

Estimativa de produtividade de cana-de-açúcar (*Saccharum officinarum* L.) por meio de técnica de análise de regressão linear múltipla

Luciana Miura Sugawara¹
Bernardo Friedrich Theodor Rudorff¹
Corina da Costa Freitas¹
Michelle Cristina Araujo Picoli¹
Marcos Adami¹

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{lmiura, bernardo}@dsr.inpe.br, {corina}@dpi.inpe.br, {michelle, adami}@dsr.inpe.br

Abstract. Crop yield estimation methods using GIS and remote sensing techniques have been developed mainly during the last decade, but have not yet been used to support operational agriculture statistics. The aim of this work is to estimate sugarcane (*Saccharum officinarum* L.) yield using a multiple linear regression model. The study was carried out over two growing seasons from 2004/2005 to 2005/2006 on 444 sugarcane fields (from 2nd to 5th cut) in the Plant of Catanduva, Sao Paulo State. Quantitative evaluated variables were: i) estimated yield by agrometeorological-spectral model (PR); ii) observed yield at previous harvest (TCHa); and iii) NDVI values from a single date Landsat/TM image (NDVI). Qualitative evaluated variables were: i) growing season year (ANO); ii) sugarcane variety (VAR); iii) sugarcane harvest stage (EST); iv) Soil Water Retention (CAD); v) sugarcane harvest method (COT); and vi) stillage application (VIN). Best result was obtained with the model that used the following variables: i) PR; ii) TCHa; iii) NDVI; iv) EST2; v) EST3 e vi) COT1. This model explained 65% of sugarcane yield variation.

Palavras-chave: sugarcane, LAI, NDVI, crop, regression model, cana-de-açúcar, IAF, NDVI, produtividade agrícola, modelo de regressão.

1. Introdução

O agronegócio registrou um saldo de US\$ 23,03 bilhões na balança comercial de janeiro a julho de 2006. O valor das exportações do setor sucroalcooleiro cresceu 40,3% nesse período e atingiram US\$ 3,55 bilhões em geração de divisas para o País. Este crescimento se deve à elevação de 60,5% dos preços do açúcar e do álcool no mercado internacional (CNA, 2006). O Brasil é o maior produtor e exportador de açúcar de cana do mundo e com os menores custos de produção, sendo o Estado de São Paulo o responsável por cerca de 60% de todo o açúcar produzido no País e por 70% das exportações nacionais deste produto (UNICA, 2006).

Com este cenário, estatísticas confiáveis sobre as safras agrícolas tornam-se um requisito fundamental para que sejam traçadas medidas para favorecer todo o processo produtivo, desde a infra-estrutura até a comercialização (Rizzi e Rudorff, 2003). No Brasil, as estimativas de produtividade agrícola dos órgãos oficiais são predominantemente realizadas de forma subjetiva. Para o aumento da confiabilidade destas estimativas há necessidade de se buscar alternativas que utilizem métodos objetivos, por exemplo, aqueles que relacionam a variabilidade meteorológica aos fatores que afetam o crescimento e desenvolvimento das culturas agrícolas e definem a sua produtividade, possibilitando inclusive associar um erro a estas estimativas.

A quantificação do efeito das variáveis meteorológicas sobre a produtividade de uma cultura agrícola pode ser realizada através do emprego de modelos matemáticos, como mostram os trabalhos de Baier (1973), Doorenbos e Kassam (1979), Rudorff (1985), Berlato (1987), Rudorff e Batista (1990), Hoogenboom (2000), Fontana et al. (2001), entre outros. Mais recentemente, foi demonstrada a possibilidade de incorporar esses modelos aos Sistemas de Informações Geográficas (SIGs), viabilizando, assim, a estimativa e o acompanhamento de

produtividade de culturas em grandes regiões (Hartkamp et al., 1999; Berka et al., 2003). Para captar a variabilidade espacial da cultura de soja, Rizzi (2004) integrou aos modelos espacializados os índices de vegetação, pois estes refletem as condições de crescimento impostas pelo efeito conjunto das variáveis meteorológicas e de cultivo ao longo do ciclo da cultura.

