

# PREVISÃO DE SAFRA DE SOJA UTILIZANDO ÍNDICES DE SATÉLITE PARA A REGÃO NORTE DE PARANÁ

<sup>1</sup>André Luiz Farias de Souza<sup>2</sup>, William Tse Horng Liu<sup>3</sup>

## RESUMO

Dados de produtividade de soja e índices derivados de satélites, entre 1981 e 1991 são utilizados no desenvolvimento de modelos de previsão de safra para as microregiões de Maringá, Londrina e Cornélio Procopio. Combinações de VCI e TCI foram utilizadas para construir e selecionar 2 modelos através do pacote estatístico SAS. Os resultados mostram que ambos os modelos estimaram bem com erros menores que 10 %, exceto 1992, que apresentou excesso hídrico na colheita. Conclui-se que os índices de satélites podem ser uma ferramenta de alta potencialidade no monitoramento e previsão da produção agrícola. Sugere-se a incorporação da previsão de tempo após a data da previsão feita pelo modelo, pois isto pode melhorar a precisão da previsão.

**PALAVRAS-CHAVES:** Soja, índices de vegetação, previsão de safra.

## INTRODUÇÃO

Entre os produtos de exportação do Brasil merece destaque a soja, não somente como fonte de divisas, mas também por ser a maior fonte de óleo vegetal e de proteína, tanto para alimentação humana, como também animal. O grão da soja possui grande diversidade de aplicações, desde a produção de óleos e margarinas, até a confecção de tintas e sabões (Costa, 1996). Em virtude dos grandes investimentos que são empregados nesta cultura, torna-se imprescindível ter em mãos uma ferramenta que possibilite uma estimativa da produtividade desta cultura, visto que a ocorrência perdas pode provocar efeitos danosos e muitos prejuízos.

Os modelos de previsão de produtividade agrícola, podem ser agrupados em dois grandes grupos: simulação de processos fisiológicos e os estatísticos. Dentro deste último grupo, podemos dividir dois subgrupos em função do tipo de dados utilizados em sua construção: estatístico - agrometeorológicos e via satélite. Recentemente, os dados de satélites, que fornecem uma fonte de dados contínuos de alta frequências temporal e espacial, estão sendo utilizados no desenvolvimento

---

<sup>1</sup> Este projeto de pesquisa é financiado pela FAPESP, Proc. N° 1997:11568-3.

<sup>2</sup> Aluno de Pós-graduação. Departamento de Ciências Atmosféricas, IAG / USP. E-mail: [andrelfs@model.iag.usp.br](mailto:andrelfs@model.iag.usp.br). Bolsista CAPES.

<sup>3</sup> Professor Doutor. Departamento de Ciências Atmosféricas, IAG / USP. E-mail: [wilthliu@usp.br](mailto:wilthliu@usp.br).

de modelos alternativos para a previsão da safra. Índices, como NDVI, VCI e TCI (Índice de Condição da Temperatura), têm sido utilizados na construção de modelos de previsão de safra (Rasmussen, 1997; Unganai e Kogan, 19988). Os modelos via satélite podem melhorar estas avaliações pois possuem resoluções temporal e espacial melhores, ou seja, pode-se estudar uma região ao nível dos píxeis.

No norte de estado do Paraná foram escolhidas três regiões que merecem destaque: O Setentrião Paranaense, polarizado por Maringá ( $\approx 4.950 \text{ km}^2$ ); O Médio Paranapanema ( $\approx 10.174 \text{ km}^2$ ), polarizado por Londrina e o Norte do Paraná, polarizado por Cornélio Procópio ( $\approx 6.785 \text{ km}^2$ ). Nas três microregiões, a soja é o principal produto agrícola e sua área total abrange cerca de  $22.000 \text{ km}^2$  (-22,5 S; -24 S; -53 W e -50,5 W). Cada microregião concentra aproximadamente 21 municípios e levou em consideração uma variação média da quantidade de municípios de cada microregião. Esta variabilidade está mais próxima da composição atual adotada pela Secretaria de Estado do Desenvolvimento Urbano do estado do Paraná.

