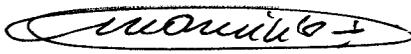


1. Publicação nº <i>INPE-3359-PRE/637</i>	2. Versão	3. Data <i>novembro, 1984</i>	5. Distribuição <input type="checkbox"/> Interna <input checked="" type="checkbox"/> Externa <input type="checkbox"/> Restrita
4. Origem <i>DDS/DTM</i>	Programa <i>TRANSF</i>		
6. Palavras chaves - selecionadas pelo(s) autor(es) <i>SENSORIAMENTO REMOTO SOLO PREPARADO</i> <i>ANÁLISE DIGITAL</i>			
7. C.D.U.: <i>528.711.7</i>			
8. Título <i>INPE-3359-PRE/637</i>		10. Páginas: <i>12</i>	
ESTUDO COMPARATIVO DA ANÁLISE DIGITAL DE ÁREAS DE SOLO PREPARADO PARA PLANTIO UTILIZANDO DIFERENTES ALGORITMOS DE CLASSIFICAÇÃO		11. Última página: <i>11</i>	
		12. Revisada por  Mário Valério Filho	
9. Autoria <i>Maurício Alves Moreira</i> <i>Getúlio Vargas de Assunção</i> <i>Antonio Roberto Formaggio</i> <i>Theлма Krug de Morais</i>		13. Autorizada por  Nelson de Jesus Parada Diretor Geral	
Assinatura responsável 			
14. Resumo/Notas <p><i>Os objetivos deste trabalho são: a) avaliar o desempenho de classificação obtidos pelos classificadores Single-Cell, Maxver e K-Médias em áreas de solo preparado para plantio; b) verificar a viabilidade do uso do método de pós-processamento chamado Uniformização de Temas-opção UNITOT sobre os resultados da classificação automática. A metodologia empregada consistiu de um estudo estatístico sobre os resultados de classificação correta e de erro de inclusão, obtidos de uma amostra selecionada aleatoriamente sobre mapas alfanuméricos, os quais continham a distribuição espacial de áreas de solo preparado obtidas de cada classificador proposto. De acordo com os resultados da pesquisa pôde-se concluir que: a) para as mesmas condições de treinamento o K-Médias foi o que apresentou pior desempenho de classificação de áreas de solo preparado; b) o Single-Cell não diferiu estatisticamente do Maxver; c) com relação ao uso do método de pós-processamento apenas as combinações de limiar-peso (5,3) (5,2) e (4,2) apresentaram resultados estatisticamente inferiores tanto para aqueles obtidos nas demais combinações selecionadas quanto para os resultados obtidos pelo Maxver sem uniformização de temas; e d) nenhuma combinação de limiar-peso, selecionada para esta pesquisa, quando usada no método de pós-processamento, mostrou melhoria nos resultados de classificação de áreas de solo preparado.</i></p>			
15. Observações <i>Este trabalho será publicado nos Anais da 4ª Reunião Anual da SELPER, Santiago, Chile, 12-17 de novembro de 1984.</i>			

ESTUDO COMPARATIVO DA ANÁLISE DIGITAL DE ÁREAS DE SOLO PREPARADO PARA PLANTIO UTILIZANDO

DIFERENTES ALGORITMOS DE CLASSIFICAÇÃO

Maurício Alves Moreira (1)\*  
Getúlio Vargas de Assunção (2)\*  
Antonio Roberto Formaggio (3)\*  
Thelma Krug de Moraes (4)\*\*

R E S U M O

Os objetivos deste trabalho são: a) avaliar o desempenho de classificação automática obtida pelos classificadores Single-Cell, Maxver e K-Médias em áreas de solo preparado para plantio; b) selecionar, para uma filtragem espacial, os parâmetros limiar-peso a serem utilizados no método de pós-processamento chamado Uniformização de Temas-opção (UNITOT), bem como estudar o seu efeito sobre os resultados da análise digital obtidos através do classificador selecionado. Para esta pesquisa selecionou-se uma área de estudo com aproximadamente 400 km<sup>2</sup>, a nordeste do Estado de São Paulo, localizada na Divisão Regional Agrícola (DIRA) de Ribeirão Preto. A metodologia empregada, tanto para avaliar o desempenho dos classificadores quanto para estudar o efeito do método de pós-processamento, consistiu de um estudo estatístico sobre os resultados de classificação correta e de erros de inclusão, obtidos de uma amostra selecionada aleatoriamente sobre mapas alfanuméricos, os quais continham a distribuição espacial de áreas de solo preparado obtidas de cada classificador proposto e de cada combinação limiar-peso selecionada. Para a análise estatística aplicaram-se os testes de análise variância e de Newman-Keuls. De acordo com os resultados da pesquisa pôde-se concluir que: a) para as mesmas condições de treinamento o K-Médias foi o que apresentou pior desempenho de classificação de áreas de solo preparado; b) o Single-Cell não diferiu estatisticamente do Maxver. No entanto, recomenda-se a utilização do Maxver por ser um algoritmo probabilístico, versátil (analisa até 32 classes ao mesmo tempo), bastante interativo e, não apresenta problemas de sobreposição de classificação, o que pode ocorrer com o Single-Cell; c) com relação ao uso do método de pós-processamento, sobre os resultados da análise digital de áreas de solo preparado, apenas as combinações de limiar-peso (5,3), (5,2) e (4,2) apresentaram resultados estatisticamente inferiores tanto para aqueles obtidos nas demais combinações selecionadas quanto para os resultados obtidos pelo Maxver sem uniformização de temas; e d) nenhuma combinação de limiar-peso, selecionada para esta pesquisa, quando usada no método de pós-processamento, mostrou melhoria nos resultados de classificação de áreas de solo preparado.

