1-1-1982

Estimating Crop Development Stages from Multispectral Data

J. C. Tilton

S. E. Hollinger

LARS Technical Report 072082

ESTIMATING CROP DEVELOPMENT STAGES FROM MULTISPECTRAL DATA

J.C. TILTON, S.E. HOLLINGER

Purdue University/Laboratory for Applications of Remote Sensing
West Lafayette, Indiana

ABSTRACT

Accurate prediction of crop yields requires knowledge of the crop's development stage at critical times during the growing season. Various meteorological models have been developed to estimate crop development. However, the crop planting date, or equivalently, the date the crop reaches a particular crop development stage generally is needed to initialize these models. Recently, efforts have been made towards developing methods for estimating crop development stage from Landsat MSS data. We have developed a technique that is designed to estimate the calendar day that a crop reaches a particular development stage early or late in the growing season. The method requires Landsat observations from the first half of the growing season to make the early season estimate and from the last half of the season for the late season estimate. Following a description of the crop model employed we describe the application of our method to Purdue Agronomy Farm data collected by a truck-mounted Exotech-100 radiometer and to Landsat MSS data. Preliminary results indicate that the method has good potential for making accurate estimates of the calendar day that a crop reaches a particular crop development stage (depending on training data).

I. INTRODUCTION

Accurate prediction of crop yields requires knowledge of the crop's development stage at critical times during the growing season. Development stage as defined here describes where the crop is in its life cycle. The Hanway scale(1) (Table 1) is commonly used to describe corn development and the Fehr-Caviness scheme(2) to describe Soybean development. There are several other scales which are used to describe other crops.

Table 1. Hanway corn development stages.

Stage Number	Stage Name
-1.00	PREPLANT
0.00	PLANTED
0.10	EMERGED
0.25	1 LEAF
1.00	4 LEAVES
2.00	8 LEAVES
3.00	12 LEAVES
4.00	16 LEAVES
4.50	TASSELED
5.00	SILKED
6.00	BLISTER
6.50	MILK
7.00	DOUGH
8.00	BEGIN DENT
9.00	FULL DENT
10.00	PHYSIOLOGIC MATURITY
10.50	HARVEST MATURITY
11.00	HARVESTED

If the calendar day a crop reaches certain development stages is known, the stress a crop is experiencing can be assessed fairly accurately from weather information. Crop yields can then be predicted from this knowledge of the level and type of stresses a crop endures during its development.

Various meteorological models have been developed to estimate the calendar day a crop reaches particular development stages. The most common methods involve the calculation of a thermal unit or a photothermal unit. The thermal unit or growing degree unit is calculated by summing the difference between the daily mean temperature and some threshold temperature. The modified growing degree unit developed by Gilmore and Rogers(3) is the most commonly used method to estimate corn development stages in the United States.

The thermal unit method of estimating crop development stages requires the planting date of a field and the temperature experienced by that field as inputs. The planting date for any given field in a large area is usually not known, so the average planting date for a state or crop reporting district (CRD) is often used. Temperature data for a given area is available from only one or at best three stations. Therefore, the mean temperature for a CRD is often used to describe the temperature regime for the entire area. This practice of using one planting date and temperature value gives an estimate of the mean development stage within a large area, but fails to fully describe the range and variation of development stages within the area. If remotely sensed spectral data from satellite or aircraft can provide an estimate of the spatial variation of planting dates and/or development stages over large areas, we would be able to make a more accurate estimate of the yield variation within a given region.

Considerable effort has been devoted to develop methods of estimating development stage using remote sensing. A spectral-temporal profile model using spectral data to describe development stage throughout the season has been developed recently by Badhwar and Henderson(4). The model has shown promise in accurately estimating development stages of corn and soybeans. However, the model requires a minimum of five acquisitions spread throughout the growing season to depict development. This becomes a problem because development stage cannot be described until after the end of the growing season and the value of the information is greatly reduced as far as assessing yield potential during the growing season is concerned. Because of this limitation, we have pursued the development of a model to give a spectral estimate of development stage early in the crop season. This model can also be used to estimate development stages late in the crop season. The model has its biggest advantage in early season development stage estimation in that observations are only required through mid-season rather than through the entire crop season. Another advantage of our model is that it does not require the computationally-intensive curve-fitting required by the Badhwar and Henderson model.

