



3 1176 00154 5467

NASA Technical Memorandum 80101

NASA-TM-80101 19790020623

A REGRESSION TECHNIQUE FOR EVALUATION AND
QUANTIFICATION FOR WATER QUALITY PARAMETERS
FROM REMOTE SENSING DATA

Charles H. Whitlock and Chin Y. Kuo

May 1979

LIBRARY COPY

MAY 1979

LANGLEY RESEARCH CENTER
LIBRARY, NASA
HAMPTON, VIRGINIA



National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23665



NF00659

1 Report No NASA TM-80101		2 Government Accession No		3 Recipient's Catalog No	
4 Title and Subtitle A Regression Technique for Evaluation and Quantification for Water Quality Parameters from Remote Sensing Data				5 Report Date May 1979	
				6 Performing Organization Code	
7 Author(s) Charles H. Whitlock and *Chin Y. Kuo				8 Performing Organization Report No	
9 Performing Organization Name and Address NASA Langley Research Center Hampton, Virginia 23665				10 Work Unit No	
				11 Contract or Grant No	
				13 Type of Report and Period Covered Technical Memorandum	
12 Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546				14 Army Project No	
15 Supplementary Notes *Chin Y. Kuo is affiliated with the Department of Civil Engineering, Old Dominion University, Norfolk, Virginia. This paper was presented at the Thirteenth International Symposium on Remote Sensing of the Environment, April 23-27, 1979, Ann Arbor, Michigan.					
16 Abstract <p>Inconsistent results have been obtained from previous experiments which have applied linear multiple regression techniques to remote-sensing data for quantification of water quality parameters. The objective of this paper is to define optical physics and/or environmental conditions under which the linear multiple regression should be applicable. To achieve this objective, an investigation of the signal-response equations was conducted and the concept was tested by application to actual remote-sensing data from a laboratory experiment performed under controlled conditions.</p> <p>From investigation of field multispectral scanner data from past experiments, it is concluded that instrument noise, ground-truth placement, and time lapse between remote-sensor overpass and water-sample operations are serious barriers to successful use of the technique. Recommendations are made which should aid in obtaining consistent results from future remote-sensing experiments.</p>					
17 Key Words (Suggested by Author(s)) Remote Sensing Water Quality Measurement Data Analysis			18 Distribution Statement Unclassified-Unlimited Subject Category 45		
19 Security Classif (of this report) Unclassified	20 Security Classif (of this page) Unclassified	21 No of Pages 15	22 Price* \$4.00		

A REGRESSION TECHNIQUE FOR EVALUATION AND QUANTIFICATION
FOR WATER QUALITY PARAMETERS FROM REMOTE SENSING DATA

by

Dr. Charles H. Whitlock
Marine Environments Branch
NASA Langley Research Center
Hampton, Virginia 23665

and

Dr. Chin Y. Kuo
Department of Civil Engineering
Old Dominion University
Norfolk, Virginia 23508

ABSTRACT

Inconsistent results have been obtained from previous experiments which have applied linear multiple-regression techniques to remote sensing data for quantification of water quality parameters. The objective of this paper is to define optical physics and/or environmental conditions under which the linear multiple-regression should be applicable. To achieve this objective, an investigation of the signal-response equations is conducted and the concept is tested by application to actual remote sensing data from a laboratory experiment performed under controlled conditions.

Investigation of the signal-response equations shows that the exact solution for a number of optical physics conditions is of the same form as a linearized multiple-regression equation, even if nonlinear contributions from surface reflections, atmospheric constituents, or other water pollutants are included. Limitations on achieving this type of solution are defined. Since the exact solution is in the form of a linear multiple-regression equation, applicability of multiple-regression techniques to remote sensing and ground-truth data is viewed as a calibration of the exact solution to account for daily variations in background constituents.

Application to laboratory data is used to demonstrate that the technique is applicable for water mixtures which contain constituents with both linear and nonlinear radiance gradients.

From investigation of field multispectral scanner data from past experiments, it is concluded that instrument noise, ground-truth placement, and time lapse between remote sensor overpass and water sample operations are serious barriers to successful use of the technique. Recommendations are made which should aid in obtaining consistent results from future remote sensing experiments.

1. INTRODUCTION

Large amounts of sediments and other pollutants are carried annually in the rivers, lakes, estuaries, and coastal waters of the United States. These sediments and pollutants are major determinates of water quality. Many agencies are investigating the potential of using remote sensing techniques to monitor various water quality parameters because of the ability of remote sensing to provide synoptic views over large areas.

Specific data needs usually vary among different user organizations (ref. 1). Generally, the desired use of remote sensing data is either identification or quantification of surface sediments and pollutants. This study is concerned with data analysis procedures for quantification of water quality parameters that have already been identified and are known to exist within the water body. Specifically, the study deals with the linear multiple-regression technique as a procedure for defining and calibrating data analysis algorithms for such instruments as spectrometers and multispectral scanners. The technique has been used by a number of authors (refs. 2-8) with apparent success. Unfortunately, results have not been completely satisfactory in that (1) analyses of subsets of data from the same experiment sometime give different correlations and algorithms (compare refs. 2, 3, and 4); (2) repeat experiments over the same water body do not always allow quantification of the same water quality parameters (compare refs. 7 and 8); (3) optimum results are not always obtained by a multiple-regression equation with radiance as the independent variable (refs. 8 and 9); and (4) final mapped results do not always agree with results of other types of analysis (compare refs. 4 and 10). From these facts, it is clear that a more complete understanding of the limitations, requirements, and precision of the linear multiple-regression technique is required before it can be applied by user agencies in an operational manner.

In an effort to provide improved scientific understanding, an analytical and laboratory analysis of the linear multiple-regression technique has been conducted (ref. 11). That analysis demonstrated that the technique is fundamentally sound, and, in principle, should apply to many environmental situations in which both the water and atmosphere contain linear and nonlinear optical effects. A number of limitations (optical, mathematical, and operational) were defined, however. One problem with the reference 11 analysis is that field results were not considered in that study. This investigation extends the reference 11 results to include actual field data. An analytical investigation of the signal response equations is conducted, and results from "mixed-brew" laboratory experiments are presented. A study of measurement errors and ground-truth operations from several past field experiments is presented, and recommendations are made concerning future investigations from both aircraft and satellite instruments.

