SECTION 105

REMOTE MEASUREMENT OF CHLOROPHYLL CONCENTRATION AND SECCHI-DEPTH USING THE PRINCIPAL COMPONENTS OF THE OCEAN'S COLOR SPECTRUM

by

James L. Mueller
Oceanography Department
Oregon State University
Corvallis, Oregon

INTRODUCTION

Several investigators have published data and theoretical arguments which suggest that a correlation exists between oceanic chlorophyll concentration and the ocean's color spectrum. There is a need however, for a more quantitative description of the nature and variability of the relationship. In this paper are presented some statistical results and conclusions, based on direct comparisons of the ocean's color and chlorophyll concentration, which may help to fill this gap.

The ocean color and ground-truth data for this analysis were collected during Mission 140 of NASA's NP3A earth resources aircraft over the period from 6 through 14 August 1970. During this experiment, chlorophyll and light attenuation data were collected by Oregon State University's R/V CAYUSE and the chartered R/V JUDY K. All sets of comparative observations are simultaneous in the sense that the ship began sampling when the aircraft came overhead.

This investigation is supported by the U. S. Naval Oceanographic Office (Spacecraft Oceanography Project) contract number N62306-70-C-041.

The use of a spectrometer and the valuable services of Mr. Richard Ramsey were provided, under subcontract, by TRW Systems Group, TRW INC., Redondo Beach, California.

DESCRIPTION OF APPARATUS

Ocean color spectra were observed with an off-plane,Ebert type, blazed diffraction grating spectrometer designed and built by TRW Inc. The spectral resolution of the spectrometer is between 5 and 7.5 nannometers. Approximately one second is required to observe one spectrum between 750 and 400 nannometers, and approximately 3 seconds elapse after the start of one spectrum until the start of the next.

The output from the TRW spectrometer was recorded both on a Sanborn strip chart recorder and on analog magnetic tape. The strip chart records were used for the present investigation.

Pigment samples were collected by filtering water from Van Dorn sampling bottles, with subsequent handling as described in Strickland and Parsons (1965,pp117-127).

Light attenuation data were observed aboard the ships both using a submersible photometer (flat-plate detector) and a standard white secchi-disk.

PROCEDURES AND METHODS OF ANALYSIS

Water samples were taken, and light attenuation measured by the crew of each research vessel simultaneously with the NP3A aircraft's arrival overhead for a station. Aboard the aircraft, the TRW spectrometer was directed away from the sun's azimuth and tilted 15° from nadir to avoid the sun glitter pattern. On each station the aircraft flew two or more data runs, each ideally being about two miles in length centered about the ship's position.

CHLOROPHYLL DATA PROCESSING

Filtered pigment samples were frozen and subsequently analysed for chlorophylls a, b, and c concentrations by the methods described in Strickland and Parsons (1965, pp117-127). The samples were also analysed for plant and animal caratenoid concentrations, but since these were uniformly less than 1 mg/m^3 we have ignored caratenoids in this first analysis.

The blue absorption bands of the three chlorophylls overlap considerably, and their distinctive red bands are overwhelmingly masked by water's aborption of red light. Therefore, we have assumed that the effect of a given concentration of total chlorophyll (a + b + c) on water color will not depend on the proportions of a, b, and c in a first approximation.

The effect of water and its contents on the daylight spectrum is a function of the pathlength followed by the light, i.e. upon some color producing depth to which light penetrates and is backscattered up into the atmosphere. It is intuitively reasonable to assume that a spectrometer responds to light backscattered from a color-producing depth about the same as the depth which a human eye can see. On this reasoning then, we have assumed color-producing depth to equal secchi-depth. Secchi-depth is the depth at which a large white disk may barely be seen by an observer above the water surface.

Hence, the measure of chlorophyll concentration which we have attempted to relate to the ocean's color spectrum is Chlorophyll (a + b + c) averaged over secchi-depth. In the sequel we will refer to this quantity simply as "chlorophyll concentration".

Our chlorophyll data were collected at discrete depths, making it necessary for us to approximate a continuous distribution by assuming a linear trend between observations. This profile was then truncated at secchi-depth and averaged as though the data had been collected continuously.

