

Considerações sobre a Modelagem Empírica de Variáveis  
Limnológicas no Reservatório de Barra Bonita Utilizando Dados  
Landsat-5 TM

CLÁUDIA ZUCCARI FERNANDES BRAGA<sup>1</sup>  
EVLYN MÁRCIA LEÃO DE MORAES NOVO<sup>2</sup>  
THELMA KRUG<sup>3</sup>

INPE - Instituto Nacional de Pesquisas Espaciais  
Caixa Postal 515

12201-970 São José dos Campos - SP - Brasil

<sup>1</sup>claudia@ltid.inpe.br

<sup>2</sup>evlyn@ltid.inpe.br

<sup>3</sup>thelma@ltid.inpe.br

**Abstract.** This work analyses optically active limnological data and Landsat-5 TM data obtained during the winter of 1989, acquired for the project SISA (Remote Sensing of Aquatic Ecosystems). Empirical models were developed using stepwise linear multiple regression analysis to estimate the concentration of limnological variables through multispectral data of the Barra Bonita Reservoir. The stations for which the correlations between in situ data and TM data were weak were eliminated from the data files. The models obtained after the file reduction showed, in most cases, improved adjusted coefficients ( $r^2$ ). The proposed method is discussed in detail and the results from the analysis of the 04 and 20 of July data are shown.

**Resumo.** Este trabalho analisa dados coletados no âmbito do Projeto SISA (Sensoriamento Remoto de Sistemas Aquáticos), referentes ao período de inverno de 1989. Foram realizadas duas missões de campo com coleta de dados limnológicos e imagens do sensor Thematic Mapper (TM) do satélite Landsat-5. As variáveis limnológicas opticamente ativas medidas nessas campanhas (Total de sólidos em suspensão - TSS, material inorgânico do TSS, material orgânico do TSS, concentração de clorofila total e profundidade Secchi) e os valores médios de reflectância extraídos das bandas TM1, TM2, TM3, TM4, das razões entre as bandas TM1 e TM2 e TM2 e TM3 e do produto das bandas TM1 e TM3, correspondentes às mesmas estações de coleta, foram tratados estatisticamente em estudos de correlação. As bandas significativamente correlacionadas com as variáveis limnológicas foram utilizadas como variáveis independentes em análises de regressão linear múltipla, visando o desenvolvimento de modelos matemáticos empíricos para predição de cada uma das variáveis indicadoras de qualidade de água. Este procedimento permite a elaboração de mapas temáticos das variáveis limnológicas bem correlacionadas com os produtos de satélite, em toda a região do corpo d'água contido na

imagem. Desta forma, as técnicas de amostragem em campo, que geram informações apenas pontuais, podem ser complementadas espacialmente. O Reservatório de Barra Bonita é formado pela junção de dois rios, o Tietê e o Piracicaba, limnologicamente bastante diferentes. Assim, os estudos de correlação e modelagem foram realizados considerando o reservatório como um todo, e particionado em três compartimentos, a saber: a região do encontro junto à barragem, o braço do rio Tietê e o braço do rio Piracicaba. Desta forma pode-se modelar um número muito maior de variáveis limnológicas do que seria possível considerando-se apenas o reservatório todo. Ainda, a análise das retas de regressão ajustadas para cada par de variáveis testado permitiu a identificação dos pontos de maior variância, que, retirados dos modelos, produziram, em geral, um melhor ajuste. O presente trabalho apresenta modelos desenvolvidos para as quatro regiões estudadas (todo o reservatório, região do encontro, braço do Tietê e braço do Piracicaba), em duas datas. Alguns procedimentos são sugeridos, visando a utilização racional das informações fornecidas pelas imagens TM, para a elaboração de modelos matemáticos empíricos a serem aplicados a estudos de qualidade de água.