2. SIGNAL-RESPONSE EQUATIONS

From reference 12, reflectance at sea level and upwelled radiance at altitude are related as follows:

$$L_z(\lambda) = T_a(\lambda) [\rho_u(\lambda) (L_{so}(\lambda)) + L_{rd}(\lambda) + L_{rs}(\lambda)] + L_a(\lambda) \quad (1)$$

where:

$L_z(\lambda)$ = apparent upwelled radiance at altitude z at wavelength λ .

$T_a(\lambda)$ = atmospheric transmission at wavelength λ .

$\rho_u(\lambda)$ = inherent upwelled reflectance slightly above water surface at wavelength λ .

$L_{so}(\lambda)$ = upwelled radiance slightly above water surface from 100 percent diffuse reflector at wavelength λ .

$L_{rd}(\lambda)$ = upwelled radiance from specular reflection of diffuse skylight at wavelength λ .

$L_{rs}(\lambda)$ = upwelled radiance from specular reflection of sunlight (sun-glitter) at wavelength λ .

$L_a(\lambda)$ = upwelled radiance from light scattered by the atmosphere (path radiance) at wavelength λ .

For a remote sensing scene with a specific solar elevation angle, the inherent component of $L_z(\lambda)$ being contributed by the water column is $T_a(\lambda)[L_u(\lambda)]$ as follows:

$$L_u(\lambda) = \rho_u(\lambda)(L_{so}(\lambda)) \quad (2)$$

$$T_a(\lambda) [L_u(\lambda)] = T_a(\lambda) [\rho_u(\lambda)(L_{so}(\lambda))] \quad (3)$$

where:

$L_u(\lambda)$ = inherent upwelled radiance slightly above water surface at wavelength λ .

A signal response model of the water may be assumed in which the remote sensing signal is expressed as the signal from the background water plus the signal change caused by some pollutant. The change caused by the pollutant may be expressed as a gradient constant times the pollutant concentration. The inherent signal component for a simplified three-constituent water mixture may be assumed as follows:

$$T_a[L_u(\lambda)] = A + BP_A + EP_B + SP_C^Q \quad (4)$$

where:

A = inherent upwelled radiance component from background water including loss due to atmospheric transmission.

BP_A = inherent upwelled radiance component caused by pollutant A.

EP_B = inherent upwelled radiance component caused by pollutant B.

SP_C = inherent upwelled radiance component caused by pollutant C (assumed to vary nonlinearly as power Q).

B, E, S = gradient constants including losses due to atmospheric transmission.

P_A, P_B, P_C = concentrations of pollutants A, B, and C, respectively.

Upwelled radiance components from surface reflection and atmospheric path radiance effects can be expressed as a value for the baseline atmosphere of a particular day plus a change caused by a variation in atmospheric pollutants over the scene as follows:

$$T_a(\lambda)[L_{rd}(\lambda) + L_{rs}(\lambda)] + L_a(\lambda) = I + LX_A^N \quad (5)$$

where:

I = surface reflection and path radiance components for the baseline atmosphere over the scene.

LX_A^N = change in surface reflection and path radiance components caused by atmospheric pollution (assumed to vary nonlinearly as power N).

L = gradient constant.

X_A = concentration of atmospheric pollutant X_A .

In most remote sensing experiments, upwelled radiance is not measured at a specific wavelength, but instead an integrated average of radiance is measured over a range of wavelengths. To differentiate between wavelength specific and integrated-average values, the term Rad will be used to denote integrated-average values for apparent upwelled radiance at altitude Z over a range or band of wavelengths. Equation (1) may be rewritten as:

$$\text{Rad} = A + BP_A + EP_B + SP_C^Q + I + LX_A^N \quad (6)$$

Equation (6) is identical to equation (4-18) in reference 11. If measurements are made with a remote sensing instrument with bands W,X,Y, and Z, then the equations for the measured values are:

$$\left. \begin{aligned} \text{Rad}_W &= A_W + B_W P_A + E_W P_B + S_W P_C^Q + I_W + L_W X_A^N \\ \text{Rad}_X &= A_X + B_X P_A + E_X P_B + S_X P_C^Q + I_X + L_X X_A^N \\ \text{Rad}_Y &= A_Y + B_Y P_A + E_Y P_B + S_Y P_C^Q + I_Y + L_Y X_A^N \\ \text{Rad}_Z &= A_Z + B_Z P_A + E_Z P_B + S_Z P_C^Q + I_Z + L_Z X_A^N \end{aligned} \right\} \quad (7)$$

where subscripts W,X,Y and Z denote values over the same wavelength ranges as bands W,X,Y, and Z.

Equation (7) is a series of four simultaneous equations which are linear in four unknowns (P_A, P_B, P_C^Q , and X_A^N). If the mathematical operations described in Appendix C of reference 11 are followed, equation (7) can be solved to produce the following solutions for the values of concentration for pollutants A and B.

$$P_A = J + K_W (\text{Rad}_W) + K_X (\text{Rad}_X) + K_Y (\text{Rad}_Y) + K_Z (\text{Rad}_Z) \quad (8)$$

$$P_B = J' + K'_W (\text{Rad}_W) + K'_X (\text{Rad}_X) + K'_Y (\text{Rad}_Y) + K'_Z (\text{Rad}_Z) \quad (9)$$

The J, J' constants are a function of the A, B, E, S, I, and L constants of the various bands, and the K, K' constants are a function of the B, E, S, and L constants. A key element in arriving at the above solution is that the degree of nonlinearity in the P_C and X_A contributions must be essentially constant over the wavelength range of bands W, X, Y, and Z or the changes must be small so that linear approximations are appropriate.

Equations (8) and (9) represent an exact solution for two constituents which have linear radiance gradients in a remote sensing scene which

(1) contains a water mixture with three constituents (one of which has a nonlinear radiance gradient), and (2) has nonlinear variations in surface reflection and path radiance contributions due to variation in some atmospheric constituent over the scene. While this model is somewhat simple, additional complications can be selectively added. Additional water pollutants (beyond pollutant C) and surface-atmospheric contribution terms can be added to equation (6) to complicate the model. As long as the degree of nonlinearity is constant over the wavelength range of interest, a series of linear algebraic equations will result and the exact solution for P_A and P_B will be of the same form as equations (8) and (9) except that additional bands of radiance values will be required. A key element of equations (8) and (9) is that values of P_A and P_B can be calculated without knowledge of each other or the values of P_C or X_A . Knowledge of the A, B, E, S, I, and L constants of equation (7) is required, however, to compute values for the J, J', K, and K' constants before Rad values can be used to compute P_A and P_B .

