

## Mapeamento da cultura da soja com imagens Landsat 5/TM utilizando algoritmos de classificação supervisionada

Juliana Rezeck Ganan<sup>1</sup>  
Jansle Vieira Rocha<sup>1</sup>  
Erivelto Mercante<sup>1</sup>  
João Francisco Gonçalves Antunes<sup>1</sup>

<sup>1</sup> Faculdade de Engenharia Agrícola – FEAGRI/UNICAMP  
Caixa Postal 6041 - 13083-875 – Campinas - SP, Brasil  
{julianar, jansle, erivelto.mercante, joao.antunes }@agr.unicamp.br

**Abstract.** The goal of this study was to produce a map of soybean cultivation areas in west region of Paraná state, through LANDSAT5/TM (Thematic Mapper) images from the 2003/2004 yield, from november 2003 to february 2004. In order to do this, there were made automatic digital classifications, which were based on regions of interest made before, in a study of monitoring that has been made in Cascavel-PR city. With the overlay of the maps generated by the classifications, it was possible to create a map with the soybean area in this region, and then to compare it with official data and selected points in the area; and also prepare a localization reference of soybean areas for the use of AVHRR/NOAA and MODIS images, of the same region, in the future.

**Palavras-chave:** crop monitoring, images digital classification, geotechnology, monitoramento agrícola, classificação digital de imagens, geotecnologia.

### 1. Introdução

Dentre os grandes produtores mundiais de soja, o Brasil figura como o país que apresenta as melhores condições de expandir a produção e prover o esperado aumento da demanda mundial. O crescimento da produção e o aumento da capacidade competitiva da soja brasileira sempre estiveram associados aos avanços científicos e à disponibilização de tecnologias ao setor produtivo (Embrapa, 2002).

A Região Sul não possui muitas áreas disponíveis para expansão da cultura, porém, poderá incrementar sua participação com o uso de tecnologias disponíveis para o aumento da produtividade e estabilidade da produção. Alguns estados brasileiros, como o Paraná, são expoentes dos níveis de competitividade e produtividade alcançados no Brasil.

A previsão de safra da cultura da soja nesta região, conseqüentemente, torna-se um elemento de fundamental importância, pois auxiliará na determinação dos preços, colaborando para que acordos prévios entre produtores sejam feitos para manutenção de estoques reguladores no mercado e desenvolvimento de políticas agrícolas.

As imagens de satélite como as do Landsat 5, podem proporcionar o monitoramento da cultura, possibilitando a estimativa da área e da produção. Sua vantagem em relação aos métodos tradicionais de previsão de safras (censos ou pesquisas) é a menor subjetividade das informações, maior rapidez na aquisição de dados e o menor custo.

O processo de classificação automática de imagens multiespectrais de sensoriamento remoto consiste em associar cada pixel da imagem a um rótulo ou classe descrevendo um objeto real. Existem duas abordagens neste processo, a primeira denomina-se classificação supervisionada, onde o usuário identifica alguns dos pixels pertencentes às classes desejadas (formando a área de treinamento) e o software localiza todos os demais pixels pertencentes àquelas classes, baseado em alguma regra estatística pré-estabelecida. A segunda chama-se classificação não-supervisionada, nesta o próprio software decide, por análise de agrupamento, quais as classes a serem separadas e quais os pixels pertencentes a cada uma dessas classes (Crósta, 1992).

Dentro da classificação supervisionada, o método do Paralelepípedo considera uma área no espaço de atributos ao redor do conjunto de treinamento, cuja forma é um quadrado ou um paralelepípedo, definido pelo conjunto de treinamento. Todos os pixels situados dentro do paralelepípedo serão classificados como pertencentes à classe definida. Já o método da Máxima Verossimilhança considera a ponderação das distâncias das médias e utiliza parâmetros estatísticos para efetuar a classificação, partindo do princípio que o usuário já conhece o bastante da imagem a ser classificada, para poder definir as classes representativas (Crósta, 1992).

