AN ALGORITHM FOR COMPUTING CHLOROPHYLL a

CONCENTRATIONS USING A DUAL-FREQUENCY FLUOROSENSOR

Janet W. Campbell
NASA Langley Research Center

INTRODUCTION

The purpose of this paper is to recommend an algorithm to be used on data from a dual-frequency fluorosensor (i.e. one using two wavelengths for excitation of chlorophyll-a fluorescence) to compute total chlorophyll-a concentration and to partition that chlorophyll between two color groups present in a mixed phytoplankton population. The recommended algorithm is based on laboratory and field-testing experience gained with the Airborne Lidar Oceanographic Probing Experiment (ALOPE) fluorosensor at NASA's Langley Research Center.

As with the single-laser fluorosensor, an assumption must be made that the fluorescence efficiency of each color group remains constant over the area being "calibrated", but the dual-frequency technique can account for a shift in the overall or net fluorescence efficiency that would result from a shift in the relative abundance of two populations having different efficiencies. Therefore, the two-frequency technique can provide more accurate total chlorophyll estimates even if there is no interest in partitioning into color groups.

Partitioning of the total chlorophyll into color group components requires knowledge of the spectral characteristics of the fluorescence excitation spectra of the color groups. Techniques used with single-laser fluorosensors to calibrate the fluorescence signal using concurrent measurements of chlorophyll a (sea truth) cannot be extended to the dual frequency technique by simply changing the dimensionality of the equations. This is because there are no reliable techniques for providing sea truth values for the chlorophyll a concentrations of the two color groups. That is, if total chlorophyll is comprised of chlorophyll from two color groups with concentrations C_1 and C_2 such that the total chlorophyll concentration is $C_T = C_1 + C_2$, sea truth does not exist for C_1 and C_2 . Conventional techniques for obtaining total chlorophyll extract the pigment from all cells indiscriminately. Microscopic phytoplankton identifications or cell counts provide information on the relative abundance of the various species, and these can be classified (hopefully) into color groups, but the information on cell size distributions and chlorophyll per cell needed to translate this into component chlorophyll concentrations is virtually nonexistent. Clearly it is beyond the capability of conventional shipboard techniques.

This paper will describe algorithms for computing C_T , C_1 , and C_2 at progressively more quantitative levels depending on the amount of information available or assumed. The first or least quantitative level is that of real-time data that can show, without sea truth, the relative variation

of C_T, C_1, and C_2 along the flight track. At a higher level, an algorithm is presented for mapping total chlorophyll using two fluorescences which is more accurate than one using a single fluorescence when varying mixtures of phytoplankton color groups are present. This requires sea truth on total chlorophyll, but no additional assumptions. Finally, given sea truth on total chlorophyll and assumptions about the fluorescence excitation characteristics of the two color groups of phytoplankton present, an algorithm is presented for computing C_1 and C_2 as well as C_T. In both algorithms that use sea truth, the criterion chosen as the basis for deriving model parameters is to select those parameters that minimize the total squared error in chlorophyll a measured at the sea truth stations. That is, if $C_{T,k}$, $k = 1,\ldots,m$, are the sea truth measurements of total chlorophyll a at m sea truth stations, and $\hat{C}_{T,k}$, $k = 1,\ldots,m$, are estimates of total chlorophyll based on the algorithm used, then the algorithm parameters are selected to minimize

$$\text{Total Error} = \sum_{k=1}^{m}(C_{T,k} - \hat{C}_{T,k})^2 \tag{1}$$

This is the conventional unweighted least-squares solution where "error" is defined in terms of total chlorophyll measurements.

GENERAL LINEAR MODEL: THEORY

Fluorometric techniques for measuring chlorophyll a concentrations in living phytoplankton cells (in vivo) are based on the assumption that if the cells in a fixed volume of water are excited by light energy, the induced fluorescent energy emitted by the chlorophyll a per unit of excitation energy will be proportional to the molecular density or volumetric concentration of chlorophyll a. The fluorescence of chlorophyll a molecules is in a narrow spectral range centered at 685 nm. If the excitation source is effectively monochromatic, such as that provided by a laser, and if the excitation light frequency is varied, a fluorescence excitation spectrum is generated. Peaks in the spectrum correspond to absorbance peaks of auxilliary pigments present in the cell. Therefore, because the four major color groups of phytoplankton are characterized by the presence or absence of the auxilliary pigments, fluorescence excitation spectra can be used as a means of classifying the color groups. It was this fact that was the basis for the development of the ALOPE fluorosensor which utilized four distinct excitation wavelengths that were selected to discriminate among the four major color groups of phytoplankton. In practice a fluorosensor flown on an aircraft at altitudes generally around 150 m (500 ft) fires light pulses from a laser into the water and senses the returned fluorescence in several spectral bands. To apply fluorometric techniques, the chlorophyll fluorescence at 685 nm must be normalized to correct for variations in the laser's penetration depth along the track. The best-known technique for accomplishing this is to divide the 685-nm fluorescence by the Raman scattering produced by the laser which, in theory, is proportional to the number of water molecules accessed

