field measurements of percent green wheat cover. Field average values of a Landsat green indicator are compared with field measured average values of percent green wheat cover in Figure 2. It is clear that there is a high degree of correlation ($r = 0.98$).

The most important test, of course, is whether Landsat data is indicative of potential yield. This hypothesis was examined by comparing field mean values of Landsat green indicators with farmers' reports of actual grain yield harvested. An example of this relationship is shown in Figure 3. The correlation between the Landsat green indicator and yield for this example is 0.80. This relationship is statistically significant.

In summary, the fundamental propositions stated earlier in this section seem to be supported by the above evidence. Therefore, we will proceed to examine other aspects of Landsat relationships with yield.

3. RELATIVE UTILITY OF LANDSAT AND ALTERNATIVE SOURCES OF INFORMATION FOR ESTIMATING YIELD

Having established that Landsat data are related to yield, an important question remaining is how yield estimates using Landsat data compare to those generated using alternative sources and types of data.

3.1 METEOROLOGICAL DATA

Meteorological conditions are important determinants of the ultimate yield of agricultural crops, including winter wheat. Historically, meteorological information has been used with some success to roughly estimate yield on a regional average basis. However, there are factors other than meteorological conditions that are also important determinants of yield. In our test sites, which are 5 x 6 miles or smaller, we found that meteorological conditions were relatively constant over each site. For example, 30 rain gauges placed throughout a site measured May rainfall as 3.76 inches, with a standard deviation of only 0.43 inches. On the other hand, the yield on the site varied substantially (21.0 bu/acre to 74.0 bu/acre) from field to field. On another test site the yield varied from 3 bu/acre to 65 bu/acre.

The reasons for such variations in yield are apparently largely related to factors other than weather, such as differences in topography, soil type, planting density, fertilization, cropping practices in a field, and irrigation, none of which are accounted for by yield models based solely on meteorological data.

---

*e.g., $\sqrt{\frac{MSS7}{MSS5}} = SQ75$; $\sqrt{\frac{MSS7 - MSS5}{MSS7 + MSS5} + .5} = TVI$; (see Section 4.0)

On the other hand, the differences in crop condition and eventual yield found in the local sites are substantially manifested in Landsat data, as indicated in Section 2. Thus, it appears that Landsat data can better account for local variations in yield than can meteorological data.

3.2 FIELD ESTIMATES OF VEGETATIVE CONDITION AND YIELD

Since we are also interested in the potential usefulness of Landsat data for inputs into existing yield models, we analyzed the ability of Landsat data to estimate wheat vegetative condition relative to an alternate field estimate. For purposes of comparison, we used carefully made ERIM measurements of percent green wheat cover as the correct values. For the two data sets where we have data for essentially all green canopies, the correlations with the ERIM measurements for subjective ASCS (Agricultural Stabilization and Conservation Service) estimates of vegetation cover and Landsat green measures are indicated in Table I.

Based on these two tests, it appears 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.

We also compared information on yield derived from Landsat data with alternative estimates of yield and stand quality. As shown in Table II, this comparison was made on three sites using subjective stand quality ratings and objective yield estimates, both of which were made by agricultural experts in the field just prior to harvest. These results suggest that Landsat indicators of yield are generally as well correlated with yield as are some alternative traditional field estimates made by agricultural experts, even for Landsat data collected well before the field estimates using alternative methods.*

3.3 CULTURAL FACTORS

Some of the factors that cause field vegetative condition and potential yield to vary in a region of similar meteorological conditions are cultural in nature, i.e., they are factors that can be affected by the individual farmer. Data on many of these variables are potentially available early in the growing season, and hence, could be used for early yield forecasting. The relative importance of some of these factors and the degree to which they can be accounted for by Landsat data are discussed in this section.

For a particular site, we investigated the importance of the factors listed in Table III. An analysis of variance was performed for the above factors by linear regression with wheat yield for the fields for which such data was available. From this analysis, it was possible to determine the percent of variance in yield accounted for separately by each of the factors. However, since high correlations exist between some of the cultural variables, the results cannot be treated as though the variables were independent of each other. The results presented in Table III indicate that individual factors associated with fertilization and irrigation account for most variance in yield.

We have performed similar analyses of cultural and Landsat variables on several sites. As a result of these analyses we have determined that individual cultural factors may account for a high amount of yield variance in one situation and very little in another situation. This is probably at least partially because of differences in the correlation of the individual cultural variables from one situation to another. It is also probably a result of the complex relationships between cultural and environmental factors and crop growth. Since these relationships are not yet fully understood, there is risk in relying on such cultural factors for predicting yield.