## METODOLOGIA

Os dados de VCI e TCI do período entre 1981 e 1991, foram utilizados na construção de dois modelos alternativos de estimativa de produtividade através do pacote estatístico SAS. Os dados de VCI e TCI foram gerados através dos dados de NOAA / AVHRR GAC 10 dias composto, cedidos pelo GSFC / NASA com as correções radiométricas e atmosféricas aplicadas (Eidenshink e Faundee, 1997). A análise de regressão linear múltipla é constituída de 10 períodos de verão (1981 - 1991), onde os períodos mais sensíveis foram selecionados. A previsão dos modelos alternativos foi aplicada para os anos de 1992, 1993 e 1994. Cada observação consiste de 18 períodos decendiais que tem o seu início em outubro de um ano e tem o seu término em março do ano seguinte.

As variáveis utilizadas neste estudo são: *Variável Prevista (DY)* – Porcentagem do desvio entre produtividade média anual de soja (kg/ha) observada na microregião composta de Maringá, Londrina e Cornélio Procópio, fornecida pelo IBGE e a tendência tecnológica de rendimento; *Variáveis Previsoras* - Dados de VCI (Índice de Condição da Vegetação) e TCI (Índice de Condição da Temperatura).

A construção dos modelos segue duas etapas. Na primeira, a tendência de rendimento é obtida considerando os dados históricos de produtividade fornecidos pelo IBGE, entre os anos de 1981 e 1991. Na segunda etapa, os dados da porcentagem do desvio (DY) entre a produtividade média observada ( $Y_M$ ) e a tendência de rendimento ( $Y_{TREND}$ ) são obtidos e esta relação é utilizada para desenvolver os modelos alternativos. Esta porcentagem é relacionada com os dados de VCI e

---

TCI e a partir daí, a estimativa de produtividade final é obtida pela aplicação dos modelos e corrigida através da tendência de rendimento. A correlação apresentada entre os dados de produtividade média e a tendência de rendimento foi de 0,62 com um nível de significância de 0,0051. Foram rejeitados os modelos que apresentassem  $R^2 < 0,5$ .

## RESULTADOS E DISCUSSÕES

Os dois modelos escolhidos para descrever DY na microregião de estudo incluíram duas variáveis e o termo de interceptação. O coeficiente de correlação quadrático e o nível de significância para cada modelo foram respectivamente: 0,87 e 0,005 e 0,86 e 0,006. As variáveis introduzidas neste modelo podem ser observadas na tabela 1 e as equações finais para os modelos de DY podem ser escritas como:

$$DY_1 = 87,133 + 0,046 * TRL_{79} - 0,066 * VT_{1214} ;$$

$$DY_2 = 89,324 + 0,044 * TRL_{79} - 0,129 * VR ;$$

Onde:  $TRL_{79}$  – Valor de TCI médio do mês de dezembro, considerando-se o primeiro, o segundo e o terceiro períodos decendiais, respectivamente, os períodos 7, 8 e 9;  $VT_{1214}$  – Valor de VCI médio do terceiro período do mês de janeiro e o primeiro e segundo períodos de fevereiro, respectivamente os períodos 12, 13 e 14;  $VR$  – Valor de VCI médio do segundo e terceiro períodos do mês de janeiro e o primeiro e segundo períodos do mês de fevereiro, respectivamente, 10, 11, 12 e 13.

De acordo com o que pode ser observado na tabela 1, no modelo 1 a variável  $TRL_{79}$  explica 38,25 % da variância de DY, respectivamente. A variável  $VT_{1214}$  explica 35,22 % da variabilidade do modelo. No modelo 2, a variável  $VR$  explica 34,76 % da variância. Através da análise da variância, pode-se estimar o comportamento do modelo construído, comparando-o com um modelo que não inclua nenhuma das variáveis previsoras, somente a média e o resíduo. Para atingir este objetivo, a estatística F é utilizada para testar a hipótese nula: todos os parâmetros são nulos (com exceção do termo de interceptação), contra a hipótese alternativa: todos os termos são não nulos. A análise de variância pode ser observada na tabela 2.

