

Áreas agrícolas em sensores com resolução espacial de 30 m estimadas a partir de dados MODIS e métricas da paisagem

Dênis de Moura Soares^{1,2}
Antônio Roberto Formaggio¹
Lênio Soares Galvão¹

¹Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{denis, formag, lenio}@dsr.inpe.br

²Diretoria de Serviço Geográfico - DSG
QGEx - Bloco F - 2º Andar – SMU - CEP 70630-901 – Brasília - DF, Brasil
dsg@citex.eb.mil.br, dsge@eb.mil.br

Abstract. The objective of this work was to evaluate the differences between crop area estimation from coarse resolution data (e.g. MODIS/Terra, with 250m) and ETM+/Landsat-7 resolution data (30m), considering different crop types and their spatial pattern quantified by landscape metrics. The analysis was applied for three different crops: corn, sugarcane and soybean. The thematic classes woodland, pasture and exposed soils were also included in the analysis. Simple (area) and multiple (area plus landscape metrics) regression models were performed using ETM+ and MODIS data. One global and three single statistical models were developed. The global approach (the three crops, simple regression) produced a coefficient of determination (R^2) of 0.46. On the other hand, the developing of statistic models for each crop (landscape metrics, multiple regression) improved the R^2 value to 0.52, 0.67 and 0.87 for corn, sugarcane and soybean, respectively. Results showed that accurate crop estimation area from coarse resolution data is much more difficult for corn than for sugarcane and soybean because of the high fragmentation of corn distribution in the study area.

Palavras-chave: coarse resolution, spatial pattern, landscape metric, regression analysis, resolução moderada, padrão espacial, métrica da paisagem, análise de regressão.

1. Introdução

O uso de técnicas de sensoriamento remoto apresenta-se como alternativa para tornar mais objetiva a estimativa de áreas agrícolas, que, no Brasil, é feita pelo Instituto Brasileiro de Geografia e Estatística (IBGE) através de preenchimento de entrevistas junto aos setores envolvidos no processo agrícola, pelo método conhecido como “subjutivo”.

As imagens do sensor ETM+/Landsat-7 ou do CCD/CBERS-2 são de resolução espacial intermediária (30 e 20 m, respectivamente), permitindo, assim, uma escala de representação compatível com os alvos agrícolas de interesse. No entanto, a resolução temporal baixa (16 dias para o ETM+ e 26 dias para o CCD) dificulta sobremaneira a obtenção de imagens livres de nuvens, principalmente na principal fase do ciclo agrícola (novembro a março).

Por outro lado, existem atualmente sensores como o *Moderate Resolution Imaging Spectroradiometer* (MODIS), presente tanto na plataforma Terra como na Aqua (NASA, 2006), que pode apresentar uma resolução temporal quase diária, sendo, assim, ideal para melhorar as possibilidades de adquirir dados livres de cobertura de nuvens. A resolução espacial das bandas 1 e 2 do MODIS é de 250m e, assim, estudos ainda são necessários para avaliação da habilidade de sensores dotados de resolução espacial moderada em detectar os elementos agrícolas de interesse, como área.

Nesse sentido, Malingreau e Belward (1992), Mayaux e Lambin (1995), Moody e Woodcock (1995), Mayaux e Lambin (1997), Moody (1998), Ponzoni et al. (2002), Millington et al. (2003), Frohn e Hao (2006) estudaram o tema para áreas florestais,

verificando o impacto da resolução espacial. Pax-Lenney e Woodcock (1997), por sua vez, fizeram uma investigação em áreas agrícolas do Egito.

É nesse contexto que se insere o presente trabalho. A representação de alvos agrícolas em imagens produzidas por sensores com baixa resolução espacial pode estar condicionada ao padrão espacial das culturas, sendo o efeito de agregação espacial um aspecto relevante na estimativa das proporções dos tipos de cobertura do solo (Moody e Woodcock, 1994; Mayaux e Lambin, 1995).

A determinação da conformação espacial dos elementos pode ser feita por meio do cálculo de métricas da paisagem, que são grandezas que quantificam a composição e configuração espacial do cenário, para um mapa ou imagem de sensoriamento remoto (Frohn e Hao, 2006; Frohn, 1998; Saura, 2002, 2004; Shen et al., 2004; Wu, 2004; Wu et al., 2000, 2002). No âmbito deste trabalho foram estudadas 20 métricas da paisagem cujas principais aparecerão no decorrer dos resultados da pesquisa.