---

Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq  
Instituto de Pesquisas Espaciais - INPE  
C.P. 515, 12.200 - São José dos Campos - SP - Brasil  
\* Engenheiro Agrônomo  
\*\* Estatística

## 1. INTRODUÇÃO

A análise digital de dados multiespectrais é definida como a utilização de sistemas computacionais que visam auxiliar o elemento humano na interpretação das informações neles contidas.

Para o uso adequado da análise digital é aconselhável que se estabeleça primeiramente uma metodologia de interpretação de tal forma que, integrando a capacidade do homem com a agilidade da máquina, se possa obter de modo mais rápido informações sobre os alvos contidos numa determinada área de interesse.

No Sistema Interativo de Análise de Imagens Multiespectrais (IMAGE-100) há vários classificadores disponíveis ao usuário, bem como métodos de pré e pós-processamento. Porém, três deles são os mais utilizados na área de recursos naturais: o Single-Cell, o Maxver e o K-Médias.

O analista de dados frequentemente tem dificuldade em decidir qual desses classificadores se destaca como melhor, no sentido da quantidade e qualidade da informação sobre determinado alvo de interesse.

Deve-se ter em mente que a seleção do melhor classificador por si só não assegura a ausência de certas grandezas de erros, de inclusão ou de omissão, durante a análise digital. Estes erros causados pela semelhança de respostas espectrais podem ser, às vezes, minizados mediante a utilização de um método de pós-processamento.

Este trabalho tem como objetivos: a) avaliar o desempenho da classificação automática obtida pelos classificadores Single-Cell, Maxver e K-Médias em áreas de solo preparado para plantio; b) selecionar para uma filtragem espacial, os parâmetros (limiar-peso) a serem utilizados no método de pós-processamento chamado *Uniformização de Temas-Opção* (UNITOT), bem como estudar o seu efeito sobre os resultados da análise digital, obtidos pela utilização do classificador selecionado.

## 2. MATERIAL

### 2.1. ÁREA DE ESTUDO

Para este estudo selecionou-se uma área com aproximadamente 400 km<sup>2</sup>, localizada na Divisão Regional Agrícola (DIRA) de Ribeirão Preto, a nordeste do Estado de São Paulo, entre os paralelos 20° 45'S a 21° 00'S e 47° 55'W a 48° 30'W.

Os solos que predominam na área, de acordo com Brasil (1960), são os Latossolos Roxos.

Os Latossolos Roxos são originários de material proveniente de rochas básicas, principalmente basaltos e diabásios. São caracterizados pela cor arroxeada e pela alta friabilidade que apresenta ao longo do perfil. Estes solos ocorrem em 175 municípios do Estado de São Paulo (Figura 1).

### 2.2. FOTOGRAFIAS AÉREAS

Em 06 de setembro de 1979 realizou-se, sobre a área de estudo, um sobrevôo utilizando a câmera métrica do tipo RC-10, com distância focal de 152mm e filme Aerochrome 2443. O vôo foi realizado com recobrimento lateral e longitudinal de 20%. Transparências positivas foram obtidas na escala de 1:20.000.

Estas fotografias aéreas serviram como suporte para identificar áreas de solo preparado durante a análise digital usando os dados do LANDSAT e também para verificar o desempenho dos três classificadores.

### 2.3. DADOS DO LANDSAT

Na análise digital de áreas preparadas para plantio foram utilizados os dados digitalizados do MSS (Multispectral Scanner System) do LANDSAT-2, correspondentes à órbita (WRS) 236, ponto 75 da passagem de 01/09/1979. Os dados do LANDSAT apresentaram uma defasagem de 5 dias em relação ao aerolevantamento.

A análise digital foi feita através do IMAGE-100 (General Electric, 1975).

### 3. METODOLOGIA

Os algoritmos implementados no IMAGE-100 e testados no presente estudo são descritos a seguir:

#### 3.1. MÉTODO DO PARALELEPÍPEDO (Single-Cell)

O Single-Cell é um algoritmo do tipo determinístico. Uma vez selecionadas as áreas de treinamento, as suas respostas espectrais são utilizadas para criar um paralelepípedo de quatro dimensões, sendo que cada lado dele corresponde à banda de cada canal.

Determinam-se os limites máximo e mínimo dos valores que cada um dos atributos (canais) dos pontos selecionados da imagem assumem. No caso de imagens, os atributos são os níveis de cinza da reflectância média da área de treinamento para cada canal.

Um "pixel" desconhecido é classificado como pertencente a uma classe conhecida se os níveis de cinza (respostas espectrais) desse "pixel", nos quatro canais, estiverem contidos no paralelepípedo criado a partir das áreas de treinamento da classe conhecida (General Electric, 1975).

#### 3.2. SISTEMA MAXVER

O Maxver é um algoritmo supervisionado, paramétrico, que destina-se à classificação ponto-a-ponto de imagens multiespectrais.

O critério usado na classificação é o de máxima verossimilhança (maximum likelihood), segundo classes escolhidas interativamente pelo analista.

Este sistema tem a capacidade de analisar até 32 classes por vez, com um máximo de 160 tomadas de área de treinamento (Velasco et alii, 1978).

#### 3.3. K-MÉDIAS (clustering)

O K-Médias é um algoritmo não-supervisionado sobre o qual o analista tem pouco controle em relação às regras de decisão do classificador.