### **Introdução e Objetivo**

O Projeto SISA (Sensoriamento Remoto de Sistemas Aquáticos) compreende o estudo da qualidade da água do Reservatório de Barra Bonita através da utilização de dados de sensoriamento remoto, a nível de campo, aeronave e orbital (Novo e Braga, 1991). Este reservatório está localizado no estado de São Paulo, sendo formado pelos rios Tietê e Piracicaba. Foram realizadas 16 coletas de dados limnológicos simultaneamente às passagens do satélite Landsat-5 sobre a região de estudo, entre maio de 1989 e outubro de 1991, abrangendo as estações de verão e de inverno. Entretanto, em somente 05 das datas amostradas, as condições atmosféricas permitiram a obtenção de imagens TM, sendo três delas correspondentes a períodos de inverno e duas a períodos de verão.

Parte dos objetivos deste

projeto consiste no desenvolvimento de modelos matemáticos empíricos para a estimativa de variáveis limnológicas oticamente ativos que são correlacionados com os dados espectrais das imagens adquiridas. Posteriormente, os modelos empíricos desenvolvidos para uma determinada variável, em cada uma das datas de trabalho, são analisados em conjunto, objetivando estabelecer-se um modelo operacional aplicável a imagens futuras, para monitoramento da qualidade da água da região de estudo.

A partir do lançamento dos primeiros satélites ambientais foram iniciadas pesquisas com o objetivo de explorar a potencialidade de imagens multi-espectrais para o estudo dos recursos naturais da Terra. Estes primeiros satélites, no entanto, não contavam com uma definição de faixas espectrais apropriada ao estudo de corpos d'água. O levantamento das características espectrais da

água, em laboratório (McCluney, 1976; Bricaud e Sathindranath, 1981; Witte et al., 1982), forneceu subsídios para a especificação de novos sensores para operar em bandas espectrais específicas para este fim, como foi o caso do Coastal Zone Color Scanner (CZCS), lançado em 1978. O CZCS opera em bandas propícias ao estudo da concentração de fitoplâncton, mas sua resolução espacial (800 m) e cobertura da terra limitam sua aplicação ao estudo de áreas costeiras e oceânicas. Exemplos de aplicações deste sensor, no desenvolvimento de modelos de estimativa de parâmetros de qualidade de água, podem ser encontrados, entre outros, em Clark (1981), Bricaud et al. (1987), Deschamps e Viollier (1987).

Em 1984, o lançamento do Landsat-5 TM, com uma banda espectral na faixa do azul e resolução espacial de 30 m, favoreceu o estudo de corpos d'água continentais e costeiros de menor área. Exemplos de quantificação de clorofila ou de sólidos em suspensão na água, utilizando dados TM, podem ser encontrados, entre outros, em Lodwick e Harrington (1985), Garcia e Caselles (1990), Ekstrand (1991) e Braga e Setzer (1991).

A análise preliminar dos resultados dos estudos de correlação do projeto SISA revelou uma grande variabilidade em relação às variáveis limnológicas significativamente correlacionadas com os dados multiespectrais, em todas as 05 datas de trabalho (Novo e Braga, 1993). Isto aumenta a dificuldade de se desenvolver modelos genéricos confiáveis, sugerindo a necessidade de um investimento mais significativo

na área de processamento de imagens digitais, como, por exemplo, correção atmosférica, teste de novas operações (razões, produtos) entre bandas do satélite, etc. Neste contexto, e dada a relevância da possibilidade de se obter mapas de estimativa de variáveis limnológicas a nível sinótico, ganha mais importância ainda a otimização dos modelos obtidos empiricamente, a fim de se minimizar os erros envolvidos nas estimativas das variáveis de interesse.

