

Auditoria de dados de campo em estimativa de áreas agrícolas com o auxílio de análise multitemporal de dados de satélite

Natalia de Almeida Crusco¹
José Carlos Neves Epiphanyo¹

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

Abstract. Crop area estimates are important for any program of crop production monitoring. There is a program under development at government level that includes remote sensing techniques in the crop area evaluation at state level. This system is based on statistical sampling, remote sensing data, and field evaluation. This paper describes a methodology of auditing the field work using multitemporal remote sensing analysis. The results show that with four years of multitemporal remote sensing data and a decision rule based on agriculture knowledge it is possible to have very accurate hints about each ground sampling land use and which needs further auditing visits. Additional aspects concerning land cover changes along the period under analysis are discussed.

Palavras chaves: sensoriamento remoto, previsão de safras, agricultura, análise espectral

1. Introdução

Nas últimas décadas, um avanço considerável foi observado na área de sensoriamento remoto agrícola. A importância da utilização de imagens de satélite nas estatísticas agrícolas já foi evidenciada por programas como MARS (Monitoring Agriculture with Remote Sensing) e pelo Serviço Nacional de Estatística Agrícola dos Estados Unidos (NASS).

No Brasil, o IBGE (Instituto Brasileiro de Geografia e Estatística) é responsável pela realização das estatísticas agrícolas, incluindo nesse procedimento a estimativa de área agrícola. Os dados gerados pelo Levantamento Sistemático da Produção Agrícola (LSPA) têm periodicidade mensal, e o método baseia-se em entrevistas aplicadas a pessoas ligadas ao meio agrícola. Este método apresenta um caráter subjetivo e não permite avaliações estatísticas nos dados.

Dada a importância do desenvolvimento de metodologias objetivas na estimativa de áreas agrícolas no Brasil, o projeto Geosafras, iniciado na safra 2003/2004, utiliza imagens de satélite como base para a definição de painéis amostrais. A partir destas imagens são geradas *imagetes* que auxiliam no trabalho de campo juntamente com o GPS. O método de seleção dos pontos amostrais foi desenvolvido por Epiphanyo et al. (2002), por meio da Amostragem Aleatória Simples (AAS), e tem como um dos princípios o fato de qualquer indivíduo n da população N , tem a mesma probabilidade de ser escolhido para fazer parte da amostra.

A coleta de dados é realizada em alguns municípios, estratificados no Estado federativo de acordo com sua área cultivada no ano anterior, para posteriormente calcular a área da cultura em questão no ano vigente. Em cada município participante do painel amostral, 100 pontos amostrais são distribuídos aleatoriamente. Após a coleta de dados em campo, cada ponto é classificado e a área cultivada é calculada por expansão direta (Luiz e Epiphanyo, 2001).

Porém, no decorrer do desenvolvimento do projeto, uma questão fundamental foi observada, relacionada à validação dos dados provenientes do campo. A estimativa de área calculada por expansão é totalmente baseada nestes dados, que podem apresentar alguns tipos de erro. Entre eles estão os erros de preenchimento por parte dos técnicos de campo, que podem não chegar ao

ponto determinado pelo sistema de amostragem, quando, por exemplo, a localização do ponto amostral estiver no limite entre dois talhões, ou em locais de difícil acesso.

A existência da relação entre os dados coletados em campo no presente e informações do uso do solo em tempos pregressos foi observada por diversos autores (Ortiz et al. 1997; Nieuwenhuis e Mícher, 1998; e Uchida, 2001). A análise do histórico de áreas com atividades agrícolas, juntamente com técnicas de mapeamento de dados multitemporais podem auxiliar na verificação da ocupação atual do solo e prever usos futuros de acordo com o histórico de ocupação da área.

Assim, o objetivo principal deste estudo é avaliar como a dinâmica de áreas agrícolas, analisada por meio de uma abordagem multitemporal, pode auxiliar o processo de previsão, auditoria e validação de dados coletados em campo.

2. Métodos

Área de estudo

A área de estudo é o município de Tuneiras do Oeste, localizado na região noroeste do Estado do Paraná (**Figura 1**). O município está entre as coordenadas geográficas: 52°41' e 53°01' de longitude oeste e 23°41' e 24°01' de latitude sul e possui uma área de 69.900 ha.