by the laser or, equivalently, the penetration depth of the laser energy. This technique also corrects for any variations in the output energy of the laser.

In a dual-excitation frequency technique corrections must also be made for spectral differences in the light penetration and for any differences in the excitation energies of the two sources. All of these corrections are accomplished if the Raman signal from each laser is used to normalize fluorescences produced by that laser. It will be assumed in the remainder of this paper that fluorescence refers to a normalized fluorescence where variations in penetration depth and excitation energy are accounted for.

The linear model used for fluorescence from a single-wavelength excitation is

$$F = b + aC_T \qquad (2)$$

where the fluorescence F is a linear function of the chlorophyll-$\underline{a}$ concentration C_T . The term b represents a background fluorescence which is not related to chlorophyll $\underline{a}$ and the slope a is related to the fluorescence efficiency of the chlorophyll $\underline{a}$. The parameters b and a are assumed to remain constant over a defined area. Concurrent measurements of C_T (sea truth) are regressed against corresponding fluorescences F , and the slope a and intercept b are estimated.

The model for a dual-frequency fluorosensor is

$$F_i = b_i + a_{i1}C_1 + a_{i2}C_2 \qquad (3)$$

where fluorescence resulting from laser i , F_i , is linearly related to the chlorophyll-$\underline{a}$ concentrations of the two color groups, C_1 and C_2 , and to a background fluorescence b_i induced by that laser. Again, the parameters b_i , a_{i1} and a_{i2} are assumed to be constants over the calibration area, and the parameters a_{i1} and a_{i2} , sometimes called cross sections, are related to the fluorescence efficiencies of the two color groups at excitation frequency i.

Experience with the ALOPE fluorosensor has shown that the ratio of fluorescences $R = F_2/F_1$ can be used as an indicator of the relative abundance of the two color groups. In cultures of a single species grown in the laboratory, where background fluorescences are assumed to be zero (i.e., $b_1 = b_2 = 0$), fluorescence ratios are equivalent to the cross-section ratio a_{2j}/a_{1j} for color group j . The two excitation frequencies 454 nm and 539 nm have been used extensively in ALOPE field tests to differentiate the golden-brown and green phytoplankton color groups, the two color groups commonly found in coastal and shelf waters. In the absence of any background fluorescence, the fluorescence ratio $F(539)/F(454)$ is approximately equal to 1.0 for a golden-brown population and 0.3 for a green population. Figure 1 is a plot of this fluorescence ratio along a 220-km long flight track that began in the Chesapeake Bay where a golden-brown diatom was the

predominant species and ended in clear offshore shelf waters where green species formed a significant component of the population. Note a shift in the ratio from values near 1.0 toward lower ratios around 0.45.

Before proceeding to a discussion of the algorithms, which assume that b_i and a_{ij} are constants, some discussion of this assumption is in order. The assumption that the fluorescence efficiencies of the color groups remain constant over the calibration area is necessary, just as it is in shipboard techniques that use flow-through fluorometers. Sea truth measurements of chlorophyll a serve essentially the same purpose as the chlorophyll extractions that are made periodically to calibrate the continuous fluorometer record.

A much more serious problem may be the assumption that background fluorescence is constant. Dissolved organic matter is known to have a broad-band emission spectrum that can significantly contribute to fluorescence at 685 nm, and, depending on the excitation wavelength used, the Raman signal can overlap the chlorophyll a fluorescence. Probably the best method for removing a variable background signal, i.e. isolating the fluorescence due to chlorophyll a, is to have sufficient spectral resolution in the emission spectrum above and below the 685 nm band. Then, as illustrated in figure 2, a varying background can be estimated and removed. In all the algorithms discussed here, b_1 and b_2 are assumed to be constants that can be estimated and removed by various techniques.