A diferença básica deste procedimento com as técnicas de classificação supervisionada é o fato de as classes não serem definidas pelo analista.

A partir de um conjunto de pontos (amostra), do número de centros iniciais (vetor média inicial para cada classe), da precisão e do número de iterações que se deseja, o sistema define as classes (grupos) e associa cada "pixel" ao centro mais próximo, segundo a função discriminante *Distância Euclidiana*.

Para verificar o efeito do uso do método de pós-processamento sobre os resultados da classificação automática de solo preparado, fez-se uma filtragem espacial utilizando o método de Uniformização de Temas-opção (UNITOT).

#### 3.4. UNIFORMIZAÇÃO DE TEMAS-OPÇÃO (UNITOT)

O método consiste em uma filtragem espacial que analisa o "pixel" central de uma matriz de 3 X 3 "pixels". Para a análise, dois parâmetros devem ser considerados: Limiar (L) e Peso (P). Estes são predeterminados pelo analista.

Inicialmente, é determinada a frequência de ocorrência de cada classe no interior da matriz, na qual a classe do "pixel" central é contada P vezes. O valor da frequência da classe de maior ocorrência é comparada com o limiar. Se o valor do limiar for menor, o "pixel" central será substituído pela classe de maior ocorrência dentro da matriz. Se o valor do limiar for maior, a classe do "pixel" central permanece sem alteração.

Além de estudar o efeito do método UNITOT sobre os resultados obtidos pela análise digital de solo preparado, procurou-se também determinar a melhor combinação limiar-peso. Utilizaram-se as seguintes combinações seguintes: (2,2), (2,3), (2,4), (2,5), (3,2), (3,3), (3,4), (3,5), (4,2), (4,3), (4,4), (4,5), (5,2), (5,3), (5,4) e (5,5).

### 3.5. ANÁLISE DIGITAL E ESTATÍSTICA

Com base em fotografias aéreas foi delimitada a área de estudo sobre os dados digitalizados do MSS do LANDSAT, mostrados no visor do IMAGE-100.

A fita que continha os dados do MSS do LANDSAT apresentou um defeito denominado "stripping" no canal 4. Segundo Cavalcanti (1975) o "stripping" é o aparecimento de pequenas faixas na imagem causadas por variações no ganho e "offset" dos detetores de um canal do MSS. Devido a este problema, o canal 4 foi excluído da análise e passou a executar os algoritmos Single-Cell, Maxver e o K-Médias com apenas três canais (5,6 e 7).

Com auxílio do cursor foram obtidas áreas de treinamento em talhões conhecidos e representativos de cada alvo de ocupação do solo a fim de realizar a análise digital. Criou-se, desta forma, um arquivo com as informações espectrais destas áreas, o que tornou possível a obtenção dos parâmetros estatísticos requeridos pelos três classificadores para a realização da classificação propriamente dita.

Com relação à análise digital que utiliza o Single-Cell, somente as áreas de treinamento de solo preparado foram necessárias na obtenção dos parâmetros de treinamento requeridos para a classificação, tendo em vista que este classificador analisa uma classe por vez.

A fim de poder comparar o desempenho de classificação obtido pelo K-Médias, que é um algoritmo não-supervisionado, com o desempenho obtido pelos classificadores supervisionados (Single-Cell e Maxver), utilizou-se o mesmo arquivo de áreas de treinamento usado para os outros classificadores na determinação dos parâmetros (vetor média inicial de cada classe) como o conjunto de ponto amostral.

Uma vez classificada a área de estudo, usando cada classificador, os resultados foram impressos em mapas alfanuméricos (printout) que contém a distribuição espacial somente de áreas de solo preparado. Desta forma, obteve-se para cada resultado de classificação um mapa alfanumérico na escala de 1:20.000, compatível com a mesma escala das fotografias aéreas.

Pela sobreposição do mapa alfanumérico em um mosaico confeccionado a partir de interpretação visual das fotos aéreas, fez-se a delimitação das áreas de solo preparado.

Para comparar quantitativamente os desempenhos dos classificadores, fez-se um estudo estatístico para os dados de classificação correta (CC)<sup>1</sup> e de erros de inclusão (EI)<sup>2</sup>, obtidos a partir de 30 segmentos de 10cm X 10cm (4 km<sup>2</sup> cada um no terreno) selecionados aleatoriamente. A localização destes segmentos foi fixada para os mapas alfanuméricos obtidos pelos três classificadores. Os dados de CC e EI foram submetidos aos testes de: a) Bartlett para a homogeneidade de variância; b) análise de variância; e c) Newman-Keuls (Zar, 1974).

A fim de verificar a viabilidade do método de pós-processamento UNITOT, escolheram-se subjetivamente os resultados de classificação, obtidos pelo algoritmo Maxver, e sobre eles aplicou-se o método UNITOT, variando a combinação limiar-peso. Os resultados obtidos de cada combinação foram também impressos em forma de mapas alfanuméricos na escala de 1:20.000. Desta maneira, foram obtidos 16 mapas alfanuméricos, os quais contém a distribuição espacial de solo preparado.

Da mesma forma da seleção dos classificadores, fez-se um estudo estatístico para os dados de CC e EI, obtidos a partir de uma amostra de 6 segmentos de 20cm X 20cm (16 km<sup>2</sup> cada um no terreno), selecionados aleatoriamente. A localização destes segmentos foi a mesma para todos os mapas alfanuméricos obtidos das 16 combinações de limiar-peso.