OCEAN COLOR SPECTRA PROCESSING

The strip charts from the TRW spectrometer system were calibrated for wavelength, and then digitized on-line to Oregon State University's CDC3300 computer. A program was developed to convert each spectrum into a vector of 70 irradiance components, each representing a 5-nm band between 400 and 750 nm. The same program corrected each spectrum for amplifier gain setting, dark-voltage error (a function of wavelength and slit number), variations in slit width, and calibrated photomultiplier response (a function of wavelength).

Data Smoothing

During the one second required to observe a single spectrum, the aircraft's forward motion caused the spectrometer to scan a strip of water about 100 meters long. This phenomenon caused each component of each spectrum to represent a distinctly different area on the ocean's surface. So if whitecaps are randomly distributed on the sea surface, or if small-scale random fluctuations in cloud cover cause fluctuations in incident irradiance, fluctuations will occur randomly in wavelength in the observed upwelled light spectra. This process is illustrated in Figure I.

Random fluctuations in cloud densities can, of course, occur on scales larger than 100 meters. This causes a variation in overall irradiance levels, but individual spectra are not contaminated internally. Variations in irradiance level make it impossible to directly compare different spectra though.

The second noise effect is routinely compensated for by normalizing each spectrum with respect to irradiance observed at some reference wavelength, e.g. 570 nm. This practice is risky in the presence of small scale fluctuations, however, for we have no assurance that the normalization wavelength is not contaminated in some non-average way. To avoid this difficulty, we departed from custom and normalized each spectrum with respect to its own mean irradiance.

Small scale whitecap and cloud induced noise is not so easily smoothed over, but we have developed a satisfactory method of doing so. If we average within each wavelength band over several spectra, then each component of the average spectrum should contain an average amount of small scale noise.

Occasionally the spectrometer viewed an unusually large whitecap, or a ship. Any such occurrence caused an anomalously large fluctuation that was not representative of the average conditions for the set of spectra being considered. The filtering problem we faced was to identify these anomalous fluctuations, and then to remove their effects from the smoothed data.

An "outlier" is an observation determined by some statistical criterion to belong to a population different from that to which the rest of the observations in the sample belong. Anomalously large fluctuations in our data were regarded as outliers, and the following iterative smoothing procedure was adopted:

1. After each spectral component was expressed as a fraction of the mean for that spectrum, natural logarithms were taken of the data to improve the normality of distribution. After this transformation, the variance behaved linearly as a function of wavelength over the regions 420-580nm and 580-700nm.

2. For each wavelength band, the variances from all data runs for a given station were pooled. Then a least squares linear regression was performed to relate variance to wavelength. The regression estimates of s^2 (sample variance) were then used to compute, for each wavelength band, the statistic:

$$O_{\pm} = \frac{\max |I - \bar{I}|}{s}$$

where I is relative irradiance for a particular wavelength band and spectrum, and $\bar{I}$ is the mean for that wavelength band and data run. Critical values of this statistic are tabulated in Halperin, et al (1955). For simplicity we assumed 25 degrees of freedom for the regression estimate of s, even though we might argue for a larger number on the strength of the large effective sample size.

3. When for any band $O_{\pm}$ exceeded the tabulated critical value, the rejected component was examined in the original, untransformed data matrix. The anomalous fluctuation associated with that outlier was then removed by substitution of values which made the offending spectrum behave locally in a manner consistent with the behavior of the rest of the spectra in the sample. This usually involved adjustment of a few data points adjacent to the outlier, even though these were not rejected on the basis of their own magnitudes.

4. The entire process was repeated iteratively until no outliers were detected. The successive smoothing of background noise allowed discovery of outliers in secondary iterations that were undetectable against the noise level of the earlier iteration(s).

The final step in smoothing the ocean color spectra was to examine each data run for linear trends which might be attributable to horizontal gradients of chlorophyll concentration. Least squares regression coefficients were calculated for each wavelength band in each data run, and then these coefficients were smoothed by taking a simple moving average over wavelength. The smoothed regression coefficients were used, finally, to adjust the smoothed mean spectrum to correspond to the position in the data run occupied by the research vessel.