II. MULTISPECTRAL CROP MODELING

The spectral response of a crop canopy, as measured by Landsat-type Multispectral Scanners (MSSs), changes in a typical manner throughout the growing season, depending on the crop type. The greenness component of the Kauth and Thomas tasseled-cap transformation(5) exhibits these changes particularly well. The greenness and brightness components of the tasseled-cap transformation are the two largest components obtained from a principle components analysis of four channel Landsat MSS data. The greenness component correlates with the amount of green vegetation present while the brightness component correlates with the overall brightness of the scene (often the brightness of the underlying soil). We will study the behavior of the greenness component of the tasseled-cap transformation for individual pixels or for field averages, which we will refer to as the green number for that particular pixel or field.

A typical plot for corn of green number versus calendar date is shown in Figure 1. Prior to planting the green number stays essentially constant at a level we call the "soil green number." After planting the green number stays at the soil green number until sufficient vegetative matter appears above the soil, usually when two or three leaves emerge from the corn plant (Hanway development stage 0.50 or 0.75). Then the green number increases with calendar date relatively quickly until the "maximum canopy green number" value is reached, usually at about tasseling or silking (Hanway development stage 4.50 or 5.00). The green number then holds fairly constant or falls slightly as subsequent development stages occur through the beginning of denting (Hanway development stage 8.00). Then the green number falls rapidly as the corn matures until it

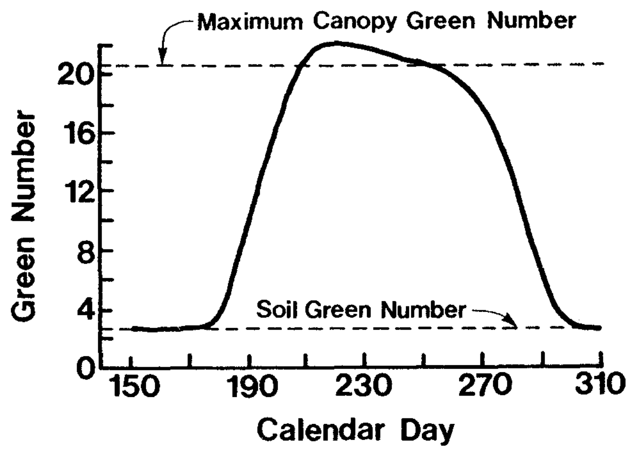

Figure 1. Typical plot of Landsat green numbers versus calendar day for a midwestern corn field.

reaches the soil green number again at harvest. Many other crops have similar green number curves with time, where the green number rises or falls rapidly over relatively short time intervals.

The crop development stage estimation technique described herein basically notes the calendar days when the green number rises and falls to values half-way between the soil green number and the maximum canopy green number. These calendar days are then correlated with particular crop development stages through the use of training data. Since the green number changes most rapidly with both time and crop development stage at these half-way rise and fall points, the development stage can be estimated most accurately at these points. Of particular interest is the crop development stage estimate for the calendar day when the green number first rises to cross the value half-way between the soil green number and maximum canopy green number. All that is required to make this estimate are MSS observations through enough of the growing season to make an estimate of the maximum canopy green number, or roughly half of the growing season. At least one observation at or prior to planting is required for the soil green number estimate. In order for good estimates to be made, the MSS observations should be made at intervals of 36 days or less.

