

RICARDO SEIXAS BRITES

**VERIFICAÇÃO DE EXATIDÃO EM CLASSIFICAÇÃO DE
IMAGENS DIGITAIS ORBITAIS: EFEITOS DE DIFERENTES
ESTRATÉGIAS DE AMOSTRAGEM E AVALIAÇÃO DE ÍNDICES
DE EXATIDÃO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Curso de Ciência Florestal, para obtenção do título de "*Doctor Scientiae*".

VIÇOSA
MINAS GERAIS - BRASIL

1996

RICARDO SEIXAS BRITES

**VERIFICAÇÃO DE EXATIDÃO EM CLASSIFICAÇÃO DE
IMAGENS DIGITAIS ORBITAIS: EFEITOS DE DIFERENTES
ESTRATÉGIAS DE AMOSTRAGEM E AVALIAÇÃO DE ÍNDICES
DE EXATIDÃO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Curso de Ciência Florestal, para obtenção do título de “*Doctor Scientiae*”.

APROVADA: 11 de dezembro de 1995

Prof. Carlos Antonio Álvares
Soares Ribeiro
(Conselheiro)

Prof. Vicente Paulo Soares
(Conselheiro)

Prof. Laércio Couto

Prof. Gilberto Chohaku Sedyama

Prof. João Carlos Chagas Campos
(Orientador)

*Esta tese é dedicada à memória do Prof.
Dr. Celestino Aspiazú, responsável pelo
início de nossas lides acadêmicas na área de
sensoriamento remoto. Que suas palavras e
ensinamentos, ainda vivos em nossa
memória, continuem a nos inspirar em nossa
busca de novos conhecimentos.*

AGRADECIMENTOS

À minha família, sacrificada por tantas ausências, pelo constante apoio e incentivo. Maria Célia, Mariana e João Diogo, vocês são as maiores razões para que a luta prossiga.

Aos membros da banca examinadora, pelas valiosas contribuições ao trabalho e pela paciência nas revisões das versões preliminares.

Ao Departamento de Engenharia Florestal da Universidade Federal de Viçosa, pela oportunidade que nos foi dada e pelo interesse neste trabalho, em particular ao colega e amigo Prof. José Carlos Ribeiro, pelo constante encorajamento.

À CAF Florestal Ltda., cuja boa vontade em ceder os dados utilizados nesta pesquisa propiciou a sua realização.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo apoio na primeira fase do curso de Doutorado, realizada na Colorado State University, em Fort Collins, CO, EUA.

Ao Prof. David R. Betters, nosso orientador na Colorado State University.

Ao Prof. João Carlos Chagas Campos, orientador de nossos trabalhos após o falecimento do Prof. Celestino Aspiazú.

Ao Grande Arquiteto do Universo, pela luz e força em todos os momentos de nossa vida.

BIOGRAFIA

Ricardo Seixas Brites, nascido a 7 de junho de 1956, em Santos, SP, graduou-se Engenheiro Agrimensor pela UFV em 1981. No ano seguinte transferiu-se para a Bahia onde, por cinco anos, trabalhou na Comissão Executiva do Plano da Lavoura Cacaueira, CEPLAC, como professor da Escola Média de Agropecuária Regional da Ceplac (EMARC), em Uruçuca, BA.

Realizou curso de especialização em Programação de Computadores e Análise de Sistemas na Federação das Escolas Superiores de Ilhéus e Itabuna / Fundação Universidade de Santa Cruz. Deixando a CEPLAC, assumiu a Diretoria de Treinamento e Desenvolvimento da Educar Informática Ltda, em Itabuna, BA.

Sua próxima posição profissional foi a de Professor do Departamento de Ciências da FESPI/FUSC, na área de matemática, acumulando esta função com a de Analista de Sistemas da Cooperativa Central dos Produtores de Cacau, COPERCACAU, em Ilhéus, BA.

Em 1988 retorna a Viçosa, ingressando no curso de Mestrado em Ciência Florestal da Universidade Federal de Viçosa, concluindo-o, sob a orientação do Prof. Celestino Aspiazú, em 1990.

Na seqüência de seu treinamento inicia, ainda em 1990, o seu doutoramento no Department of Forest Sciences, da Colorado State University, em Fort Collins, CO, EUA, sob a orientação do Prof. David R. Betters. Após a conclusão de seus créditos naquela instituição, dois anos depois, retorna ao Brasil, já contratado pela UFV, onde, paralelamente às suas atividades como professor do Departamento de Engenharia Florestal, viria a defender tese em dezembro de 1995.

Além de suas funções regulares como docente, Ricardo Seixas Brites é o Presidente da Associação Brasileira de Laboratórios de Sensoriamento Remoto e Coordenador do Núcleo de Geoprocessamento Prof. Celestino Aspiazú - NUGEO, da UFV.

CONTEÚDO

	Página
EXTRATO	viii
ABSTRACT	x
1. INTRODUÇÃO	1
1.1. Generalidades	1
1.2. O Problema	2
1.3. Objetivos	4
2. REVISÃO DE LITERATURA	5
2.1. Transformações Especiais Aplicadas às Imagens Orbitais	5
2.1.1. A Análise de Componentes Principais	5
2.1.2. A Transformação <i>Tasseled Cap</i>	10
2.2. O Algoritmo de Classificação por Máxima Verossimilhança	14
2.3. Técnicas de Amostragem	18
2.3.1. Os Padrões de Amostragem	19
2.3.2. A Amostragem Blocada	21
2.3.3. A Amostragem Totalmente Casualizada	21
2.3.4. A Amostragem Sistemática	22

	Página
2.3.5. A Amostragem Sistemática Estratificada Não-Alinhada	24
2.4. A Exatidão da Classificação	24
3. MATERIAL E MÉTODOS	33
3.1. Localização da Área de Estudo	33
3.2. Dados Utilizados	35
3.2.1. Imagem Orbital	35
3.2.2. Fotografias Aéreas	35
3.2.3. Mapas	36
3.2.4. Dados de Campo	36
3.3. Sistemas Computacionais e Equipamentos Utilizados	36
3.3.1. Periféricos	36
3.3.2. Equipamentos Especiais	36
3.4. Metodologia	37
3.4.1. A Coleta das Amostras de Treinamento	37
3.4.2. O Algoritmo de Classificação	38
3.4.3. As Combinações de Bandas e Transformações Especiais	38
3.4.4. A Amostragem sobre os Dados de Referência	39
3.4.5. O Cálculo dos Índices e de Suas Variâncias	47
3.4.6. As Comparações dos Índices de Exatidão	49
4. RESULTADOS E DISCUSSÃO	50
4.1. As Classificações Geradas	50
4.2. O Comportamento do Índice Kappa	56
4.3. Comparações entre os Índices de Exatidão Global, Kappa e Tau	63
4.4. A Significância das Diferenças	67
5. RESUMO E CONCLUSÕES	71
6. RECOMENDAÇÕES	74
REFERÊNCIAS BIBLIOGRÁFICAS	75
APÊNDICE	80

EXTRATO

BRITES, Ricardo, S., D.S., Universidade Federal de Viçosa, janeiro de 1996. *Verificação de Exatidão em Classificação de Imagens Digitais Orbitais: Efeitos de Diferentes Estratégias de Amostragem e Avaliação de Índices de Exatidão*. Professor Orientador: João Carlos Chagas Campos. Professores Conselheiros: Carlos Antônio Álvares Soares Ribeiro e Vicente Paulo Soares.