A coleta de dados na água se dá num período de aproximadamente 04 horas, enquanto que a obtenção das imagens se dá em menos de 15 segundos. As condições superficiais da água, no momento do imageamento pelo satélite, são, portanto, desconhecidas. Ocorrências tais como plantas aquáticas flutuantes ou vegetação morta, por exemplo, podem influenciar os valores de radiância medidos pelo sensor remoto, tornando-os não representativos das condições de qualidade da água no local. Além disto, problemas intrínsecos ao sensor (Mather, 1987) podem introduzir ruídos no sistema, descaracterizando a verdadeira resposta espectral dos alvos. A identificação destes valores tendenciosos nos arquivos de dados é feita através da análise das retas de regressão ajustadas para cada par de variáveis correlacionadas, através da busca dos pontos de maior variância, e na frequência com que estes mesmos pontos se mostram mal ajustados no conjunto inteiro de correlações observadas para cada arquivo de dados.

Neste trabalho são discutidos os modelos desenvolvidos a partir das correlações

encontradas entre os dados orbitais e limnológicos, referentes às duas primeiras datas de inverno (04 e 20 de julho de 1989), objetivando obter estimativas das variáveis modeladas.

### **Materiais e Métodos**

Nos dias 04 e 20 de julho de 1989, datas de passagem do satélite Landsat-5 sobre a região de estudo, foram estimadas em 30 estações à superfície, as seguintes variáveis: temperatura, profundidade Secchi (SEC), nutrientes e variáveis limnológicas opticamente ativas (total de sólidos em suspensão (TSS), material orgânico em suspensão (MO), material inorgânico em suspensão (MI) e concentração de clorofila total (CLTOT)). A medida de profundidade Secchi é um estimador da turbidez da água e, por estar relacionado à cor da mesma, apresenta, via de regra, uma boa correlação com dados de sensoriamento remoto. Neste estudo os dados de sensoriamento remoto considerados foram: valores médios de reflectância extraídos das bandas TM1, TM2, TM3, TM4, razões entre as bandas TM1 e TM2 (RB1B2), entre as bandas TM2 e TM3 (RB2B3) e o produto das bandas TM1 e TM3 (MB1B3). Através de estudos de correlação entre os dados de campo e os dados das imagens, identificaram-se aqueles pares de variáveis cujo coeficiente de correlação,  $r$ , era maior ou igual a 0,59. Em cada data de trabalho, tratou-se o reservatório como um todo (considerando-se, ao todo, 30 estações), e também como sendo formado por três compartimentos limnologicamente independentes, quais sejam, o braço do rio Tietê (11 estações), o braço do rio

Piracicaba (11 estações) e a região do encontro, o lago principal do reservatório (08 estações). As duas datas de coleta utilizadas no presente trabalho constituem-se das 3ª e 4ª datas do projeto SISA. Foram analisados, portanto, os seguintes arquivos: dados de 04 de julho de 1989, considerando-se o reservatório como um todo (BB3), a região do encontro (BB3E), o braço do Piracicaba (BB3P), e o braço do Tietê (BB3T). Os arquivos análogos correspondentes à data de 20 de julho de 1989 são referenciados por BB4, BB4E, BB4P e BB4T.

O desenvolvimento dos modelos parte dos resultados de correlação. As bandas significativamente correlacionadas com uma determinada variável opticamente ativa são introduzidas em análises de regressão linear múltipla como variáveis independentes, e a variável que se deseja estimar como variável dependente. Adotou-se um nível de confiança de 95%. O modelo final inclui parcimoniosamente as bandas que melhor descrevem o comportamento da variável estimada. Em outras palavras, são incluídas no modelo aquelas bandas que oferecem o melhor ajuste aos dados observados, avaliado em termos do coeficiente de determinação,  $r^2$ . Para se otimizar estes modelos, partiu-se da análise do ajuste das retas de regressão entre os pares de variáveis de campo e espectrais significativamente correlacionadas. Foram identificadas as estações cujos valores observados eram mais discrepantes em relação aos valores estimados pelas retas de regressão. As estações que apresentaram valores considerados discrepantes em pelo menos 50% dos casos de correlações estatisticamente significativas

observadas para um determinado arquivo de dados, foram retiradas destes mesmos arquivos. Eliminadas estas estações, os cálculos de correlação foram refeitos. Utilizando-se teste de hipóteses, foram identificadas as variáveis limnológicas que apresentaram melhoras significativas nos coeficientes de correlação com as variáveis espectrais. Para estas variáveis limnológicas foram refeitas as análises de regressão linear múltipla, e os novos modelos desenvolvidos foram comparados aos anteriores, em termos de seus coeficientes de determinação.

