Larry D. Lathrop
Dr. S. P. Pennypacker
Department of Plant Pathology
The Pennsylvania State University
University Park, PA 16802

Spectral Classification of Tomato Disease Severity Levels

It is demonstrated that the reflectance within the spectral regions 380 to 510 nm and 600 to 690 nm is significantly correlated with selected severity levels of tomato early blight disease.

Abstract: *Spectral reflectance measurements were made of randomly sampled plants in a 0.4 ha block of determinant-growth tomatoes (*Lycopersicon esculentum*) having various levels of defoliation caused by the disease early blight. Spectral measurements were made over the 380 to 800-nm region in 10-nm bandwidths. Ordinary least-squares regression was utilized to determine the relative significance among the 43 10-nm bands and the severity levels of defoliation. Two spectral regions of strong statistical significance (380 to 510 and 600 to 690 nm) were identified and subsequently integrated to simulate two broad wavebands. The utility of these wavebands was demonstrated through the use of several computationally simple discriminant functions.*

Introduction

The assessment of plant disease is of major concern to plant pathologists and agronomists. The responsibilities of these professionals may include (1) evaluating the effectiveness of disease control measures, (2) monitoring disease epidemics, (3) determining yield-loss relationships from experimental field plot studies, and (4) estimating the annual regional and national crop yield-loss values for various economically important diseases. Present methods for quantitatively assessing plant disease are laborious, and surveys requiring a large sample size or involving large areas are very time consuming. These surveys might be done more efficiently if remote sensing techniques were developed to determine efficiently and accurately the levels of disease severity.

The traditional approach to assessing disease severity by remote sensing is the "approach of assumption," which is to assume that the broad near-infrared waveband is associated with disease severity (Colwell, 1970; Wallen *et al.*, 1973). A preferrable approach would be to characterize the spectral signatures associated with the disease severity for a specific host-pathogen system.

Pearson *et al.* (1976) developed a hand-held instrument to estimate non-destructively and efficiently green gramineous biomass using radiometric measurements. Since defoliating diseases are evidenced by a reduction in green biomass, a similar type of hand-held instrument would be of value in the assessment of defoliating plant diseases during surveys of small areas. Regional and national surveys could be more efficiently conducted through the use of airborne sensors, which have been shown to be effective in disease incidence surveys (Colwell, 1970; Wallen *et al.*, 1973).

The objectives of this study were to

- Determine from *in situ* measurements which spectral reflectance wavelengths are significantly correlated with selected severity levels of tomato early blight disease, and
- Investigate the usefulness of these reflectance measurements in the classification of disease severity through the use of computationally simple discriminant functions.

Crop

Tomato, *Lycopersicon esculentum*, the crop chosen for this study, is one of the leading vegetable crops in the United States today. Every year approximately 8 million tonnes or $750 million worth are produced commercially (USDA, 1975).

In the northeast, the disease early blight, caused

Photogrammetric Engineering and Remote Sensing,
Vol. 46, No. 11, November 1980, pp. 1433-1438.

by the fungus *Alternaria solani*, is the principal defoliating disease of tomato. The disease characteristically starts on the lower leaves with the appearance of dark, target shaped lesions. Defoliation becomes pronounced as the number of leaves affected, number of lesions, and lesion size increase.

Recent results (Figure 1) indicated that foliar lesions and subsequent defoliation caused by early blight during tomato "fruit-set to fruit full-size prior to fruit ripening" (a particular tomato growth stage) is associated with a reduction in yield (Madden, 1978).

Methods

Tomato seedlings (ca. 18 cm) of the cultivar "Merit," a determinant-growth cultivar, i.e., axial growth does not occur indefinitely, were transplanted into a 0.4-ha block. The plants were set in 5.5-m rows; the rows were spaced 1.8-m apart. Within the rows, the plants were spaced 23 cm apart. These plants were infected with naturally occurring inoculum of *A. solani*.

