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

SATELLITE REMOTE SENSING: AN INTEGRAL TOOL IN ACQUIRING GLOBAL CROP PRODUCTION INFORMATION

F.G. HALL

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

ABSTRACT

The state-of-the-art in satellite remote sensing of crop type, condition, and stage in foreign regions is reviewed. A very general approach for automatically distinguishing and identifying various classes of cultural vegetation is described. This approach is based on Landsat MSS data and in particular the Greenness-Temporal Profile models of Badhwar and incorporates a sound physical understanding with which to apply this technology to general cultural and noncultural vegetation types.

The Temporal-Profile approach to identifying and estimating the area of corn and soybeans near harvest has been tested over the U.S. Corn Belt and Mississippi Delta regions and demonstrated to produce unbiased estimates of corn and soybean area with very low variance. These techniques are completely automated. The next logical step is the extension of this capability to foreign regions and to mid and early season.

Research in advanced information extraction techniques to identify crops is also reviewed. Several key issues are discussed; improving preprocessing and registration techniques; improving the understanding of the statistical distributions of the profile parameters; understanding how registration errors and mixed pixels affect these distributions; understanding how to model these distributions in the digital image scene and how to incorporate the spatial component of the digital image data.

Technology for estimating stage of development for corn, soybeans and small grains is reviewed and found capable of predicting stage to within about 1 week of ground observed occurrences of these stages. In addition, early results from the use of spectral data to monitor key plant processes such as evapotranspiration and photosynthesis are reviewed.

Work to explore the additional spectral regions made available by Thematic mapper, and future microwave sensors is in its early stages. However, current findings indicate that these spectral regions are responding to physical phenomena different than is sensed at VIS/NIR wavelengths and should add significantly to our capability to identify vegetation and monitor its condition.

Finally, future research directions in vegetation mapping are discussed.

I. INTRODUCTION

Since its inception in the 1960's, a major thrust of the National Aeronautics and Space Administration (NASA) program of research in remote sensing has been to advance the state-of-the-art in machine processing of satellite acquired multispectral data. The specific focus for this research has been to investigate the use of multispectral data to identify type, to monitor condition, and to estimate the ontogenetic stage of cultural vegetation.

This effort has been confined largely to crops and has been embodied in a number of major programs. Among these are: The Corn Blight Watch conducted in 1970 (ref. 1), which established the feasibility of digital, maximum likelihood classification of aircraft acquired multispectral scanner data to identify crops over a large (multistate) area in a timely manner; the Crop Identification Technology Assessment for Remote Sensing (CITARS) in 1972 (ref. 2), which established the feasibility of automating the digital classification approach to provide rapid, repeatible processing of Landsat Multispectral Scanner System (MSS) data over large areas; and the Large Area Crop Inventory Experiment (LACIE) in 1974-1978 (ref. 3), which demonstrated that Landsat MSS data in foreign regions, in the absence of ground observations, could be processed utilizing manually assisted machine processing techniques, along with multistage sampling, weather data and crop condition models, to identify crops, to estimate their areal extent, stage of maturity, condition, yield and production on a global basis. These efforts stimulated a follow-on program of research (AgRISTARS)(ref. 4) designed to address the technical issues defined by LACIE, to investigate other portions of the electromagnetic spectrum, and in addition to expand the LACIE small grains technology to corn, soybeans, and other key crops in several important producing regions of the world.

This current paper will review the state-of-the-art of the technology used to make remote sensing crop production estimates in foreign regions. This assessment will be largely confined to the technology resulting from those efforts described above, and in particular, will focus on those achievements of the Supporting Research Project, one of several projects within the AgRISTARS program.

II. LANDSAT DATA ACQUISITION

The major inputs to a remote sensing crop survey system are radiometric image data in each of several regions of the electromagnetic spectrum along with meteorological and historic agricultural statistics. These data permit the identification of the type, the condition, and the stage of development of vegetative canopies. Ground observations of crop type and condition cannot be utilized, unless a reliable and objective source of these data can be established. Thus, for foreign crop production survey technology much of the remote sensing survey research has concentrated on crop identification techniques which do not rely on ground observations.

As with conventional, ground-based crop surveys, satellite-based crop surveys rely on statistical sampling techniques. Sampling of the total Landsat coverage from the survey region maximizes the survey accuracy for a fixed quantity of data processing resources. Results from the LACIE showed that a 2% areal sample of a survey region would achieve a precision of roughly 2% (ref. 5). Thus, processing and Landsat data from the remaining 98% of the survey region would add little to the sampling precision, but would almost certainly degrade area estimation accuracy. For a fixed resource base, degraded accuracy would result since processing resources per sample unit would decrease and therefore very likely increase classification error.

A discussion of the state-of-the-art of sampling techniques for remote sensing agricultural surveys is beyond the scope of this current review. Many of the techniques such as stratification, allocation, inference, etc., developed for ground surveys, apply rather straightforwardly to the satellite survey. Suffice it to say, that this technology is rather well developed and was verified in LACIE. Comprehensive reviews of sampling technology as applied to remote sensing may be found in various articles within the LACIE symposium literature (ref. 6).

To date the research to develop remote sensing crop survey systems has primarily focused on the MSS orbiting aboard the Landsat spacecraft (ref. 7). To be launched in 1983 is the Thematic Mapper (TM) with a projected improvement in performance. The thrust of the crop survey research has concentrated on the visible and near-infrared (VIS/NIR) portion of the electromagnetic spectrum. This research has shown that vegetative appearance in the VIS/NIR at a single time is not sufficient to reliably distinguish and identify different vegetation types over large areas. At least 3 to 4 passes during the growing season are needed to distinguish and identify types based upon their developmental difference through time (ref. 8).

Only very modest efforts have explored the Middle Infrared (Mid-IR), Thermal Infrared (Thermal IR), and microwave portions of the electromagnetic spectrum. There are strong indications that these regions will be most useful for identifying and evaluating the condition of vegetation. While the VIS/NIR portions respond primarily to the chlorophyll content and cellular structure of the leaves of a vegetative canopy (ref. 9), the Mid-IR responds primarily to the total water content of the canopy (ref. 10), the Thermal IR to the kinetic temperature of its leaves, and the microwave, particularly the active microwave, to the overall canopy structure (ref. 11).

