United States
Environmental Protection
Agency

Environmental Monitoring
Systems Laboratory
P.O. Box 15027
Las Vegas, NV 89114

Research and Development

EPA LANDSAT MSS Classification Of Near-Shore Water Quality In Lake Ontario

August 18,1981
EPA Region V

TS-AMD-82095
April 1983

LANDSAT MSS CLASSIFICATION OF NEAR-SHORE WATER QUALITY IN LAKE ONTARIO

August 18, 1981

by

T. H. Mace
Environmental Programs
Lockheed Engineering and Management Services Company, Inc.
Las Vegas, Nevada 89114

Contract No. 68-03-3049

Project Officer

G. A. Shelton
Advanced Monitoring Systems Division
Environmental Monitoring Systems Laboratory
Las Vegas, Nevada 89114

ENVIRONMENTAL MONITORING SYSTEMS LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
LAS VEGAS, NEVADA 89114

NOTICE

This document is currently undergoing a peer and administrative review within EPA. It is a draft document and is subject to revision. Therefore, it should not be cited nor reproduced for further distribution.

ABSTRACT

At the request of the *U.S. Environmental Protection Agency's Region V, Great* Lakes Program Office, the Environmental Monitoring Systems Laboratory at Las Vegas, Nevada, performed computer-assisted classifications of Landsat multispectral scanner data covering near-shore portions of Lake Ontario. A best subsets multiple linear regression procedure was used to relate samples of the Landsat data with concurrent water quality samples collected on July 30 and August 18, 1981, from stations located near the Rochester, Niagara, and Oswego Harbors.

Relationships were developed between the satellite data and Secchi depth, turbidity, conductivity, dissolved oxygen, alkalinity, NH_3 + NH_4^+-nitrogen, NO_2^- + NO_3^--nitrogen, total phosphorous, Chlorophyll a and Phaephyton a. Classification maps of the near-shore areas around Rochester and Niagara were produced for NO_2^- + NO_3^--nitrogen, total phosphorous, and conductivity.

CONTENTS

FIGURES

TABLES

INTRODUCTION

The Landsat satellite system has been continually recording image data in four spectral bands since July 1972. Flying in a sun-synchronous polar orbit, Landsats 1, 2, and 3 have imaged most of the earth every 18 days with an instantaneous field-of-view of approximately 79 by 79 m. Each continuous path (from north to south) is divided into scenes of approximately 185 by 185 km consisting of scan lines every 79 m and elements every 56 m. (Figures 1 and 2 illustrate the image form of the digital data from Band 4 of the two scenes used in this study.) Table 1 lists the multispectral scanner (MSS) bands and wavelength ranges for Landsats 1, 2, and 3. The bands are labeled 4-8, as a 3-channel return beam videcon system occupied bands 1-3.

TABLE 1. LANDSAT 1, 2, AND 3 MSS BANDS

Band	Wavelengths (µm)	Spectral Region
4	0.5- 0.6	Green
5	0.6- 0.7	Red
6	0.7- 0.8	Near Infrared
7	0.8- 1.1	Near Infrared
8*	8.0-14.0	Thermal Infrared

*Landsat 3 only.

Landsat 4, launched in July of 1982, is currently orbiting on a 16-day repeat cycle. This satellite possesses the conventional MSS Bands 4, 5, 6, and 7 (re-named 1, 2, 3, and 4) as well as the experimental 7-band Thematic Mapper scanner with 30 by 30 m resolution.

Previous work on remote sensing of lake eutrophication (summarized by Witzig and Whitehurst, 1980) has indicated that the Landsat MSS system may be used to extend simultaneous water measurements over large regional areas. Additionally, work

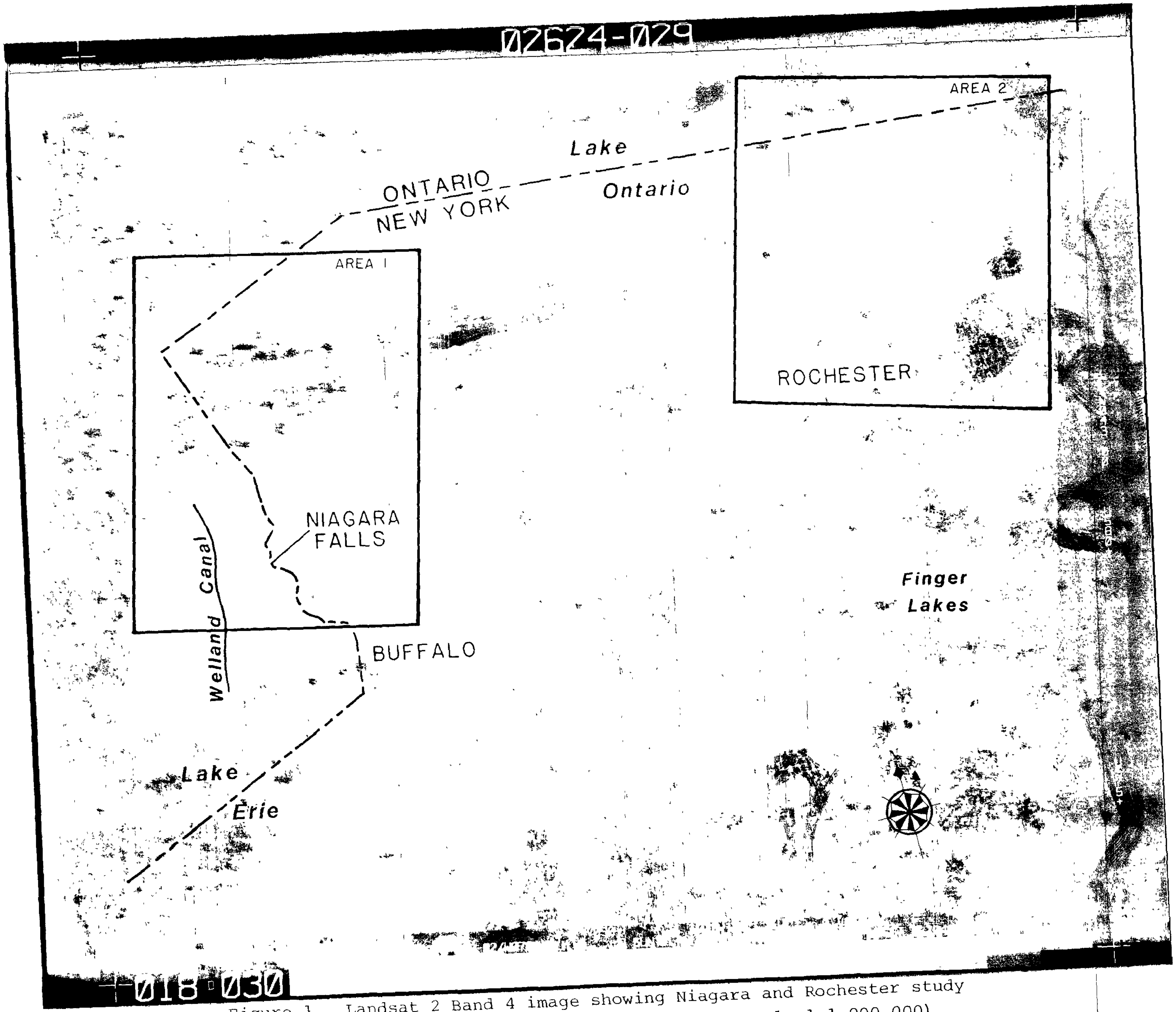

Figure 1. Landsat 2 Band 4 image showing Niagara and Rochester study areas - August 18, 1981 (approximate scale 1:1,000,000).

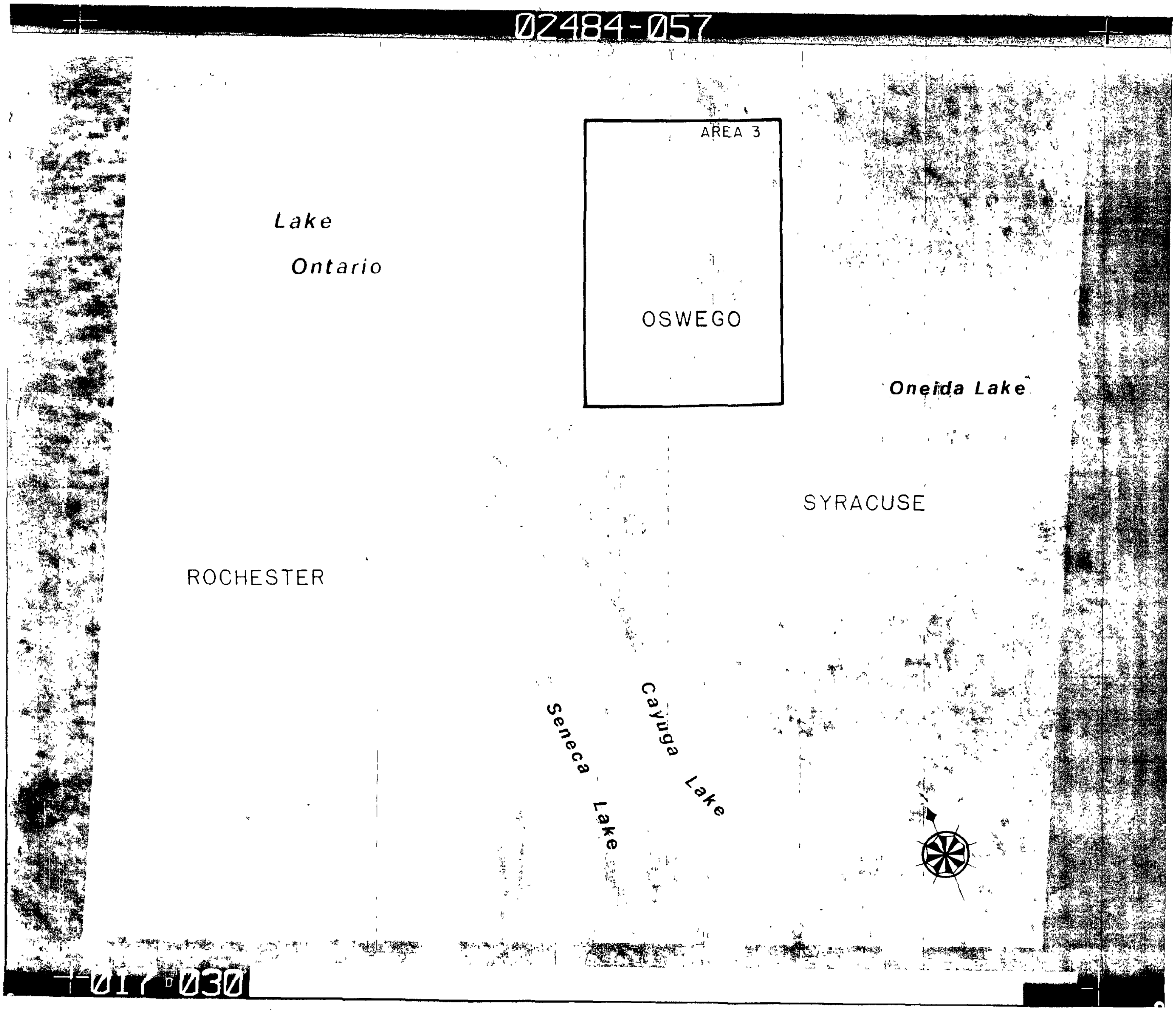

Figure 2. Landsat 2 Band 4 image showing Oswego study area - July 30, 1981 (approximate scale (1:1,000,000).

performed at the U.S. Environmental Protection Agency's (EPA) Environmental Monitoring Systems Laboratory at Las Vegas (EMSL-LV) has indicated that airborne multispectral scanner data may be used within the framework of a multiple linear regression procedure to quantitatively map a variety of water quality-related parameters (Mace, 1982 and Mace, 1983). Therefore, at the request of EPA's Region V, Great Lakes Program Office, the EMSL-LV analyzed portions of two Landsat images using water quality data from the STORET data base system. The object of this study was to use archival Landsat MSS data to quantitatively extend the existing near-shore cruise data to the portions of Lake Ontario near Niagara, Rochester, and Oswego Harbors (Figures 1 and 2) using a regression-based technique.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

Based on the findings of this report and review of the appropriate literature it is concluded that:

- The Landsat satellite multispectral scanner data may be used to map a variety of water quality parameters from point sample data, provided that the sampling stations are distributed such that positional errors do not have relatively large effects.
- Same-day data sets are required for large lakes, due to the dynamic nature of these large water bodies.
- Existing water sample data may be combined with existing Landsat data in a regression-based procedure to extend the point measurements over large areas.
- The regression-based procedure is a robust method and produces quantitative statistical information regarding the validity of its use in extending water quality measurements.

