Modeling Inland Water Quality Using Landsat Data

D. J. CARPENTER

Department of Engineering Physics, Research School of Physical Sciences, Australian National University, P. O. Box 4, ACT 2601, Australia

and

S. M. CARPENTER

Division of Mathematics and Statistics, CSIRO, P. O. Box 1965, ACT 2601, Australia

The water quality parameters turbidity and algal pigment concentration of freshwater lakes have been modeled and predicted using Landsat multispectral scanner data as multiple linear predictors. Satellite data for an area in South East Australia from seven occasions during 1978 and 1979 were used along with concurrent ground-based measurements from sampling sites on three lakes covering a wide range of water quality regimes. Date-independent models for turbidity and algal pigment were obtained using the satellite multispectral data and the water quality data from up to 21 sampling sites on one lake on six occasions. The sun elevation at the time of satellite overpass was included in the models to account for differences between dates, and the time of sample collection was included to compensate for diurnal variations in pigment fluorescence. These models were used to successfully predict these water quality parameters for this lake on a new occasion and for the two other lakes on three occasions.

I. Introduction

In an evaluation of Landsat multispectral scanner (MSS) data for water resources applications (Carpenter, 1982), water quality data were collected from three lakes in the Murrumbidgee River catchment of New South Wales, Australia, at times coincident with data collections by the Landsat satellites. The data collections took place between May 1978 and October 1979 and produced seven sets of coincident MSS and ground-based data.

Early in the Landsat program, it was demonstrated that sediment loads of inland waters could be estimated from Landsat MSS data (Kritikos et al., 1974), and, more recently, there have been moves toward general models for this parameter (Holyer, 1978). Munday and Alfoldi (1979) have discussed several such models and have suggested a nonlinear one based on diffuse reflectance, which appears to have considerable generality. Strong (1974) used Landsat MSS data to examine algal blooms and other studies have included MSS data in trophic state inventory or assessment programs for inland waters (Wezernak et al., 1976; Blackwell and Boland, 1979; Scarpace et al., 1979), in which sediment load, or turbidity, and algal pigment concentration were key indicators. From these studies, it is clear that turbidity and algal pigment concentration are water quality parameters accessible to remote sensing by Landsat MSS, although sophisticated techniques, including multidate analysis, multilinear regression, and data transformation by principal components analysis, must be applied to the digital data to obtain useful results.

0034-4257/83/$3.00

In this study, the relationships between the observable but subtle variations in the MSS data and measured changes in the water quality parameters were quantified using multiple linear regression analyses. In these, the dependent variables were turbidity and algal pigment concentration ("pigment") and the satellite data were the predictor variables. Six of the data sets from the largest and most variable lake included in the surveys were initially used to obtain models for turbidity and pigment for each of the individual sampling occasions. These individual date models were generally successful, but were limited in their applicability to the date and data range of each occasion. In order to produce models which could be used for predictive purposes and which were more general in applicability, the data from these six occasions were pooled and the sun elevation at the time of satellite overpass included to compensate for differences between dates. The models obtained from this data set were date independent and covered larger data ranges. The predictive abilities of these models were tested by applying them to the seventh set of data for this lake and to three sets of data from the other two lakes included in the ground sampling program.

2. Water Quality Data

The data were obtained from three lakes: Lake Burley Griffin, an urban, essentially eutrophic and highly turbid lake of 794 ha, located in Canberra, the National Capital of Australia (nine sampling sites); Lake Ginninderra, a second urban lake, of 105 ha, also in Canberra, usually clear and potentially eutrophic (six sites); and Burrinjuck Reservoir, a large (5300 ha) flood control and irrigation storage dam with a history of algal blooms during low river flows and of inflows of highly turbid water following heavy rains. The water level varies by typically 20 m during the cycle of summer irrigation demand and winter runoff recharge (21 sites).

The water quality data consisted of measurements of nephelometric turbidity (NTU) and algal pigment concentrations (chlorophyll *a* and pheopigments), measured as the fluorescence of extracted pigments (Strickland and Parsons, 1968), from two depths at 36 sampling sites over the three lakes. The two depths corresponded to the 100% (immediately subsurface) and the 50% downwelling quanta irradiance, obtained from the irradiance profiles. For the analyses the turbidity and pigment values from the two depths were averaged; there did not appear to be any consistent stratification of the water over that depth (often less than 2 m), and it was felt that the averaged values more properly represented the layer of water in which the backscattered irradiance measured by the satellite sensors was generated. Turbidities measured varied from less than 1 NTU to over 100 NTU, algal pigment concentration varied from less than 1 mg m^{-3} to almost 100 mg m^{-3}.