III. CROP DEVELOPMENT STAGE ESTIMATION TECHNIQUE

A. AGRONOMY FARM DATA

Spectral data acquired over experimental plots at the Purdue University Agronomy Farm near West Lafayette, Indiana were used in the early development work on our crop development stage estimation technique. The data were collected using a truck-mounted Exotech-100 radiometer which has the same wavelength bands as the Landsat MSS. The data were calibrated and corrected for sun-angle effects. Various agronomic measures, including crop development stage, were recorded simultaneously with each radiometric observation. For a complete description of the data see Bauer, et al(6). The Kauth-Thomas greenness component of the Exotech-100 wavelength bands (B1, B2, B3 and B4) is given by the following transformation(7):

$$\text{Green number} = -0.4894*B1 - 0.6126*B2 + 0.1729*B3 + 0.5854*B4.$$

A key step in estimating crop development stages from the green number values is a process through which the soil green number and the maximum canopy green number are estimated from the data. This estimation process requires green number estimates at regular time intervals throughout the growing season. This regular interval was chosen to be nine days to make the Agronomy Farm data look more like Landsat MSS data. The shortest time interval over which repeat Landsat MSS data may be available is generally nine days. Since the MSS observations are generally available at irregular time intervals, interpolation must be used to obtain green number estimates for every nine days. The interpolator employed should be conservative (not prone to wide oscillations) since we do not expect green number variations due to crop development to have wide oscillations. Such an interpolator is the quasi-Hermite spline interpolator contained in the IMSL(C) mathematical and statistical subroutine software package. This interpolator is designed to approximate a curve drawn manually through the data points.

As noted above, Landsat MSS data are generally available no more frequently than every nine days. Purdue Agronomy Farm data may be available more frequently. Data at intervals less than about nine days may contain misleading short-term fluctuations due to such things as changes in illumination level and crop moisture level (e.g. it rained between observations). Where such short-term fluctuations occur, even the conservative quasi-Hermite spline interpolator produces green number estimates with unrealistic oscillations. (Data at widely spaced intervals also contain fluctuations due to short-term events, but these fluctuations are about a long-term trend and are not interpreted as high frequency oscillations by an interpolator.)

Because of the problems with short-term fluctuations in the data, the calculated green number values are smoothed (or filtered) to dampen out the high frequency variations suggested by the short-term fluctuations in the data. This smoothing also serves to make the Agromony Farm data look more like Landsat MSS data, since Landsat MSS data generally cannot contain fluctuations of shorter term than nine days. This smoothing is accomplished by a time-domain convolution of the data with a sinc**2 function ((sin(pi*x)/pi*x)**2). Since such a convolution makes sense only for stationary data and since the green number values for the entire growing season cannot be considered stationary, the convolution is only performed over an eighteen day window. The data can be considered to be approximately stationary over a time span of about eighteen days or less. A sinc**2 function with zeros nine days before and nine days after its peak has been found to perform well. The function is in effect set to zero by the eighteen day window for times earlier than nine days before the central peak and for times later than nine days after the central peak. Convolving with such a function does not affect green number values calculated from observations which are nine days or more apart.

The soil green number and the maximum canopy green number are estimated from the smoothed and interpolated green number estimates. These estimates are first normalized so that the minimum green number value is zero and the maximum green number value is twenty. These minimum and maximum values are arbitrary, but they are roughly the minimum and maximum values typically found in the Agronomy Farm data.

The soil green number is estimated as follows: The normalized green numbers are ordered from smallest to largest. Initially, the smallest green number is considered to be the soil green number estimate. The next largest green number is tested against the current soil green number estimate using a one-sided chi-square test with one degree of freedom. This test gives the probability that the tested green number is not an observation of the soil green number. We will refer to this as the probability that the tested green number is "above" the current soil green number estimate. If the tested green number has a probability of 50% or less of being above the soil green number estimate, the tested green number is considered to be an additional observation of the soil green number and averaged with the other soil green number observations to produce a new soil green number estimate. The above process is repeated for the next largest green number. If the tested green number has a probability of more than 50% of being above the current soil green number estimate, the current soil green number estimate is considered to be the final soil green number estimate.

The maximum canopy green number is estimated in a similar way to the method for estimating the soil green value. Here, however, the normalized soil green numbers are ordered from maximum to minimum, and the next smallest green number is tested against the current maximum canopy green number estimate for being "below" the maximum canopy estimate. The threshold probability here is taken to be 90% rather than 50% since the maximum canopy green number observations tend to be more variable than the soil green number observations. We should note that if the green number observations were not normalized, or if they were normalized differently, we would obtain different estimates for the soil green number and the maximum canopy green number because of the nature of the chi-square test, unless we would adjust our threshold probabilities appropriately.