During a two-day period, *in situ* spectroradiometric measurements were made on 30 randomly identified tomato canopies in the growth stage "fruit-set to 25 percent of fruit full-size prior to fruit ripening," a growth stage when individual plants could no longer be detected and before the canopy crown opened. All spectral measurements were made on days with 0 percent cloud cover and minimal Mie scattering (Heller, 1970), during the time period when the solar zenith angle varied from 18 to 26 degrees.

The spectroradiometric measurements were made with an EG&G Model 580/585 High Sensitivity Spectroradiometer and fiber optic probe (EG&G Inc., Electro-Optics Division, Salem, Massachusetts, 01970). The field-of-view of the sensor is equal to

$$\cos^{1.5} Y$$

where

Y = angular deviation from normal axis.

This instrument configuration measured the spectral radiant flux density in absolute units of watts per square centimetre-nm.

Spectral measurements were made in 10-nm bandwidths over the 380 to 800-nm waveband of the electromagnetic spectrum. The fiber optic probe was placed directly over the tomato canopy at the required height to provide a 0.25 square metre area-of-view at the Earth's surface. The canopy spectral exitance, M (W/cm^2 – nm), was measured by orienting the probe perpendicular to and facing the Earth's surface. The tomato canopy was the only surface within the 0.25 square metre area-of-view, i.e., the fields-of-view did not include exposed areas between plant rows. Direct spectral irradiance, Ed (W/cm^2 – nm), incident on the canopy was measured at hourly intervals by turning the probe 180 degrees from the position whence the canopy was spectroradiometrically measured. Reflectance was estimated by the ratio, M/Ed, at each of the 43 10-nm bands. Reflectance relationships among the spectral wavebands relative to tomato early blight disease severity were investigated in this study.

Following each series of spectral radiant flux density measurements, the canopy's early blight disease severity was assessed by the Horsfall-Barratt (H-B) rating system (Horsfall and Cowling, 1978). The H-B rating system is based upon the "Weber-Fechner" law, which states that the human eye distinguishes according to the logarithm of light stimulus. This law indicates that a visual disease rating system should be based on equal ability to distinguish, not on equal percent disease severity intervals. In essence, the eye tends to see the amount of diseased tissue (lesions) and defoliation when the percent severity is less than 50 percent. When the percent disease severity is greater than 50 percent, the eye tends to see the disease free tissue. The H-B rating system uses 50 percent as a midpoint, with the severity levels differing by a factor of two in either direction of 50 percent (Table 1). The visual ability to distinguish 0-3 percent disease severity therefore equals the ability to distinguish 25-50 percent disease.

Ordinary least-squares regression (OLSR) was used to determine the significant relationships among the reflectance measurements in the 43 10-nm bands and disease severity. The reflectance measurements over the 380 to 800-nm spectral region were not asymptotic with a decrease in disease severity. Therefore, simple linear regression was utilized for the growth stages used in this study, i.e.,

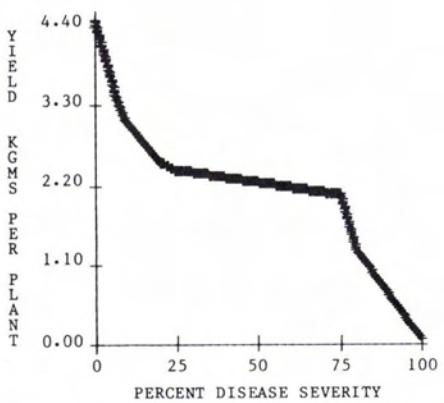

Fig. 1. Relationship between early blight disease severity during the tomato plant growth stage "fruit full-size prior to fruit ripening" and yield (kilograms of fruit harvested per plant) (Madden, 1978)

TABLE 1. HORSFALL-BARRATT RATING SCALE (HORSFALL AND COWLING, 1978)

Categories	Categories in Percent Interval	Category Interval Mean
0	0	0
1	0–3	1.5
2	3–6	4.5
3	6–12	9.0
4	12–25	18.5
5	25–50	37.5
6	50–75	62.5
7	75–88	81.5
8	88–94	91.0
9	94–97	95.5
10	97–100	98.5
11	100	100.0

$$REFL = REFOL + (SL \times DS)$$

where

$REFL$ = canopy reflectance at wavelength L,
$REFOL$ = regression derived estimate of the $REFL$-intercept at wavelength L,
SL = regression derived estimate of the slope at wavelength L, and
DS = disease severity level in percent.