Quatro diferentes estratégias de amostragem foram utilizadas sobre os dados de referência para verificação de exatidão em cinco classificações obtidas a partir de combinações de bandas espectrais e transformações especiais, com o objetivo de verificar se haveria influência dos tipos de amostragem sobre os resultados reportados por três índices de exatidão: Global, Kappa e Tau. À parte disso foi também conduzido um estudo sobre o comportamento dos três índices, uns em relação aos outros. Foram utilizadas as seguintes amostragens: blocada, casualizada, sistemática estratificada e sistematizada, sendo as três últimas realizadas pixel a pixel. O algoritmo de máxima verossimilhança foi aplicado às seguintes combinações das bandas espectrais do sensor TM, a bordo do satélite Landsat 5: TM3, TM4 e TM5; TM3, TM4 e TM7; TM2, TM3 e TM4. Além dessas combinações também foram classificadas as

composições baseadas nas três primeiras componentes principais resultantes das seis bandas reflexivas do TM e as três componentes resultantes da transformação *Tasseled Cap*. Dentre os principais resultados obtidos, podem ser destacados: pelas comparações de cada classificação com todas as combinações possíveis de amostragens, conclui-se que para nenhuma delas houve diferença significativa, pelos três índices, quando foram usadas amostragens não-blocadas, o que sugere ser a amostragem blocada a principal fonte de superestimação dos índices; o comportamento dos índices G, K e T em todas as classificações e em todas as amostragens é bastante semelhante no que se refere à variação dos mesmos de uma classificação para outra; o índice T, ao se situar entre G e K, parece ser o ideal, mostrando tornar possível a correta consideração das concordâncias real e casual, para a situação na qual é assumida a equiprobabilidade de ocorrência das categorias informacionais; dentro de uma mesma estratégia de amostragem, à exceção das amostragens sistemática e blocada, os três índices discriminaram as diferenças entre classificações de maneira idêntica; à medida que aumenta o grau de interferência do analista nos procedimentos de amostragem, aumenta a sensibilidade dos índices no que se refere ao estabelecimento de diferenças entre classificações independentes; não houve diferença significativa entre as estratégias não-blocadas de amostragem, considerados os três índices utilizados; de acordo com a conclusão anterior, parece ser a amostragem sistemática a mais vantajosa, do ponto de vista da facilidade de operacionalização, quando se objetiva quantificar melhor a exatidão dos processos de classificação, desde que as feições de interesse não apresentem caráter de periodicidade na paisagem.

ABSTRACT

BRITES, Ricardo, S., D.S., Universidade Federal de Viçosa, January, 1995. *Accuracy Assessment upon Digital Orbital Imagery: Effects of Different Sampling Approaches and Evaluation of Accuracy Indexes*. Adviser: Prof. João Carlos Chagas Campos. Committee Members: Prof. Carlos Antônio Álvares Soares Ribeiro and Prof. Vicente Paulo Soares.

Four different sampling approaches were utilized upon reference data to assess the accuracy of five classifications obtained from combinations of spectral bands and special transformations, in order to verify the influence of the sampling strategies over the reported results of three accuracy indexes: Global, Kappa and Tau. A study about the behavior of the three indexes in relation to each other was also conducted. The following sampling approaches were used: blocked, randomized, systematic stratified, and systematic, being the last three based on individual pixels. The maximum likelihood algorithm was applied to the following combinations of spectral bands: TM3, TM4, TM5; TM3, TM4, TM7 and TM2, TM3, TM4. The composition based on the three first principal components resulting from the six TM reflected channels, and the components resulted from the Tasseled Cap transformation, produced

two more classifications. The results showed: by the comparisons of each classification with all the possible sampling combinations, it was concluded that for none of the classifications there was a significant difference, by the three indexes, when pixel-based samplings were taken, what indicates the blocked sampling as the main source of the indexes superestimation; the behavior of the indexes Global, Kappa and Tau over all the classifications and in all the sampling approaches is very similar concerning to the way each one varies from one classification to the other; the index Tau, with its values always between Global and Kappa seems to be the ideal, showing to be possible the correct consideration of real and casual agreements, for the situation where the equiprobability of occurrence of the informational categories is assumed; considering one sampling approach at a time, excluding the systematic and blocked ones, all the three indexes discriminate the differences among classifications in the very same way; the greater the interference of the analyst in the sampling procedure, the greater the sensitivity of the indexes in what is concerned to the establishment of differences among independent classifications; there was no significant difference among the pixel-based samplings, considered the three indexes; the systematic sampling procedure could be the recommended, from the stand point of implementation, when the objective is to better quantify the accuracy of the classification processes, provided that the features of interest are not in phase with the sampling grid.

1. INTRODUÇÃO

1.1. *Generalidades*

O desenvolvimento da civilização hodierna é feito, em sua maior parte, com base na ocupação do espaço físico na superfície da terra. Dessa forma, revestem-se de especial importância as tecnologias que permitem os levantamentos de como esta superfície vem sendo utilizada ou de como partes dela poderiam vir a ser úteis à saga do desenvolvimento. As necessidades prementes de conservação de recursos naturais ameaçados de extinção fazem com que o planejamento do uso da terra contemple uma grande gama de elementos ambientais que necessitam ser visualizadas e estudadas dentro do contexto em que estão inseridas. O estudo dos fenômenos da natureza não pode ser conduzido considerando-os de forma isolada e, sim, dentro de um sistema onde seja possível estimar os efeitos que modificações em um de seus componentes possam vir a produzir em outros.

O surgimento dos modernos satélites de observação da Terra trouxe novas perspectivas às atividades de monitoramento e levantamento dos recursos naturais. O seu uso, inicialmente restrito às instituições de pesquisa e organismos governamentais, vem se expandindo de forma

extremamente rápida, fazendo com que hoje as suas imagens se tornem parte do dia-a-dia do cidadão comum, a exemplo do que ocorre com os produtos oriundos dos satélites meteorológicos.

As aplicações para as tecnologias baseadas em sensores a bordo de plataformas orbitais tendem a crescer muito mais à medida que novos satélites sejam lançados. Poder-se-ia dizer que há uma tendência de que as imagens orbitais venham, em termos de resolução espacial, a se equiparar aos sensores fotográficos, possuindo sobre estes as vantagens da repetibilidade do recobrimento, da grande resolução espectral e do baixo custo.

O Quadro 1 indica os satélites com previsão de serem lançados até o ano de 2004, resumindo algumas das características dos sensores que eles carregam.

1.2. O Problema

As imagens orbitais destinam-se, primariamente, à confecção de mapas dos mais diversos tipos, que são gerados a partir da transformação das categorias espectrais, presentes nelas, em categorias ou classes informacionais, mediante processos de classificação. Estes mapas servem de base a diversas atividades que variam em suas utilizações, desde simples ilustrações até produtos de alta qualidade final, que servirão de base a atividades críticas de planejamento e monitoramento de recursos.

Dependendo de qual seja a utilização dos dados oriundos dos sensores remotos, é esperado um determinado nível de exatidão do processo de classificação, mensurado por índices destinados a este fim, que dará uma idéia da confiabilidade do produto no que se refere à fidedignidade com que representa as feições presentes na região imageada.