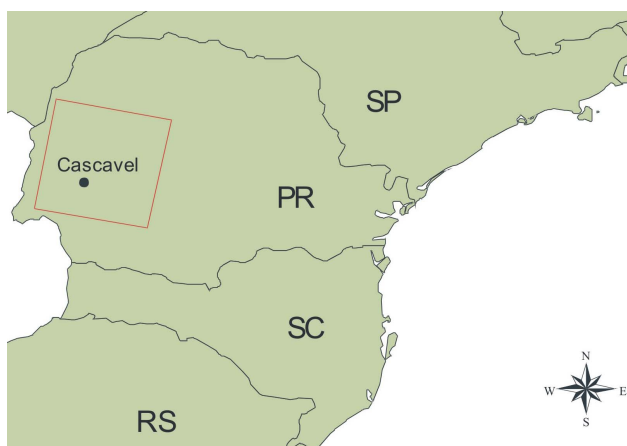
Segundo Congalton (1992), a coleta de dados que representa a situação real de campo na época da obtenção da imagem, é uma parte essencial de qualquer projeto de classificação e mapeamento envolvendo dados obtidos por meio de sensoriamento remoto. Esses dados são usados para verificar a acurácia da classificação, bem como detectar distinção entre classes e aperfeiçoar o processo de refinamento da classificação.

Um dos parâmetros mais utilizados e eficientes para quantificar a exatidão de levantamentos de uso da terra é o índice Kappa, que é obtido mediante a adoção de uma referência para comparação dos mapeamentos produzidos. Neste caso, a análise da exatidão dos mapeamentos é obtida através de matrizes de confusão ou matrizes de erro, sendo posteriormente calculado o índice Kappa de concordância (Cohen, 1960).

Neste contexto, o trabalho tem como objetivo mapear as áreas da soja na região Oeste do Estado do Paraná com imagens Landsat 5/TM, por meio de classificações digitais supervisionadas realizadas no software Envi 4.0, tendo como base uma área agrícola monitorada próximo ao município de Cascavel – PR.

## 2. Material e Métodos

A área de estudo é a cena do satélite Landsat 5, sensor TM (Thematic Mapper), órbita 223 / ponto 77 com resolução espacial de 30 metros, a mesma, está localizada na região Oeste do Paraná, a qual é representativa de grande volume de produção da soja no Estado, ilustrada na **Figura 1**.



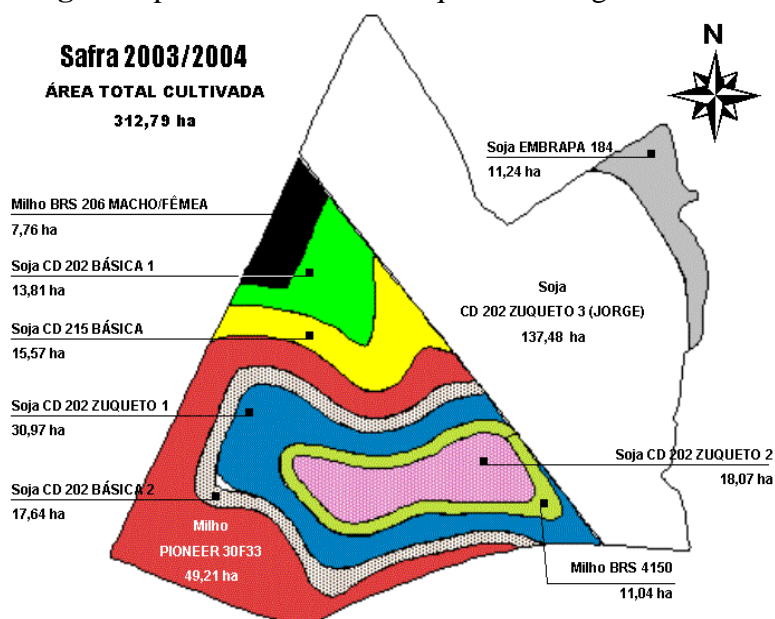
**Figura 1.** Abrangência da Cena 223/77 do Landsat 5.

Foram utilizadas cinco imagens adquiridas nas datas de 21/11/2003, 07/12/2003, 23/12/2003, 24/01/2004 e 09/02/2004, sendo selecionadas de forma a abranger todo o período do ciclo de desenvolvimento (estádios vegetativos e reprodutivos) da cultura da soja na região estudada.

A partir de composições coloridas das bandas do espectro infravermelho próximo, infravermelho médio e vermelho (RGB-453 falsa-cor), foram realizadas interpretações visuais (regiões de interesse) e posteriormente as classificações digitais supervisionadas para a

geração de um mapa temático contendo as áreas ocupadas com a cultura da soja, convencionalmente chamado de máscara da soja.