The soil green number and maximum canopy green number estimation process assumes that each green number observation is of equal importance. This assumption is satisfied if green number estimates are taken at equal time intervals. If the green number observations are left at irregular time intervals, the estimation process would require weighting each observation according to its relative importance. The relative importance or weight of each observation could be determined by the time interval between the observation and the previous and following observations. (If the observation is the first or last observation, we would have to make some reasonable assumption.) These relative weights should be used when these green number observations are averaged together to give estimates of the soil green number or maximum canopy green number. In addition, the relative weight of each observation being tested in the chi-square test would have to be incorporated into the test. It is much easier to interpolate green number estimates at regular time intervals (as we have done) than to resort to such observation weighting.

Now that we have estimates of the soil and maximum canopy green numbers, we can make estimates of the time certain growth stages occur. A fairly typical graph of the processed green numbers is shown in Figure 2. This graph is for a plot of corn that was planted later than most other corn on the Purdue Agronomy Farm. Such late plantings typically exhibit a non-crop "green-up" such as shown here at about calendar date 160. The field was tilled and planted on calendar date 163. For this corn plot, the calendar dates that the processed green values crossed the half-way rise and fall value between the soil green number and maximum canopy green number were day 196 and day 276. The Hanway growth stage at day 196 was about 1.75 (7 leaves) and the Hanway growth stage at day 276 was about 9.00 (full dent).

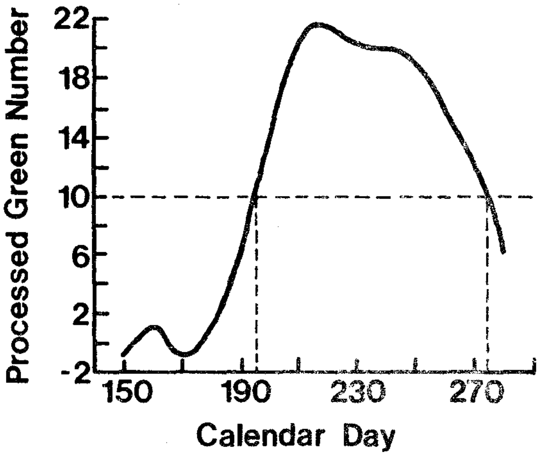

Figure 2. A fairly typical plot of processed green numbers exhibiting early non-crop "green-up." In this case ten is the green number value half-way between the soil green number and maximum canopy green number. The calendar dates where the processed green numbers cross this half-way value are indicated.

B. LANDSAT MSS DATA

The crop development stage estimation technique was also tested on three 9.3 by 11.1 km Landsat data segments, two from 1978 and one from 1979. Field observations of crop development stages were made for selected fields in the segments at several times during the growing season. The Landsat-2 and -3 observations were calibrated to each other and corrected for sun angle. The Kauth-Thomas greenness component of the calibrated Landsat wavelength bands (B1, B2, B3 and B4) is given by(5):

$$\text{Green number} = -0.283*B1 - 0.660*B2 + 0.577*B3 + 0.3884*B4.$$

It is reasonable to assume that the crop observed in each pixel of a particular field should have approximately the same development stage. For this reason, field averages are taken for each Landsat spectal band. Also, estimating crop development stages for field averages rather than for each individual pixel is much more cost effective. Sample standard deviations are also calculated for each field for each Landsat channel as an indication of the variability of the crop within the field. The field green number standard deviation is calculated from the individual Landsat wavelength band standard deviations (SD1, SD2, SD3 and SD4) as follows:

$$\text{Green number standard deviation} = [(0.283*SD1)^2 + (0.660*SD2)^2 + (0.577*SD3)^2 + (0.3884*SD4)^2]^{.5}.$$

Occasionally two sets of Landsat MSS observations are available which are separated by only one day where adjacent orbital paths give overlapping coverage. These observations from adjacent orbital paths may sometimes give noticeably different green number values due to factors besides crop development stage such as atmospheric changes, changes in crop moisture level, a different sun angle, and different sensor look angle. Because of this, the Landsat MSS green numbers need to be filtered (smoothed) in the same manner that the Agronomy Farm green numbers were smoothed to dampen out misleading high frequency variations suggested by the short-term fluctuations. As noted in the Agronomy Farm data discussion, this filtering does not affect observations taken nine or more days apart. The green number standard deviations are filtered in the same way the green number field averages are.