Utilizou-se o resultado obtido pelo Maxver sem a uniformização de temas como referência (*teste munha*) para o estudo do método de pós-processamento e seleção da combinação de limiar-peso. Desta forma, a análise estatística foi realizada sobre 17 tratamentos e os testes estatísticos empregados foram a análise de variância e o de Newman-Keuls.

---

<sup>1</sup> CC =  $\frac{\text{área de solo preparado classificada corretamente pelo IMAGE-100}}{\text{área de solo preparado obtida através de fotografias aéreas}}$

<sup>2</sup> EI =  $\frac{\text{área de outros alvos classificada como solo preparado pelo IMAGE-100}}{\text{área de solo preparado estimada pelo IMAGE-100}}$

#### 4. RESULTADOS E DISCUSSÃO

Além do solo preparado, outros temas interpretados a partir de fotografias aéreas foram: *cana nova, cana cortada, cana intermediária, cana adulta, pastagem, áreas com drenagem deficiente e água.*

A interpretação de todos os alvos da área de estudo teve como objetivo verificar quais deles poderiam confundir-se espectralmente com o solo preparado, nos dados do LANDSAT.

Uma análise visual das imagens do MSS do LANDSAT mostrou que os temas cana nova e cana cortada, embora facilmente discrimináveis do solo preparado em nível de aeronave, teriam de ser incluídos como solo preparado em nível orbital para a análise de uma única data. Isto porque as partes aéreas da cana nova e o palhiço da cana cortada cobriam pouco a superfície do solo; não chegando porém, a influenciar espectralmente. Tal opinião é consistente com Assunção e Duarte (1983) em estudo com a mesma finalidade.

A Tabela 1 contém, para os segmentos amostrados, os dados de densidade de classificação de solo preparado, de áreas obtidas através da interpretação visual de fotografias aéreas, de áreas classificadas corretamente pelos três classificadores a partir do LANDSAT e de áreas de outros alvos que foram classificados como solo preparado (erro de inclusão).

A densidade de classificação do solo preparado é a relação percentual entre a área desta classe obtida nas fotos aéreas e a área total de cada segmento (400 ha cada).

Conforme pode ser observado na Tabela 1, a densidade de classificação de solo preparado variou de 4,42 a 83,78%. Isto mostra que para esta data (01/09/1979) de passagem do LANDSAT (início de preparo do solo), apesar de a área de estudo se situar numa região de intensa atividade agrícola, apresentou variações na quantidade de solo preparado para os segmentos amostrados.

A tabela 2 contém os dados percentuais de classificação correta (CC), de erro de inclusão (EI) e de diferença relativa (DR)<sup>3</sup>, obtidos com a utilização dos três classificadores selecionados.

Pela análise da Tabela 2 observa-se que, de maneira geral, a classificação correta foi maior quando se empregou o classificador MAXVER e que o K-Médias foi o algoritmo que apresentou os mais baixos índices de CC e de EI.

Uma das possíveis explicações de o K-Médias ter apresentado valores menores de CC e de EI poderia estar relacionada com a decisão do algoritmo para a definição das classes (Distância Euclidiana). Desta forma, as classes definidas pelo analista, a partir das áreas de treinamento podem não responder àquelas definidas pelo K-Médias. Em outras palavras, este algoritmo usa estas áreas adquiridas pelo analista (áreas de treinamento) como uma amostra e, a partir do fornecimento deste conjunto de pontos, do número de centros iniciais e da precisão, ele define novas classes que não respondem exatamente àquelas caracterizadas de maneira supervisionada.

Por exemplo, algumas áreas que na realidade eram de solo preparado, as quais apresentaram uma pequena diferença espectral do solo típico, não foram classificadas pelo K-Médias como pertencentes a esta classe; ou seja, tais áreas foram agrupadas em outras classes. Assim, a percentagem de classificação correta diminuiu. Por outro lado, áreas de pastagem degradada e de drenagem deficiente (com solo exposto), as quais foram classificadas pelo SINGLE-CELL e pelo MAXVER como solo preparado, não foram classificadas pelo K-MÉDIAS, diminuindo assim o erro de inclusão.

A classificação correta de alguns segmentos foi prejudicada pela defasagem de cinco dias entre a obtenção dos dados do satélite (01/09/1979). Algumas áreas apresentaram-se diferentemente nas duas datas de aquisição dos dados. Áreas que foram preparadas para plantio, ou áreas com cana adulta que foram cortadas neste intervalo, foram identificadas nas fotos aéreas como solo preparado ou cana cortada; ao passo que nos dados do satélite não foram classificadas como pertencentes a este tema. Estas áreas aumentaram o erro de omissão diminuindo, em consequência, a classificação correta.

---

<sup>3</sup>DR =  $\left( \frac{\text{área estimada pelo IMAGE-100}}{\text{área estimada através de foto aérea}} - 1 \right) \times 100$

#### 4.1. ANÁLISE ESTATÍSTICA

A análise estatística foi efetuada tanto para os dados de classificação correta quanto para os de erros de inclusão, contidos na Tabela 1, para os três classificadores propostos.

Segundo ZAR (1974), para aplicar a análise de variância, as variâncias estimadas das amostras obtidas sobre os resultados dos diferentes classificadores devem ser iguais e todas as amostras devem provir de populações normais. Desta forma, empregou-se o teste de Bartlett para verificar a homogeneidade de variância dos dados de CC e EI.

Com a aplicação do teste de Bartlett para homogeneidade de variância, não foi possível rejeitar a hipótese de igualdade de variância ao nível de significância de 5%.