Principal Component Analysis

For any sample of N-component vectors, the N N-component eigenvectors of the sample covariance matrix for the bases of an orthogonal coordinate system into which the original observations may be transformed without any loss of information. When extracted, this eigensystem will always be aligned so that the first eigenvector, e_{1j} $(j = 1,2,...N)$, defines the direction of maximum sample variance. Further, e_{2j} defines the direction of the next largest amount of sample variance orthogonal to e_{1j}, and so forth with each successive eigenvector representing a direction of lesser variance than any of its predecessors. Because of this property, if the sample variations occur in definite modes, i.e. if the original N-variables vary together in definable ways, then most of the variation may be accounted for in terms of only the first few eigenvectors.

An original observation vector X_j $(j = 1,2,....,N)$, may be transformed into its eigensystem representation through the k equations:

$$Y_i = e_{ij} (X_j - E_j); \text{ where } j = 1,2,....,N; \; i = 1,2,....,k \qquad (1)$$

and where E_j is the sample mean vector, e_{ij} is the i^{th} eigenvector, and Y_i is called the i^{th} principal component of X_j. Thus we effect the transformation:

$$(X_1,X_2,.....,X_N) \longrightarrow (Y_1,Y_2,.....,Y_k)$$

where k is selected less than or equal to N to retain the desired proportion of sample variance. When k is considerably less than N, we gain a significant reduction in the number of variables in exchange for a defined, and presumably acceptable, loss of sample variation.

Any original vector X_j may be recovered from its principal component representation Y_i through the N equations

$$X_j = E_j + Y_1 e_{1j} + Y_2 e_{2j} + ...+ Y_k e_{kj} \qquad (2)$$

where $j = 1,2,.....,N$.

Morrison (1967) gives a thorough, yet readable, introduction to principal component analysis. Simonds (1963) and Church (1966) discuss examples similar to the present application, i.e. analysis of data recorded in the form of a curve. This method is alternativel referred to as "principal component analysis", "eigenvector analysis", or "characteristic vector analysis" in the literature.

It is possible to apply regression results obtained using the principal components of one sample to subsequently observed samples. Simply regard the mean vector of the original sample (E_j) as the origin of the coordinate

system and apply equation (1), using the eigenvectors (e_{ij}) obtained from the original sample. For example, to estimate chlorophyll concentration, using equation (3) below, from any observed ocean color spectrum, use the values of E_j, e_{1j}, and e_{2j} given in Figure II. With these values determine Y_1 and Y_2 from equation (1) and substitute into equation (3).

Multiple Regression Analysis

Linear least squares multiple regression analyses were performed using standard methods. The first "k" principal components Y_j were treated as independent variables, with alternatively chlorophyll concentration (C) or secchi-depth (Z_s) as dependent variable, yielding equations of the form

$$C = A + B_1Y_1 + B_2Y_2 + \ldots\ldots + B_kY_k$$

The squared multiple correlation coefficient (R^2) was calculated to estimate the percentage of variability in the dependent variable accounted for by the regression equation.

We also calculated, as a measure of precision, the residual standard deviation

$$S_{c.y} = \left[\frac{\sum_{m=1}^{n}(C - \overset{*}{C})_m^2}{n - q}\right]^{\frac{1}{2}}$$

where C and $\overset{*}{C}$ are respectively the observed value and regression estimate of the dependent variable, n is the number of observations, and $q = k + 1$ is the total number of variables in the regression equation.

Finally, simultaneous confidence limits were calculated for the regression coefficients. When such a confidence interval contained zero, it was regarded as strong evidence of no significant correlation between the associated independent variable and the dependent variable, allowing that term to be dropped from the equation.

DISCUSSION OF RESULTS

A principal component analysis was performed on the sample of 31 trend-adjusted mean color spectra, each having 55-components between 420 and 695nm. The mean vector (E_j) and first two eigenvectors (e_{ij}; i = 1,2; j = 1,2,.....,55) are given in Figure II.

The first eigenvector (e_{1j}) accounts for 75% of sample variance, and the second eigenvector (e_{2j}) accounts for an additional 20%. Thus, 95% of the total variance in the sample of 31 55-component color vectors ($X_1, X_2, \ldots\ldots, X_{55}$) is contained in 31 pairs (Y_1, Y_2).

