Reprinted from

Eighth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Crop Inventory and Monitoring

July 7-9, 1982

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

LARGE AREA RELATION OF SATELLITE SPECTRAL DATA TO WHEAT YIELDS

T.L. BARNETT, D.R. THOMPSON

National Aeronautics and Space Administration/Johnson Space Center
Houston, Texas

I. INTRODUCTION

Since the advent of the first Landsat satellite in the early 1970s, the prospect of using satellite-derived spectral information in making large-area estimates of crop yields has been attractive. However, it has remained an elusive goal, possibly because the data sets analyzed have been too restricted. Efforts have focused primarily on Landsat since its high resolution permits acquisition of spectral data from individual fields. For surveys of large-area general crop conditions, however, this high resolution may not be necessary. The investigations reported here used Landsat data averaged over 5nmi x 6nmi sample segments throughout the entire U.S. Great Plains winter and spring wheat regions for two years. They provide evidence of a real linear relationship between wheat yield and satellite spectral data over these large regions. Tests at both a smaller scale on sets of field-level spectral data and yield, and at a larger scale on 25mi x 25mi gridded spectral data from the NOAA-6 environmental satellite appear to confirm the relation.

A great deal of effort has been directed toward relating spectral information to readily measurable properties of crops such as biomass, leaf area index, or percent cover (e.g., Wiegand et al.10; Tucker et al.9, and Aase and Siddoway1). Such parameters are potentially available from satellite observations and would be useful inputs to yield models. Early efforts attempted to correlate Landsat spectral data directly to grain yield (e.g., Colwell et al.3). Generally, strong correlations were found in the development data set, but these did not hold up well when extended in space or time. Idso et al. 4 suggested using albedo at ripe stage to predict wheat yield. This, however, is rather late in the season to be useful, and would require high resolution satellite data.

The most promising progress reported to date in direct use of spectral data in yield modeling has been that of Schubert and Mack 6 who used Landsat data to calculate percent cover indices - the percentages of pixels within an area of particular vegetation type (wheat or all vegetation) that were classified as having a closed green canopy - for three phenological stages labeled "emerged," "headed," and "ripe." In regressions of these indices to yields over three years (1973-1975) at four test sites in different soil/climate regions of western Canada, the correlations were quite high for the first two crop stages. A test of predicted vs. published yields in a later year (1976) showed satisfactory results only for the wheat index. Application would thus require identification and classification of wheat fields using high resolution satellite data.

II. DEVELOPMENT OF THE YIELD-TO-GREENNESS RELATION

The relation of yield to spectral data was developed from a large set of Landsat observations at "sample segments" - 5 nmi x 6 nmi sites at which Landsat data were acquired and processed in the U.S. Great Plains states of Texas, Oklahoma, Colorado, Kansas, Nebraska, South Dakota, North Dakota, Minnesota, and Montana during 1978 and 1979. A total of 414 segments selected for analysis had enough clear acquisitions to accurately track the spectral response of winter and spring wheat throughout the growing season. This large number of useable segments was made possible by the two Landsat satellites in orbit simultaneously, providing a nine-day coverage cycle.

The Landsat data set consisted of Green Index Numbers (GIN) for the sample segments. GIN represents the percent of Landsat pixels in the sample segment with "greenness" greater than a specified value. Greenness was defined by Kauth and Thomas 5 to be the following linear combination of the four Landsat MultiSpectral Scanner (MSS) bands:

$$g = -.29MSS1 - .56MSS2 + .60MSS3 + .49MSS4.$$

A "soil line," the minimum value of g, had been subtracted off the above definition of greenness. In calculating greenness values for the test cases cited later, this soil line value was not subtracted off. However, the term "greenness" is used in both cases because the conclusions should be independent of the precise definition of greenness and should, in fact, apply equally well to a variety of spectral parameters. Values of GIN5, GIN10, GIN15, GIN20, and GIN30 (representing percent of pixels with greenness greater than 5, 10, 15, 20, and 30 respectively) provided information on the greenness distribution.

