

AVALIAÇÃO DO DESEMPENHO DA CLASSIFICAÇÃO E ESTIMATIVA DA PROPORÇÃO DA
CULTURA DA CANA-DE-AÇÚCAR PELO PROCEDIMENTO-INPE UTILIZANDO
DADOS DO LANDSAT

Bernardo F.T. Rudorff e Getúlio T. Batista

Instituto de Pesquisas Espaciais - INPE
Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq
12200 - São José dos Campos, SP, Brasil

RESUMO

Este estudo teve como objetivo determinar o efeito da concentração da cultura, o período de aquisição de imagens de satélite e o número de canais no desempenho da classificação e estimativa da proporção da cultura da cana-de-açúcar em segmentos amostrais (9x11km, aproximadamente). Foram utilizados dados multitemporais e multiespectrais do LANDSAT/MSS analisados no IMAGE-100 através do PROCEDIMENTO-INPE. Quatro passagens registradas do LANDSAT/MSS para cada um dos quatro segmentos localizadas em regiões de alta (dois segmentos) e baixa (dois segmentos) concentração da cultura canavieira foram analisadas. Todas as possíveis combinações de datas foram feitas para as diversas classificações utilizando o classificador de máxima verossimilhança. O procedimento requer a rotulação de 100 pontos alocados aleatoriamente no segmento. Tanto a rotulação quanto a posterior avaliação do desempenho da classificação foi feita utilizando cartas temáticas confeccionadas a partir de fotografias aéreas infravermelhas e de levantamento topográfico. Os resultados mostraram que houve um desempenho similar de classificação para as diversas combinações das datas de passagens do satélite. O efeito da concentração da cultura no desempenho da classificação parece ser significativo. O algoritmo de *correção de tendência* melhorou a estimativa da proporção da cultura da cana-de-açúcar.

ABSTRACT

The objective of this study was to determine the effect of crop density, satellite data acquisition period and number of spectral bands on the classification performance and proportion estimate of sugar cane in sampling segments (9x11km approximately). Multitemporal and multispectral LANDSAT/MSS data were analysed in the IMAGE-100 (INPE's Image Analysis System) using the PROCEDURE-INPE. Four LANDSAT/MSS registered passes for each one of the four studied segments located in regions of high density (two segments) and low density (two segments) of sugar cane were analysed. All possible combinations of four acquisition dates were classified

using the maximum likelihood decision rule. This procedure requires the labeling of 100 pixels randomly distributed in the segment. Both labeling and classification performance assessment were done using thematic maps obtained either from CIR aerial photographs or from topographic charts. Results showed similar classification performance for all combinations of date acquisitions. The effect of crop concentration on classification performance seems to be significant. The algorithm for bias correction improved the sugarcane proportion estimation.

1. INTRODUÇÃO

A disponibilidade de informações sobre a área plantada com culturas agrícolas é de fundamental importância para um sistema de previsão de safra. Os dados LANDSAT/MSS têm se mostrado de grande utilidade para fornecer informações relevantes, de forma rápida e precisa. No entanto, ainda carecem de estudos que tornem operacional a utilização destes dados para fins de previsão de safra.

A cultura da cana-de-açúcar tem sido muito estudada através dos dados LANDSAT, especialmente no que se refere ao mapeamento e à estimativa de área plantada por meio de técnicas de interpretação visual, utilizando imagens multiespectrais e multitemporais (Mendonça et alii, 1981). Técnicas de análise automática, para a estimativa de área plantada com cana-de-açúcar, eram utilizadas de forma restrita no Instituto de Pesquisas Espaciais (INPE) (Mendonça, 1980) uma vez que não se dispunha de um procedimento que utilizasse o aspecto multitemporal dos dados LANDSAT para classificação automática.

Recentemente foi desenvolvido no INPE um procedimento automático para a estimativa de área plantada com culturas agrícolas, em segmentos de imagens do LANDSAT de aproximadamente 9x11km, denominado PROCEDIMENTO-INPE (Câmara et alii, 1982).

A metodologia proposta para a estimativa de área de uma determinada cultura, através deste procedimento, é a de amostragem estratificada. Assim a região em estudo é dividida em zonas homogêneas ou estratos, nos quais são escolhidos de forma aleatória unidades amostrais denominadas segmentos. Através do PROCEDIMENTO-INPE é determinada a proporção da cultura em cada segmento. As proporções da cultura nos diversos segmentos são agregadas e obtém-se a estimativa da área da cultura no estrato. Por sua vez as áreas estimadas nos estratos são agregadas para obter as estimativas para toda a região.

O objetivo geral deste estudo consiste em avaliar o desempenho do PROCEDIMENTO-INPE, ao nível de segmento amostral, para fornecer subsídios a futuros trabalhos de previsão de safras que envolvem grandes áreas. Os objetivos específicos são: determinar o efeito da concentração da cultura sobre o desempenho do procedimento; definir as melhores datas ou períodos para aquisição de imagens; determinar o número de canais mais apropriados para a classificação e verificar o efeito da *correção de tendência* do classificador na estimativa da proporção da cultura.

2. METODOLOGIA

O estudo foi desenvolvido nas regiões de Lençóis Paulista e de Araçatuba, ambas situadas no Estado de São Paulo. O município de Lençóis Paulista caracteriza-se pelo intenso cultivo da cana-de-açúcar, enquanto no município de Araçatuba a pecuária de corte é uma exploração tradicional e só recentemente, através do programa PROALCOOL, o setor canavieiro começou a ser implantado, condicionando consideráveis mudanças no uso atual do solo. Entretanto, a cana-de-açúcar ainda se encontra em baixa concentração na região (Batista et alii, 1984).

Foram selecionados dois segmentos de 9x11km aproximadamente, em cada região, os quais foram superpostos digitalmente (registrados) em quatro passagens do LANDSAT/MSS. As datas de aquisição das imagens do satélite são mostradas na Tabela 1.