Quadro 1 - Satélites de observação da terra com lançamento previsto até o ano de 2004

País	Prop.	Programa	Data	Tipo	Res. (m)			# Be	Estéreo
					P	M	R		
França	G/O	Spot 5B	'04	P&M	5	10	4	F/A	
USA	G/O	EOS AM-2 / L-8	'04	P&M	10	30	7		
França	G/O	Spot 5A	'99	P&M	5	10	4	F/A	
Índia	G/O	IRS-1 D	'99	P&M	10	20	4	C/T	
USA	C/O	Space Imaging	'98	P&M	1	4	4	F/A	
Coréia	G/O	KOMSAT	'98	P&M	10	10	3	F/A	
USA/Japão	G/O	EOS AM-1	'98	M	15	15	14	F/A	
USA	G/O	Landsat-7	'98	P&M	15	30	7		
ESA	G/O	ENVISAT	'98	R			30		
USA	C/O	Space Imaging	'97	P&M	1	4	4	F/A	
USA	C/O	Eyeglass	'97	P	1			F/A	
França	G/O	Spot 4	'97	P&M	10	20	4	C/T	
USA	C/O	Earth Watch	'97	P&M	1	4	4	F/A	
USA	C/O	Earth Watch	'96	P&M	3	15	3	F/A	
USA	G/E	CTA Clark	'96	P&M	3	15	3	F/A	
USA	G/E	TRW Lewis	'96	P&M	5	30	384		
Rússia	G/O	Almaz 2	'96	R			5		
Japão	G/O	ADEOS	'96	P&M	8	16	4	C/T	
China/Brasil	G/O	CBERS	'96	P&M	20	20	7	C/T	
Canadá	G/O	Radarsat	'95	R			9		
Índia	G/O	IRS-1 C	'95	P&M	10	20	4	C/T	
China/Brasil	G/O	CBERS	'95	P&M	20	20	7	C/T	
Rússia	G/O	Resours-02	'95	M		27	3		

M = somente multispectral; R = radar; O = operacional; F/A = linha; P&M = pancromático e multiespectral; G = governamental; E = experimental; C/T = lado; P = somente pancromático; # BE = número de bandas espectrais; e C = comercial (extraído de “Land Satellite Information in the Next Decade”, prospecto para a conferência promovida pela American Society for Photogrammetry and Remote Sensing de 25-28 de setembro de 1995 em Vienna, Virginia, USA).

Para a determinação desses índices, e existem vários, é necessário que a região em estudo seja amostrada, em trabalhos de campo e, ou, com base em levantamentos fotogramétricos e outros dados ancilares, a fim de proporcionar elementos que permitam a comparação entre o que se chama de “verdade de campo” e o resultado do processo de classificação. Dessa forma, quanto maior o número de coincidências, maior será o valor do índice de exatidão, e vice-versa.

O problema reside em saber de que forma a estratégia de amostragem, utilizada sobre os dados de referência, pode vir a afetar os valores desses índices e, se confirmado o efeito da amostragem, tentar quantificá-lo.

Essa preocupação não é recente entre os pesquisadores que se dedicam ao sensoriamento remoto, sendo levantada em vários artigos científicos, a exemplo de CONGALTON (1984) e CONGALTON (1988), que apontam a escassez de trabalhos nesta área.

1.3. *Objetivos*

O objetivo principal deste estudo é testar os efeitos de diferentes estratégias de amostragem, aplicadas aos dados de referência, sobre três índices utilizados como mensuradores de exatidão (Exatidão Global, Kappa e Tau) na classificação de imagens orbitais. Mais especificamente, também constituem objetivos:

- testar o comportamento do índice Tau, comparado aos índices tradicionalmente utilizados em sensoriamento remoto (Global e Kappa);
e
- comparar diferentes combinações de bandas e transformações especiais, no sentido de verificar qual delas seria mais adequada para a utilização em classificações de regiões densamente vegetadas.

2. REVISÃO DE LITERATURA

2.1. *Transformações Especiais Aplicadas às Imagens Orbitais*

Há diversas transformações especiais que, aplicadas aos dados oriundos dos sensores remotos, podem produzir efeitos que variam desde a redução da dimensionalidade destes dados até o realce de feições específicas presentes na cena a ser analisada. Dentre elas podem ser destacadas a análise de componentes principais e a transformação *Tasseled Cap*.

2.1.1. *A Análise de Componentes Principais*

Esta técnica é comumente utilizada em análise de dados oriundos de sensores remotos. Os dados multiespectrais obtidos pela maioria dos sensores remotos exibem altas correlações entre as suas bandas ou canais. Como exemplo podem ser citadas as bandas 1, 2 e 3 do Mapeador Temático (TM), a bordo dos satélites Landsat 4 e 5. Dessa forma, o processamento dos dados produzidos pelas bandas disponíveis nos sensores envolve um alto grau de redundância. Este fato implica aumento dos custos do processamento destes dados, especialmente quando são trabalhadas passagens de datas diferentes (FUNG e LEDREW, 1987). De

acordo com LILLESAND e KIEFER (1987), quando esta técnica é empregada antes do processamento digital das imagens, ela, geralmente, aumenta a eficiência computacional dos processos de classificação pelo fato de propiciar a redução na dimensionalidade dos dados. Em outras palavras, o propósito de se utilizar a técnica seria o de comprimir a informação contida no conjunto original das n bandas do sensor em um número menor de bandas ou componentes.

Diferentemente das razões entre bandas que requerem uma cuidadosa escolha das combinações de bandas a fim de enfatizar características espectrais específicas, a transformação em componentes principais não requer informação *a priori* e, dessa forma, pode ser aplicada de maneira menos direcionada. Ela é normalmente aplicada a todas as bandas disponíveis do sensor com o qual se está trabalhando. As imagens resultantes enfatizam distinções entre os diferentes materiais existentes na superfície, mas, em virtude do fato de cada nova imagem ser uma combinação linear de todas as bandas consideradas, as componentes, individualmente, ou por meio de composições feitas com as mesmas, não podem ser interpretadas da mesma maneira que se faria com o produto resultante da razão entre bandas (WILLIAMS JR., 1983).

De acordo com HARALICK e FU (1983), a análise de componentes principais consiste de uma técnica estatística padronizada para a seleção de um subespaço com dimensões tais que nele resida a maior parte da variância dos dados. Se x_1, \dots, x_n são os vetores de dados amostrais, μ o vetor das médias das amostras e Σ a matriz de variância-co-variância das amostras, a melhor dimensão K na qual podem ser projetados os dados será aquele subespaço K -dimensional abrangido pelos K autovetores de Σ que possuam o maior autovalor. Assim, se T é uma matriz cujas K linhas são estes autovetores, as K principais componentes de x_1, \dots, x_n serão Tx_1, \dots, Tx_n , com cada Tx_n sendo um vetor K -dimensional.

A transformação em componentes principais é uma técnica estatística multivariada que, de acordo com LOUGHLIN (1991), seleciona combinações lineares não-correlacionadas de variáveis de uma forma tal que cada sucessiva combinação linear extraída possui uma variância menor. A variância estatística nas imagens multiespectrais é relacionada com a resposta espectral dos vários materiais presentes na superfície da terra, como rochas, solos e vegetação, sendo também influenciada pela dimensionalidade dos dados da imagem. Quando canais multiespectrais da imagem são tratados como variáveis e sujeitos à transformação, a geração ordinal das componentes principais é influenciada pela abundância espacial dos vários materiais presentes na superfície imageada e pelas características estatísticas da imagem. Dessa forma, as estatísticas da cena podem ser tomadas com vantagem por poderem ser medidas e ajustadas, objetivando forçar a transformação a produzir informações concernentes à distribuição espacial e relativa abundância de materiais ou feições específicas presentes na cena, isto é, usa-se a análise de componentes principais como uma técnica de interrogação.

A análise de componentes principais é considerada uma técnica exploratória que pode ser extremamente útil na obtenção de um melhor entendimento do inter-relacionamento entre as variáveis, podendo ser usada, inclusive, para testes de normalidade. Se as componentes principais não são normalmente distribuídas, as variáveis originais também não o serão (AFIFI e CLARK, 1990).

De acordo com SINGH e HARRISON (1985), a transformação em componentes principais possui diversas características, algumas das quais são de especial interesse em sensoriamento remoto:

- 1) A variância total é preservada na transformação, isto é:

$$\sum_{i=1}^N \mathbf{s}_i^2 = \sum_{i=1}^N \mathbf{l}_i$$

em que σ_i^2 representa as variâncias das bandas originais e λ_i as variâncias das componentes principais.

2) Esta transformação minimiza a aproximação dos erros médios quadráticos.

3) É a única transformação que gera coeficientes não-correlacionados. No sentido geométrico, ela rotaciona as feições altamente correlacionadas no espaço N-dimensional para uma orientação mais favorável, fazendo com que estas feições se posicionem ortogonalmente umas às outras, de tal forma que a variância seja considerada em ordem decrescente de magnitude na ordenação das componentes. Esta técnica vem sendo vista como de compressão de informações em um número menor de componentes pelo descarte da informação redundante nas componentes de mais alta ordem.