RECOMMENDATIONS

Based on the findings of this study, it is recommended that:

- Landsat black-and-white imagery be used to plan future water sampling programs.

- Where feasible, water sampling should be conducted during satellite overpasses to maximize the archival value of the data.

- Water sampling stations should be evenly distributed with respect to the parameters described as well as with respect to the satellite resolution elements.

- It would be of value to investigate placing these analyses into a geographic information system, to allow further manipulation for modeling purposes.

- Future near-shore studies should include monitoring stations near the Welland Canal and its associated plumes to more fully describe inputs to the lake.

METHODOLOGY

Landsat 2 uncorrected image tapes covering Niagara and Rochester harbors on August 18, 1981, and Oswego Harbor on July 30, 1981, were acquired from the EROS data center, Sioux Falls, South Dakota. Additionally, geometrically corrected negatives for Band 4 of the same images were obtained for overview purposes (Figures 1 and 2). The reason for obtaining the tapes in uncorrected format was that the geometric corrections may introduce radiometric errors in the multispectral data. These errors are negligible for pictorial representations but may be quite significant to quantitative applications. Therefore, the uncorrected MSS data were used.

The Earth Resources Laboratory Applications Software package (ELAS) was used to display the Band 7 Landsat data on a cathode ray tube and to extract MSS statistics over corresponding water quality sample stations. Mean values in each of the four bands were calculated for the water quality stations sampled on July 30, 1981, at Oswego (Figure 3) and on August 18, 1981, at Rochester and Niagara (Figures 4 and 5). These mean values were then converted to radiance values following the procedures described by Robinove (1982).

Figure 6 illustrates the image processing flow for the nitrate-nitrogen classification. All of the water quality parameters mapped followed this same flow. MSS statistics were extracted from the raw Landsat data and passed to program P9R (Best Subsets Multiple Linear Regression) of the Biomedical Computer Programs - P Series (Dixon, 1977) where they were regressed against water quality variables from the STORET data base system. Mallow's Cp statistic was used to determine the best subset of channels to use as predictors, and the transformation equation yielding the lowest Cp was calculated.

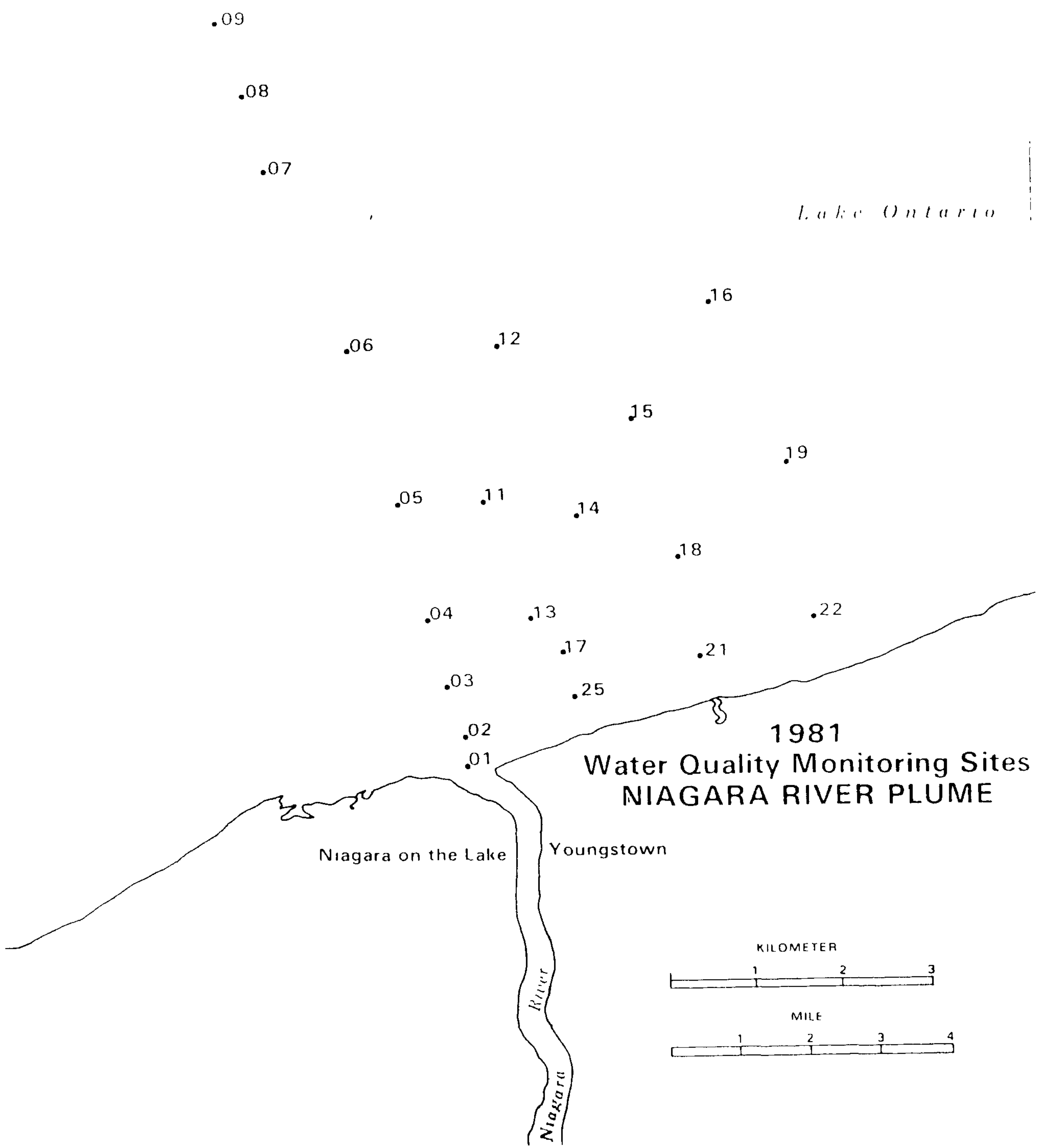

Figure 3. Niagara water quality monitoring stations (Great Lakes Program Office).

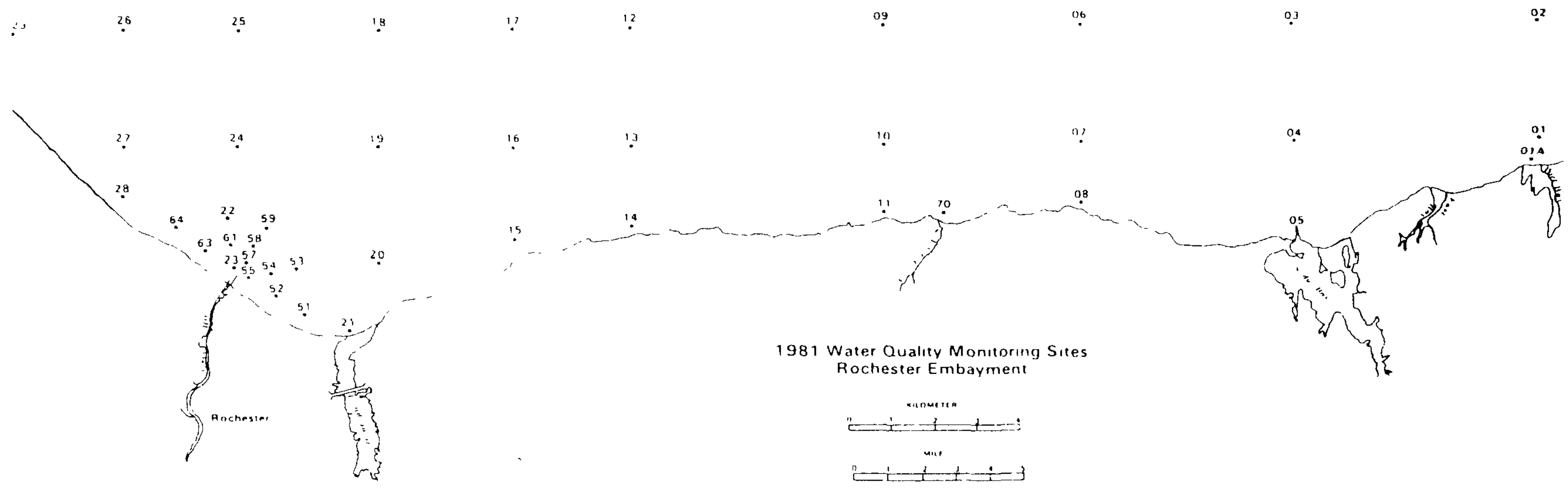

Figure 4. Rochester water quality monitoring stations (Great Lakes Program Office).

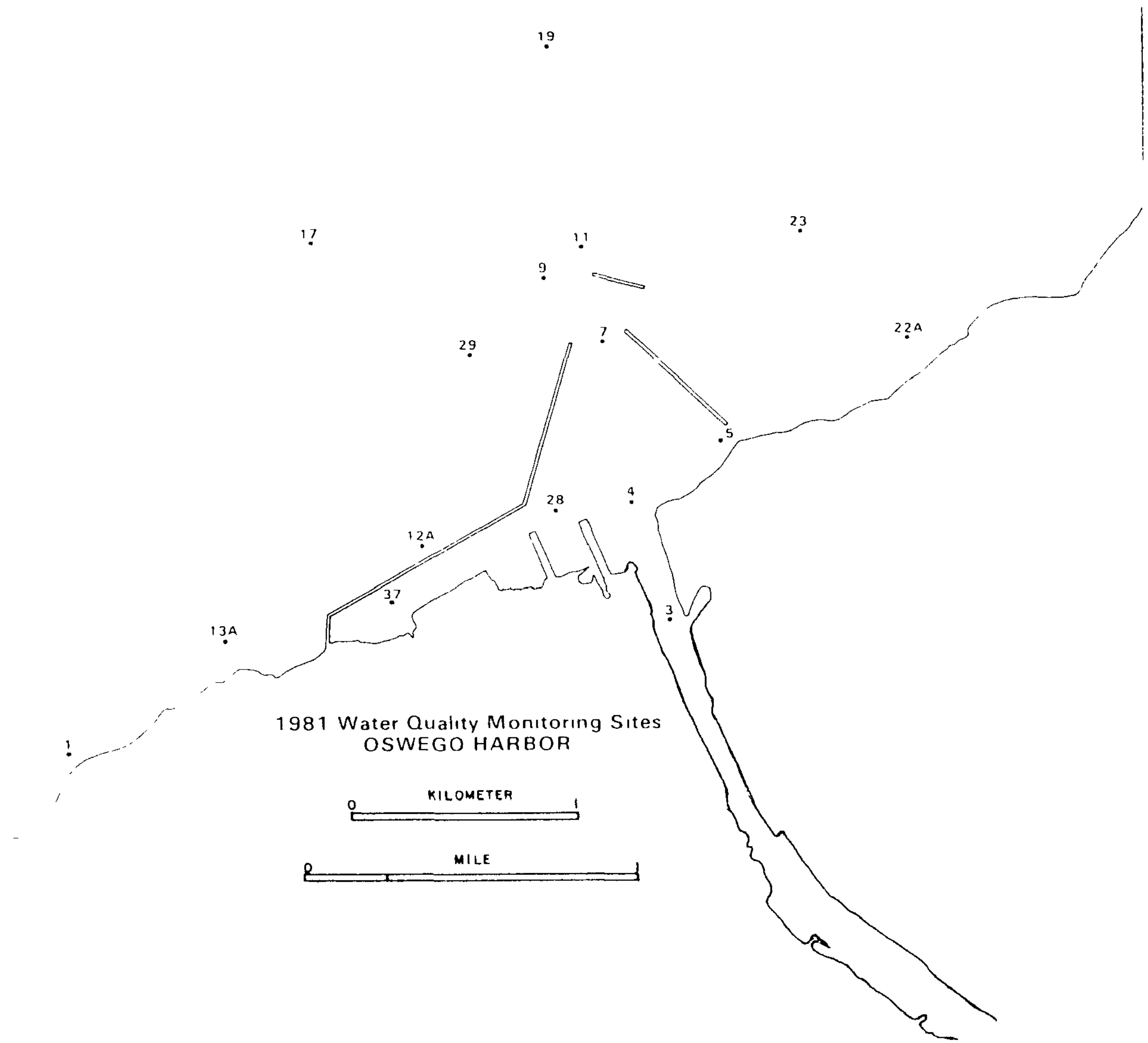

Figure 5. Oswego water quality monitoring stations (Great Lakes Program Office).