Thompson and Wehmanen 7,8 correlated GIN15 to serious drought conditions in the U.S Great Plains in 1978 and 1979 with impressive results. GIN15 appeared to be a reasonable parameter to relate to yield. This was done in a preliminary paper (Barnett and Thompson 2). However, GIN15 has the drawback that it is available only from a high resolution sensor like the Landsat MSS, and then only at considerable effort. In order to derive a more generally useful relation and to permit comparison to other spectral data sets, the GIN values were converted for this study to estimates of segment-average greenness for all 414 segments. The technique used to convert GIN values to greenness is described in Appendix 1.

The extensive spectral data set offered a much greater amount and range of data than had been examined previously for such a purpose. A relation was sought between the average greenness for sample segments and the yields for the counties in which the segments were located. Since the sample segments had been chosen during the Large Area Crop Inventory Experiment (LACIE) in major crop areas, the spectral responses were generally dominated by wheat but also included substantial amounts of other vegetation and soil in the scene.

Greenness at heading appeared to be the most appropriate parameter for large-area correlation to yield. Heading is easily detected on a plot of greenness vs. day-of-year (DOY) since it represents the peak of the curve. For each of the 414 winter and spring wheat segments greenness was plotted vs. DOY. Heading was determined rather accurately in each case by correlating information published in the weekly Crop Weather Bulletins (USDA state publications) with the best of the curves in the vicinity which exhibited strong greenness peaks. This procedure permitted estimation of greenness at heading even for those cases that exhibited an indistinct peak or no peak at all due to substantial amounts of vegetation other than wheat in the scene.

A. WINTER WHEAT 1978

Winter wheat for 1978 contained the most observations and presented the widest range of yield, stress, and geography. It consisted of 196 sample segments in the winter wheat areas of Texas, Oklahoma, Colorado, Kansas, Nebraska, South Dakota, and Montana. Weekly maps of the Crop Mositure Index in the Weekly Weather and Crop Bulletin (USDA publication) indicated that much of the winter wheat region in 1978 experienced good weather and little stress. However a drought covered parts of the southern Great Plains throughout the winter and most of the growing season, ranging from moderate in southwestern Nebraska, the western third of Kansas, and parts of Colorado, to severe in the Oklahoma and Texas panhandles, to near-disastrous in the Texas plains and plateau regions. Yields per harvested acre ranged from 40 Bu/A in the best areas to less than 6 Bu/A in the most seriously affected Texas counties. In Figure 1 the county yields are plotted against segment-average greenness for those counties containing segments across the 1978 winter wheat region. Although these is a wide range of yields with quite a few points at low yield/low greenness, the majority of points are clustered in the high yield/high greenness region. A least squares fit to these county-level points is dominated by the latter cluster.

To reduce the scatter and the preponderance of points at high yield/high greenness, the segment-average greenness values were averaged over the Crop Reporting Districts (CRDs) in which they were located for Texas, Oklahoma, and Kansas. For Colorado, Nebraska, South Dakota, and Montana there were too few segments to make reasonable estimates of CRD-level greenness. Greenness values were averaged over these states. The yields at CRD or state level were plotted against greenness at heading in Figure 2, giving a much more reasonable distribution of points. A linear least squares fit to these points gave

$$\hat{y} = -2.10 + 1.66\ g,$$

with correlation coefficient $r = 0.88$, standard deviation of fit 2.0 Bu/A, and 90% confidence intervals on the intercept (-2.10 ± 5.36) and slope (1.66 ± 0.14).

B. SPRING WHEAT 1979

Spring wheat for 1979 also offered a large range of yield and stress in 64 sample segments across Minnesota, North Dakota, South Dakota, and Montana. The plot of county yields vs. segment-average greenness at heading for counties containing the segments is given in Figure 3. There were not enough sample segments to permit meaningful averaging to CRD level; however, the county-level scatter was relatively small and the distribution was fairly even. A least squares fit to these points gave

$$\hat{y} = -6.77 + 1.62\ g,$$

with a correlation coefficient $r = 0.87$, a standard deviation of fit 1.6 Bu/A, and 90% confidence intervals on the intercept (-6.77 ± 4.07) and slope (1.62 ± 0.19).

