

Modelos para estimativa da concentração de sedimentos em suspensão em rios amazônicos de águas brancas via sensoriamento remoto

Otávio Cristiano Montanher
Evlyn Márcia Leão de Moraes Novo
Cláudio Clemente Faria Barbosa

Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{otaviocm, evlyn}@dsr.inpe.br; claudio@dpi.inpe.br

Abstract. This paper presents empirical models for suspended sediment concentration (SSC) retrieval from Landsat 5 images in several Amazon Rivers. In view of great gaps in the time series of in situ SSC database in Amazon Rivers, estimates based on historical orbital data may be an option to open new research possibilities. The models are based on a database composed of 504 *in situ* samples and near-simultaneously Landsat images. Two approaches are tested: i) using the entire database and ii) regionalizing the data according to environmental features of the watersheds they belong. The results show that the use of the whole database does not provide accurate SSC estimates. The regional modeling provides better estimates by fragmenting the data into five clusters. All these models display p-values $\approx 1 \times 10^{-6}$, R^2 values ranging from 0,83 to 0,91. The cross validation LOOCV and relative error values also showed their robustness. The models are very accurate, mainly for low SSC levels, between 0 to 200 mg/l. As the concentration increases, the absolute error increases too, but relative errors remain low (up to 7%).

Palavras-chave: Landsat 5, Amazon Rivers, empirical models, Landsat 5, rios amazônicos, modelos empíricos.

1. Introdução

A concentração de sedimentos em suspensão (CSS) é um parâmetro importante para a biota de ecossistemas aquáticos (Wetzel, 2001), para o desenvolvimento de planícies de inundação (Pierce e King, 2008) e em muitos problemas de engenharia, como no assoreamento de reservatórios (Thornton, 1990). No sistema fluvial do Rio Amazonas os sedimentos em suspensão são fortes reguladores de processos ecológicos nos rios e planícies de inundação (McClain e Naiman, 2008; Junk et al. 2011) e são indicadores de processos geomorfológicos nas bacias de drenagem, em especial na cadeia de montanhas dos Andes (McClain e Naiman, 2008).

A amostragem da CSS nos rios da Bacia Amazônica é caracterizada por baixa frequência espacial e temporal. Os dados de domínio público fornecidos pela ANA (Agência Nacional de Águas, disponíveis em: www.ana.gov.br/portalsnirh) mostram isso: são 97 estações amostrais em toda a bacia amazônica com 2,72 amostragens por ano para cada estação, em média. Os dados disponibilizados pelo programa HYBAM (<http://www.ore-hybam.org/>) possuem uma maior frequência de amostragem, mas possuem uma menor cobertura temporal e abrangem apenas os principais rios amazônicos. Tais características dos dados *in situ* limitam o estudo do transporte da carga sedimentar nos rios amazônicos. Uma forma alternativa para o estudo da carga suspensa em rios amazônicos é com uso de sensoriamento remoto, por exemplo, com dados MODIS (Espinoza Villar et al. 2012; Martinez et al. 2009).

A série de dados MODIS possui uma boa calibração radiométrica e uma alta frequência de imageamento. No entanto, sua resolução espacial não permite a avaliação de muitos corpos hídricos, tornando sua aplicação limitada aos canais principais, mesmo na Amazônia. Outra limitação é a restrição temporal (2000 – atual). Levando em conta estas características, este trabalho apresenta modelos que foram desenvolvidos a partir de dados Landsat 5, sensor TM, visando suprir essa lacuna de informação. Optou-se por trabalhar com esta série de dados, tanto pelo ao acesso gratuito, como pela abrangência temporal de décadas.

2. Metodologia de trabalho

Os modelos foram gerados por meio de uma abordagem empírica, tomando-se a base de dados *in situ* como dados de calibração e valores de reflectância derivados a partir das imagens TM de diversas órbitas ponto (Figura 1.A). A base de dados *in situ* é formada por 23 estações em muitas das principais drenagens amazônicas (Figura 1.A). Estes dados foram adquiridos tanto da base da ANA quanto do HYBAM.