Thus, adding these new bands, will permit additional physical phenomena to be observed and significant improvements in identifying and monitoring the condition of crops can be anticipated as the Mid, Thermal IR and Microwave portions of the spectrum are explored in the next few years.

III. REGISTRATION AND PREPROCESSING

The need for multidate acquisitions to reliably identify crops implied the need for accurate registration and preprocessing of Landsat data. Registration was necessary to place subsequent Landsat acquisitions into spatial congruence. Preprocessing was needed to remove within-pass and pass-to-pass radiometric and geometric variability.

Geometric preprocessing operations are intended to correct the digital images for geometric distortions introduced by scanner irregularities, topography, earth curvature, etc. The state-of-the-art in geometric preprocessing to correct for spacecraft attitude and ephemeris differences is well developed (ref. 12). Success has also been reported (ref. 13) for aircraft scanner data, which in general is more difficult because of the less stable platform. However, more work is needed to geometrically correct for differences between digital image scenes from disparate sensor types, such as differences between optical scanner and synthetic aperture radar images. In addition, correction for terrain effects must be more fully developed.

Radiometric preprocessing operations are intended to "normalize" the radiometric values for each scene for scene-to-scene differences such as sun angle differences from one season to the next, for differences in instrument calibration from one acquisition to the next, and for differences in atmospheric haze from one day to the next.

Radiometric corrections for detector-to-detector calibration differences in Landsat MSS scanners are extremely important since incorrect procedures can introduce considerable scene variance into the digital image maps, and cause acquisition-to-acquisition differences in radiometric values. Detector-to-detector adjustments were made for the Landsat MSS (ref. 14) and in general were reasonably successful. Acquisition-to-acquisition adjustments require that a stable calibration source be periodically viewed. An in-flight calibration of the Landsat MSS was accomplished by viewing an on-board calibration lamp each scan. The stability of the on-board calibration was checked on each Landsat pass over the Earth's polar regions as solar flux was reflected into the MSS detectors. A key problem with linear array systems envisioned in SPOT and future US satellites will be the inability to accurately account for detector-to-detector calibration differences.

A key area where little success has been realized is the correction of Landsat data for within scene and between scene atmospheric differences. A number of atmospheric models exist for estimating the effects of the atmosphere (refs. 15 and 16) on radiance at spacecraft altitudes, and these models are in general realistic. The key problem, however, is the estimation of the input parameters to the models. These models generally require at least atmospheric optical depth or visibility estimates. A number of techniques have been suggested to estimate these parameters, such as the yellowness component of the Kauth-Thomas transformation (ref. 17), and dark-level subtraction techniques (ref. 18) to estimate path radiance over relatively dark targets such as water bodies. However, these techniques are approximate and themselves add noise to the scene. Evaluations of these techniques are inconclusive as to whether they add or detract from classification accuracy. Research is needed which will improve the understanding of the magnitude of the within and between scene variability in parameters such as atmospheric optical depth, and their effects on classification accuracy.

Upon completion of preprocessing, multidate acquisitions from a segment are registered to a preselected reference acquisition. This completely automatic process begins by processing the image to locate "edges" or boundaries such as highways or agricultural field edges in both the reference image and the image to be registered (ref. 19). Many state-of-the-art approaches accomplish this by defining edge pixels as those contained in a neighborhood in which there is a strong gradient in radiometric value. Thus, the gradient image is used to specify an edge image for each acquisition. The edge images are spatially correlated to the reference image and the position of maximum correlation chosen to register the images. Based on these concepts registration techniques developed and implemented at the Earth Resources Research Division (ERRD) are capable of 0.5 pixel rms registration accuracy (ref. 20). High accuracies are extremely important because, in order to follow the temporal development of a crop canopy, it is necessary to be able to measure the radiance from a single point on the ground over the course of an entire growing season. Misregistration, particularly near field boundaries, can cause measurements to shift from one vegetative class to another during the season. A simple model has shown (ref. 21) that for the same pixel from two different acquisitions to have 50 percent overlap or more, an rms registration accuracy of 0.2 pixels rms is required. Research is underway to improve current accuracies beyond the currently obtainable 0.5 pixel. The central issues being pursued are (1) improved identification and location of image edges and lines to subpixel accuracies using the techniques defined by Haralick and others (ref. 22); (2) improved techniques for correlating the image features from one image to the next; and (3) application of existing image processing techniques (refs. 23 and 24) to improve image resolution and geometric fidelity prior to registration.

IV. DATA TRANSFORMATION

Since for the MSS, one acquisition results in 4 channels of information, 4 such acquisitions result in each pixel being represented by a 16-dimensional vector or radiance measurements. With Thematic Mapper the dimensionality will increase. Analysis in a 16-dimensional vector space, however, has proved to be impractical both because of computational difficulties as well as the large number of parameters which must be estimated when using techniques such as parametric maximum likelihood classification. In addition, principle component along with feature selection analyses of Landsat data have shown that operating in such a large vector space is not necessary since most of the vector information for a given Landsat pass is largely contained in a subspace of 2 dimensions. The problem, however, was to find the directions of the axes of the information bearing subspaces.

A breakthrough in this problem came in 1975 by Kauth and Thomas (ref. 25) when they discovered that a fixed 2-dimensional subspace of Landsat channel space contained most of the spectral information for a very large range of agricultural, seasonal, and meteorological conditions. Further, they showed that changes in value along one of the axes of this subspace was related to changes in the level of scene albedo or brightness and that changes along a perpendicular axis were related to changes in the amount of leafy matter in vegetative canopies. They proposed axes that were related to the 4 Landsat bands by the linear transformation.

$$Z_j = \sum_{i=1}^{4} R_{ij} X_i , \tag{1}$$

$$R = \begin{pmatrix} 0.33231 & -0.28317 & -0.89952 & -0.01594 \\ 0.60316 & -0.66006 & 0.42830 & 0.13068 \\ 0.67581 & 0.57735 & 0.07592 & -0.45187 \\ 0.26278 & 0.38833 & -0.04080 & 0.88232 \end{pmatrix}$$

where (X^1, X^2, X^3, X^4) represent the original Landsat channel values.