The slopes for both the 1978 winter wheat and the 1979 spring wheat fits at regional level are nearly identical. The intercepts are statistically indistinguishable, although the confidence interval is large enough to leave the intercept as an open question. There appears to be a common relation between yield and greenness at regional level for wheat, at least when stress is an important factor.

C. WINTER WHEAT 1979

Winter wheat in 1979 was represented by 65 sample segments in Texas, Oklahoma, Kansas, Nebraska, and South Dakota for which good estimates of greenness at heading were available. A plot of county-level yield vs. segment-average greenness at heading for counties containing sample segments is shown in Figure 4. No reasonable estimates at CRD level were possible because of the relatively small number of points, the considerable scatter, and the relatively small yield range, due to generally good weather in 1979.

For these reasons a linear least squares fit to the data did not appear appropriate. However, as shown in Figure 4, the linear fit for 1978 winter wheat, with slope 1.66 (Bu/A)/(unit greenness) and intercept -2.10 Bu/A, fits the 1979 winter wheat data reasonably well. the best fit would actually be a parallel line a few Bu/A higher in yield. The intercept of such a line would fall within the confidence interval on the 1978 winter wheat fit, so the question of intercept cannot be settled. One possible reason for the apparent shift of the intercept value may be the fact that the sample segments in 1979 tended to be concentrated more in the eastern parts of the winter wheat region than was the case in 1978.

B. SPRING WHEAT 1978

Spring wheat in 1978 was represented by 88 sample segments in Minnesota, North Dakota, South Dakota, and Montana for which greenness at heading could be estimated. Figure 5 displays the plot of county-level yield vs. segment-average greenness for counties containing segments. There is a relatively small range of yield, much scatter, and no discernable relation between yield and greenness. This was apparently due to the excellent weather and soil moisture conditions over the entire spring wheat region and the consequent lack of stress. This particular set of data indicates that a yield/greenness relation can be meaningful only when there is at least a moderate degree of stress in at least part of the region of interest.

III. TESTING AND VERIFICATION

The hypothesis to be tested was that there exists a common relation between large-area yield and greenness for wheat. The first test was conducted on a test set at a smaller scale than the development set. Sets of field-level yields, as reported by the farmer or other reliable sources, were collected as part of the LACIE program at selected sample segments for winter and spring wheat from 1975 through 1978. These were correlated with field-level Landsat data for clear acquisitions during these years, forming a "spectral-met" data base.

Although fields within a given segment experienced nearly the same weather, different levels of irrigation and fertilization produced large ranges of yields and greenness. The purity of the spectral data for the fields removed the complicating factor of other vegetation or soil in the scene. This was a particularly strong test for two reasons. First, the greenness values were taken at the Landsat acquisition date nearest to heading, and were not interpolated to heading. This tested the applicability of the relation at stages other than precisely heading. Second, many of the yields for individual fields were in the range 50-70 Bu/A with a few greater than 80 Bu/A. This tested the yield/greenness relation at yields up to twice those in the development set.

Fifteen sample segments, including winter and spring wheat over several years, displayed a field yield range of 25 Bu/A or more and had at least twelve fields in the segment. These were considered to represent a set of test cases. Landsat MSS band averages over the fields were converted to estimates of greenness through the Kauth-Thomas formula. The "soil line" minimum greenness value could not be calculated from the available information and was not subtracted off. For each of the fifteen test cases the field-level yields were plotted against the field greenness values. A line of slope 1.66 (Bu/A)/(unit greenness), corresponding to the 1978 winter wheat fit, was superimposed on each plot and shifted up or down in yield to give the best fit. Since there were several biases built into the test cases-the observation dates were not all at heading, the greenness values did not have the soil line removed, and there were variations between field yields from different sources - the intercepts were not expected to match the intercept of the 1978 regional fit.