The 10-nm bands included in the spectral regions of strong statistical significance were integrated to simulate broad bands by the following approximation (Tucker and Maxwell, 1976):

$$REFD = [DL \sum_{i=1}^{n} (REF(i))^2]^{1/2}$$

where

$REFD$ = reflectance in band from $L(1)$ to $L(n)$,
$REF(i)$ = reflectance at $L(i)$, and
DL = 10 nm.

Discriminant functions were used to determine the usefulness of the simulated broad band reflectance measurements in the classification of disease severity. The nearest neighbor, the k-nearest neighbor, and the generalized squared distance discriminant functions were selected on the basis of their computational simplicity (Tou and Gonzales, 1974; Barr *et al.*, 1976).

(1) Nearest neighbor (NN):

$$\mathrm{DIS}^2(\mathbf{X},\mathbf{Y}) = (\mathbf{Y}-\mathbf{X})^{\mathrm{T}}\,\mathbf{S}_i^{-1}\,(\mathbf{Y}-\mathbf{X})$$

where

$\mathrm{DIS}^2(\mathbf{X},\mathbf{Y})$ = the Mahalanobis distance between $\mathbf{Y}$ and its nearest neighbor, $\mathbf{X}$, when $\mathbf{S}$ is the category covariance matrix or it is the Euclidean distance when $\mathbf{S}$ is an identity matrix;
$\mathbf{X}$ = the observation that yields the smallest DIS^2 $(\mathbf{X},\mathbf{Y})$;
$\mathbf{Y}$ = the observation vector to be classified; and
$\mathbf{S}_i$ = the covariance matrix for category i or the identity matrix when the Euclidean distance is desired.

(2) K-nearest neighbor (K-NN):

The K-NN classification rule uses a posterior probability function based upon the NN function. The posterior probability function

$$P_i = N_i / \sum_{J=1}^{K} J$$

where

P_i = posterior probability of membership in category i;
N_i = the number of neighbors that correspond to category i; and
K = chosen by operator, its value (an integer) determines the number of smallest DIS^2 $(\mathbf{X},\mathbf{Y})$ from NN to be considered.

(3) The generalized squared distance (GSD):

The GSD classification rule utilizes a posterior probability function based upon the results from the GSD function

$$G^2(\mathbf{X}_i,\mathbf{Y}) = (\mathbf{Y}-\mathbf{X}_i)^{\mathrm{T}}\,\mathbf{S}_i^{-1}\,(\mathbf{Y}-\mathbf{X}_i) + \mathrm{Ln}\,|\mathbf{S}_i|$$

where

$G^2(\mathbf{X}_i,\mathbf{Y})$ = the generalized distance between category mean vector $\mathbf{X}$ of category i and $\mathbf{Y}$;
$\mathbf{X}_i$ = mean vector of category i; and
$|\mathbf{S}_i|$ = the determinant of the covariance matrix

when

$\mathbf{S}$ = the pooled covariance matrix,

then

$$G^2(\mathbf{X}_i,\mathbf{Y}) = (\mathbf{Y} - \mathbf{X}_i)^T \mathbf{S}^{-1} (\mathbf{Y} - \mathbf{X}_i).$$

The posterior probability function

$$P_i = \text{Exp}[-0.5(G^2(\mathbf{X}_i,\mathbf{Y}))] / \sum_{i=1}^{N} \text{Exp}[-0.5(G^2(\mathbf{X}_i,\mathbf{Y}))]$$

where

N = total number of disease categories, $N = 4$ in our study.