This discovery by Kauth and Thomas reduced the dimensionality problem by a factor of two, provided Landsat parameters that were physically relatable to known properties of soils and vegetation, and parameters that were relatively insensitive to certain background effects such as row direction. In these regards, various studies have shown that greenness is highly correlated to percent ground cover by the canopy (ref. 26) and Leaf Area Index (LAI), defined to be the ratio of total canopy leaf surface area to horizontal projected area extended by the canopy. Studies have also shown that radiance measurements from different soils in a scene tend to be distributed along the brightness axes, which being perpendicular to the greenness axis, renders the greenness axis relatively insensitive to soil background effects. For the same reason, greenness has been shown to be relatively insensitive to other brightness dominated effects such as row direction with respect to sun azimuth (ref. 27). An important problem now under investigation is to ascertain where the information regarding vegetation and soils is in TM channel space and whether or not there is new information present in the Mid-IR and Thermal IR regions sensed by Landsat TM.

V. DATA MODELING

Although the Kauth-Thomas discovery had reduced the Landsat dimensionality problem by a factor of 2, there still remained a problem at 6 to 8 dimensionality. Further, it was known that the trajectory of greenness and brightness through temporal space was different for different crops. The problem remained as to how to relate these temporal trajectories to agrophysical properties of the crops such as differences in development rates, growing season lengths, planting dates, etc. The first solution to this problem was obtained by Badhwar (ref. 28) in 1980 when he showed on the basis of empirical considerations that for corn, soybeans, and small grains, the greenness profile in time could be approximated by

$$G(t) = G_0 (t/t_0)^{\alpha} e^{-k(t-t_0)^2} \tag{2}$$

where $G(t)$ is the greenness profile, G_0 is the bare soil greenness, t_0 is the date of emergence, α is a crop specific parameter related to the rate of change of greenness in early season, and k is another crop specific parameter related to the rate of onset of senescence. This formulation was extremely important because it permitted the dimensionality of the data to be further reduced and related the Landsat data to agrophysical parameters such as emergence date, rate of green development, rate of senescence and total length of growing season. To estimate the parameters of eq.(2) from the Landsat data, preprocessed, multidate Landsat data is transformed by (1) into Kauth-Thomas space, then greenness is modeled by the Badhwar model by fitting to each pixel the function of eq.(2) using the technique described in ref. 29. Thus for each pixel, the 16-dimensional vector (assuming 4 acquisitions) is replaced by the 3-dimensional parameter set α, k, t_0. Since the original work by Badhwar, a more general functional form for G(t) has been investigated by him. He related the average rate of growth, k, of greenness and a growth retardation factor, r, to the rate of change of greenness through the differential equation

$$dG/dt = kG(t) - rE(G) \tag{3}$$

where E(G) is the relationship between growth retardation and greenness (i.e., plant population density). Choosing E(G) to be G^2, i.e., the retardation rate per individual plant proportional to the population size, equation (3) has the solution,

$$G = \frac{G_{max}}{1 + (\frac{G_{max}}{G_0} - 1)e^{-\int_{t_0}^{t} k(t)dt}} \tag{4}$$

where G_{max} is the maximum canopy greenness.

The form of equation (4) has several advantages over the earlier form (2) proposed by Badhwar. First, the parameters of (4) can be more directly related to the canopy morphology. The parameter k is equal to the birth rate minus the death rate of leaves in a canopy, which can be established for various species under various conditions by either empirical observations or by phenological models such as the one developed by Ritchie (ref. 30). Thus given 4 Landsat observations of an unknown crop, equation (4) may be fit to that sequence of observations, the parameter k estimated from the fit, and the crop identified by matching the Landsat observed k to the k predicted by a phenological model.

The second advantage of the model form (4) is that it overcomes a weakness of Badhwar's previous form. To wit (Bauer, et al. (ref. 31), and Crist (ref. 32)) showed that there is a general flattening of the greenness profile around its peak. Badhwar's earlier form, eq.(2) was highly peaked at t = sqrt(a/2k) and thus did not adequately represent the true greenness profile at this time. However, with the form of eq.(4), when k is positive (leaf birth rate exceeds death rate) greenness is increasing. When k is zero, greenness remains stationary and when k is negative, greenness decreases. Thus, eq. (4) is consistent with the phenology of a crop and corrects for the weaknesses noted with Badhwar's earlier model. Equation (4) can be simplified further in cases where the form of the greenness profile can be assumed invariant with respect to planting date. This leads to an invariant model of greenness

$$G(t) = G_0 + (G_{max} - G_0)(\frac{2\beta e}{\alpha})^{\alpha/2} \cdot (t-t_0)^{\alpha} e^{-\beta(t-t_0)^2} \tag{5}$$

which has a peak at $t_p - t_0 = \sqrt{(\alpha/2\beta)}$ and has a maximum greenness value $G_{max} - G_0 = A(\alpha/2\beta e)^{\alpha/2}$. Moreover, this profile has two inflection points t_1 and t_2, such that

$$\sigma^2 = (t_2 - t_1)^2 = 1/2\beta + \alpha/\beta(1 - (1-1\alpha)^{1/2} \approx 1/\beta \tag{6}$$

Badhwar has shown (ref. 33) that t_1 corresponds closely to the onset of the reproductive phase of a crop and that t_2 corresponds to the onset of senescence. This implies that t_1, t_2 and thus σ can be predicted using crop calendar models and provide an important feature for labeling crop types.

The key problems to be addressed in the profile modeling research are (1) the investigation of the temporal behavior of the brightness component of the Kauth-Thomas transformation, (2) application of optimum statistical approaches to profile fitting, (3) investigation of the effects of plant condition variations (e.g., drought, disease) on the form of temporal profile, (4) investigation of the profile form at TM wavelengths, (5) verification of the underlying model form of eq.(4), particularly the relationship of k to the birth-dath rates of leaves and (6) the use of σ, t_1, t_p and G_{max} in early season crop identification prior to full season data set availability.