The 1978 winter wheat slope of 1.66 used in this manner satisfied all fifteen of the test cases quite well. Four examples are shown in Figure 6 a-d, ranging from the best cases (Figures 6 a and 6 c) to average (Figure 6 b) to worst case (Figure 6 d). This appears to provide strong evidence of a real physical or agronomic relation between yield and greenness, at least as regards the slope of the relation.

A second test of the yield/greenness relation at a larger scale than the development data set was provided by 1981 spectral data from the NOAA-6 Advanced Very High Resolution Radiometer (AVHRR). This sensor has an instantaneous resolution of about one kilometer in several spectral bands, including one in the visible corresponding approximately to Landsat bands MSS1 + MSS2, and one in the near infrared corresponding to MSS3 + MSS4. The spectral parameter Environmental Vegetation Index (EVI) is defined as the band 2 signal minus the band 1 signal, and corresponds closely to the definition of greenness. The 1981 spectral data set consisted of EVI values averaged over 25 mi x 25 mi grid cells covering the wheat regions of Kansas, Oklahoma, and Texas. For this test, EVI grid cell values were averaged over CRDs in these states and converted to estimates of greenness as described in Appendix 2. The yield/greenness relation from 1978 winter wheat,

$$\hat{y} = 2.10 + 1.66\ g,$$

was then applied to the estimated CRD greenness values to estimate 1981 yields at CRD level. A scatter diagram of the results, Figure 7, plotting predicted 1981 yields vs. reported USDA yields shows that the 1978 relation does a remarkably good job. While a single good test cannot be regarded as proving the relation as a yield model, it certainly provides further support for the relation.

IV. CONCLUSIONS

Analysis of spectral data from Landsat for winter and spring wheat across the U.S. Great Plains in 1978 and 1979 indicates that a linear relation exists between yield and greenness, common to both types of wheat, whenever there is an appreciable degree of stress. Tests on both smaller and larger scales strongly support the concept of a common slope but are less clear with regard to the intercept of the relation.

Of particular importance is the fact that the yield/greenness relation seems to apply to spectral data of all vegetation in a large scene, and not merely to wheat pixels carefully identified and extracted from high resolution satellite data. The potential applications to global surveys of yield potential reasonably early in the season are obvious.

Further testing is required to determine how the intercept is affected by vegetation background or by different levels of yield potential such as might be encountered in foreign regions, to determine if the slope has the same value for regions other than the Great Plains, and to determine if similar relations exist for other crops. If further analysis supports the yield/greenness relation, yield models involving significant spectral inputs may become a reality.

APPENDIX 1

CONVERSION OF GIN TO GREENNESS

Assume we are given values of GIN5, GIN10, GIN15, GIN20, and GIN30. GIN5 represents the the percent of pixels in the sample segment with greenness greater than 5, GIN10 the percent of pixels with greenness greater than 10, etc. One may interpret this as saying that there are (GIN30)% of the pixels with $g > 30$, (GIN20 - GIN30)% with $30 > g > 20$, (GIN15 - GIN20)% with $20 > g > 15$, (GIN10 - GIN15)% with $15 > g > 10$, (GIN5 - GIN10)% with $10 > g > 5$, and (100 - GIN5)% with $5 > g > 0$. If we make the assumption that the average greenness of those pixels with $g > 30$ is 32, the average greenness of those pixels with $30 > g > 20$ is $\frac{1}{2}(30+20) = 25$, etc., then we can estimate the average greenness of the sample segment as:

$$g = [32 \times GIN30 + 25 \times (GIN20-GIN30) + 17.5 \times (GIN15-GIN20) + 12.5 \times (GIN10-GIN15) + 7.5 \times (GIN5 - GIN10) + 2.5 \times (100 - GIN5)]/100.$$

APPENDIX II

RELATION OF GREENNESS TO EVI

No information on the relation of greenness to EVI over large areas was known to exist. An empirical relation was therefore established as a straight line between the following two (greenness, EVI) points:

a. values for essentially bare soil (pre-emergence acquisition) for 1978 segment-average greenness and 1981 EVI 25 mi x 25 mi grid data, and
b. 1978 greenness at heading and 1981 EVI at heading averaged over all observations in Kansas and Oklahoma (both years represented good growing conditions in these states).