---

*Traditional methods using trained field personnel can certainly be more precise measures of field condition 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.

In all sites which we have examined, a Landsat green measure was found to account for a large amount of variance in yield. This finding lends support to our expectation that a Landsat green measure will account for the combined effects of the complex factors that influence crop growth, and that a Landsat green measure will, therefore, be a good indicator of potential grain yield in a variety of situations.

3.4 COMBINATIONS OF DATA FOR PREDICTING YIELD

In the previous section we discussed the usefulness of various individual cultural variables for predicting yield. In this section we address the question of predicting yield using data from selected combinations of sources.

Table IV gives the results for one site we have examined. Note that, together, all of the cultural variables (1-6) account for a substantial amount of yield variance (75%). Nevertheless, the Landsat green indicators for the four dates (variables 7-10) for which we have processed Landsat data account for even more variance in individual field yield (87%) than all of the cultural variables. The combination of all Landsat and cultural variables accounts for almost all of the variance in yield (94%).

We previously suggested that field condition as measured by Landsat may account for the integrated effects of the factors governing crop growth and potential yield, including the cultural factors. Cultural factors are mostly accounted for by Landsat data in this site. That is, addition of all six cultural factors to the four dates of Landsat transforms increased the variance accounted for by only 6.3%.

In some situations using Landsat data by itself may be sufficient to predict wheat yield with acceptable accuracy on a regional basis. Consider the standard error of estimates of yield. Using the Landsat green measures from the four dates in the previous example the standard error is 4.8 bu/acre on the above test site. If this performance could be achieved on 100 randomly selected fields, with a normal distribution of yields about the mean, the average yield on the 100 fields could thus be estimated to within $\pm$0.48 bu/acre, a significant potential accomplishment.

While Landsat data alone may be sufficient to estimate yield in certain situations, some combination of Landsat, meteorological, and ancillary data will probably improve yield prediction performance. In such situations, the appropriate combination of data sources will depend on the cost of obtaining and using such data, compared to the benefits.

4. YIELD PREDICTION EXTENSION

Thus far in this discussion we have confined our analyses to the Landsat relations with yield on a given site and time. In some sense, these analyses indicate the best performance we might achieve on another site with identical conditions. However, other sites will seldom exhibit identical conditions, and attempts to extend a yield prediction relation generally produce results that are not quite as good as those achieved locally.

The need to correct for conditions which differ from one site to another has led to investigation of Landsat data transforms (green measures) which retain the maximum of information about green vegetation and potential yield, and the minimum of other information (noise). In our tests, the green measures tended to measure green cover and yield well (retaining most of the yield information present in the original 4 Landsat bands), and had some effect in reducing variation due to other causes. However, no single green measure was always superior to the others tested.

We carried out tests of extensions of wheat prediction by developing yield-predictive relations on one site and applying them on another. Each relation was formed using one of the Landsat green measures or using the four Landsat bands. The results of the tests are shown in Table V.

In one test, a Landsat wheat yield relationship developed on May 21 Landsat data was applied to May 20 Landsat data collected on the same site. The May 21/20 test shows that there is only a modest reduction of local yield-predictive information by use of either green measure transform (SQ75, TVI), as evidenced by their slightly larger local RMS error. However, when extending a relationship from one date to another, the non-local (prediction extension) RMS error for individual field yield is less for the transformed data than for the untransformed data. In addition, the mean value of predicted yield is substantially in error using the untransformed four bands of data (5 bu/acre), whereas there is very little bias using the transforms. In other words, the Landsat green measure transforms are better for the extension of yield relations in this test.

In another test, a Landsat wheat yield relationship developed on 18 April Landsat data from one site was applied to 18 April Landsat data from a different site. Again, there is only a small loss of local yield information using either of the transforms. However, both individual field yields and average yield are predicted more accurately by the combination of all four individual non-transformed bands than by either Landsat green measure transformation, as evidenced by the smaller non-local RMS errors and smaller bias.

Additional tests of yield prediction extension have been performed, and they have indicated variable results from one test to another. More testing is being done in an effort to gain more insight into possible sources of error. It may be that procedures that are generally optimum can be discovered only by development of a large base of tests of candidate procedures.

5. DIRECT WHEAT PRODUCTION FORECASTS

Thus far we have discussed only the ability to forecast wheat yield (per acre) using Landsat data. By itself, this information would be valuable as part of a system for forecasting wheat production. However, our work to this point has suggested a method for utilizing the relationship between Landsat data and yield, together with other relationships, to effect direct Landsat forecasts of winter wheat production which may overcome certain troublesome problems in some of the existing approaches.