Desta forma, o objetivo deste trabalho foi avaliar as diferenças obtidas nas estimativas de áreas de culturas agrícolas quando são utilizados sensores de resolução espacial moderada (p.ex, MODIS/Terra, com 250m), ao invés de resolução fina (ETM+/Landsat-7, com 30m). As culturas abordadas foram soja, cana-de-açúcar e milho, que são produtos de interesse nacional e apresentam padrões de distribuições espaciais distintos. Além dessas culturas, foram ainda consideradas na análise as classes temáticas mata, pastagem e solo exposto, para compor um modelo estatístico mais abrangente e facilitar a compreensão geral do comportamento das métricas da paisagem. A região correspondente aos municípios de Ipuã, Guará e São Joaquim da Barra (coordenadas W 47° S 22°), norte paulista, constituiu-se na área estudo. O padrão das parcelas agrícolas nesta área é considerado representativo das condições encontradas, predominantemente, na agricultura paulista e do centro-sul brasileiro (Luiz, 2003; Sanches, 2004).

Para atender ao objetivo supracitado, usando imagens reflectância de superfície ETM+/Landsat-7 (cena 220/74) e MODIS/Terra obtidas na mesma data (05 de Janeiro de 2002), buscou-se atingir os seguintes objetivos específicos:

- verificar as relações de regressão simples entre as estimativas de áreas obtidas com os dados de resolução espacial de 30m (ETM+/Landsat-7) e 250m (MODIS/Terra), para as seis classes temáticas (soja, cana-de-açúcar, milho, mata, pastagem e solo exposto) estudadas em conjunto e separadamente;
- construir um modelo de regressão múltipla que incorpore informação sobre o arranjo espacial das classes temáticas para melhorar a precisão da estimativa de suas áreas com dados de baixa resolução espacial (MODIS/Terra, 250m), selecionando as métricas de paisagem mais adequadas para tal finalidade.
- Traçar um paralelo entre o desempenho de modelos específicos (para cada classe, separadamente) e a performance de modelos gerais (abrangendo todas as classes).

2. Metodologia

Uma classificação digital não-supervisionada k-médias com 15 classes e 20 iterações, conforme proposto por Frohn e Hao (2006), foi aplicada sobre as bandas 3 e 4 do sensor ETM+/Landsat-7 (cena 220/74, de 05 de janeiro de 2002) e bandas 1 e 2 (250m), da mesma data, do MODIS. Tais produtos foram a base para confecção de um modelo estatístico capaz de, a partir da imagem de resolução espacial de 250m, predizer a área de classes temáticas extraídas com o uso de um sensor de 30m.

Os modelos de regressão simples e múltipla, descritos a seguir, foram precedidos por uma análise exploratória dos dados, envolvendo a análise da normalidade da variável dependente

(p.ex., teste de normalidade Shapiro-Wilk) e a necessidade de eventuais transformações matemáticas das variáveis preditivas e da dependente.

2.1. Modelo de Regressão Simples ($\text{ÁREA}_{\text{ETM+}}$ versus $\text{ÁREA}_{\text{MODIS}}$)

Esta operação teve como objetivo avaliar o desempenho de um modelo de predição de área baseado somente em medidas de área, sem considerar, portanto, o padrão de distribuição espacial das diferentes culturas. A idéia final foi comparar o desempenho deste modelo com o de regressão múltipla, verificando o ganho da incorporação do padrão espacial dos elementos na predição de áreas de sensores de resolução espacial fina a partir de dados de resolução espacial moderada.

Foram gerados 216 subconjuntos (blocos) de dimensão 125×125 pixels para a imagem ETM+, que na imagem MODIS corresponde à dimensão 15×15 pixels. Assim, com o recorte das imagens classificadas, originou-se um total de 1296 observações, uma vez que foram consideradas seis classes: cana-de-açúcar, soja, milho, solo exposto, mata e pastagem. Do quantitativo total, foram separadas 240 observações (40 para cada classe) para validação da equação de regressão simples.