Para escolha das imagens definiu-se um limiar de tempo transcorrido entre as coletas *in situ* e o imageamento. Em condições de baixos níveis d'água, aceitou-se até nove dias de diferença, já em imagens da estação chuvosa, com maiores níveis de CSS, não foram selecionadas cenas com mais de quatro dias de diferença. A pesquisa e aquisição das imagens foram feitas em: www.dgi.inpe.br/CDSR. Ao total, foram selecionadas 504 imagens, das quais 85% têm uma defasagem temporal de até quatro dias.

Nas imagens, a área de água mais próxima às estações *in situ* foram submetidas a uma amostragem dos pixels. Neste procedimento evitou-se a coleta em áreas contaminadas por nuvens e sombras, barras arenosas, plantas aquáticas e efeitos de turbulência, bem como efeitos de adjacência (Figura 1.B).

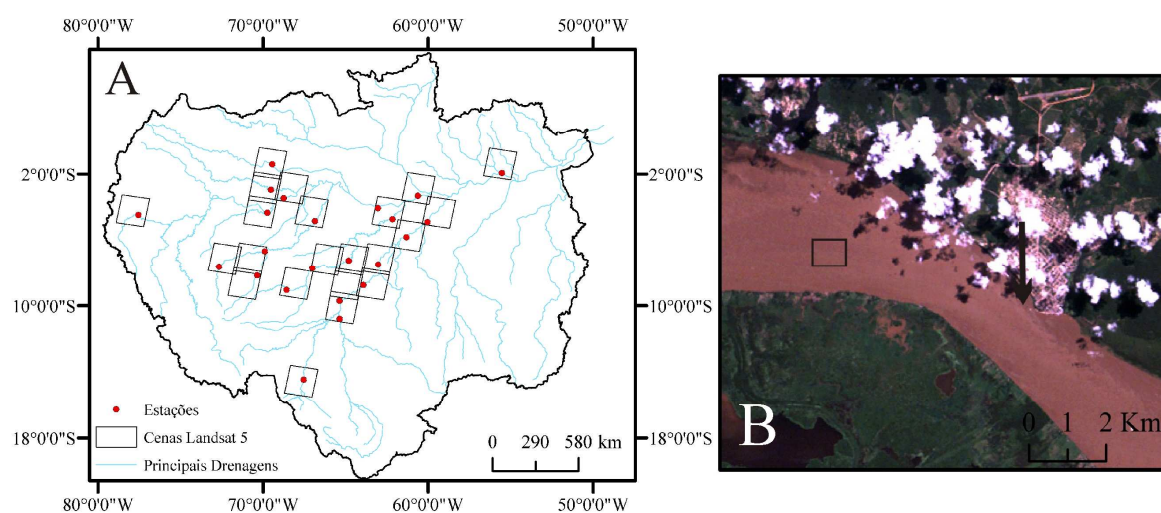


Figura 1. Localização das estações de amostragem de CSS e das órbitas ponto das cenas Landsat utilizadas (A). Exemplo de recorte de uma imagem Landsat 5 TM, composição colorida natural. O local da seção transversal da estação de Óbidos, Rio Amazonas, é indicado pela seta. A área selecionada para amostragem dos pixels está representada pelo quadrado preto.

Após o recorte das cenas, os números digitais (ND) foram convertidos para uma série de valores de reflectância (ρ). A primeira conversão foi feita de ND para $\rho_{apa(\lambda)}$, segundo Chander et al. (2009). A partir dos valores de $\rho_{apa(\lambda)}$ foram calculados valores de $\rho_{sup(\lambda)}$ com uso do código computacional 6S (Vermote et al. 1997). A partir dos valores de $\rho_{sup(\lambda)}$ foram feitas correções dos efeitos da interface ar/água, derivando-se, por meio de duas técnicas diferentes, dois valores de $\rho_{ag(\lambda)}$. Foram feitos testes com todos estes conjuntos de dados, pois isso permite avaliar se os processamentos estão realmente sendo úteis para aumentar a precisão dos modelos.

Após esta primeira etapa de amostragem da cena e a série de conversões físicas, os pixels foram novamente submetidos a uma amostragem. A distância interquartil entre o limite superior do primeiro quartil e o limite superior do terceiro quartil foi utilizada para definição de outliers. Isto foi feito para minimização de efeitos de ruído radiométrico ou mesmo de efeitos aleatórios, não reconhecidos na etapa de interpretação visual. Após a exclusão dos outliers foi calculada a média em cada banda, em cada nível de correção.