De acordo com uma tabela de distribuição F, o valor de F maior que  $F(2, 7) = 4,12$  possibilita a rejeição da hipótese nula com um nível de significância igual a 5 % em ambos os modelos. Os parâmetros utilizados na equação de regressão podem ser observados na tabela 3. Nesta tabela encontra-se o valor do teste t-Student aplicado no teste de hipóteses sobre as variáveis, o nível de significância que permite a rejeição da hipótese nula, a tolerância e o fator de inflação da variância de para cada variável (Larsen e Marx, 1990).

Dos valores da estatística t-Student para cada variável, pode-se concluir que a rejeição da hipótese nula ocorre com um nível de significância ( $\text{Prob}>T$ ) inferior à 9 %, em todas as variáveis do modelo 1 e 5 % para o modelo 2. A tolerância e o fator de influência da variância encontram-se próximos de 1, e isto fornece uma estimativa de que não ocorrem altas correlações entre as variáveis previsoras do modelo.

Uma amostra da dinâmica dos modelos pode ser observada nas tabelas 4, onde os resíduos studentizados e o índice de influência D de Cook são apresentados. Estes parâmetros auxiliam na interpretação das amostras dos dados utilizados na construção dos modelos e através destes parâmetros, pode-se observar qual a amostra de dados desempenhou maior influência durante o ajuste do modelo (Larsen e Marx, 1990). Baseando-se nestas informações, nota-se que as amostras de 1986 e 1989 foram as que maior peso apresentaram no ajuste de ambos os modelos. Estes anos representam as safras 85/86 e 87/88, que são as amostras 5 e 7, respectivamente. Fisicamente, a observação 5 apresentou altos valores de VCI e baixos valores de TCI, durante os períodos considerados pelas equações de regressão para DY. A observação 7 apresentou baixos valores de VCI e altos valores de TCI. As produtividades observadas e as obtidas através dos modelos alternativos podem ser observadas na tabela 5 e os erros relativos na tabela 6. A figura 1 apresenta as produtividades observadas e previstas pelos modelos.

Os resultados da validação dos dois modelos alternativos desenvolvidos, mostram que ambos estimaram bem as produtividades dos anos de 1993 e 1994. Os erros reais nestes períodos variaram de - 2,6 à 7,5 %. Todavia, a estimativa de produtividade do ano de 1992 foi superestimada por ambos os modelos, apresentando um erro da ordem de 40 %. O excesso de precipitação nos meses de fevereiro e março de 1992, coincidiu com os estágios de maturação e colheita da soja. Isto influenciou a queda de produtividade da safra deste ano. Devido aos modelos utilizarem dados de VCI e TCI nos meses anteriores ao terceiro período de fevereiro, os efeitos danosos do clima posteriores a este influenciam contribuir na estimativa final da produtividade. Os resultados obtidos neste trabalho para a região produtora de soja do norte do estado do Paraná, estiveram de acordo com resultados obtidos por Liu e Kogan (1999), para o estado do Paraná, especialmente para a estimativa do ano 1992.

## **CONCLUSÕES**

Conclui-se que os índices de satélites podem ser uma ferramenta de alta potencialidade no monitoramento e previsão da produção agrícola. Sugere-se a incorporação da previsão de tempo após a data da previsão feita pelo modelo, pois isto pode melhorar a precisão da previsão.

## ACKNOWLEDGE:

The authors wish to thank the Distributed Active Archive Center (Code 902.2) at the Goddard Space Flight Center, Greenbelt, MD, 20771, for producing the data in their present form and distributing them. The original data products were produced under the NOAA/NASA Pathfinder program, by a processing team headed by Ms. Mary James of the Goddard Global Change Data Center; and the science algorithms were established by the AVHRR Land Science Working Group, chaired by Dr. John Townshend of the University of Maryland. Goddard's contributions to these activities were sponsored by NASA's Mission to Planet Earth program.