Como forma de análise inicial dos dados, as áreas das classes extraídas entraram, então, num procedimento de regressão, para determinar a relação existente entre as medidas obtidas nas imagens de resolução espacial fina (ETM+/Landsat-7) e moderada (MODIS/Terra).

O coeficiente de explicação obtido da regressão simples foi um parâmetro de comparação com o modelo de regressão múltipla descrito a seguir.

2.2 Modelo de Regressão Múltipla com incorporação de Métricas de Paisagem ($\text{ÁREA}_{\text{ETM+}}$ versus $\text{ÁREA}_{\text{MODIS}}$ + Métricas)

A importância da teoria da regressão linear está na possibilidade de sua aplicação em diversas áreas de conhecimento. Moody e Woodcock (1995) afirmam que a análise de regressão pode ser aplicada para investigar a influência do padrão espacial de culturas na estimativa de área com sensores de resolução espacial moderada. Para tanto, faz-se necessária a criação de um modelo que relacione a estimativa real (variável dependente), com os diversos padrões que se julguem necessários, além da estimativa extraída da imagem estudada (variáveis independentes).

Dessa forma, foi desenvolvido um modelo de regressão múltipla para estimativa da área de classes temáticas obtidas com sensor de 30m de resolução espacial (ETM+/Landsat-7), a partir de dados de área medidos com imagem de 250m (MODIS/Terra), considerando-se o uso potencial das métricas de paisagem, que denotam padrão de distribuição espacial das culturas agrícolas na cena MODIS. A **Tabela 1** apresenta as principais métricas envolvidas nos estudos.

Tabela 1 – Principais métricas da paisagem estudadas.

Área	Forma de Classe (LSI)	Dimensão Fractal (PAFRAC)	Interposição e Justaposição (IJI)	Trama (MESH)
Número de Polígonos (NP)	Forma de Classe Normalizada (NLSI)	Número de Fragmentos Nucleares (NDCA)	Porcentagem de Adjacência (PLADJ)	Fragmentação (SPLIT)
Maior Fragmento (LPI)	Área Core Total (TCA)	Coesão (COHE)	Divisão (DIV)	Agrupamento (Clump)

Semelhante ao procedido para a regressão simples foram inseridas na modelagem 1056 observações, deixando-se separadas as 240 restantes para validação do modelo. Critérios estatísticos, tais como C_p , R_p^2 , R_a^2 , foram utilizados para seleção das melhores variáveis para compor o modelo final.

2.3. Comparação dos Modelos de Regressão Simples e Múltipla: Abordagem Geral (Todas as Culturas) e Abordagem Específica (por Cultura)

A elaboração de modelos de regressão simples e múltipla teve por objetivo avaliar o benefício causado pela adição do padrão espacial, para estimativa de área de classes temáticas com sensores de resolução moderada.

Adicionalmente, foi feita uma análise comparativa entre modelos estatísticos multivariados estabelecidos, especificamente, para cada cultura agrícola e aquele obtido para classes temáticas genéricas. Tal procedimento teve por fim a discussão sobre o procedimento mais adequado para se obter área de culturas agrícolas com sensores de resolução moderada, uma vez que se perceberam comportamentos característicos para o conjunto de dados de cada classe.

3. Resultados e Discussão

3.1. Constituição do Modelo de Regressão Simples

A Tabela 2 apresenta os parâmetros da regressão simples entre a raiz quadrada da área medida na imagem do sensor ETM+ e a área obtida pelo MODIS. O cálculo da raiz da métrica original (área ETM+) teve por objetivo normalizar a variável dependente, o que causou efeitos no mesmo sentido para os resíduos. Não houve necessidade, entretanto, de transformações na área obtida pelo MODIS (variável independente).

A regressão simples apresentou um coeficiente de determinação (R^2) de 0,46, com 95% de confiança. Sua ponderação foi necessária tendo em vista a inconstância da variância dos resíduos da regressão original, o que acarreta volatilidade na estimativa dos parâmetros. Foi diagnosticada, ainda, a inexistência de *outliers* influentes nos resultados, bem como a normalidade dos resíduos. A fase de validação mostrou a adequação dos parâmetros apresentados na Tabela 2.

