81-30 1981

# COMPARISON OF VEGETATION INDICES BASED ON SATELLITE-ACQUIRED SPECTRAL DATA

Lyle F. Lautenschlager and Charles R. Perry, Jr.
U.S. Department of Agriculture, Statistical Reporting Service*

ABSTRACT

Since the launching of Landsat I in 1972, investigators have derived numerous formulae for the reduction of multispectral scanner (MSS) measurements to a single value (vegetation index) for predicting and assessing vegetative characteristics such as plant leaf area, total biomass and general plant stress and vigor. This report summarizes the origin, motivation, and derivation of some four dozen vegetation indices. Empirical, graphical, and analytical techniques are used to investigate the relationships among the various indices. It is concluded that many vegetative indices are very similar, some being simple algebraic transforms of others.

## 1. INTRODUCTION

Current and accurate information on a global basis regarding the extent and condition of the world's major food and fiber crops is important in today's complex world. Traditional sampling techniques for estimating crop conditions, based on field collection of data, are time consuming, costly, and not generally applicable to foreign regions. An alternate approach is remote sensing - the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation [Lillesand and Kiefer (1979)].

A series of earth resources technology satellites (Landsats) have provided a way to monitor worldwide crop conditions since 1972. The sensor system onboard the Landsats, the multispectral scanner (MSS), measures the reflectance of the scene in four wavelength intervals (bands or channels) in the visible and near-infrared portions of the spectrum. The spectral measurements are influenced by the vegetation canopy, soil type, and atmospheric condition.

Investigators have developed techniques for qualitatively and quantitatively assessing the vegetative canopy from spectral measurements. The objective has been to reduce the four bands of Landsat spectral data to a single number for predicting or assessing such canopy characteristics as leaf area, biomass, percent ground cover, and plant population.

This report summarizes and references the origin, derivation, and motivation for some four dozen of these formulae which are referred to as vegetation indices (VIs). The VIs are categorized on the basis of statistical correlations and algebraic similarities. This analysis reveals the similarities of many vegetation indices.

## 2. LANDSAT DATA CHARACTERISTICS

Three Landsats have been launched since the summer of 1972, with Landsats 2 and 3 still operational. Each satellite is capable of providing 18-day repetitive coverage of the earth's surface. Each Landsat's onboard four-channel MSS system measures reflectance in four bands (fig. 1). The measurements are converted to digital counts and transmitted to receiving stations. Landsat MSS images cover an area of 185 by 185 kilometers and are composed of 7,581,600 picture elements (pixels). [Watkins and Freeden (1979)].

Typical reflectance patterns for herbaceous vegetation and soil are compared in figure 1. Dead or dormant vegetation has higher reflectance than living vegetation in the visible spectrum and lower reflectance in the near-infrared. Soil has higher reflectance than green vegetation and lower reflectance than dead vegetation in the visible, whereas in the near-infrared, soil has lower reflectance than green and dead vegetation [Tappan (1980)]. Jackson et al. (1980), Tucker and Miller (1977), and Deering et al. (1975) provide an extensive discussion of reflectance properties. Three papers of historical interest are Jordan (1969), Knipling (1970), and Pearson and Miller (1972).

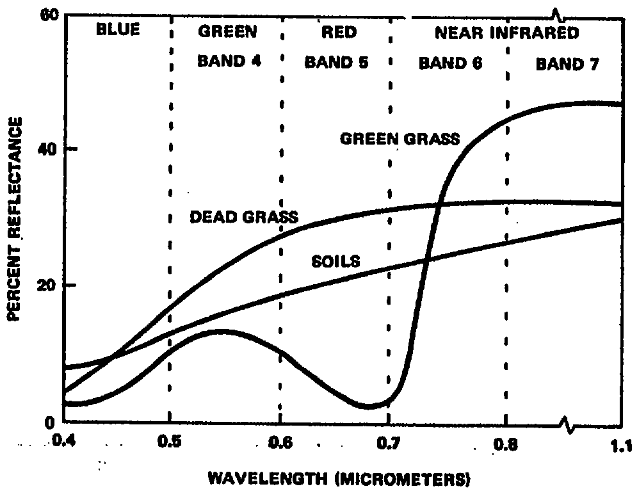



Figure 1. Typical reflectance of herbaceous vegetation and soil from 0.4 to 1.1 micrometers.

## 3. DEVELOPMENT OF VEGETATION INDEX FORMULAE

Numerous vegetation indices have been used to make quantitative estimates of leaf area index, percent ground cover, plant height, biomass, plant population, and other parameters [Pearson and Miller (1972) and Wiegand et al. (1974)]. The formulae are based on ratios and linear combinations of the MSS bands.

The individual Landsat bands (CH4, CH5, CH6, CH7) have been used to estimate percent ground cover and vegetative biomass [Wiegand et al. (1974) and Seevers et al. (1973)]. The correlation coefficients reported ranged from 0.295 for CH7 with crop cover to 0.877 for CH6 with leaf area index. Similar correlations were reported by Tucker (1979).



1981 Survey Section Proceedings, ASA

Ratios of the Landsat bands have been used to estimate and monitor green biomass, etc. [Rouse et al. (1973, 1974), Carneggie et al. (1974), Johnson (1976), and Maxwell (1976)]. The obtained coefficients of determinations were slightly higher than those for the corresponding band differences. The twelve pairwise ratios (six of which are inverses of the other six) will be denoted by R45 = CH4/CH5, R46 = CH4/CH6, etc.