The existing approaches tend to separate the task of forecasting into two separate subsystems consisting of: (1) wheat acreage determination; and (2) regional average determination of per acre yield. The approach discussed below could make it possible to determine production on a pixel-by-pixel basis, using early-season Landsat data, with a single processing step. Thus, it may become possible to survey large areas such as a state or country much more economically than at present, and achieve more timely information. What follows is a discussion of the rationale of the suggested approach, and a demonstration of its initial implementation.

One of the ideas behind the direct wheat production approach using Landsat data is that an appropriate value of production can be determined for each pixel in the scene, perhaps without even the need to specify whether the pixel is wheat.

We have previously shown that several Landsat transforms are good measures of green vegetative cover, and that cover in turn is strongly related to wheat yield. Given the knowledge of the area covered by a pixel the estimate of yield on a per pixel basis can be directly converted to production. An additional fact is that in winter wheat regions such as Kansas, wheat tends to develop significant green cover sooner than most non-wheat fields and can therefore be easily distinguished. (Wheat classification accuracies of 92 and 94% were achieved on two Kansas sites using only the Landsat SQ75 green measure.)

Thus, if a production-predictive relation (developed on wheat fields) is applied to non-wheat pixels, a very low production indication would be expected, and might be a negligible source of error. If applied to pixels falling on a boundary between wheat and non-wheat, an appropriate intermediate value of green cover, and thus, intermediate average production would be estimated. This intermediate value of production could approximate the total amount of wheat production represented by the pixel, which covers an area only partially planted to wheat. Thus, pixels would tend to contribute only their fair share of the total production estimate.

As a part of this procedure it is necessary to establish the production-predictive relationship on an area where ground truth information is available.* With the relationship established, the present approach is to select a threshold below which no wheat production is assigned to a given pixel. The need for such a threshold is dictated by the fact that, in general, some non-wheat pixels generate Landsat green measures which fall above those of some low production wheat pixels. The threshold value is selected to cause errors of omission and commission to compensate.

As an initial test of the direct production forecast procedure, the above approach was employed using the SQ75 green measure on a portion of the 6 May 1976 Landsat data for Site A. Employing the resulting relationship on all of Site A a production forecast of 42,700 bushels was made. This compared favorably with the actual production of 40,600 bushels for this site, an error of only 5.2%. In addition we applied the same procedure to the same site using 18 April 1976 Landsat data, and to a different site (Site B) using 6 May 1976 Landsat data. The resulting production estimates for these tests are shown in Table VI. Note that the total production estimated for the two May 6 tests was within 1.6 percent of the correct total production, well within LACIE desired accuracy.** Whether the compensating effect of apparently random errors in estimating production would prevail over a larger sample of test sites awaits further investigation.

Preliminary indications based on the three test results give encouragement that the direct wheat production approach using early-season Landsat data is worth pursuing. Many more tests in different situations need to be carried out in order to assess the consistency in performance.

In any event, the approach does address some problems that may exist in present methods. The difficulty in locating field boundaries on Landsat data for determination of wheat acreage is alleviated since all pixels can potentially be included in the proposed new technique. Small or irregularly shaped fields can contribute to the production estimate even if not a single pixel falls completely within the field boundary. Furthermore, large bare areas within wheat fields will be assigned little or no production, thereby giving approximately the correct production, without a decision necessarily having to be made as to whether the area should be assigned to wheat acreage or not. Finally, marginal wheat fields, ones which are not likely to be harvested, will not be included in early-season production forecasts if they fall below the green measure threshold.

Present indications are that these desirable features of the direct wheat production approach are being fulfilled. For example, there were several wheat fields in our Site A test for which no "pure" pixels could be obtained. That

---

*In an operational environment, several carefully selected sites and data from previous years should satisfy the need for training.

**MacDonald, R. B., F. G. Hall, and R. B. Erb, 1975. "The Large Area Crop Inventory Experiment (LACIE) -- An Assessment After One Year of Operation", Proceedings of Tenth International Symposium on Remote Sensing of Environment, Environmental Research Institute of Michigan, Ann Arbor, Michigan.

is, all pixels covering these fields overlapped the field boundary, or very nearly so. One such field had a farmer-reported production of 1001 bushels and an area of 32.7 acres. Even though not a single pure pixel was present, production of 732 bushels was estimated for this field using the direct production procedure, based just on the pixels whose centers fell within the field boundaries.