ALGORITHMS

Two situations will be considered in the algorithms that are presented. First, no assumptions will be made about the fluorescence excitation spectra of the two color groups. In this situation the algorithm can provide total chlorophyll, either its relative variation (without sea truth) or absolute variation (with sea truth). Second, it will be assumed that the ratios $R_j = a_{2j}/a_{1j}$ for both color groups (j = 1 and j = 2) are known constants. In this situation both relative and absolute estimates of C_1 and C_2 can be derived.

Situation 1. Estimation of Total Chlorophyll a

Without Assumptions About Cross-Section Ratios

Without sea truth data, plots of F_1 and/or F_2 versus distance provide information about the relative concentration of chlorophyll a along the flight track. Since a shift in the relative abundance of the two color groups with different fluorescence efficiencies can affect the assumed linearity of fluorescence with respect to total chlorophyll, a plot of F_2/F_1 versus distance can be used to delineate the portions of the flight track over which linearity can be assumed. That is, either F_1 or F_2 will be a valid relative measure of total chlorophyll over a region, provided F_2/F_1 is fairly constant.

420

Note that to compute fluorescence ratios the data must consist of fluorescence pairs representing the same locations along the flight track. Since the lasers are fired sequentially rather than simultaneously some preliminary interpolation is required to estimate the fluorescence return of the unfired laser to pair with a measured return.

Once sea truth values for total chlorophyll are available, it is recommended that a multiple linear regression equation

$$\hat{C}_T = \beta_0 + \beta_1 F_1 + \beta_2 F_2 \tag{4}$$

be derived rather than a simple linear regression of C_T on F_1 or F_2 alone. The reason for this is illustrated in figure 3. The symbols represent a hypothetical situation in which sea truth values of total chlorophyll concentration $C_T = C_1 + C_2$ are plotted against a fluorescence computed from the model

$$F_i = 1.0 + 1.2C_1 + 0.3C_2 \tag{5}$$

Line I represents the linear relationship that would exist if the population were exclusively from color group 1, and line II that for color group 2. Any mixture of groups 1 and 2 would result in a point lying between the two lines.

The cluster of points in figure 3 in the range $7 \leq C_T \leq 10$ was sampled from a patch that was predominantly color group 2, whereas the cluster near $C_T = 3$ was from a patch which was predominantly group 1. Since the efficiency of group 2 is lower than that of group 1, the increase in fluorescence resulting from a tripling of the chlorophyll concentration was offset almost entirely by an overall reduction in fluorescence efficiency. Clearly a regression line drawn through these data would be a poor representation of the true relationship, particularly outside the range of the measured chlorophyll, or if a high chlorophyll patch of color group 1 or a low chlorophyll region of color group 2 were encountered.

A multiple regression of C_T on F_1 and F_2 would prevent such errors. At each sea truth station, the relationship between $C_{T,k}$ and F_{1k}, F_{2k} should be modeled as

$$C_{T,k} = \beta_0 + \beta_1 F_{1k} + \beta_2 F_{2k} \tag{6}$$

for $k = 1,\dots,m$. The least-squares solution for the coefficients that minimizes errors in $C_{T,k}$ (see equation (1)), is given by

$$\beta = [F^t F]^{-1} F^t C_T \tag{7}$$

where $\beta = (\beta_0, \beta_1, \beta_2)^t$; F is the $m \times 1$ column vector of sea truth chlorophylls $C_{T,k}$. More explicitly, the formulas for the regression coefficients are

$$\beta_1 = \frac{S_{C_T}(\rho_1 - \rho_0\rho_2)}{S_{F_1}(1 - \rho_0^2)}$$

$$\beta_2 = \frac{S_{C_T}(\rho_2 - \rho_0\rho_1)}{S_{F_2}(1 - \rho_0^2)} \tag{8}$$

$$\beta_0 = \bar{C}_T - \beta_1\bar{F}_1 - \beta_2\bar{F}_2$$

where $\bar{C}_T$ and S_{C_T} are the mean and standard deviation, respectively, of the sea truth chlorophylls; $\bar{F}_i$ and S_{F_i} are the mean and standard deviation, respectively, of the corresponding F_i sample; ρ_0 is the linear correlation coefficient of F_1 and F_2 at sea truth stations; and ρ_i is the linear correlation coefficient of F_i and C_T at the truth stations.