In actual field experiments, values for the constants A, B, E, S, I, and L are seldom known. In fact, the number of water and atmospheric parameters which have a significant influence on the total upwelled radiance signal is usually unknown. As a result of this situation, empirical methods must be used to determine values for the J, J', K, and K' constants in equations (8) and (9). Fortunately, equations (8) and (9) are of the same form as the statistician's traditional multiple-regression equation:

$$P = J + \sum_{i=1}^{p-1} K_i \text{Rad}_i + \epsilon \quad (10)$$

where.

P = dependent variable.

Rad_i = independent variable.

J, K_i = coefficients obtained by least squares fitting techniques.

* ε = error.

p = total number of J, K coefficients.

In order to evaluate J, K coefficients by empirical methods, certain environmental restrictions must apply. All data used in the fitting process must be from a situation with identical J, K coefficients. This means that values for the A, B, E, S, I, and L constants must be equal for each data point. This condition is often satisfied if one has a single remote sensing scene with multiple ground-truth points, and atmospheric transmission values do not vary by a large amount over the scene. (Surface reflections and path radiance are allowed to vary.) Values for the J, K coefficients generally will not be equal for each data point if one has the situation of multiple remote sensing scenes with a data point in each. In that case, each scene has a different solar elevation angle and atmospheric condition which causes different values for $T_a(\lambda)$, $L_{so}(\lambda)$, $L_{rd}(\lambda)$, $L_{rs}(\lambda)$, and $L_a(\lambda)$ in equation (1). This means that the A, B, E, S, I, and L constants in equations (6) and (7) will be different for each data point. Since A, B, E, S, I, and L determine values for the J, K coefficients, the J, K coefficients will be different for each data point. In such a case, it is impossible to obtain an accurate estimate of the J, K coefficients if radiance and ground-truth concentration values used in the least-squares multiple-regression process are from different

remote sensing scenes. This partly explains why linearized multiple regression with radiance as the independent variable failed to give a good data-reduction algorithm in reference 9. In that experiment, the operational technique was multiple overpasses with a single ground-truth point in each scene.

3. LEAST-SQUARES AND STATISTICAL METHODS

Estimation of J, K Coefficients

The regression task is to estimate the J, K coefficients in which Rad_1 is assumed as the independent variable. In many observations, the independent variables are correlated with each other as well as with the dependent variable which makes results difficult to interpret (ref. 13, p. 398). For the remote sensing situation, high correlations between the independent variables (Rad_1) should be expected if the pollutant of interest has a broad spectral signal over the wavelength range.

As noted previously, least-squares procedures are used to estimate the J, K coefficients using a number of ground-truth points where radiance-constituent concentration data pairs are available. Details of the least-squares procedure are summarized in reference 11. In performing the process on remote sensing data, three major assumptions are involved (ref. 14). They are:

1. The correct form of the equation has been chosen (Rad_1 is linear with concentration for all bands involved).
2. The data are representative of the whole range of environmental combinations in the remote sensing scene.
3. The observations of the dependent variables (ground-truth concentration values) are uncorrelated and statistically independent.

Three minor assumptions are

1. All observations of the dependent variable (concentration) have the same (but unknown) variance, σ^2 .
2. The distribution of uncontrolled error is normal.
3. All independent variables (Rad_1 values) are known without error.

One problem is that measurements of the independent variable (Rad_1) do contain errors. Reference 14 indicates that errors in the independent variable cause estimates of the J, K coefficients to be biased. As a rule-of-thumb, it is recommended that remote sensing experiments be designed such that the variance of radiance about mean values for the ground truth locations ($\sigma_{Rad_1}^2$) be at least 10 times the variance of data noise ($\sigma_{N_1}^2$). This rule-of-thumb may be referred to as Daniel's Criteria (see ref. 11). In terms of standard deviation values:

$$\sigma_{Rad_1}^2 \geq 10.0 \sigma_{N_1}^2 \quad (11)$$

or

$$\sigma_{Rad_1} \geq 3.16 \sigma_{N_1} \quad (12)$$

Equation (12) states that ground-truth locations should be selected within the remote sensing scene in such a manner that the standard deviation of the change

in upwelled radiance for the ground-truth points should be at least 3.16 times the standard deviation of data noise for the particular remote sensing instrument being used if least-squares procedures are to be used in the analysis of data.

Measures of Precision

Before an experiment, it is not known how many bands (or in what combination) will be required to separate the desired water-quality parameter from the total mix of factors which contribute to apparent upwelled radiance. The usual process is one of first calculating a regression equation for the best single band of radiance data and then successively defining multiple-regression equations for the best two bands, three bands, etc. As additional bands are utilized, a number of statistical parameters may be used as indicators of the precision of each new multiple-regression equation. A number of factors must be considered when viewing these parameters to select an optimum multiple-regression equation.

One popular statistical parameter used as a measure of precision is the correlation coefficient, r . The proportion of total variation that is not explained by the regression equation is $1 - r^2$. (An r value equal 0.9 means that 19 percent of the signal variation is not explained by the multiple-regression equation.) Reference 15 states that r is not a good measure of precision as the number of estimated coefficients approach the number of experimental observations. This implies that either the number of ground-truth observations should exceed the number of instrument bands by a wide margin or the number of bands included in the multiple-regression equation should be limited.

The standard deviation, σ , is a second measure of precision of the least-squares process for estimating J, K coefficients. The standard deviation is assumed to represent a value within which 68 percent of all errors are expected to fall if (1) there is an infinite number of observations, and (2) there is minimal error in the independent variables (Rad_1). Unfortunately, most remote sensing experiments have only a limited number of ground-truth observations and Rad_1 measurements do contain errors.

The F-test is a third method of evaluating the adequacy of the least-squares process. The calculated value of F (see eq. (5-14) in ref. 11) must be greater than a critical value (F_{cr}) taken from F-distribution tables in order to be judged significant. If the multiple-regression equation is to be used for predictive purposes, calculated F should be at least 4 times the tabular critical F value (ref. 15). One problem with this parameter is that a level of confidence must be arbitrarily selected before F/F_{cr} can be calculated.