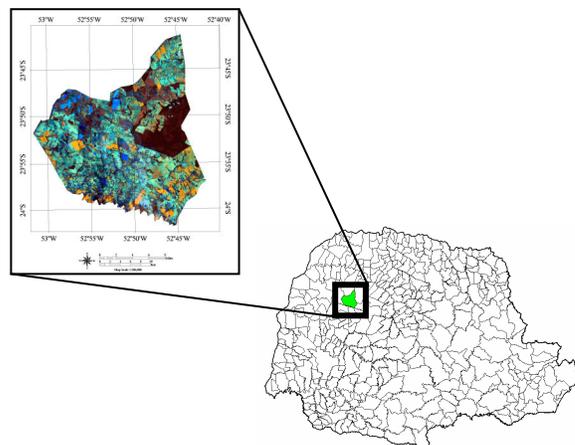


Figura 1 - Localização da área de estudo no Estado do Paraná.

O total de área plantada corresponde a 41% da área total do município. Nesta porcentagem não está incluída a porção de terras destinadas ao pasto, que compreendem mais 36% da área. Ou seja, cerca de 77% da área de estudo é dedicada ao uso agrícola e às pastagens cultivadas ou naturais. As lavouras temporárias (anuais e semi-perenes) são responsáveis pela maior área cultivada (99,13%) no município. A cultura da soja é responsável por mais de 50% do total de área plantada em todo o município, seguida pela cana-de-açúcar, que possui 7.481 ha de área plantada (IBGE, 2006).

Pré-processamento

Para a geração dos perfis espectro-temporais foram utilizadas 24 imagens do satélite Landsat-5 e 7, sendo sete obtidas pelo sensor ETM+ e 17 pelo sensor TM. Essas imagens compreendem aos anos de 2002, 2003, 2004 e 2005 para a cena 223/77.

Todas as imagens adquiridas foram registradas, com a transformação RST (rotation, scaling and translation) e aplicação do algoritmo de interpolação vizinho mais próximo. Após o registro, as imagens foram importadas, em formato GeoTIFF, para o banco de dados presente no SPRING,

versão 4.1. A correção atmosférica, baseada na metodologia de subtração do pixel escuro (SPE), desenvolvida por Chaves (1988), foi aplicada somente a uma imagem do banco de dados - a imagem referência (Silva et al., 2003). Após a correção da imagem referência, as outras imagens do banco de dados foram uniformizadas pelo processo de regressão linear proposto por Gürtler (2003), com o uso de cinco alvos claros e seis escuros.

Feito o procedimento de uniformização, as imagens foram transformadas de nível digital para valores de reflectância. Esta transformação se faz necessária quando se estuda o comportamento espectral de alvos por uma análise multitemporal.

Análise do perfil espectro-temporal

O projeto Geosafra apresenta um caráter amostral, em que cada ponto (amostra) é representado na imagem por um pixel. Neste trabalho, optou-se como unidade básica de estudo o segmento envolvido pelo ponto amostral, que na maioria das vezes é representado por um talhão ou um fragmento florestal. A escolha de estudar segmentos está baseada no fato de que o erro obtido no registro teria menos impacto sobre segmentos, do que sobre pixels isolados. Após alguns testes de variação dos limiares de similaridade e área do algoritmo de segmentação, optou-se pelos valores 8 níveis digitais para similaridade e 22 pixels para área.

Para todos os anos, foi analisado o comportamento das bandas e do índice de vegetação NDVI em todas as imagens, assim como sua resposta visual, pela interpretação das imagens em composições coloridas para os 100 pontos amostrais. Foram gerados gráficos representativos do comportamento espectral dos segmentos ao longo do ano, tendo como base os valores de média de reflectância de cada polígono.

Após a análise de cada segmento, estes foram divididos em classes, correspondentes à sua ocupação predominante no ano (soja, cana-de-açúcar, pasto e mata). E para cada classe foi gerado um gráfico referente à média de reflectância dos segmentos pertencentes à classe. Esse processo foi realizado repetidamente para os quatro anos avaliados (2002, 2003, 2004 e 2005).

Previsão

Após a definição da classe de cada ponto amostral em todos os anos estudados, foi realizada uma previsão com base neste histórico de ocupação. Para a realização da previsão de ocupação dos pontos amostrais, foram atribuídos diferentes valores aos anos avaliados. Maiores valores foram atribuídos aos anos mais recentes, pois se acredita que a influência na ocupação dos pontos esteja mais fortemente relacionada à ocupação no momento mais atual t_1 do que no momento anterior t_2 , em que t é o ano avaliado. Porém, todos os anos possuem alguma influência.