Situation 2. Estimating C_T, C_1, and C_2

With Assumptions About Cross-Section Ratios

It is now assumed that the ratio of cross-sections a_{2j}/a_{1j} are known constants R_1 and R_2 for color groups 1 and 2, respectively. The equations governing the fluorescences are now

$$F_1 = b_1 + a_{11}C_1 + a_{12}C_2$$

$$F_2 = b_2 + R_1 a_{11}C_1 + R_2 a_{12}C_2 \tag{9}$$

At this point some further assumptions must be made regarding the background terms b_1 and b_2. One cannot solve for a_{11}, a_{12}, b_1, and b_2 using total chlorophyll sea truth data alone. One solution is to find a minimum F_1 and F_2 in the entire data set, assume that chlorophyll is zero at this location, and set

$$b_1 = \min F_1$$

$$b_2 = \min F_2 \tag{10}$$

Another solution is to assume a fixed ratio $R_0 = b_2/b_1$ based on known spectral characteristics of, say, dissolved organic material.

Assume first that b_1 and b_2 are estimated by other means (e.g., equation (10)) and then subtracted from the fluorescences. The resulting chlorophyll a fluorescences would then be given by

$$F_1 = a_{11}C_1 + a_{12}C_2$$

$$F_2 = R_1 a_{11}C_1 + R_2 a_{12}C_2 \qquad (11)$$

Solving these equations for C_1 and C_2 gives

$$C_1 = \frac{1}{a_{11}} \left[\frac{R_2 F_1 - F_2}{R_2 - R_1} \right]$$

$$C_2 = \frac{1}{a_{12}} \left[\frac{F_2 - R_1 F_1}{R_2 - R_1} \right] \qquad (12)$$

and summing these gives

$$C_T = \frac{U_1}{a_{11}} + \frac{U_2}{a_{12}} \qquad (13)$$

where

$$U_1 = \frac{R_2 F_1 - F_2}{R_2 - R_1}$$

and $\qquad (14)$

$$U_2 = \frac{F_2 - R_1 F_1}{R_2 - R_1}$$

Although sea truth values are required to derive estimates of a_{11} and a_{12} note that U_1 and U_2 are relative measures of C_1 and C_2, respectively, accurate to within scale factors $1/a_{11}$ and $1/a_{12}$. Thus, plots of U_1 and U_2 along a flight track can be computed to provide the relative variation of C_1 and C_2. The sum of U_1 and U_2, however, is not necessarily a good measure of total chlorophyll because the scale factors may differ significantly in magnitude. Like F_1 or F_2, the sum $U_1 + U_2$ is a good relative measure of total chlorophyll only in regions where F_2/F_1 is fairly constant, indicating a constant $C_2:C_1$ ratio.

Given sea truth $C_{T,k}$, $k = 1,\ldots,m$, and corresponding values of U_1 and U_2

$$U_{1k} = \frac{R_2 F_{1k} - F_{2k}}{R_2 - R_1}$$

$$\qquad (15)$$

$$U_{2k} = \frac{F_{2k} - R_1 F_{1k}}{R_2 - R_1}$$

the least-squares solution for $1/a_{11}$ and $1/a_{12}$ is

$$
\begin{bmatrix} a_{11}^{-1} \\ a_{12}^{-1} \end{bmatrix} = \left[U^t U \right]^{-1} U^t C_T
\tag{16}
$$

where U is the $m \times 2$ matrix whose kth row is $(U_{1k}\ U_{2k})$, and C_T is the $m \times 1$ column vector of sea truth chlorophyll measurements.

More explicitly, the solution for a_{11} and a_{12} is

$$
a_{11} = \frac{\Sigma U_1^2 \Sigma U_2^2 - (\Sigma U_1 U_2)^2}{\Sigma C_T U_1 \Sigma U_2^2 - \Sigma C_T U_2 \Sigma U_1 U_2}
$$

$$
a_{12} = \frac{\Sigma U_1^2 \Sigma U_2^2 - (\Sigma U_1 U_2)^2}{\Sigma C_T U_2 \Sigma U_1^2 - \Sigma C_T U_1 \Sigma U_1 U_2}
\tag{17}
$$

where all sums are over $k = 1, \ldots, m$. With these parameters, then, the U_1 and U_2 values can be converted to $C_1 = U_1/a_{11}$ and $C_2 = U_2/a_{12}$.