The vectors, $\mathbf{X}$ and $\mathbf{Y}$, contain the simulated broad band reflectance measurements. The $\mathbf{X}$ vector has been classified into a disease severity interval by visual assessment. The $\mathbf{Y}$ vector is the observation to be classified to a disease severity interval based upon its association to a vector, $\mathbf{X}$, as dictated by the discriminant function. The NN rule classifies $\mathbf{Y}$ into the category corresponding to the observation, $\mathbf{X}$, which yields the smallest $\text{DIS}^2(\mathbf{X},\mathbf{Y})$. The K-NN rule classifies $\mathbf{Y}$ into the category having the largest posterior probability, P. The GSD rule also uses P and classifies $\mathbf{Y}$ into the class in which P is largest. An observation is not classified in the event two P values from either the K-NN or the GSD function are tied for the largest. All classification rules assume normality and equal probability of an observation belonging to any of the categories.

In order to make efficient use of the data samples, a method was used to obtain unbiased confidence intervals (Lachenbrock, 1967) for the probability of correct classification for the computationally simple discriminant functions previously described. This method classifies each observation by removing that observation from the total number of observations, N, and then from the $N - 1$ observation the statistics are calculated for use by the discriminant functions. After the removed observation has been classified, it is replaced and then the procedure repeats itself until all observations have been classified. The unbiased confidence intervals for the true probability of correct classification, P^*, is then obtained from the roots of the equation:

$$N(P - P^*)^2/P^*(1 - P^*) = Z_{\alpha/2}$$

where

N = Total number of observations;

$P = \sum_{i=1}^{N} Y(i)/N$ where $Y(i) = 1$ if the ith observation is classified correctly and 0 if it is not;

$Z_{\alpha/2}$ = is the $100 \times \alpha/2$ percentile of the normal distribution; and

α = significance level.

Results and Discussion

The 30 visual disease severity assessments utilized in this study ranged from 1.5 to 98.5 percent. The mean and standard deviation of these assigned values were 42.2 percent and 33.4 percent, respectively.

Reflectance variation associated with changes in solar zenith angle and changes in canopy morphology due to transpiration could not be avoided in this study. An attempt was made to minimize the variation associated with these variables by making measurements during a short time period and through randomization. The reflectance variation associated with these variables serves only to decrease the degree of association among the 43-10 nm reflectance wavebands and disease severity.

The "F" values from the analysis of variance resulting from regressing reflectance at 43 10-nm wavebands on disease severity are graphically shown in Figure 2. The large "F" values indicate wavelengths with the most reflectance variability among the disease severity levels. The plot of "F" values is bimodal with the larger "F" values occurring in the spectral regions 380 to 510 and 600 to 690 nm. Within these two regions the 600 to 690-nm band has the larger "F" values. The reflectance values within these two spectral regions, which are associated with chlorophyll pigment absorption (Rabideau, et al., 1946) were integrated to simulate broad bands which can be recorded by the conventional optical camera systems most commonly used for remote sensing purposes.

The infrared (700 to 800 nm) reflectance, which is traditionally used to assess plant vigor (Colwell, 1970; Wallen *et al.*, 1973), did not vary with the level of disease. This lack of association was attributed to the ability of infrared radiation to pene-

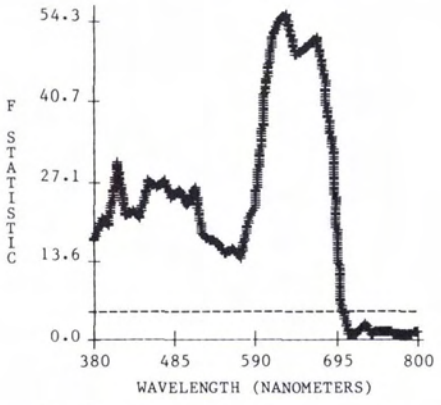

Fig. 2. "F" values for the 43-10 nm wavebands correlated with early blight disease severity. (---) "F" statistic at the 5 percent significance level.