O cálculo da probabilidade de ocupação do ponto foi realizado por classe, ou seja, cada ponto possui uma probabilidade de ocupação para cada classe. Se o ponto não foi ocupado pela classe em questão no ano determinado, receberá o valor 0. Se foi ocupado em 2002 recebe 0,1, em 2003 recebe 0,2, em 2004 recebe 0,3 e em 2005 recebe 0,4.

Ao final da atribuição dos valores por ponto para cada classe, somam-se os valores por classe. Obtém-se, assim, uma probabilidade de ocupação para cada ponto, e a classe que obtiver o valor mais alto será definida como a previsão de ocupação para o próximo ano. No caso em que o ponto for ocupado pela mesma classe em todos os anos, o somatório dos valores será 1, ou seja, a chance de que este ponto pertença a esta mesma classe no próximo ano é muito alta.

Trabalho de campo e Auditoria

Foi realizado um trabalho de campo entre os dias 5 e 12 de março de 2006, com a finalidade de validação dos dados gerados pela previsão de uso do solo e para a definição dos pontos passíveis de auditoria.

Feito o trabalho de campo, foi possível a verificação de quais os pontos passíveis de serem auditados. Essa auditoria está baseada essencialmente no histórico de ocupação da área.

Pelo confronto dos dados de previsão de ocupação e dos dados coletados em campo, foi possível definir quais os pontos não apresentaram coerência no uso do solo entre as duas fontes de informação e, conseqüentemente, quais seriam indicados ao processo de auditoria.

3. Resultados e Discussão

A identificação das classes de uso do solo refere-se ao uso predominante ou mais importante durante o ano. Por exemplo, a classe soja é representada pelos pontos que apresentam soja em Janeiro e, em alguns casos, em Dezembro, mas ao longo do ano tem outros usos. Para o caso da cana-de-açúcar, foram observadas diversas épocas de plantio e colheita. No entanto, o fator predominante na identificação desta classe é a duração de seu vigor vegetativo, a resposta espectral e o formato dos talhões. As classes mata e pasto foram definidos como aqueles que apresentam padrões relativamente estáveis de cobertura do solo durante todo ano (**Figura 2**).

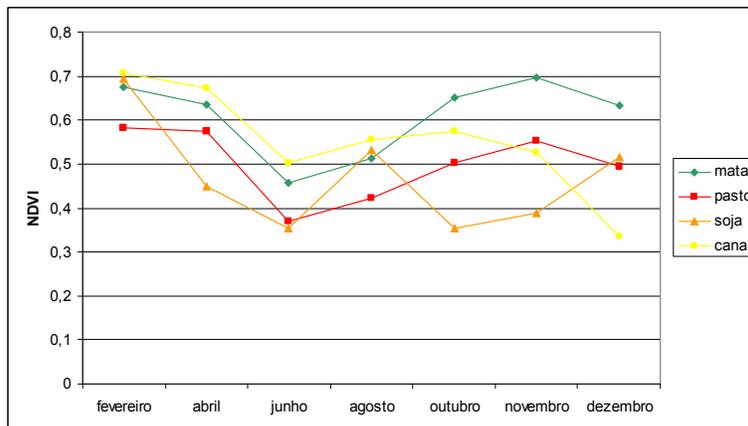


Figura 2 – Comportamento temporal médio dos alvos ao longo do ano.

A classe soja apresenta um comportamento bem característico ao longo de um ano. Sua resposta espectral é observada no início do ano (imagem de Fevereiro), a colheita já ocorrida em Abril e o reaparecimento da resposta espectral de nova cultura no final do ano, quando a soja foi plantada novamente. Outro padrão evidente para os talhões de soja foi que em meados de Agosto outra cultura podia ser plantada para o melhor aproveitamento da terra. Neste caso, pela análise do calendário agrícola, é possível que a cultura fosse o trigo ou aveia preta, plantada logo após a colheita da soja, em Abril. Este segundo padrão pode ser observado na Figura 2. A identificação da soja é facilitada, por esta ser uma espécie com resposta espectral bem distinta dos outros usos, principalmente na fase de maior desenvolvimento da cultura. Na composição RGB 453 TM/Landsat-5, as áreas com presença de soja ficam nítidas, pois estas apresentam uma coloração amarelada/alaranjada. As culturas que passam a ocupar os talhões de soja, no meio do ano, têm uma coloração mais avermelhada.