Rouse et al. (1973, 1974) proposed using the normalized difference of Landsat channels 7 and 5 for monitoring vegetation, which will be referred to as ND7. Deering et al. (1975) added 0.5 to ND7 to avoid negative values and took the square root of the result in hopes of stabilizing the variance. This index is referred to as the transformed vegetation index and will be denoted by TVI7. Similar formulae using channels 6 and 5 were proposed.

$$\mathrm{ND6} = (\mathrm{CH6} - \mathrm{CH5})/(\mathrm{CH6} + \mathrm{CH5})$$

$$\mathrm{ND7} = (\mathrm{CH7} - \mathrm{CH5})/(\mathrm{CH7} + \mathrm{CH5})$$

$$\mathrm{TVI6} = (\mathrm{ND6} + 0.5)^{1/2}$$

$$\mathrm{TVI7} = (\mathrm{ND7} + 0.5)^{1/2}$$

Our experience has been that the addition of 0.5 does not eliminate all negative values. We suggest the following computationally correct formulae:

$$\mathrm{TVI6} = (\mathrm{ND6} + .5)/\mathrm{ABS}(\mathrm{ND6} + .5)[\mathrm{ABS}(\mathrm{ND6} + .5)]^{1/2}$$

$$\mathrm{TVI7} = (\mathrm{ND7} + .5)/\mathrm{ABS}(\mathrm{ND7} + .5)[\mathrm{ABS}(\mathrm{ND7} + .5)]^{1/2}$$

where ABS denotes absolute value, and 0/0 is set equal 1. In section 6, it is shown that these formulae are equivalent for decision making to the basic ratios R65 and R75. Therefore, their use can only be justified if either they improve the regression fit or they normalize the regression errors [Draper and Smith (1966)].

Kauth and Thomas (1976) proposed an orthogonal transformation of the original Landsat data space to a new four-dimensional space. They christened this transformation the tassel cap transformation and named the four new axes soil brightness (SBI), green vegetation (GVI), yellow stuff (YVI), and non-such (NSI). The names attached to the new axes indicate the characteristics the indices were intended to measure.

$$\begin{aligned}
\mathrm{SBI} &= .332\ \mathrm{CH4} + .603\ \mathrm{CH5} + .675\ \mathrm{CH6} + .262\ \mathrm{CH7}\\
\mathrm{GVI} &= -.283\ \mathrm{CH4} - .660\ \mathrm{CH5} + .577\ \mathrm{CH6} + .388\ \mathrm{CH7}\\
\mathrm{YVI} &= -.899\ \mathrm{CH4} + .428\ \mathrm{CH5} + .076\ \mathrm{CH6} - .041\ \mathrm{CH7}\\
\mathrm{NSI} &= -.016\ \mathrm{CH4} + .131\ \mathrm{CH5} - .452\ \mathrm{CH6} + .882\ \mathrm{CH7}
\end{aligned}$$

Wheeler et al. (1976) and Misra et al. (1977) applied principal component analysis to Landsat data. The structure of the resulting transformation and the interpretation of the principal components are similar to those for the Kauth-Thomas transformation.

$$\begin{aligned}
\mathrm{MSBI} &= .406\ \mathrm{CH4} + .600\ \mathrm{CH5} + .645\ \mathrm{CH6} + .243\ \mathrm{CH7}\\
\mathrm{MGVI} &= -.386\ \mathrm{CH4} - .530\ \mathrm{CH5} + .535\ \mathrm{CH6} + .532\ \mathrm{CH7}\\
\mathrm{MYVI} &= .723\ \mathrm{CH4} - .597\ \mathrm{CH5} + .206\ \mathrm{CH6} - .278\ \mathrm{CH7}\\
\mathrm{MNSI} &= .404\ \mathrm{CH4} - .039\ \mathrm{CH5} - .505\ \mathrm{CH6} + .762\ \mathrm{CH7}
\end{aligned}$$

Misra et al. (1977) proposed another linear transform, based on the idea of spectral brightness and contrast. Generalizations of spectral brightness and contrast were defined in spectral density space, then transformed back to count space. The first two components of the resulting transformation are similar to the first two components of the two preceding transformations.

$$\begin{aligned}
\mathrm{SSBI} &= .437\ \mathrm{CH4} + .564\ \mathrm{CH5} + .661\ \mathrm{CH6} + .233\ \mathrm{CH7}\\
\mathrm{SGVI} &= -.437\ \mathrm{CH4} - .564\ \mathrm{CH5} + .661\ \mathrm{CH6} + .233\ \mathrm{CH7}\\
\mathrm{SYVI} &= -.437\ \mathrm{CH4} + .564\ \mathrm{CH5} - .661\ \mathrm{CH6} + .233\ \mathrm{CH7}\\
\mathrm{SNSI} &= -.437\ \mathrm{CH4} + .564\ \mathrm{CH5} + .661\ \mathrm{CH6} - .233\ \mathrm{CH7}
\end{aligned}$$

Richardson and Wiegand (1977) used the perpendicular distance to the "soil line" as an indicator of plant development. The "soil line", a two-dimensional analogue of the Kauth-Thomas SBI, was estimated by linear regression. Two perpendicular vegetation indices were proposed.