In Site B there was a wheat field which was not harvested because the stand was too sparse. Every pixel within that field boundary had a green transform value less than the minimum threshold. Therefore, even though the field was wheat, it did not contribute to the production estimate, which is the desired result in this case since no wheat was produced on this field.

6. CONCLUSIONS

As a result of this study, we draw the following interim conclusions:

1. Landsat data can be effectively used to estimate certain variables which are required in existing yield models (such as LAI or percent cover).

2. Landsat indicators of yield are as highly correlated with individual field yield as are estimates using traditional field sampling methods, even when using Landsat data collected several weeks before the field samples are made.

3. A considerable amount of the variance in individual field yield which is not explainable by meteorological data can be accounted for by Landsat data.

4. In order for Landsat data to be of maximal use in an operational system, improvements in the ability to remove the external effects (particularly atmospheric effects) are required.

5. It may be possible in certain situations to make direct wheat production forecasts using early-season Landsat data.

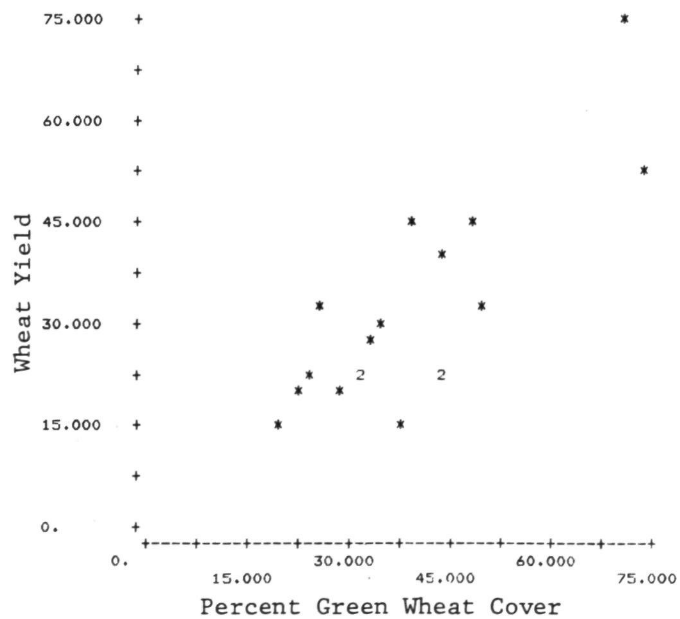

FIGURE 1. ERIM MEASUREMENTS OF PERCENT COVER VS WHEAT YIELD
(Combined 1976 and 1975 Data)