Reference 14 recommends the statistic, C_p , as a measure of the sum of the squared biases plus the sum of the squared random errors (see eq. (5-17) in ref. 11). Given a multiple-regression equation with p estimated J, K coefficients, a low value of C_p in combination with a C_p/p ratio ≤ 1.0 is considered to indicate a good fit with negligible bias. The regression equation should then be useful for predictive purposes. The C_p/p ratio is the only one of the above parameters which is indicative of bias in the fitted equation (ref. 14)

For future remote sensing experiments, it is recommended that multiple-regression equations be computed for all combinations of bands for which upwelled radiance values are available. Values for $r, \sigma, F/F_{cr}, C_p$, and C_p/p should be computed for each equation. The prime basis for selecting an "optimum" regression equation should be $C_p/p \leq 1.0$ for minimum bias and $F/F_{cr} \geq 4.0$ for predictive utilization., Values for r should approach 1 and σ

should approach zero. Daniel's Criteria (eq. 12)) should also be satisfied for all bands in the regression equation. When all five conditions are met, the multiple regression equation with the minimum number of bands should be selected for calculation of pollutant concentration.

4. LABORATORY VALIDATION EXPERIMENTS

It is desirable to validate data analysis techniques with actual remote sensing data under controlled conditions. To achieve this result, tests were conducted with filtered-deionized tapwater in the Marine Upwelled Spectral Signature Laboratory at the NASA Langley Research Center. A sketch of the laboratory setup is shown in figure 1. (More complete descriptions of the system and test procedures can be found in references 11 and 16.) Only partial results from one series of tests will be presented in this paper for reasons of brevity. Additional laboratory test cases are presented in reference 11.

It was desirable to test the multiple-regression technique with data from water mixtures which contained constituents with both linear and nonlinear radiance gradients. Single-constituent tests were conducted on a number of materials. From these data, it was concluded that both Ball Clay and Feldspar soils have near linear radiance gradients for concentrations between 4 and 173 ppm. Rhodamine WT dye has a nonlinear gradient for concentrations between 17 and 1052 ppb. With this knowledge, a series of three-constituent tests were conducted with 25 different water mixtures. Table I shows the concentrations of Ball Clay, Feldspar, and Rhodamine WT that were present in the filtered-deionized tapwater for each test. Also shown are radiance values for the following 5 wavelength bands:

Band Number	Wavelength Range (nm)	Center Wavelength (nm)
1	340-500	420
2	460-620	540
3	540-700	620
4	620-780	700
5	700-860	780

The radiance values shown are in terms of relative units obtained by dividing power/bandwidth measurements over the water by gray-card diffuse reflectance measurements of the input light source.

A multiple-regression analysis was performed for Ball Clay using 12 of the 25 tests in table I to simulate "ground-truth" values. The 12 "ground-truth" values were tests 1, 3, 5, 6, 8, 10, 13, 15, 18, 20, 21, and 23. The standard deviation of the change in upwelled radiance was next calculated for each band for this "ground-truth" data set. Values are compared with data noise from the laboratory measurements as follows:

Band Number	σ_{Rad}	σ_N (Ref. 11)	$\sigma_{\text{Rad}}/\sigma_N$
1	0.1372	0.0343	4.00
2	0.1125	0.0343	3.23
3	0.1414	0.0343	4.12
4	0.1431	0.0343	4.17
5	0.0992	0.0343	2.89

These figures indicate that Daniel's Criteria (eq. (12)) is satisfied for all bands except number 5. Since the signal-to-noise standard deviation ratio

for band 5 is only slightly below the value of 3.16, it was decided not to exclude that band from the regression analysis in this particular case. Regression equations for all combinations of bands were next calculated for Ball Clay concentration. Estimated values of the J, K coefficients and various statistical measures of precision are shown in table II. Review of table II indicates that the lowest value of total squared error (C_p) is obtained for band combinations 2, 3, and 4. The regression equation for this combination is:

$$P_A = -4.1 + 234.4 (\text{Rad}_2) - 613.7 (\text{Rad}_3) + 918.0 (\text{Rad}_4) \quad (13)$$

Statistical estimates of precision for this equation are:

$$r = 0.99$$

$$\sigma = 6.8 \text{ ppm}$$

$$(F/F_{cr})_{.95} = 105.7$$

$$C_p = 3.0$$

$$C_p/p = 0.8$$

Since values for r , σ , F/F_{cr} , and C_p/p are acceptable, it is assumed that equation (13) has good predictive capability. To test this assumption, the radiance values for bands 2, 3, and 4 from table I were applied to equation (13). Values for Ball Clay concentration (P_A) were calculated and are compared with actual values in figure 2. Shaded symbols denote the 12 "ground-truth" points used in the least-squares fit, and the open symbols represent calculations for the remaining 13 independent test points. (Some open symbols are hidden under the shaded points.) Since all points fall within $\pm 3.9\sigma$ of the true value, it is concluded that equation (13) has good predictive capability and that linear multiple-regression analysis procedures have the potential for quantification of constituents with linear radiance gradients in water mixtures which also contain nonlinear constituents. A satisfactory linear regression equation was also found for the concentration of Feldspar soil in these same mixtures (ref. 11).

5. REVIEW OF FIELD DATA

Analytical analysis and laboratory test cases have been used to perform a limited validation of linearized multiple-regression analysis for quantification of marine constituents. No matter how many controlled tests are conducted, final validation of the technique must come through use of field experiments. As noted previously, results from past experiments have not been completely satisfactory. This section examines the data and operational conditions of past tests in an attempt to define problem areas which require increased attention for future experiments.

Instrument Noise

Unfortunately, remote sensing data always contain error and random noise. Most operational instruments contain onboard calibration lamps and black bodies to minimize radiance error. Noise has been somewhat more difficult to eliminate because it is caused by many components in the instrument system. After an experiment, instrument noise can usually be evaluated by calculating the standard deviation of calibration lamp and/or black body count values and applying the appropriate calibration constants to convert to radiance. If onboard calibration sources are not available, flat or near-constant radiance portions of the remote sensing scene may be examined. Both of these procedures have been used to examine noise in the data for the field experiments described

in references 4, 5, 6, and 9. Figure 3 presents estimates of noise standard deviation values for each band for which data were obtained. These data indicate that a wide range of noise values have been observed in previous experiments.