3. Satellite Data

The Landsat MSS data used were identified in the World Reference System (WRS) as Path 96, Row 84. The data were received and used as sets of multispectral digital radiance values on computer compatible tapes (CCTs), with accompanying single-band photographs for location guides and cloud cover assessment. These data were acquired on 5

May, 20 June, 8 July 1978, and 13 January, 13 March, 10 October, and 28 October 1979.

The Landsat MSS obtains data from four spectral bands: band 4, 500–600 nm (visible green); band 5, 600–700 nm (visible red); band 6, 700–800 nm; and band 7, 800–1100 nm (both near infrared). Clear water is a strong absorber of near infrared radiation so that only at wavelengths corresponding to bands 4 and 5 is there sensible penetration of the water by incident sunlight. This light is then backscattered over the depth of penetration to give rise to the volume reflected light eventually sensed at the satellite. However, extremely turbid water can reflect strongly in band 6 and, to a lesser extent, band 7, from near surface suspended particulate matter.

The MSS data as received invariably displayed considerable six-band striping over water areas which tended to obscure variations in the water reflectance. This striping was a result of residual calibration differences between the six detectors within each of the spectral bands and was frequently caused by these detectors having different forbidden or missing values following the calibration and decompression processes (Thomas, 1977). These effects can dominate at the (low) radiance levels reflected by water, especially given the small ranges of radiances (grey levels) from water bodies.

A simple recalibration procedure was developed to improve the homogeneity of the data over water areas by more closely aligning the detector responses for the range of grey levels of interest in each spectral band. The data were first returned to a close approximation to the raw (i.e., uncalibrated and decompressed) state by passing them through the inverse of the calibrated decompression look-up tables. Adjustments were then made to the detector offsets, in unit grey levels, until the best result (i.e., most homogeneous) was obtained for water areas, using one detector as a reference in each spectral band. All data from that band were then calibrated by the reference detector gain and offset and decompressed via one look-up table.

Radiance values from the four spectral bands were obtained by defining primitives of several pixels around each of the sampling sites located in the Landsat data and using the mean values of the four-band radiances of the primitives. The primitives were defined so as to exclude shorelines, to confidently include the actual sampling site, and to not overlap. The primitives contained up to 20 pixels and in all cases the observable changes in the satellite data occurred over distances larger than the typical dimensions of the primitives. For simplicity, all pixels were given equal weights. This procedure simultaneously overcame the problems of registration of the pixel matrix with the positions of the sampling sites and of any residual detector imbalance (or nonuniformity) within the spectral bands. The uncertainty in registration arises since the pixel matrix is not fixed relative to the ground and there is uncertainty not only in the location of the sampling site on the lake, but also in the location of the lake in the satellite data.

4. Multiple Linear Regressions

The six data sets were used to form linear models for the water quality parameters from the satellite data using a multiple regression procedure. The dependent variables were the measured values of turbidity and pigment, averaged over the two sample depths, and the predictor variables were the spectral radiance values from the MSS bands. Individual models were fitted for the Burrinjuck Reservoir, and, following an initial examination of the residuals, the transformation to logarithm of pigment and turbidity was used to reduce the variance of the larger values. Initially all four bands were included as predictors to allow the comparison of models between dates, but in all cases a more efficient model was obtained using fewer bands; there were usually strong correlations between the bands, and the small variation in band 7 over water meant that it could contribute little to the variance of the water quality parameters. The time of sample collection was also tested as a linear predictor for pigment since previous studies (Parker and Tranter, 1981) indicate that fluorescence for a given pigment concentration varies diurnally, with a minimum at noon. All combinations of the predictor variables were tested so as to find the most efficient models. These models fitted the data well, but as each was restricted in applicability to the data set used to generate it, they could not be used for prediction purposes for new data.

In an attempt to create a single model for each water quality parameter that was more generally valid, and which could thus be used for prediction, the six individual sampling occasion data sets were combined and the sine of the sun elevation at the satellite overpass time was included as a predictor to compensate for changes in overall brightness between scenes. In these models, band 7 was omitted as before, and the other three bands tested in combinations for inclusion. The time of sample collection was again also tested as a linear predictor for pigment.