Para cada nível de correção da reflectância, foram utilizadas as bandas espectrais de 1 a 5 e uma série de razões entre bandas. Além destas variáveis contínuas, foram introduzidas nos modelos as seguintes variáveis categóricas: estação, rio e região. As duas primeiras são respectivamente a estação e o rio onde foi feita a coleta do dado, e a “região” foi uma definição adotada apenas para os rios Purus e Juruá, para diferenciação dos trechos: Alto, Médio e Baixo curso. Isto foi necessário porque há uma diferenciação evidente nos níveis de CSS entre estes trechos dos rios. Algumas destas variáveis podem ser redundantes ou mesmo não fornecer nenhuma informação. Portanto procedeu-se a seleção de atributos com uso do algoritmo Greedy disponível no software livre Weka 3.6.7 (Hall et al. 2009), que leva em conta efeitos de colinearidade entre as variáveis e a significância de cada uma.

A geração dos modelos é dividida em duas partes: a geração de um modelo global e a geração dos modelos particionados. Inicialmente buscou-se a definição de um modelo único, o que facilitaria a aplicação do mesmo para estimativa da CSS em outras imagens. Mas também foi testado se a partição por conjuntos de estações em bacias hidrográficas similares resultaria em uma melhora na estimativa dos modelos. Esta regionalização das estações *in situ* foi feita com base nas principais características do meio físico que regulam a produção de sedimentos e materiais orgânicos: geologia, geomorfologia, precipitação e cobertura do solo. Esta análise resultou em cinco agrupamentos (Figuras 2 e 3, Tabela 1).

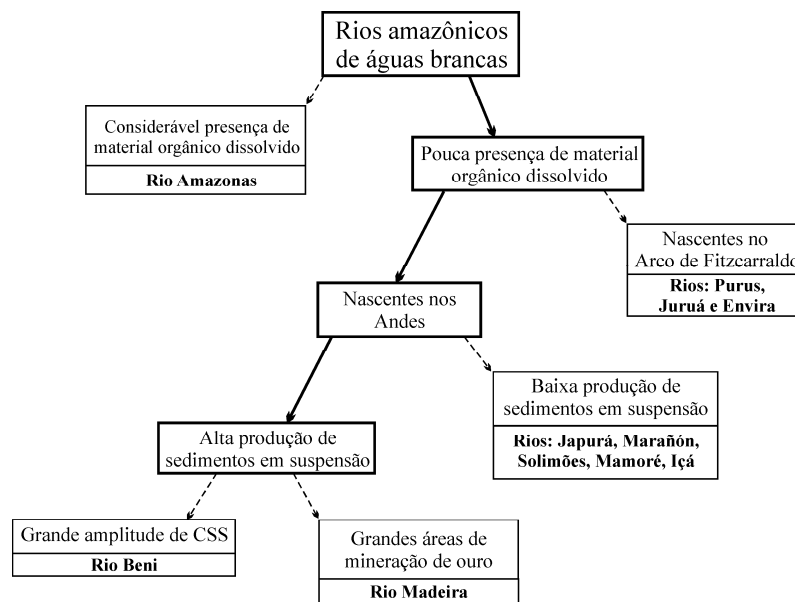


Figura 2. Critérios adotados para diferenciação das estações de coleta *in situ*.

Como a distribuição dos dados de CSS é extremamente desproporcional ao longo de toda a amplitude de CSS os valores foram normalizados por uma função logarítmica para o modelo global e para os dados da estação do Rio Beni, pois estes abrangem o maior intervalo de CSS (0 – 3561 mg/l). Para o restante dos modelos particionados uma transformação quadrática foi aplicada, pois este conjunto possui uma amplitude menor (0 – 1280 mg/l). Esta diferença nas transformações deve-se ao fato de que grandes transformações na variável resposta podem causar efeitos indesejados de alteração dos níveis de variância e distribuição dos erros do modelo (Kutner et al., 2005). Os modelos foram validados com uso da validação cruzada LOOCV (Leave-one-out cross validation). Esta validação fornece uma boa idéia da precisão dos modelos ao estimar valores de CSS em imagens sem dados *in situ*.