Tabela 2 – Parâmetros da regressão simples ponderada.

	B	Desv. Pad.	p-valor
Intercepto	8,8649	0,0698	0,0000
Área	0,0228	0,0002	0,0000

3.2. Constituição do Modelo de Regressão Múltipla

A Tabela 3 apresenta o resultado da modelagem estatística confeccionada com a inclusão das métricas de paisagem. É notável o fato de os três critérios estatísticos para seleção de variáveis (C_p , R_p^2 , R_a^2) resultarem em um único modelo, o que traz indício de sua adequação. Novamente houve necessidade de ponderação da regressão.

Tabela 3 – Parâmetros da regressão múltipla ponderada.

	B	Desv. Pad.	p-valor
Intercepto	8,8858	0,2369	0,0000
Área	0,0275	0,0007	0,0000
raiz LPI	-0,6813	0,0841	0,0000
LSI	-0,9707	0,0808	0,0000
TCA	-0,0467	0,0041	0,0000
PLADJ^2	0,0007	0,0001	0,0000
IJI	0,0348	0,0029	0,0000

Observou-se um R^2 de 0,49 para a regressão múltipla (5% de significância), cuja ponderação também foi necessária. Notem-se as transformações aplicadas para as variáveis

PLADJ e LPI. Adicionalmente, observou-se a inexistência de *outliers* influentes e a normalidade distribuição dos resíduos. Na validação, os parâmetros estabelecidos foram aceitos.

3.3. Análise Comparativa entre a Abordagem Geral versus a Específica Utilizando Modelos de Regressão Simples e Múltipla

Com a abordagem geral (todas as classes consideradas em conjunto), a regressão simples apresentou um coeficiente de determinação (R^2) de 0,46, enquanto que a regressão múltipla mostrou um valor de R^2 de 0,49. Dessa forma, a inclusão dos índices que denotam o padrão espacial das classes temáticas aumentou o potencial de explicação, pelo sensor MODIS, da área obtida pela imagem ETM+.

Este incremento de cerca de 7%, apesar de não ser um valor tão alto, fornece um indício de que, pela inclusão da informação de padrão espacial das classes temáticas, há possibilidade de melhoria no processo de estimativa de área com sensores de resolução espacial moderada.

Assim, em uma modelagem exploratória paralela, para melhorar a visualização do efeito que se pretende mostrar, foram processadas apenas as culturas agrícolas, cujo resultado é apresentado na **Figura 1**. Nota-se que cada cultura apresentou-se agrupada de maneira característica (**Figura 1a**), o que motivou a confecção das **Figuras 1b, 1c e 1d**, caracterizando os modelos específicos.

