ERIM
FORMERLY WILLOW RUN LABORATORIES THE UNIVERSITY OF MICHIGAN

WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA

TYPE II PROGRESS REPORT

16 May 1977 - 15 October 1977

114800-37-L

(E78-10009) WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA Progress Report, 16 May - 15 Oct. 1977 (Environmental Research Inst. of Michigan) 13 p HC A02/MF A01 CSCL 02C G3/43

N78-10534

Unclas
00009

NASA Contract No. NAS5-22389

Prepared by
Richard F. Nalepka - Principal Investigator
John Colwell - Co-Principal Investigator
Daniel P. Rice
Patricia A. Bresnahan

Mr. G. R. Stonesifer, NASA Technical Officer/Code 902

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt Road
Greenbelt, Maryland 20771

WHEAT PRODUCTIVITY ESTIMATES USING LANDSAT DATA

TYPE II PROGRESS REPORT

16 May 1977 - 15 October 1977

The following report serves as the ninth 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 winter wheat yield:
   a. for irrigated and for non-irrigated test sites
   b. for two different years with varying weather conditions.

1.0 PROBLEMS

No significant problems were encountered during this reporting period.

2.0 SIGNIFICANT RESULTS

Large area Landsat yield estimates have been generated. These results have been compared with estimates computed using the NOAA Center for Climatic and Environmental Assessment (CCEA) meteorological yield model. Both of these estimates have been compared with Kansas Crop and Livestock Reporting Service (KCLRS) estimates of yield, in an attempt to assess the relative and absolute accuracy of the Landsat and CCEA estimates. The results to date are inconclusive.

A large area direct wheat prediction procedure has been implemented. This procedure potentially overcomes many of the serious problems (e.g., small fields and cloud cover over specific sites) being faced by other available approaches. Initial results have produced a wheat production estimate comparable with the KCLRS estimate.

3.0 ACTIVITIES

The activities we were involved in during this reporting period concerned two general topics: 1) large area yield estimates; and 2) direct Landsat wheat production estimates. The following material is a cursory description of those activities. More details will be available in the final report.

4.0 LARGE AREA YIELD ESTIMATES

In order to make a comparison between Landsat estimates of yield and meteorological (CCEA) yield model estimates of yield, we decided to make both kinds of estimates over a large area for which adequate "true" values of average yield were available. We chose to use the Central Crop Reporting District of Kansas as our basis for comparison. Yield estimates for all counties in the Central CRD are available from the Kansas Crop and Livestock Reporting Service, and they were used as the "correct" values.

In the following material, we will first discuss Landsat estimates of yield over the Central CRD, and then CCEA meteorological yield model estimates of yield. The two methods will then be compared.

4.1 LANDSAT LARGE AREA YIELD ESTIMATES

The Central Crop Reporting District (CRD) of Kansas was chosen for a Landsat large area yield prediction demonstration. One reason this area was chosen is that, of the Kansas CRD's, it best satisfied the requirement for adequate "training" data. Information on individual field yield which is necessary in order to calibrate a Landsat wheat yield relation was available for three sites within the Central CRD.

It was decided to carry out the yield prediction test using early May 1976 Landsat data. Landsat yield prediction on the test sites was based on a regression relation between the Landsat green measure, SQ75, and farmers' combine weight estimates of yield per harvested acre on the training sites.

Because of the possible variation in external effects such as atmospheric haze over the training and test sites, it is possible that a correction for such factors would be required. Accordingly, the procedure of training and testing a Landsat yield algorithm over the Central CRD was repeated using data that was corrected for amount of haze in the atmosphere by a recently developed ERIM haze normalizing program called XSTAR.

When the training algorithms were applied to the test sites, the yield predictions that resulted are shown in Table 1. The Landsat estimates appear to be sensitive to yield variation, since the uncorrected Landsat county average estimates have a variance of 4.49 compared with a variance of 6.35 for KCLRS estimates. The individual correlation between Landsat yield estimates of a particular test site and county KCLRS average yields is not large for either the uncorrected (r = 0.25) Landsat data or for the XSTAR corrected (r = 0.08) Landsat data.