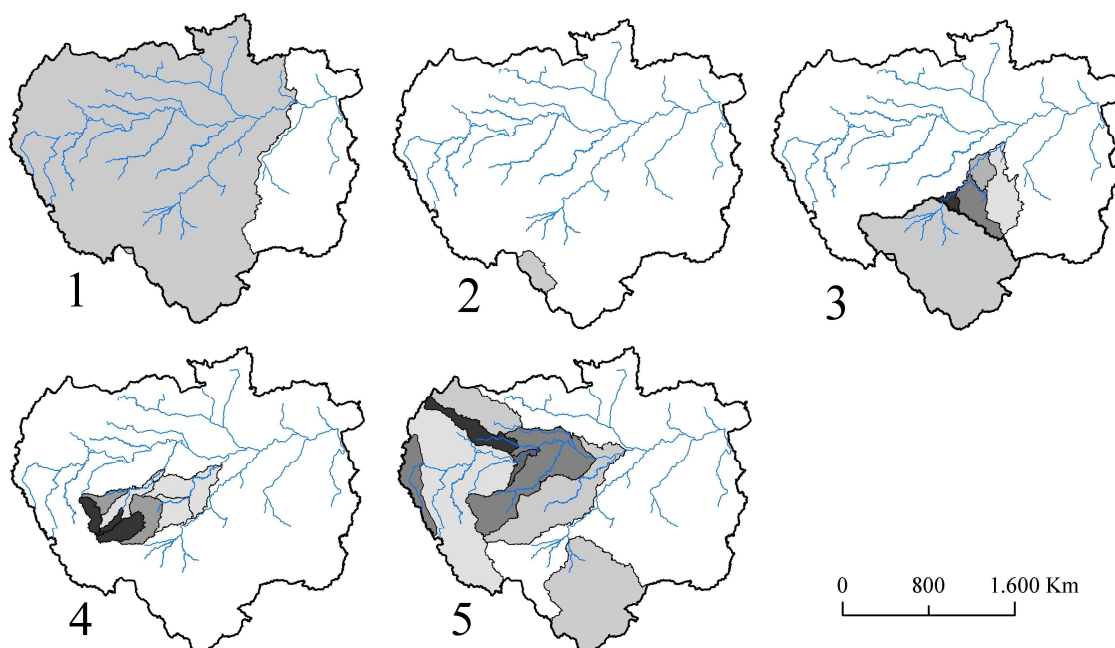


Figura 3. Bacias hidrográficas das estações de cada conjunto. 1 – Rio Amazonas; 2 – Rio Beni; 3 – Rio Madeira; 4 – Arco de Fitzcarraldo; 5 – Andes - Baixa produção.

Tabela 1. Métricas descritivas dos valores de CSS (mg/l) dos modelos regionais. μ = média; σ = desvio padrão.

Modelo	μ	σ	Mín. – Max.	Rios / n° de estações
Andes – baixa produção	136,85	111,13	0 – 728,4	Içá, Japurá, Marañón, Solimões, Mamoré / 8
Rio Madeira	230,17	205,79	0,39 – 881,02	Madeira / 5
Rio Beni	997,28	1103,11	28,3 – 3561,72	Beni / 1
Rio Amazonas	56,4	40,53	10,4 – 180,09	Amazonas / 1
Arco de Fitzcarraldo	197,07	207,21	7 – 1280,99	Purus, Juruá, Envira / 8

3. Resultados e Discussão

A Tabela 2 apresenta os resultados obtidos a partir da modelagem com todo o conjunto de dados, em todos os quatro níveis de correção.

Tabela 2. Estatísticas referentes aos modelos globais em cada nível de correção. Valor p de todos os modelos $\approx 1 \times 10^{-6}$. N = 504; ξ : μ = Erro médio; ξ : σ = desvio padrão.