Aplicou-se, a seguir, a análise de variância aos dados de CC e EI, que forneceram os resultados mostrados nas Tabelas 3 e 5.

Obteve-se para classificação correta um valor  $F = 20,04$ , que foi significativo ao nível de 5% de probabilidade. Desta forma, rejeita-se a hipótese de que a classificação correta obtida dos três classificadores não diferiram estatisticamente entre si. A rejeição da hipótese não implica, entre tanto, que todas as médias sejam diferentes uma das outras. A análise de variância não fornece elementos que tornem possível a localização da diferença dentre as três médias populacionais.

Uma vez rejeitada a hipótese de igualdade das médias, aplicou-se o teste de variação múltipla (multiple range test) de Newman-Keuls, conforme ZAR (1974).

A aplicação do teste de Newman-Keuls para detecção de quais médias, dentre as possíveis, eram diferentes entre si (Tabela 4), indicou que Single-Cell foi igual ao Maxver e que o K-Médias foi diferente do Single-Cell e do Maxver.

Com relação ao erro de inclusão, a análise de variância (Tabela 5) forneceu valor  $F = 16,56$ , de forma a permitir a rejeição da hipótese de igualdade de médias ao nível de significância de 5%.

Da mesma forma, como se procedeu com relação à classificação correta, foi empregado o teste de Newman-Keuls, apresentado na Tabela 6, que forneceu o seguinte resultado: K-Médias foi diferente do Single-Cell, o Single-Cell foi igual ao Maxver.

Pelos resultados dos testes estatísticos efetuados nota-se que são poucas as inferências que se pode fazer com respeito ao desempenho dos três classificadores em questão.

Por outro lado, observando a diferença relativa (DR) obtida para cada segmento amostrado (Tabela 2), em termos de valores absolutos, pode-se afirmar o seguinte: a) para os casos observados, as diferenças relativas obtidas para o MAXVER foram superiores às do Single-Cell e às do K-Médias em cerca de 63% e 60%, respectivamente; b) em 60% dos casos observados o Single-Cell foi pior do que o K-Médias.

A diferença relativa média obtida da amostra para os três classificadores foram: (+ 3,99%) para o Single-Cell, (+ 3,22%) para o MAXVER e (- 12,08%) para o K-Médias.

Com o objetivo de verificar a viabilidade do método de pós-processamento, aplicou-se o UNITOT sobre os resultados de classificação obtidos pelo MAXVER, variando a combinação limiar-peso, conforme consta na Seção 3.4.

A análise estatística foi efetuada sobre os dados de CC e EI, obtidos de uma amostra de 6 segmentos, em cada um dos "printouts" e os testes estatísticos aplicados foram a análise de variância e de Newman-Keuls. Para a aplicação dos testes utilizaram-se, como piloto, os resultados de classificação do MAXVER sem uniformização de temas.

A hipótese de igualdade de médias para os dados de classificação correta foi rejeitada ( $F = 11,02$ ) ao nível de significância de 5% com a aplicação da análise de variância (Tabela 7).

Conforme foi mencionado anteriormente, a análise de variância não fornece elementos que tornem possível a localização das diferenças dentre as 17 médias populacionais. Para isto, aplicou-se o teste de variação múltipla de Newman-Keuls para detecção de quais médias dentre as possíveis, eram diferentes entre si.

A aplicação dos testes de Newman-Keuls indicou que somente as combinações de limiar-peso, (5,2), (5,3) e (4,2), diferiram estatisticamente das demais ao nível de significância de 5%. Em outras palavras, estas três combinações foram as que apresentaram os piores resultados de classificação, conforme pode ser observado na Tabela 8. No entanto, as demais combinações não alteraram estatisticamente os resultados da classificação, quando comparadas com o MAXVER sem a uniformização de temas.

A hipótese de igualdade de médias para os dados de erros de inclusão não pode ser rejeitada ( $F = 0,48$ ) ao nível de significância de 5% (Tabela 9). Desta forma, o teste de Newman-Keuls não foi aplicado tendo em vista que não se observou diferença significativa entre as 17 médias testadas.

## 5. CONCLUSÕES

De acordo com os resultados desta pesquisa chegou-se às seguintes conclusões:

- 1) Nos levantamentos de solo preparado para plantio, em regiões com intensa atividade agrícola, como é o caso da área de estudo, as informações coletadas no campo ou por aeronave devem ser obtidas na mesma data de aquisição dos dados do satélite. A defasagem na coleta dos dados pode acarretar: a) um aumento no erro de inclusão quando os dados de campo e/ou de aeronave forem obtidos antes da data de passagem do satélite; b) um decréscimo da classificação correta quando os dados de campo e/ou de aeronave forem obtidos depois da data da passagem do satélite.
- 2) Os temas que apresentaram maior confusão espectral com áreas de solo preparado para plantio foram as áreas com pastagem degradada e as de drenagem deficiente com vegetação rala.
- 3) Não foi possível separar as áreas com cana nova e cana cortada daquelas de solo preparado para plantio, utilizando dados do MSS/LANDSAT de uma única data.
- 4) Em termos de valores absolutos, para os casos observados, as diferenças relativas obtidas pelo uso do MAXVER foram superiores às obtidas pelo Single-Cell e pelo K-Médias em cerca de 63% e 60%, respectivamente.
- 5) Em 60% dos casos observados, as diferenças relativas obtidas pelo Single-Cell foram piores do que aquelas determinadas pelo uso do K-Médias.
- 6) Usando as mesmas áreas de treinamento, o K-Médias foi o algoritmo que apresentou o pior desempenho de classificação.
- 7) As áreas de solo preparado para plantio apresentaram um alto contraste espectral em relação aos alvos vizinhos. Assim, as áreas de treinamento adquiridas sobre este tema foram bastante representativas e isto fez com que não houvesse uma superioridade marcante entre os classificadores Single-Cell e o Maxver.
- 8) Embora o Maxver tenha apresentado um desempenho de classificação tão bom quanto o Single-Cell, recomenda-se a sua utilização tendo em vista que este algoritmo é probabilístico, mais versátil (analisada até 32 classes ao mesmo tempo), bastante interativo e não apresenta o problema de sobreposição de classificação, o que ocorre com o Single-Cell.
- 9) Das combinações de limiar-peso selecionadas para esta pesquisa, somente as combinações (5,3), (5,2) e (4,2) foram estatisticamente diferentes das demais. No entanto, as demais combinações não alteraram estatisticamente os resultados da classificação, quando comparadas com o Maxver sem a uniformização de temas.