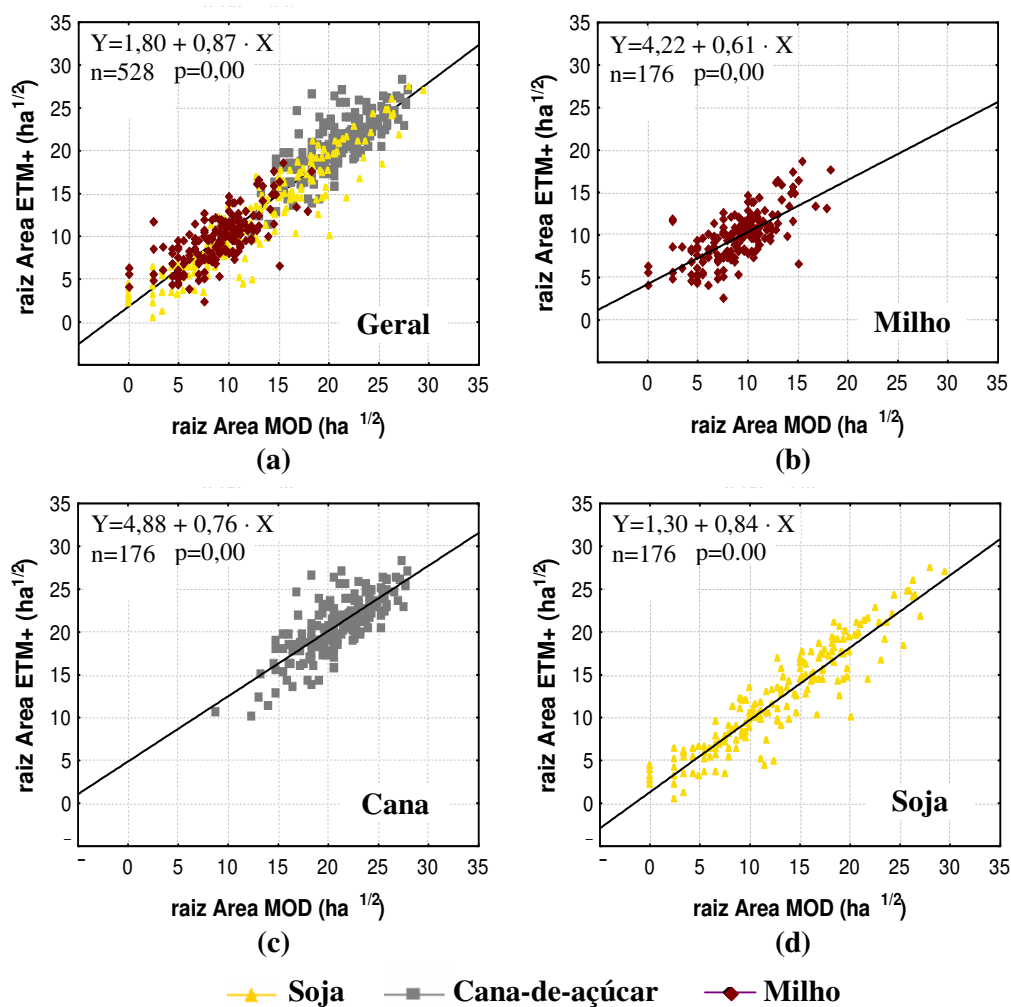


Figura 1 – Regressão simples entre áreas estipuladas pelos sensores ETM+ e MODIS: (a) Modelo Geral; Modelos específicos para (b) Milho; (c) Cana-de-açúcar; (d) Soja.

É possível observar que o milho (**Figura 1b**) é composto por pequenas áreas, quando comparado ao modelo geral. Ainda seguindo esse paralelo, a inclinação da reta de regressão é bastante inferior, donde se conclui que o milho é superestimado em maior escala, o que está associado à sua alta fragmentação.

Para a cana (**Figura 1c**), observou-se o predomínio de maiores valores de área, sendo sua superestimação inferior àquela observada para o milho, no entanto mais severa do que o percebido para o modelo geral. Finalmente, a soja (**Figura 1d**) foi a cultura cujo modelo mais se aproximou da modelagem geral, uma vez que a inclinação e o intercepto das retas de regressão foram muito próximos. É nítida a existência de observações de tamanhos variados, englobando todo o domínio estudado.

Assim, evidenciada a habilidade de estudos específicos, foram estabelecidos modelos de regressão simples para cada classe agrícola, cujo resultado é observado na **Tabela 4**. Os modelos específicos de regressão múltipla, considerando o arranjo espacial das classes agrícolas, são apresentados nas **Tabelas 5, 6 e 7**.

As inclinações e interceptos das retas de regressões simples são bastante diferentes (**Tabela 4**), o que confirma o comportamento particular de cada cultura. Na mesma linha, observa-se que para cada cultura houve uma seleção de variáveis bastante distintas, gerando modelos muito diferentes.

Tabela 4 – Parâmetros das regressões simples específicas.

Cultura	Soja		Cana		Milho	
	B	Desv. Pad.	B	Desv. Pad.	B	Desv. Pad.
Intercepto	6,1425	0,3006	12,5811	0,5670	6,6294	0,2802
Área	0,0289	0,0010	0,0183	0,0012	0,0337	0,0026

Tabela 5 – Parâmetros da regressão múltipla específica para o milho.

Milho	B	Desv. Pad.	p-valor
Intercepto	4,4179	0,7084	0,0000
NP	0,2555	0,0754	0,0009
raiz LPI	-3,1221	1,5268	0,0424
CLUMP	0,0255	0,0092	0,0062
IJI	-0,0254	0,0097	0,0096
raiz 4 MESH	9,4358	2,6824	0,0006

Tabela 6 – Parâmetros da regressão múltipla específica para a soja.