The problem of data noise has been widely recognized, and pixel averaging is often employed as a method of reducing the error. To illustrate the effects of this process, calibration lamp values from the reference 4 data were averaged in a manner to simulate pixel averaging. New values for noise standard deviation were calculated and are compared with single-pixel values in figure 4. Large reductions in instrument noise may be achieved by averaging even small size (4 by 4 or less) pixel arrays. Use of array sizes larger than 7 by 7 produced only small reductions in noise for the reference 4 data set.

Pixel averaging effectively increases the size of the ground resolution element that is observed by the remote sensing instrument. If the array size becomes too large, average radiance values of the enlarged pixels may not be representative of ground-truth values because of spatial variability in pollutant concentration. As pixels are averaged, noise error will be reduced but new error is introduced because of scene dynamics or hydraulic changes. Carried to extreme, hydraulic features that depict smaller-scale pollutant transport may be erased by the averaging process. If averaging is necessary, then the following criteria should be used to establish credibility of the enlarged pixel radiance values.

$$\left[\sigma_{\text{Rad}_{\text{pt}}} \right]_{\text{nxn}} \approx \left[\sigma_{\text{N}} \right]_{\text{nxn}} \quad (14)$$

where

$\left[\sigma_{\text{Rad}_{\text{pt}}} \right]_{\text{nxn}}$ = standard deviation of enlarged pixel about average radiance value (calculated from individual pixel radiances within n by n array surrounding each ground truth point).

When $\left[\sigma_{\text{Rad}_{\text{pt}}} \right]_{\text{nxn}}$ is much larger than $\left[\sigma_{\text{N}} \right]_{\text{nxn}}$, one has an indication that scene dynamics have introduced large errors into the process that offset the noise reduction benefits gained by pixel averaging.

Ground Truth Placement

Previous discussion has indicated that ground-truth points should be located within the remote sensing scene in such a manner that Daniel's Criteria (eq. 12) are satisfied for all bands in the regression equation. To review this aspect of previous experiments, values of σ_{Rad} have been calculated using radiance values over each set of ground-truth points for the field experiments of references 4, 5, 6, and 9. Pixel averaging had been employed to eliminate signal changes caused by the ground-truth boat, time differences and location uncertainty as well as to reduce noise in each data set. New estimates of σ_{N} have been made from calibration lamp or background water data to simulate the pixel averaging process for each experiment. Values of $\sigma_{\text{Rad}}/\sigma_{\text{N}}$ have been computed and are compared with Daniel's Criteria in figure 5. It appears that ground-truth placement was adequate such that Daniel's Criteria were satisfied after pixel averaging for most experiments. Approximate calculations indicate that not a single one of these experiments would satisfy Daniel's Criteria on a 1 by 1 pixel array basis, however. Ground-truth locations for future remote sensing experiments should be selected in such a manner that σ_{Rad} has the largest possible value to minimize the requirement for pixel averaging. Use of past remote sensing images of the area to estimate relative radiance differences between proposed locations is probably the best method of selecting a distribution with maximum standard deviation.

Physical Consistency

An unwritten assumption in the multiple-regression process is that all data used in the correlation are "good" without physical deficiencies. When remotely monitoring water-quality parameters, this generally means that a number of criteria should be satisfied concerning both environmental conditions and physical operations of the experiment. Only limited discussion justifying physical consistency of the data is available in references 4 through 9. For that reason, the remainder of this review will discuss best engineering practice with the experience from past experiments described for only those items discussed in the literature.

Relative to environmental criteria, two desirable conditions are:

1. Water depth greater than the Secchi depth.
2. Constant vertical concentration gradient within the remote sensing penetration depth (the depth above which 90 percent of the upwelled radiance originates, ref. 17).

It is sometimes assumed that the maximum remote sensing penetration depth is the same as the Secchi depth. Calculations in reference 18 as well as unpublished data from the James River in Virginia indicate that maximum remote sensing penetration depth may be on the order of 20 to 50 percent of Secchi depth, depending on absorption and scattering characteristics of the mixture. Knowledge of the maximum remote sensing penetration depth is required so that water samples are not obtained below the zone that is causing the remotely-sensed signal. One method to estimate the maximum remote sensing penetration depth is that of lowering a flat black plate with less than 1 percent diffuse reflectance into the water and noting its depth of disappearance (see figs. 6 and 7 in ref. 18).

In terms of operational criteria, a major problem is the time lapse between overpass of the remote sensor and collection of individual water samples. Time lapse can cause the ground-truth data to be hydraulically inappropriate if there are significant wind or tidal influences on the water body being observed. It may be possible to correct ground-truth data to account for small time lapses (ref. 19), but such procedures have not yet been widely demonstrated in field experiments. In principal, all data should be synchronous with the remote sensor overpass, but that has not been achieved in past experiments, as shown below

Reference	Geographical Area	Number of Ground Truth Points in Regression Analysis	Maximum Time Lapse Between Water Sampling and Remote Sensor Overpass
4	James River in Virginia	21	2.0 hrs
5	New Jersey Coast	22	3.0 hrs
6	New York Bight	10	0.3 hr
7	Saginaw Bay in Michigan	27	8.0 hrs
8	Saginaw Bay in Michigan	16	8.0 hrs
9	Delaware Shelf	7	0.5 hr

Some experiments have a small time lapse, but in other cases it has been assumed that constituents in a water pixel remain constant for 2 to 8 hours. Considering that algae tend to migrate depthwise with changing light intensity (solar elevation angle), large time lapses should be justified with quantitative water sample data even when wind and tidal effects are such that flow conditions are stable.

In the operational area, another major problem is often encountered. Large-scale experiments with multiple vessels often require either some water sample analysis onboard ship or laboratory analysis of the same parameter by different organizations. Consistency of laboratory results between different laboratories is a longstanding problem, and variations for certain parameters may be larger than that of the remote sensing scene. It is recommended that future experiments maintain consistent handling and laboratory analysis procedures.