## 6. BIBLIOGRAFIA

- ASSUNÇÃO, G.V.; DUARTE, V. *Avaliação de áreas preparadas para plantio (solo nu) utilizando dados do satélite LANDSAT*. São José dos Campos INPE, 1983. (INPE-2637-TDL/113).
- BRASIL. Ministério da Agricultura. Centro de Pesquisas Agronômicas. *Levantamento de reconhecimento dos solos do Estado de São Paulo*. Rio de Janeiro, 1960. 630p. (Boletim 12).
- CAVALCANTI, L.A. *Correção radiométrica dos dados enviados pelo sistema MSS do satélite ERTS*. São José dos Campos, INPE, dez. 1975. (INPE-786-NTI/039).

GENERAL ELECTRIC COMPANY. *Image-100 - Interactive Multispectral Image Analysis System*. Daytona, 1975.

VELASCO, F.R.D.; PRADO, L.O.C.; SOUZA, R.C.M. *Sistema MAXVER: manual do usuário*. São José dos Campos, INPE, jul. 1978. (INPE-1315-NTI/110).

ZAR, J.H. *Two-sample hypotheses*. In: *Biostatistical analysis*. Englewood Cliffs, N.J.; Prentice-Hall, 1974. cap. 12; p.151-181.

SEGMENTO (n)	ÁREA OBTIDA DE FOTO AÉREA (ha)	DENSIDADE DE SOLO PREPARADO (%)	ÁREA DE SOLO PREPARADO CLASSIFICADA CORRE- TAMENTE A PARTIR DOS DADOS DO LANDSAT (ha)			ÁREA DE OUTROS ALVOS CLASSIFICADA COMO SOLO PREPARADO (ha)		
			SINGLE-CELL	MAXVER	K-MÉDIAS	SINGLE-CELL	MAXVER	K-MÉDIAS
1	258,41	64,40	203,19	215,07	200,95	22,47	25,03	5,78
2	150,55	37,64	132,57	130,97	114,27	10,27	3,21	2,56
3	93,73	23,43	88,92	88,92	79,28	43,98	32,42	19,26
4	182,00	45,50	163,39	156,01	122,30	56,17	40,12	19,58
5	335,12	83,78	327,10	332,88	322,93	22,79	20,54	12,19
6	69,98	17,49	58,25	61,31	54,25	12,84	7,70	1,61
7	254,87	63,72	217,64	215,71	215,71	71,26	51,36	26,64
8	186,50	46,62	165,96	164,99	156,01	37,87	35,31	20,22
9	285,37	71,34	269,96	277,88	271,88	9,95	10,59	15,09
10	122,94	30,73	83,14	102,72	95,02	19,26	22,79	12,52
11	180,08	45,02	138,64	139,64	117,80	31,46	30,81	28,89
12	234,65	58,66	165,96	178,48	167,24	8,67	9,63	9,63
13	17,66	4,42	13,48	13,80	11,55	19,90	11,87	1,93
14	96,94	24,23	87,95	83,14	76,39	4,81	2,56	1,61
15	204,48	51,12	170,45	176,23	164,99	18,94	19,58	4,17
16	275,74	68,93	256,80	243,32	225,98	17,98	23,75	19,58
17	90,52	22,63	69,98	76,40	62,59	40,77	34,34	23,11
18	258,08	64,52	249,10	248,45	244,92	14,13	24,07	28,25
19	103,36	25,84	95,98	96,62	93,73	16,37	24,71	11,23
20	208,97	52,24	132,57	146,38	110,10	8,67	9,30	5,14
21	150,23	37,55	131,61	131,93	129,36	15,73	20,86	20,86
22	170,77	42,69	162,10	163,39	157,29	11,88	18,29	14,12
23	187,46	46,86	128,40	126,15	120,05	27,93	33,06	28,25
24	253,59	63,39	232,08	236,58	226,95	6,42	5,70	2,00
25	192,60	48,15	181,69	178,80	170,77	81,53	60,05	35,31
26	159,86	39,96	137,71	151,19	133,85	38,52	40,76	15,09
27	89,56	22,39	69,66	62,27	48,15	11,88	9,36	7,06
28	162,43	40,61	147,34	144,13	140,92	22,47	24,72	12,84
29	211,22	52,80	208,33	205,76	198,05	29,53	38,52	27,93
30	305,59	76,40	243,32	290,32	248,13	20,22	22,47	16,69

TABELA 1 - DENSIDADE DE SOLO PREPARADO, ÁREAS OBTIDAS A PARTIR DE FOTOS AÉREAS E DO LANDSAT E ERRO DE INCLUSÃO DETERMINADOS PARA OS SEGMENTOS AMOSTRADOS.