TABELA 1

DATAS DE AQUISIÇÃO DE IMAGENS

PERÍODO	LENÇÓIS PAULISTA Órbita 221-Ponto 75	ARAÇATUBA Órbita 222-Ponto 75
1	25 fevereiro 1983	16 fevereiro 1983
2	16 maio 1983	07 maio 1983
3	17 junho 1983	10 julho 1983
4	20 agosto 1983	27 agosto 1983

A primeira data corresponde à fase de intenso crescimento vegetativo da cultura. Os outros alvos frequentemente apresentam resposta espectral similar à da cana-de-açúcar nas imagens do satélite (ex.: pastagens). As demais datas correspondem à fase de maturação e ao período de colheita da cana-de-açúcar, e a confusão com outros alvos nas imagens tende a diminuir (Mendonça et alii, 1981).

A Figura 1 mostra as curvas de crescimento da cultura da cana-de-açúcar e o acúmulo de matéria seca ao longo do período de crescimento. São mostrados também os períodos de aquisição das imagens LANDSAT em relação às fases de crescimento da cultura.

A Figura 2 mostra a distribuição espacial e a proporção da cultura canavieira, obtidas a partir de mapas temáticos considerados como verdade terrestre para as regiões de Lençóis Paulista e Araçatuba, nos quatro segmentos selecionados.

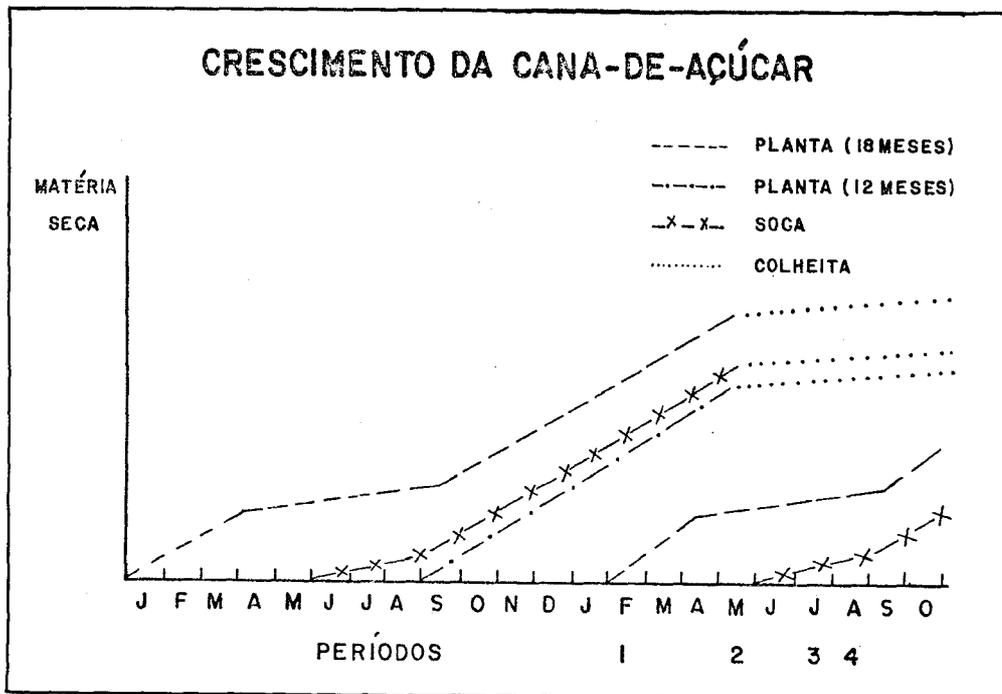


Fig. 1 - Curvas de crescimento da cultura da cana-de-açúcar e períodos de aquisição das imagens LANDSAT.

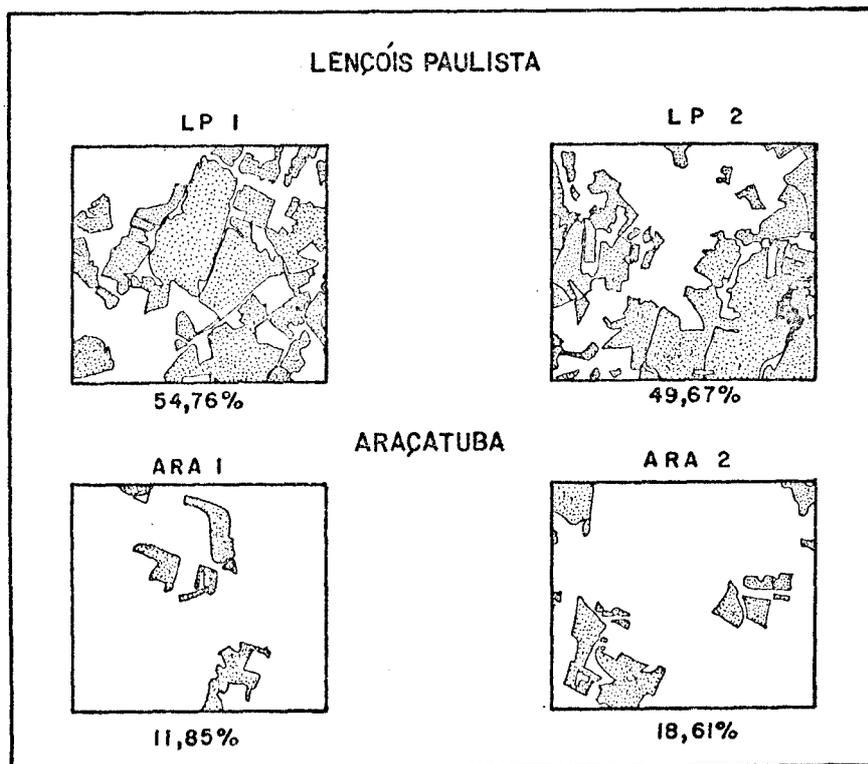


Fig. 2 - Distribuição espacial da cana-de-açúcar nos segmentos de Lençóis Paulista e Araçatuba e respectivas proporções da cultura.

O PROCEDIMENTO-INPE, utilizado para obter a proporção da cultura nos segmentos, envolve a participação do analista (agrônomo-fotointérprete) e do computador, que trabalham de forma interativa, em quatro fases distintas (Figura 3).