An important additional problem to be addressed with respect to profile parameters is the problem of feature selection. Which of the features permitted by the temporal greenness profiles provides the best separability and identifiability of crop types? How much information is retained in transforming from the original Landsat space to the profile parameter space? In this context, separability is defined as, for example, the Bhattacharyya distance (or some other measure such as probability of misclassification) between the segment probability distributions functions $F(\alpha,k,.../c)$ pairwise between the crops c of interest in the segment. Identifiability does not have a precise mathematical definition but refers to the ability to specify a unique map between parameter values of the marginal distribution functions $F(\alpha,k,.../c)$ and the crop types in the object scene. The separability problem has been extensively addressed in a more general context by Swain, Decell, Guseman et al. (refs. 34, 35, 36), but their techniques have not been applied to the specific problem of selection of optimum profile parameters. All of the work to date regarding feature selection with profile parameters has been heuristic/empirical in nature. Linear combinations of parameter features have not been investigated. For corn from soybeans, and for small grains as a class from pasture, at harvest separability in the profile parameter feature space just discussed is reasonably high for pure pixels, in the range of 90 to 95 percent probability of correct classification.

VI. PROPORTION ESTIMATION

Given a profile feature representation of each pixel in the digital image scene, the problem remains as to how to use this representation to identify the various cover types in the object scene and to estimate their areal proportions.

To date, most approaches to processing Landsat MSS data for crop identification have relied primarily on the spectral information contained in the digital image scene, and very little on the spatial information (e.g., texture, contextual structure), with the exception of spatial clustering approaches such as ECHO (ref. 37), BLOB (ref. 38) and AMOEBA (ref. 39), and the contextual classifier of Swain (ref. 40).

Historically, most of the spectral approaches have relied on ground observations to train a pixel classifier of some type such as the maximum likelihood classifier. Pixels classified into the various cover types were counted and the ratio of the count of each cover type to the whole utilized as an estimator of the areal proportions P^i for each cover type, i.e.,

$$P^i = n^i/N \qquad (7)$$

However, this is a biased estimator in the presence of misclassification error, and the bias is given by

$$E(P^i - P^t) = P(w/o) - [P(O/w) + P(w/o]\ P^t \qquad (8)$$

where P^t is the true proportion of target crop w. Because the errors of omission $P(o/w)$ and commission ($P(w/o)$ do not balance, and in general are not correlated to the ground cover proportions, the bias of eq.(8) can be large. As the errors of omission and commission approach zero, the bias approaches zero. Thus, for well separated cover classes, the bias in the proportion estimate will be small. For such features, the per-pixel classifier approach can work well as demonstrated by Badhwar in an experiment utilizing segments in the U.S. Corn Belt, where bias in estimating corn and soybeans proportions were less than one percent (ref. 41). To what extent features can be found such that a per-pixel classifier leads to acceptably small biases depends of course on the degree of accuracy required.

Research is currently underway to obtain unbiased estimates of crop proportion even in the face of significant "overlap" in the crop spectral distribution functions. This approach relies on a technique called "Mixture Decomposition" because it involves a decomposition of the segment distribution function $P(X)$ into its component marginals. In this approach the segment distribution function is modeled as the linear combination of the marginal distribution functions of the individual crop types in the object scene.

$$P(X) = \Sigma\alpha^i P(X/C^i) \qquad (9)$$

where α^i is the areal proportion of the i^{th} crop in the scene and $P(X/C^i)$ is the marginal distribution function for the i^{th} crop. Teicher (ref. 42) showed that $P(X)$ can be resolved uniquely in terms of the α's and the marginals, given that the marginals form a linearly independent set of functions.

Using a maximum likelihood estimator this result provides a means for obtaining unbiased estimates of the proportions, α^i using the following approach: $P(X)$ can be estimated directly from the set of spectral observables $<X^i>$. α^i, and $P(X/C^i)$ can be estimated using iterative numerical techniques to ascertain a maximum likelihood fit to eq.(9). By assuming a parametric form for $P(X/C^i)$, such as normally distributed marginals,

$$P(X/C^i) \sim N(\mu^i, \Sigma^i) \qquad (10)$$

maximum likelihood estimates for the parameter set $<\mu^i, \Sigma^i>$ may be obtained.

To what degree this approach will be successful in obtaining accurate proportion estimates is under study. Clearly, the accuracy will be influenced by the degree to which reality is approximated by the model assumptions. Are the true marginals really linearly independent? To what extent are these distributions normal?

Perhaps a more serious concern with the mixture decomposition approach is the fact that the mixture decomposition model, eq.(9), assumes that each pixel in the digital image scene is monitoring radiance from a single cover type. Pitts and Badhwar (ref. 43) have shown that in the U.S. roughly 60% of the pixels in a Landsat MSS scene will be mixed. While it is clear that the mixture decomposition model is not the correct one for other than pure pixels, it is not known to what degree this model represents the true distribution including mixed pixels, or for that matter, how to write down the proper model to include mixed pixels. In general, there are two possible approaches to dealing with the mixed pixels.

One approach is to separate the scene into mixed and pure pixels and estimate the crop proportions using different methods for the mixture and the pure pixels. This has the advantage that existing methods, such as the mixture decomposition approach can be applied to the pures. A separate technique would be required to estimate proportions for the mixed pixels. However, this dual approach requires a method for detecting which pixels are mixed and which are pure. Approaches similar to the ones being investigated for edge detection in registration may be applicable.

Another approach to handling the mixed pixel problem is to develop a scene mixture model which models the distributions for both pures and mixed. Whether such a model could be solved uniquely for the crop proportions will not be known until the precise form of the model is known. A strong possibility is that such a model form will not have unique solutions for crop proportions.