Neste sentido, podem traduzir os estágios de desenvolvimento da cana-de-açúcar, que é uma cultura semi-perene, plantada em dois sistemas de cultivo, dando origem à cana planta de ano, cuja colheita se dá em torno de um ano após o plantio e à cana planta de ano-e-meio, que é colhida cerca de um ano e seis meses após o plantio. O tempo de permanência da cana planta no campo influencia diretamente a produtividade em sua primeira colheita (Primeiro corte ou Estágio 1), como pode se verificar pelos valores de produtividade média apresentados na **Figura 1**.

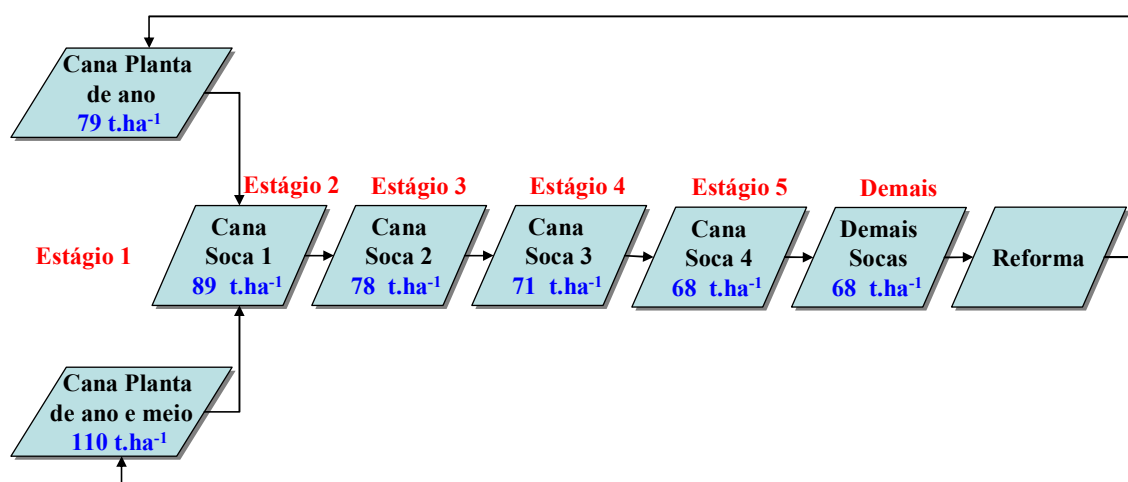


Figura 1: Estágios de corte da cana-de-açúcar e os valores de produtividade média para os anos safra de 1998 a 2003.

Fonte: Adaptado de CTC (2004).

A produtividade das canas socas tende a diminuir gradativamente com o aumento dos estágios de corte. Quando a produtividade atinge um nível que não é mais interessante economicamente, o que ocorre, em geral, cinco anos após o plantio, a cultura passa pelo processo de reforma, ou seja, a cana é retirada e uma nova muda é plantada em seu lugar (**Figura 1**).

Dentro deste contexto o presente trabalho teve como objetivo avaliar o uso de um modelo de regressão linear múltipla para estimar a produtividade agrícola da cana-de-açúcar nos estágios de soca através de variáveis espectrais, meteorológicas e agronômicas, para os anos-safra 2004/2005 e 2005/2006 na área de abrangência da Usina Catanduva.

2. Material e Métodos

O estudo foi realizado nas áreas de cultivo de cana-de-açúcar da Usina Catanduva, cuja sede se localiza no município de Ariranha no Estado de São Paulo, próximo aos limites com os municípios de Palmareis Paulista e Pindorama. A Usina Catanduva possui lavouras de cana em diversos municípios desta região. A **Figura 2** mostra a distribuição espacial da cultura de cana-de-açúcar nas safras 2004/2005 (Rudorff et al., 2004) e 2005/2006 (Sugawara et al., 2006), bem como a distribuição espacial dos talhões de cana avaliados neste trabalho.