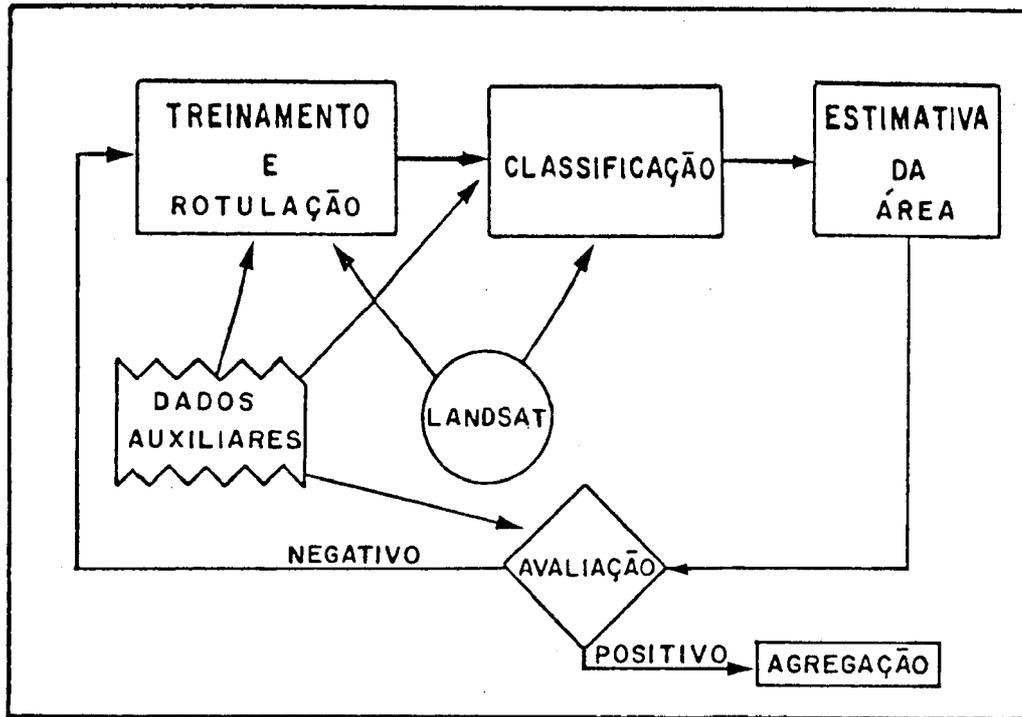


Fig. 3 - Fluxograma do PROCEDIMENTO-INPE.

a) Treinamento e rotulação

Nesta fase o analista rotula manualmente pontos pré-allocados aleatoriamente sobre a imagem com dois objetivos: 1) fornecer os parâmetros necessários para a classificação; 2) dar subsídios para testar a estimativa da proporção e corrigir erros de tendência do classificador na estimativa desta proporção.

Neste experimento os pontos foram rotulados usando informações da verdade terrestre.

b) Classificação

A classificação separa cada ponto do segmento em duas classes: cultura e não-cultura, de acordo com a regra de máxima verossimilhança.

c) Estimativa de proporção

O resultado da classificação é usado para a estimativa da proporção da cultura no segmento. Nesta fase também é feita uma correção das eventuais tendências ocorridas na classificação, utilizando os subsídios fornecidos na fase de treinamento e rotulação.

d) Avaliação e aprovação

O resultado da estimativa da proporção é submetido ao analista para aprovação e, a seu critério, os passos do procedimento poderão ser feitos ou o segmento estará pronto para fornecer a proporção da cultura plantada.

Para testar o desempenho deste procedimento foram feitas todas as combinações possíveis, nos quatro períodos de aquisição de imagens, utilizando os canais 5 e 7 do MSS. Adicionalmente foram feitas classificações em períodos isolados, utilizando os canais 4, 5, 6 e 7. Esta sistemática resultou em 19 classificações por segmento.

Para diminuir eventuais tendências de incluir ou omitir erroneamente áreas canavieiras na classificação, causadas pela regra de decisão do classificador, é utilizada a seguinte equação:

$$P = \frac{11}{11+01} \times \frac{N1}{1} + \frac{10}{00+10} \times \left(1 - \frac{N1}{N}\right), \quad (1)$$

onde:

P = proporção da cultura corrigida,

11 = pontos rotulados pelo analista como cultura e classificados pelo PROCEDIMENTO-INPE como cultura,

01 = pontos rotulados como não-cultura e classificados como cultura,

00 = pontos rotulados como não-cultura e classificados como não-cultura,

10 = pontos rotulados como cultura e classificados como não-cultura,

N1 = número de "pixels" classificados como cultura,

N = total de "pixels" no segmento.

A exatidão da classificação é feita baseada em 60 pontos rotulados pelo analista, os quais são sobrepostos ao resultado da classificação, fornecendo o número de pontos classificados corretamente e o número de pontos omitidos e incluídos na classificação. Estes 60 pontos rotulados são selecionados ao acaso do total de 22932 "pixels" do segmento. O formulário utilizado para calcular a exatidão da classificação foi proposto por Kalensky e Wightman (1976), o qual é descrito a seguir:

a) Exatidão da Classe (EC)

$$EC(\%) = \frac{N}{N+E} \times 100(\%), \quad (2)$$

onde:

N = número de "pixels" classificados corretamente,

E = número de "pixels" classificados erroneamente ("pixels" omitidos + "pixels" incluídos).

b) Exatidão Geral das Classes Cana e Não-Cana

$$EG(\%) = \frac{NC.ECC(\%) + NNC.ECNC(\%)}{NC + NNC}, \quad (3)$$

onde:

NC = número de "pixels" classificados como cana,

NNC = número de "pixels" classificados como não-cana,

ECC = exatidão da classe cana,

ECNC = exatidão da classe não-cana.

A diferença relativa (DR) entre a proporção da cultura, fornecida pelo classificador, e a proporção da cultura na verdade terrestre é dada por:

$$DR(\%) = \frac{\text{Proporção da classificação} - \text{Proporção da Verdade Terrestre}}{\text{Proporção da Verdade Terrestre}} \times 100. \quad (4)$$

3. RESULTADOS

3.1 - ESTIMATIVA DE ÁREA

Os resultados do desempenho da estimativa da proporção da cultura, obtidos pelo PROCEDIMENTO-INPE, são apresentados em termos de diferença relativa percentual (DR%). A Figura 4 mostra comparativamente as diferenças relativas obtidas pela classificação feita utilizando todos os canais (4, 5, 6 e 7) e apenas os canais 5 e 7 do MSS. A pequena diferença existente entre os resultados obtidos com a classificação dos canais 4, 5, 6 e 7 e a classificação dos canais 5 e 7, para uma data, confirma as conclusões de Hixson et alii, (1980) e Doughtry e Hixson, (1979), os quais afirmam ser suficiente a utilização dos canais 5 e 7, de cada data, para a classificação de culturas agrícolas. Assim, as demais classificações das combinações entre datas foram feitas utilizando apenas os canais 5 e 7.