$$\mathrm{PVI7} = [(.355\ \mathrm{CH7} - .149\ \mathrm{CH5})^2 + (.355\ \mathrm{CH5} - .852\ \mathrm{CH7})^2]^{1/2}$$

$$\mathrm{PVI6} = [(-.498 - .457\ \mathrm{CH5} + .498\ \mathrm{CH6})^2 + (2.734 + .498\ \mathrm{CH5} - .543\ \mathrm{CH6})^2]^{1/2}$$

Evidently a minor error was made in the derivation of PVI6. The formula for PVI6 should be:

$$\mathrm{PVI6} = [(-2.507 - .457\ \mathrm{CH5} + .498\ \mathrm{CH6})^2 + (2.734 + .498\ \mathrm{CH5} - .543\ \mathrm{CH6})^2]^{1/2}$$

These formulae are computationally inefficient and do not distinguish right from left of the "soil line" (water from green stuff). The standard formula from analytic geometry for the perpendicular distance from a point to a line solves this difficulty [Salas and Hille (1978)].

$$\mathrm{PVI6} = (1.091\ \mathrm{CH6} - \mathrm{CH5} - 5.49)/(1.091^2 + 1^2)^{1/2}$$

$$\mathrm{PVI7} = (2.4\ \mathrm{CH7} - \mathrm{CH5} - .01)/(2.4^2 + 1^2)^{1/2}$$

The difference vegetation index (DVI), suggested by Richardson and Wiegand (1977) as computationally easier than PVI7, is essentially a rescaling of PVI7.

$$\mathrm{DVI} = 2.4\ \mathrm{CH7} - \mathrm{CH5}$$

The Ashburn vegetation index [Ashburn (1979)] was suggested as a measure of green growing vegetation. The doubling of CH7 is to make the scale compatible; CH7 is 6-bit data and has one-half the range of the other three bands which are 8-bit data.

$$\mathrm{AVI} = 2.0\ \mathrm{CH7} - \mathrm{CH5}$$

Colwell et al. (1979) proposed a vegetation indicator called greenness above bare soil (GRABS). This was another attempt to develop an indicator for which a threshold value could be specified for detecting green vegetation. The calculations were made using the Kauth-Thomas tassel cap transformation applied to sun angle- and haze-corrected data. The resulting index is quite similar to the GVI, since the contribution of SBI is less than 10 percent of GVI.

$$\mathrm{GRABS} = \mathrm{GVI} - .09178\ \mathrm{SBI} + 5.58959$$

Kanemasu et al. (1977) regressed winter wheat leaf area measurements on MSS band ratios and produced the following regression equation.

$$\mathrm{ELAI} = 2.68 - 3.69\ \mathrm{R45} - 2.31\ \mathrm{R46} + 2.88\ \mathrm{R47} + 0.43\ \mathrm{R56} - 1.35\ \mathrm{R57} + 3.07[\mathrm{R45} - (.5\ \mathrm{R47})(\mathrm{R45})]$$

Pollack and Kanemasu (1979) later used a larger data set plus stepwise regression and obtained another regression equation.

$$\mathrm{CLAI} = .366 - 2.265\ \mathrm{R46} - .431(\mathrm{R45} - \mathrm{R47})(\mathrm{R45}) + 1.745\ \mathrm{R45} + .057\ \mathrm{PVI7}$$

Separate regression equations were also obtained for CLAI values above and below 0.5.

LAI = 1.903 - 1.138 R56 - .071(R45 - R47)R45 - .016 PVI6, if CLAI is less than 0.5

LAI = -5.33 + .036 PVI7 + 6.54 TVI6, if CLAI is greater than 0.5

The Foreign Crop Condition Assessment Division (FCCAD) of the Foreign Agricultural Service (FAS), Houston, Texas uses another leaf area model. We have been unable to find any reference to the development of this model.

OLAI = 41.325 R45 - 42.45 R46

Badhwar (1981) proposed a ratio of GVI to SBI as an indicator of crop discrimination. It will be shown in section 6 that this index is a generalization of a normalized difference.

GVSB = GVI/SBI

Craig Wiegand (personal communication) suggested converting reflectance values to radiances. Linear transformations were used to change from reflectance to radiance values. Ratio and normalized difference formulae were also created using the radiance values.

| | |
|---|---|
| RAD5 = 0.0157 CH5 | for Landsat 1 |
| = 0.0134 CH5 + 0.06 | for Landsat 2 |
| = 0.0139 CH5 + 0.03 | for Landsat 3 |
| RAD7 = 0.0730 CH7 | for Landsat 1 |
| = 0.0603 CH7 + 0.11 | for Landsat 2 |
| = 0.0603 CH7 + 0.03 | for Landsat 3 |

RADR75 = RAD7/RAD5

NDRAD = (RAD7 - RAD5)/(RAD7 + RAD5)

Thompson and Wehmanen (1978) proposed a technique utilizing transformed Landsat digital data to indicate when agricultural vegetation is undergoing moisture stress. The screening number or green number (GIN) was proposed to estimate the percentage of land in an area with a "healthy" cover of vegetation. A "soil line" is determined by inspecting the channel data and discarding data not considered reasonable for agricultural data. The "soil line" is then evaluated as the minimum value remaining in CH5 and subtracted from GVI to obtain GIN.

GIN = GVI - soil line

The data sets included in this study did not permit the computation of GIN. However, GIN is a linear transformation of GVI.