The final case considered assumes $R_0 = b_2/b_1$ to be a known constant. Here the fluorescences are

$$
F_1 = b_1 + a_{11} C_1 + a_{12} C_2
$$

$$
F_2 = R_0 b_1 + R_1 a_{11} C_1 + R_2 a_{12} C_2
\tag{18}
$$

Solving for C_1, C_2, and C_T gives

$$
C_1 = \frac{U_1}{a_{11}} - \frac{b_1}{a_{11}} \left[\frac{R_2 - R_0}{R_2 - R_1} \right]
$$

$$
C_2 = \frac{U_2}{a_{12}} - \frac{b_1}{a_{12}} \left[\frac{R_0 - R_1}{R_2 - R_1} \right]
\tag{19}
$$

and

$$
C_T = \frac{U_1}{a_{11}} + \frac{U_2}{a_{12}} + \beta_0
\tag{20}
$$

where β_0 is the sum of the constant terms in equation (19).

A multiple linear regression of sea truth chlorophyll values on U_1 and U_2 similar to that described above in equations (6) to (8) would provide solutions for β_0, $\beta_1 = 1/a_{11}$ and $\beta_2 = 1/a_{12}$.

Explicitly, the solutions are

$$a_{11} = \frac{S_{C_T}(\rho_1 - \rho_0\rho_2)}{S_{U_1}(1 - \rho_0^2)}$$

$$a_{12} = \frac{S_{C_T}(\rho_2 - \rho_0\rho_1)}{S_{U_2}(1 - \rho_0^2)} \tag{21}$$

$$b_1 = \frac{\bar{C}_T - \bar{U}_1/a_{11} - \bar{U}_2/a_{12}}{U_0/a_{11} + (1 - U_0)/a_{12}}$$

where $U_0 = (R_0 - R_1)/(R_2 - R_1)$, and ρ_i are linear correlation coefficients for U_i and C_T at sea truth stations.
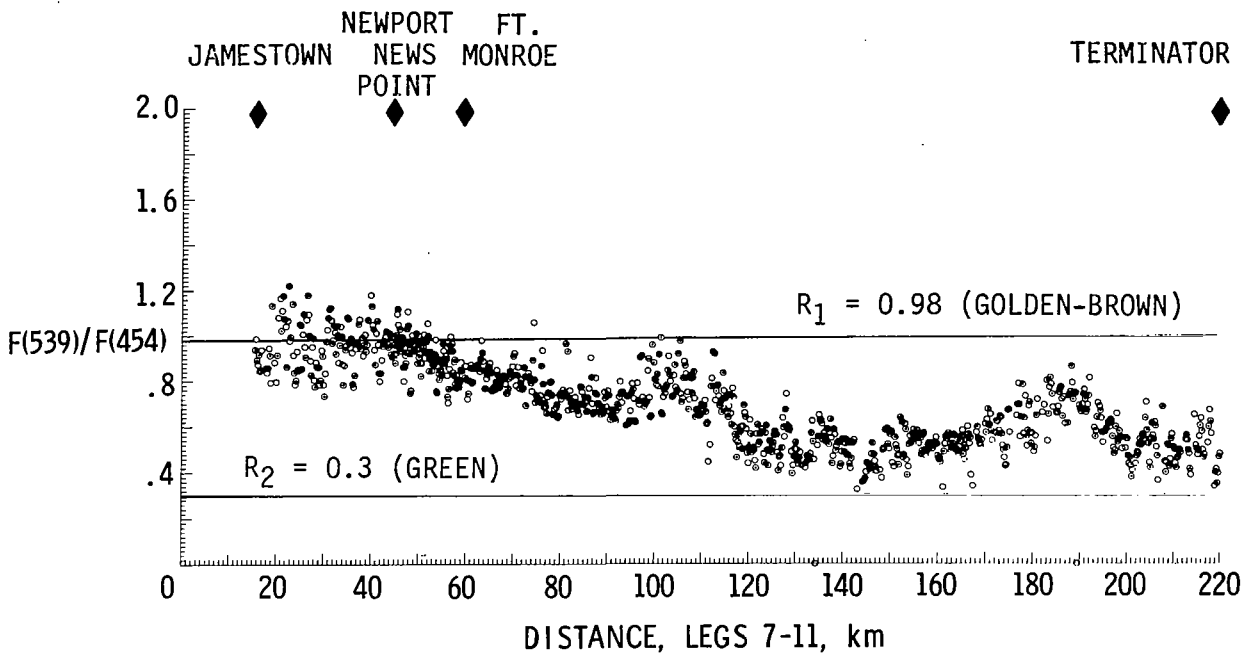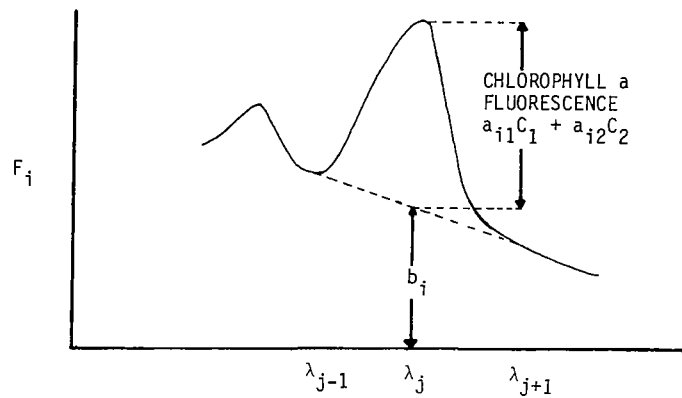
Figure 1.- Fluorescence ratios for the March 17, 1980 James River/shelf mission.