A Figura 5 mostra os resultados das diferenças relativas das proporções da cultura para as classificações unitemporais e multitemporais (canais 5 e 7) para os quatro segmentos estudados.

A Tabela 2 mostra as diferenças relativas (%) da estimativa da proporção da cultura para os quatro segmentos analisados considerando a classificação feita com os quatro períodos de aquisição, observando-se um erro médio da DR(%) de 1,7(%) com um desvio padrão de 12,4(%)

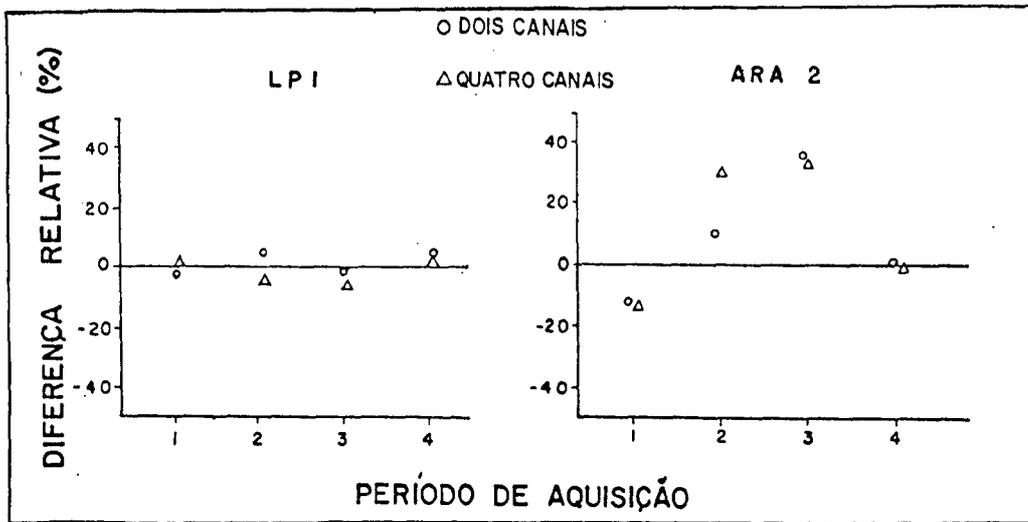


Fig. 4 - Diferença Relativa (%) da proporção da cultura para classificação com uma data utilizando quatro e dois canais.

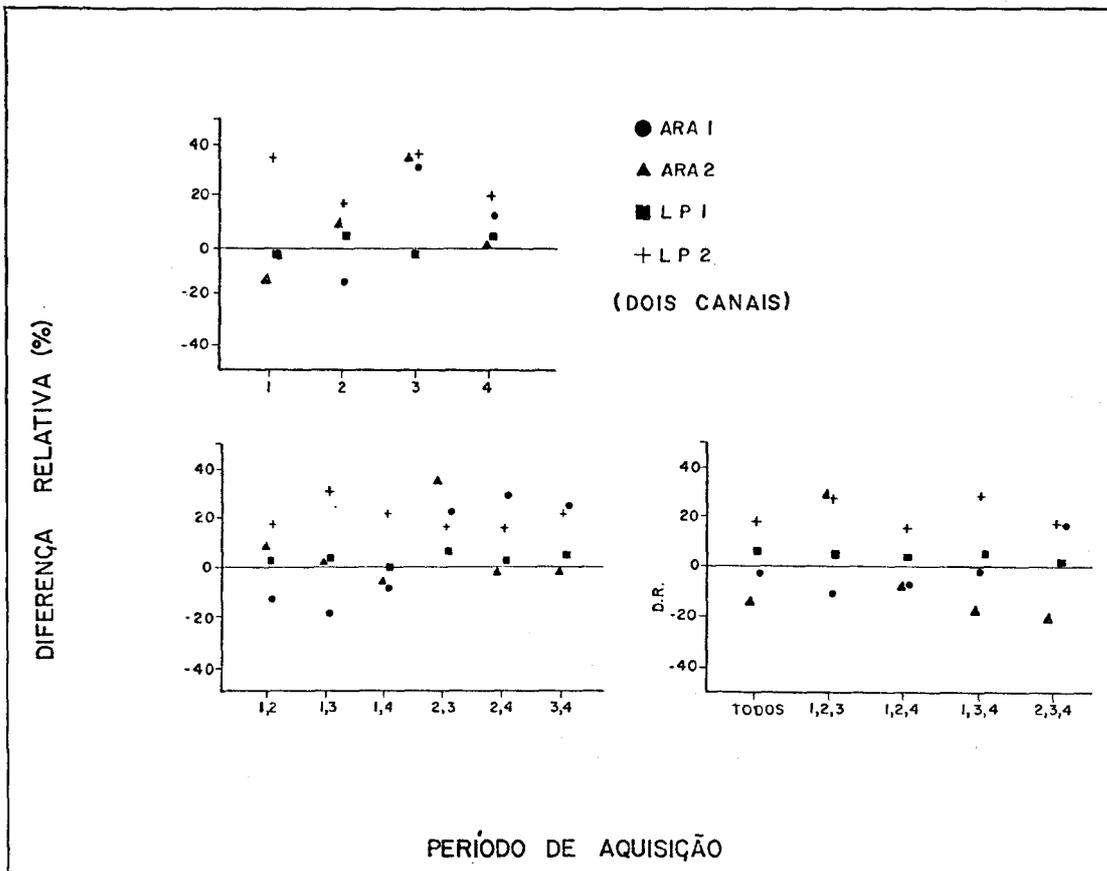


TABELA 2

DIFERENÇA RELATIVA (%) DA PROPORÇÃO DA CULTURA, PARA OS QUATRO SEGMENTOS, OBTIDA DA CLASSIFICAÇÃO COM OS QUATRO PERÍODOS DE AQUISIÇÃO

SEGMENTO	PROPORÇÃO DA CULTURA %	PROPORÇÃO DA VERDADE TERRESTRE %	D.R. %
ARA 1	11,4	11,8	-3,6
ARA 2	15,8	18,6	-15,1
LP 1	58,8	54,8	7,4
LP 2	58,6	49,7	18,1
Média	36,1	33,7	1,7

O número de segmentos (dois para cada região) não foi suficiente para fazer uma análise estatística que fornecesse, com um nível de confiança desejável (0,95), meios para tomar decisões seguras sobre a escolha da melhor data ou combinação de datas para a classificação. Com base na Figura 5 foram escolhidas as combinações que tiveram melhores resultados para os quatro segmentos, dentre as 15 combinações analisadas no experimento.