6. SUMMARY OF RECOMMENDED CONDITIONS FOR TECHNIQUE APPLICATION

An investigation of the linear multiple-regression technique with remotely sensed radiance as the independent variable has been conducted. Signal response equations have been analyzed, and results from "mixed brew" laboratory tests are presented. Results from these studies indicate that the technique is fundamentally sound and should apply in many environmental situations in which both the water and atmosphere contain linear and nonlinear optical effects. Conditions which limit application of the technique have also been discussed. A review of previous field experiments has served to emphasize additional limitations which must be considered in future experiments. A summary of recommended conditions for use of the technique is given in figure 6. From this listing, it is clear that the linearized multiple-regression analysis should never be applied blindly to a set of data without background knowledge concerning the constituent of interest, hydraulics of the water body, typical Secchi and maximum remote sensing penetration depth values, and measurement uncertainties from various sources. The technique has strong theoretical foundation and careful application should yield useful results. It is particularly appropriate for use in small regions to validate either mathematical or hydraulic models of pollutant transport and diffusion. Present economics for obtaining multiple ground-truth points within a single remote sensing scene limit its usefulness for many routine monitoring missions, however.

7. CONCLUDING REMARKS

The study described herein is part of a continuing effort to define data-reduction techniques and their appropriate application so that increased benefits can be derived from both aircraft and satellite remote sensing data. The goal of the present study was a more complete understanding of limitations, requirements, and precision of the linear multiple-regression technique with radiance as the independent variable. Environmental and optical physics conditions have been defined for which an exact solution to the signal response equations is of the same mathematical form as the statistician's traditional multiple-regression equation. In such a case, use of linearized multiple regression is merely an empirical correlation to obtain coefficients for the exact solution to the signal-response equations. Additional analytical investigations are desirable to more completely define atmospheric limitations and to consider the problem of bottom reflection in optically shallow waters.

One problem with use of the regression technique is that the independent variables (upwelled radiances) contain errors and are often correlated with each other. This requires consideration of a number of statistical parameters when performing the regression analysis.

Review of past field experiments indicates that data noise was of such magnitude that data smoothing was required before Daniel's Criteria could be satisfied for the least-squares multiple-regression process. Improved selection of ground-truth locations to maximize variance is recommended to minimize data smoothing requirements and the physical errors associated with that process.

Time lapse between remote sensor overpass and water sample collection appears to have been a problem in past experiments. Economic consideration will always result in only a limited number of water samples obtained synchronously with the remote sensor overpass. Additional studies to develop and demonstrate techniques for correcting nonsynchronous data for remote sensing use are desirable.

ACKNOWLEDGMENT

The authors are indebted to Mr. Stuart R. LeCroy of the Vought Corporation for computer programming support and the handling of computer compatible tapes for extraction of data noise statistics.

REFERENCES

1. Kuo, Chin Y., and Cheng, Robert Y. K.. Laboratory Requirements for In-Situ and Remote Sensing of Suspended Material. NASA CR-145263, 1976.
2. Johnson, Robert W Quantitative Suspended Sediment Mapping Using Aircraft Remotely Sensed Multispectral Data. NASA Earth Resources Survey Symposium, vol. I-C Land Use-Marine Resources, NASA TM X-58168, 1975, pp. 2087-2098.
3. Johnson, Robert W Application of Aircraft Multispectral Scanners to Quantitative Analysis and Mapping of Water Quality Parameters in the James River, Virginia. COSPAR Space Research, vol. XVII, M. J. Rycroft and A. C. Strickland, eds., Pergamon Press, Inc., 1977, pp. 25-31.
4. Johnson, Robert W., and Bain, Gilbert S. Quantitative Analysis of Aircraft Multispectral-Scanner Data and Mapping of Water-Quality Parameters in the James River in Virginia. NASA TP-1021, 1977.
5. Johnson, Robert W. Mapping the Hudson River Plume and an Acid Waste Plume by Remote Sensing in the New York Bight Apex. Results from the National Aeronautics and Space Administration Remote Sensing Experiments in the New York Bight - April 7-17, 1975. J. B. Hall and A. O. Pearson, eds., NASA TM X-74032, 1977.
6. Johnson, R. W., Duedall, I. W., Glasgow, R. M.; Proni, J. R.; and Nelsen, T. A.. Quantitative Mapping of Suspended Solids in Wastewater Sludge Plumes in the New York Bight Apex. J. Water Pollut. Control Fed., vol. 49, no. 10, Oct. 1977, pp. 2063-2073.
7. Rogers, R. H., Shah, N. J., McKeon, J. B.; Wilson, C., Reed, L.; Smith, V. Elliott, and Thomas, Nelson A.: Application of Landsat to the Surveillance and Control of Eutrophication in Saginaw Bay. Proceedings of the Tenth International Symposium on Remote Sensing of Environment, vol. I, Environ. Res. Inst. of Michigan, Oct. 1975, pp. 437-446.
8. Rogers, R. H.; Shah, N. J., McKeon, J. B.; and Smith, V. E.: Computer Mapping of Water Quality in Saginaw Bay with LANDSAT Digital Data. Proceedings of the ACSM-ASP Convention, Washington, D.C., February 1976.
9. Ohlhorst, C. W. Quantitative Mapping by Remote Sensing of an Ocean Acid-Waste Dump. NASA-TP 1275, 1978.
10. Bressette, Walter E. Aerial Photographic Water Color Variations from Pollution in the James River. Proceedings of the ACSM-ASP 1978 Fall Convention, Albuquerque, New Mexico, October 1978.
11. Whitlock, Charles H.: Fundamental Analysis of the Linear Multiple Regression Technique for Quantification of Water Quality Parameters from Remote Sensing Data. NASA TM X-74600, 1977.
12. Miller, J. R.; Jain, S. C.; O'Neill, N. T.; McNeil, W. R.; and Thomson, K.P.B.: Interpretation of Airborne Spectral Reflectance Measurements Over Georgian Bay. Remote Sensing of Environment. vol. 6, no. 3, 1977, pp. 183-200.