It is not essential that these county estimates be highly correlated for the technique to be working, since a small sample in a county may not be representative of the whole county. What is hoped, however, is that these county samples, when appropriately aggregated, will be good indicators of average yield over the entire Central CRD. In order to investigate this possibility, the individual county yield estimates were weighted by the number of harvested acres of wheat for the respective county, and aggregated to determine an average value of yield for the Central CRD.

The Landsat average value of weighted county yields was then compared with the KCLRS average yield, using a t-test. The hypothesis was that the means are identical. This hypothesis was barely accepted at the 5 percent level for the uncorrected Landsat estimate of average yield. There appears to be a bias in the Landsat estimates of yield since most Landsat estimates were too high using both XSTAR data (+4.2 bu/acre) and using uncorrected data (+2.9 bu/acre). Apparently the source of bias was not one that could be corrected by only accounting for atmospheric effects (haze).

TABLE 1

KCLRS Actual Yields and Landsat Predicted Yields (unweighted)

County	Actual Yield (KCLRS)	Landsat Yield Uncorrected	Landsat Yield XSTAR Corrected
Saline	27.5	37.8	38.0
Ellis	30.6	33.7	34.6
Marion	29.3	30.7	31.3
McPherson	28.5	34.1	36.3
Rush	30.8	35.1	32.5
Rice	34.3	35.2	37.8
Russell	34.5	32.9	33.5
Ellsworth	30.5	32.9	36.1
County Average	30.8	33.7	34.7
Standard Deviation	2.5	2.1	2.5

4.2 AGROMETEOROLOGICAL MODEL YIELD ESTIMATES

Since agrometeorological yield models are so frequently used, they are in some sense a yardstick with which to evaluate alternative approaches. In the following sections we will describe the results of implementation of an agromet yield model, and we will subsequently compare those results with Landsat results.

The agrometeorological yield model which was implemented for this project was a model developed to operate in Kansas by the Center for Climatic and Environmental Assessment (CCEA) of the National Oceanographic and Atmospheric Administration*. The model was implemented for the Central Crop Reporting District (CRD) of Kansas using all readily available data from meteorological stations scattered through the central CRD. We chose to implement the model for May truncation, since we intended to examine late April and early May Landsat data, and since for April truncation no CCEA model was available.

After the CCEA estimates were calculated for each meteorological station, an average value was obtained for each county with more than one meteorological station.

It was possible to get complete weather data from ten of the meteorological stations located in the central CRD. CCEA agromet model estimates of yield were calculated for the ten meteorological stations and compared with KCLRS county estimates. The unweighted CCEA estimates and the KCLRS estimates were found to have a non-significant correlation ($r = 0.09$). Less than 1 percent of the variance in KCLRS estimates is accounted for by the agromet (CCEA) estimates.

The CCEA estimates are very stable, or conservative. The variance in county CCEA estimates is 1.01, whereas the variance in KCLRS county estimates is 6.35. One might have expected the point samples (CCEA estimates generally from single meteorological stations) to be more variable than large area averages (KCLRS county estimates). However, the

* NASA/JSC, 1975, "Wheat Yield Models for the United States", LACIE-00431, JSC-11656.

CCEA agromet perturbation model is not very sensitive to changes in weather. An additional example of this relative insensitivity is that if there had been no precipitation between August and February, the CCEA model would have predicted a yield reduction from normal yield of only 3.7 bu/acre. In reality, such a situation would likely have had catastrophic effects in yield.

The individual county sample estimates of yield were subsequently weighted by the wheat acreage harvested (in 1976) in the county corresponding with the meteorological station(s). The estimates were then aggregated to a single estimate for the central CRD, as was done using Landsat estimates. Despite the apparent insensitivity of the CCEA model to meteorological variations (or perhaps because of it), and despite the low correlation between CCEA and KCLRS estimates, the average weighted CCEA value of yield is not far removed from the KCLRS estimate. The difference is 1.6 bu/acre, which has a P-value of 0.18. Therefore, we accept the estimate of yield as being not statistically significantly different from yield.