## 4. EVALUATION OF VEGETATION INDICES

### 4.1 Background

Richardson and Wiegand (1977) correlated eight VIs (GVI, DVI, SBI, PVI6, PVI7, TVI6, TVI7, and R57) with four plant component variables (crop cover, shadow cover, plant height, and leaf area index). The correlation coefficients obtained by plant component with the VIs (excluding SBI) were very similar. Later, Wiegand et al. (1979) correlated leaf area indices for winter wheat fields to five VIs (TVI7, TVI6, PVI7, PVI6, and GVI). The correlation coefficients by field and even between fields were similar.

Aaronson et al. (1979) studied the similarities and differences among seven VIs (AVI, DVI, GVI, OLAI, PVI7, TVI7, and KVI). The obtained correlation coefficients ranged from 0.8 to 1.0 and were stable from spring greenup to harvest. Aaronson and Davis (1979) later used a large data set, which included vegetation measurements and several VIs, to study interrelationships. The VIs (AVI, DVI, GVI, OLAI, KVI, PVI6, PVI7, TVI6, and TVI7) were correlated against each other and against vegetation measures such as plant height from tillering through harvest. The correlation coefficients between the VIs ranged from 0.81 to 1.00, and those between VIs and vegetation measures were similar.

### 4.2 Cluster Analysis of VIs

The similarity between the VIs was first studied using the BMDP program P1M, cluster analysis of variables. The absolute value of the bivariate correlations was used as the measure of distance between VIs, and the average distance between elements was used as the between cluster distance. Similar results were obtained using other standard distance measures.

This procedure separated the VIs into two large clusters plus a number of small clusters. One large cluster contained VIs based on MSS bands 5 and 7, which included AVI, PVI7, R75, TVI7, and ND7. The other large cluster contained VIs, based on MSS bands 5 and 6, and a few VIs involving three or all four bands, which included GRABS, CLAI, OLAI, R65, TVI6, ND6, GVI, MGVI, PVI6, and SGVI. The VIs within these two clusters had absolute simple linear correlations greater than 0.90, with most greater than 0.95. The elements of these two large clusters are correlated at 0.8 or higher. Three smaller clusters readily apparent were: (NSI, R76), (R64, R74), and (SBI, MSBI, SSBI, SNSI). This clustering is applicable to the period from spring greenup to harvest. There are some clusters, however, which have high correlations for the whole season, especially those involving bands 5 and 7. The cluster trees on which this discussion is based are included in a more detailed report by Lautenschlager and Perry (1981).

Some VIs were not used in the cluster analysis because of their known relationships to others. The inverse ratios R54, R46, R47, R56, R67, and R57 were not used. DVI was discarded because of its relationship to PVI7, as were RAD5, RAD7, RADR75, and NDRAD because of the linear relationships to CH5, CH7, R75, and ND7.

## 5. VEGETATION INDICES EQUIVALENCE

In this section, a definition of VI equivalence will be developed. This permits a natural categorization of the VIs. VIs are functions which associate a real value to the four-dimensional Landsat reflectance measurement vector, (MSS4, MSS5, MSS6, MSS7). Thus, it will be convenient to employ standard function notation: f:S1--S2 denotes a function from the set S1 into the set S2; f(X), the value of f at the point (X) of S1; Dom(f), the domain of f; Ran(f), the range of f; and $f^{-1}$:S2--S1, the inverse of f when it exists. The inverse exists if, and only if, f is one-to-one and onto. The composition of two functions has an inverse if, and only if, both functions have inverses; in which case

$(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.

It might seem that VI equivalence should correspond to function equality; i.e., V1 = V1 if, and only if, V1(X) = V1(X) for each Landsat reflectance value X. However, this requirement is too restrictive because it involves only the VIs output and ignores the decisions made on the basis of this output. Since vegetation indices are formulae used in making decisions about crop characteristics and conditions, it seems appropriate to say that two VIs are equivalent if the same decision results regardless of the VI employed. This means that two VIs, V1 and V2, are equivalent for making the set of decisions D if, and only if for every decision rule $d1:Ran(V1)\text{--}D$, there corresponds a decision rule $d2:Ran(V2)\text{--}D$ such that the decision, based on d2 and V2, is the same as the decision based on d1 and V1 for all Landsat reflectance measurements X; that is, $d1(V1(X)) = d2(V2(X))$ for each X. It is easy to see that the two vegetation indices, V1 and V2, are equivalent if, and only if, there exists a one-to-one onto function $T:Ran(V1)\text{--}Ran(V2)$ such that $T \circ V1 = V2$. This implies that a decision d results from the same set of Landsat reflectance regardless of which VI is used; that is

$$V1^{-1}[T^{-1}(d)]=(T \circ V1)^{-1}(d) = V2^{-1}(d) \qquad \text{(Eq. 1)}$$

for each decision d in D, where the superscript -1 indicates the inverse image of d under the given function. The relationship defined is an equivalence relation on the set of vegetation indices.

A number of studies have investigated the transformed vegetation indices TVI6 and TVI7 and the corresponding ratios R65 and R75 as predictors of biomass, leaf area index, plant height, and percent ground cover. The predictive ability of TVI6 and R65 or TVI7 and R75 are similar as evidenced by the estimated correlation coefficient. We now show that the transformed vegetation index and its generalizations are equivalent to the corresponding ratios. This example makes clear not only the algebraic and geometric meaning of VI equivalence but also demonstrates the utility and appropriateness of this definition.