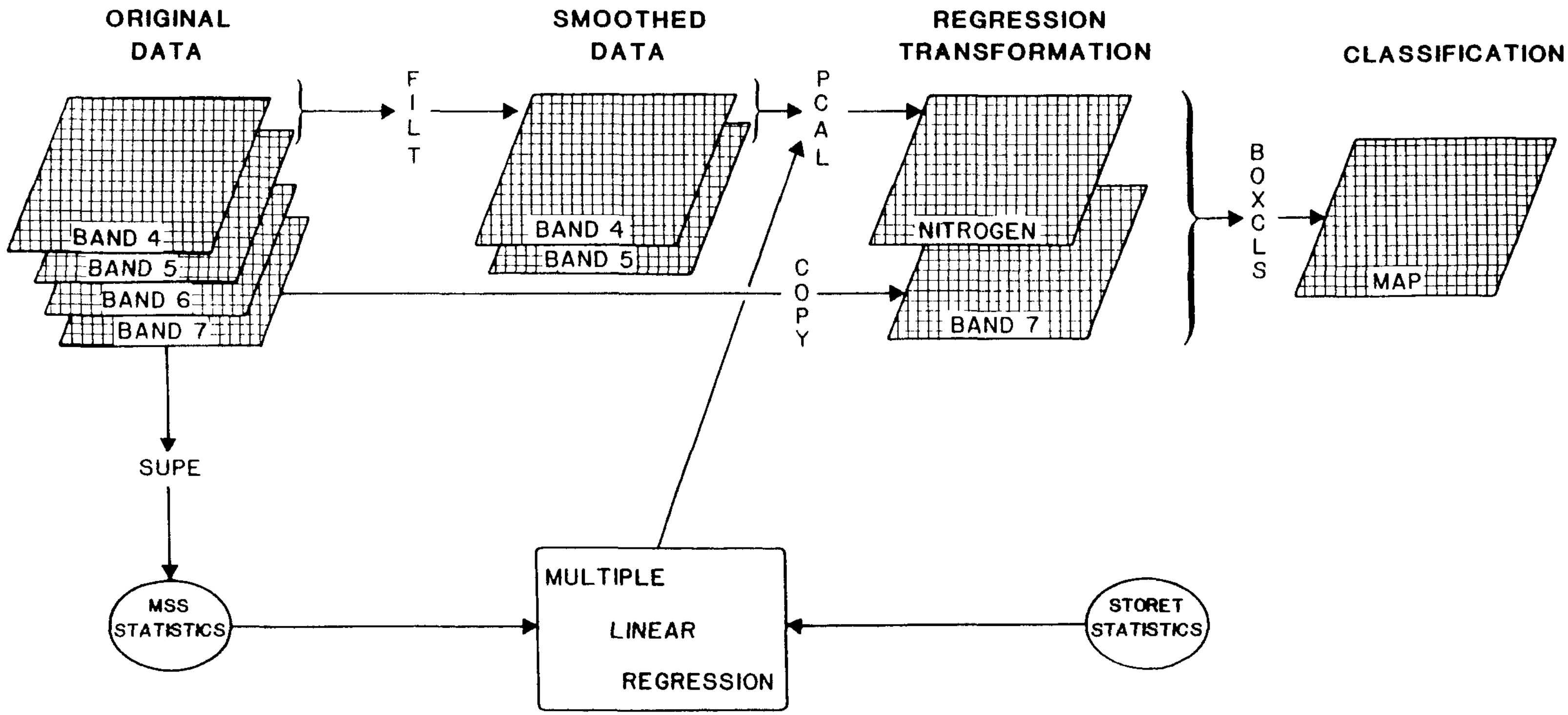

Figure 6. Data processing flow for the NO_2^- + NO_3^--nitrogen classification.

Next, the optimum set of bands were extracted and a spatial filter (FILT) consisting of a 3 by 3-pixel "moving window" was applied to the data. This smoothing operation removed some of the high frequency noise associated with the Landsat MSS system without substantially degrading the imagery.

The programmable calculator overlay (PCAL) was then used to transform the Landsat data into 8-bit integer representations of the mapped water quality parameter. Additionally, the COPY overlay was used to copy Band 7 of the original data set into the transformed file.

A parallelepiped classifier (BOXCLS) was then used to classify the water areas (defined by low values in Band 7) into ranges of the water quality parameter. The resulting classification array (map) was color-coded on a CRT and photographically reproduced using Polaroid film and a Dunn Instruments film recorder.

Image transformation equations were calculated for Secchi depth, turbidity, conductivity, dissolved oxygen, alkalinity, total NH_3 + NH_4^+-nitrogen, total NO_2^- and NO_3^--nitrogen, total phosphorous, Chlorophyll a, and Phaeophyton a. Transformations for NO_2^- + NO_3^--nitrogen, total phosphorous, and conductivity were performed for Area 1 (Niagara) and Area 2 (Rochester) 1024 by 1024-pixel data sets. No transformations of the Area 3 (Oswego) data set were made due to poor correlations between the MSS and STORET statistics.

RESULTS

REGRESSION ANALYSIS

The results of the regression analysis are presented in Tables 2 through 4. Four criteria for evaluating the regression are adapted from the procedures described by Whitlock, Kuo, and Le Croy (1982). These criteria are as follows:

R^2 - the coefficient of determination. This is a number between zero and one which, when multiplied by 100, gives the percent of the total dependent variable described by the multiple linear regression equation. Ideally, its value should approach one. However, values of 0.7 or above may indicate a useful regression for predictive purposes. Its square root, R, is the multiple correlation coefficient.

Cp/P - Mallow's Cp statistic divided by P, the number of predictors. Values of one or below indicate lack of bias in the independent variables. The regression program chooses the best subset of predictors to minimize the value of Cp.

SE - the standard error of the estimate. Sixty-eight percent of the errors should fall within one standard error.

Significance - the probability that the regression equation is nonlinear or has a slope of zero as determined through an F-test. For this study, the maximum allowable significance was set at 0.001 (99.9 percent confidence level).

TABLE 2. REGRESSION SUMMARY, ROCHESTER-NIAGARA DATA SET

Dependent Variable	R^2	Cp/P	SE	Significance (F-Test)	Predictors	Number of Points
*Total Phosphorous	0.974	0.73	0.079	0.0002	Band 5	6
*NO_2^- + NO_3^--Nitrogen	0.936	1.35	0.027	<0.0000	Bands 4 & 5	15
Alkalinity ($CaCO_3$)	0.917	1.33	2.908	<0.0000	Bands 4, 5, & 6	15
Phaeophyton a	0.810	1.06	0.445	0.0068	Bands 5 & 6	9
Secchi Depth	0.486	1.16	1.0	0.0184	Bands 4 & 5	15
Conductivity	0.875	1.10	36.111	<0.0000	Bands 4 & 6	16
Dissolved	0.117	0.67	0.573	0.4526	Band 6	7
Turbidity	0.040	0.34	3.770	0.4733	Band 5	15
NH_3 + NH_4^+-Nitrogen	0.101	0.03	0.010	0.2674	Band 6	14
Chlorophyll a	0.321	0.27	2.677	0.1119	Band 6	9

TABLE 3. EFFECT OF LOG_{10} TRANSFORMATIONS, ROCHESTER-NIAGARA DATA SET

Dependent Variable	R^2	Cp/P	SE	Significance (F-Test)	Predictors	Number of Points
Total Phosphorous	0.945	1.32	2.148	0.0012	Band 4	6
NO_2^- + NO_3^--Nitrogen	0.853	1.07	0.023	<0.0000	Bands 4 & 5	15
Alkalinity ($CaCO_3$)	0.902	1.33	2.797	<0.0000	Bands 4, 5, & 6	15
Phaeophyton a	0.662	0.91	0.524	0.0076	Band 5	9
Secchi Depth	0.538	1.28	1.2	0.0098	Bands 4 & 5	15
*Conductivity	0.863	1.06	27.84	<0.0000	Bands 4 & 6	14
Chlorophyll a	0.345	0.25	2.498	0.0961	Band 6	9

TABLE 4. REGRESSION SUMMARY, OSWEGO DATA SET

Dependent Variable	R^2	Cp/P	SE	Significance (F-Test)	Predictors	Number of Points
Total Phosphorous	0.491	1.01	0.026	0.0173	Bands 4 & 6	15
NO_2^- + NO_3^--Nitrogen	0.124	0.03	0.050	0.1989	Band 6	15
Secchi Depth	0.482	1.23	0.667	0.0192	Bands 4 & 6	15
Conductivity	0.500	1.10	232.89	0.0155	Bands 4 & 6	15
Turbidity	0.405	1.00	1.42	0.0444	Bands 4 & 6	15
NH_3 + NH_4^+-Nitrogen	0.253	0.88	0.0218	0.0560	Band 6	15

*Image-transformation performed.

The STORET summary data are presented in Tables 5 and 6. The minimum and maximum values are particularly of interest, as they describe the range of values over which the regressions were performed. While extrapolations beyond these limits are possible, they should be used with caution, as errors become undefined.

No laboratory error analysis data were provided, but it is assumed that the laboratory errors fell within limits prescribed by standard EPA quality assurance programs.

TABLE 5. STORET DATA SUMMARY, ROCHESTER-NIAGARA DATA SET

Dependent Variable	Mean	Minimum	Maximum	Significance Deviation	Units
Total Phosphorous	0.2018	0.0110	1.1060	0.4430	mg/l
NO_2^- + NO_3^--Nitrogen	0.134	0.080	0.490	0.100	mg/l
Alkalinity ($CaCO_3$)	92.40	88.00	124.00	8.93	mg/l
Phaeophyton a	2.26	1.30	4.30	0.88	μg/l
Secchi Depth	3.13	1.50	5.00	1.31	m
Conductivity	331.7	299.0	679.0	95.2	μmhos/cm
Dissolved	9.34	8.60	10.10	0.56	mg/l
Turbidity	5.25	1.10	10.10	3.71	Hach FTU
NH_3 + NH_4^+-Nitrogen	0.0147	0.0010	0.0320	0.0097	mg/l
Chlorophyll a	4.278	1.900	10.200	3.038	μg/l

TABLE 6. STORET DATA SUMMARY, OSWEGO DATA SET

Dependent Variable	Mean	Minimum	Maximum	Significance Deviation	Units
Total Phosphorous	0.0449	0.0130	0.0930	0.0332	mg/l
NO_2^- + NO_3^--Nitrogen	0.113	0.020	0.200	0.052	mg/l
Secchi Depth	2.17	1.00	3.50	0.86	m
Conductivity	563.2	315.0	1006.0	305.1	μmhos/cm
Turbidity	3.03	1.50	6.20	1.70	Hach FTU
NH_3 + NH_4^+-Nitrogen	0.0238	0.0040	0.0700	0.0244	mg/l

It is often useful to present the results of multiple linear regressions graphically. Figures A1 through A10 provide scatter plots of predicted versus observed values for the regressions presented in Table 2.

IMAGE CLASSIFICATIONS

Figures 7 through 12 show, in image form, the results of image tranformation and classification of total phosphorous, NO_2^-+ NO_3^--nitrogen, and conductivity. Only the Rochester and Niagara areas are presented, as no significant relationships were found for the Oswego data set.

The areas classified are as delineated on Figure 1. Slight differences exist between the geometry of Figure 1 and Figures 7 through 12. This is because Figure 1 was generated from geometrically corrected imagery, and Figures 7 through 12 were generated from uncorrected imagery to maximize radiometric accuracy.