Samples were assumed to be independent but, for the overall models in particular, this assumption may not have been true, and the adjusted multiple correlation coefficient (R^2) might then be inflated. A better way to judge the success of the models was to use them to predict the water quality parameters for other occasions and lakes; a new data set for the Burrinjuck Reservoir and data from two other lakes on three occasions were used to evaluate these models.

5. Results

Generally, the six individual sampling date models fitted the data well, with adjusted multiple correlation coefficients (R^2) in the range 0.59–0.95 for turbidity and 0.50–0.85 for pigment; the most efficient models for each occasion are shown in Table 1. The particularly poor fit for pigment for the June data may have been a result of the low values on that date (< 1.0 mg m^{-3}), compounded by the very high turbidity (> 30 NTU) and the small sample size of 8 (half the lake was cloud-covered on that date). The July pigment and turbidity values were much the same as in June as there had been an enormous influx of turbid runoff water in early June and this took several weeks to settle; the July models were also not very successful. The time of sample collection was generally found to be a significant predictor for pigment.

The general models obtained by including the sun elevation as a predictor were very successful, fitting the data extremely well. This indicated that scene-to-scene (date-to-date) variations could be compensated for, leading to the generation of

TABLE 1 Multiple Linear Regression Models for Burringjuck Reservoir Data; Individual Data Sets

DATE	MODEL[a,b]	% VARIANCE ACCOUNTED FOR
6 May 1978[c]	$T = -0.252(0.671)+0.211(0.169)B4+0.133(0.0919)B5$	90.2
20 June 1978	$T = 3.33(0.349)+0.0636(0.339)B4-0.128(0.0459)B6$	59.0
	$P = 9.289(4.36)-1.48(0.507)B6-0.00290(0.00289)TIME$	50.3
8 July 1978	$T = 3.42(0.461)-0.305(0.0741)B4+0.190(0.0615B5+0.302(0.106)B6$	64.1
	$P = 3.59(1.25)+0.218(0.183)B4-0.00486(0.000664)TIME$	80.3
13 January 1979	$T = -2.31(0.171)+0.300(0.0193)B5-0.0518(0.0196)B6$	94.6
	$P = 8.03(1.87)-0.0837(0.0706)B4-0.00477(0.000906)TIME$	83.0
26 March 1979	$T = -1.03(0.638)+0.21660(0.0850)B4+0.155(0.0513)B6$	87.9
	$P = -1.96(1.28)+0.213(0.236)B4+0.0387(0.141)B5$	85.1
10 October 1979	$T = -5.19(0.797)+0.501(0.0652)B4$	75.3
	$P = 2.19(3.17)+0.284(0.216)B4-0.00350(0.0000731)TIME$	74.8

[a] $T = \log_e$(turbidity in NTU); $P = \log_e$(pigment in mg m^{-3}); TIME = time of satellite overpass.
[b] Standard errors shown in parenthesis.
[c] Pigment data not available for 6 May 1978.

date-independent models. In fact, these overall models were more successful than most of the individual date models, although for pigment it was found necessary to include sampling time to obtain the best fit since this was also a significant predictor. The most efficient models were:

$$\log(T) = 4.51(0.174)+0.304(0.0447)*B4 -0.0727(0.0327)*B5 +0.0534(0.0326)*B6 -10.5(0.517)*SUN,$$

$$R^2 = 0.963,$$

$$\log(P) = 5.48(0.519)-0.114(0.0766)*B4 -0.0546(0.0537)*B5 +5.12(0.845)*SUN -0.00479(0.000360)*TIME,$$

$$R^2 = 0.901,$$

where B4, B5, and B6 refer to the MSS spectral band data (range 0–127) SUN refers to the sine of the sun elevation (range 17–47°) at the time of satellite overpass, and TIME to the time of day of sample collection (range 0820–1330 h); the standard errors of the coefficients are shown in parentheses in the equations.