### Resultados e Discussão

A repetição dos estudos de correlação, após a eliminação de algumas estações dos arquivos de dados, nem sempre apresentou resultados significativamente melhores. Isto sugere que a identificação das amostras de maior variância em relação às retas de regressão deve ser feita de maneira mais específica, fixando-se, por exemplo, a variável limnológica que se deseja modelar, e estudando-se as correlações desta com as variáveis espectrais. A justificativa para este procedimento se baseia na possibilidade de ocorrer erro na medida das variáveis limnológicas, muito embora tenha-se assegurado a existência de réplicas para análise laboratorial, de forma a permitir a análise da consistência dos dados. No entanto, quanto menos genérico for o comportamento de uma determinada estação amostral, em termos dos desvios do modelo proposto, maior a chance de eliminá-la indevidamente do modelo.

Com exceção de BB3 e BB3T, todos os outros conjuntos de dados apresentaram melhoras significativas nos coeficientes de correlação dos novos modelos. A Tabela 1 mostra as variáveis limnológicas modeladas, as variáveis espectrais significativamente correlacionadas com as mesmas (com os respectivos valores de  $r$  e o número de observações,  $n$ ), as variáveis inseridas no modelo final, e o coeficiente de determinação,  $r^2$ , para cada arquivo de dados, antes da eliminação das estações de coleta. A Tabela 2 apresenta as mesmas informações para os modelos desenvolvidos após a redução dos arquivos de dados. Ressalta-se que estas tabelas apresentam apenas os casos em que houve melhora significativa nos resultados dos estudos de correlação, refeitos através do procedimento proposto.

A análise dos resultados apresentados nas Tabelas 1 e 2 mostra que no modelo para estimativa de TSS, utilizando os dados de BB4E, a variável R12 não apresentou melhora significativa na correlação com TSS, após a redução do arquivo. Verificaram-se duas novas correlações, com R23 e com M13. Apesar da correlação entre TSS e R23 não ser alta, R23 foi incluída no modelo final, que apresenta um excelente ajuste ( $r^2=0,92$ ). O mesmo ocorreu com o modelo para MI, também com os dados de BB4E. A variável R23 apresentou fraca correlação com MI após a redução do arquivo, mas foi também incluída no modelo final.

Na modelagem para estimativa de CLTOT, utilizando dados de BB4T, verificou-se um comportamento anômalo: as duas únicas correlações significativas encontradas entre CLTOT e as

variáveis espectrais (TM2 e R12), antes da redução do arquivo, não se reproduziram após a redução do mesmo, quando CLTOT passou a se correlacionar com outras quatro variáveis (Tabela 2). O modelo final proposto, no entanto, é melhor que o primeiro. Este é um caso que justifica a necessidade de se estudar, de forma mais restrita, os desvios das observações com relação às retas de ajuste. Neste caso, por exemplo, considerando-se apenas as correlações de CLTOT com os produtos espectrais, conforme sugerido acima.

A análise comparativa das duas tabelas revela resultados diferentes para os arquivos referentes a uma mesma região, em cada data. Na maioria dos casos, além de um aumento significativo nos coeficientes de correlação entre as variáveis limnológicas e os dados espectrais, outras correlações significativas foram encontradas, após a redução dos arquivos. Os dados de TSS referentes a BB3P, assim como os dados de CLTOT, referentes a BB4T, apresentaram correlações significativas com os dados espectrais, o que não se verificara antes da redução dos arquivos. Para BB3P desenvolveu-se um modelo de estimativa da concentração de clorofila com um ajuste de 0,70 que, apesar de moderado, significa a possibilidade de mapear um parâmetro extremamente importante para avaliação de qualidade da água. Em BB4T, as correlações do TSS com os produtos espectrais foram baixas e não produziram um modelo razoável ( $r^2=0,28$ ).