Para realizar as classificações, o primeiro passo foi selecionar regiões de interesse referentes à cultura da soja, essas regiões foram analisadas primeiramente em uma área agrícola monitorada que se encontra dentro da cena do Landsat 5 estudada. Vale ressaltar que essa área faz parte de um projeto paralelo a este, em andamento no Laboratório de Geoprocessamento da Faculdade de Engenharia Agrícola - FEAGRI da Universidade Estadual de Campinas – UNICAMP em parceria com a Universidade Estadual do Oeste do Paraná - UNIOESTE. As regiões de interesse tomaram por base a evolução da cultura da soja na safra 2003/2004 na área agrícola monitorada a qual está situada próxima ao município de Cascavel – PR. Na **Figura 2** pode ser visto um croqui da área agrícola monitorada.



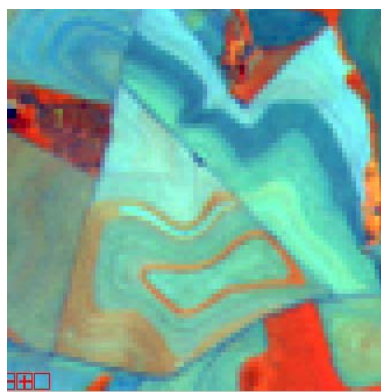
**Figura 2.** Área agrícola monitorada com culturas de soja e milho.

Fonte: LabGeo/FEAGRI - UNICAMP

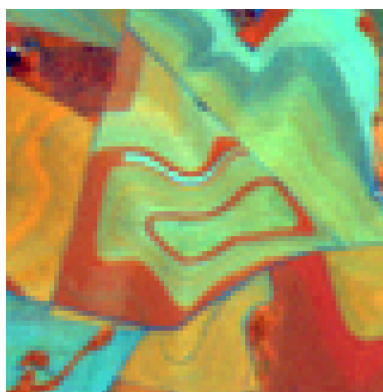
A partir do acompanhamento desta área agrícola (**Figura 2**), que caracteriza a maioria das áreas da região Oeste do Paraná, pelo fato de se cultivar ao mesmo tempo culturas de soja e milho em diferentes talhões (espaços delimitados que caracterizam a rotação de cultura de uma safra para outra). Como na área encontram-se talhões com as culturas de soja (cores: verde, amarelo, cinza, azul, rosa e branco) e de milho (cores: preto, vermelho e verde musgo), pode-se então acompanhar toda a evolução espectral dessas culturas ao longo dos seus ciclos de desenvolvimento, auxiliando para distinguir o mapeamento das áreas que continham soja das de milho.

As **Figuras 3, 4, 5, 6 e 7** ilustram o comportamento (ou a cor característica) da cultura da soja e do milho na área agrícola monitorada, em cada uma das datas de aquisição da imagem do Landsat 5. Observa-se nestas figuras como o comportamento da cultura de soja se diferencia do comportamento da cultura de milho, ao longo do desenvolvimento dessas culturas. Desta forma, os dados da área serviram de referência para delimitar os padrões e regiões de interesse no restante da cena Landsat 5 em cada data de aquisição.

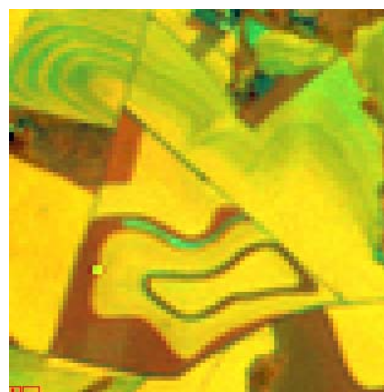
Em função das regiões de interesse coletadas, efetuaram-se então as classificações correspondentes aos algoritmos de classificação supervisionada do software ENVI 4.0 (RSI, 2003).