Nível de correção	R ²	R ² LOOCV	ξ : μ (mg/l)	ξ : σ (mg/l)
$\rho_{apa(\lambda)}$	0,76	0,748	99,32	266,21
$\rho_{sup(\lambda)}$	0,693	0,65	117,49	357,87
$\rho_{ag_1(\lambda)}$	0,687	0,643	115,77	317,73
$\rho_{ag_2(\lambda)}$	0,7	0,658	113,16	313,85

Além destas estatísticas, os gráficos de dispersão para cada nível de correção também foram avaliados, os quais não mostram diferentes distribuições de erros, apenas mudanças nos níveis de variância. Dos quatro níveis de correção, a $\rho_{apa(\lambda)}$ forneceu as estimativas mais precisas. A validação LOOCV mostra que os modelos possuem uma resposta muito similar ao estimar valores de CSS para casos em que os dados não fazem parte do modelo. Porém, tanto a média quanto o desvio padrão dos erros são muito altos. A dispersão entre valores de CSS reais e estimados mostra que a modelagem global não tem boas respostas, principalmente para níveis de CSS acima de 500 mg/l (Figura 3).

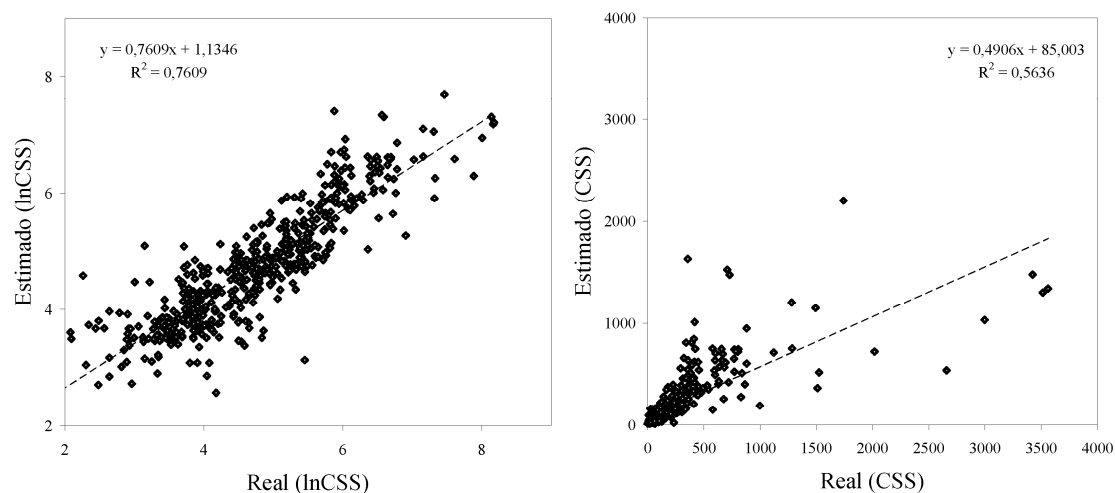


Figura 3. Dispersão entre CSS real e estimada pela modelagem global e com uso da $\rho_{apa(\lambda)}$.

Como a $\rho_{apa(\lambda)}$ forneceu as melhores estimativas, a modelagem regionalizada foi feita apenas com esta unidade. Os resultados da modelagem regional estão apresentados na Tabela 3. A dispersão entre valores reais e estimados de cada modelo está apresentada na Figura 4.

Tabela 3. Resultados dos modelos regionais. Valor $p \approx 1 \times 10^{-6}$; k = número de variáveis preditoras. $\xi: \mu$ = Erro médio; $\xi: \sigma$ = desvio padrão; $\xi: \%$ = erro relativo.

Modelo	R ²	R ² LOOCV	N	k	$\xi: \sigma$	$\xi: \mu$	$\xi: \%$
Andes – baixa produção	0,83	0,78	112	10	27,47	29,07	3,9
Rio Madeira	0,87	0,83	168	10	57,23	52,83	5,9
Rio Beni	0,91	0,79	34	6	286,22	247,41	7
Rio Amazonas	0,89	0,86	83	7	7,04	9,06	5,3
Arco de Fitzcarraldo	0,88	0,81	107	10	50,08	50,05	3,9

Apesar de que a heterocedasticidade foi satisfatoriamente corrigida nos modelos com uso das transformações (Figura 4 à esquerda), os gráficos de dispersão dos valores reais (Figura 4 à direita) mostram que à medida que aumentam os níveis de CSS há um aumento dos erros. Isto ocorre devido à natureza não linear da relação entre CSS e reflectância (Ritchie e Shiebe, 2000), de forma que quanto maiores os níveis, maior é o efeito da forma discreta de captação da radiação pelo sensor. Portanto, este é um problema inerente às limitações da forma de aquisição dos dados radiométricos, especialmente porque o sensor TM possui uma baixa resolução radiométrica. Apesar disso, podem ser feitas estimativas muito precisas, especialmente para níveis abaixo de 200 mg/l.