CHLOROPHYLL CONCENTRATION

A regression analysis was performed using $\ln(0.508 + C)$ as the dependent variable, and Y_1 and Y_2 as independent variables, with the result

$$\ln (0.508 + C) = 1.0085 - 0.5149\ Y_1 + 1.2790\ Y_2 \qquad (3)$$

where C is chlorophyll concentration in mg/m^3, and the arbitrary constant 0.508 is the ratio of the average specific absorption coefficients of water and chlorophyll. Equation (3) is significant at the 0.005 level and accounts for 77% of the variability in the observed chlorophyll data (over the range from 0.00 mg/m^3 to 8.43 mg/m^3). The residual standard deviation was ± 1.62 mg/m^3. These results are collected in Table I and illustrated in the scatter diagrams of Figure III. In Figure III note that residual deviation is measured from the plane of the two regression lines; a glance at only one marginal distribution can be misleading.

SECCHI-DEPTH

Ocean color is produced by wavelength selective processes acting over color producing depth, which we have assumed to be equal to secchi-depth. Therefore we expect a strong correlation between secchi-depth and the ocean color spectrum.

As a first step we plotted the scatter diagrams shown in Figure IV. These graphs suggest a partitioning of the sample, with classification according to which research vessel collected the ground truth data. This distinction is a result of the different ocean environments in which the two vessels operated. The JUDY K operated exclusively within a few miles of the mouht of the Columbia River, where we may assume uniformly high densities of suspended particles. The CAYUSE on the other hand, operated in a more oceanic regime well away from the river mouth, where we may assume relatively low densities of suspended particles. We have tentatively and arbitrarily described the two environments as respectively a particle-scattering dominated and an absorption dominated ocean color system.

A multiple regression analysis of secchi-depth (Z_s) versus Y_1 and Y_2 for the particle-scattering dominated subsample yielded

$$Z_{s(p)} = 5.483 + 1.768\ Y_1 \qquad \text{meters} \qquad (4)$$

where the regression coefficient of Y_2 was shown to not be significantly different from zero by simultaneous confidence intervals. Equation (4) accounts for only 55% of the variability in secchi-depth, but the sample range was very narrow (2.5 to 6.5 meters). The residual standard deviation is only $\pm$ 1 meter. These results are collected in Table IIa and illustrated in Figure IVa.

A similar analysis of the absorption dominated subsample yielded

$$Z_{s(a)} = 9.214 - 7.833\ Y_2 \qquad \text{meters} \qquad (5)$$

where the regression coefficient of Y_1 was shown to be not significantly different from zero by simultaneous confidence intervals. Equation (5) accounts for 82% of the variability in secchi-depth over a range from 6 to 20 meters. Residual standard deviation is $\pm$ 2.16 meters. These results are collected in Table IIb and depicted in Figure IVb.

INTERPRETATION AND APPLICATION

Through joint use of equations 3 through 5 it is possible to estimate chlorophyll concentration averaged over secchi-depth, estimate secchi-depth (i.e. the depth of the layer in which we are estimating chlorophyll), and say somethingabout particle concentrations (at least qualitatively).

For routine application of these results, it should be possible to estimate Y_1 and Y_2 from irradiance measurements in three narrow wavelength bands (I_1, I_2, I_3).[2] By solving the three equations

$$\ln I_j = \ln \bar{I} + E_j + Y_1 e_{1j} + Y_2 e_{2j}$$

for Y_1 and Y_2 ($\ln \bar{I}$ is also unknown, but it need not be solved for explicitly), estimates may be obtained for substitution in equations (3) through (5). Values for E_j, e_{1j} and e_{2j} are to be taken from Figure II for the appropriate wavelengths.