Soja	B	Desv. Pad.	p-valor
Intercepto	3,3741E+00	4,3846E-01	0,0000
Área	1,9741E-02	2,8739E-03	0,0000
NP	7,3158E-01	1,0536E-01	0,0000
TCA	-1,8665E+03	9,0471E+02	0,0406
DCAD	2,6122E+04	1,2665E+04	0,0407
COHE^4	7,1136E-08	1,8484E-08	0,0002

Tabela 7 – Parâmetros da regressão múltipla específica para a cana-de-açúcar.

Cana	B	Desv. Pad.	p-valor
Intercepto	-38,8831	15,83294	0,015062
Área	0,0337	0,00619	0,000000
LSI	3,9220	0,87243	0,000013
CLUMP	43,4461	12,96585	0,000993
IJI	-0,0748	0,01637	0,000009
NLSI	44,5468	15,27085	0,004010

A **Figura 2** ratifica as diferenças apontadas para cada modelagem, mostrando o desempenho de cada modelo, quanto ao quesito coeficiente de determinação. Pode-se perceber que sempre a modelagem específica é mais eficiente, inclusive para a cultura agrícola mais fragmentada, que é o milho. Para a classe de distribuição espacial mais compacta, que é a soja, o incremento chegou a cerca de 80%, para regressão múltipla, utilizando a abordagem específica, quando comparada à regressão múltipla com abordagem geral. Quando se confronta a regressão múltipla específica com a regressão simples geral obtém-se aumento no R^2 de 13%, 90%, 46% para o milho, a soja e a cana, respectivamente.

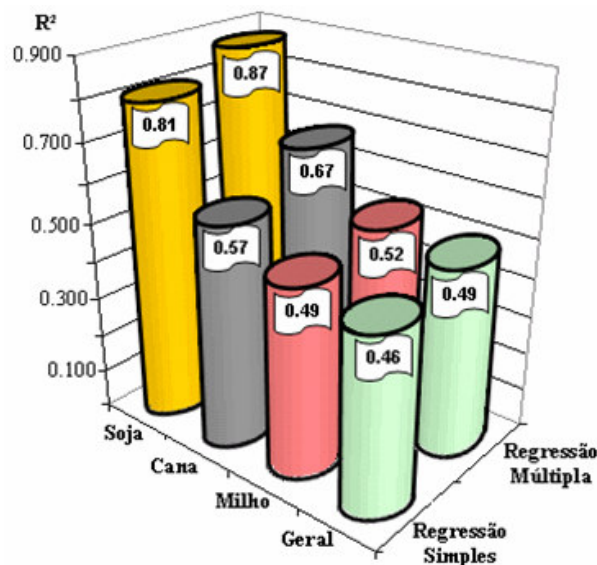


Figura 2 – Variações no coeficiente de determinação (R^2) para os modelos de regressão simples e múltipla, geral e específico.

4. Conclusões

Os resultados obtidos neste trabalho demonstraram que sensores de resolução espacial moderada (p. ex., MODIS/Terra; 250m) podem ser utilizados para prever áreas com precisões semelhantes às possíveis com o uso de sensores de resolução mais fina (ETM+/Landsat-7; 30m), para culturas agrícolas que não apresentem fragmentação acentuada, como soja e cana-de-açúcar, como nas condições encontradas na área de estudo. Essa habilidade potencializa-se com a adoção de modelos estatísticos específicos para as diferentes culturas agrícolas e que incorporem métricas de paisagem. Estatisticamente, os coeficientes de determinação (R^2) obtidos de 0,52, 0,67 e 0,87 para o milho, a cana e a soja, respectivamente, com a utilização de regressão múltipla e de índices espaciais, superaram o valor de 0,49 atingido pela abordagem geral equivalente. A fragmentação acentuada da distribuição das parcelas de milho na área de estudo é um obstáculo para o uso de sensores de resolução moderada na estimativa da área vista por instrumentos com melhor resolução espacial.

Agradecimentos

À Ieda Del'Arco Sanches e ao Alfredo José Barreto Luiz por permitirem acesso irrestrito a suas bases de dados.