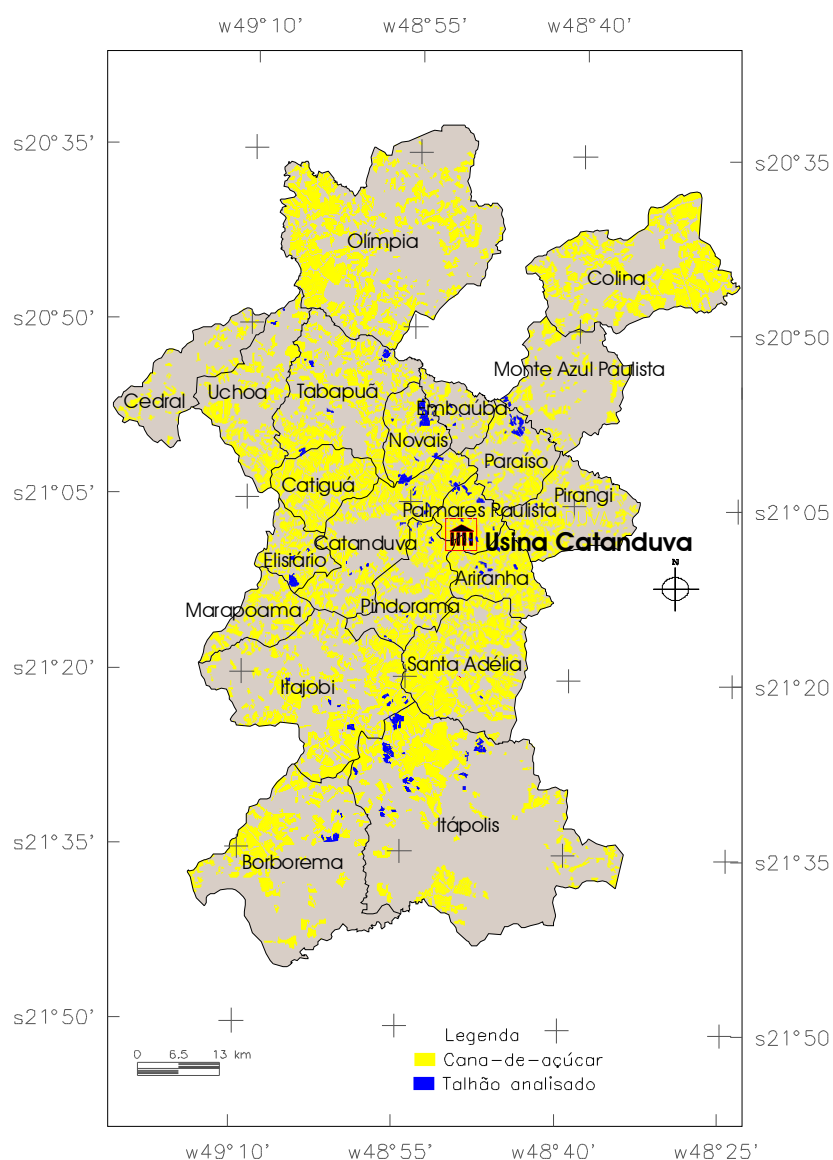


Figura 2: Distribuição espacial da cultura de cana-de-açúcar nas safras 2004/2005 e 2005/2006 na região de Catanduva e os talhões de cana amostrados pertencentes à Usina Catanduva.

Adaptado de Rudorff et al. (2004) e Sugawara et al. (2006).

O trabalho realizado por Picoli (2006) utilizou variáveis agrometeorológicas e espectrais integradas em um modelo agrônomo-meteorológico-espectral (Berka et al., 2003; Rizzi, 2004) para o cálculo da produtividade agrícola da cana. Picoli (2006) utilizou os resultados obtidos pelo modelo, juntamente com os dados quantitativos e qualitativos fornecidos pela Usina Catanduva, bem como os valores do NDVI obtidos através das bandas 3 e 4 do sensor Thematic Mapper (TM) do satélite Landsat-5 e os integrou através do uso de um modelo de agregados de redes neurais artificiais para estimar a produtividade agrícola da cana.

No presente trabalho, a integração dos mesmos dados utilizados por Picoli (2006) foi feita através da técnica de regressão linear múltipla em 444 talhões de cultivo de cana nos estágios 2 a 5 (EST2 a EST5), distribuídos entre os anos safra de 2004/2005 e 2005/2006.

O modelo de estimativa de produtividade de cana-de-açúcar para a Usina Catanduva foi obtido pela análise das seguintes variáveis quantitativas: i) produtividade estimada pelo modelo agrônomo-meteorológico-espectral (PR); ii) produtividade observada na safra anterior (TCHa); iii) valor do NDVI calculado com o uso de imagens Landsat do mês de

fevereiro (NDVI). As variáveis qualitativas utilizadas foram: i) ano safra (ANO); ii) variedade de cana-de-açúcar (VAR); iii) estágio de corte da cana-de-açúcar (EST); iv) capacidade de água disponível no solo (CAD); v) tipo de colheita (COT); vi) aplicação de vinhaça (VIN).