Na análise espectral dos segmentos de mata houve algumas variações relacionadas aos períodos mais chuvosos e menos chuvosos. No comportamento temporal do NDVI, que é um índice sensível ao vigor vegetativo, observou-se uma queda nos valores de Fevereiro a Junho, mais acentuada no intervalo entre os meses de Abril e Junho, quando começa o período menos chuvoso. A partir de Agosto os valores começam a aumentar, permanecendo constantes nos últimos meses do ano. Os valores de NDVI variam na faixa de 0,45 a 0,67 nos meses de Junho e Novembro, respectivamente. Para as bandas do visível, as variações de reflectância são de 0,05 a 0,1. A estrutura da floresta é a principal responsável pelos menores valores de reflectância na época menos chuvosa (imagens de Junho e Agosto). Nesta época, sempre há perda de folhas, o que pode aumentar os espaços na estrutura florestal e assim diminuir a reflectância da mata nesse período.

Pela grande variabilidade do comportamento da cana-de-açúcar dentro de um mesmo ano, optou-se por analisar esta classe com os dados de todos os anos agrupados, ou seja, os pontos de cana-de-açúcar foram avaliados no intervalo de quatro anos. Os polígonos representativos dessa classe apresentaram características facilitadoras em sua identificação. Por ser uma espécie da família *Gramineae*, a cana-de-açúcar apresenta folhas erectófilas, com resposta espectral distinta de espécies com disposição foliar planófila (Jackson e Pinter, 1986), como é o caso da soja. A coloração de um talhão de cana na composição RGB 453 é avermelhada e sua textura é bem homogênea, assim como a maioria das culturas agrícolas. Outro aspecto muito relevante na identificação de um talhão de cana é sua fenologia, bem diferenciada de culturas de verão, podendo ter duração de um ano a um ano e meio por ciclo, em que se observam altos valores de NDVI ao longo de alguns meses, caracterizando uma cultura de ciclo longo. O tipo de cultivo também pode caracterizar a cultura, pois é possível uma boa percepção visual de carregadores e talhões bem delimitados.

Mesmo com sua variabilidade característica, foi possível identificar um padrão específico de comportamento do pasto durante o ano. Em se tratando de NDVI, seu comportamento temporal é bem semelhante ao observado para os polígonos de mata (Figura 2), porém o pasto apresenta valores mais baixos de NDVI em magnitude (de 0,37 a 0,58), enquanto que para a mata os valores de NDVI variaram de 0,45 a 0,67. As variáveis banda 5 e banda 7, também apresentaram o mesmo comportamento espectro-temporal para as duas classes, porém com valores de reflectância mais altos associados à classe pasto.

Avaliação do padrão de uso do solo em quatro anos

A maior diferença encontrada entre as classes quanto à área utilizada no município foi para o pasto, que apresentava uma área de 34.944 ha em 2002 e caiu para 25.858 ha em 2005. Esta diferença é justificada pelo avanço da agricultura nos últimos anos, quando as áreas de pasto foram transformadas em áreas de plantio de diversas culturas (Brandão et al., 2005). Como evidência deste fato, além da diminuição dos pontos de pasto, está o aumento da área de soja entre 2002 e 2005, que passou de 9.784 ha para 14.676 ha. A área de mata permaneceu a mesma e nas mesmas posições geográficas durante todos os anos avaliados. O ponto amostral de milho (ponto 69) que aparece em 2005, foi um ponto com soja nos três anos anteriores. A área de cana-de-açúcar sofreu um pequeno incremento entre 2002 e 2005, passando de 16 para 19 pontos, respectivamente.

A **Figura 3** mostra a dinâmica dos valores de NDVI em todas as imagens do banco de dados para o ponto 47, que é um ponto representativo do grupo cana. Os menores valores, indicados por círculos, estão associados aos períodos de palhada da cultura, que ocorrem entre agosto e outubro

para todos os anos analisados. Os maiores valores ocorrem entre o final e o início de cada ano, como indicam as elipses vermelhas, no período de maior vigor vegetativo da cultura. Há uma percepção de que os valores de pico de NDVI são decrescentes à medida que os anos passam. Embora não haja comprovação empírica, esse comportamento pode indicar uma possibilidade de monitoramento do decréscimo produtivo da cana em safras sucessivas.