This resulted in a large-scale relation of greenness to EVI of

$$\hat{g} = 0.54 + 0.15\ EVI.$$

REFERENCES

1. Aase, J.K. and Siddoway, F.H. (1981), Assessing winter wheat dry matter production via spectral reflectance measurements. Remote Sensing of Environment 11:267-277.
2. Barnett, T.L. and Thompson, D.R. (1982), The use of large-area spectral data in yield estimation, submitted to Remote Sensing of Environment.
3. Colwell, J.E.; Rice, D.P.; and Nalepka, R.F., (1977), Wheat yield forecasts using Landsat data. Proceedings of the Eleventh International Symposium on Remote Sensing of Environment, pp. 1245-1254.
4. Idso, S.B.; Hatfield, J.L.; Reginato, R.J.; and Jackson, R.D. (1978), Wheat yield estimates by albedo measurement, Remote sensing of Environment, 7:273-276.
5. Kauth, R.J. and Thomas, G.S. (1976), The tasselled cap - a graphic description of the spectral-temporal development of agricultural

crops as seen by Landsat. Proceedings of the Symposium on Machine Processing of Remote Sensing at Purdue University, pp. 4B-41 to 4B-51.

6. Schubert, J. and Mack, A.R. (1977), Global agricultural productivity evaluation from Landsat data. Proceedings of the W. Nordberg Memorial Symposium of the Twentieth Plenary Symposium of COSPAR (Tel Aviv, Israel). Published by Pergamon Press for the Committee on Space Research, pp. 95-99.
7. Thompson, D.R. and Wehmanen, O.A., (1979), Using Landsat digital data to detect moisture stress, Photogrammetric Engineering and Remote Sensing, 45: 201-207.
8. Thompson, D.R. and Wehmanen, O.A., (1980), Using Landsat digital data soybean growing regions, Photogrammetric Engineering and Remote sensing, 46: 1087-1093.
9. Tucker, C.J.; Holben, B.N.; Elgin, J.H. Jr.; and McMurtrey, J.E. III, (1980), Relationship of spectral data to grain yield variation. Photogrammetric Engineering and Remote Sensing, 46: 657-666.
10. Wiegand, C.L.; Richardson, A.L.; and Kanemasu, E.T., (1979), Leaf area index estimates from Landsat and their implications for evapotranspiration and crop modeling, Agronomy Journal, 71: 336-342.

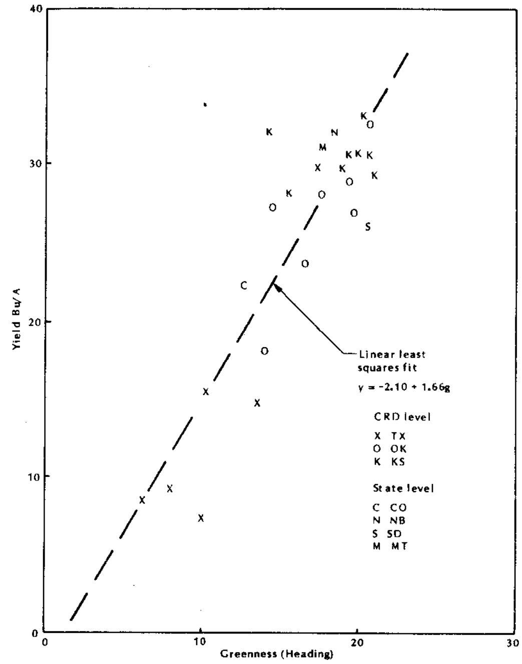

Figure 2. 1978 winter wheat yields vs. greenness at heading for CRD and state levels.