- 13 Snedecor, G. W., and Cochran, W. G.: Statistical Methods. Sixth ed., The Iowa State University Press, 1967
14. Daniel, C.; and Wood, F. S. Fitting Equations to Data. Wiley-Interscience, 1971.
15. Draper, N. R., and Smith, H. Applied Regression Analysis. John Wiley and Sons, 1966.
16. Whitlock, Charles H., Usry, J. W.; Witte, William G.; and Gurganus, E. A.: Laboratory Measurements of Upwelled Radiance and Reflectance Spectra of Calvert, Ball, Jordan, and Feldspar Soil Sediments. NASA TP-1039, 1977.
17. Gordon, H. R.; and McCluney, W. R.. Estimation of the Depth of Sunlight Penetration in the Sea for Remote Sensing. vol. 14, no. 2, February 1975, pp. 413-416.
18. McCluney, W. R.. Estimation of Sunlight Penetration in the Sea for Remote Sensing. NASA TMX-70643, 1974.
19. Kuo, C. Y.; and Blair, C. H. Rational Method for Correcting Sea Truth of Suspended Sediment Concentration Related to Remote Sensing. Old Dominion University Technical Report 76-C2, 1976.

Test Number	Soft Clay (ppm)	Feldspar (ppm)	Rhodesmine WT (ppb)	Rad ₁	Rad ₂	Rad ₃	Rad ₄	Rad ₅
1	9	17	34	0.094	0.090	0.094	0.053	0.028
2	9	17	1052	0.039	0.093	0.140	0.078	0.022
3	9	35	34	0.105	0.100	0.097	0.054	0.033
4	17	35	34	0.121	0.127	0.120	0.080	0.047
5	9	35	1052	0.044	0.100	0.143	0.078	0.029
6	17	35	1052	0.054	0.101	0.154	0.099	0.039
7	17	35	190	0.116	0.134	0.135	0.091	0.045
8	17	32	190	0.138	0.148	0.147	0.101	0.052
9	17	32	535	0.099	0.148	0.190	0.110	0.050
10	32	32	535	0.117	0.190	0.252	0.170	0.090
11	17	32	1052	0.059	0.105	0.145	0.102	0.044
12	32	32	1052	0.113	0.133	0.212	0.154	0.080
13	173	129	34	0.426	0.428	0.437	0.370	0.260
14	173	129	190	0.388	0.347	0.429	0.380	0.254
15	32	129	535	0.232	0.230	0.304	0.218	0.128
16	129	129	535	0.359	0.323	0.439	0.382	0.220
17	173	129	535	0.359	0.300	0.440	0.407	0.254
18	32	129	1052	0.147	0.195	0.245	0.188	0.095
19	129	129	1052	0.246	0.213	0.355	0.313	0.188
20	173	129	1052	0.299	0.263	0.444	0.426	0.254
21	173	173	34	0.459	0.470	0.475	0.395	0.267
22	173	173	190	0.432	0.379	0.475	0.404	0.267
23	173	173	535	0.387	0.322	0.444	0.413	0.267
24	129	173	1052	0.260	0.225	0.376	0.334	0.197
25	173	173	1052	0.298	0.250	0.428	0.388	0.244

Table I. - Concentration and radiance data for three-constituent tests.

Bands Used	J	K ₁	K ₂	K ₃	K ₄	K ₅	r	σ	FW _{cr} ¹	C _p	C _b
1	-27.1	485.74	0	0	0	0	0.95	25.1	17.1	114	58.2
2	-36.1	0	313.33	0	0	0	0.96	32.0	9.7	194	91.2
3	-54.6	0	0	484.88	0	0	0.97	17.7	36.26	54	27.1
4	-29.6	0	0	491.89	0	0	0.99	12.7	22.6	24	12.0
5	-15.5	0	0	0	0	704.15	0.99	8.4	169.5	4	2.9
1,2	-22.3	633.02	-164.23	0	0	0	0.95	25.9	9.4	114	57.9
1,3	-52.0	114.74	0	377.49	0	0	0.98	17.8	21.2	50	14.7
1,4	-31.5	168.00	0	0	394.35	0	0.99	11.8	49.1	19	6.3
1,5	-12.3	-79.96	0	0	0	814.89	0.99	8.1	104.7	4	1.9
2,3	-34.4	0	54.73	438.98	0	0	0.97	18.4	19.6	54	14.1
2,4	-34.5	0	98.08	0	483.48	0	0.99	11.7	33.4	18	6.7
2,5	-13.2	0	-34.87	0	0	748.85	0.99	8.0	93.0	7	2.4
3,4	-22.3	0	0	-123.09	648.93	0	0.99	13.1	34.6	25	8.2
3,5	-2.3	0	0	-144.38	0	913.15	0.99	8.0	109.6	5	1.8
4,5	-15.7	0	0	0	5.18	708.79	0.99	8.8	88.8	8	2.8
1,2,3	-48.6	234.28	-183.15	371.27	0	0	0.98	18.5	13.6	50	12.6
1,2,4	-31.8	33.68	71.86	0	485.54	0	0.99	12.3	31.5	20	5.0
1,2,5	-13.8	-147.41	62.61	0	0	833.39	0.99	8.4	66.6	7	1.8
1,3,4	-4.4	186.46	0	-388.60	708.97	0	0.99	15.8	48.2	12	3.0
1,3,5	1.5	-83.90	0	-150.11	0	1033.46	0.99	7.5	85.5	5	1.2
1,4,5	-5	-172.28	0	0	-281.25	1308.85	0.99	7.8	88.9	5	1.4
2,3,4	-4.1	0	234.43	-613.65	914.09	0	0.99	6.8	103.7	3	0.8
2,3,5	-1.1	0	-74.48	-138.14	0	935.17	0.99	8.3	70.3	7	2.7
2,4,5	-6.1	0	-60.72	0	-104.93	924.76	0.99	8.0	91.7	9	2.2 ¹
3,4,5	-4	0	0	0	-227.38	180.35	0.99	7.8	73.2	4	1.4
1,2,1,4	-4.3	-39.46	285.79	-434.14	954.08	0	0.99	7.1	72.4	5	1.8
1,2,1,5	2.0	-880.93	287.82	-143.38	0	1111.39	0.99	7.3	66.4	5	1.1
1,2,4,5	-2.2	-815.07	44.73	0	-845.23	1242.72	0.99	8.2	54.2	7	1.4
1,3,4,5	1.9	-105.77	0	-122.21	-63.78	1113.88	0.99	8.0	58.7	7	1.4
2,3,4,5	-1.0	0	193.82	-352.38	785.25	147.37	0.99	7.2	68.7	5	1.0
1,2,1,4,5	-4	-131.38	211.89	-488.68	832.89	364.38	0.99	7.1	58.5	4	1.8

Table II. - Estimates of J, K coefficients and measures of precision for Ball Clay sediment in three-constituent mixtures.