After smoothing, green number estimates are interpolated every nine days with the same quasi-Hermite spline interpolator used with the Agronomy Farm data. Interpolated standard deviation estimates are obtained by running the interpolator directly on the green number standard deviations calculated from the observations (after smoothing). This is a reasonable approach if we consider the green number standard deviation to be an inherent characteristic of a field which may increase or decrease throughout the season depending on several factors including crop development stage.

In the Landsat MSS data case, where the field averages and standard deviations of green numbers are estimated, the soil green number is estimated in a manner similar to that described above for the Agronomy Farm data. The only difference is that instead of a chi-square test, a test is employed that exploits the standard deviation information. This is a test designed to solve the Behrens-Fisher Problem(10), i.e. the problem of testing two samples of normal populations with unequal variances (or standard deviations) against each other for having identical means. Besides the field mean and standard deviation estimates of the green number, this test also requires knowledge of the number of pixels used to estimate the mean and standard deviation of the green number in the field in question. In this case, appropriate one-sided probability thresholds are 75% for both the soil green number estimate and the maximum canopy number estimate. Two different thresholds are not required, because this test already takes into account the variability of the green number estimates through the standard deviation information. Since standard deviation information is exploited by the test for soil green number and maximum canopy green number, the test will give the same results whether or not the green number estimates are normalized as is done in the Agronomy Farm data case.

As we shall see in the results section below, fields in particular geographic areas tend to have processed green number values that cross the half-way rise and fall value at characteristic development stages early and late in the growing season. We will call these estimates of development stages, respectively, early and late season estimates of the crop development stage. The characteristic development stage estimates vary somewhat from one geographic area to another and from one year to the next, so training data for a particular geographic area and/or year is needed to establish these characteristic development stages for the geographic area and/or year in question.

IV. EVALUATION OF RESULTS

A. AGRONOMY FARM DATA

The method for estimating crop development stages was first tested on Purdue Agronomy Farm data. The method gives the calendar day when the normalized green number value rises and falls to cross the half-way value. To get crop development stage estimates, we would have to use training fields to correlate the half-way rise and fall calendar days with crop development stages. Since we do not have enough reference data to divide it into an adequate number of mutually exclusive test and training fields, we chose to use a test of the potential of this method rather than a direct test. (We plan to do direct tests later.)

Our test for the potential of our method is as follows: First we use our method to find the calendar day the normalized green number rises to cross the half-way value and the calendar day it falls to cross the half-way value for each plot. Then we estimate from the reference data the actual crop development stage the crop was at for the indicated calendar days. (We generally have to estimate the crop development stage by interpolation because in most cases the indicated calendar day did not happen to fall on a day a ground observation was made.) Then we calculate the mean and standard deviation of the early season crop development stage estimates and the mean and standard deviation of the late season development stage estimates. A small standard deviation for both cases would indicate that this method has good potential for making accurate estimates of crop development stage early and late in the growing season, given adequate training data. We can compare the mean values of the estimated development stages for data sets from different years and locations to get an indication of how sensitive the training is to changes in years and geographic location.

We tested Purdue Agronomy Farm corn plot data from both 1979 and 1980 in this way. With 36 test plots in 1979 we found the observed average Hanway crop development stage was 2.04 for the early season estimate, with a standard deviation of 0.36. (For convenience we write this result 2.04 ± 0.36.) The observed Hanway stage for the late season estimate was 9.51 ± 0.40. (Complete reference and spectral data sets were available for the late season estimate for 27 out of the 36 plots.) With 52 corn test plots in 1980 we found observed Hanway stages of 2.02 ± 0.26 and 8.99 ± 0.08. (Complete data were available for the late season estimate for 40 out of the 52 plots.) See Figure 3 for histogram of these results. For both the 1979 and 1980 Agronomy Farm data, we find standard deviations of less than 0.50, which is close to the commonly accepted error bound for observing development stages in the field. This indicates that our method does have potential for making reasonably accurate estimates of crop development stages. The closeness of the mean values for the two years may indicate that the training may not be very critical for different years at this location.