For example, if irradiances are measured in 3 10nm-wide wavelength bands centered on 495nm (j=1), 547.5nm (j = 2) and 602.5nm (j=3), then

$$Y_1 = \frac{1}{0.300}\left\{\ln\left(\frac{I_1}{I_2}\right) + 0.475\right\}$$

$$Y_2 = \frac{1}{0.208}\left\{\ln\left[\left(\frac{I_2}{I_1}\right)\left(\frac{I_1}{I_3}\right)^{\frac{1}{2}}\right] - 0.330\right\}$$

which is nearly as simple as attempting to measure chlorophyll from only a single measured ratio of irradiances.

CONCLUDING REMARKS

We have shown by empirical means that it is possible to measure chlorophyll concentration in the ocean as a linear function of the first two principal

components of the ocean's color spectrum. The residual standard deviation of the estimate is about $\pm$ 1.6 mg/m^3.

The chlorophyll concentration estimated is the average over secchi-depth, which may also be estimated: a) within about 25% in very turbid waters as a linear function of Y_1, and b) within about 15% in relatively clear ocean waters as a linear function of Y_2.

For routine application of these results, Y_1 and Y_2 may be estimated easily from irradiance measurements in only three narrow wavelength bands. A 3-channel measurement will not suffice for future investigations aimed at broadening or improving these results, however. A full spectrum analysis by methods similar to those presented here is recommended for that task.

The relatively poor fit to the chlorophyll data ($\pm$ 1.6 mg/m^3) suggests that we next conduct a covariance analysis to examine the joint effects of chlorophyll, particle scattering and secchi-depth on ocean color.

REFERENCES

Church, Alonzo Jr. (1966). Analysis of data when the response is a curve. TECHNOMETRICS, Vol 8 No 2 pp 229-246

Halperin, M., S.W. Greenhouse, J. Cornfield, and J. Zalokar (1955) Tables of percentage points for the studentized maximum absolute deviate in normal samples. AMER. STATIST. ASSN. JOUR. March 1955, pp185-195

Morrison, D. F. MULTIVARIATE STATISTICAL METHODS McGraw Hill, New York (1967) 338p

Simonds, J. L. (1963) Application of characteristic vector analysis to photographic and optical response data. JOUR. OPT. SOC. AM. Vol 53 No. 6 pp 968-974

Strickland, J.D.H., and Parsons, T.R. (1965). A manual of sea water analysis. FISH RES. BD. CANADA BULL. No. 122, 172p

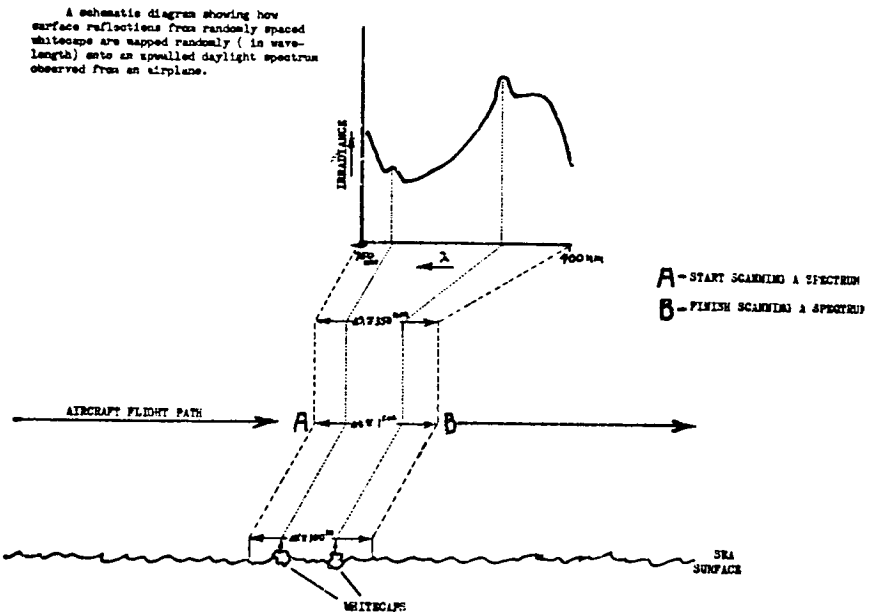
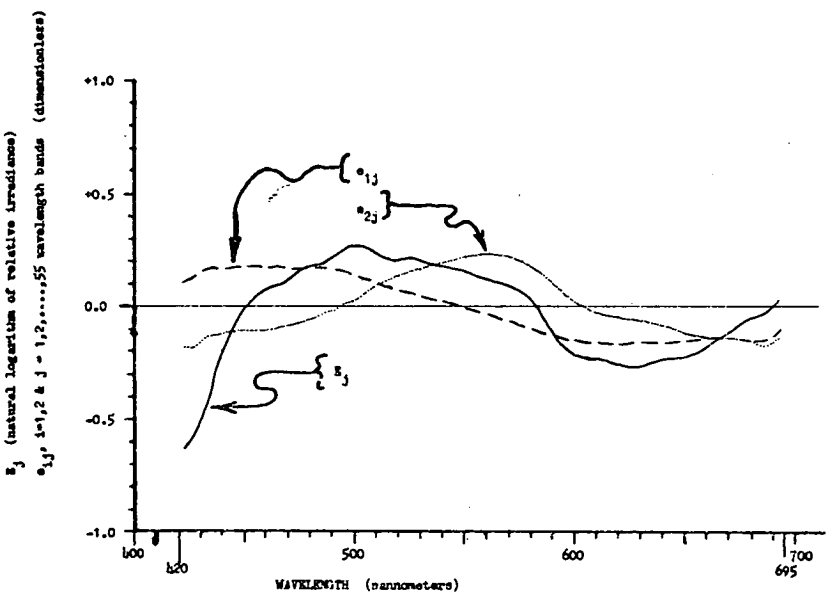