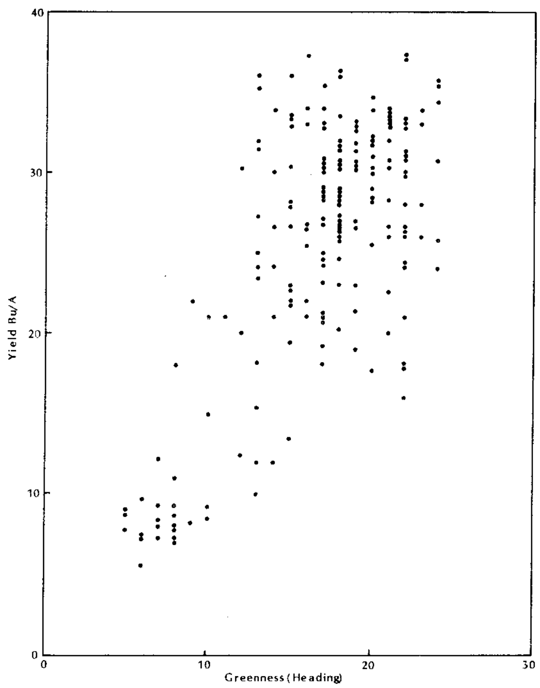

Figure 1. 1978 winter wheat county yields vs. segment-average greenness at heading for TX, OK, CO, KS, NB, SD, and MT.

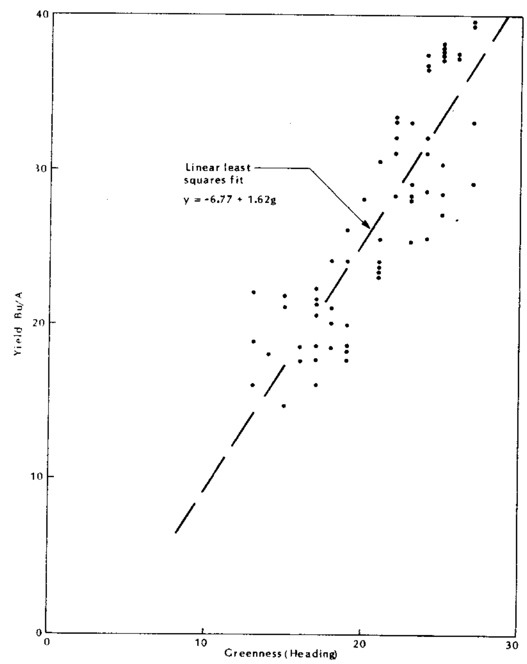

Figure 3. 1979 spring wheat county yields vs. segment-average greenness at heading for ND, SD, MN, and MT.

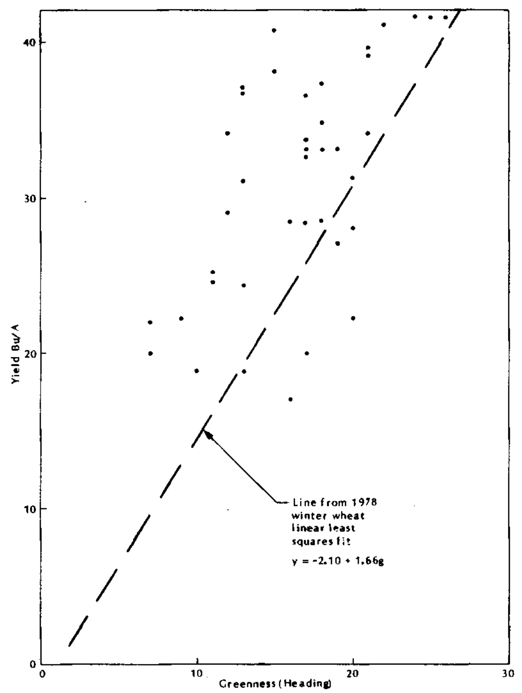

Figure 4. 1979 winter wheat county yields vs. segment-average greenness at heading for TX, OK, KS, NB, and SD.