A Tabela 3 mostra os resultados para as melhores combinações selecionadas e ainda que o desempenho geral do resultado para 2, 3 ou 4 datas é o mesmo. Esta Tabela apresenta também a proporção do analista, obtida a partir da rotulação de 60 "pixels" escolhidos aleatoriamente do total de 22932 "pixels" do segmento. O desempenho geral da proporção do analista praticamente não difere dos resultados da classificação, quando se considera os quatro segmentos. Davidson (1980), utilizando um procedimento similar ao utilizado neste trabalho, concluiu que o resultado da classificação não foi melhor do que aquele obtido pelo analista, através da rotulação de "pixels" amostrados no segmento, para as culturas de milho e soja nos Estados Unidos. Entretanto, examinando os segmentos individualmente, o resultado da classificação é melhor, quando comparado com a proporção do analista, exceto para o segmento LP2.

Houve uma ligeira tendência do classificador em superestimar a proporção da cultura na região de Lençóis Paulista (alta concentração da cultura), enquanto na região de Araçatuba (baixa concentração da cultura) os resultados foram mais heterogêneos para os dois segmentos estudados nas diversas combinações de datas de aquisição das imagens LANDSAT/MSS analisadas. Batista et alii, (1982), trabalhando com as culturas de milho e soja, concluíram que a concentração da cultura é uma das principais características da cena que afeta a classificação digital.

O resultado de todas as classificações para os quatro segmentos melhorou com a *correção de tendência* que visa diminuir eventuais tendências de omitir ou incluir erroneamente áreas canavieiras na classificação.

As Figuras 6 e 7 mostram a DR(%) das proporções obtidas, sem e com *correção de tendência*, para um dos segmentos de Lençóis Paulista e um de Araçatuba, respectivamente.

TABELA 3

ESTIMATIVA DA PROPORÇÃO (%) DE CANA-DE-AÇÚCAR NOS QUATRO SEGMENTOS
PARA AS MELHORES DATAS E COMBINAÇÕES DE DATAS

PERÍODO SEGMENTO	2	4	1,2	1,4	1,2,4	TODAS	PROPORÇÃO DO ANALISTA	PROPORÇÃO DA VERDADE TERRESTRE
ARA 1	10,2	13,6	10,4	11,0	10,8	11,4	8,3	11,8
ARA 2	20,6	18,8	19,5	17,6	17,1	15,8	11,7	18,6
LP 1	57,6	55,7	56,1	54,9	57,8	58,8	46,7	54,8
LP 2	59,1	60,6	58,4	60,9	57,9	58,6	56,7	49,7
Média	36,9	37,2	36,1	36,1	35,9	36,1	30,8	33,7
D.R. %	9,5	10,4	7,1	7,1	6,5	6,5	7,1	-

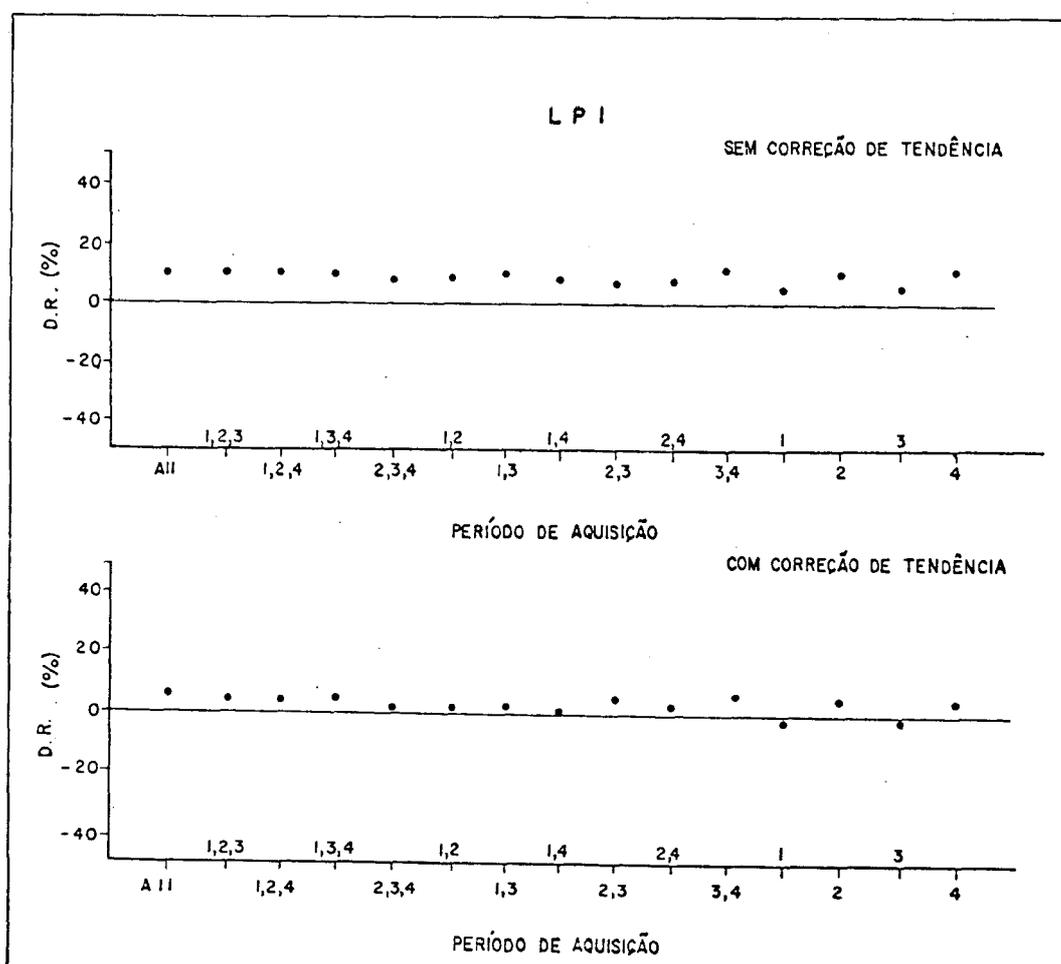


Fig. 6 - Diferença relativa (%) sem e com *correção de tendência*, entre a proporção da cana-de-açúcar da classificação e da verdade terrestre para as diversas combinações de da tas do segmento LP1.