Segundo JENSEN (1986), a transformação em componentes principais pode ser melhor conceituada considerando-se a distribuição bidimensional dos valores dos pixels obtidos em duas bandas do sensor TM (mapeador temático), a partir de agora chamadas simplesmente de X_1 e X_2 . Um diagrama de dispersão dos valores de brilho associados a cada pixel em cada banda, bem como a localização de suas médias μ_1 e μ_2 , é mostrado na Figura 1. A dispersão ou variância na distribuição dos pontos é um indicativo da correlação e qualidade da informação associada com ambas as bandas. Se todos os pontos estivessem agrupados em uma zona extremamente estreita no espaço bidimensional, estes dados, provavelmente, seriam de pequena utilidade para a análise.

A posição inicial dos eixos X_1 e X_2 pode não ser, necessariamente, o melhor arranjo no espaço multiespectral para analisar os dados destas duas bandas. O objetivo é o de se usar a análise de componentes principais para produzir uma translação ou rotação nos eixos originais, de forma que os valores de brilho originais plotados contra os dois eixos

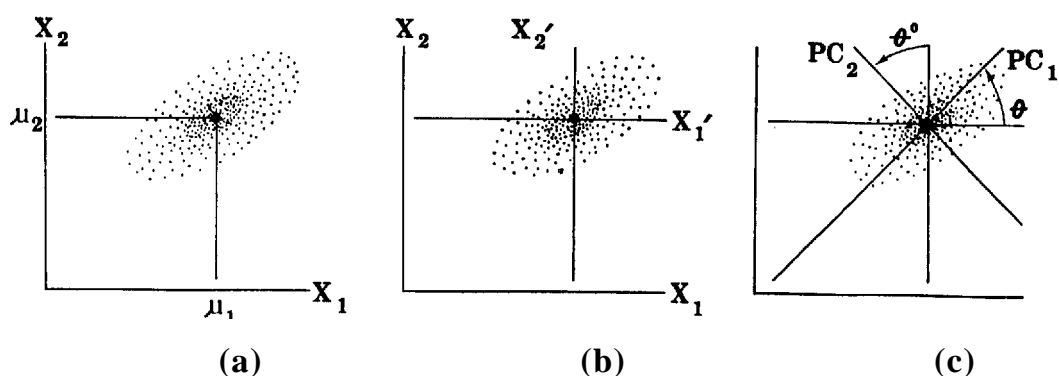


Figura 1 - Representação diagramática do relacionamento espacial entre as duas primeiras componentes principais (Adaptado de JENSEN, 1986).

sejam reprojctados em um novo conjunto de eixos X'_1 e X'_2 . Como pode ser depreendido da análise da Figura 1-b, os dois novos eixos passam a ter a sua origem no par coordenado μ_1, μ_2 . Os novos valores de coordenadas para os pontos passam a ser expressos da seguinte forma:

$$X'_1 = X_1 - \mu_1, \text{ e}$$

$$X'_2 = X_2 - \mu_2.$$

Na seqüência o sistema X' é rotacionado em torno de sua origem, segundo um ângulo θ , de forma que a primeira componente principal (PC_1) seja orientada de acordo com o semi-eixo maior da distribuição dos pontos e que a variância da segunda componente principal (PC_2) seja máxima (Figura 1-c). Para que isso ocorra, PC_2 deve ser perpendicular a PC_1 . As próximas componentes principais contêm as quantidades monotonicamente decrescentes das variâncias existentes no conjunto de dados. Usualmente é comum que a primeira componente contenha algo em torno de 90% da variância total do conjunto, enquanto a segunda responda por alguma coisa em torno de 5% deste total.

As variâncias associadas a cada uma das componentes principais são chamadas de autovalores, raízes características, raízes latentes ou valores próprios (AFIFI e CLARK, 1990).

Segundo DAVIS (1986), tem havido uma certa confusão, especialmente entre geólogos, com relação ao uso da terminologia. Em muitos dos estudos publicados por eles, o que se denomina análise fatorial consiste, na verdade, de análise de componentes principais. Alguns autores têm achado tão confusa esta situação que resolveram chamar a análise fatorial de “análise fatorial verdadeira”, a fim de distingui-la da análise de componentes principais.

2.1.2. A Transformação *Tasseled Cap*

Um dos mais importantes índices de vegetação é o chamado Tasseled Cap Transformation, desenvolvido por Kauth e Thomas (1976), citados por JENSEN (1986). Ele é baseado nas técnicas de ortogonalização seqüencial de Gram-Schmidt, que produzem uma transformação ortogonal do espaço de dados das quatro bandas originais do MSS em um novo espaço de quatro dimensões. Esse índice tem sido rigorosamente testado e usado extensivamente em pesquisas agrícolas.

Pesquisas mais recentes vêm avaliando o Mapeador Temático (*Thematic Mapper* - TM) a bordo dos satélites Landsat 4 e 5, para determinar se ele seria capaz de prover informações tão úteis como aquelas obtidas a partir do MSS. Crist e Cicone (1984), citados por JENSEN (1986), verificaram que as seis bandas refletidas do TM são dispersas em um espaço tridimensional, definindo dois planos perpendiculares e uma zona de transição entre os dois. Áreas completamente vegetadas definem o "plano de vegetação", enquanto solos expostos definem o "plano de solos". Entre estes dois existem áreas parcialmente vegetadas, nas quais a vegetação e os solos são visíveis. Dessa forma, durante a estação de crescimento das plantas, uma dada área sob atividade agrícola deve

começar no plano de solos, migrar para a zona de transição, chegar ao plano de vegetação próximo ao final do desenvolvimento da cultura e, depois, mover-se de volta ao plano de solos durante a colheita ou senescência (Figura 2).

A partir dos coeficientes mencionados a seguir, são obtidos três componentes, um relativo ao verdor (greenness), um relativo ao brilho (brightness) e um terceiro que está relacionado com o solo, incluindo o seu teor de umidade. Essa importante fonte de informações a respeito dos solos torna-se disponível em virtude da inclusão de bandas na região espectral do infravermelho intermediário, disponíveis no sensor TM (não existentes no MSS).

O quadro extraído de JENSEN (1986) mostra os coeficientes a serem aplicados às seis bandas reflexivas do Mapeador Temático.

De acordo com o trabalho de CRIST e CICONE (1984), as componentes brilho, verdor e umidade podem ser assim caracterizadas:

BRILHO: é a soma ponderada de todas as seis bandas reflexivas do Mapeador Temático (TM). Como tal, ela contém respostas relativas às mudanças na refletividade total e àqueles processos físicos que venham a afetar esta mesma refletividade total. Dessa forma, diferenças nas características dos solos, como distribuição dos tamanhos das partículas, serão claramente expressas na componente “brilho”, enquanto aumentos na densidade de vegetação, que tendem a aumentar a resposta na região espectral do infravermelho próximo enquanto a diminuem na região visível, irão causar mudanças menos substanciais nesta componente.