A técnica de análise de discriminantes foi utilizada para reduzir o número de classes das variáveis qualitativas através do agrupamento de classes com características semelhantes ponderada pela produtividade observada dos talhões (TCHobs), que é a variável dependente do modelo a ser obtido. O critério utilizado para a união das classes em um mesmo grupo foi a probabilidade das mesmas serem iguais com $\alpha=5\%$.

Após esta análise prévia, foi aplicada a técnica de “Best Subsets” (Neter et al., 1996), que avalia todas as combinações possíveis entre as variáveis para encontrar o melhor conjunto de variáveis para estimar a TCHobs, com base no valor de R^2 . Com a escolha do conjunto de variáveis que formaram o modelo, foram feitos testes para verificar os efeitos de interação entre elas.

Em seguida, foi feito o estudo dos resíduos para avaliar a homocedasticidade, a correlação dos erros com as variáveis independentes e a normalidade dos mesmos. Além disso, foi feita uma análise para avaliar a presença de multicolinearidade nos dados e verificar se os erros eram autocorrelacionados. A presença de “outliers” foi avaliada pelo gráfico de normalidade dos resíduos e pelo teste de resíduos padrões.

A avaliação do modelo ajustado foi feita através da análise de 38 talhões previamente selecionados aleatoriamente, que não foram incluídos na análise inicial. Nestes talhões foi aplicado o modelo ajustado (TCHj) e estimada a produtividade. Em seguida, uma análise de regressão linear simples foi realizada entre a TCHobs e a TCHj. O modelo também foi avaliado por um índice d (Willmott, 1985) que verifica a concordância entre os dados observados e os estimados. Este índice varia de 0 a 1, em que 1 indica total concordância e 0 total discordância entre o valor observado e o valor estimado. Além do índice d, também são fornecidos os valores de erro sistemático (Es) e erro aleatório (Ea), componentes do erro absoluto médio (EAM), que é uma medida da magnitude média das diferenças entre os valores estimados e observados.

3. Resultados e Discussão

A **Tabela 1** resume o resultado do agrupamento das variáveis qualitativas. A análise de discriminantes demonstrou que as variedades da variável VAR foram agrupadas em dois grupos (VAR1 e VAR2).

Tabela 1: Redução do número de elementos por variável qualitativa através de análise de agrupamentos.

Variável	Número de elementos	
	Inicial	Final
VAR	17	2
EST	4	3
CAD	8	4
COT	4	2
Total	33	11

A variável EST formou três grupos, EST2, EST3 e EST4e5, em que o Estágio 5 foi agrupado ao Estágio 4, pois as produtividades médias de ambos são semelhantes (**Figura 1**). Os valores da variável CAD foram reduzidos à metade (CAD1, CAD2, CAD3 e CAD4), de modo que os solos com as características semelhantes de retenção de água foram incluídos no mesmo grupo. Os tipos de colheita (COT) foram agrupados em dois grupos de corte: manual

(COT1) e mecanizado (COT2). Deste modo, o número de elementos das variáveis qualitativas foi reduzido de 33 para 11, o que facilitou as análises realizadas.

A técnica de “Best Subset” apontou que o melhor custo-benefício seria obtido com o uso de seis variáveis, pois os valores de R^2 tendem a se estabilizar, ou seja, a TCHobs não seria mais bem explicada com o uso de mais variáveis no modelo de regressão (**Figura 3**).