Referências

- Frohn, R. **Remote sensing for landscape ecology: New metric indicators for monitoring, modeling, and assessment of ecosystems.** Boca Raton, FL: Lewis, 1998. 99 p.
- Frohn, R. C.; Hao, Y. Landscape metric performance in analyzing two decades of deforestation in the Amazon Basin of Rondonia, Brazil. **Remote Sensing of Environment**, v. 100, n. 2, p. 237-251, Jan.2006.
- Luiz, A. B. Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto. dez. 2003. (INPE--TDI). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003.
- Malingreau, J. P.; Belward, A. S. Scale considerations in vegetation monitoring using AVHRR data. **International Journal of Remote Sensing**, v. 13, n.12, p.2289-2307, Aug.1992.
- Mayaux, P.; Lambin E. F. Estimation of tropical forest area from coarse spatial resolution data: a two-step correction function for proportional errors due to spatial aggregation. **Remote Sensing of Environment**, v. 53, n. 1, p. 1-15, July 1995.
- Mayaux, P.; Lambin E. F. Tropical forest area measured from global land-cover classifications: inverse calibration models based on spatial textures. **Remote Sensing of Environment**, v. 59, n. 1, p. 29-43, Jan. 1997.
- Millington, A. C.; Velez-Liendo, X. M.; Bradley, A. V. Scale dependence in multitemporal mapping of forest fragmentation in Bolivia: implications for explaining temporal trends in landscape ecology and applications to biodiversity conservation. **ISPRS International Journal of Photogrammetry and Remote Sensing**, v. 57, n. 4, p. 289-299, Feb. 2003.
- Moody, A. Using landscape spatial relationships to improve estimates of land-cover area from coarse resolution remote sensing. **Remote Sensing of Environment**, v. 64, n. 2, p. 202-220, May 1998.
- Moody, A.; Woodcock, C. E. Scale-dependent errors in the estimation of land-cover proportions – implications for global land-cover datasets. **Photogrammetric Engineering and Remote Sensing**, v. 60, n. 5, p. 585-594, May 1994.
- Moody, A.; Woodcock, C. E. The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. **Landscape Ecology**, v.10, n. 6, p. 363-379, Dec. 1995.
- National Aeronautics and Space Administration (NASA). **MODIS Web**, 2006. Disponível em: <<http://modis.gsfc.nasa.gov>>. Acesso em: 13 out. 2006.
- Pax-Lenney, M.; Woodcock, C. E. The effect of spatial resolution on the ability to monitor the status of agricultural lands. **Remote Sensing of Environment**, v. 61, n. 2, p. 210-220, Aug.1997.
- Ponzoni, F. J.; Galvão, L. S.; Epiphanyo, J. C. N. Spatial resolution influence on the identification of land cover classes in the Amazon environment. **Anais da Academia Brasileira de Ciências**, v. 74, n. 4, p. 717-725, Dec. 2002.
- Sanches, I. D. Sensoriamento remoto para o levantamento espectro-temporal e estimativa de área de culturas agrícolas. 2004. 172 p. (INPE-10290-TDI/909). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004.
- Saura, S. Effects of minimum mapping unit on land cover data spatial configuration and composition. **International Journal of Remote Sensing**, v. 23, n. 22, p. 4853– 4880, Nov. 2002.
- Saura, S. Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. **Landscape Ecology**, v.19, n. 2, p. 197-209, Feb. 2004.
- Shen, W., Jenerette, G. D., Wu, J., & Gardner, R. H. Evaluating empirical scaling relations of pattern metrics with simulated landscapes. **Ecography**, v.27, n. 4, p. 459– 469, Aug. 2004.
- Wu, J.; Shen, W.; Sun W.; Tueller, P. T. Empirical patterns of the effects of changing scale on landscape metrics. **Landscape Ecology**, v.17, n. 8, p. 761-782, Nov. 2002.
- Wu, J. Effects of changing scale on landscape pattern analysis: Scaling relations. **Landscape Ecology**, v. 19, n. 2, p. 125– 138, Mar. 2004.
- Wu, J., Jelinski, D. E., Luck, M., & Tueller, P. Multiscale analysis of landscape heterogeneity. *Geographic Information Sciences*. v. 6, n. 1, p. 6– 19, Jun. 2000.