VERDOR: é o contraste entre as bandas na região da radiação visível e aquelas da região do infravermelho. Esta componente é responsável pela resposta obtida pela combinação da alta absorção nas bandas visíveis (em virtude da pigmentação das plantas, particularmente

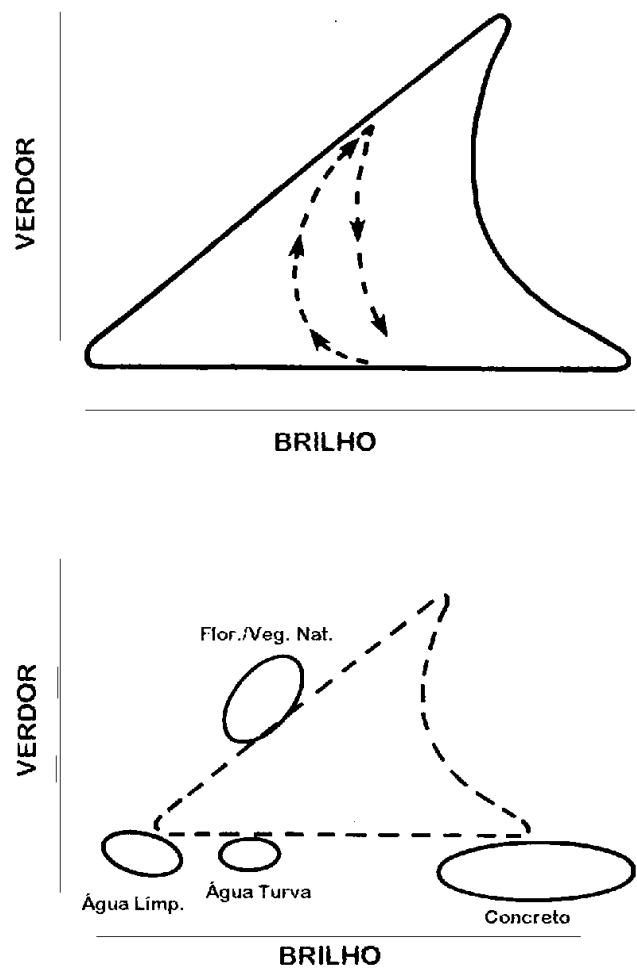


Figura 2 - A transformação *Tasseled Cap* (Adaptada de JENSEN, 1986).

Quadro 2 - Coeficientes aplicados às bandas reflexivas do sensor mapeador temático, a bordo do Satélite Landsat 5, para obtenção da transformação *Tasseled Cap*

	1	2	3	4	5	7
BRILHO	0,33183	0,33121	0,55177	0,42514	0,48087	0,25252
VERDOR	-0,24717	-0,16263	-0,40639	0,85468	0,05493	-0,11749
UMIDADE	0,13929	0,22490	0,40359	0,25178	-0,70133	-0,45732

clorofila) e aquela resultante da alta refletividade na região do infravermelho próximo (em virtude da estrutura interna das folhas e do resultante espalhamento da radiação nesta região do espectro, características da vegetação verde. As componentes “brilho” e “verdor” definem o “plano de vegetação” no espaço da transformação *Tasseled Cap*.

UMIDADE: esta componente contrasta a soma das bandas do visível e infravermelho próximo com aquelas da região do infravermelho intermediário. O nome “umidade” se deve ao fato de as bandas do infravermelho intermediário serem mais susceptíveis ao teor de água no solo. Poderia ser esperado que mudanças na quantidade de umidade poderiam afetar estas bandas de maneira mais acentuada que na região do visível ou do infravermelho intermediário e, dessa forma, que um contraste entre estes dois conjuntos de bandas iria salientar características da cena relativas à umidade. As componentes “brilho” e “umidade” definem o chamado “plano de solos” no espaço da transformação *Tasseled Cap*. Nas análises de dados reais e simulados, o teor de umidade do solo revelou ser a característica primária expressa na componente “umidade”. Por outro lado, somente uma quantidade limitada da variação em teor de água nas plantas conseguiu ser representada no conjunto de dados.

A literatura especializada reporta diversos trabalhos que mostram a utilidade desse tipo de transformação, a exemplo de LAUVER e WHISTLER (1993) que, em seu trabalho para identificação de pradarias e áreas de *habitat* de espécies raras, utilizam a técnica em virtude da capacidade que possuem as componentes “brilho” e “verdor” de distinguir diferenças na vegetação, bem como da componente “umidade” de distinguir condições de turgidez entre os diferentes tipos de vegetação. O que, a princípio, pode parecer uma contradição com o parágrafo anterior não o é, pois, embora a componente umidade não seja tão efetiva com a vegetação quanto o é com os solos, ela ainda se presta bem para as finalidades de discriminação quando as diferenças em umidade são relativamente altas. MASSASATI (1990) afirma que as aplicações para os dados do Mapeador Temático podem ser expandidas com a adição das informações providas pela componente “umidade” no reconhecimento dos vários estádios do desenvolvimento agrícola em relação a outros elementos ambientais, tais como água e materiais produzidos pelo homem. Na mesma linha, YOSHIMURA e OSHIMA (1991) concluem que o monitoramento das condições da vegetação pelo uso da transformação *Tasseled Cap* foi confirmado como sendo um dos métodos mais efetivos.

2.2. O Algoritmo de Classificação por Máxima Verossimilhança

Dentre os diversos algoritmos utilizados nos procedimentos de classificação supervisionada, merece destaque, pela sua eficiência, o de máxima verossimilhança. Este classificador, de acordo com LILLESAND e KIEFER (1987), avalia a variância e a covariância dos padrões de resposta da categoria espectral quando classifica um pixel desconhecido. Para que isto possa ser feito, é assumido que a distribuição do grupamento de pontos que formam os dados de treinamento para a categoria seja normal. Esta condição é geralmente satisfeita para a maior parte das respostas espectrais dos materiais na superfície da terra. Seguindo esta

linha, a distribuição dos padrões de resposta espectral pode ser completamente descrita pelo vetor de médias e pela matriz de covariância. Dados estes parâmetros, pode ser computada a probabilidade estatística de um dado pixel, por seu valor de brilho, pertencer a uma ou outra categoria particular de uso do solo. Funções de densidade de probabilidade são utilizadas para classificar um pixel desconhecido computando a probabilidade de que ele venha a pertencer a uma dada categoria. Depois de avaliar a probabilidade para cada categoria, o pixel é atribuído àquela categoria cuja probabilidade tenha sido a mais alta, ou a nenhuma delas no caso de os valores de probabilidade estarem abaixo de um determinado limite estipulado pelo analista.

A popularidade do classificador por máxima verossimilhança é devida a várias características, dentre elas podem ser citadas: a regra de decisão deste classificador é intuitivamente apelativa em virtude da classe mais provável ser a escolhida dentre todas as outras; a regra de decisão possui uma fundamentação teórica bem desenvolvida e, para dados normalmente distribuídos, é estatisticamente desejável e matematicamente tratável; esse algoritmo pode, sem problemas, acomodar dados covariantes, ocorrência comum em dados provenientes de imagens orbitais; e, finalmente, o classificador de máxima verossimilhança tem-se provado eficiente sob diferentes condições de cobertura do solo e sensores orbitais (BOLSTAD e LILLESAND, 1991).

CAMPBELL (1987) salienta que a regra de decisão por máxima verossimilhança, implementada quantitativamente para considerar diversas classes ou categorias e diversos canais espectrais, simultaneamente, gera uma poderosa técnica de classificação. Ela requer cálculos intensivos, necessitando, por esta razão, de mais recursos computacionais do que os outros classificadores.

Nesta regra de decisão cada pixel possuindo um padrão de medidas X é atribuído a uma determinada classe, cujos elementos que a compõem

sejam os mais prováveis de terem dado origem a este vetor de medidas X . JENSEN (1986) estabelece a conceituação matemática da teoria que embasa o algoritmo.

Seja M_C o vetor das médias dos valores de brilho dos pixels da classe C e V_C a matriz de covariância da classe C contemplando todas as bandas espectrais. De acordo com Anuta, 1977, citado por JENSEN, 1986, a regra de decisão, aplicada ao, até então, não-classificado vetor de valores X , é: atribua X à classe C se, e somente se

$$p_C \geq p_i, \text{ em que } i = 1, 2, 3, \dots, m \text{ possíveis classes}$$

em que p_C representa a probabilidade de o vetor X ser atribuído à classe C e p_i a probabilidade de o vetor X ser atribuído a qualquer outra classe. Continuando, tem-se:

$$p_C = [-0,5 \log_e (\det(V_C))] - [0,5(X - M_C)^T (V_C)^{-1} (X - M_C)]$$

em que $\det(V_C)$ é o determinante da matriz de covariância V_C . Dessa forma, para classificar o vetor X , a regra de decisão de máxima verossimilhança computa o valor p_C para cada classe ou categoria, assinalando, então, o pixel na classe que possuir o maior valor. A equação acima considera que as classes sejam equiprováveis no terreno. Não é difícil perceber que, para a maioria das aplicações em sensoriamento remoto, existe uma alta probabilidade de encontrar algumas classes mais freqüentemente do que outras. É possível se incluir esta informação, dita *a priori*, no processo de decisão.