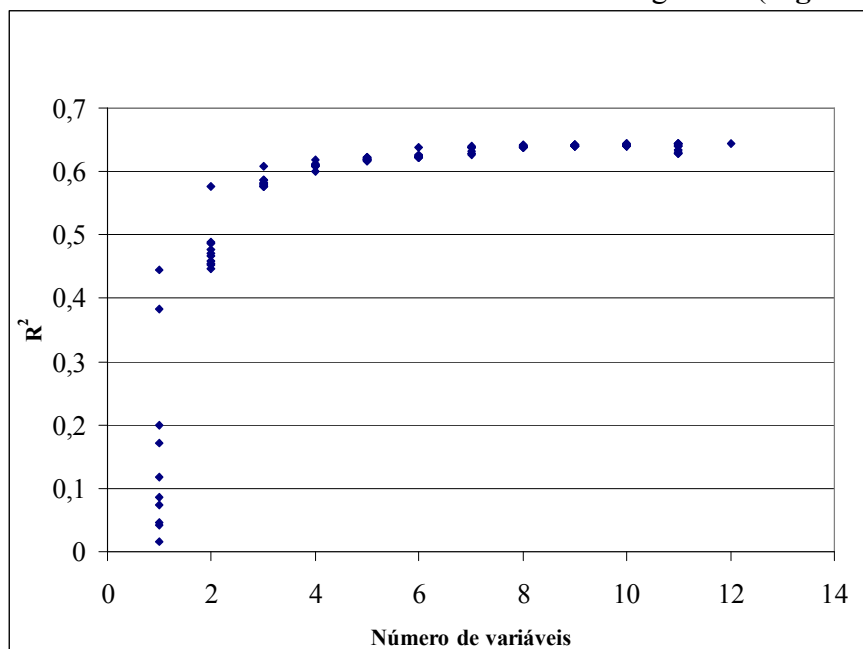


Figura 3: Valores de R^2 obtidos de acordo com o número de variáveis utilizados no modelo de regressão com a aplicação da técnica de “Best Subset”.

Assim, o modelo sugerido pela técnica de “Best Subset” levou em conta as seguintes variáveis quantitativas: i) PR; ii) TCHa; iii) NDVI; e as variáveis qualitativas (dummy): i) EST2; ii) EST3 e iii) COT1. Com a análise da matriz de correlação (**Tabela 2**) entre estas variáveis se observou que a TCHobs estava mais correlacionada com as variáveis TCHa e NDVI, contudo, a contribuição da variável PR não poderia ser descartada.

Tabela 2: Matriz de correlação entre as variáveis utilizadas no modelo e a TCHobs

	PR	TCHa	NDVI	EST2	EST3	COT1	TCHobs
PR	1	0,275	0,282	0,031	-0,007	0,092	0,351
TCHa	0,275	1	0,442	0,424	-0,097	0,214	0,632
NDVI	0,282	0,442	1	0,382	-0,009	0,331	0,669
EST2	0,031	0,424	0,382	1	-0,697	0,46	0,459
EST3	-0,007	-0,097	-0,009	-0,697	1	-0,3	-0,059
COT1	0,092	0,214	0,331	0,46	-0,298	1	0,414
TCHobs	0,351	0,632	0,669	0,459	-0,059	0,414	1

A avaliação do teste t (**Tabela 3**) mostrou que todas as variáveis selecionadas no modelo têm importância significativa ao nível de 5%. Com a análise dos valores de β , foi verificado que a variável que contribuiu menos no modelo foi a PR, mas esta variável é de suma importância, pois nela estão representadas as variações decorrentes do clima (ocorrência de seca, por exemplo), do manejo da cultura, entre outras (Berka et al., 2003; Rizzi, 2004).

Tabela 3: Parâmetros estimados pela análise de regressão, erros e valores do teste t.

	β	Erro Padrão	B	Erro Padrão	t(339)	valor p
Intercepto			-87,421	11,7254	-7,45565	0
NDVI	0,33523	0,04083	121,302	14,7748	8,21005	0
TCHa	0,31909	0,03959	0,29380	0,03646	8,06021	0
EST2	0,25832	0,05908	10,8414	2,47936	4,37265	0,000016
EST3	0,20429	0,05022	8,24250	2,02643	4,06750	0,000059
COT1	0,16303	0,03687	6,84230	1,54730	4,42208	0,000013
PR	0,14693	0,03480	1,09040	0,25823	4,22251	0,000031

A **Tabela 4** mostra que o modelo obtido pode explicar 65% da variação da TCHobs e o teste F mostrou que a relação de regressão entre a variável dependente e as independentes é significativa. Ou seja, as variáveis independentes podem explicar a variável dependente (TCHobs). Além disso, a regressão linear simples realizada entre os valores de TCHobs e os de TCHj para a avaliação do modelo resultou em um coeficiente de determinação semelhante ao obtido pelo modelo, com um erro padrão da mesma ordem de magnitude, o que demonstra que o modelo está coerente. O índice d foi igual a 0,87, o que indica que apesar do modelo ajustar 65% da variância dos dados, o modelo concorda com 87% dos dados na validação. O erro absoluto estimado foi de 10,22 ton.ha⁻¹, o erro sistemático foi de 11,16 ton.ha⁻¹ e o erro aleatório foi de 5,33 ton.ha⁻¹.