If fluorescence emission spectrum is measured at wavelengths λ_{j-1}, $\lambda_j = 685$ nm, λ_{j+1}, then an estimate of b_i = background is:

$$b_i = \frac{(\lambda_{j+1} - \lambda_j)F_{i,j-1} + (\lambda_j - \lambda_{j-1})F_{i,j+1}}{\lambda_{j+1} - \lambda_{j-1}}$$

Figure 2.- Use of neighboring bands in emission spectrum to estimate background.
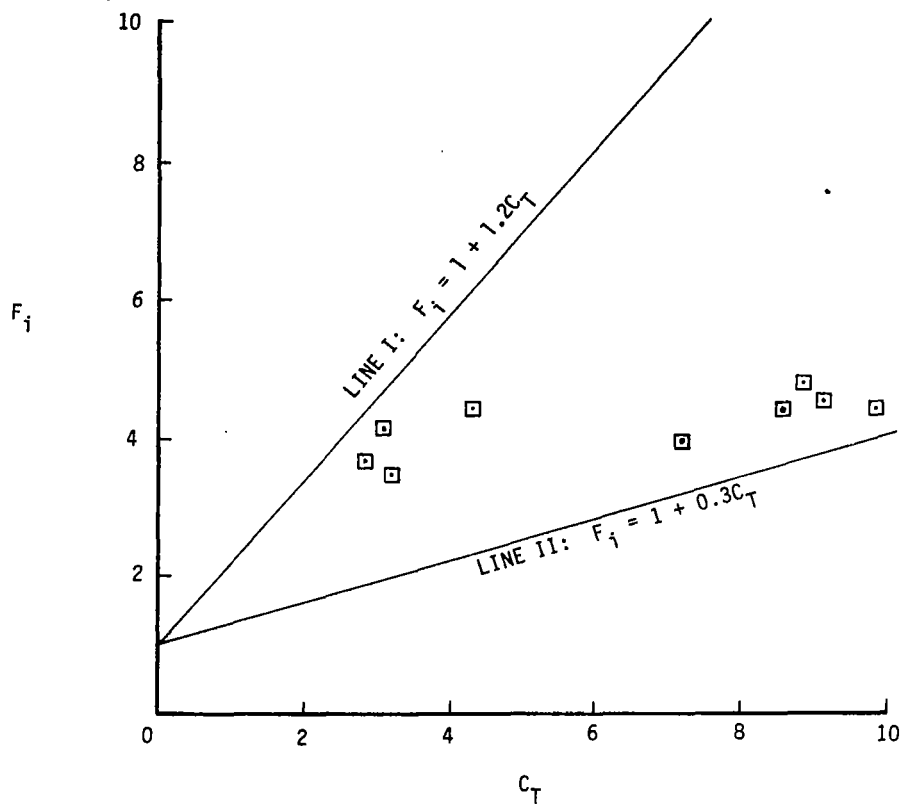
Figure 3.- Illustration of the relationship between fluorescence and total chlorophyll when there is a variation in the relative abundance of two color groups with different fluorescence efficiencies.