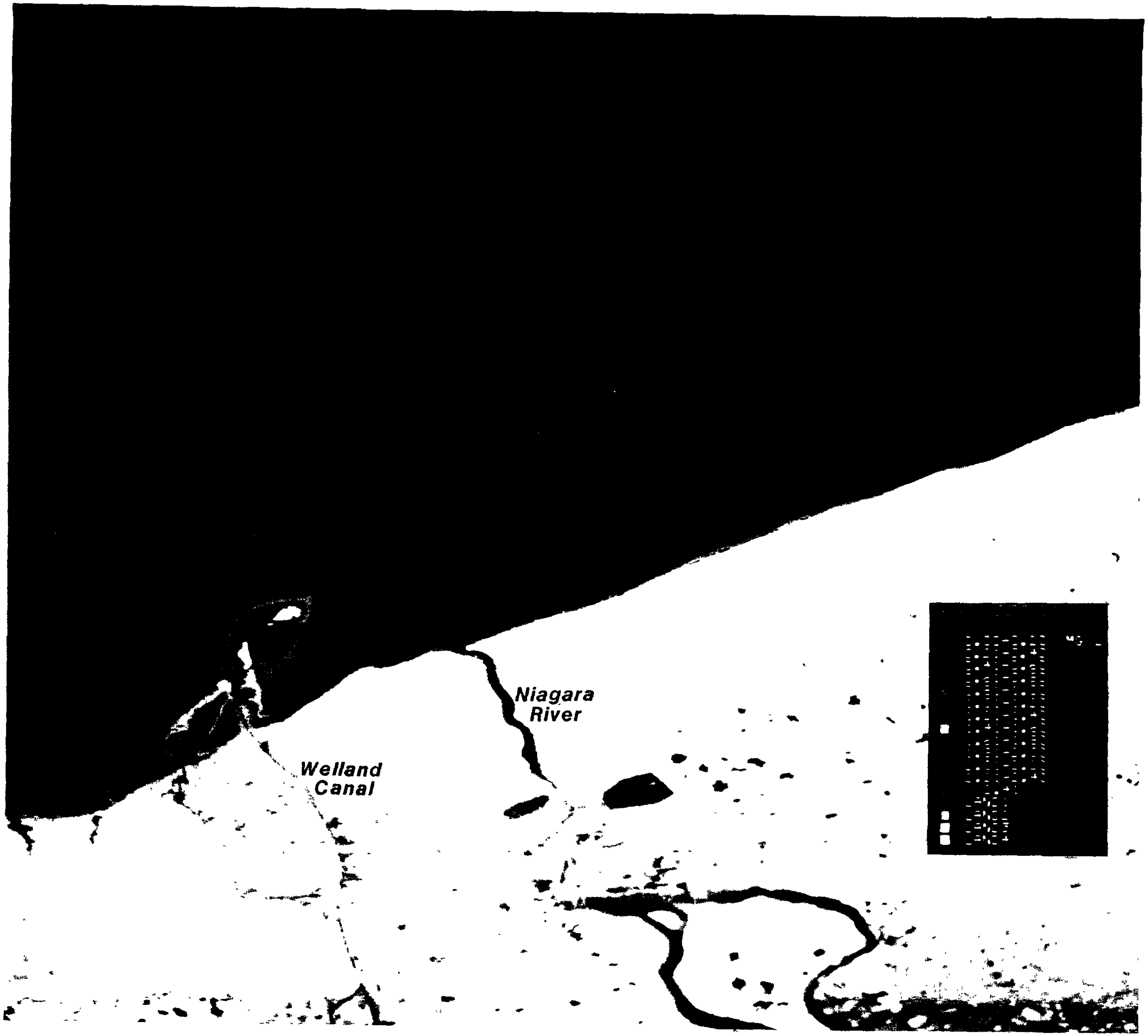

Figure 7. Total phosphorous (mg/l) August 18, 1981 - Area 1, Niagara.

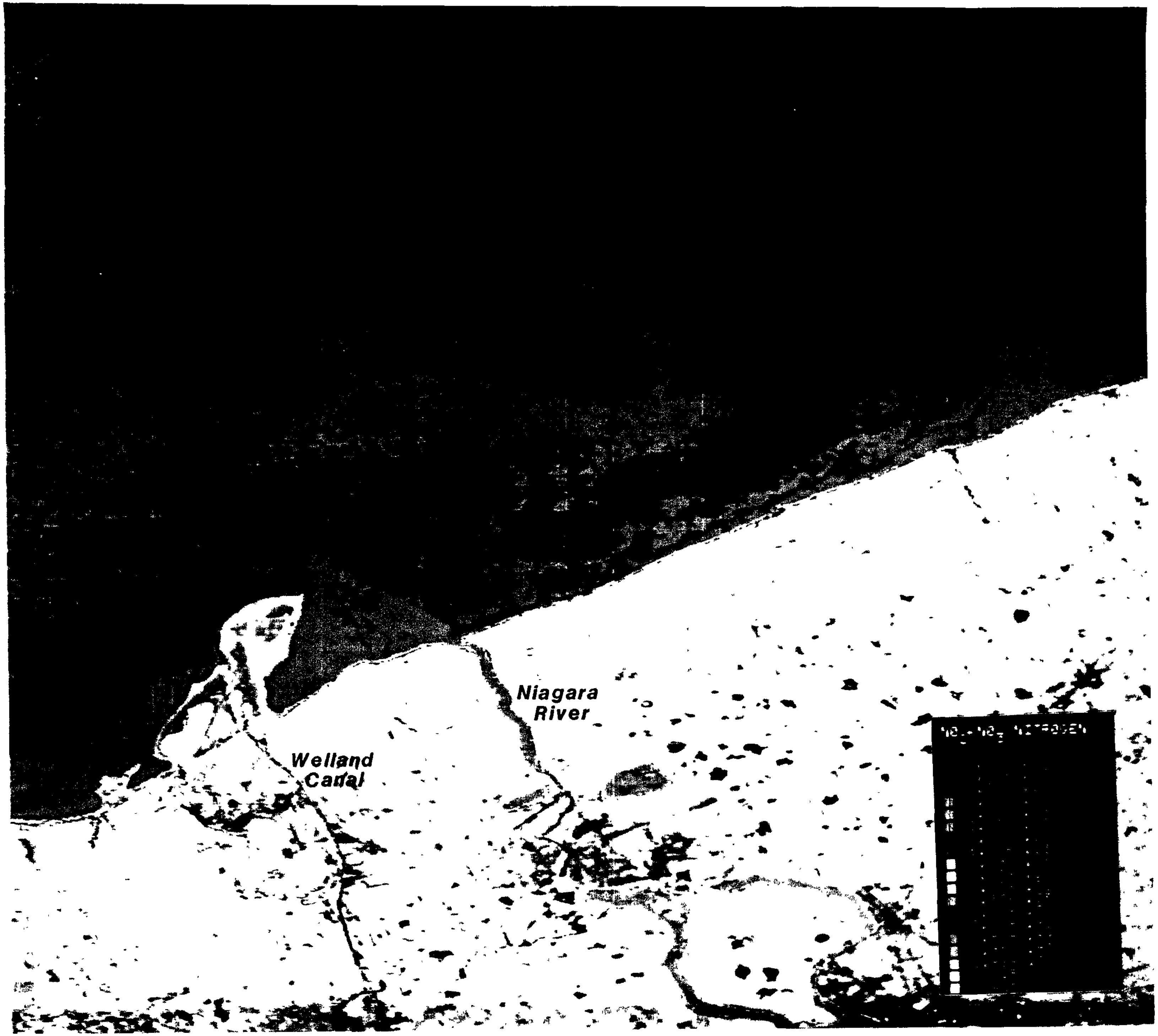

Figure 8. NO_2^- + NO_3^--nitrogen (mg/l) August 18, 1981 - Area 1, Niagara.

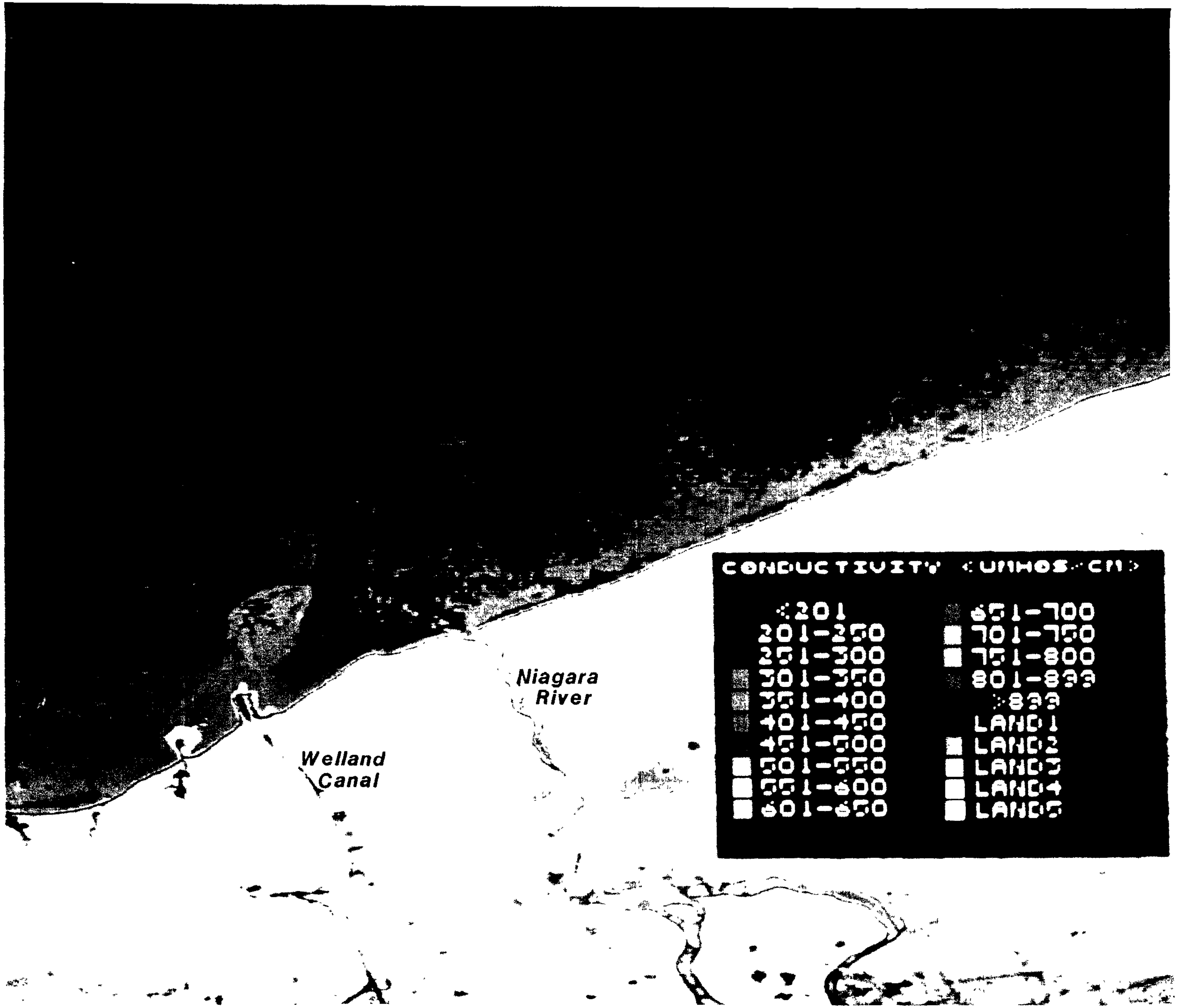

Figure 9. Conductivity (μmohs/cm) August 18, 1981 - Area 1, Niagara.

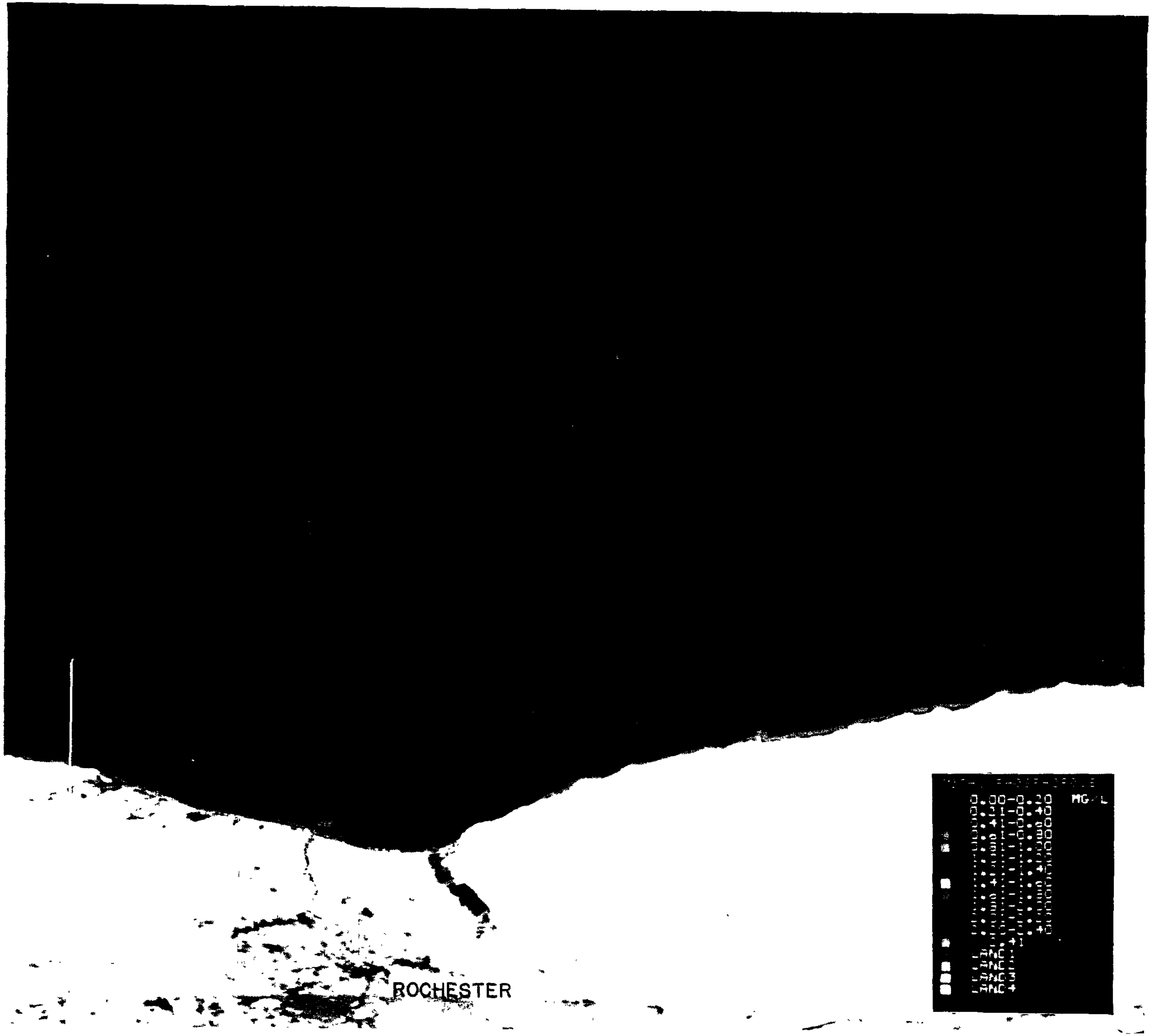

Figure 10. Total phosphorous (mg/l) August 18, 1981 - Area 2, Rochester.