### Conclusão

O método proposto para se

melhorar o ajuste de modelos empíricos para a estimativa da concentração de variáveis limnológicas opticamente ativas através da utilização de imagens Landsat-5 TM mostrou ser adequado, produzindo resultados bastante promissores. Coeficientes de determinação da ordem de 0,60, por exemplo, atingiram valores em torno de 0,75, 0,85, ou até mesmo 0,90, em alguns casos, após a aplicação da metodologia proposta. Além da melhoria dos ajustes, o método possibilitou o mapeamento de variáveis limnológicas que não eram significativamente correlacionadas com os dados espectrais, antes da redução dos arquivos de dados.

### Agradecimentos

As autoras agradecem o Centro de Recursos Hídricos e Ecologia Aplicada - CREA/USP, entidade conveniada com o INPE, responsável pela realização das coletas e análises laboratoriais de dados limnológicos do projeto SISA, a Companhia Energética de São Paulo - CESP, que também fornece apoio para as coletas em campo e dados de suas estações climatológicas e hidrológicas, e a Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP, financiadora do referido projeto (processo 88/4016-5) durante dois anos.

### Referências

- Braga, C.Z.F.; Setzer, A.W. Modeling suspended solids concentration based on TM/Landsat-5 images at Guanabara Bay, Rio de Janeiro, Brazil. Presented at the 24<sup>th</sup> International Symposium on Remote Sensing

- of Environment, ERIM, Rio de Janeiro, Brazil, 27-31 May 1991. Summaries. ERIM, Ann Arbor, MI. pp. 59, 1991.
- Bricaud, A.; Sathindranath, S. Spectral signatures of substances responsible for the change in ocean colour. Signatures Spectrales D'objects em Teledetection. Avignon, 8-11 sep. 1981.
- Bricaud, A.; Morel, A.; Andre, J.M. Spatial/temporal variability of algal biomass and potential productivity in the mauritanian upwelling zone, as estimated from CZCS data. Advances in Space Research. 7(2):53-62, 1987.
- Clark, D.K. Phytoplankton pigment algorithms for the Nimbus-7 CZCS. Oceanography from Space, New York, J.I.R. Gower, 1981. pp.227-237.
- Deschamps, P.Y.; Viollier, M. Algorithms for ocean color from space and application to CZCS data. Advances in Space Research, 7(2)11-19, 1987.
- Ekstrand, S. Quantification of chlorophyll-a in coastal waters using Landsat TM. Presented at the 24<sup>th</sup> International Symposium on Remote Sensing of Environment, ERIM, Rio de Janeiro, Brazil, 27-31 May 1991. Summaries. ERIM, Ann Arbor, MI. pp. 55-56, 1991.
- Garcia, M.J.L.; Caselles, V. A multi-temporal study of chlorophyll-a concentration in the Albufera lagoon of Valencia, Spain, using Thematic Mapper data. International Journal of Remote Sensing, 2(2):301-311, 1990.
- Lodwick, G.D.; Harrington, R.D. Anais do VII SBSR, 1993
- Deriving sediment information for Lake Athabasca using principal component analysis of Landsat data. Canadian Journal of Remote Sensing, 11(1):4-16, 1985.
- Mather, P.M. Computer processing of remotely-sensed images: an introduction. New York, NY, John Wiley, 1987.
- McCluney, W. Remote measurement of water color. Remote Sensing of Environment, 5:3-33, 1976.
- Novo, E.M.L.M.; Braga, C.Z.F. Use of TM/Landsat data to estimate chlorophyll concentration and turbidity in the Barra Bonita reservoir. Presented at the 24<sup>th</sup> International Symposium on Remote Sensing of Environment, ERIM, Rio de Janeiro, Brazil, 27-31 May 1991. Summaries. ERIM, Ann Arbor, MI. pp. 55-56, 1991.
- Novo, E.M.L.M.; Braga, C.Z.F. II Relatório do Projeto Sensoriamento Remoto de Sistemas Aquáticos. São José dos Campos, 1993 (no prelo).
- Witte, W.G.; Whitlock, C.H.; Harris, R.C.; Usry, J.W.; Poole, L.R.; Houghton, W.M.; Morris, W.E.; Gurganus, E.A. Influence of dissolved organic material on turbid water optical properties and remote sensing reflectance. Journal of Geophysical Research, 87(C1):441-446, 1982.