Tabela 4: Resultados da Regressão de acordo com os talhões de cana analisados para o ajuste e a validação do modelo através da análise de regressão simples.

Estatística de regressão	Ajuste do modelo	Validação do modelo
R ²	0,6513	0,6031
F	105,5175	54,7053
F - valor p	0,0000	0,00000001
Erro padrão	12,0066	11,4760
Número de talhões	346	38

O modelo final de estimativa de produtividade agrícola da cana foi constituído pelas seguintes variáveis: i) PR; ii) TCHa; iii) NDVI; iv) EST2; v) EST3 e vi) COT1. A **Tabela 5** apresenta o modelo final, para todas as combinações das variáveis qualitativas.

Tabela 5: Modelo de regressão e a sua decomposição em função das variáveis qualitativas.

Estágio	Corte	Modelo
-	-	$TCHj = -87,42 + 1,09 \cdot PR + 0,29 \cdot TCHa + 121,30 \cdot NDVI + 10,84 \cdot EST2 + 8,24 \cdot EST3 + 6,84 \cdot COT1$
EST2	COT1	$TCHj = (-87,42 + 10,84 + 6,84) + 1,09 \cdot PR + 0,29 \cdot TCHa + 121,30 \cdot NDVI$
EST2	COT2	$TCHj = (-87,42 + 10,84) + 1,09 \cdot PR + 0,29 \cdot TCHa + 121,30 \cdot NDVI$
EST3	COT1	$TCHj = (-87,42 + 8,24 + 6,84) + 1,09 \cdot PR + 0,29 \cdot TCHa + 121,30 \cdot NDVI$
EST3	COT2	$TCHj = (-87,42 + 8,24) + 1,09 \cdot PR + 0,29 \cdot TCHa + 121,30 \cdot NDVI$
EST4e5	COT1	$TCHj = (-87,42 + 6,84) + 1,09 \cdot PR + 0,29 \cdot TCHa + 121,30 \cdot NDVI$
EST4e5	COT2	$TCHj = (-87,42) + 1,09 \cdot PR + 0,29 \cdot TCHa + 121,30 \cdot NDVI$

Os talhões de cana no Estágio 2 tem uma produtividade de 10,84 t.ha⁻¹ maior do que os talhões nos Estágios 4 e 5. Já os talhões no Estágio 3 tem uma produtividade de 8,24 t.ha⁻¹

maior do que os talhões dos Estágios 4 e 5. A variável corte manual apresentou um incremento de 6,84 t.ha⁻¹ na produtividade da cana quando comparada à variável de corte mecanizado. Neste caso específico, se tem o conhecimento que a implementação do sistema de corte mecanizado começou a ocorrer há poucos anos atrás. A mudança do sistema de corte manual para o mecanizado, requer uma reestruturação dos talhões. No corte manual os talhões são menores para facilitar a entrada de caminhões para a retirada da cana colhida. No caso de uma colheita mecanizada, a eficiência se dá quando se tem grandes talhões. Com a mudança do formato dos talhões, áreas que antes eram carreadores se tornaram áreas de cultivo de cana. A inclusão destas áreas no sistema de cultivo resulta em redução na produtividade agrícola, pois possuem solos compactados e de baixa fertilidade. Entretanto maiores estudos devem ser realizados a este respeito.

4. Conclusões

De acordo com os resultados obtidos é possível estimar a produtividade agrícola da cana através de um modelo de regressão linear múltipla que leva em conta as seguintes variáveis: produtividade estimada pelo modelo agrônomo-meteorológico-espectral, produtividade real da safra anterior, NDVI, estágios de corte da cana e o tipo de corte realizado na colheita.

Estudos mais detalhados devem ser realizados para verificar a influência da mudança do tipo de corte manual para o mecanizado.