VII. LABELING

With estimates of the areal proportions, α^i, the remaining problem is to associate (label) the marginal distributions with ground cover types. Research is currently focused on comparing the maximum likelihood estimates of μ_i and Σ^i to predictions of these parameters from agrometeorological models. For example, the G_{max} of Badhwar's model, eq.(6) is an obvious candidate for the mixture decomposition approach. Using this parameter to characterize the seasonal behavior of each pixel would result in a specific form of eq.(9) given by

$$P(\sigma) = \sum_i \sigma^i * P(G_{max}{}^i / C^i) \qquad (11)$$

The mixture decomposition approach would yield estimates of the $G_{max}{}^i$ for each marginal distribution (as well as the areal proportion). Because the $G_{max}{}^i$ are known to represent the peak greenness for the ith crop, expected values for $G_{max}{}^i$ for each cover type suspected to reside in the object scene can be predicted using a priori knowledge of crop characteristics. For example, if corn and soybeans were the only crops in the scene, and each were represented by a single normal distribution of G_{max}, the prior knowledge of the characteristics of these crops would tell the analyst that

$$E(G_{max}) \text{ soy} \geq E(G_{max}) \text{ corn} \qquad (12)$$

and thus the α^i corresponding to the marginal $P(X/C^i)$ with the largest value for $E(G_{max})$ would represent the estimate of the areal proportion for soybeans in the segment.

For parameters where such stable relationships may not exist, and indeed may be functions of region and weather, models which predict these parameters as functions of these exogeneous variables (agrometeorological models) may be necessary to obtain estimates of these parameters for the purpose of labeling. The predicted values can be mathematically matched to the maximum likelihood estimates obtained from the Landsat data to label the several marginal distribution functions derived by mixture decomposition.

The labeling problem is not a solved problem and research is currently ongoing to develop automatic labeling approaches. Research which may bear on this problem is the current work in the area of Expert Systems (ref. 44). Even though this research is in its early stages one thing is clear. The labeling process will most likely require an ability to predict crop development stage and crop condition from meteorological and ancillary data such as historical agricultural statistics.

VIII. DEVELOPMENT STAGE MODELS

Crop development stage models (more properly, ontogenetic models) serve two very important functions in the production forecast system. As just discussed, these models will almost certainly be key to the crop identification problem. Perhaps even more important, however, is their role in crop yield forecasts. Crop yield is dependent on plant condition which is in turn dependent on weather and more specifically at what point in the life cycle the key yield influencing weather events occur. For example, in corn high temperatures during the pollination period can have a devastating effect on yield, whereas the same high temperatures on either side of this critical period will have little or no effect.

To date crop development stage models have been developed for several major crops, including wheat (ref. 45), barley (ref. 46), corn (ref. 47) and soybeans (ref. 48). The development stages of these crops depend primarily on date of planting, temperature and precipitation. These models predict for a local site, the calendar date at which half the crop area in the scene has reached a particular point on a numerically described scale of development: for example, a scale such as the one by Robertson in which the stage designations range numerically from 2.0 at crop emergence to 6.0 when the crop is mature. These models must be initiated using either observed or predicted date of planting (by field or scene median, whichever is appropriate). In foreign applications, where ground observed planting date is not readily available, planting date models are needed to start the crop stage models.

Input requirements for some of the simpler stage of development models such as that of Robertson are daily maximum and minimum temperatures collected from ground meteorological stations, available worldwide from the World Meteorological Organization (WMO) network). Some of the more sophisticated models can in addition utilize Landsat spectral data, solar radiation data and daily precipitation.

The models generally advance and then accumulate the development stage on a daily basis as does Robertson according to the relation

$$S^i = [A^1 (P - A^o) + A^2(P - A^o)^2] * [B^1 (T^{max} - B^o) + B^2(T^{max} - B^o)^2 + C^1 (T^{min} - B^o) + C^2 (T^{min} - B^o)^2] \quad (13)$$

where P is the daily photoperiod calculated for the particular latitude and longitude of the area, T^{max} and T^{min} are the measured daily maximum and minimum temperature for the area and all other terms are coefficients determined by an iterative regression technique, from a data base of temperatures and development stages collected over several years from experimental plots located at agricultural experiment stations.

The median stage of development after N days have elapsed since planting would then be given by

$$S^N = \sum_{i=1,N} S^i \quad (14)$$

Since the coefficients of eq.(13) themselves depend on the development stage, the period of accumulation in eq.(14) is divided into periods during which the coefficients are considered uniform. For example in Robertson's model there are four such periods; emergence (2.0 to jointing (3.0), 3.0 to heading (4.0), 4.0 to soft dough (5.0), and 5.0 to ripe (6.0).

The Robertson model was extensively evaluated during LACIE and was found to predict the median date to within about 7 days of the ground observed date (ref. 49). A key weakness in this model was its lack of response to moisture deficit conditions. Doraiswamy and Thompson (ref. 50) investigated this problem in AgRISTARS and developed a modification to the Robertson model which attempts to account for the effects on rate of crop development of moisture deficits. A test of their model over several U.S. test sites indicates some improvement over the Robertson model. However, a limiting factor in evaluating the improvements in inaccuracies in the ground observations of crop stage. Fields are visited only each 9 days at the maximum, and the observations for a field are not based on an objective survey. Before work can progress further in this area, more precise ground observations of stage are required. This is a key effort in AgRISTARS field measurements program in 1982.

All of the agrometeorological models require a knowledge of either the planting or emergence date in order to start the model clock running. Recently Badhwar and Henderson (ref. 51) have developed an emergence date prediction model based on Badhwar's spectral profile model. This technique amounts to nothing more than estimating t_0 from eq.(2) and letting this represent the emergence date. Of course, t_0 is the date at which emergence is spectrally detectable, which occurs when the greenness of the canopy is significantly different than the soil greenness. An empirical evaluation of this technique showed that t_0 agreed with ground reported emergence to $\sim\pm 6$ days. This technique does not require that the crop identity of a pixel be known, thus it can be used in conjunction with the agrometeorological models to predict hypothetical distributions of development stage within a segment and then compared to spectrally observed stages for labeling using the approach described above in the APEP discussion. It should be pointed out, however, that currently this spectral emergence date does require a full season's data set, although potentially a mid-season technique is certainly a possibility.

Another key issue in need of further study in crop stage of development research is the accuracy requirements. In some instances, such as predicting the effects on corn yields of high temperatures during pollination, accuracies of as little as 1 day may be required. However, a sensitivity study has not been conducted to evaluate the effects of crop stage prediction errors on either crop yield or crop identification errors.