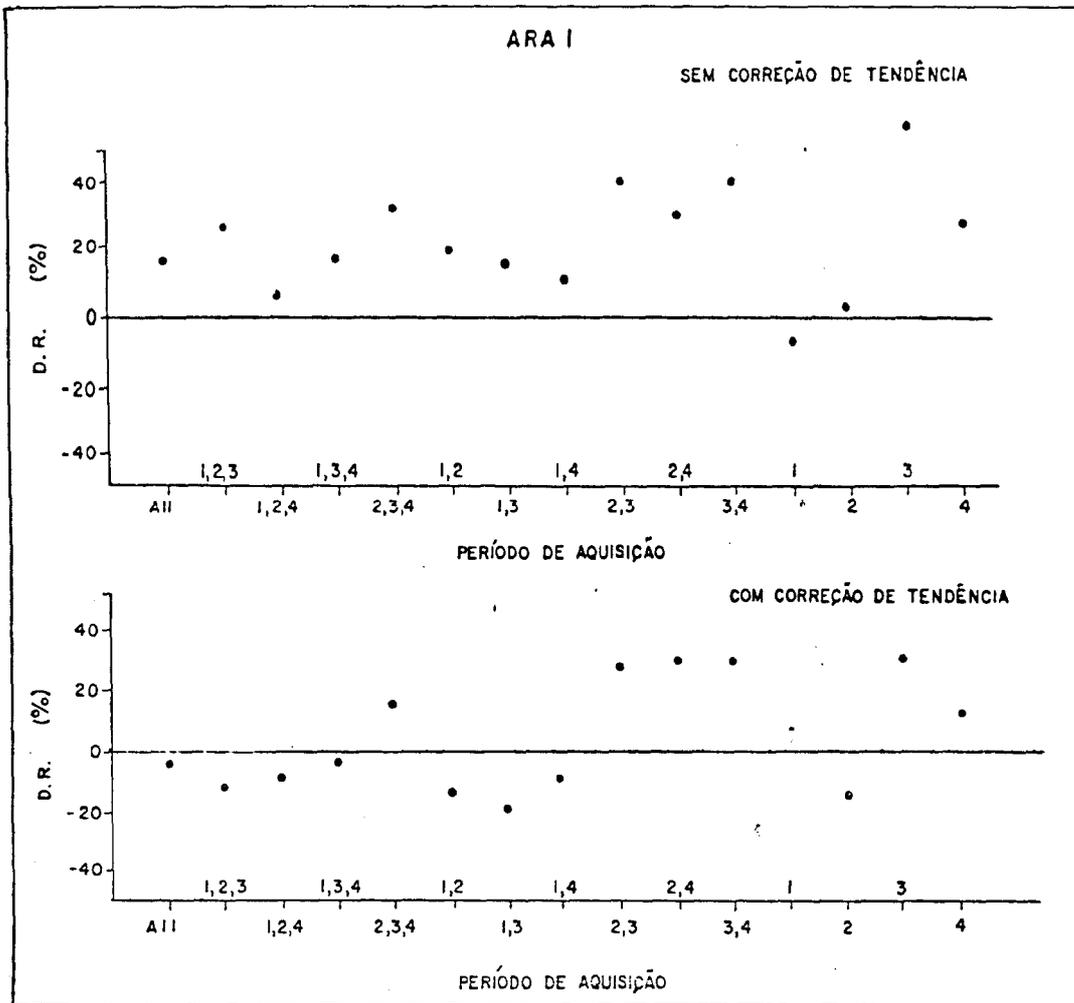


Fig. 7 - Diferença relativa (%) sem e com *correção de tendência*, entre a proporção de cana-de-açúcar da classificação e da verdade terrestre para as diversas combinações de datas do segmento ARA1.

3.2 - EXATIDÃO DE MAPEAMENTO DA CLASSIFICAÇÃO

Os resultados da exatidão de mapeamento da classificação (EC%) para a classe cana e para a classificação geral (EG%) são mostrados nas Figuras 8 e 9. A exatidão da classificação leva em conta a localização das áreas canavieiras, e a avaliação é feita baseada em 60 pontos do segmento rotulados pelo analista que, neste experimento em particular, esteve apoiado em informações de verdade terrestre para a rotulação. Os cálculos são feitos utilizando as Equações 2 e 3 para a exatidão da classe cana e exatidão geral, respectivamente.

Na região de Lençóis Paulista o desempenho do procedimento para exatidão de mapeamento variou em torno de 50% a 70% para as melhores combinações de datas de tomada dos dados LANDSAT selecionados na Seção 3.1. Devido à alta concentração da cultura nestes segmentos (54,76% e 49,67%), o desempenho da classificação para a classe geral foi em torno de 5% inferior ao da classe cana.