Let a and b be positive constants, and define the functions f, g, and T by

$$f(X5,X7) = (aX7 - bX5)/(aX7 + bX5)$$
$$g(X5,X7) = X7/X5$$
$$T(y) = (b/a)[(1 + y)/(1 - y)]$$

for X5 and X7 positive and ABS (y) less than one. Observe that T is invertible; in fact

$$T^{-1}(z) = (az - b)/(az + b) \text{ for } z \text{ positive}$$

Thus, f and g are equivalent and the values of f can be computed from the values of g and vice versa.

$$(T \circ f)(X5,X7) = g(X5,X7)$$
$$(T^{-1} \circ g)(X5,X7) = f(X5,X7)$$

Let k and p be real, and define the functions $G:(-1,1)\text{--}(k-1,k+1)$ and $H:(k-1,k+1)\text{--}(L,U)$ by

$$G(v) = v + k$$

$$H(w) = w[ABS(w)]^{P-1}, \text{ for}$$

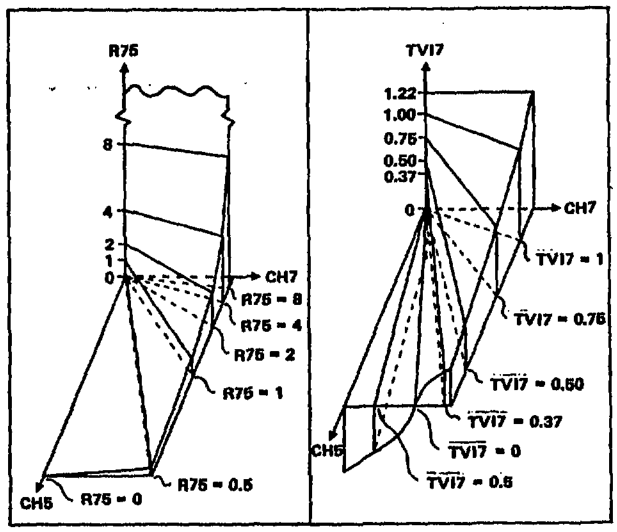
w between k-1 and k+1, $L = (k-1)[ABS(k-1)]^{P-1}$, $U = (k+1)[ABS(k+1)]^{P-1}$, ABS(v) less than one, and 0/0 defined as 1. It is easy to verify that G and H are one-to-one and onto and that

$$(H \circ G \circ T^{-1} \circ g)(X5,X7) = (f(X5,X7) + k)[ABS(f(X5,X7) + k)]^{P-1}.$$

Taking k = p = 1/2 and a = b = 1 show that the transformed vegetation index, TVI7, is equivalent to the seven-five ratio, R75.

$$(H \circ G \circ T^{-1})\ R75 = TVI7$$

Equivalence of VIs means their response surfaces determine precisely the same partition of the reflectance measurement space (equation 1). Elements of this partition are referred to as decision classes. Representive response surfaces and equivalence classes associated with TVI7 and R75 are illustrated in figures 2a and 2b. The nonlinear algebraic relationships exhibited above between R75, TVI7, and ND7 are illustrated graphically in figure 3.




(a) Associated with R75 (b) Associated with TVI7

Figure 2. Response surface and equivalence classes.

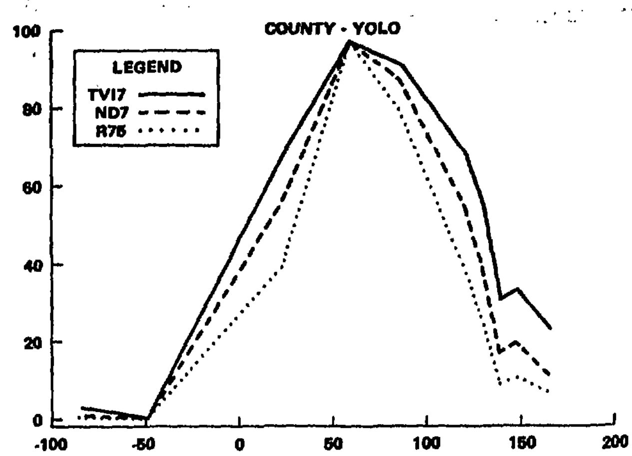



Figure 3. TVI7, ND7, and R75 versus time using data listed in Lautenschlager and Perry (1981). All VI values have been rescaled 0 to 100.

As a further illustration of the utility of VI equivalence, GVSB is shown to be approximated by ND6. Thus, the more complicated GVSB can be expected to provide approximately the same information about crop condition as the simple ratio R65.

Using Landsat data, the following estimates were obtained [Lautenschlager and Perry (1981)].

GRANT AREA DATA / N = 6084

| Variable | N | Mean | Std. Dev. |
|---|---|---|---|
| CH4 | 6084 | 23.2 | 7.2 |
| CH5 | 6084 | 26.7 | 10.0 |
| CH6 | 6084 | 41.4 | 15.9 |
| CH7 | 6084 | 17.5 | 6.3 |

CORRELATION COEFFICIENTS

| Variable | CH4 | CH5 | CH6 | CH7 |
|---|---|---|---|---|
| CH4 | 1.00 | | | |
| CH5 | 0.86 | 1.00 | | |
| CH6 | 0.73 | 0.64 | 1.00 | |
| CH7 | 0.67 | 0.50 | 0.96 | 1.00 |

From these estimates, one easily obtains the regression equations

CH7 = .4100 CH6 + .5100
CH4 = .6236 CH5 + 6.564

Naively substituting into the formulae for GVI and SBI gives the following formulae.