IX. CROP CONDITION MODELS

As with crop development stage models, crop condition models are also important to both crop identification and crop yield estimation. By crop condition models we generally mean mathematical relationships which are necessary to the estimation of crop yield. Crop condition models quantitatively describe yield influencing processes such as the ability of the plant to move water (evapotranspiration), photosynthetic rate and temperature effects on the reproductive or grain filling process. The literature contains many condition models for various crop types. We will not attempt to review these models here. Suffice it to say, that the models range from physical models which simulate key plant processes at a very detailed level, to less detailed models, which because of the simplicity of their input requirements, can be used in the prediction of crop yields over large geographic regions. AgRISTARS work funded by NASA has consisted of investigating the use of Landsat MSS data to improve estimates of solar input to photosynthesis and evapotranspiration (ET).

In one such study Daughtry et al. (ref. 52), utilized the Energy Crop Growth Variable (ECG) model of Dale (ref. 53) to evaluate the improvements in corn yield prediction resulting from the use of MSS data to estimate SRI. The ECG model is of the form

$$ECG = \sum_{\substack{i=\text{plant,}\\ \text{mature}}} (SR^i)(SRI^i)(ET^i/PET^i)(FT^i) \qquad (15)$$

where SR^i is the daily solar radiation intercepted by the canopy, PET^i is the daily potential evapotranspiration, and FT^i is a daily temperature function which relates growth rate to temperature. To estimate SRI, the authors used the empirically derived relationship between SRI and Kauth-Thomas Greenness G.

$$SRI = -0.1613 + 0.0811\,G - 0.0015\,G^2 \qquad (16)$$

Reflectance data were collected using a Landsat MSS band Exotech 100 radiometer at the Purdue Agronomy Farm, throughout two growing seasons.

To provide a data set in which most of the yield variation could be ascribed to differences in seasonal LAI, and further to evaluate the effects of soil color on the ability to measure SRI, the experimental plots consisted of two completely randomized blocks within three plant populations planted on three different dates, at biweekly intervals. All rows were north-south in direction to assure uniform lighting conditions. Significant water stress was not evident among the treatments. Plots were fertilized to prevent nutrient related growth limitations.

SRI as estimated from eq.(16) was used as an input to the ECG model, along with required meteorological inputs, to estimate yields for the agronomy farm plots. The study showed a correlation between the model estimated and measured yields of 0.7. SRI was not measured directly in this experiment, therefore, the authors could not tell whether their inability to completely explain the observed yield variations is an effect of an incompleteness in the ECG model, or errors in the measurements of input parameters to the model: for example, errors in the inference of SRI from eq.(16). However, the study indicates a potential approach to using satellite data to improve yield estimates over large areas.

In another AgRISTARS funded study over a six-state area, Mohiuddin et al. (ref. 54), at Kansas State University, evaluated the use of Landsat MSS data as an input to the Kanemasu computer simulation yield model to improve the yield estimates from that model over several AgRISTARS test sites. The Kanemasu model consists of two parts--an ET model and a growth model which computes the daily gross photosynthesis, respiration and dry matter accumulation. Robertson's biometeorological time scale is used to predict the wheat growth stages, and the dry matter is accumulated for these stages to estimate kernal number, kernal weight, and grain yield. The Landsat MSS digital count data for 11 AgRISTARS test sites in six states was transformed into LAI using regression equations developed by Pollock and Kanemasu which are based on ratio of MSS channel values and vegetative indices developed by Richardson and Weigand (ref. 55). Kanemasu's yield model was run with and without Landsat estimated LAI, using an assumed, fixed temporal curve for LAI when MSS data were not used. Yield estimates from ground observers and from the model predictions were compared on a field-by-field basis over the test sites for approximately 200 fields. Although the procedures for obtaining ground estimated yields were subject to considerable observational and sampling error, the differences between the ground observed and Kanemasu model estimates using Landsat measured LAI had significantly lower mean square error (793) than did the model estimates (1296) without the Landsat input. The correlation of the Kanemasu model estimates with the ground observed estimates for both experimental cases was about 0.65.

The Daughtry study and the Kanemasu study reported above indicate the preliminary nature of the work in condition modeling using Landsat data. But they also indicate the tremendous potential. As the Mid and Thermal IR regions become available through Thematic Mapper, bands sensitive to the canopy water content and the temperature of the plant, both key inputs to the estimation of Evapotranspiration, should greatly expand the ability to monitor plant condition.

X. OUTLOOK

At this point in agricultural remote sensing, a very general technique for distinguishing and identifying various classes of cultural vegetation, based on temporal profile models and greenness-brightness transformations has been developed. These techniques have been successfully demonstrated for corn and soybeans near harvest, will likely be demonstrated in the very near future for these crops at mid-season, and show good potential for application to the small grains. Furthermore these techniques are based on a sound physical understanding between spectral features and the agrophysical features of the crops. Such an understanding is crucial to extending the techniques to other regions and to other vegetation classes.

Perhaps the most significant result of these techniques is that the Holy Grail of remote sensing, namely signature extension, has been achieved over very broad areas and across several years. The decision boundary used in these techniques for separating and labeling corn and soybeans was developed over 4 segments for one crop year and has been successfully applied across the U.S. Corn Belt and into the Mississippi Delta for three crop years. The next logical step is extension to foreign regions.

Successful signature extension was essential to completely automated procedures. The corn and soybeans techniques discussed above have been automated and require absolutely no manual intervention. Further, they can be made to run with great efficiency, requiring seconds of CPU time per sample unit. The techniques therefore provide a first capability for efficiently and cost-effectively identifying corn and soybeans over large areas and are ready to be evaluated in an application test. These steps, however, are outside the charter of the Supporting Research Project.

Another very important result of this research is that a very general approach for vegetation identification, based on the profile parameter techniques, has been developed. The extension of these techniques to small grains, rangelands and important classes of noncultural vegetation should be relatively straightforward developments.

Very clearly, the new portions of the spectral region which will become available through Thematic Mapper and the Shuttle Imaging Radar (SIR) experiments should be rapidly explored. The focus here will be to reiterate and improve the approach employed by Kauth-Thomas for Landsat MSS, to find the information axes in the expanded multispectral space offered by these new bands. This should be approached in an integrated fashion, adding the Mid and Thermal IR as well as the Microwave to the VIS/NIR bands already in use.