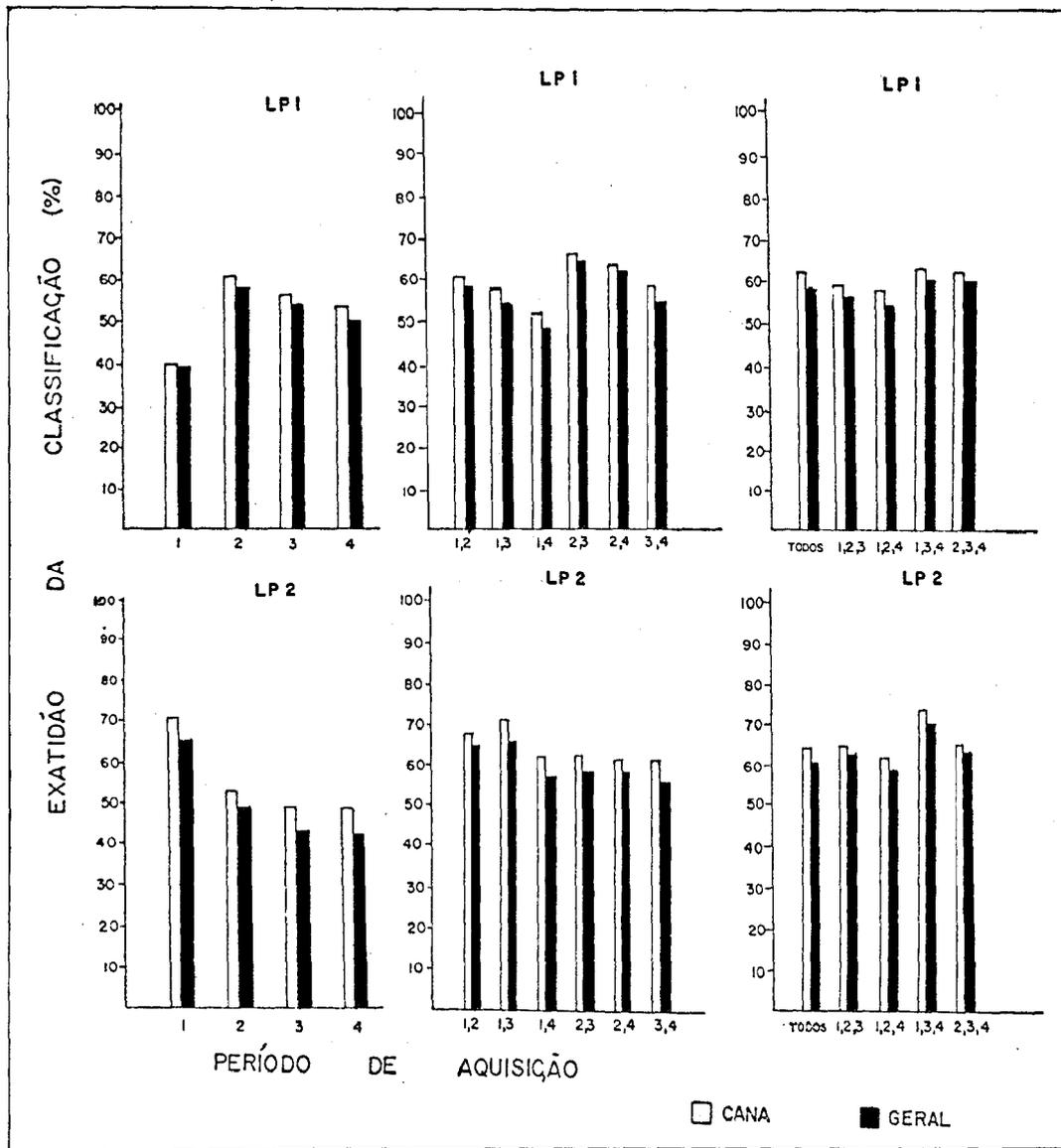


Fig. 8 - Exatidão de mapeamento da classificação (%) dos segmentos de Lençóis Paulista para as diversas combinações de datas de aquisição.

A classe cana teve um baixo desempenho na região de Araçatuba, a qual variou de aproximadamente 20% a 50% para as melhores combinações de datas selecionadas na Seção 3.1. Por outro lado, o desempenho da classificação geral foi alto e variou entre aproximadamente 80% a 90%. Atribui-se este fato à baixa densidade da cultura (11,85% e 18,61%) nos segmentos de Araçatuba.