TABLE 2. REFLECTANCE DESCRIPTIONS FOR THE FOUR TOMATO DISEASE SEVERITY CATEGORIES UTILIZED IN THIS CLASSIFICATION STUDY

Category	Observations Per Category	Reflectance Means Per Waveband 380-510 nm A	600-690 nm B	Covariance Matrix A	B
0–12	8	9.5	16.6	A 0.5	0.8
				B	2.8
12–25	4	11.5	19.9	A 0.4	0.2
				B	0.1
25–75	10	12.6	23.0	A 0.2	0.4
				B	3.1
75–100	8	15.4	27.0	A 0.4	0.7
				B	5.0

trate the tomato canopy and interact with the canopy background. Radiation in the visible (400 to 700 nm) region interacts more with the canopy and the degree of variation is influenced by properties such as canopy morphology, leaf area, and chlorophyll concentration.

In another portion of this study (Lathrop, 1979), early blight disease severity and corresponding reflectance measurements were made on 31 randomly identified tomato canopies of the cultivar "Merit" during the growth stage "25 percent fruit full-size to 50 percent of fruit ripe." The "F" values from the analysis of variance resulting from regressing reflectance at 43-10 nm wavebands on disease severity were not significant at the 5 percent significance level. Several 10 nm wavebands were significant at the 10 percent significance level in the 380-510 and 600-690 nm spectral regions. This lack of significance is attributed to the low canopy density and to the variable canopy crown morphologies associated with this tomato growth stage. The canopy morphology of field grown processing type tomato cultivars becomes quite variable after the growth stage "fruit-set to 25 percent of fruit full-size." The canopy crowns after this growth stage varies from closed, where soil, stems, and fruit under the canopy can not be viewed from a vertical position over the canopy, to an open crown, where the soil, stems, and fruit can be viewed from the vertical position over the canopy.

Preliminary analysis revealed adjacent Horsfall-Barratt rating categories (Lathrop, 1979) could be combined into four categories having percent disease intervals 0-12, 12-25, 25-75, and 75-100 percent. The range of disease severity included in each of the four categories is based upon Madden's preliminary yield-loss studies (Madden, 1978). The results of his initial investigation indicated that reduction in yield was associated with 0-12, 12-25, and 75-100 percent disease severity levels (Figure 1). The early blight severity assessments used in his study were made during the growth stage "fruit full-size prior to fruit ripening." The statistics for each of the four categories and the two simulated broad regions are presented

TABLE 3. A SUMMARY OF PERCENT CORRECT CLASSIFICATION (LACHENBRUCK, 1967) WITH CORRESPONDING 95 PERCENT CONFIDENCE INTERVALS (CI) FOR EACH DISCRIMINANT FUNCTION UTILIZED IN THIS STUDY.

Discriminant Function	Percent Correct Classification	95% CI Lower Bound	95% CI Upper Bound
Nearest Neighbor Euclidean Distance	77	59	88
Nearest Neighbor Mahalanobis Distance	73	54	86
K-Nearest Neighbor Euclidean Distance	70	52	84
K-Nearest Neighbor Mahalanobis Distance	73	54	86
Generalized Square Distance Within Covariance Matrix	70	52	84
Generalized Square Distance Pooled Covariance Matrix	70	52	84

in Table 2. There is a direct correlation between reflectance in the visible region and disease severity. This association is the result of the decreased chlorophyll concentration caused by the early blight disease (Lathrop, 1979).