These models using the three-band data and sun elevation, and sampling time for pigment, fit the data extremely well and appeared to be satisfactory for predictive purposes. This was tested by applying them to a new data set from the Burrinjuck Reservoir and to data from Lakes Burley Griffin and Ginninderra on three occasions. This was not an exhaustive test since the data were from the same lake as that used to generate the models, or came from other lakes on the same dates and in the same scene as that from the Burrinjuck Reservoir, and these lakes all lie in the same river system. However, Lakes Burley Griffin and Ginninderra do not show obvious similarities with the Burrinjuck Reservoir, particularly in a date-by-date examination. There were 18 data points in the new Burrinjuck data (from 28 October 1979), 27 in total for Lake Burley Griffin, and nine in total for Lake Ginninderra (from 13 January, 13 March, and 28 October 1979). All of the new data lay within the ranges of turbidity and pigment used in the generation of the regression models. The correlations between the observed and predicted data for the new data set for the Burrinjuck Reservoir were 0.88 for log(turbidity) and 0.73 for log(pigment), indicating that the models could accurately predict the water quality for this lake on a new occasion from the satellite data. For Lake Burley Griffin these correlations were 0.90 and 0.61, respectively, while for Lake Ginninderra the correlation for turbidity was 0.92, but there was no significant correlation for the pigment data, indicating that the turbidity model could reliably predict the values in other lakes but that the model for pigment was not general enough to be useful for all other lakes, failing to predict the values for Lake Ginninderra.

6. Discussion

Landsat MSS data are not well suited for the remote sensing of water quality since the spectral bands of the multispectral scanner are mostly located in that part of the spectrum where water absorbs strongly. Also, the sensitivity of the detectors is set for the generally more strongly reflecting land covers such as soil and vegetation. These problems are exacerbated by the data processing and presentation methods generally employed, which are again designed for land cover reflectances. However, by using a simple recalibration technique specific to data obtained over water, and by using multiple linear regressions based on data from small groups of pixels, it has been shown that both turbidity and algal pigment concentration can be modeled and predicted from the digital MSS data.

Turbidity and pigment are parameters widely associated with water quality and are frequently measured in water resource assessment or monitoring programs. These parameters were well modeled from Landsat MSS for one lake in the Murrumbidgee River catchment of NSW using data spanning 18 months and representing a large variety of water quality regimes. It was necessary to include the time of sample collection in the pigment models to take into account a diurnal variation in pigment fluorescence; this was a result of the time taken to collect the ground reference data compared to the almost instantaneous collection by the satellite. It was not necessary to include all four spectral bands in the models since there are strong correlations between them, and in any case band 7 can have only a very weak response to variations in water color and turbidity. The inclusion of the sun elevation was seen to adequately account for the overall brightness variations between scenes, allowing the use of data from several dates in the generation of date-independent regressions suitable for prediction. A more sophisticated atmosphere correction scheme, such as that adopted by Scarpace et al. (1979), was not required, given the success of the models. Similarly, linear models appeared to be adequate, and the use of more complex models involving nonlinear terms, such as those suggested by Munday and Alfoldi (1979), could not be justified.

The models were also used to predict water quality parameters in this lake on a new occasion and in two other lakes in the same river catchment during this period. Turbidity was well predicted in all cases, and pigment was well predicted for the Burrinjuck Reservoir. However, for the other lakes pigment was not so well predicted, with the model failing to predict pigment values for Lake Ginninderra. This may have been a result of the difficulty in extracting truly representative radiance values for this small lake, i.e., values that were not affected by the surrounding land, bottom reflections, or macrophytes; the lake is shallow and was generally relatively clear. On one of the occasions, 8 July 1978, there was the possibility that clouds adjacent to this lake could have affected the radiance values extracted.

The models generated allowed the extrapolation of point-sampled data to entire lakes so that water sample measurements could be used to produce complete distribution maps for the lakes. Figure 1 shows an example of a distribution map for turbidity at the Burrinjuck Reservoir with quantized turbidity levels shown by various grey levels; the most turbid areas are brightest, the least turbid areas are darkest, and the turbidity range of 2.4–4.8 NTU is represented by 32 grey levels. The models could also be used to predict turbidity and, less accurately, pigment for these lakes from new satellite data, assuming that similar regimes were present.

Although the results given here are not generally applicable to all inland water bodies, they suggest that the MSS satellite data can be a useful tool in monitoring water quality following a relatively limited collection of ground-based data. The limi-