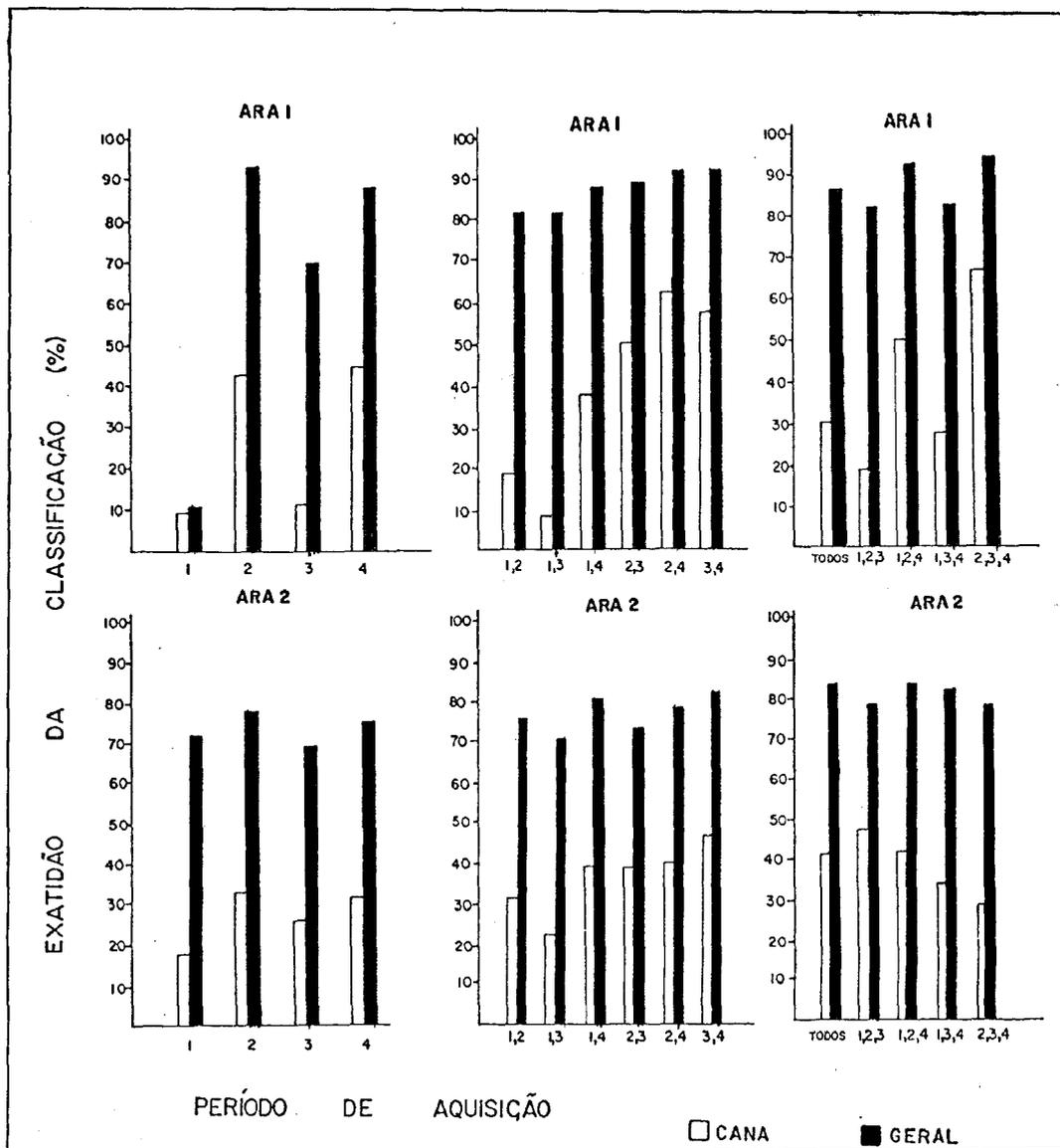


Fig. 9 - Exatidão de mapeamento da classificação (%) dos segmentos de Araçatuba para as diversas combinações de datas de aquisição.

4. CONSIDERAÇÕES FINAIS

Este trabalho representa o primeiro esforço de testar o PROCEDIMENTO-INPE para a cultura da cana-de-açúcar. Embora de caráter preliminar, as seguintes observações devem ser ressaltadas:

- 1) O erro médio da diferença relativa da estimativa da proporção para os quatro segmentos, considerando a classificação feita com os quatro períodos de aquisição, foi de 1,7% com um desvio padrão de 12,4%.
- 2) O procedimento para a estimativa da proporção da cana-de-açúcar para as melhores combinações de três, dois e um único período de aquisição (1,2,4; 1,2; 1,4; 2; 4) teve desempenho similar.

- 3) As proporções de cana-de-açúcar foram subestimadas para a região de baixa concentração de cultura, exceto para o período de aquisição 4. Por outro lado, as proporções foram superestimadas para a região de alta concentração de culturas.
- 4) O procedimento de treinamento para a classificação parece ser inadequado para a caracterização da classe cultura para a região de baixa concentração.
- 5) A exatidão de mapeamento para a classificação da cana-de-açúcar foi tipicamente de 60% e 40% para as regiões de alta e baixa concentração da cultura, respectivamente. Por outro lado, a exatidão geral de mapeamento foi tipicamente de 55% e 85% para as regiões de alta e baixa concentração da cultura, respectivamente.
- 6) O procedimento de *correção de tendências* melhorou significativamente a estimativa da proporção de cana-de-açúcar nos segmentos.
- 7) Os erros de estimativa da proporção de cana-de-açúcar feita pelo analista, através da rotulação de uma amostra aleatória de 60 pontos e pelo classificador, são semelhantes.

A experiência adquirida neste trabalho indica a necessidade de investigar outras abordagens de treinamento do PROCEDIMENTO-INPE, assim como a necessidade de estudo do comportamento espectral da cultura da cana-de-açúcar ao longo do seu ciclo vegetativo para a definição das datas de aquisição mais indicadas para maximizar o desempenho da classificação.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BATISTA, G.T.; LIMA, A.M.; RUDORFF, B.F.T.; MENDONÇA, F.J.; PINTO, S.A.F.; CHEN, S.C.; DUARTE, V. *Desenvolvimento de teste de um sistema de fiscalização das operações de crédito agrícola ao nível de município e de propriedades agrícolas utilizando técnicas de sensoriamento remoto*. São José dos Campos, INPE, ago., 1984. (INPE-3239-RPE/462).
- BATISTA, G.T.; HIXSON, M.M.; BAUER, M.E. Corn and soybeans LANDSAT-MSS classification performance as a function of scene characteristics. In: SYMPOSIUM ON MACHINE OF REMOTELY SENSED DATA WITH SPECIAL EMPHASIS ON CROP INVENTORY AND MONITORING, 8., West Lafayette, IN, July 7-9, 1982. *Proceedings*. West Lafayette, IN, Purdue University, 1982, p.178-187.
- CÂMARA, G.; VELASCO, F.R.D.; OLIVEIRA, M.O.B. *Um procedimento automático para classificação de áreas agrícolas em imagens de satélite, para uso em previsão de safras*. São José dos Campos, INPE, jul., 1982. (INPE-2482-PRE-166).
- DAUGHTRY, C.S.T.; HIXSON, M.M. Processing Techniques Development, Part I: Crop Inventory Techniques, Final Report, West Lafayette, IN. Purdue University, 1979, V.2 (LARS contract Report 112979).
- DAVIDSON, S.A. *Segment-level evaluation of the simulated aggregation test: U.S. corn and soybean exploratory experiment*. Huston, 1980. (Ag-RISTARS ISC-16820).

- HIXSON, M.M.; BAUER, M.E.; SCHOLZ, D.K. An assessment of data acquisition history on identification and area estimation of corn and soybeans. In: SYMPOSIUM ON MACHINE PROCESSING OF REMOTELY SENSED DATA, 6., West Lafayette, IN., June 3-6, 1980. *Proceedings*. West Lafayette, IN., Purdue University, 1980, p 72-77.
- KALENSKY, Z.; WIGHTJAN, J.M. Automatic forest mapping using remotely sensed data. In: INTERNATIONAL UNION OF FORESTRY RESEARCH ORGANIZATIONS. *Remote sensing in forestry*; Proceedings of the Symposium held during the XVI IUFRO world Congress, Oslo, 21-26 June 1976. Freiburg, University Freiburg, 1976, p.115-135.
- MENDONÇA, F.J.; LEE, D.C.L.; TARDIN, A.T.; SHIMABUKURO, Y.E.; CHEN, S.C.; LUCHT, L.A.M.; MOREIRA, M.A.; LIMA, A.M.; MAIA, F.C.S. *Levantamento da área canavieira do Estado de São Paulo, utilizando dados do LANDSAT ano safra 1979/80, VI*. São José dos Campos, INPE, mar., 1981. (INPE-2021-RPE/288).
- MENDONÇA, F.J. *Uso de dados do LANDSAT para avaliação de áreas ocupadas com cana-de-açúcar no Estado de São Paulo*. São José dos Campos, INPE, abr., 1980. (INPE-1713-TDL/923).