The above discussion indicates both the advantages and disadvantages of an agromet perturbation model of the type implemented. Its stability and relative freedom from a constant bias generally guarantees that it will not be far in error in reasonably "normal" years. However, its conservativeness also precludes it from adequately reflecting the effects on yield of large deviations from normal weather.

4.3 COMPARISON OF AGROMET AND LANDSAT ESTIMATES

The preceding discussion (Sections 4.1, 4.2) may not furnish us with definitive answers that reflect the general relative utility of agromet and Landsat yield estimates. For example, whether individual county estimated of yield using the two techniques are correlated with KCLRS yield or not may not be terribly relevant, because of the sampling schemes used. Similarly, the accuracy of prediction of weighted average yield is not necessarily definitive. This is due to the fact that most of the "information" in this particular test seems to be in the acreage weighting factors, which have substantially larger coefficient of variation than do the yield estimates (Table 2). Therefore, the county with the largest harvested wheat acreage tends to have the largest weighted yield estimate, regardless of the type of yield estimate (Landsat, KCLRS, or CCEA), and conversely for the county with the smallest harvested wheat acreage. This situation results in, for example, unweighted CCEA estimates having a correlation with KCLRD estimates of 0.09, and the corresponding weighted estimates having a correlation of 0.92.

Despite these difficulties in interpretation, the results do shed some light on characteristics of the two approaches that might be fairly general in nature. Specifically, the agromet model is characterized by relative lack of consistent yield error (bias) and by insensitivity to large changes in yield. The present Landsat model is characterized by potentially large yield bias and by high sensitivity to changes in yield. In other words, either approach has advantages and disadvantages. Either approach might be modified to reduce its disadvantages. Also, agromet and Landsat information could be used together to estimate yield. These possibilities are being further explored, and will be documented in the final report.

5.0 DIRECT LANDSAT WHEAT PRODUCTION ESTIMATES

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

TABLE 2

Coefficient of Variation (σ/m) for Production-Related Parameters from Counties Within the Central CRD

Parameter	σ/m
KCLRS	0.08
CCEA	0.03
LANDSAT	0.06
Acreage	0.22

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 specifiy 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 percent 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

---

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

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 percent. In addition we applied the same procedure to a different site (Site B) again using 6 May 1976 Landsat data. The resulting production estimates for these tests are shown in Table 3. Note that the total production estimated for the two tests with separate training was within 1.6 percent of the correct total production, well within LACIE desired accuracy*.

A further test of the Landsat direct wheat production approach was performed over ten of the eleven counties of the entire central CRD** using mid-April data. In this case, training was accomplished using six sites, and the "test" data set consisted of a 7.7 percent sample of the available Landsat pixels over the ten counties of the central CRD***. The resulting Landsat production estimate for the ten counties was within 2.6 percent of the final KCLRS estimate. This error is less than the spread of the several preliminary KCLRS estimates made in the months following April.

---

* MacDonald, R. B., Hall, F. G., and Erb, R. B., "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, 1975.

** No Landsat data was available for the Southeastern-most county.

*** Further details will be given in the final report.

TABLE 3

Initial Result from ERIM Direct Wheat Production Forecast Procedure
(Two LACIE Intensive Test Sites)

Site	LANDSAT Overpass	True Production	ERIM Production Forecast	Error (%)
A	6 May 76	40,600 bu	42,700 bu	5.2
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

The above tests of the ERIM Landsat wheat production estimation system are certainly not definitive. Many more tests in different situations need to be carried out in order to assess the consistency of performance. However, the preliminary indications based on our limited tests give encouragement that the direct wheat production approach using early-season Landsat data is worth pursuing.

6.0 FUTURE PLANS

Technical efforts of this project are in the final stages. These efforts will be fully documented in a final report, which is currently being prepared.

7.0 PUBLICATIONS/PRESENTATIONS

Mr. R. F. Nalepka and Dr. J. E. Colwell attended the Landsat Crop Condition and Yield Briefing held at NASA Headquarters on September 27, 1977. Mr. Nalepka presented a review of our activities entitled "World-wide Wheat Production Forecasts Using Landsat Data".