## REFERÊNCIAS BIBLIOGRÁFICAS

- COSTA, J. A., 1996. Cultura da Soja. Ed. Ivo Manica e J. A. Costa. Porto Alegre – RS. 233 pp.
- EIDENSHINK, J. C. AND J. L. FAUNDEEN, 1997. The 1 km AVHRR global land data set: first stages in implementation. *Int. J. Rem. Sens.* 51: 39 – 56.
- LARSEN, R. J., and M. L. MARX, 1990. Statistics. Ed. Prentice-Hall, New Jersey. 771 pp.
- LIU, W. T. H, and KOGAN, F. 1999. Crop Yield Prediction Using Satelite Indices. (A ser submetido apresentado no XI Congresso Brasileiro de Agrometeorologia).
- RASMUSSEN, M. S., 1997. Operational yield forecasting using AVHRR NDVI data: reduction of enviromental and inter-annual variability. *Int. Jour. Rem. Sens.* 18: 1059 1077.
- UNGANAI, L. S. and F. N. KOGAN, 1998. Drought Monitoring and Corn Yield Estimation in Southern Africa from AVHRR Data. *Ren. Sens. Env.* 68: 219 – 232.

Tabela 1: Variáveis predictoras da equação de regressão e  $R^2$  parcial de cada variável (stepwise).

Variável Incluída	$R^2$ Parcial
Modelo 1	
TRL79	0,3825
VT1214	0,3522
Modelo 2	
TRL79	0,3825
VR	0,3476

Tabela 2: Análise de variância geral.

Fonte	Grau de Liberdade	Soma dos Quadrados	Média dos Quadrados	Valor de F
Modelo 1				
Modelo	2	946,298	473,143	22,221
Erro	7	1491,521	213,074	
Total	9	2437,806		
Modelo 2				
Modelo	2	957,689	478,844	22,265
Erro	7	1480,116	211,445	
Total	9	2437,806		

Tabela 3: Parâmetros estimados, erro padrão das estimativas, teste de hipóteses sobre os parâmetros, tolerância e fator de influência da variância para cada variável incluída no modelo.

Variável	Parâmetro Estimado	Erro Padrão	t p/ H <sub>0</sub> Parâmetro = 0	Prob > T	Tolerância	Fator de Influência da Variância
<b>Modelo 1</b>						
INTERCEP	87,133	28,072	3,104	0,0072	,	0,000
TRL79	0,046	0,022	2,104	0,0034	0,969	1,031
VT1214	-0,066	0,264	-0,251	0,0090	0,969	1,031
<b>Modelo 2</b>						
INTERCEP	89,323	27,323	3,218	0,0047	,	0,000
VR	-0,129	0,377	-0,343	0,0019	0,978	1,022
TRL79	0,044	0,022	2,026	0,0024	0,978	1,022

Tabela 4: Resíduo studentizado e índice de influência D de Cook, para cada observação e modelos 1 e 2.

<b>Modelo 1</b>				<b>Modelo 2</b>			
Obs	ANO	Resíduo Studentizado	Cook's D	Obs	ANO	Resíduo Studentizado	Cook's D
1	1982	0,83	0,068	1	1982	1,059	0,130
2	1983	0,80	0,026	2	1983	0,950	0,072
3	1984	-0,91	0,061	3	1984	-0,913	0,061
4	1985	-0,36	0,017	4	1985	-0,308	0,015
5	1986	-2,08	1,253	5	1986	-2,131	1,685
6	1987	0,46	0,028	6	1987	0,152	0,005
7	1988	0,25	0,006	7	1988	0,148	0,001
8	1989	1,97	1,021	8	1989	1,893	0,887
9	1990	-0,49	0,127	9	1990	-0,394	0,034
10	1991	-0,67	0,031	10	1991	-0,653	0,024

Tabela 5: Produtividades observadas (IBGE) e previstas pelos modelos alternativos para os três anos de teste: 1992, 1993 e 1994.