**Figura 3:** Área monitorada em 21/11/2003.



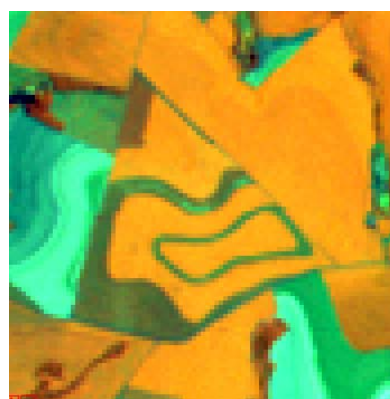
**Figura 4:** Área monitorada em 07/12/2003.



**Figura 5:** Área monitorada em 23/12/2003.



**Figura 6:** Área monitorada em 24/01/2004.



**Figura 7:** Área monitorada em 09/02/2004.

Fonte: LabGeo/UNICAMP

Após a geração da máscara de soja, foram realizados os cálculos da área de soja classificada para oito municípios produtores, que fazem parte da imagem. Esses municípios foram escolhidos, pelo fato de juntamente com outros municípios dentro do Estado do Paraná já estarem sendo monitorados em um projeto chamado GeoSafras, dirigido pela Companhia Nacional de Abastecimento – CONAB em parceria com outras instituições. Um dos objetivos desse projeto é a geração de metodologias que possam subsidiar estimativas de safras com caráter menos subjetivos. Dentro das metodologias utilizadas no projeto, uma delas foi a metodologia chamada de “plano amostral” desenvolvida por pesquisadores do Instituto Nacional de Pesquisas Espaciais – INPE. Resumidamente essa se caracteriza pela distribuição aleatória de 100 pontos dentro dos limites dos municípios sorteados, com o objetivo de conferir quais são as coberturas do solo (soja, milho, mata, etc) que contém em cada um desses pontos. Essa metodologia tem como finalidade, calcular as áreas ocupadas por determinadas culturas de forma menos subjetiva.

Diante dos 100 pontos amostrados nos oito municípios para a safra 2003/2004, foi possível realizar a avaliação da qualidade (exatidão) das informações geradas pela classificação, a qual foi feita por meio do índice Kappa (Cohen, 1960). O Kappa indica a qualidade da classificação, variando de 0 a 1, sendo que quanto mais se aproxima do valor 1, mais a classificação se aproxima da realidade. Esse coeficiente de concordância vem sendo muito utilizado como uma medida da exatidão da classificação temática.