These investigations should also focus on defining new sensor and mission parameters. The questions of spatial resolution, registration accuracy, overpass frequency and time of day for overpass have yet to be fully explored. These questions will be particularly relevant in the setting of crop condition monitoring. If we are on the verge of having a solution to the vegetation identification problem, we are just getting underway in understanding how to use spectral data to monitor key plant processes such as evapotranspiration and photosynthesis, and thus to quantify the effects on yield of drought stress, winterkill, and disease. These areas are ripe for exploration and should yield significant dividends in the future.

The key here will be to utilize the existing knowledge base specifying which biophysical characteristics are associated with various plant conditions and learn how to spectrally observe these biophysical characteristics, e.g., leaf area index, biomass, ontogenetic stage, canopy water content, etc. This will require the use of canopy reflectance models and a strong supporting field measurement program to develop these observational techniques. While some progress has been made in the area of canopy reflectance modeling, the resources have been inadequate in the past to pursue this area fully. NASA has taken steps to increase the research in this area by initiating, this year, a program of Fundamental Research in the area of Scene Radiation and Atmospheric Effects Characterization. However of great concern is the effect of reduced budgets on the field measurements programs.

Another key area will be the investigation of improved information extraction techniques. The key issues here will be improved preprocessing and registration techniques, and improved understanding of the statistical distributions of the profile parameters, how registration errors and mixed pixels affect these distributions, how to model these distributions in the digital image scene and how to take advantage of the spatial component of the data. NASA has also initiated a program of Fundamental Research in the area of Pattern Recognition and Image Analysis.

We are on the verge of an explosion of techniques and capability for vegetation and crop monitoring. It is critical that the U.S. maintain this leadership position in the technology of crop monitoring from space.

XI. REFERENCES

1. Corn Blight Watch Experiment, Summary Report, Vol. 3, Experiment Results. (NASA SP-353, NTIS, Springfield, VA, 1974).

2. Hall, F.G., M.E. Bauer, and W.A. Malila: "First Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS)," Proceedings of the Ninth International Symposium on Remote Sensing of the Environment, Ann Arbor, Michigan, April 1974.

3. MacDonald, R.B. and F.G. Hall, "Global Crop Forecasting," Science, Vol. 208, 670-679, May 1980.

4. AgRISTARS Technical Program Plan. NASA JSC, January 15, 1980, unnumbered document, for copy contact R. Eason, Earth Resources Applications Division-SH, NASA JSC, Houston, Texas 77058.

5. LACIE, Symposium, Proceedings of Technical Sessions (NASA JSC-16016), NTIS, Springfield, VA, 1978).

6. IBID (see 5.)

7. ERTS Reference Manual, General Electric Corp., Valley Forge Space Center, P.O. Box 8555, Philadelphia, PA 19101.

8. IBID (see 5.).

9. Sinclair, T.R., R.M. Hoffer, and M.M. Schreiber: Reflectance and Internal Structure of Leaves from Several Crops During the Growing Season. Agronomy J., Vol. 63, 864-868, Nov-Dec. 1971.

10. IBID (see 9.).

11. Lang, R.H.: Backscattering from a Leaf Canopy. TASS Workshop, Feb. 1980, George Washington University, Washington, D.C.

12. Dye, R.H.: Environmental Research Institute of Michigan, private communication.

13. Ungar, S.G.: "Scanner Imaging Systems in Aircraft," Proceedings of the NASA Workshop on Registration and Rectification, N.A. Bryant editor, JPL Publication, 82-73, May 1982.

14. Wacker, A.G.: Empirically Determined Calibration Differences Between MDP-LIVES and LACIE Processed Data, AgRISTARS Report SR-J1-04133, NASA Johnson Space Center, June 1981.

15. Turner, R.E.: "Atmospheric Effects in Multispectral Remote Sensor Data," NASA CR-ERIM 109600-15-F, 1975.

16. Dave, J.F.: Extensive Data Sets of the Diffuse Radiation in Realistic Atmospheric Models with Aerosols and Common Absorbing Gases. Solar Energy 21:361-369 1978.

17. Lambeck, P.E.: Revised Implementation of the XSTAR Haze Correction Algorithm and Associated Preprocessing Steps for Landsat Data, ERIM Memo IS-PFL-1916, Nov. 1977.

18. Fraser, R.S.: "Computed Atmopheric Corrections for Satellite Data." Proceedings of the Society of Photo-Optical Instrumentation Engineers, August 1974.

19. Wolfe, R.H. and R.D. Juday: Inter-Image Matching. N.A. Bryant, editor, JPL Publication, May 1982.

20. Wolfe, R.H., R.D. Juday, A.D. Wacker, and K. Toyohisa: Landsat Image Registration for Agricultural Applications. Proceedings of the 1982 International Symposium on Machine Processing of Remotely Sensed Data. LARS-Purdue.

21. Trichel, M.D.: NASA Johnson Space Center, Earth Resources Applications Division-SF. Private Communication.

22. Haralick, R.: "Edge and Region Analysis for Digital Image Data," Computer Graphics and Image Processing, 12, 60-73, 1980.

23. Rosenfeld, A. and A.C. Kak: Digital Picture Processing, Academic Press, New York, 1976.

24. Anuta, P.E.: Geometric Corrections of ERTS-1 Digital Multispectral Scanner Data, Information Note 103073, Purdue University, Nov. 1973.

25. Kauth, R.J. and G.S. Thomas: System for Analysis of Landsat Agricultural Data. Environmental Research Institute of Michigan, NASA CR-ERIM 109600, 1976.

26. Bauer, M.E., L.L. Biehl, and B.F. Robinson: Field Research on the Spectral Properties of Crops and Soils, AgRISTARS Report SR-PO-04022, Purdue University, Nov. 1980.

27. IBID (see 26.).

28. Badhwar, G.D., J.G. Carnes, and W.W. Austin: Use of Landsat-Derived Temporal Profiles for Corn-Soybean Feature Extraction and Classification, Remote Sensing of Environment, 12:57-79, 1982.