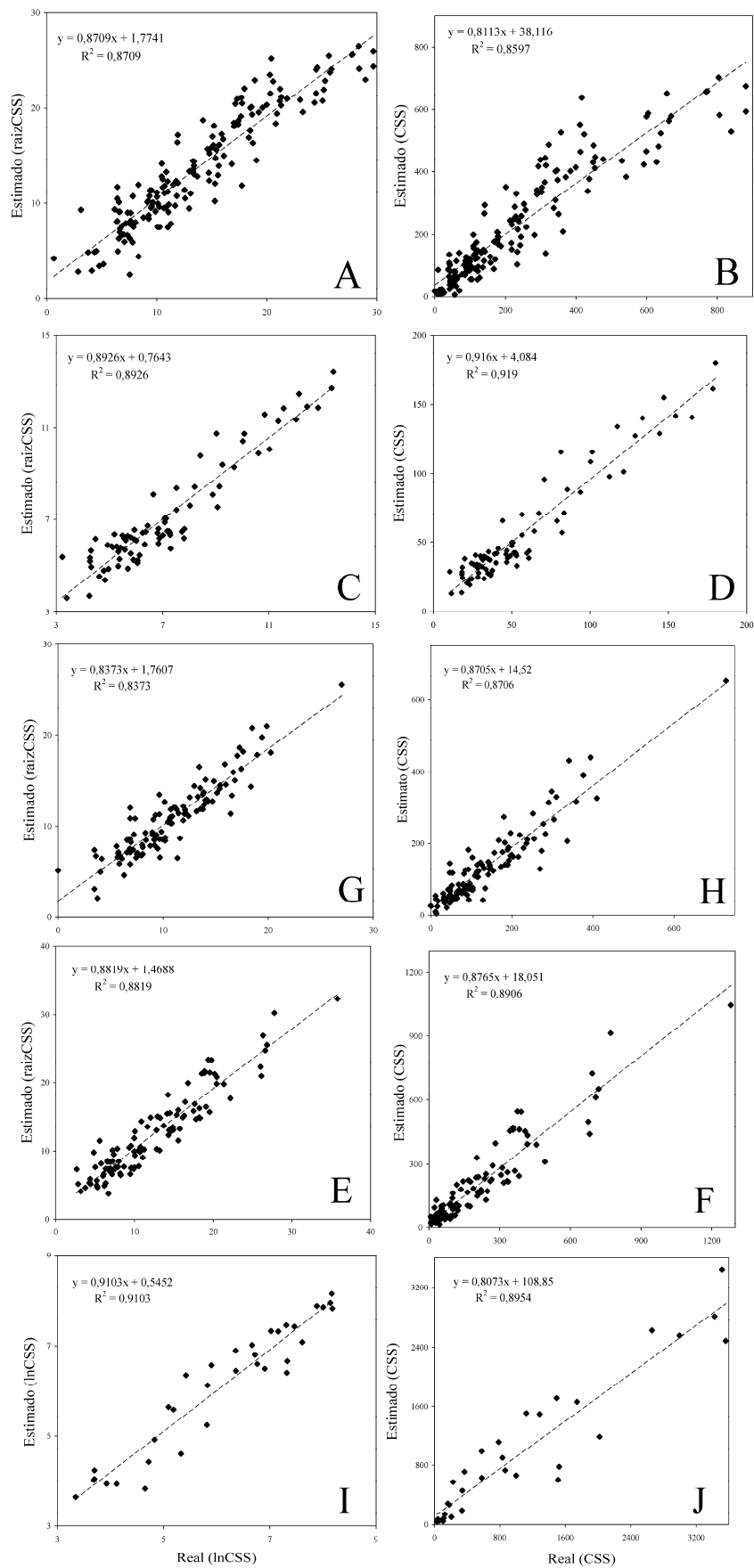


Figura 4. Dados reais e estimados para os modelos particionados: Rio Madeira (A e B), Rio Amazonas (C e D), Arco de Fitzcarraldo (E e F), Andes – baixa produção (G e H) e Rio Beni (I e J).