SEGMENTO (n)	CLASSIFICAÇÃO CORRETA(%)			ERRO DE INCLUSÃO			DIFERENÇA RELATIVA ENTRE AS ÁREAS ESTIMADAS PELO I-100 E AS ESTIMADAS POR FOTO AÉREAS(%)		
	SINGLE-CELL	MAXVER	K-MÉDIAS	SINGLE-CELL	MAXVER	K-MÉDIAS	SINGLE-CELL	MAXVER	K-MÉDIAS
1	78,63	83,23	77,76	9,95	10,42	2,79	-12,67	- 7,08	- 19,99
2	88,06	86,99	75,90	7,19	2,39	2,19	- 5,12	-10,87	- 22,39
3	94,17	94,86	84,58	33,25	26,72	19,54	41,09	29,45	5,13
4	89,77	85,72	67,20	25,58	20,45	13,80	20,63	7,76	- 22,04
5	97,61	99,33	96,36	6,51	5,81	3,63	4,40	5,46	0,00
6	83,48	87,61	77,52	18,02	11,16	2,88	1,80	- 1,38	- 20,17
7	85,39	84,63	84,63	24,66	19,23	10,99	13,35	4,78	- 4,91
8	88,98	88,47	83,65	18,57	17,63	11,47	9,29	7,39	- 5,50
9	94,59	97,29	95,27	3,55	3,67	5,25	- 1,91	1,01	- 0,01
10	67,63	83,55	77,28	18,80	18,16	11,64	-16,70	2,09	- 12,52
11	77,18	77,54	65,41	18,45	18,07	19,69	- 5,34	- 5,34	- 18,54
12	70,73	76,06	71,27	4,96	5,12	5,44	-25,58	-19,83	- 24,62
13	76,33	78,14	65,40	59,62	46,24	14,31	89,01	45,35	- 23,66
14	90,72	85,76	78,80	5,18	2,99	2,06	- 4,31	-11,96	- 19,53
15	83,35	86,18	80,68	10,00	10,00	2,46	- 7,37	- 4,24	- 17,27
16	93,13	88,24	81,95	6,54	8,89	7,97	- 0,01	- 0,03	- 10,94
17	77,31	84,40	69,14	36,81	31,00	26,97	22,34	22,34	- 5,32
18	96,52	96,27	94,70	5,37	8,83	10,34	1,99	5,59	5,84
19	92,86	93,47	90,68	14,57	20,36	10,69	8,69	17,38	1,54
20	63,44	70,05	52,69	6,14	5,97	4,46	-32,14	-25,50	- 44,85
21	87,60	87,82	86,11	10,67	13,65	13,88	- 1,92	1,70	0,00
22	94,92	95,68	92,11	6,82	10,06	8,24	1,88	6,38	0,01
23	68,49	67,29	64,04	17,87	20,76	19,05	-16,60	-15,06	- 20,88
24	91,51	93,29	89,49	2,69	2,35	0,01	- 5,95	- 4,45	- 9,72
25	94,33	92,83	88,66	30,97	25,14	17,13	36,66	24,01	6,99
26	86,14	94,57	83,74	21,86	21,23	10,13	10,24	-20,07	- 4,95
27	77,78	69,53	53,76	14,57	13,07	12,78	- 8,95	20,02	- 38,35
28	90,71	88,73	86,75	13,23	14,64	8,35	4,54	3,95	- 5,34
29	98,63	97,41	93,76	12,41	15,76	12,36	12,61	15,65	6,98
30	79,62	95,17	81,19	7,67	7,18	6,30	-13,76	2,35	- 13,34
$\bar{x}$	85,32	87,80	79,68	15,74	14,56	9,80	3,99	3,23	- 12,08

TABELA 2 - DADOS DE CLASSIFICAÇÃO CORRETA, ERRO DE INCLUSÃO E DIFERENÇA RELATIVA DETERMINADOS PARA OS SEGMENTOS AMOSTRADOS.

CAUSA DE VARIACÃO	G.L.	S.Q.	Q.M.	F
TRATAMENTO	2	2.291,02	1.145,51	275,09*
SEGMENTO	29	461.436,84	15.911,61	19,80*
ERRO	58	3.354,79	56,29	
TOTAL	89	2.664.900,00		

TABELA 3 - ANÁLISE DE VARIÂNCIA PARA OS DADOS DE CLASSIFICAÇÃO CORRETA.

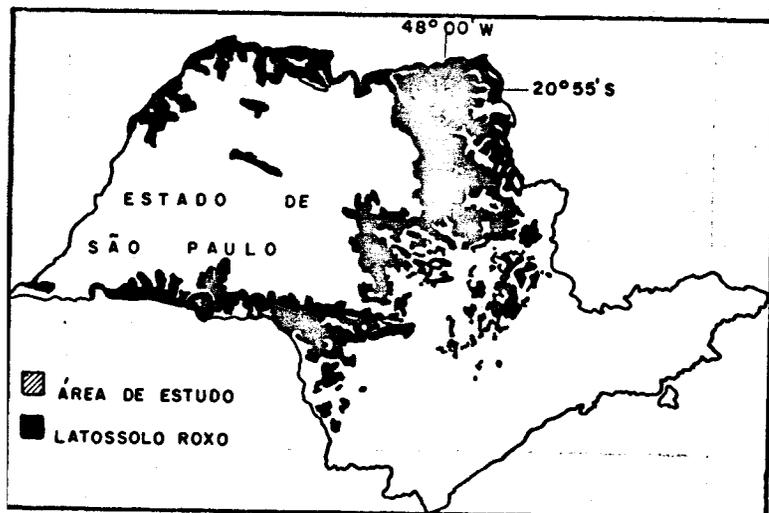


Fig. 1 - Localização da área de estudo e distribuição espacial dos Latossolos Roxos no Estado de São Paulo.  
Fonte: Brasil (1960)