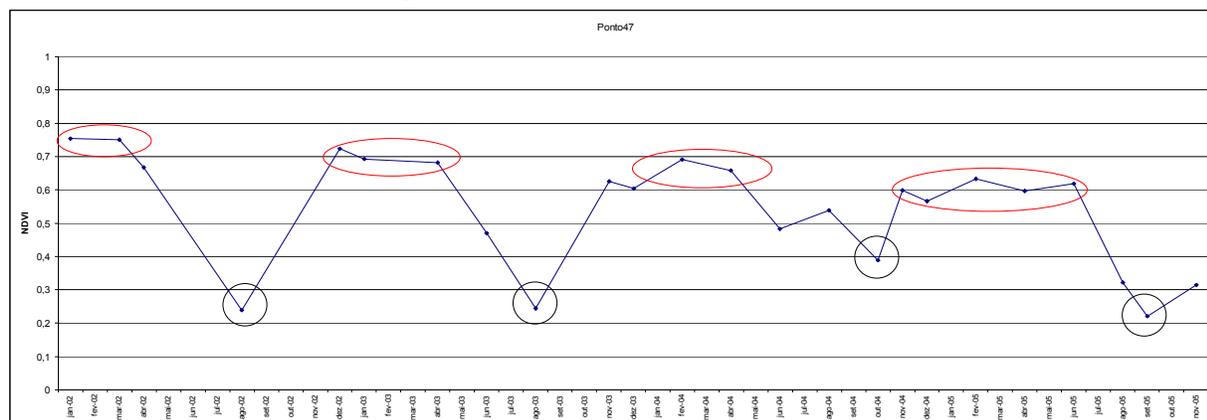


Figura 3 – Comportamento do NDVI para o ponto 47 ao longo de quatro anos (2002 a 2005)

Previsão

Foi desenvolvido um processo de cálculo de probabilidade de cada classe vir a ocupar um ponto. O somatório dos valores de cada ano para cada classe por ponto define o grupo de probabilidade. Esses grupos foram divididos como mostra a **Tabela 1**:

Tabela 1 – Grupos de probabilidade de ocupação de acordo com o histórico apresentado por cada ponto amostral.

Somatório	Probabilidade de ocupação
0	Muito baixa
0,1 a 0,3	Baixa
0,4 a 0,6	Média
0,7 a 0,9	Alta
1	Muito alta

O grupo *muito baixa* só ocorre quando o ponto não foi ocupado pela classe em todos os anos avaliados, e o grupo *muito alta* representa os pontos onde a ocupação foi a mesma durante todos os anos. Para o grupo *baixa*, as situações de uso possíveis são: 2002, 2002/2003, 2003 e 2004. O grupo *média* ocorre quando as seguintes combinações de uso foram observadas: 2002/2004, 2002/2003/2004, 2002/2005, 2003/2004, 2003/2005 e 2005. E para o grupo *alta*, as combinações de uso possíveis são: 2004/2005, 2003/2004/2005, 2002/2003/2005 e 2002/2004/2005.

É necessário deixar claro que o grupo *muito alta*, não significa 100% de certeza de ocupação do ponto pela classe, ou seja, um ponto com valor 1 na classe soja, tem uma chance muito alta de ser soja no próximo ano; porém, pode haver mudança de ocupação devido a outros fatores que fogem ao acompanhamento dos dados do histórico da área, e que levem a uma mudança na ocupação.

Este procedimento de previsão foi a chave para a definição dos pontos que serão auditados após o recebimento pela planilha de campo preenchida pelos técnicos.

Trabalho de campo e Auditoria

Após a definição das probabilidades de ocupação para cada classe em todos os pontos amostrais, foram escolhidas os grupos de probabilidade de ocupação que representariam os pontos que deveriam passar por auditorias. Ou seja, após o recebimento da planilha de campo verifica-se se o preenchimento do dado de referência coletado em campo está coerente com a previsão realizada. Os grupos de probabilidade de ocupação definidos como críticos para a realização de auditoria foram os grupos *baixa* e *muito baixa*.

O trabalho de campo foi realizado entre os dias 5 e 12 de Março de 2006. Foram visitados 81 pontos, dos quais três (pontos 53, 81 e 100) se encaixaram nos grupos baixo e muito baixo de probabilidade de ocupação, ou seja, pontos passíveis de auditoria. O histórico de ocupação de alguns destes pontos será descrito a seguir.

O ponto 53, em todo o decorrer da análise, foi descrito como soja, tanto nas análises das imagens como nos dados coletados em campo pelos técnicos em 2004 e 2005. Porém, no trabalho de campo realizado em 2006, foi observada uma área de pasto na área em questão. Neste caso, a ocupação por pasto no ponto foi calculada como sendo muito baixa, já que esta classe não havia sido contemplada nos anos anteriores.