Thomas et al. (1987), citados por FOODY et al. (1992), expressam a regra de decisão do classificador de uma forma diferente da anterior, considerando, agora, probabilidades *a priori*. Segundo eles,

$$P(x_k|i) = [1 / (2\pi)^{n/2} |V_i|^{1/2}] \exp [-1/2(x_k - u_i)^T V_i^{-1} (x_k - u_i)],$$

em que $p(x_k | i)$ é a função de densidade de probabilidade para um pixel x_k vir a ser um membro da classe i , n é o número de bandas ou canais espectrais, x_k é o vetor de dados para o pixel em todos os canais, u_i é o vetor de médias para a classe i , considerando-se todos os pixels, e V_i é a matriz de variância-covariância para a classe i . O termo

$$(x_k - u_i)^T V_i^{-1} (x_k - u_i)$$

é a distância de Mahalanobis entre o pixel considerado e o centróide da classe i , a qual representa a medida de quanto o pixel é típico naquela classe; a tipicidade é negativamente relacionada com a distância entre o pixel e o centróide da classe. As tipicidades podem ser derivadas das medidas obtidas pela distância de Mahalanobis com referência a uma distribuição F , embora uma aproximação de χ^2 seja freqüentemente utilizada. Isso pode ser percebido se a tipicidade for entendida como sendo a probabilidade de uma área na cauda da distribuição associada a um determinado caso para uma classe em particular (McKay e Campbell, 1982; Campbell, 1984; citados por FOODY et al., 1992).

O uso dessas probabilidades *a priori* vem sendo feito com sucesso, como um meio de incorporar os efeitos de relevo e outras características do terreno na melhoria da exatidão obtida pela classificação (Strahler, 1980, citado por JENSEN, 1986).

2.3. Técnicas de Amostragem

De acordo com o que é discutido por EBERT e LYONS (1983), amostragem é um processo cuidadoso de inspeção de uma pequena parte representativa de uma área no que se refere às características ou aos parâmetros nos quais se está interessado e que podem, então, ser projetados para uma área maior. De maneira semelhante, KELLY (1970)

define amostragem como sendo a seleção de uma parte de um agregado que representa todo o agregado. De acordo com os conceitos mencionados, é possível perceber a importância da representatividade da amostra. Se de antemão é sabido ser homogênea a população a ser amostrada, o processo de amostragem não necessita de grandes sofisticções, uma vez que qualquer amostra selecionada representará fidedignamente a população. Por outro lado, se a variabilidade da população é alta, situação normalmente encontrada em sensoriamento remoto, este processo reveste-se de toda uma série de técnicas e precauções para que as amostras a serem consideradas possam representar, de forma conveniente, o universo da população.

A coleta dos dados de referência, ou seja, daqueles que representam 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 exatidão da classificação, para descobrir confusão entre classes e como entrada no processo de refinamento da classificação (CONGALTON e BIGING, 1992).

Como expresso por WARREN et al. (1990), muito pouco tem sido escrito no que se refere à representatividade das amostras de campo em relação à interpretação das imagens orbitais. A extração casual de amostras no campo é um importante fator na confiabilidade estatística dos dados resultantes da interpretação. Todavia, em virtude de os diferentes tipos de cobertura de solo ou as características a serem consideradas não ocorrerem de forma casual na natureza, uma simples amostragem casualizada pode não representar adequadamente todas as categorias de cobertura de solo espectralmente discerníveis. A estratificação da amostragem pode melhorar as chances para que todas as categorias sejam adequadamente representadas. De forma semelhante, Card (1980), citado por CONGALTON (1984), comenta que não vêm sendo feitos estudos

comparativos entre amostragens sistemáticas em comparação com amostragens casualizadas. Na verdade, em razão dos altos custos envolvidos nos processos de coleta de dados e do tempo despendido neste tipo de tarefa, não têm sido feitas comparações diretas entre quaisquer esquemas de amostragem usados na verificação da exatidão dos mapas derivados dos dados obtidos por sensores remotos (CONGALTON e BIGING, 1992).

O que se depreende do parágrafo anterior pode ser extrapolado eliminando-se, por razões de custo e outras, as visitas a campo e, em vez delas, utilizar-se de recobrimentos aerofotogramétricos, desde que os mesmos tenham sido realizados em escala compatível com o objetivo do trabalho e em época suficientemente próxima àquela da obtenção da imagem orbital a ser utilizada, em concordância com o que afirmam CONGALTON e BIGING (1992), segundo os quais os dados de referência são freqüentemente coletados a partir de fotointerpretação de fotografias aéreas em grande escala ou sobrevoando-se a área em um helicóptero ou outra aeronave, tomando-se notas durante o percurso.

2.3.1. Os Padrões de Amostragem

Entende-se por padrão de amostragem a maneira, em termos espaciais, em que as amostras são tomadas sobre uma dada área. Em sensoriamento remoto, estes padrões são de fundamental importância, uma vez que, a depender deles, poder-se-ia estar incorrendo na incorporação de amostras com alta correlação espacial.

Como afirmado por CAMPBELL (1987), a correlação espacial é a tendência de as medidas em um dado local geográfico se assemelharem às tomadas em outros locais que se encontram na vizinhança imediata do primeiro. Esse conceito é importante em amostragem porque, idealmente, é desejável que cada observação forneça informações independentes. Se as observações estiverem pouco espaçadas entre si, elas estarão mensurando

informações espacialmente correlacionadas e, em virtude disso, fornecerão menos informações do que seria desejável. A respeito disso, GONG e HOWART (1990) recomendam a amostragem pixel a pixel como uma forma de minimizar os efeitos da correlação espacial. O que se pretende, em última análise, é que o procedimento de amostragem a ser utilizado deverá minimizar os efeitos da correlação espacial, bem como assegurar que todas as categorias de interesse sejam adequadamente amostradas (DICKS e LO, 1990). A amostragem por pontos, como preconizam Kendall e Buckland (1972), citados por LUND (1982), consiste simplesmente de um método de amostragem de uma área geográfica pela seleção de pontos em seu interior, mais especificamente pela escolha casual ou sistemática destes pontos em um mapa, uma fotografia aérea ou no campo. Estes pontos servirão, primariamente, de indicadores dos locais onde os dados serão coletados. O mesmo autor ainda comenta que é freqüente o não-entendimento do que, exatamente, seja a amostragem por pontos. Segundo ele, a falta de entendimento se daria por três razões: uso na literatura da mesma terminologia para designar diferentes métodos de amostragem; aversão à amostragem por pontos quando ela é, erroneamente, associada a um tipo particular de levantamento ou intensidade de amostragem; e, mais comumente, o não-reconhecimento da contribuição da amostragem por pontos quando ela está sendo aplicada.

Para a determinação do número ideal de pontos para serem testados em classificações de uso do solo, JENSEN (1986) recomenda uma relação oriunda da teoria de probabilidade binomial. A fórmula é expressa por:

$$N = [(4p)(q\sim)] / E^2$$

em que p é o percentual esperado de exatidão, q~ é a diferença entre 100 e p, E é o erro permissível e N é o número de pontos a serem amostrados.