B. LANDSAT MSS DATA

Thus far we have completed a limited test with Landsat MSS data on selected fields in only three segments. We tested 10 fields each in two segments of 1978 data. For segment 127 (located in Montgomery Co., Indiana) we found the late season estimate of the Hanway stage to be 9.28 ± 0.78. Ground observations were not taken early enough in the growing season to make an early season test. For segment 862 (located in Calhoun Co., Iowa) we found Hanway stage early season estimate of 4.58 ± 0.37 and late season estimate of

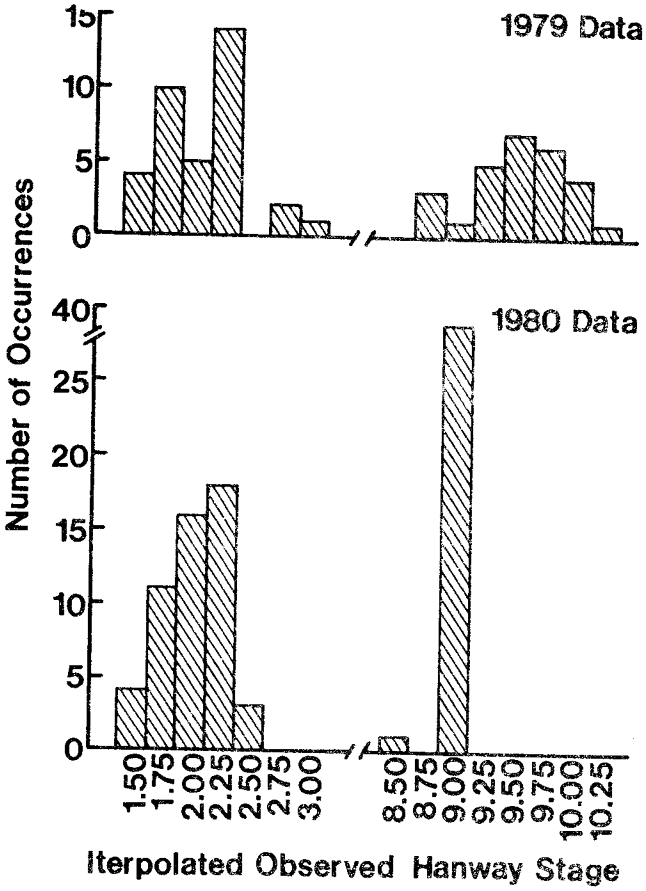

Figure 3. Histograms of the interpolated observed Hanway stages at the early and late season estimates (calendar days the green number rose and fell to cross the half-way value) for the 1979 and 1980 Agronomy Farm data.

9.62±0.32. (Only 4 fields were used in the late seaon estimate due to insufficient data.) We tested 15 fields in one segment of 1979 data. In 1979 a different development stage scale was used for ground observations of development stage (see Table 2). For segment 892 (located in Shelby Co., Iowa) we found development stages of 3.44±0.19 (using 6 fields) and 6.20±0.31 (using all 15 fields) for the early and late season estimates. These development stages correspond roughly to the Hanway stages 3.00 and 10.00, respectively. As with the Agronomy Farm results, the standard deviations for these estimates are below 0.50 (except for one case), indicating that our method has good potential for making accurate crop development stage estimates from Landsat MSS data. The fairly wide differences in early and late season estimates for the segments tested indicate separate training may be necessary for data sets from different geographic areas and different years.

Table 2. Corn development stage coding used for 1979 Landsat reference data.

Stage Number	Stage Name
1.0	PLANTING
2.0	EMERGED
3.0	SIX LEAVES
4.0	TASSELS EMERGED
5.0	BLISTER
6.0	PHYSIOLOGIC MATURITY
7.0	HARVEST

V. CLOSING REMARKS

The preliminary tests indicate that, given sufficient training data, our method should be able to make accurate estimates of the calendar date a crop reaches a particular early season crop development stage using Landsat MSS observations from the first half of a growing season with minimal computation cost. An estimate of the calendar date a crop reaches a particular late season crop development stage can be made using Landsat MSS observations from the last half (or all) of a growing season.