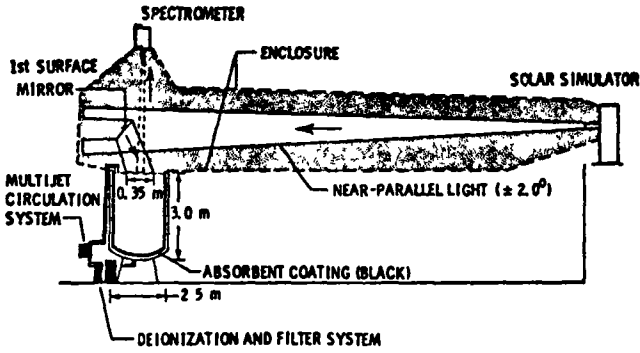


Figure 1. - Laboratory setup.

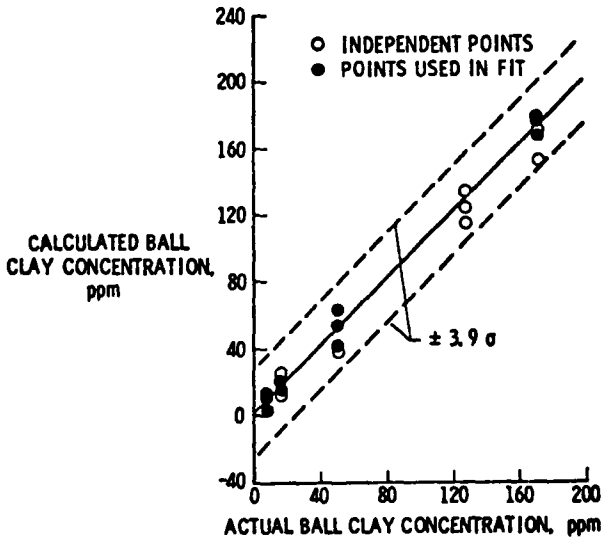


Figure 2. - Comparison of calculated and actual Ball Clay concentrations for three-constituent mixture tests.

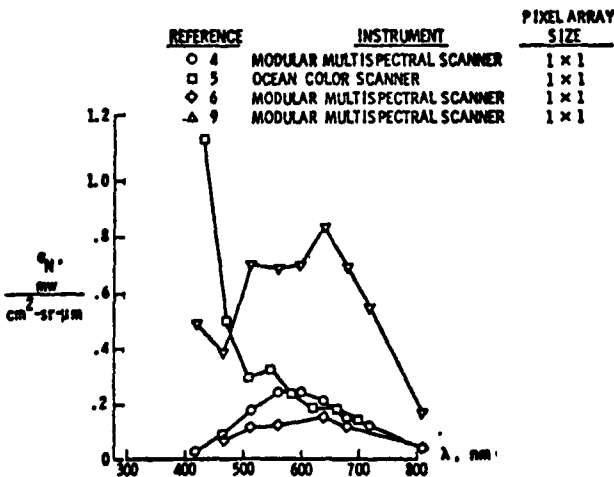


Figure 3. - Estimated standard deviation of data noise for several field experiments.

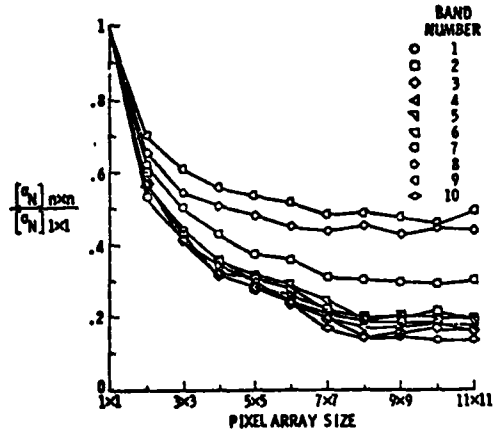


Figure 4. - Effects of pixel averaging for reference 4 data.

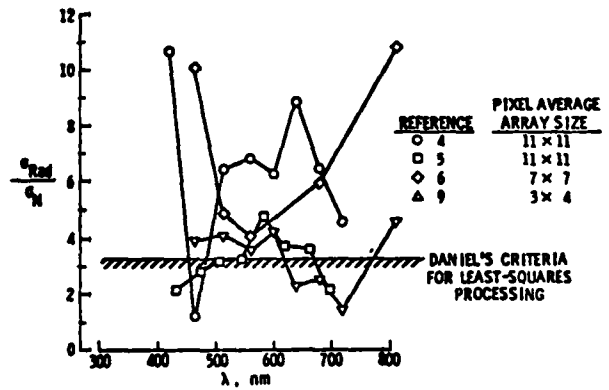


Figure 5. - Signal to noise standard deviation ratio for several field experiments.

ENVIRONMENTAL	<ul style="list-style-type: none"> Linear radiance gradient for constituent of interest. Different radiance spectra from other constituents. Degree of nonlinearity of other contributors constant over wavelength. Small changes in atmospheric transmission over remote sensing scene. Water depth greater than remote sensing penetration depth. Near-constant vertical concentration gradient within remote sensing penetration depth.
GROUND TRUTH	<ul style="list-style-type: none"> Single remote sensing scene with boundaries that give maximum change in upwelled radiance. Multiple water sample locations with near-uniform radiance distribution. Multiple sample points in all plumes and water masses in scene. Hydraulically appropriate sampling sequence. Total number of water sample locations much greater than number of remote sensor bands in regression equation. Constant sample depth less than remote sensing penetration depth. Consistent handling and laboratory analysis of water samples.
DATA	<ul style="list-style-type: none"> $\sigma_{Rad} / \sigma_N \geq 118$ $C_p / p \leq 1.0$ $FF_{CR} \geq 4.0$ $\sigma = 0$ $r = 1.0$ Minimum pixel averaging with $(\sigma_{Rad})_{AVG} = (\sigma_N)_{REF}$

Figure 6. - Recommended conditions for use of linear multiple-regression technique with radiance as the independent variable.

End of Document