Figure 11. NO_2^- + NO_3^--nitrogen (mg/l) August 18, 1981 - Area 2, Rochester.

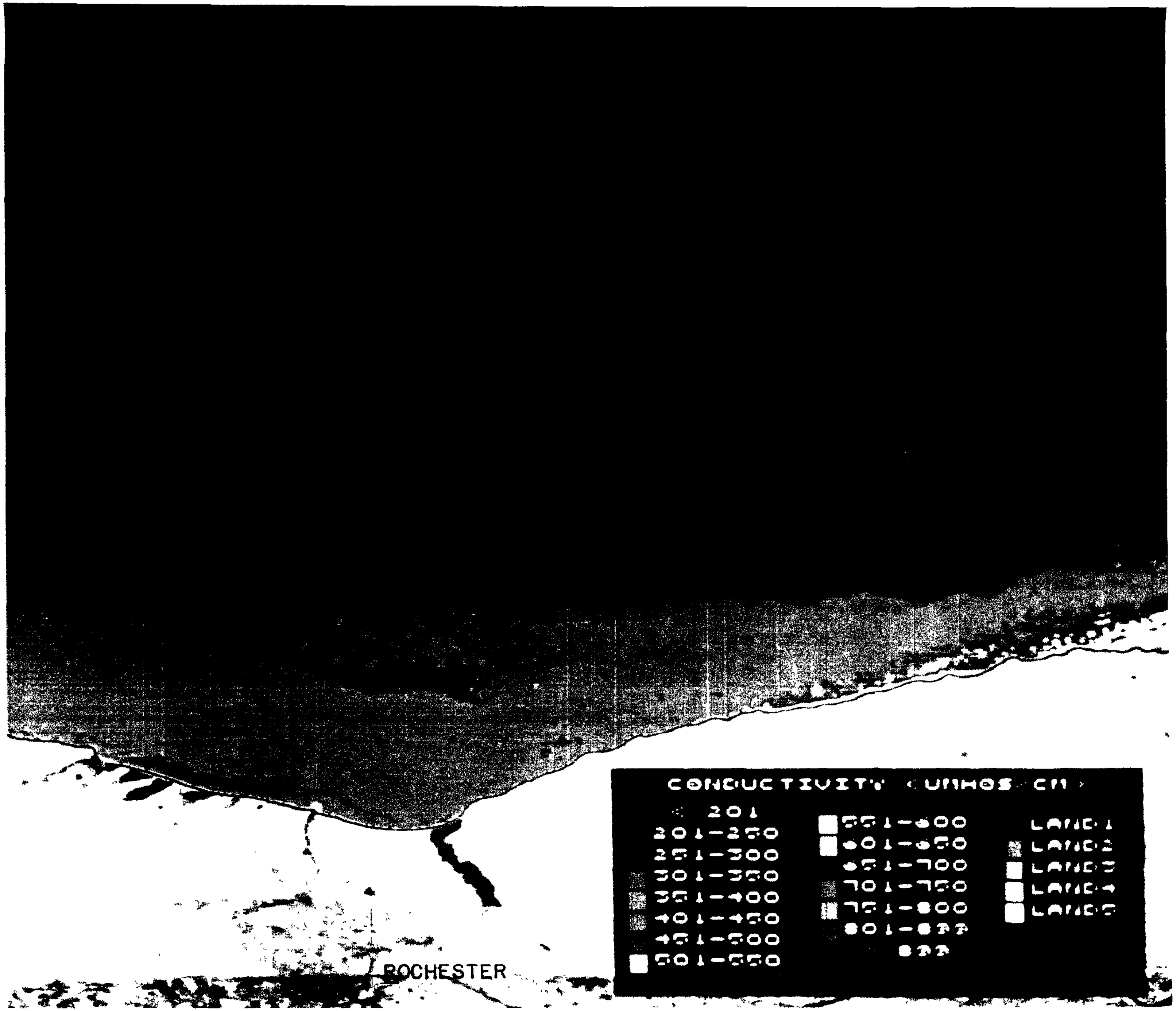

Figure 12. Conductivity (µmhos/cm) August 18, 1981 - Area 2, Rochester.

DISCUSSION

Image Analysis

Visual comparison between the Rochester-Niagara image (Figure 1) and the Oswego image (Figure 2) shows qualitatively the dynamic nature of Lake Ontario. Light-toned areas in the lake correspond to areas of higher concentrations of suspended matter. Differences in the pattern of these light-toned areas from Figure 1 to Figure 2 made it apparent that only same-day water samples would be useful in this analysis. This greatly reduced the number of samples available for use in the regression calculations.

Nevertheless, significant relationships were found between several of the water quality parameters and the Landsat data for the Rochester-Niagara data set. Total phosphorous, nitrate (NO_2^- + NO_3^-)-nitrogen, alkalinity, and conductivity possessed high coefficients of determination, relatively low standard errors, Cp/P ratios approaching unity, and statistical significances better than 0.001. However, Figures A1, A2, A3, and A6 show that the regression equations are heavily dependent upon a single data point. While this condition is less than optimal, experience at EMSL-LV has shown that the data transformations are still operationally usable.

Analyses of the final classifications for the Niagara site show significantly higher inputs to the lake from the Welland Canal. Three plumes are visible along the Canadian portion of the shore. The cause of these plumes is not revealed within this imagery, but the values along the Welland Canal seem to be elevated downstream from Welland, Ontario.

Slight plumes exist along the New York shoreline emanating from the Niagara River and Eighteen Mile Creek at Olcott. Another small plume from the Genesee River can be seen in the Rochester classifications. Most of the other patterns visible in the imagery appear to be the result of shore erosion and lake circulation patterns.

Errors

As with all measurement systems, the regression-based image classifications are subject to error. The best measure of this error is the standard error of the estimate (SE). Most of the interpolated errors will occur within one standard error. However, extrapolated values may exceed the standard error. Therefore, calculated values beyond the ranges specified in Table 5 should be used with caution.

The interpolated errors may be divided into three basic types: (1) errors in the water chemistry measurements (including boat position errors), (2) errors in sampling the Landsat imagery, and (3) radiometric errors inherent in the Landsat MSS system.

Ship position was identified using Loran-C. According to Dull and Clerke (1981), the probable position errors generated using this instrument are 20 to 400 m with possible errors of up to 463 m (0.25 nmi). Even if water chemistry measurement errors are considered to be negligible, improper placement of water sample points could combine with errors in placement of the Landsat samples to introduce substantial variance in the regression equations. This apparently was not a problem with the Rochester-Niagara image, but could have been the reason that the Oswego analysis failed to provide significant results. The Oswego area consisted of sample points that were clustered extremely close together. Any misregistration between water samples and Landsat data samples could have seriously jeopardized the subsequent regression analysis. This is particularly true for areas in which strong gradients are present (such as the mouth of a river).

Less important, but more evident, are the errors generated by the scanning process. Because we are conducting these analyses at the limit of the MSS system capabilities, spatially oriented noise is very apparent in the final classifications. High-pass spatial filtering tends to reduce the effects of noise, but horizontal banding due to detector calibration inconsistencies is still apparent. While visually disturbing, this banding should account for no more than 2 percent of the scene variance and is more a nuisance than a serious source of error.

Future Considerations

While it is apparent that Landsat imagery can be analyzed for a variety of water quality parameters, it is also apparent that much care must be taken in the water sampling phase. For dynamic systems, such as Lake Ontario, it is imperative that same-day data be used. It is also important that the range of observed values be as large as possible and that their distribution is uniform throughout that range.

Bands 4 and 5 black-and-white images of the Landsat data are excellent planning tools and are quite cost-effective. Even if remote sensing extensions of the water quality data are not planned, the eventual value of water quality data collected with possible remote environmental monitoring in mind is greatly increased. It is strongly recommended that, whenever possible, EPA water sampling programs should be planned using latest available satellite imagery. The imagery could be used to plan sampling locations to minimize the number of points necessary to characterize lake water quality and to assure that their distribution matches both the variability inherent in the water and the resolution of the remote sensing systems. Then, should the need arise, the United States' land satellite system could be used to augment data acquired through the point sampling process. In this way, the power of the developing geographic information system technology could be used to more completely understand the spatial domain of environmental problems.

REFERENCES

Dixon, W. J. (ed) 1977. Biomedical Computer Programs - P Series. University of California Press, Berkeley, California.

Dull, C. W. and W. H. Clerke. 1981. Loran-C Navigation as an Aid to Aerial Photographic Operations. Technical Papers of the American Society of Photogrammetry: ASP-ACSM Fall Technical Meeting. ASP, Falls Church, Virginia. :531-541.

Mace, T. H. 1982. Characterization of Lake Water Quality Parameters with Airborne Multispectral Scanner Data: Flathead Lake, Montana. Technical Papers of the American Society of Photogrammetry: ASP-ACSM 48th Annual Meeting. ASP, Falls Church, Virginia. :375-387.

Mace, T. H. 1983. Multispectral Scanner Characterization of Acid Rain Effects in Freshwater Lakes - Adirondack State Park, New York. Unpublished paper presented at the ASP-ACSM 49th Annual Meeting, Washington, D.C.

Robinove, C. J. 1982. Computation of Physical Values from Landsat Digital Data. Photogrammetric Engineering and Remote Sensing. 48(5):781-784.

Whitlock, C. H. , C. Y. Kuo, and S. R. LeCroy. 1982. Criteria for the Use of Regression Analysis for Remote Sensing of Sediment and Pollutants. Remote Sensing of Environment. (12):151-168.

Witzig, A. S. and C. A. Whitehurst. 1981. Literature Review of the Current Use and Technology of MSS Digital Data for Lake Trophic Classification. Technical Papers of the American Society of Photogrammetry: ASP-ACSM Fall Technical Meeting. ASP, Falls Church, Virginia. :1-20.