Para o ponto 81, uma área de conversão de pasto para soja, foi observado um talhão de milho na fase inicial. Pela data do campo, pode ser que a colheita da soja tivesse sido realizada anteriormente. Contudo, não se observou nenhum vestígio de soja sob o solo, o que não era de se esperar, já que a prática de plantio direto é muito utilizada na região. Para este ponto, também é recomendada auditoria, pois a classe milho apresenta uma probabilidade de ocupação muito baixa.

4. Conclusão

Dos 100 pontos amostrais presentes na área, 17 tiveram seu uso alterado em quatro anos de avaliação. Na maior parte destes pontos a mudança foi de áreas de pasto transformadas em áreas cultivadas com soja.

As regras de probabilidade de ocupação na área de estudo, definidas com base o histórico de ocupação de cada ponto amostral mostraram-se muito adequadas. Assim como, a definição de classes de probabilidade de ocupação em função do ano em que dado cultivo ocorreu. A demonstração disso é que a previsão realizada pelas regras de probabilidade de ocupação confirmou-se em 79 dos 81 pontos visitados em campo.

Além disso, foi possível a definição de pontos passíveis de auditoria. Estes pontos só puderam ser definidos com a análise das imagens de satélite quanto ao uso do solo em anos passados, que evidenciaram erros cometidos no campo. Essa auditoria pode ser realizada de duas formas:

- Pela utilização de imagens de satélite, quando disponíveis, em que pode ser feita a verificação do uso do solo nos pontos amostrais;
- Pelo trabalho de campo para a verificação e validação dos dados gerados, quando não houver disponibilidade de imagens no momento adequado.

5. Bibliografia

Brandão, A.S.P.; Rezende, G.C.; Marques, R.W.C. Crescimento agrícola no período 1999-2004, explosão da área plantada com soja e meio ambiente no Brasil. **Texto para discussão nº 162**. Rio de Janeiro: IPEA, 2005. 30p.

Chaves, Jr., P.S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. **Remote Sensing of Environment**, v.24, p.459-479, 1988.

Epiphanyo, J.C.N.; Luiz, A.J.B.; Formaggio, A.R. Estimativa de áreas agrícolas municipais, utilizando um sistema de amostragem simples sobre imagens de satélite. **Bragantia**, v.61, n.2, p.187-197, 2002.

Gürtler, S. **Estimativa de área agrícola a partir de sensoriamento remoto e banco de pixels amostrais**. 2003. 179p. (INPE – 9774-TDI/858). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2003.

Instituto Brasileiro de Geografia e Estatística – IBGE. Sistema IBGE de Recuperação Automática – **SIDRA**. [online]. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: Janeiro, 2006.

Jackson, R.D.; Pinter, P.J. Spectral response of architecturally different wheat canopies. **Remote Sensing of Environment**, v.20, n.1, p.43-56, 1986.

Luiz, A.J.B.; Epiphanyo, J.C.N. Amostragem por pontos em imagens de sensoriamento remoto para estimativa de área plantada por município. In: Simpósio Brasileiro de Sensoriamento Remoto, 10, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2001. p.111-118. CD-ROM.

Nieuwenhuis, G.J.A.; Múcher, C.A. Satellite remote sensing and crop growth monitoring. In: Gudmandsen (Ed.), **Future trends in remote sensing**. Balkema, Rotterdam, p.251-262, 1998.

Ortiz, M.J.; Formaggio, A.R.; Epiphanyo, J.C.N. Classification of croplands through integration of remote sensing, GIS and historical database. **International Journal of Remote Sensing**, v.18, n.1, p.95-105, 1997.

Silva, P.G.; Graça, P.M.A.; Santos, J.R.; Shimabukuro, Y.E. Avaliação da correção atmosférica e retificação radiométrica em estudos multitemporais. In: Simpósio Brasileiro de Sensoriamento Remoto, 11. Belo Horizonte, 2003. **Anais...**São José dos Campos: INPE, 2003. p.2135-2142.

Uchida, S. **Discrimination of agricultural land use using multi-temporal NDVI data**. In: 22nd Asian Conference on Remote Sensing, 05-09 Novembro 2001, Singapura. [online]. Disponível em: <<http://www.crisp.nus.edu.sg/~acrs2001/pdf/061UCHIDA.PDF>>. Acesso em: Setembro, 2005.