Referências

- Baier, W. Crop-Weather Analysis Model: Review and Model Development. **Journal of Applied Meteorology**, v. 12, n. 6, p. 937-947, 1973.
- Berka, L.M.S., Rudorff, B.F.T., Shimabukuro, Y.E. Soybean yield estimation by an agrometeorological model in a GIS. **Scientia Agricola**, v.60, n.3, p.433-440, 2003.
- Berlato, M. A. **Modelo de relação entre o rendimento de grãos de soja e o déficit hídrico para o Estado do Rio Grande do Sul**. São José dos Campos. 94p. Dissertação (Doutorado em Meteorologia) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 1987.
- Confederação da Agricultura e da Pecuária do Brasil (CNA). **Soja, açúcar e álcool recuperam exportações agrícolas**. <<http://www.cna.org.br/cna/publicacao/noticia.wsp?tmp.noticia=12741>> Acesso em: 22 agosto 2006.
- Centro de Tecnologia Copersucar – CTC. **Censo Varietal Copersucar 2003**. Piracicaba: Centro de Tecnologia Copersucar, 2004. 16 p.
- Doorenbos, J.; Kassam, A. H. **Yield response to water**. Rome: Food and Agriculture Organization of the United Nations, 1979. 193 p. (FAO-Irrigation and Drainage Paper n. 33).
- Fontana, D. C.; Berlato, M. A., Lauschner, M. H.; Mello, R. W. Modelo de estimativa de rendimento de soja no Estado do Rio Grande do Sul. **Pesquisa Agropecuária Brasileira**, v. 36, n. 3, p. 399-403, 2001.
- Hartkamp, A. D.; White, J. W.; Hoogenboom, G. Interfacing geographic information systems with agronomic modelings: A review. **Agronomy Journal**, v. 91, n. 5, p. 761-772, 1999.
- Hoogenboom, G. Contribution of agrometeorology to the simulation of crop production and its applications. **Agricultural and Forest Meteorology**, v. 103, n. 1-2, p. 137-157, 2000.
- Neter, J.; Kutner, M.H.; Nachtsheim, C.J., Wasserman, W. **Applied Linear Statistical Model**. New York: McGraw-Hill, 1996. 1408 p.
- Picoli, M. C. A. Estimativa da produtividade agrícola da cana: Estudo de caso Usina Catanduva. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2006. (no prelo).
- Rizzi, R. **Geotecnologias em um sistema de estimativa da produção de soja: estudo de caso no Rio Grande do Sul**. 214 p. (INPE-12271-TDI/983). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2004. Disponível em:

<<http://mtc-m12.sid.inpe.br/rep-/sid.inpe.br/jeferson/2004/07.02.14.17>>. Acesso em: 23 ago. 2006.

Rizzi, R.; Rudorff, B. F. T. Imagens Landsat na estimativa da área plantada com soja em municípios do Rio Grande do Sul In: Simpósio Brasileiro de Sensoriamento Remoto, 11., Belo Horizonte, 2003. **Anais...** São José dos Campos: INPE, 2003.

Rudorff, B. F. T. **Dados Landsat na estimativa da produtividade agrícola da cana-de-açúcar.** São José dos Campos. 114p. (INPE-3744-TDL/202). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 1985.

Rudorff, B. F. T.; Batista, G. T. Yield estimation of sugarcane based on agrometeorological-spectral models. **Remote Sensing of Environment**, v. 33, n. 3, p. 183-192, 1990.

Rudorff, B. F. T.; Berka, L. M. S.; Moreira, M. A.; Duarte, V.; Rosa, V. G. C. **Estimativa de área plantada com cana-de-açúcar em municípios do estado de São Paulo por meio de imagens de satélites e técnicas de geoprocessamento: ano safra 2004/2005.** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2004. 54 p. (INPE-11421-RPQ/762).

Sugawara, L. M.; Rudorff, B. F. T.; Vieira, R. M. S. P.; Rafaelli, D. R.; Rezende, A. C. P.; Afonso, A. G.; Moreira, M. A.; Duarte, V. **Imagens de satélites na estimativa de área plantada com cana na safra 2005/2006 – Região Centro-Sul.** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2006. 73 p. (no prelo).

União da Agroindústria Canavieira de São Paulo (UNICA). **Cana-de-açúcar: Produtos.** Disponível em: <http://www.unica.com.br/pages/cana_produtos.asp> Acesso em: 22 agosto 2006.

Willmott, C. J.; Ackleson, S. G.; Davis, R. E.; Feddema, J. J.; Klink, K. M.; Legates, D. R.; O'Donnell, J.; Rowe, C. M. Statistics for the evaluation and comparison of models. **Journal of Geophysical Research**, v. 90, n. C5, p.8995-9005, 1985.