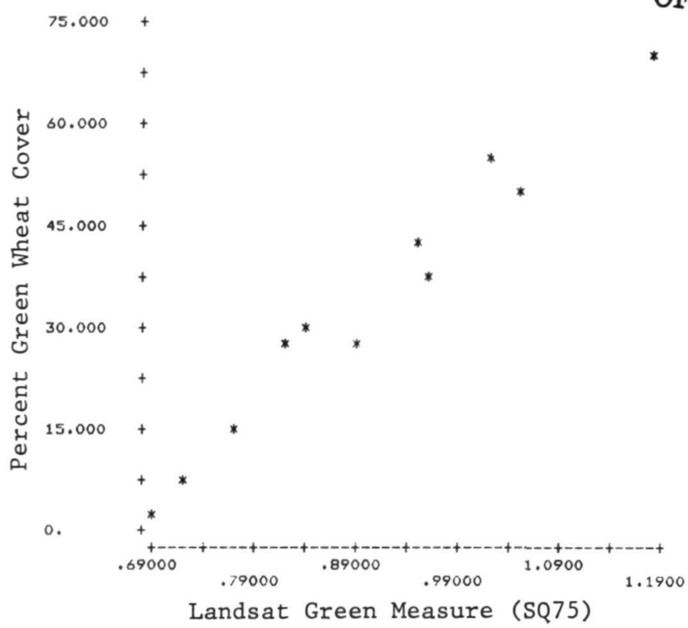

FIGURE 2. LANDSAT GREEN MEASURE VS ERIM MEASUREMENTS OF PERCENT GREEN WHEAT COVER

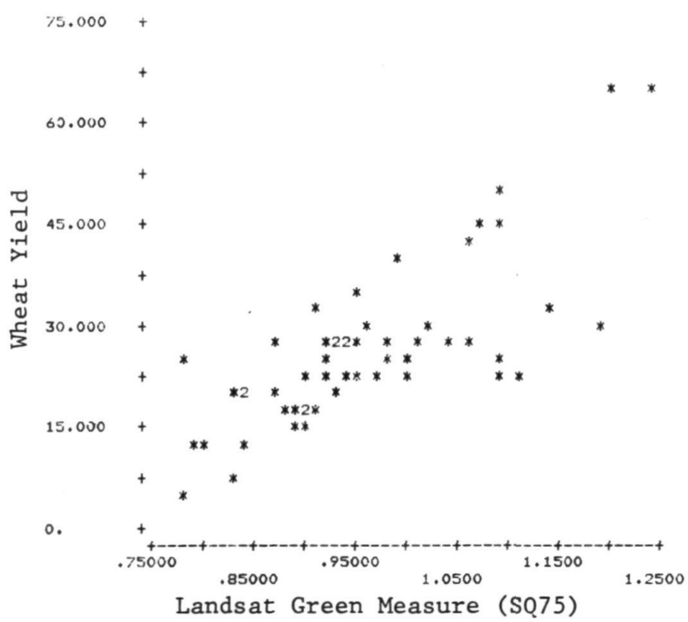

FIGURE 3. LANDSAT GREEN MEASURE VS WHEAT YIELD

TABLE I. CORRELATION OF ERIM MEASUREMENTS OF PERCENT GREEN WHEAT COVER WITH TWO OTHER GREEN COVER MEASURES

	Site A	Site B
ASCS	0.52	0.71
Landsat Green Measure	0.93	0.97

TABLE II. CORRELATIONS OF FARMERS' YIELD WITH FIELD ESTIMATES AND LANDSAT ESTIMATES OF YIELD

Yield Estimator	Site A	Site B	Site C	Average
FCIC*	0.95[1]	0.26[1]	0.74[1]	0.65
Stand Quality**	0.47[1]	0.78[1]	0.89[1]	0.71
Landsat (4 Bands)	0.94[2]	0.80[4]	0.79[3]	0.84
Landsat (TVI)	0.93[2]	0.79[4]	0.64[3]	0.79

Dates when estimators were available:

[1]Pre-harvest (mid-late June); [2]15 April; [3]21 May; [4]6 May

*Federal Crop Insurance Corporation objective estimates.

**Agricultural Stabilization and Conservation Service subjective estimates.

TABLE III. PERCENT OF VARIANCE IN YIELD ACCOUNTED FOR SEPARATELY BY SEVERAL CULTURAL FACTORS

Cultural Factors	Percent of Variance
Planting Date	0.1
Wheat Variety	10.6
Fallow Previous Year (yes/no)	35.8
Irrigation (yes/no)	56.3
Fertilization (yes/no)	55.0
Amount Fertilization (lb/acre)	57.4

TABLE IV. PERCENT OF VARIANCE IN YIELD ACCOUNTED FOR BY SEVERAL COMBINATIONS OF CULTURAL AND LANDSAT VARIABLES

Variables	Percent Variance	Standard Error
1-6 (all cultural vars)	74.9	6.89
7-10 (all Landsat vars)	87.3	4.78
4,5,7,10 (optimum four vars)	90.7	4.10
1-10 (all vars)	93.6	3.65

Variable Key:

1 = variety	6 = amount fertilizer
2 = irrigation	7 = SQ75 (May 6)
3 = fertilization	8 = SQ75 (June 2)
4 = planting date	9 = SQ75 (June 12)
5 = cropping	10 = SQ75 (April 18)

TABLE V. TWO TESTS OF EXTENSIONS OF LANDSAT WHEAT YIELD PREDICTION

From	To	Landsat Predictor	RMS Error[1] Local	RMS Error[1] Non-Local	Bias[2]
21 May Site A	20 May Site A	4 Bands	4.40	6.70	-5.00
		SQ75[3]	5.24	5.08	0.00
		TVI[4]	5.03	4.88	0.02
18 April Site A	18 April Site B	4 Bands	7.41	9.10	-0.23
		SQ75[3]	8.12	10.18	2.15
		TVI[4]	7.98	9.29	1.17

[1] On field by field basis, in bushels.

[2] Average difference between actual and predicted yield, in bushels.

[3] $\sqrt{MSS7/MSS5}$

[4] $\sqrt{(MSS7-MSS5)/(MSS7+MSS5)+0.5}$

TABLE VI. RESULTS FROM SIMPLE DIRECT WHEAT PRODUCTION ESTIMATION PROCEDURE

Site	Landsat Overpass	True Production	ERIM Estimate	Error (%)
A	6 May 76	40,600 bu	42,700 bu	5.2
A	18 Apr 76	40,600 bu	42,800 bu	5.4
B	6 May 76	27,900 bu	24,700 bu	11.5
A+B	6 May 76	68,500 bu	67,400 bu	1.6