APPENDIX

SCATTER PLOTS OF PREDICTED VERSUS OBSERVED VALUES OF INVESTIGATED WATER QUALITY PARAMETERS FOR THE ROCHESTER-NIAGARA DATA SET

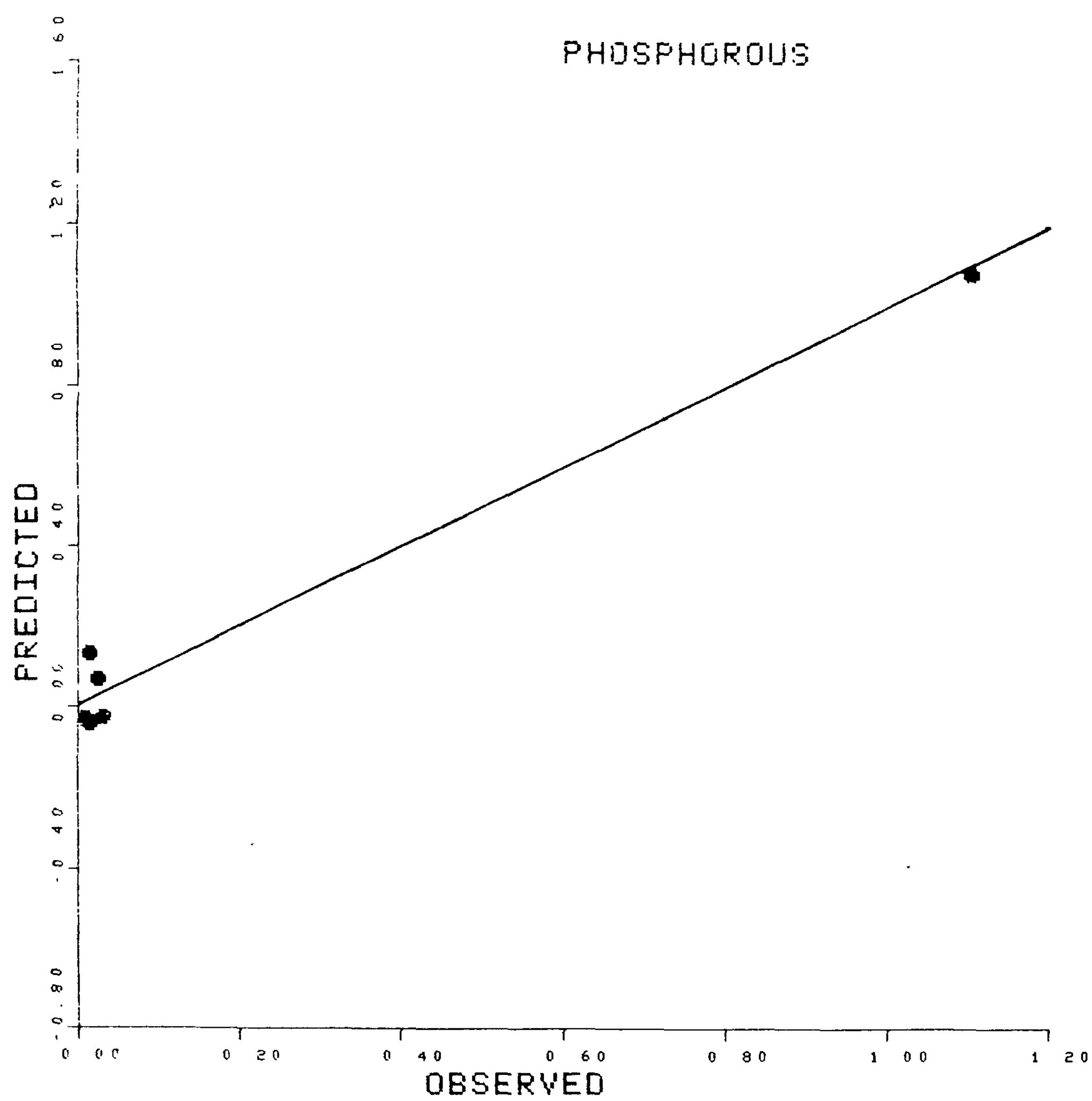

Figure A1. Predicted versus observed total phosphorous (mg/l).

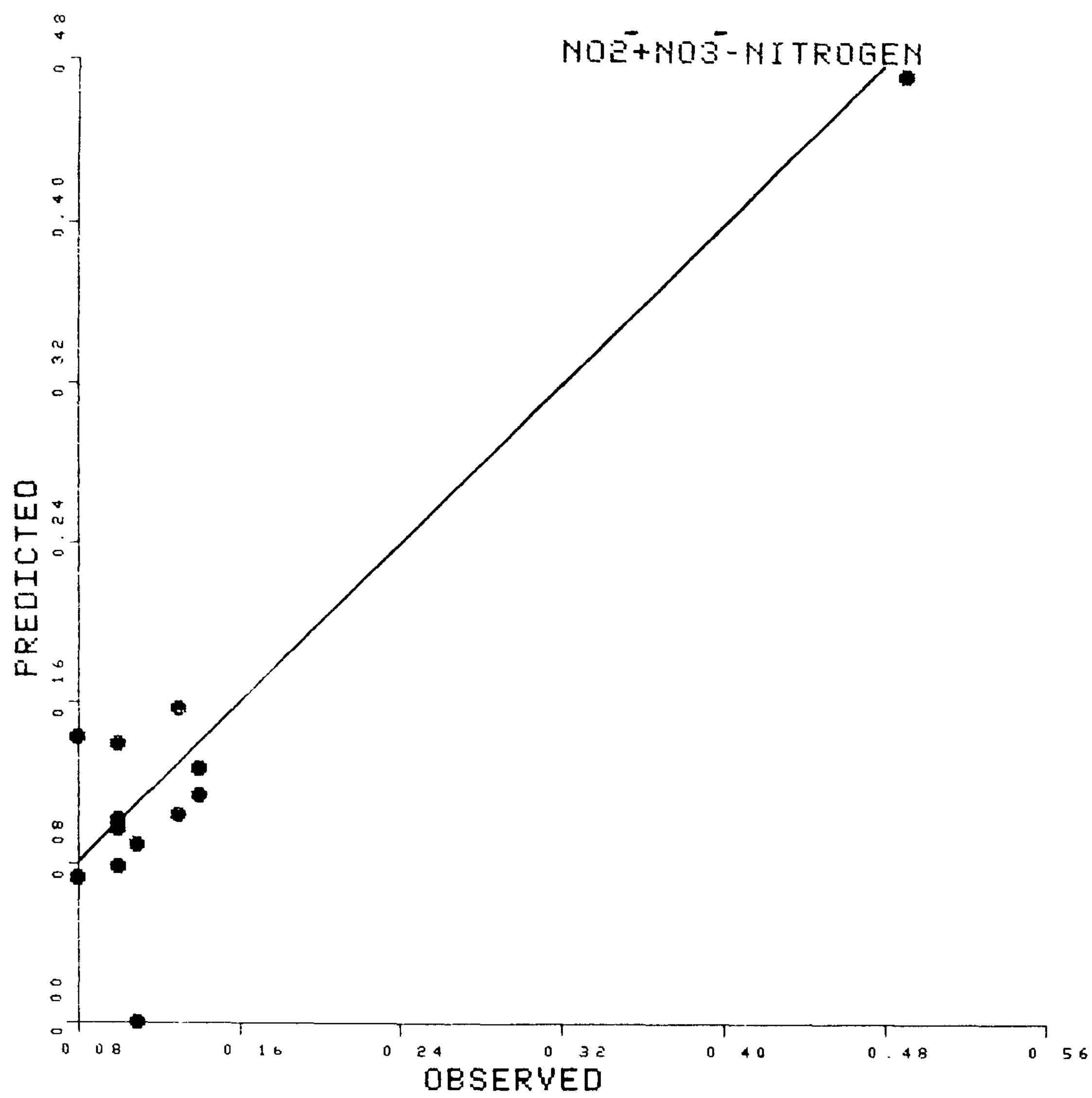

Figure A2. Predicted versus observed NO_2^- + NO_3^--nitrogen (mg/l).

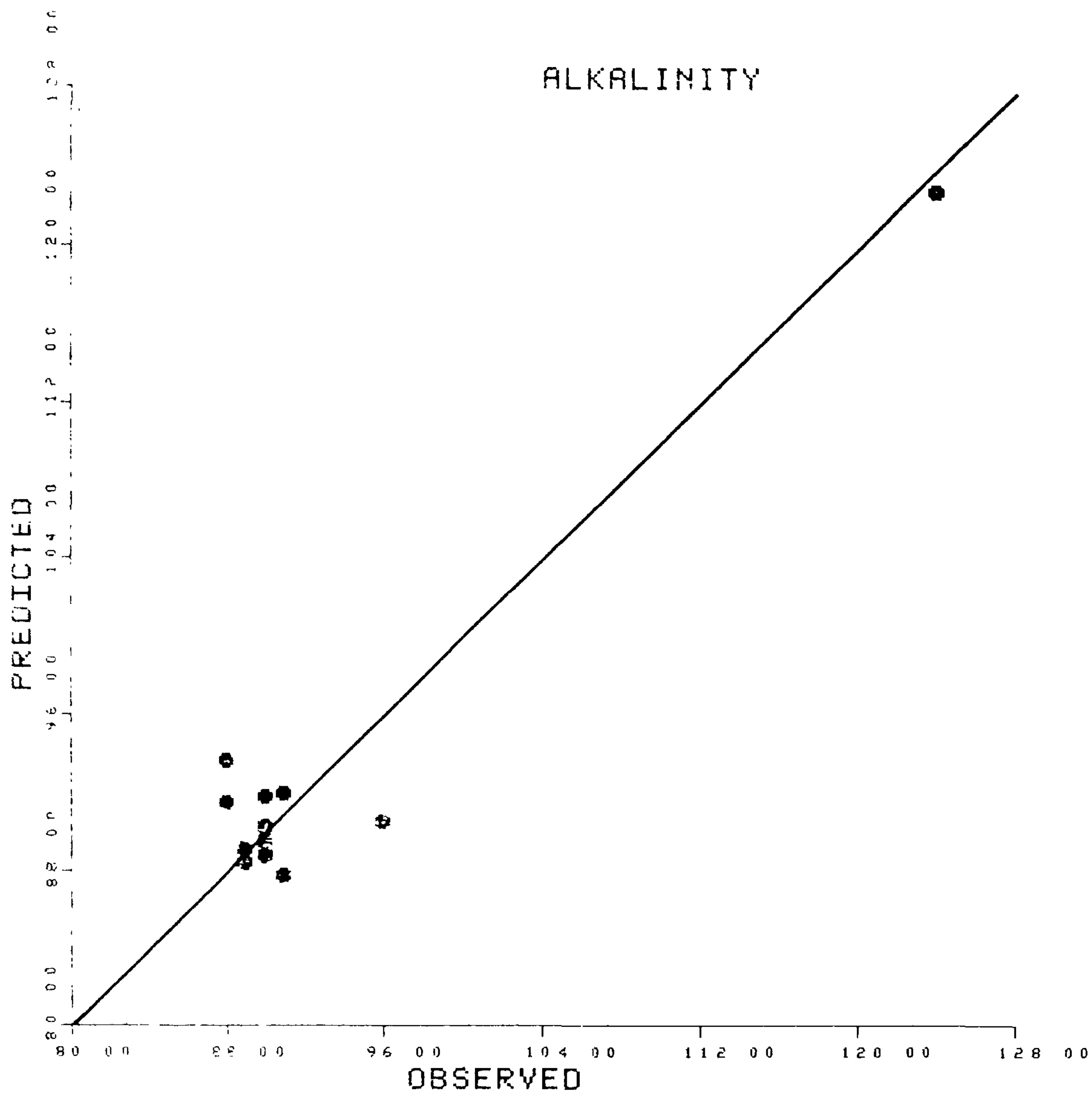

Figure A3. Predicted versus observed alkalinity (mg/l).