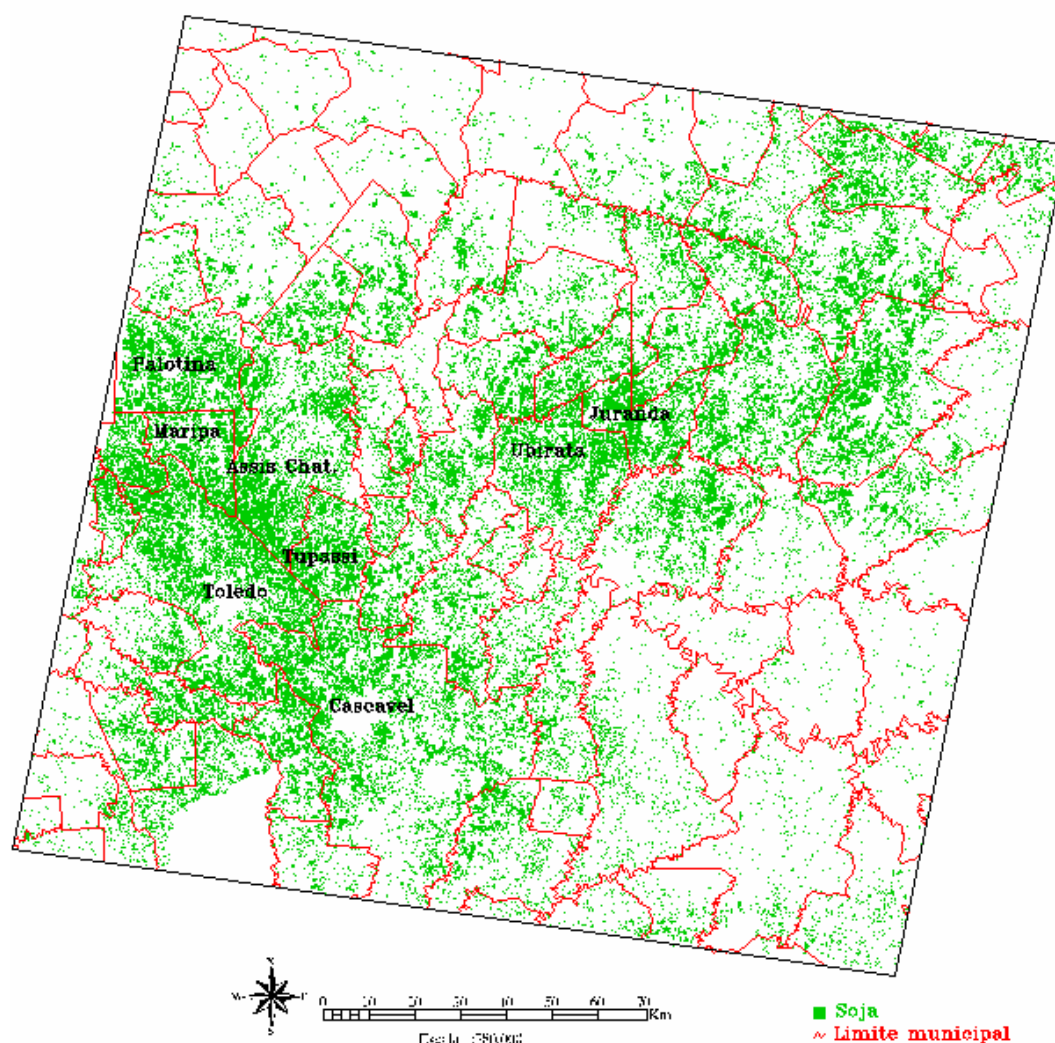
### 3. Resultados

A partir das classificações realizadas em cada uma das cinco imagens Landsat 5, utilizando os algoritmos supervisionados Paralelepípedo e Máxima Verossimilhança, a máscara da soja foi construída pela sobreposição de cada uma dessas imagens classificadas, através do software Envi 4.0.

Vale destacar que as classificações do algoritmo de Máxima Verossimilhança foram feitas utilizando um limiar de aceitação de 90%, o que significa dizer que os pixels que não alcançaram uma probabilidade de 90% de pertencer à classe soja, pré-determinada pela amostra de treinamento, foram então associados à classe de pixels classificados, como não soja.

As datas das imagens contemplaram todo o ciclo da cultura, isso foi muito importante pelo fato de que a região estudada caracteriza-se por altas variações nas datas de plantio e de colheita da soja, existindo variações no ciclo de desenvolvimento da cultura, ou seja, algumas áreas poderiam em determinada imagem (data) estarem no estágio vegetativo e outras no estágio reprodutivo próximo a colheita. Desta maneira, foi possível classificar todas as áreas que continham soja nas imagens.

A **Figura 8** mostra a máscara da soja, realizada com a sobreposição das classificações, destacando alguns dos municípios mais produtores de soja situados na região Oeste do Paraná.



**Figura 8.** Máscara da soja gerada para a cena do Landsat 5 223/77.

Com a máscara da soja e os dados de campo dos oito municípios, fez-se então, a avaliação da qualidade das informações geradas, por meio da exatidão global e do índice Kappa para cada um dos municípios. Ainda, como forma de comparação com um dado oficial de estimativa de área agrícola para a cultura de soja, para os oito municípios foram efetuadas as comparações entre a área de soja calculada pela classificação supervisionada e a área estimada pela Secretaria da Agricultura e Abastecimento do Paraná – SEAB e pelo Departamento de Economia Rural – DERAL.

A **Tabela 1** apresenta os valores de área calculados por meio do resultado da classificação e os estimados pela SEAB/DERAL para safra 2003/2004, os valores obtidos para a exatidão global e o índice Kappa em cada um dos oito municípios.

**Tabela 1.** Área e valores da Exatidão Global e do Índice Kappa para os oito municípios

Município	Área (ha)		Exatidão Global	Índice Kappa
	máscara de soja	SEAB/DERAL - PR		
Assis Chateaubriand	40780,44	72000,00	0,70	0,6968
Cascavel	64025,37	82000,00	0,83	0,8287
Mambore	26812,62	54200,00	0,67	0,6700
Maripá	14615,19	21500,00	0,67	0,6608
Palotina	30329,82	43300,00	0,63	0,6176
Rancho A. do Oeste	9555,57	15400,00	0,68	0,6790
Toledo	55436,49	68600,00	0,74	0,7356
Ubiratã	27155,43	51200,00	0,70	0,6968

Nota-se que a classificação (máscara de soja) subestimou a área estimada para os oito municípios quando comparados com os dados de estimativa de área da SEAB/DERAL – PR. Entretanto, a máscara de soja teve comportamento semelhante aos dados oficiais, ou seja, os municípios com maiores áreas de soja foram os mesmos para as duas estimativas.