2.3.2. *A Amostragem Blocada*

A questão relativa à forma como as amostragens são realizadas sobre a imagem orbital vem há muito sendo razão de discussão entre os especialistas da área de sensoriamento remoto. O que se faz corriqueiramente é coletar essas amostras em regiões da imagem onde se tem certeza de que a feição que ali se encontra corresponde àquela que se tem em mente. Esse tipo de abordagem, largamente utilizada, denominada amostragem blocada, peca pela falta de aleatoriedade do procedimento. Esse método viola o requisito de independência entre as amostras e, em consequência, faz com que os dados coletados para representar as diversas classes sejam menos representativos (GONG e HOWART, 1990).

2.3.3. A Amostragem Totalmente Casualizada

Neste caso específico, todos os indivíduos ou eventos possuem igual chance de serem selecionados. Dessa forma, uma vez determinado o número de indivíduos a serem amostrados, procede-se ao sorteio dos mesmos, seja por geração automática de números casualizados, seja pela utilização de tabelas apropriadas ou qualquer outro método que assegure a casualização do procedimento.

O principal problema com relação a essa abordagem está relacionado com a questão da uniformidade na distribuição das amostras, cuja importância aumenta na medida em que também aumenta a variação dos elementos da população presente na área de interesse. Não há nada que garanta esta uniformidade, fazendo com que, na maioria das vezes, algumas partes da área fiquem subamostradas e outras, superamostradas. No caso de imagens orbitais isto pode não representar um grande problema, desde que de antemão se saiba que a correlação espacial é bastante grande. Caso contrário, os resultados obtidos poderão ser de qualidade duvidosa.

2.3.4. A Amostragem Sistemática

A maioria dos autores de textos em estatística não recomenda este tipo de amostragem quando se pretende a extração de amostras não-tendenciosas e independentes para a representação de uma dada população (CONGALTON, 1988). Como afirma CAMPBELL (1987), em virtude de a seleção do ponto de partida condicionar o posicionamento de todos os outros pontos amostrais, os dados derivados deste tipo de amostragem não cumprem as exigências da estatística de inferência para a seleção casualizada de pontos. Existe uma grande diferença de opiniões a respeito dos valores das estimativas obtidas a partir da amostragem sistemática. A experiência prática tem mostrado que a amostragem sistemática, em muitos casos, produz estimativas úteis, do ponto de vista prático, e que a locação sistemática de linhas ou parcelas é, ao mesmo tempo, eficiente, de baixo custo e adequada aos processos de mapeamento (LOETSCH e HALLER, 1964). CONGALTON (1988) afirma ser a sua maior desvantagem o fato de que todos os indivíduos não possuem a mesma chance de serem incluídos na amostra.

Por outro lado, de acordo com STEHMAN (1992), amostras sistemáticas são amostras de igual probabilidade porque cada indivíduo possui a mesma chance de ser incluído na amostra. Para ilustrar o seu ponto de vista, ele exemplifica considerando um simples caso de amostragem sistemática de um universo discreto de sete unidades: y_1, y_2, \dots, y_7 . Se o intervalo para a amostragem sistemática for $k = 3$, uma das três possíveis amostras:

$$\text{amostra 1} = y_1, y_4, y_7$$

$$\text{amostra 2} = y_2, y_5$$

$$\text{amostra 3} = y_3, y_6$$

será selecionada, dependendo de que valor a casualização do ponto de partida assumir (1, 2 ou 3, respectivamente). Nesse caso, a probabilidade de que uma determinada unidade seja incluída na amostra será a probabilidade de que a amostra contendo a unidade seja selecionada. Em

virtude de todas as três amostras possuírem a probabilidade de 1/3 de serem selecionadas, todas as sete unidades terão a mesma probabilidade (1/3) de serem selecionadas. No caso da amostragem sistemática bidimensional de pixels, numa malha quadrada, o raciocínio se aplicaria da mesma forma. Com relação à não-representatividade quando o intervalo de amostragem está em fase com a periodicidade na população, o mesmo autor postula ser possível que uma determinada amostra, dentre outras, não seja representativa, porém essa asserção poderia ser aplicada a qualquer outro padrão de amostragem.

Os geógrafos defendem ponto de vista semelhante, de acordo com DOZIER e STRAHLER (1983), postulando que, no caso de correlação espacial e com um pequeno número de pontos a serem amostrados, a amostragem sistemática será representativa e mais precisa desde que não haja periodicidade nos dados coincidindo com o espaçamento sistemático das unidades da amostra.

2.3.5. A Amostragem Sistemática Estratificada Não-Alinhada

Este padrão de amostragem combina propriedades da amostragem sistemática e da amostragem estratificada, propiciando a cobertura de toda a área de interesse e, simultaneamente, preservando a casualidade do procedimento. Como explica CAMPBELL (1987), a área de estudo é dividida em células uniformes, geralmente por meio de um reticulado quadrado. As células deste retículo representam a porção sistemática do processo, enquanto a casualização é proporcionada pela localização das observações dentro de cada parcela. O resultado da localização destas observações dentro da área de estudo possui várias características favoráveis. O uso do reticulado de células como base para a estratificação significa que as observações são distribuídas de maneira a contemplar toda a extensão da área. A componente de casualização não permite o rígido alinhamento, característico da amostragem sistemática pura, introduzindo

um elemento de chance que aumenta a probabilidade de que as observações venham a contemplar as categorias presentes na região.

O procedimento para o lançamento dos pontos nesse tipo de amostragem se dá, inicialmente, dividindo a região a ser amostrada em um número de sub-regiões igual ao número de pontos a serem gerados. Na seqüência, é gerado casualizadamente um ponto em cada uma das sub-regiões. Percebe-se aí a importância do estabelecimento do número de pontos. Um número muito pequeno irá implicar uma sub-região maior para ser representada por um único ponto.

2.4. A Exatidão da Classificação

O teste estatístico mais simples para se julgar a exatidão dos mapas derivados de dados de sensores orbitais consiste em determinar um número ótimo de unidades de amostra a ser tomado no terreno e assumir uma margem de erros permissível. Então, pela comparação dos pontos amostrados no terreno com aqueles correspondentes obtidos no mapa gerado pelo processo de classificação, o número de erros pode ser determinado. Se o número de erros é menor ou igual à margem assumida, o mapa é aceito como exato dentro dos intervalos de confiança especificados (CONGALTON, 1984).

De acordo com CAMPBELL (1987), os erros estão presentes em qualquer tipo de classificação. Tais erros, quando decorrentes de processamento automatizado, são resultantes de interações complexas entre as estruturas espaciais da paisagem, a resolução do sensor, os algoritmos de pré-processamento e os procedimentos da própria classificação. Provavelmente as mais simples das causas de erro são relacionadas com o errôneo assinalamento de categorias informacionais nas categorias espectrais. Todavia, as fontes da maioria das causas de erros são mais complexas.

Ainda segundo o mesmo autor, a forma padronizada para reportar erros em locais específicos é a chamada matriz de erros, também conhecida como matriz de confusão, por identificar não somente o erro global da classificação para cada categoria, mas também como se deram as confusões entre categorias. Essencial para qualquer estudo sério sobre a exatidão dos processos de classificação, uma vez que por ela é possível a identificação, para cada categoria, dos erros de omissão e comissão, esta matriz é um exemplo específico de uma classe mais genérica de matrizes conhecidas como tabelas de contingência.

A verificação quantitativa da matriz de erros é geralmente necessária após a realização de uma inspeção inicial para se detectar a natureza genérica dos erros presentes. Alguns dos procedimentos que foram desenvolvidos para a análise de tabelas de contingência podem ser aplicados ao exame da matriz de erros.

Como sugerem GONG e HOWART (1990), um dos procedimentos que vem sendo usado para mensurar a exatidão das classificações temáticas é o índice Kappa. Ele vem sendo recomendado como uma medida apropriada da exatidão por representar inteiramente a matriz de confusão. Ele toma todos os elementos da matriz de confusão em consideração, em vez de apenas aqueles que se situam na diagonal da mesma, o que ocorre quando se calcula a exatidão global da classificação.