The method can be used to initialize a meteorological development stage model to provide estimates of the calendar day a particular field reaches any given development stage. The meteorological model could be run forward in time from the development stage provided by our method to the critical development stages for yield estimation. The meteorological model could even be run backwards to give estimates of planting dates (possibly for comparison with other methods).

The examples cited and the experimental results given were for corn only. However, this method should be applicable for any other crop that exhibits a similar peaking of green numbers towards the middle of the growing season.

REFERENCES

1. J. J. Hanway (1966). "How a corn plant develops." Iowa State University Cooperative Extension Service Special Report #48, Ames, Iowa.

2. W. R. Fehr and C. E. Cavines (1977). "Stages of Soybean Development." Iowa State University Cooperative Extension Service Special Report #80, Ames, Iowa.

3. E. C. Gilmore Jr. and J. S. Rogers (1958). "Heat Units as a Method of Measuring Maturity in Corn." Agronomy Journal, Vol. 50, pp. 611-615.

4. G. D. Badhwar and K. E. Henderson (1981). "Estimating Development Stages of Corn from Spectral Data: An Initial Model." Agronomy Journal, Vol. 73, pp. 748-755.

5. R. J. Kauth and G. S. Thomas (1976). "The Tasselled Cap - A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat." Proceedings of the 1976 Machine Processing of Remotely Sensed Data Symposium (IEEE Catalog No. 76 CH 1103-1 MPRSD), pp. 4B-41 through 4B-51, June 1976.

6. M. E. Bauer, L. L. Biehl, C. S. T. Daughtry, B. F. Robinson and E. R. Stoner (1979). Agricultural Scene Understanding and Supporting Field Research. Laboratory for Applications of Remote Sensing (LARS), Purdue University, West Lafayette, Indiana. NASA Contract NAS9-15466.

7. W. A. Malila and J. M. Gleason (1977). Investigations of Spectral Separability of Small Grains, Early Season Wheat Detection, and Multicrop Inventory Planning. Environmental Research Institute of Michigan (ERIM), Ann Arbor, Michigan. ERIM Report 122400-34-E.

8. P. J. Bickel and K. A. Doksum (1977). Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day, Inc., San Francisco, pp. 218-219.

James C. Tilton is presently a Postdoctoral Research Associate at the School of Electrical Engineering at Purdue University performing research for Purdue's Laboratory for Applications of Remote Sensing (LARS). He has recently accepted a position as Senior Engineer with Computer Sciences Corporation, System Services Division, Silver Springs, Maryland. B. A. cum laude, Rice University, 1976, in electrical engineering, environmental science and engineering, and anthropology; M. E. E., Rice University, 1976; M. S., optical sciences, University of Arizona, 1978; and Ph. D., electrical engineering, Purdue University, 1981. He is a member of Phi Beta Kappa, Tau Beta Pi and Sigma Xi honoraries and a member of the IEEE. Dr. Tilton's primary research interest is pattern recognition as applied to remote sensing. Of particular interest is working with application disciplines (Agronomy, Environmental Sciences, Forestry, Geology, Soil Sciences, etc.) in designing optimal analysis procedures for specific tasks in remote sensing, and developing techniques to integrate dissimilar data types such as in Georeferenced Information Systems.

Steven E. Hollinger, a research Agronomist at LARS/Purdue, holds a B. S. in Agronomy from Colorado State University, M. S. in Agricultural Meteorology, and a Ph. D. in Agronomy/Agricultural Meteorology from Purdue University. He was raised on a farm and has expertise as an Area Extension Agent in Colorado. While working on his degrees at Purdue, he helped to develop and improve physiological and meteorological based corn and soybean crop models, and to incorporate the models into a large area crop inventory system. He served as project manager for the crop inventory system development, overseeing the field and model work. Since joining the LARS staff, he has worked to incorporate spectral inputs into growth and yield models, and has assisted with the various tasks associated with assessing crop condition.