Os modelos estão apresentados a seguir:

Andes – baixa produção:

$$\sqrt{CSS} = 38,733 - 3,8706(B2) + 4,3951(B3) - 1,6578(B5) + 29,5835(B2/B1) - 38,8503(B4/B3) - 263,0106(B3/(B2+B1)) + 115,2593((B4+B3)/(B2+B1)) + 0,8221(\text{Estação} = \text{Ipiranga Velho, Vila Bittercourt, Itapeuá, Guarajá-Mirim, Marañón, São Paulo de Olivença, Teresina}) - 1,8823(\text{Estação} = \text{Guarajá-Mirim, Marañón, São Paulo de Olivença, Teresina}) + 3,1785(\text{Estação} = \text{Marañón, São Paulo de Olivença, Teresina})$$

Rio Madeira:

$$\sqrt{CSS} = -131,0751 - 6,6876(B1) + 2,1762(B2) - 2,1762(B5) + 187,9055(B3/B2) - 173,24(B3/B1) + 114,4793(B2/B1) + 45,2776(B4/B3) + 2,8114(\text{Estação} = \text{Faz. Vista Alegre, Manicoré, Porto Velho}) - 1,6418(\text{Estação} = \text{Manicoré, Porto Velho}) + 1,635(\text{Estação} = \text{Porto Velho})$$

Rio Beni:

$$\ln CSS = -50,0503 - 0,3135(B5) + 150,0787(B3/B2) + 57,7208(B2/B1) - 16,5347(B4/B3) - 355,5992(B3/(B2+B1)) + 44,3064((B4+B3)/(B2+B1))$$

Arco de Fitzcarraldo:

$$\sqrt{CSS} = -62,8656 - 1,9198(B2) + 2,9848(B3) - 1,3095(B5) + 165,8008(B3/B2) + 82,9254(B2/B1) - 402,8298(B3/(B2+B1)) + 15,5996((B4+B3)/(B2+B1)) + 2,202(\text{Conjunto} = \text{Médio Purus Juruá, Alto Purus Juruá}) + 6,787(\text{Rio} = \text{Envira}) + 1,114(\text{Estação} = \text{Gavião, Lábrea, Seringal Caridade, Cruzeiro do Sul, Feijó})$$

Rio Amazonas:

$$\sqrt{CSS} = 25,2644 - 2,6239(B2) + 1,8769(B3) + 1,7261(B4) - 0,9222(B5) + 27,9049(B3/B1) - 11,7893(B4/B3) - 81,2363(B3/(B2+B1))$$

Em que:

$$B1 = \rho_{apa} \text{ na banda TM1}$$

$$B2 = \rho_{apa} \text{ na banda TM2}$$

$$B3 = \rho_{apa} \text{ na banda TM3}$$

$$B4 = \rho_{apa} \text{ na banda TM4}$$

$$B5 = \rho_{apa} \text{ na banda TM5}$$

Apesar do uso da $\rho_{apa(\lambda)}$ nas estimativas, os modelos apresentam sistematicamente duas características que indicam que lidam com efeitos não relacionados com a CSS. A primeira é em relação à banda 5. Como nesta faixa espectral o volume d'água absorve praticamente toda radiação incidente, o sinal que retorna é proveniente da interação da radiação eletromagnética com a interface ar/água. Portanto, quanto maior é o sinal proveniente da interface (conhecido na literatura como *Sun glint*), maior é a interferência sobre a estimativa da CSS. A banda 5 foi significativa para todos os cinco modelos, possuindo coeficientes negativos, ou seja, atuam minimizando a interferência do sinal da interface.