ANO	Y <sub>OBS</sub> (kg/ha)	YM 1(kg/ha)	YM 2 (kg/ha)
1992	1638,0	2328,8	2282,3
1993	2208,0	2329,1	2374,4
1994	2409,0	2350,4	2345,9

Tabela 6: Erros relativos entre as produtividades observadas e previstas pelos modelos alternativos para os três anos de teste: 1992, 1993 e 1994.

ANO	ER REL YM1 (%)	ER REL YM2 (%)
1992	42,17	39,33
1993	5,49	7,54
1994	-2,43	-2,62

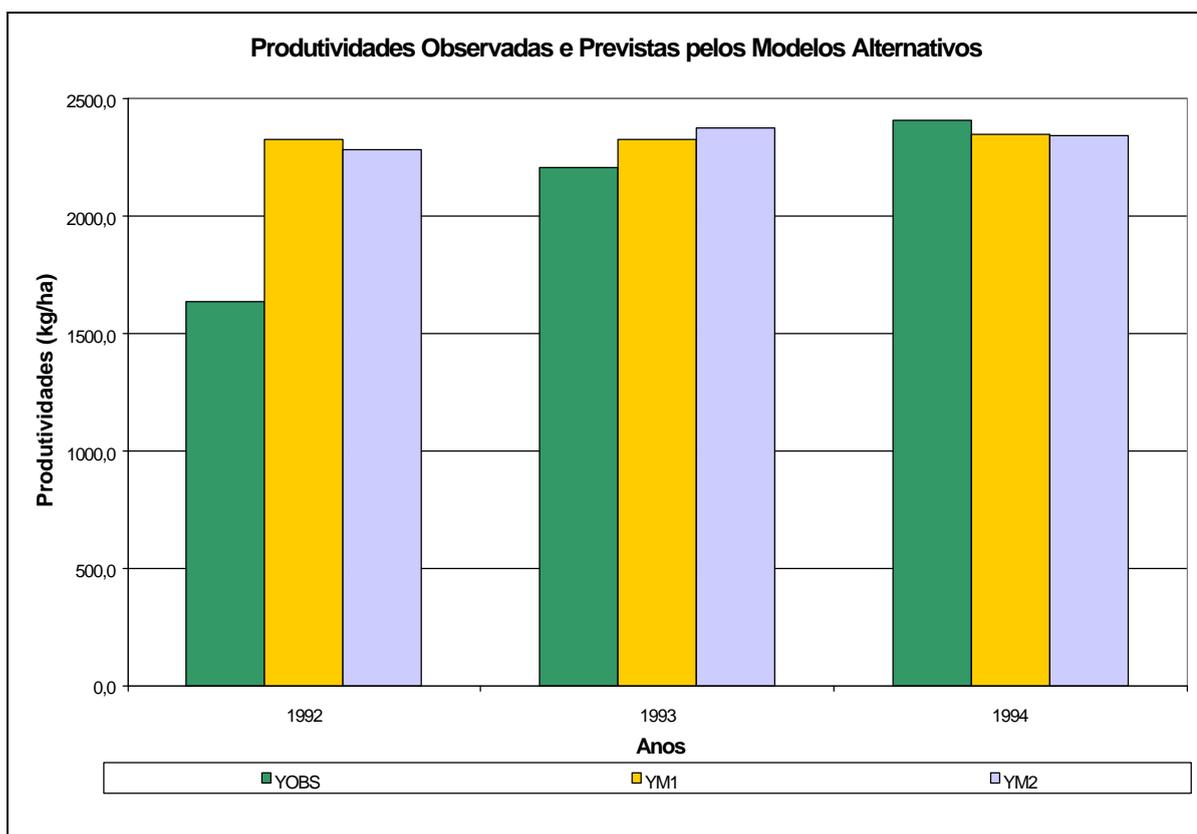


Figura 1: Produtividades observadas (IBGE) e previstas pelos modelos alternativos.