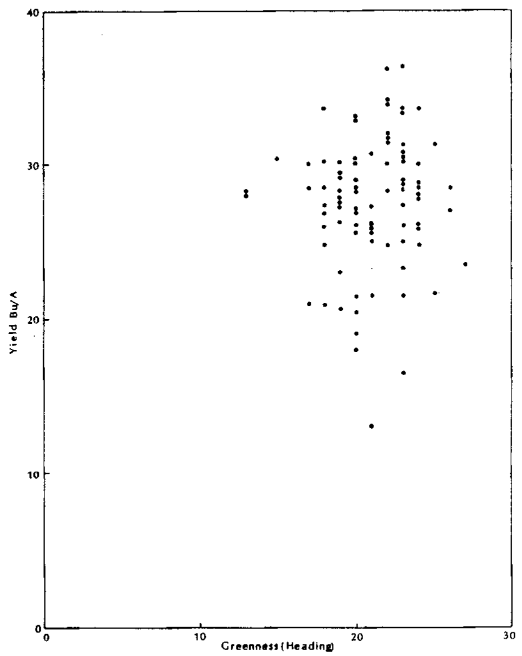

Figure 5. 1978 spring wheat county yields vs. segment-average greenness at heading for ND, SD, MN, and MT.

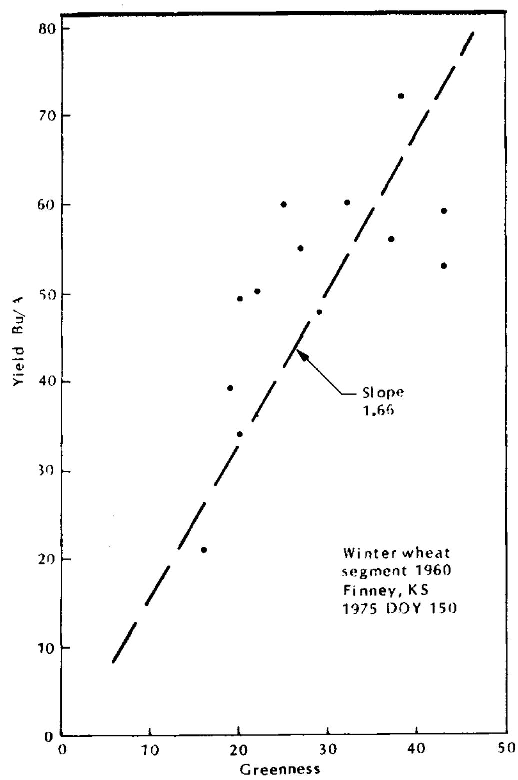

Figure 6 a. Field-level test for winter wheat, Finney Co. KS, 1975.

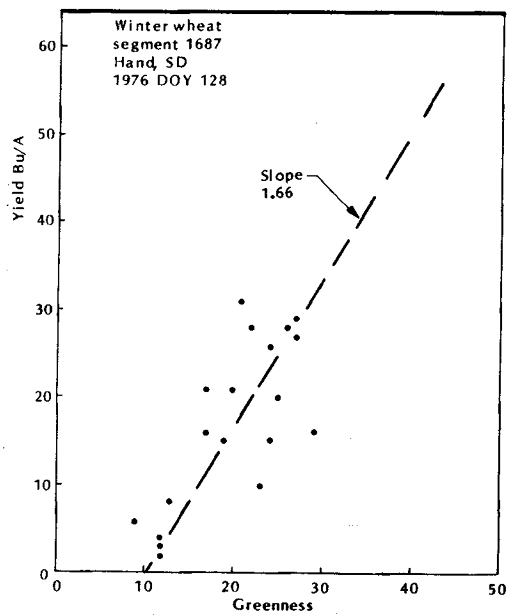

Figure 6 b. Field-level test for winter wheat, Hand Co. SD, 1976.