GVI = .74 (CH6 - 1.14 CH5 + .03)
SBI = .78 (CH6 + 1.03 CH5 + 2.96)

Using the information in the above tables pertaining to the expected range of the data, it is easy to see that a rough approximation for GVSB is:

EGVSB = (CH6 - 1.14 CH5)/(CH6 + 1.03 CH5)

which is approximately ND6. In fact, let

$h(v) = (b + vd)/(a - vc)$
$k(x,y) = (ax - by)/(cx + dy)$
$r(x,y) = x/y$, then $h(k(x,y)) = x/y = r(x,y)$

Thus, the estimate, EGVSB, is equivalent to R65 and ND6. These relationships are illustrated graphically in figure 4.

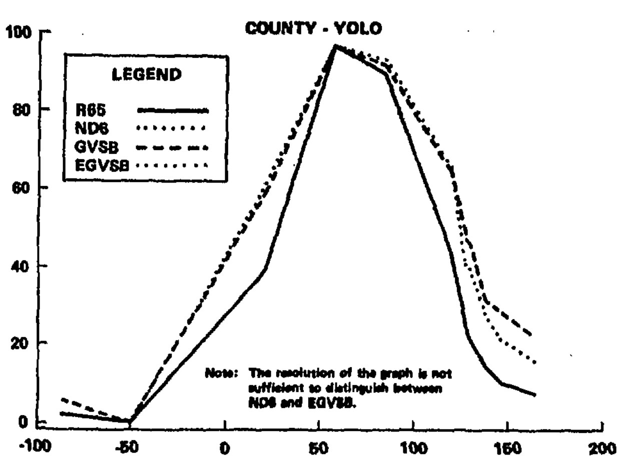



Figure 4. R65, ND6, GVSB, and EGVSB versus time using data listed in Lautenschlager and Perry (1981). All VI values have been rescaled 0 to 100.

## 6. SUMMARY AND CONCLUSIONS

Other researchers have studied the relationships among a few of the VIs considered in this report. Past work has been based exclusively on correlation analysis. Aaronson and Davis (1979) showed conclusively that, during the spring greenup to harvest phase of the crop season, the VIs used operationally by The Foreign Agriculture Service (FAS)/Foreign Crop Condition Assessment Division (FCCAD) were highly correlated and had similar correlations with various plant components such as biomass, plant height, etc.

This study extends analysis to include all VIs found in the literature. Techniques used to investigate relationships between the VIs included variable clustering by correlation, graphical presentations, and functional equivalence for decision making. Variable clustering separated out two large clusters of VIs. One cluster contained those VIs which used channels 5 and 7 data. The other cluster contained VIs using channels 5 and 6 data plus some VIs using all four channels of data. The variable clustering technique also showed that these two clusters were highly correlated. The relationships were stable during the spring greenup to harvest period of the crop season. Graphical presentations reinforced the clustering results, illustrating the relationships over time and through response surfaces. Mathematical techniques were used to formalize the idea of VI equivalence. This equivalence was used to confirm relationships observed earlier and to investigate less apparent relationships.

## LITERATURE CITED


Aaronson, A. C. and L. L. Davis, 1979, "An Evaluation of Relationships between Vegetative Indices, Soil Moisture and Wheat Yields", Technical Memo. No. 9, Crop Condition Assessment Division, USDA/FAS, Houston, TX.

Aaronson, A. C.; L. L. Davis; and G. A. May, 1979, "Results of the Vegetative Index Correction Study", Tech. Memo. No. 6, Crop Condition Assessment Division, USDA/FAS, Houston, TX.

Ashburn, P., 1979, "The Vegetative Index Number and Crop Identification", The LACIE Symposium.

Badhwar, G. D., 1981, "The Use of Parameters to Separate and Identify Spring Small Grains", Quarterly Technical Interchange Meeting, NASA-JSC, Houston, TX, July 8.

Carneggie, D. M.; S. D. de Gloria; and R. N. Colwell, 1974, "Usefulness of ERTS-1 and Supporting Aircraft Data for Monitoring Plant Development and Range Conditions in California's Annual Grassland", BLM Rpt. 53500-CT3-266 (N).

Colwell, R. N.; C. M. Hay; C. A. Kuretz; J. B. Odenweller; E. Scheffner; and B. Wood, 1979, "Development of AI Procedures for Dealing with the Effects of Episodal Events on Crop Temporal Spectral Response", AgRISTARS SR-B9-00434, Contract NAS 9-14565.

Deering, D. W.; J. W. Rouse; R. H. Haas; and J. A. Schell, 1975, "Measuring 'Forage Production' of Grazing Units from Landsat MSS Data", Proc. of the 10th Intl. Symp. on Remote Sensing of Environ., II:1169-1178.

Draper, N. R. and H. Smith, 1966, "Applied Regression Analysis", John Wiley and Sons, NY.

Jackson, R. D.; P. J. Pinter, Jr.; R. J. Reginato; and S. B. Idso, 1980, "Hand-held Radiometry - A Set of Notes Developed for Use at the Workshop on Hand-held Radiometry", Phoenix, AZ, AgRISTARS EW-U1-04052, JSC-17118.

Johnson, G. R., 1976, "Remote Estimation of Herbaceous Biomass", M. S. Thesis, Colorado State University, Ft. Collins.

Jordon, C. F., 1969, "Derivation of Leaf-area Index from Quality of Light on the Forest Floor", Ecology, Vol. 50, No. 4.