A segunda característica remete à banda de menor comprimento de onda de cada modelo. Os efeitos de espalhamento atmosférico molecular e por aerossóis são mais intensos em comprimentos de onda mais curtos. Da mesma forma que o *Sun glint*, o espalhamento atmosférico também adiciona sinal à radiação emergente da água captada pelo sensor o qual não possui relação com a CSS. Desta forma, o coeficiente negativo para a banda de menor comprimento de onda de cada modelo pode indicar que estes também estão lidando com o espalhamento atmosférico, pois quanto maior é a reflectância nestas bandas (TM1 e TM2), menores são as estimativas da CSS. Isto não vale apenas para o modelo "Rio Beni", que possui apenas razões entre bandas.

Considera-se que tais padrões ocorrem devido ao grande número amostral dos modelos, de forma que o método de mínimos quadrados ajustou-se, empiricamente, aos efeitos não relacionados à CSS.

4. Conclusões

Os resultados permitem concluir que é possível a estimativa da CSS de rios amazônicos de águas brancas a partir de sensoriamento remoto, mesmo utilizando dados de baixa resolução radiométrica como os do sensor TM. Também notou-se que a série de correções aplicadas nos valores de reflectância não aumentaram a precisão das estimativas, de forma que a unidade mais simples é capaz de fornecer boas estimativas. O tratamento de todos os dados conjuntamente não produziu um modelo adequado, com erros muito altos, mas a regionalização das estações de coleta e a geração de cinco modelos independentes permitiram uma melhora nas estimativas. Por fim, apesar do caráter empírico dos modelos, os padrões encontrados de coeficientes negativos para as bandas de menor comprimento de onda e a banda 5 sugerem que estes estão lidando com efeitos não relacionados com o nível de CSS.

Agradecimentos

Os autores agradecem a Capes pela concessão de bolsa ao primeiro autor e à política de disponibilização gratuita das imagens Landsat do INPE.

Referências bibliográficas

Chander, G.; Markham, B.L. Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. **Remote Sensing of Environment**, v. 113, n. 5, p. 893–903, 2009.

Espinoza Villar, R.; Martinez, J.M.; Guyot, J.L.; Fraizy, P.; Armijos, E.; Crave, A.; Bazán, H.; Vauchel, P.; Lavado, W. The integration of field measurements and satellite observations to determine river solid loads in poorly monitored basins. **Journal of Hydrology**, n. 444-445, p. 221–228, 2012.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v. 11, n. 1, 2009.

Junk, W.; Piedade, M.T.F.; Schöngart, J.; Cohn-Haft, M.; Adeney, J.M.; Wittmann, F. A classification of major naturally-occurring Amazonian lowland wetlands. **Wetlands**, v. 31, n. 4, p. 623-640, 2011.

Kutner, M.H.; Nachtsheim, C.J.; Neter, J.; Li, W. **Applied Linear Statistical Models**. Boston: McGraw-Hill, 2005. 1398 p.

Martinez, J.M.; Guyot, J.L.; Filizola, N.; Sondag, F. Increase in suspended sediment discharge of the Amazon River assessed by monitoring network and satellite data. **Catena**, v. 79, n. 3, p. 257-264, 2009.

McClain, M.E. ; Naiman, R.J. Andean Influences on the Biogeochemistry and Ecology of the Amazon River. **BioScience**, v. 58, n. 4, p. 325–338, 2008.

Pierce, A.R.; King, S.L. Spatial dynamics of overbank sedimentation in floodplain systems. **Geomorphology**, v. 100, n. 4, p. 256-268, 2008.

Ritchie, J.C.; Shiebe, F.R. **Water Quality**. In: Shultz, G. A.; Engman, E. T. (Edi). **Remote Sensing in Hydrology and Water Management**. Springer, 2000.

Thornton, K.W. **Perspectives on reservoir limnology**. In K. W. Thornton, B. L. Kimmel, Payne, F. E. (Edi). **Reservoir limnology: ecological perspectives**. New York: Wiley, 1990. cap. 1, 1 – 13.

Vermote, E.F; Tanre, D.; Deuze, J.L.; Herman, M.; Morcrette, J.J. Second Simulation of the satellite signal in the solar spectrum, 6S: An overview. **IEEE Trans. Geosc. And Remote Sens.**, v. 35, n. 3, p. 675-686, 1997.

Wetzel, R.G. **Limnology: Lake and River Ecosystems**. San Diego: Academic Press, 2001. 1006 p.