A schematic diagram showing how surface reflections from randomly spaced whitecaps are mapped randomly (in wavelength) onto an upwelled daylight spectrum observed from an airplane.
IRRADIANCE
λ
400 nm
A - START SCANNING A SPECTRUM
B - FINISH SCANNING A SPECTRUM
AIRCRAFT FLIGHT PATH
A
B
SEA SURFACE
WHITECAPS

FIGURE I.



THE EIGENSTRUCTURE OF OCEAN COLOR, as determined from a sample of 31 mean spectra.

E_j = EIGENSYSTEM ORIGIN, taken to be the SAMPLE MEAN VECTOR

e_{ij} = FIRST (i = 1) & SECOND ORTHONORMAL BASIS VECTORS of the EIGENSYSTEM, computed as the first two EIGENVECTORS OF THE SAMPLE COVARIANCE MATRIX

FIGURE II.

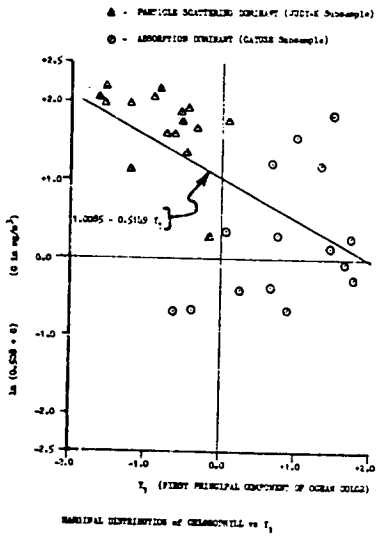


FIGURE III(a)

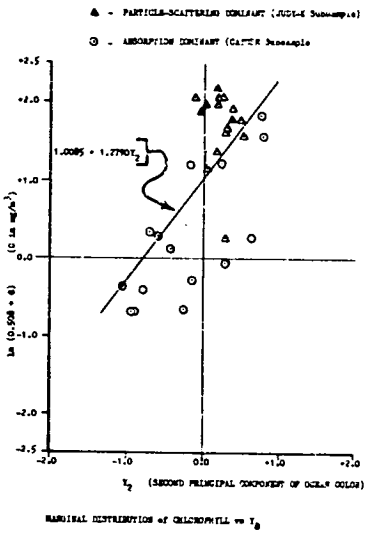


FIGURE III (b)

TABLE I

CHLOROPHYLL CONCENTRATION and the PRINCIPAL COMPONENTS OF OCEAN COLOR

MULTIPLE REGRESSION ANALYSIS RESULTS

$\ln (0.508 + C) = 1.0085 - 0.5149\ Y_1 + 1.2790\ Y_2$

Significant at the 0.005 level.