Kanemasu, E. T.; J. L. Heilman; J. O. Bagley; and W. L. Powers, 1977, "Using Landsat Data to Estimate Evapotranspiration of Winter Wheat", Environ. Mgt., 1:515-520.

Kauth, R. J. and G. S. Thomas, 1976, "The Tassel Cap - A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as seen by Landsat", Proc. of the Symp. on Machine Processing of Remotely Sensed Data, LARS, Purdue, IEEE Cat. No. 76, Ch. 1103-1 MPRSD.

Knipling, E. B., 1970, "Physical and Physiological Basis for the Reflectance of Visible and Near-Infrared Radiation from Vegetation", Remote Sensing of Environment 1, pp. 155-159.

Lautenschlager, L. F. and C. R. Perry, Jr., 1981, "An Empirical, Graphical, and Analytical Study of the Relationships Between Vegetative Indices", AgRISTARS EW-J1-04150, JSC-17424.

Lillesand, T. M. and R. W. Kiefer, 1979, "Remote Sensing and Image Interpretation", John Wiley and Sons, NY.

Maxwell, E. L., 1976, "Multivariate System Analysis of Multispectral Imagery", Photogrammetric Engineering and Remote Sensing, Vol. 42.

Miller, G. E.; L. F. Lautenschlager; and C. R. Perry, Jr., 1981, "A Partial Literature Review of Landsat Vegetation Indices, Their Origins and Development", AgRISTARS EW-L1-04134, JSC-17413.

Misra, P. N. and S. G. Wheeler, 1977, "Landsat Data from Agricultural Sites - Crop Signature Analysis", Proc. of the 11th International Symp. on Remote Sensing of the Environment, ERIM.

Misra, P. N.; S. G. Wheeler; and R. E. Oliver, 1977, "Kauth-Thomas Brightness and Greenness Axes", IBM personal communication, Contract NAS 9-14350, RES 23-46.

Pearson, R. L. and L. D. Miller, 1972, "Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie", Eighth International Symposium on Remote Sensing of Environment, Univ. of Michigan, Ann Arbor.

Pollock, R. B. and E. T. Kanemasu, 1979, "Estimating Leaf-area Index of Wheat with Landsat Data", Remote Sensing of Environment, 8:307-312.

Richardson, A. J. and C. L. Wiegand, 1977, "Distinguishing Vegetation from Soil Background Information", Photogrammetric Engineering and Remote Sensing, Vol. 43, No. 12, pp. 1541-1552.

Rouse, J. W.; R. H. Haas; J. A. Schell; and D. W. Deering, 1973, "Monitoring Vegetation Systems in the Great Plains with ERTS", Third ERTS Symp., Vol. I.

Rouse, J. W.; R. H. Haas; J. A. Schell; D. W. Deering; and J. C. Harlan, 1974, "Monitoring the Vernal Advancement and Retrogradation (Greenwave Effect) of Natural Vegetation, NASA/GSFC Type III Final Report, Greenbelt, MD.

Salas, S. L. and E. Hille, 1978, "Calculus: One and Several Variables - With Analytic Geometry", 3rd Edition, John Wiley and Sons, NY.

Seevers, P.M.; D. T. Lewis; and J. V. Drew, 1973, "Applications of ERTS-1 Imagery in Mapping and Managing Soils and Range Resources in the Sand Hills of Nebraska", Proceedings of the Symposium on Significant Results from ERTS-1, Vol. 1.

Tappan, Gray, 1980, "The Monitoring of Green Vegetation Cover in the Kansas Flint Hills from Landsat Data", KARS, Univ. of Kansas, Lawrence.

Thompson, D. R. and O. A. Wehmanen, 1978, "Detecting and Monitoring Agricultural Vegetative Water Stress Over Large Areas Using Landsat Digital Data", NASA/JSC-13737, LACIE-00506.

Tucker, C. J. and L. D. Miller, 1977, "Soil Spectra Contributions to Grass Canopy Spectral Reflectance", Photogrammetric Engineering and Remote Sensing, Vol. 43, No. 6, pp. 721-726.

Tucker, C. J. 1979, "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation", Remote Sensing of Environment, No. 8.

Watkins, A. H. and S. L. Freeden, 1979, "Landsat Data Users Handbook", Revised Ed., U.S.G.S., Washington, D.C.

Wiegand, C. L.; H. W. Gausman; J. A. Cuellar; A. H. Gerberman; and A. J. Richardson, 1974, "Vegetation Density as Deduced from ERTS-1 MSS Response", Third ERTS-1 Symposium, Vol. I.

Wiegand, C. L.; A. J. Richardson; and E. T. Kanemasu, 1979, "Leaf Area Index Estimates for Wheat from Landsat and Their Implications for Evapotranspiration and Crop Modeling", Agron. J. 71:336-342.

Wheeler, S. G.; P. N. Misra; and A. Q. Holmes, 1976, "Linear Dimensionality of Landsat Agricultural Data with Implications for Classification", Proc. of the Symp. on Machine Processing of Remotely Sensed Data, LARS, Purdue.

*Contribution of the Early Warning/Crop Condition Assessment (EW/CCA) project within the Agriculture Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program, a joint program of USDA, USDC, NASA, and USDI. EW/CCA is located at 1050 Bay Area Blvd., Houston, Texas 77058.