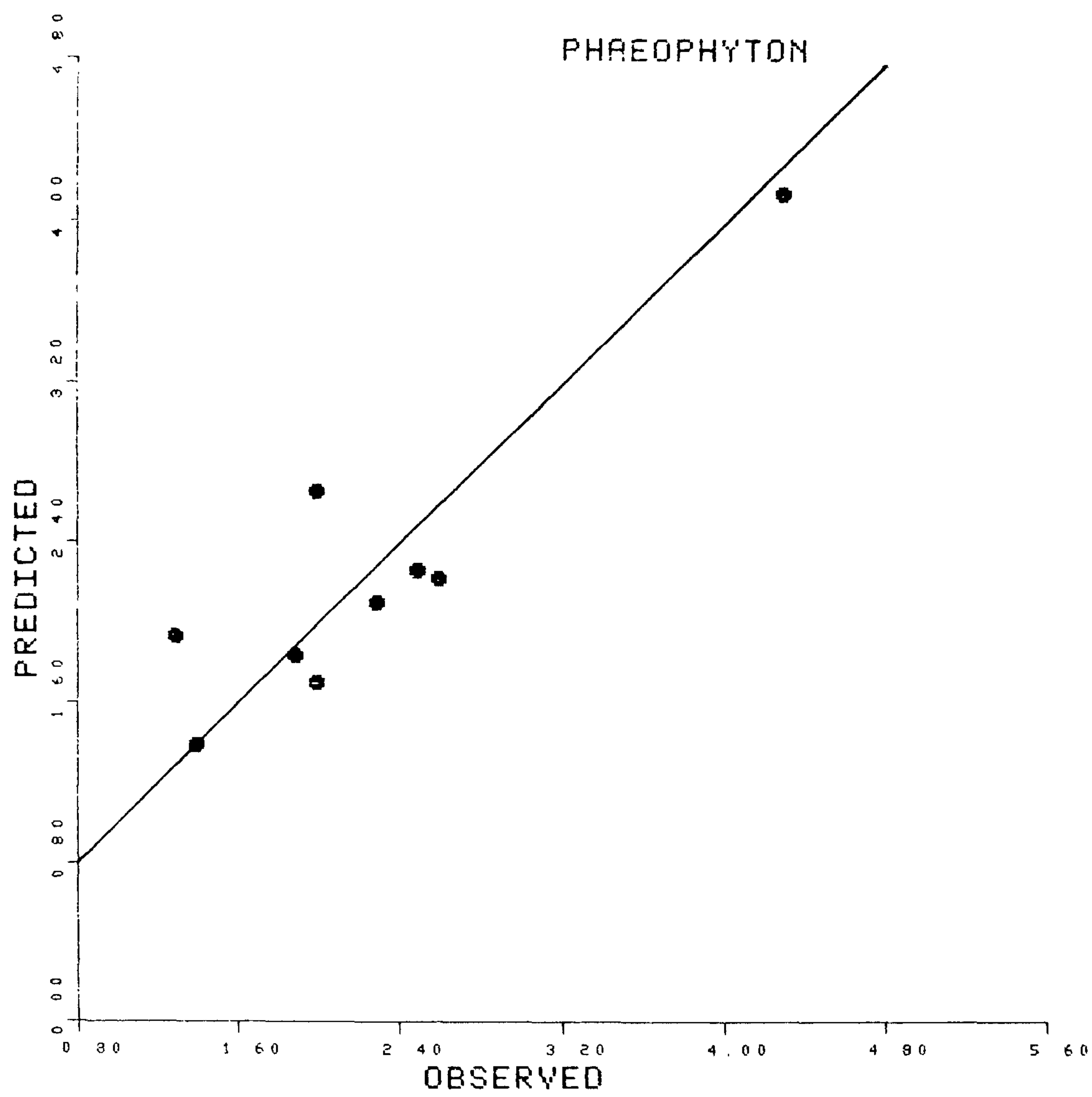

Figure A4. Predicced versus observed Phaeophyton a (μg/l).

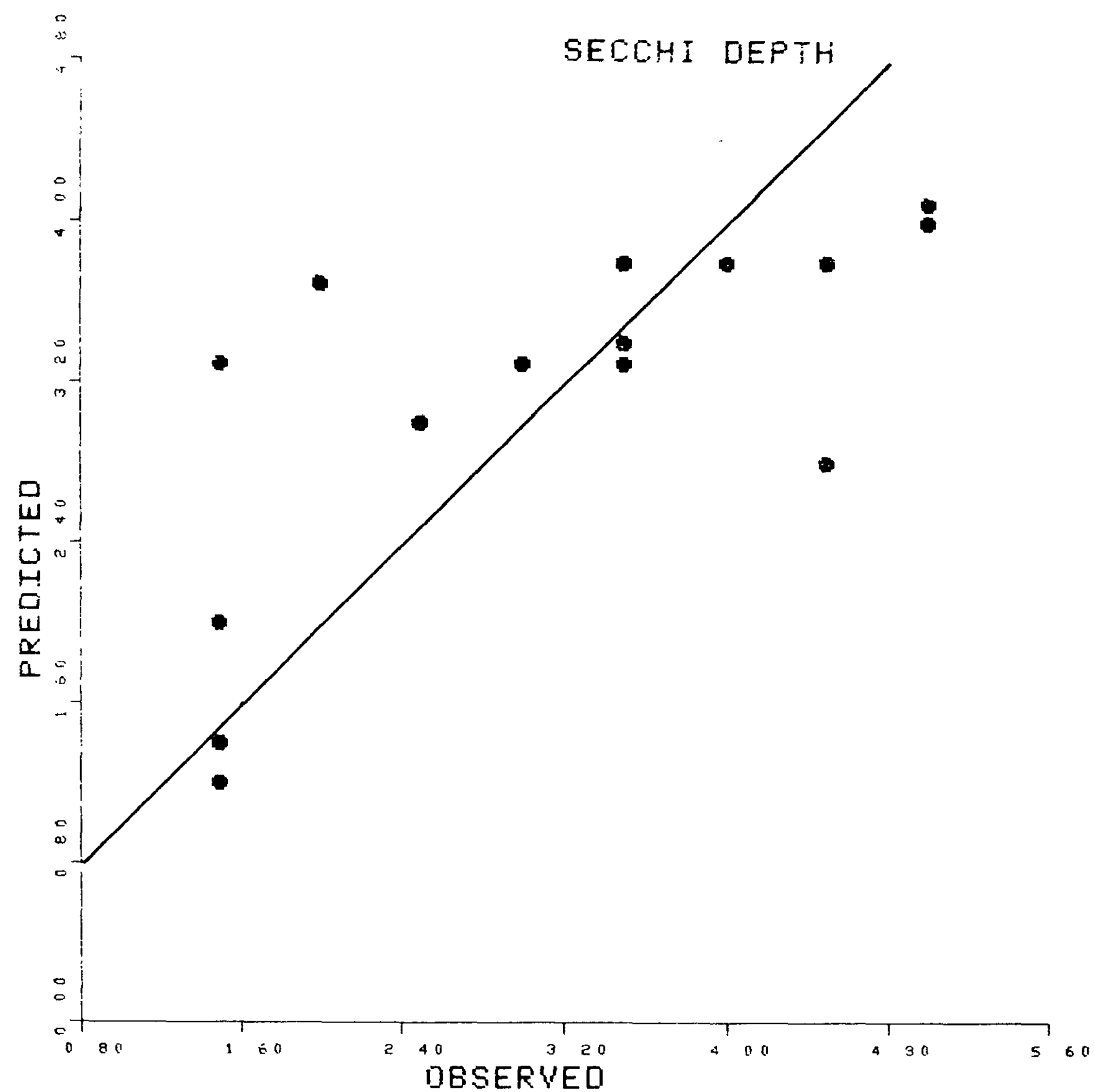

Figure A5. Predicted versus observed Secchi depth (m).

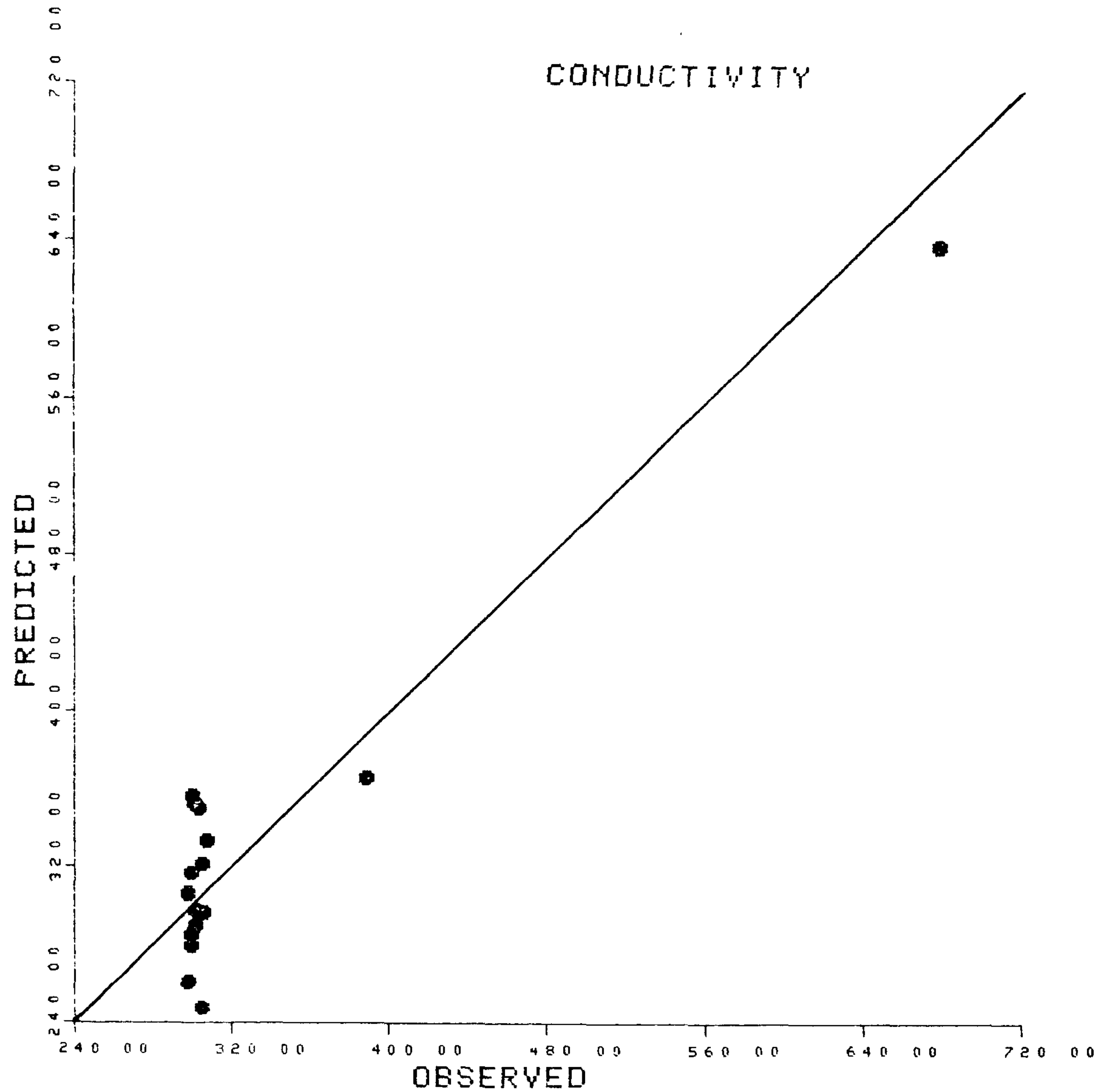

Figure A6. Predicted versus observed conductivity (μmhos/cm).

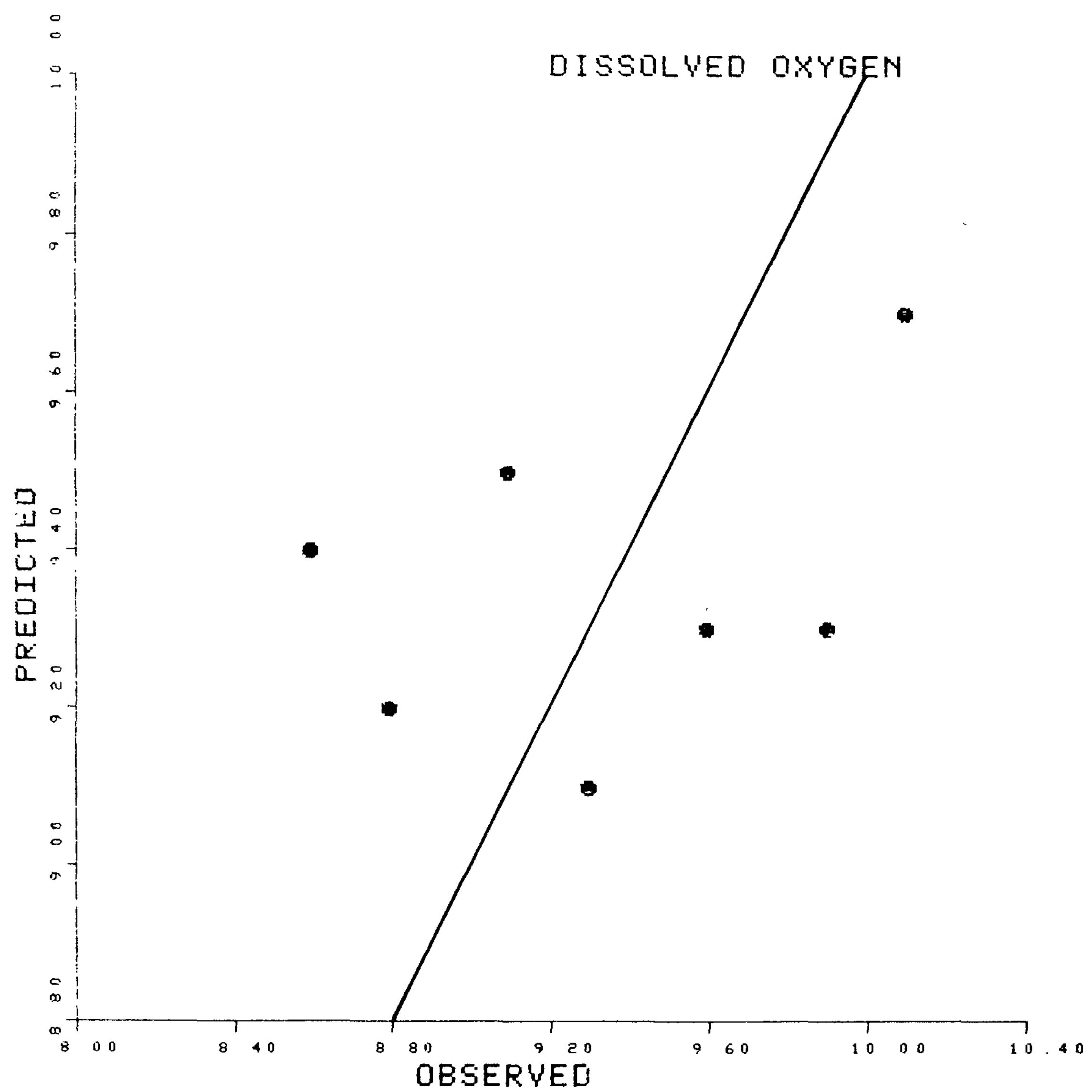

Figure A7. Predicted versus observed dissolved oxygen (mg/l).