$R^2 = 0.77$ (percent of variance in chlorophyll concentration accounted for by the regression equation)

$R = 0.88$

$S_{c.y} = \pm 1.62$ mg-chl/m^3 (standard deviation from regression plane)

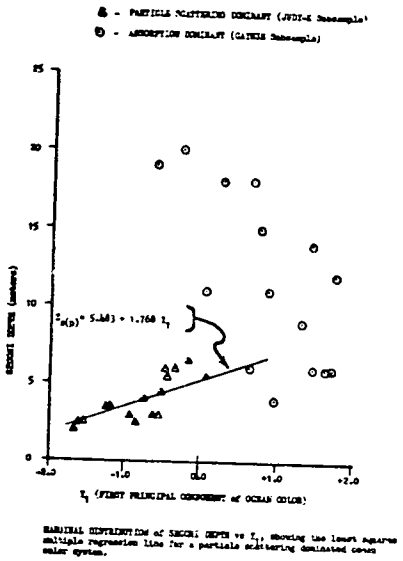
$S_{c.y} = 19\%$ of sample range.

Simultaneous Confidence Limits on Regression Coefficients:

97.5% CI on β_1 : $-0.9487 \leq \beta_1 \leq -0.0811$

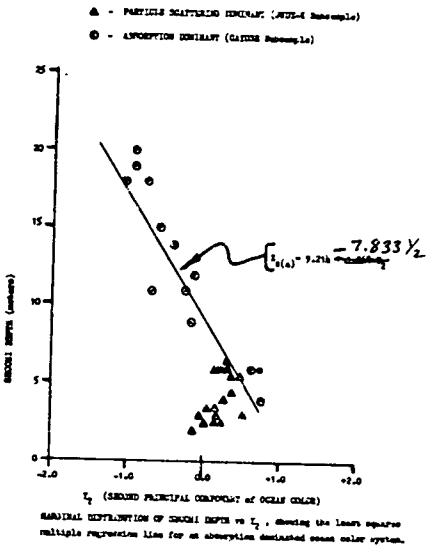
97.5% CI on β_2 : $0.4487 \leq \beta_2 \leq 2.1093$

C = Sum of chlorophylls a, b, and c averaged over secchi depth, which is assumed equal to color producing depth.



MARGINAL DISTRIBUTION of SECCHI DEPTH vs Y_1, showing the least squares multiple regression line for a particle scattering dominated ocean color system.

FIGURE IVa



MARGINAL DISTRIBUTION OF SECCHI DEPTH vs Y_2, showing the least squares multiple regression line for an absorption dominated ocean color system.

FIGURE IVb

TABLE II

(a)

MULTIPLE REGRESSION ANALYSIS RESULTS

$Z_{s(p)} = 5.883 + 1.768\ Y_1$ (meters)

Significant at the 0.01 level.

$R^2 = 0.55$ (percent of sample variance in $Z_{s(p)}$ accounted for by the regression equation)

$R = 0.74$

$S_{e.y} = \pm 0.98$ meters (standard deviation from regression line)
$S_{e.y} = 25\%$ of sample range.

Simultaneous Confidence Limits on Regression Coefficients:

90% on β_1: $0.341 \leq \beta_1 \leq 3.195$

90% CI on β_0 included zero, indicating that secchi depth is not significantly correlated with Y_1 in a particle scattering dominated ocean color system.

(b)

MULTIPLE REGRESSION ANALYSIS RESULTS

$Z_{s(a)} = 9.214 - 7.833\ Y_2$

Significant at the 0.005 level.

$R^2 = 0.82$ (percent of variance in $Z_{s(a)}$ accounted for by the regression equation.)

$R = 0.90$

$S_{e.y} = \pm 2.152$ meters (standard deviation from regression line.)
$S_{e.y} = 13\%$ of sample range

Simultaneous Confidence Limits on Regression Coefficients:

90% CI on β_1 included zero, indicating that secchi depth is not significantly correlated with Y_1 in an absorption dominated ocean color system.

95% on β_2: $-11.991 \leq \beta_2 \leq -1.675$