1981

DISCUSSION

C. H. Proctor, North Carolina State University

#1. Kalsbeek, Mendoza, and Budescu: New Model

The proposed new model is an expression that gives travel expenses as a function of size of interviewer assignment area, No. of PSU's in area, No. of PSU's to be visited in one trip and No. of callbacks to be made. It extends work of HH&M and is most welcome. The derivation seems eminently reasonable and actual survey expenditures should be found to follow the functional form, although further experiences or a review of existing ones would be required to establish the conformity between actual expenditures and the expression.

The authors compare recommended optimum PSU sizes based on three expressions for travel expenses: the simple one, the HHM and the new one. It seems that the recommendation based on ignoring travel expenses, the simple one, calls for too small PSU's, although there is no very serious loss of precision until the survey is taken to cover all of the US. In practice one may prefer to use the optimizing formula based on the simple model but also use some judgement in changing $C_1^{(S)}$ and $C_2^{(S)}$ so as to take account of travel expenses. This judgement could be sharpened by applying the authors' vision of how interviewers travel about.

#2. L. R. Ernst: Controlled Selection

The paper furnishes a way of tightening the control of the two-way stratification method given by Bryant, Hartley and Jessen (1960). I have been calling their method "merging random permutations" because of the way I carry it out. That is, a two-way stratification design selection can be exhibited as two columns of strata identifiers one for each "way." The two identifiers in each row point to a cell where a selection is to be made. By permuting the second column the cell selections are changed but the marginal selection numbers are "controlled." If there are, for example, two or more identifiers for strata in both ways then cells may be hit none, one, or two times and this may constitute too much loss of "control."

The author's method, if a solution exists, allows cells to be hit zero or one times, or one or two times, or two or three times, etc., but with no more flexibility relative to cell quotas. This method may be called "deep control," in parallel to the terminology "deep stratification" that describes multi-way subdivision of the population. I wonder if my merging random permutations approach could not be used after satisfying cell quotas up to none or one additional selection.

#3. Drummond: Workload Bias

The paper describes a variety of options for scheduling field work with a sympathetic appreciation for the realities of enumerating. The title of the paper suggests that imposed randomization might combat bias. Although I found expressions for inclusion probabilities, I don't believe there was even an expression for the estimator, much less its bias or variance. Since there is some cost to randomize, if only the looking at a random number, there ought to be some reduction in bias, if only a half of one percent. Some judgement of the probable amount of improvement would help in deciding whether to advocate the method.

In my own sample design practice I try to create subsamples (A series, B series, etc.) both replicated in Deming's sense and interpenetrating in Mahalanobis'. One instruction to the field workers would be to finish the A series before doing the B's, etc. A question may be posed as to the optimum number of subsamples to form with a sample of size n. There may be one, of size n, or two of size n/2, or three of size n/3 up to n of size n/n = 1. With r subsamples each of size n/r the instruction would be to assign the labels A, B, etc. randomly and then enumerate A series, B series, etc. One stops when money or time runs out and throws away the data on the incompleted subsample. Bias is always zero. The waste would be least for r = n but travel costs would also be maximized. What value of r is best?

#5. **Charles R. Perry: Information**

The author deals with recovery of ground-based data from photo interpretation of a satellite image. The data he uses as illustrative are qualitative, crop types, and he shows how Fisher's measure of information can be applied to characterize the quality of the photo interpretation. I confess I had not known how Fisher had introduced his notion of information and I enjoyed the author's presentation of Fisher's viewpoint. There have been questions raised as to the relative appropriateness of Fisher information as compared to the "n $\ell$og n" or communication theory type of information measure as used in Information Theory and Statistics by Kullback (1959). For example, Fisher's information rather unfortunately goes infinite as p goes to zero or one, while the communication theory type of quantity rises to zero as its maximum.

We could continue discussing "appropriateness" without settling much. What is needed is a clear statement of the problem and then we would be led to calculate some "best" estimate which might lead us to one or the other measure of information. When I described the problem to myself as one of having many, many photo interpreted pixels along with a few ground-based measurements and wishing to estimate the ground-based measurements over all many, many pixels, the sampling design was then recognized as the two-phase one, also called double sampling. Having named it, I looked into the JASA index and sure enough the problem had been answered for binary data by Aaron Tenebein (1970) 65: 1350.

Tenebein suggested a quantity K, the square of the correlation coefficient between the ground-based and the photo interpreted zero-one data, as a measure of quality of the photo interpretation. The variance formulas and optimum allocation of effort between phases become very simple expressions in terms of K. I suspect that there are still fertile fields of statistical investigation open to extend this model to polytomous (not just binary) data and also to three-phase (aerial photo too) sampling. For the present application, a particularly important extension would be the case in which estimates are needed for a number of



1981 Survey Section Proceedings, ASA

strata although ground-based measurements are available from some but not all strata. These extensions also direct our attention more to the proportions of various kinds of misclassifications as well as to a summary measure of agreement.

#6. Lautenschlager and Perry: Comparison of Vegetation Indices

The paper furnishes background information on remote sensing using the Landsat bands that I found most fascinating. The listing of indices was less gripping, but their clustering was of some reasonableness. Then the authors introduce the concepts of decision rules and equivalence classes that seem very close to the notions of test in statistical inference. I began to look for a comparison of indices in terms of, say, their asymptotic relative efficiencies, but couldn't find it. Actually the paper seemed to stop in mid-argument. It was marked "Working Draft" and perhaps the final version will arrive at some comparison of power or of efficiency.