29. Marquardt, D.W.: "An Algorithm for Least-Squares Estimation of Non-Linear Parameters," J. Soc. Indus. Eng. 11:431, 1963.

30. Personal Communication: Dr. Joe Ritchie, USDA-ARS, Grassland Soil & Water Research Lab., Box 748, Temple, Texas 76503.

31. Bauer, M.E., C.S.E. Daughtry, and V.C. Vanderbilt: Spectral Agronomic Relationships of Corn, Soybean and Wheat Canopies, Proceedings of International Colloguium on Signatures of Remotely Sensed Objects, Avignon France, 1981. Purdue LARS Publication.

32. Crist, E.D.: Association of Spectral Development Patterns with Development Stages of Corn. AgRISTARS Report IT-E2-04235, Environmental Research Institute of Michigan, Feb. 1982.

33. AgRISTARS Semi-Annual Program Review, Supporting Research, SR-J2-04271, NASA Johnson Space Center, April 1982.

34. Swain, P.H. and R.C. King: Two effective Feature Selection Criteria for Multispectral Remote Sensing. LARS Info. Note 042673, Laboratory for Application of Remote Sensing, Purdue University, W. Lafayette, Ind., April 1973.

35. Decell, H.P. and L.F. Guseman: Linear Feature Selection with Applications. Pattern Recognition, Vol. 11:55-63, 1978.

36. Guseman, L.F., B.C. Peters, and H.F. Walker: On Minimizing the Probability of Misclassification for Linear Feature Selection. Annals of Statistics, Vol. 3, No. 3, 661-668, 1975.

37. Rettig, R.L. and D.A. Landgrebe: "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects," Proceedings of the Conference on Machine Processing of Remotely Sensed Data, Purdue University, 1975.

38. Kauth, R.J., A.P. Pentland, and G.S. Thomas: BLOB: An Unsupervised Clustering Approach to Spatial Preprocessing of MSS Imagery. Proceedings of the 11th International Symposium on Remote Sensing of Environmenta, ERIM, Ann Arbor, MI, April 25-29, 1977, vol. 2, pp. 1309-1317.

39. Bryant, J.: AMOEBA: A Spatial Clustering Algorithm for Multitemporal Landsat Data, Report No. 15, NASA Contract NAS9-14689-4S, Texas A&M University, Dept. of Mathematics. August 1977.

40. Swain, D.H., S.B. Vardeman, and J.C. Tilton: "Contextual Classification and Multispectral Image Data," Pattern Recognition, Vol. 13, No. 6, 429-441.

41. IBID (see 33.).

42. Teicher, H.: Identifiability of Finite Mixtures, Ann. of Math. Stat. 34, 1265, 1963.

43. Pitts, D.E. and G. Badhwar: Field Size, Length and Width Distributions Based on LACIE Ground Data, Remote Sensing of Environment 10:202-213, 1980.

44. Feigenbaum, E.A.: The Art of Artificial Intelligence Themes and Case Studies of Knowledge Engineering. Proc. of 5th IJCAI, 5, 1014-1029.

45. Robertson, G.G.: 1968. A Biometeorological Time Scale for a Cereal Crop Involving Day and Night Temperatures and Photoperiod. Int. J. Biometeorol., 12:

191-223.

46. Williams, C.D.W.: 1974. Deriving a Biometeorological Time Scale for Barley. Int. J. Biometeorol., 18:57-69.

47. Cross, H.A. and M.S. Zaber; 1972. Predictions of Flowering Dates in Maize Based on Different Methods of Estimating Thermal Units. Agro. J., 351-355.

48. Major, D.J., D.R. Johnson, and V.D. Luedders; 1975. Evaluation of Eleven Thermal Unit Methods for Predicting Soybean Development. Crop Sci., 15:172-174.

49. Whitehead, V.S., D.E. Phinney, and W.E. Crea; 1978. Growth Stage Estimation in Proceedings of Technical Sessions. The LACIE Symposium. JSC-16015, NASA-JSC, Houston, TX, p. 109-114.

50. Doraiswamy, P.C. and D.R. Thompson: 1982. A Crop Moisture Stress Index for Large Areas and Its Application in the Prediction of Spring Wheat Phenology. Ag. Meteorology J. (accepted).

51. Badhwar, G.D. and D.R. Thompson; 1982. Estimating Emergence Date of Spring Small Grains Using Only Landsat Spectral Data. Agronomy J. (submitted).

52. Daughtry, C.S.T., K.P. Gallo, and M.E. Bauer: Spectral Estimates of Solar Radiation Intercepted by Corn Canopies, NASA Contract NAS9-15466, NASA Johnson Space Center, 1982.

53. Dale, R.F.: An Energy-Crop Growth Variable for Identifying Weather Effects Upon Maize Yield. Proc. Symp. Agrometeor. Maize Crop, p. 240-248. WMO No. 481, July 1976, Iowa State University, Ames, Iowa.

54. Mohiuddin, S. and E.T. Kanemasu: Predicting Winter Wheat Yields With and Without Spectral Data from Landsat Using Kanemasu's Model. AgRISTARS Report SR-M2-04241, NASA Johnson Space Center, Feb. 1982.

55. Richardson, A.J. and C.L. Weigand; Distinguishing Vegetation from Soil Background Information. Photogram. Eng. Remote Sensing, 43:1541-1552, 1977.

Forrest G. Hall received the B.S. in Mechanical Engineering from the University of Texas in 1963, the M.S. in Physics in 1968, and the Ph.D. in Physics in 1970 from the University of Houston. Dr. Hall, a Mathematical Physicist, has published models of the Lunar Atmosphere, Microwave Sea Surface Emission properties, and has conducted and published theoretical investigations in stochastic theory. In 1978, Dr. Hall joined the JSC Earth Observations Division ERTS investigation effort. He has served as Project Leader of CITARS (1972), Project Scientist for LACIE (1974-1978 for which he received NASA's medal for Exceptional Scientific Achievement), Chief Scientist for the Earth Observations Division (1979-1980), and currently serves as Manager of the AgRISTARS Supporting Research Project.