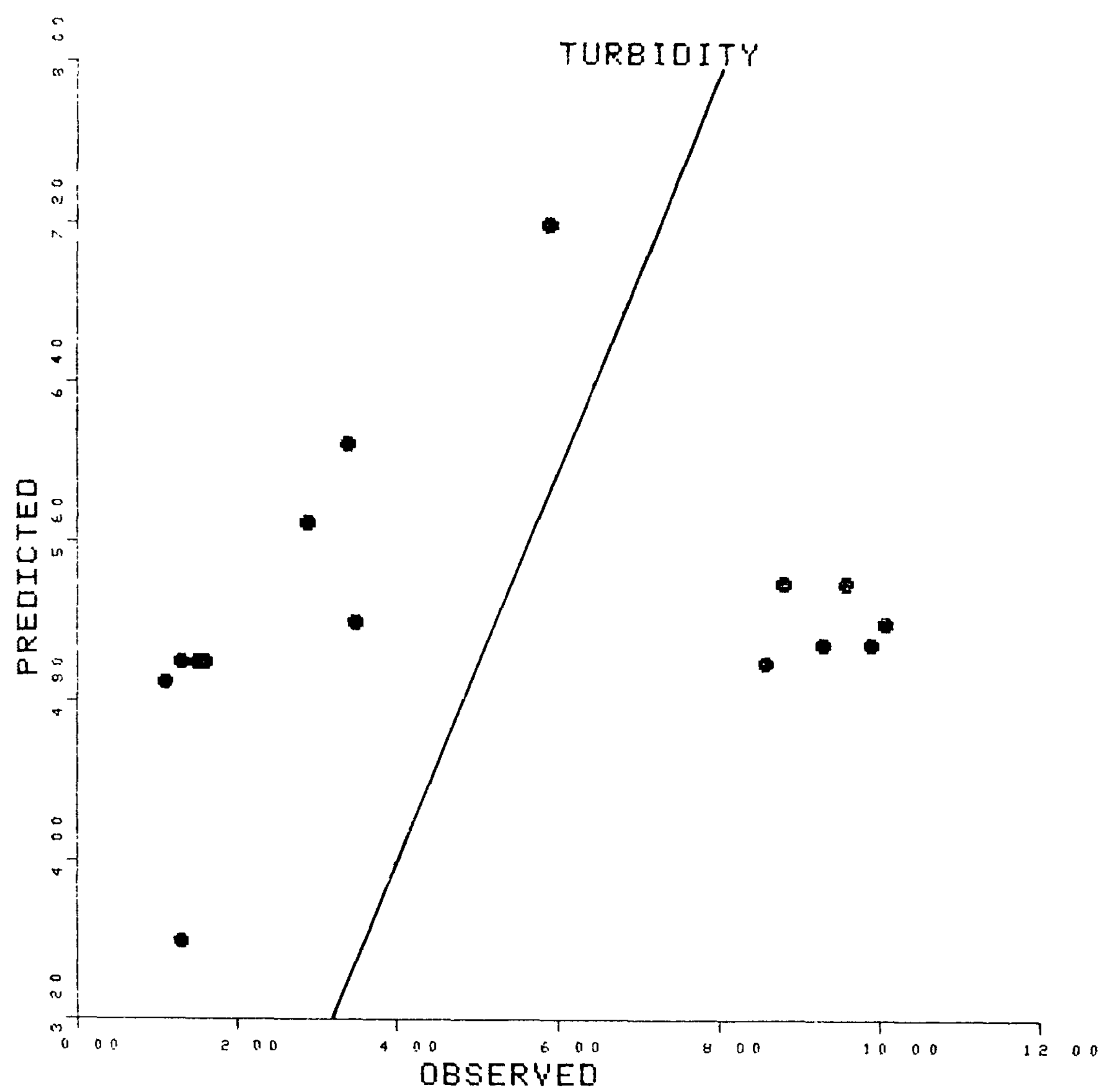

Figure A8. Predicted versus observed turbidity (Hach, Ftu).

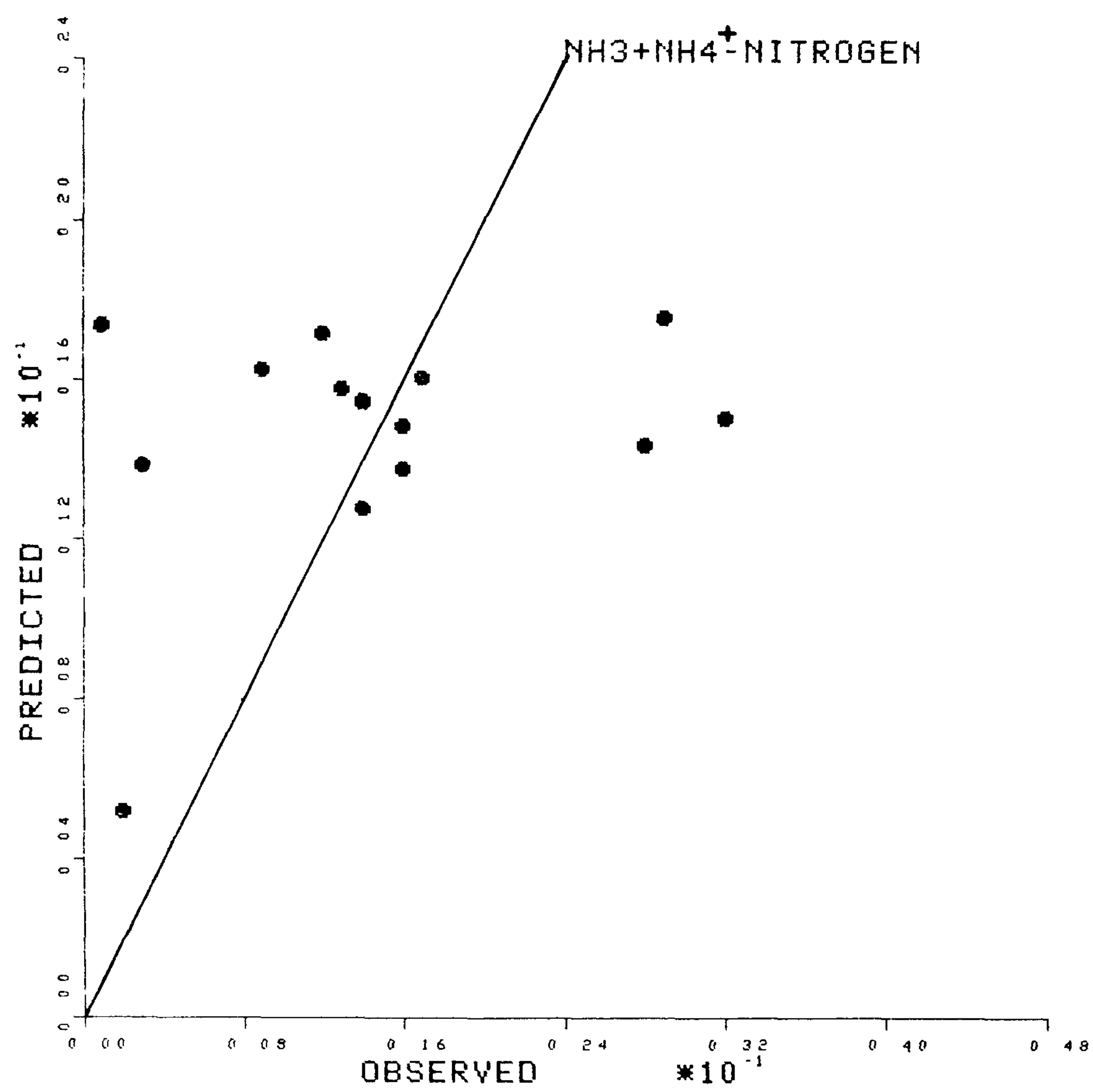

Figure A9. Predicted versus observed NH_3 + NH_4^+-nitrogen (mg/l).

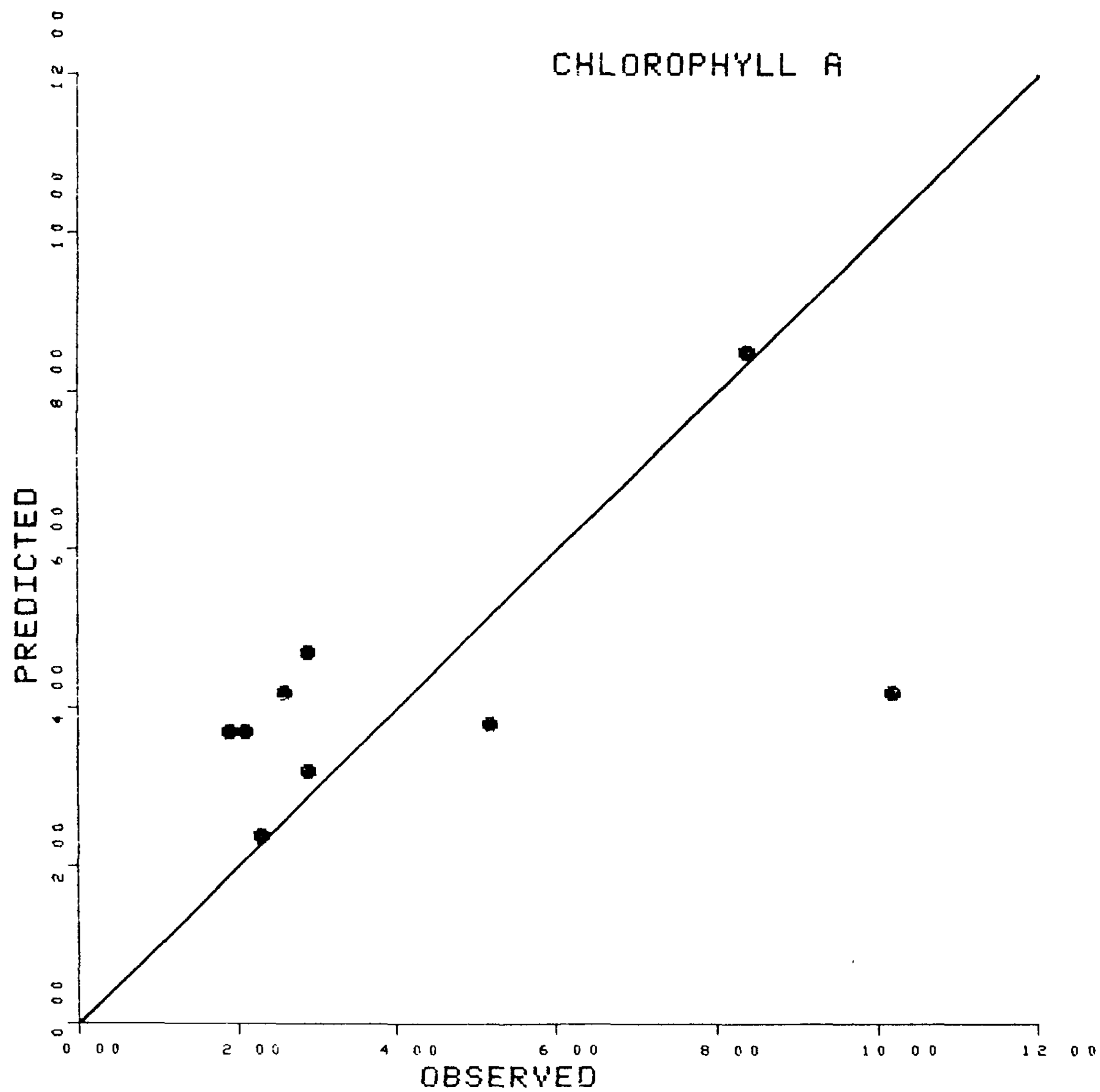

Figure A10. Predicted versus observed Chlorophyll a (μg/l).