Observa-se que o classificador apresentou limitações ao efetuar o reconhecimento da cultura, apesar do padrão de soja ter sido bem delimitado nas amostras de treinamento. Os dados da exatidão global, que representa o número de pontos coincidentes classificados como soja ou não-soja na máscara quando comparados com os 100 pontos amostrais de cada município, teve em média o valor de 70%. Esses valores influenciaram nas estimativas de área por parte da máscara de soja, como pode ser constatado nos municípios de Mambore (exatidão global de 0,67) e Ubiratã (exatidão global de 0,70), onde pelos dados da SEAB/DERAL o primeiro município continha maior área de soja do que o segundo, ao passo que para a máscara de soja a situação inverteu-se, ou seja, o município de Ubiratã passou a ter maior área com soja do que o município de Mambore.

Os resultados obtidos para o índice Kappa destes municípios foram comparados com os valores da **Tabela 2**, a qual descreve a avaliação da qualidade de classificação proposta por Landis e Koch (1977).

Os valores do índice Kappa encontrados para todos os municípios estão acima de 0,61, portanto considerados como classificação muito boa, segundo os autores. O município de Cascavel foi o que apresentou melhor índice Kappa, com o valor de 0,8287, considerado como uma classificação excelente. Já o município de Palotina apresentou o menor índice Kappa com o valor de 0,6176.

**Tabela 02.** Qualidade da classificação associada aos valores do índice Kappa

Índice Kappa	Qualidade
0.00	Péssima
0,01 a 0,20	Ruim
0,21 a 0,40	Razoável
0,41 a 0,60	Boa
0,61 a 0,80	Muito Boa
0,81 a 1,00	Excelente

Fonte: Landis & Koch (1977 p.165)

#### 4. Conclusão

O uso da área agrícola monitorada como base foi de grande utilidade, pois se obteve informações prévias sobre o comportamento das culturas de soja e milho na região estudada ao longo do ciclo de desenvolvimento das culturas, ajudando na seleção das melhores áreas de treinamento.

Por meio da sobreposição das classificações geradas em cada uma das datas de aquisição das imagens, foi possível contemplar as áreas de soja nos diferentes estádios de desenvolvimento (vegetativo e reprodutivo), compondo assim a máscara de soja final.

A máscara de soja gerada para a região Oeste do Estado do Paraná, poderá ser utilizada futuramente em outros trabalhos, como uma referência para a localização das áreas com a cultura da soja em outras imagens de satélite com menor resolução espacial (sensores NOAA/AVHRR, MODIS), caracterizando a técnica de passagem de escala entre sensores.

Os resultados obtidos com a classificação supervisionada da área da soja não tiveram o objetivo de ser qualificado como melhor ou pior, quando comparados com as estimativas oficiais, mas sim o objetivo de se buscar novos métodos que venham a ter menor subjetividade nas informações adquiridas, com maior rapidez na aquisição dos dados e o menor custo nos levantamentos.

#### 5. Agradecimentos

Ao CNPq pela bolsa de iniciação concedida e ao proprietário da área agrícola monitorada pela contribuição ao trabalho.

#### 6. Referências Bibliográficas

COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**. v. 20, n. 1, p. 37-46, 1960.

Congalton, R.G. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. **Photogrammetric Engineering and Remote Sensing**. Bethesda. 54, 1669-1671, 1992.

Crósta, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Campinas, SP: Instituto de Geociências/UNICAMP, 1992.

Embrapa Soja. **Tecnologia de produção de soja – Região Central do Brasil**. Sistemas de Produção n. 1. Londrina – PR: Embrapa soja. 2002. 1 ed. 195 p.

Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v.33, n.1, p.159-174, 1977.

RSI - RESEARCH SYSTEMS, INC. **User's Guide. ENVI - version 4.0** - The Environment for Visualizing Images. 2003.