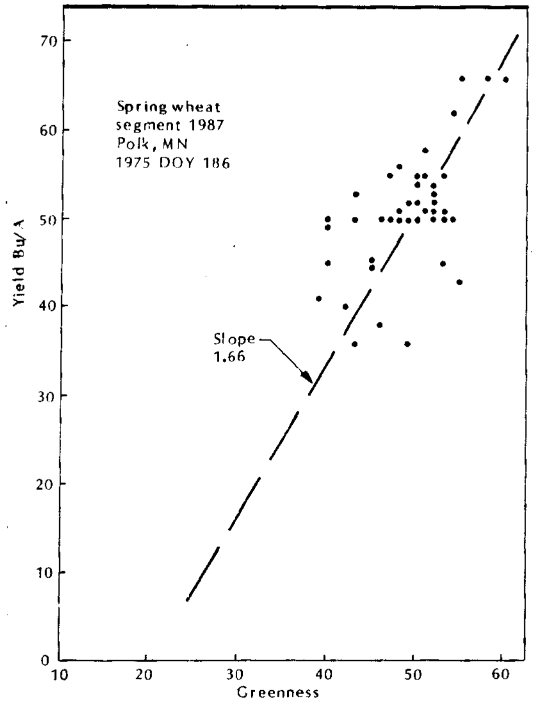

Figure 6 c. Field-level test for spring wheat, Polk Co. MN, 1975.

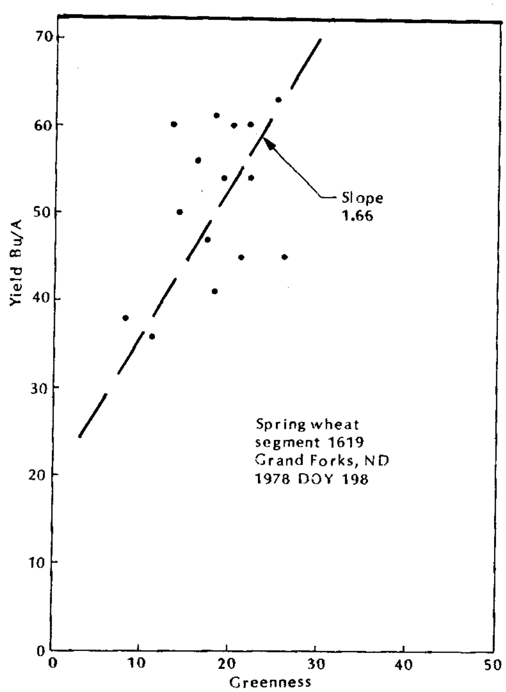

Figure 6 d. Field-level test for spring wheat, Grand Forks Co. ND, 1978.

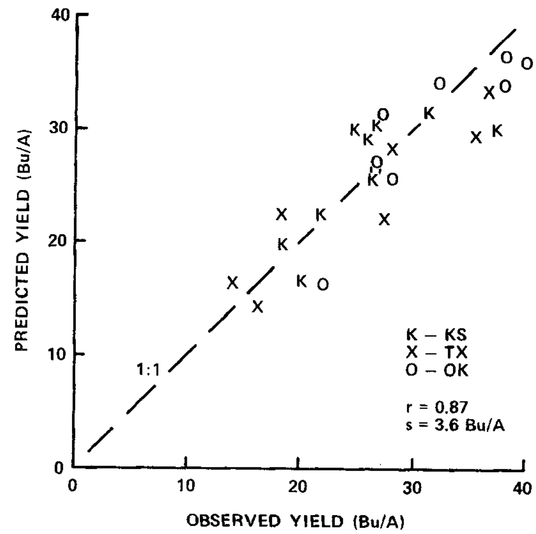

Figure 7. Plot of CRD-level yields predicted from 1981 NOAA-6 satellite data vs. 1981 SRS reported yields.

Dr. Thomas L. Barnett received his B.S. in Physics from West Virginia University and M.S. and Ph.D. in Physics from Michigan State University. He has worked for NASA - Johnson Space Center since 1967 where he has been involved in the Atmospheric Physics, Skylab, and Earth Observations programs, including LACIE and AgRISTARS. He is currently assigned to the AgRISTARS Yield Model Development Project at Columbia, MO, working on development of spectral inputs to yield estimation.

David R. Thompson received B.S., M.S. and Ph.D. degrees in agronomy from Texas A & M University. He worked several years for the USDA Soil Conservation Service. Since 1975 he has been an agronomist at the NASA Johnson Space Center where he conducted research on the effect of drought on the spectral response of crops during the Large Area Crop Inventory Experiment. Present research involves the development of crop development stage and yield models utilizing Landsat MSS data as part of the AgRISTARS program, and application of spectral data for soils research.