Tabela 1: Modelos empíricos desenvolvidos para a estimativa de variáveis limnológicas.

ARQUIVO PAR. ESTIM.	VARIÁVEIS CORRELACIONADAS					VARIÁVEIS NO MODELO		$r^2$
(n)	r							
BB3E	CLTOT	TM1	TM3	R23	M13	R23		0,67
(8)	-0,64	-0,80	0,85	-0,80				
BB3P	TSS	TM3	TM4	R23	M13	TM4		0,46
(11)	0,73	0,72	-0,69	0,71				
BB4	MI	TM3	R23	M13		R23		0,68
(30)	0,70	-0,83	0,60					
	SEC	TM3	R23			R23		0,69
(30)	-0,76	0,84						
BB4E	TSS	TM2	TM3	TM4	R12	TM3		0,44
(8)	0,73	0,72	0,70	-0,72				
	MI	TM2	TM3	TM4	R12	M13	TM3	0,85
(8)	0,86	0,93	0,87	-0,61	0,86			
BB4P	TSS	R23				R23		0,64
(11)	-0,80							
	MI	TM3	TM4	R23	M13	R23	M13	0,89
(11)	0,70	0,64	-0,90	0,63				
	SEC	R23				R23		0,62
(11)	0,79							
BB4T	CLTOT	TM2	R12			R12		0,67
(11)	-0,81	0,84						

Tabela 2. Modelos empíricos desenvolvidos para a estimativa de variáveis limnológicas, após a redução dos arquivos de dados.

ARQUIVO/PAR. ESTIM.		VARIÁVEIS CORRELACIONADAS						VARIÁVEIS NO MODELO				r <sup>2</sup>
(n)		r										
BB3E	CLTOT (7)	TM1 -0,72	TM3 -0,94	R23 0,91	M13 -0,94							0,86
BB3P	TSS (8)	TM2 0,59	TM3 0,81	TM4 0,77	R12 -0,60	R23 -0,86	M13 0,83	TM3 TM4	R12 R23	M13	0,99	
	CLTOT (8)	TM1 -0,71	TM2 -0,79	TM3 -0,59	R12 0,74	M13 -0,65					0,70	
BB4	M1 (28)	TM3 0,76	R12 -0,60	R23 -0,86	M13 0,66							0,74
	SEC (28)	TM3 -0,86	R12 0,73	R23 0,89	M13 -0,79					0,87		
BB4E	TSS (7)	TM2 0,77	TM3 0,85	TM4 0,87	R12 -0,71	R23 -0,59	M13 0,78	TM2 TM3	TM4	0,44		
	M1 (7)	TM2 0,86	TM3 0,94	TM4 0,89	R12 -0,72	R23 -0,59	M13 0,92	TM2 TM3	TM4 R23	0,85		
BB4P	TSS (9)	TM3 0,75	TM4 0,79	R23 -0,91	M13 0,67							0,94
	M1 (9)	TM3 0,88	TM4 0,85	R12 -0,71	R23 -0,97	M13 0,82					0,93	
	SEC (9)	TM3 0,79	TM4 -0,79	R12 0,63	R23 0,92	M13 -0,72					0,94	
BB4T	CLTOT (9)	TM3 0,91	TM4 0,65	R23 -0,95	M13 0,74							0,67
	TSS (9)	TM1 -0,61	TM2 -0,64	R23 -0,61							0,28	