COMPARAÇÃO	DIFERENÇA ( $\bar{X}_A - \bar{X}_B$ )	SE	q	P	$q_{0,05,58p}$	CONCLUSÕES
3 Vs. 1	11,94	1,36	8,78	3	3,39	REJEITA HO
3 Vs. 2	3,56	1,36	2,62	2	2,83	ACEITA HO
2 Vs. 1	8,38	1,36	6,16	2	2,83	REJEITA HO
CONCLUSÃO FINAL				1- K-MÉDIAS		
K-MÉDIAS ≠ SINGLE-CELL = MAXVER				2- SINGLE-CELL		
				3- MAXVER		

TABELA 4 - RESULTADOS DO TESTE DE NEWMAN-KEULS PARA OS DADOS DE CLASSIFICAÇÃO CORRETA.

CAUSA DE VARIACÃO	G.L.	S.Q.	Q.M.	F
TRATAMENTO	2	1.850,10	925,05	16,85*
SEGMENTO	29	15.317,27	528,18	9,62*
ERRO	58	3.183,66	54,89	
TOTAL	89	61.192,66		

TABELA 5 - ANÁLISE DE VARIÂNCIA PARA OS DADOS DE ERRO DE INCLUSÃO

COMPARAÇÃO	DIFERENÇA ( $\bar{X}_A - \bar{X}_B$ )	SE	q	P	$q_{0,05,58p}$	CONCLUSÕES
3 Vs. 1	10,19	1,36	7,49	3	3,39	REJEITA HO
3 Vs. 2	1,38	1,36	1,01	2	2,83	ACEITA HO
2 Vs. 1	8,81	1,36	6,47	2	2,83	REJEITA HO
CONCLUSÃO FINAL					1 - K-MÉDIAS	
K-MÉDIAS ≠ SINGLE-CELL = MAXVER					2 - SINGLE-CELL	
					3 - MAXVER	

TABELA 6 - RESULTADOS DO TESTE DE NEWMAN-KEULS PARA OS DADOS DE ERRO DE INCLUSÃO.

CAUSA DE VARIACÃO	G.L.	S.Q.	Q.M.	F
TRATAMENTO	16	302.776,37	18.923,52	11,02*
SEGMENTO	5	61.782.000,00	12.356.000,00	7.193,40*
ERRO	80	137.418,32	1.717,00	
TOTAL	102	571.470.000,00		

TABELA 7 - ANÁLISE DE VARIÂNCIA PARA OS DADOS DE CLASSIFICAÇÃO CORRETA.

CLASSIFICAÇÃO DAS MÉDIAS AMOSTRAIS EM ORDEM CRESCENTE DE MAGNITUDE																	
COMBINAÇÕES DE LIMITE-PESO																	
PARÂMETROS	$\bar{X}(5,2)$	$\bar{X}(5,3)$	$\bar{X}(4,2)$	$\bar{X}(3,5)$	$\bar{X}(4,3)$	$\bar{X}(4,5)$	$\bar{X}(5,4)$	$\bar{X}(0,0)$	$\bar{X}(4,4)$	$\bar{X}(2,5)$	$\bar{X}(3,4)$	$\bar{X}(5,5)$	$\bar{X}(3,3)$	$\bar{X}(2,3)$	$\bar{X}(3,2)$	$\bar{X}(2,4)$	$\bar{X}(2,2)$
	2047,33	2184,83	2194,47	2225,83	2220,33	2236,17	2238,50	2240,33	2744,83	2745,50	2750,17	2253,67	2257,50	2278,00	2278,83	2281,00	2291,67
k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$q_{0,05,60p}$	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,45	4,65	4,73	4,81	4,88	4,94	5,00	5,06	5,11	5,16
$R_p$	47,88	57,53	63,28	67,34	70,39	72,92	75,12	76,98	78,68	80,03	81,38	82,57	83,58	84,60	85,61	86,46	87,31
CONCLUSÃO FINAL																	
	$\bar{X}(5,2) \bar{X}(5,3) \bar{X}(4,2)$		$\bar{X}(3,5) \bar{X}(4,3) \bar{X}(4,5) \bar{X}(5,4) \bar{X}(0,0) \bar{X}(4,4) \bar{X}(2,5) \bar{X}(3,4) \bar{X}(5,5) \bar{X}(3,3) \bar{X}(2,3) \bar{X}(3,2) \bar{X}(2,4) \bar{X}(2,2)$														

TABELA 8 - RESULTADOS DO TESTE DE NEWMAN-KEULS PARA OS DADOS DE CLASSIFICAÇÃO CORRETA.

CAUSA DE VARIACÃO	G.L.	S.Q.	Q.M.	F
TRATAMENTO	16	140.113,51	8.757,09	0,48 n.s
ERRO	86	1.538.700,00	18.102,00	
TOTAL	102	10.161.000,00		

TABELA 9 - ANÁLISE DE VARIÂNCIA PARA OS DADOS DE ERRO DE INCLUSÃO.