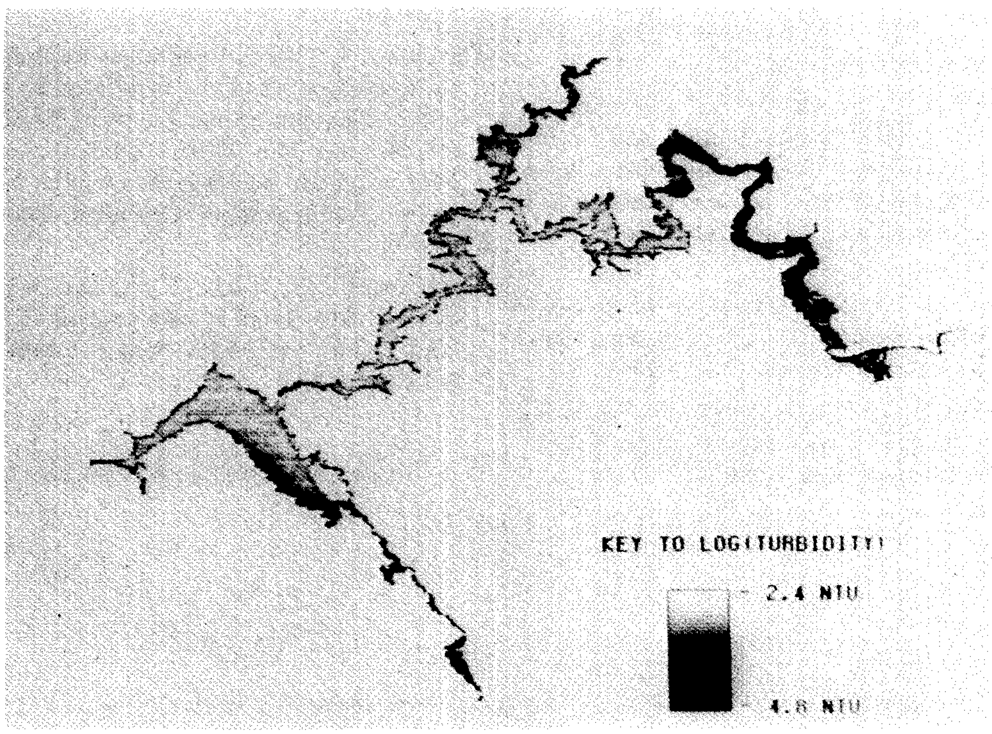

FIGURE 1. Turbidity distribution in Burrinjuck Reservoir from Landsat MSS data, 28 October 1979. Turbidity estimated from model:

$$\log(\text{turbidity}) = 4.51 + 0.304(B4) - 0.0727(B5) + 0.0534(B6) - 10.5(SUN).$$

tations of the current MSS data will be largely overcome with the advent of improved satellite sensors with increased spatial, spectral and radiometric resolutions, e.g., the Landsat-4 Thematic Mapper. The assessment and monitoring of inland water quality from satellite data can then be expected to become reliable and accurate.

This research was supported by an Australian Water Resources Council grant from the Department of National Development and Energy of the Australian Federal Government.

References

Blackwell, R. J., and Boland, D. H. P. (1979), Trophic classification of selected Colorado lakes, JPL Publication 78-100, JPL, Pasadena, California.

Carpenter, D. J. (1982), An Assessment of Landsat MSS Satellite data for inland water resources applications, Australian Water Resources Council Technical Paper No. 72. Australian Government Publishing Service, Canberra, 1982.

Holyer, R. J. (1978), Toward universal multispectral suspended sediment algorithms, *Remote Sens. Environ.* 7:323–338.

Kritikos, H., Yorinks, L., and Smith, H. (1974), Suspended solids analysis using ERTS-A data, *Remote Sens. Environ.* 3:69–80.

Munday, J. C., and Alfoldi, T. T. (1979), Landsat test of diffuse reflectance models for aquatic suspended solids measurement, *Remote Sens. Environ.* 8:169–183.

Parker, R. R., and Tranter, D. J. (1981), Estimation of algal standing stock and growth parameters using *in vivo* fluorescence, *Austr. J. Marine Freshwater Res.* 32:629–638.

Scarpace, F. L., Holmquist, K. W., and Fisher, L. T., (1979), Landsat analysis of lake quality, *Photogramm. Eng. Remote Sens.* 45:623–663.

Strickland, J. D. H., and Parsons, T. R. (1968), *A Practical Handbook of Sea Water Analysis*, Fisheries Research Board of Canada, Ottawa, Canada.

Strong, A. E. (1974), Remote sensing of algal blooms by aircraft and satellite in Lake Erie and Utah Lake, *Remote Sens. Environ.* 3:99–107.

Thomas, V. L. (1977), Generation and physical characteristics of the Landsat 1, 2, and 3 MSS computer compatible tapes, Technical Memorandum 78018, Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, Maryland.

Wezernak, C. T., Tanis, F. J., and Bajza, C. A. (1976), Trophic state analysis of inland lakes, *Remote Sens. Environ.* 5(3):147–152.

Received 14 June 1982; revised 2 December 1982.