The three discriminant rules were compared using the four categories described in Table 2. The summary of percent correct classification (Table 3) is based upon the results of classifying the observations according to the Lachenbruck method (Lachenbruck, 1967). These data are thought significant for the following reasons: the individual tomato canopy locations were themselves randomly identified prior to any measurements, physical or visual; the tomato canopy morphology during the growth stage "fruit-set to 25 percent of fruit full-size" was found consistent in all locations within the 0.4-ha block used in this study; and errors in the visual classification were minimized by using the H-B rating system and by combining several adjacent H-B categories to produce broader disease severity levels.

All three discriminant functions utilizing the selected wavebands (380 to 510 and 600 to 690 nm) in all of their configurations, correctly classified 70-77 percent of the observations, with 95 percent confidence intervals (CI) ranging from a lower bound of 52 percent to an upper bound of 88 percent (Table 3). As indicated by the overlapping CI, none of the discriminant functions were able to utilize the selected wavebands more effectively than the others. The CI's, which were calculated according to the Lachenbruch method (Lachenbruch, 1967), provides conservative estimates of the lower and upper bounds for the probability of correct classification.

This study demonstrates that: the reflectance within the spectral regions 380 to 510 and 600 to 690 nm is significantly correlated with selected severity levels of tomato early blight disease; and the reflectance from these spectral regions can be used to classify early blight disease severity into the appropriate severity category. Also, the approach used in this study to determine the spectral signatures associated with severity levels of early blight disease of tomato, is more pragmatic than the "approach of assumption," which is based upon studies dealing with phenomena other than those in question. The "approach of assumption," i.e., near-infrared reflectance is associated with disease severity, has limited the successful use of remote sensing in the area of plant disease severity assessment.

ACKNOWLEDGMENT

This paper is Contribution No. 1058, Department of Plant Pathology, The Pennsylvania Agriculture Experiment Station. Authorized for publication January 15, 1979 as Journal Series Paper No. 5663.

REFERENCES

Barr, Anthony J., James H. Goodnight, John P. Sall, and Jane T. Helwig, 1976. *A User Guide to SAS-76.* Sparks Press, Raleigh, N.C. 329 pages.

Colwell, Robert N., 1970. Applications of remote sensing in agriculture and forestry. In *Remote Sensing.* National Academy of Sciences. pp 164-179.

Heller, R. C., 1970. Imaging with photographic sensors. In *Remote Sensing.* National Academy of Sciences. pp. 35-72.

Horsfall, J. G., and E. B. Cowling, 1978. Gathometry: the measurement of plant disease. In *Plant Disease, an Advanced Treatise,* Vol. II. Ed. J. G. Horsfall and E. B. Cowling. Academic Press, Inc., New York, N.Y. pp. 119-136.

Lachenbruck, Peter A., 1967. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics* 23:639-645.

Lathrop, Larry D., 1979. *Relationships among spectral reflectance and selected tomato* (Lycopersicon esculentum) *canopy variables.* M.S. Thesis, The Pennsylvania State University, University Park.

Madden, Laurence V. 1978. Unpublished data.

Pearson, Robert L., Lee D. Miller, and Compton J. Tucker, 1976. Hand-held spectral radiometer to estimate gramineous biomass. *Appl. Opt.* 15:416-418.

Rabideau, G. S., C. S. French, and A. S. Holt, 1946. The absorption and reflection spectra of leaves, chloroplast suspensions, and chloroplast fragments as measured in an Ulbright sphere. *Am. J. Bot.* 33:769-777.

Tou, J. T., and R. C. Gonzales. 1974. *Pattern Recognition Principles.* Wesley Publishing Company, Reading, Mass. 377 pages.

Tucker, C. J., and E. L. Maxwell, 1976. Sensor design for monitoring vegetation canopies. *Photogram. Eng. Remote Sensing.* 42:1399-1410.

USDA Agricultural Statistics. 1975. Washington. Pages 185-187.

Wallen V. R., D. Galway, H. R. Jackson, and L. E. Philpotts, 1973. Aerial survey for bacterial blight, 1970. *Can. Plant Dis. Sur.* 53:96-98.

(Received 30 May 1979; revised and accepted 7 April 1980)