ROSENFELD e FITZPATRICK-LINS (1986) comentam ser o índice Kappa um coeficiente de concordância para escalas nominais que mede o relacionamento entre a concordância, além da casualidade, e a discordância esperada. A estimativa de Kappa (\bar{K}) é a proporção de concordância depois que a concordância em virtude da casualidade é retirada de consideração, ou seja:

$$\bar{K} = (p_0 - p_c) / (1 - p_c)$$

em que

p_0 = proporção de unidades que concordam; e

p_c = proporção de unidades que concordam por casualidade.

Simplificando o raciocínio citado, CAMPBELL (1987) propõe:

$$\bar{K} = (\text{observado} - \text{esperado}) / (1 - \text{esperado})$$

Neste caso, entenda-se por **observado** o valor global para a percentagem correta, ou seja, o somatório dos elementos presentes na diagonal da matriz dividido pelo total de elementos. Por **esperado** entenda-se os valores calculados usando-se os totais de cada linha e de cada coluna da matriz, designados pelo autor como marginais. Os produtos dos marginais de linhas pelos das colunas estimam o número de elementos (pixels) assinalados para cada célula da matriz, dado que esses pixels são assinalados por chance em cada categoria. Com a finalidade de ilustrar o procedimento, mostra-se, em seguida, um exemplo adaptado daquele citado por CAMPBELL (1987):

Seja uma matriz de confusão **X**, a seguir explicitada, em que as linhas dizem respeito aos dados de referência (“verdade de campo”) para quatro categorias e as colunas se referem aos dados obtidos após a classificação da imagem, para as mesmas quatro categorias. Evidentemente, cada elemento X_{ij} desta matriz representa o número de pixels coincidentes entre as categorias dos dados de referência e aquelas geradas pela classificação, cujo elemento se encontra na interseção da linha com a coluna. Assim sendo, para o cálculo do observado correto, tem-se:

	A_C	B_C	C_C	D_C	T_R
A_R	35	14	11	1	61
B_R	4	11	3	0	18

C_R	12	9	38	4	63
D_R	2	5	12	2	21
T_c	53	39	64	7	

total geral (nº de pixels total) = 163

total correto (soma dos elementos na diagonal) = 86

observados corretos = 86/163 = 0,528

percentual correto = 53%

Da análise da matriz citada percebe-se que os valores sombreados em tom mais claro (diagonal) representam para cada categoria o número de pixels corretamente classificados, enquanto as células mais escuras contêm os totais para as linhas e para as colunas (totais marginais).

Para o cálculo dos esperados corretos, são utilizados os totais marginais para cada categoria, o da linha e o da coluna. A justificativa estatística para se multiplicar estes dois totais reside na seguinte propriedade das probabilidades:

$$p(A+B) = (p(A)).(p(B))$$

ou seja, a probabilidade de um conjunto de pixels estar em uma mesma categoria em ambos os conjuntos de dados (referência e classificação) é o produto das probabilidades isoladas de estarem presentes nesta categoria em cada um dos conjuntos de dados.

Calculando-se os produtos dos totais marginais, tem-se:

3233	2379	3904	427
954	702	1152	126
3339	2457	4032	441
1113	819	1344	147

Concordância esperada por casualidade = soma dos elementos da diagonal / total:

$$= 8114 / 26569 = 0,305$$

A partir destes cálculos, calcula-se o índice Kappa:

$$\bar{K} = (\text{observado} - \text{esperado}) / (1 - \text{esperado})$$

$$\bar{K} = (0,528 - 0,305) / (1 - 0,305) = 0,321$$

Com base neste último índice, que pode variar de 0 a 1, podem ser feitas comparações de eficiência entre diferentes algoritmos de classificação, desde que os resultados por eles produzidos sejam comparados com um mesmo conjunto de dados de referência.

O trabalho de FOODY et al. (1992) afirma que, de acordo com a relação

$$\bar{K} = (\text{observado} - \text{esperado}) / (1 - \text{esperado}),$$

o grau de concordância por chance pode ser superestimado pelo fato de ele incluir, além da concordância por chance, a concordância real. Em consequência disso, a magnitude de Kappa não estaria refletindo a concordância presente na classificação, descontada a concordância resultante somente da casualidade. Considerando-se que não sejam fixadas probabilidades *a priori* para as categorias, a proporção marginal para cada uma delas seria dada por $1/m$, em que m é o número de categorias na classificação. Conseqüentemente, a probabilidade de concordância por chance será $1/m$, gerando um outro índice para medir a exatidão da classificação, que seria definido por:

$$K_m = (\text{observado} - 1/m) / (1 - 1/m).$$

Em recente pesquisa, MA e REDMOND (1995) concordam com o que propõe FOODY et al. (1992), para o caso de não se estabelecerem probabilidades *a priori* e estabelecendo outro procedimento quando estas probabilidades são diferentes. Os autores batizam este índice como Tau, T_e , quando ocorre a equiprobabilidade entre as categorias. Dessa forma, T_e torna-se um ajuste do percentual de concordância (aquele que considera somente os elementos da diagonal da matriz de erros, também chamado de concordância global) pelo número de categorias. À medida que este número aumenta, T_e se aproxima da percentagem de concordância global.

O índice T_e é expresso por:

$$T_e = \frac{P_0 - 1/M}{1 - 1/M}, \text{ em que}$$

$$P_0 = \frac{\sum_{i=1}^M n_{ii}}{N} \quad \mathbf{e}$$

M = número de categorias informacionais.

Neste caso, e de acordo com a matriz constante do exemplo anterior, o valor do índice Tau, considerando-se as quatro categorias ou classes presentes na matriz de erro, seria:

$$T_e = \frac{0,528 - 1/4}{1 - 1/4} = 0,37$$

Ainda de acordo com MA e REDMOND (1995), para o caso de se desejar testar diferentes classificações, as seguintes condições devem ser satisfeitas:

- a) os pixels deverão ser amostrados independentemente;

b) as categorias de cobertura de solo deverão ser independentes e mutuamente exclusivas; e

c) as classificações deverão ser processadas independentemente.

Em virtude de a concordância global (P_0) seguir uma distribuição binomial quando o número de pixels usado para construir a matriz de erros é grande ($N > 100$), ela irá se aproximar de uma distribuição normal (Steel e Torrie, 1960, citados por MA e REDMOND, 1995). Uma vez que a distribuição de P_0 se aproxima da normalidade, a distribuição de Tau também se aproximará (Cohen, 1960, citado por MA e REDMOND, 1995).

Considerada a normalidade, uma forma geral para um intervalo de confiança de 95%, para um coeficiente de exatidão (C), pode ser expressa por

$$\text{intervalo de confiança (95\%)} = C \pm 1,96\sigma .$$

Para testar a significância entre dois coeficientes de exatidão (C_1 e C_2), derivados de duas classificações independentes, o teste estatístico tomará a forma:

$$Z = \frac{C_1 - C_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} .$$

Em virtude de a concordância por casualidade (P_c) ser independente dos elementos da matriz de erros e poder ser calculada antes que se processe a classificação, ela pode ser tratada como uma constante. A variância de Tau pode ser derivada da seguinte maneira:

$$s^2(P_c) = 0$$

e

$$s^2(P_0) = \frac{1}{N} P_0(1 - P_0) ,$$

tal que

$$\begin{aligned}\sigma^2 (T_e) &= \sigma^2 \frac{P_0 - P_c}{(1 - P_c)^2} \\ &= \frac{\sigma^2 (P_0 - P_c)}{(1 - P_c)^2} \\ &= \frac{\sigma^2 (P_0) - \sigma^2 (P_c)}{(1 - P_c)^2} \\ &= \frac{P_0(1 - P_0)}{N(1 - P_c)^2}\end{aligned}$$

Obtida a variância para Tau, testes podem ser feitos para determinar se o coeficiente é significativamente maior que zero ou se dois coeficientes Tau, derivados de classificações independentes, são significativamente diferentes.

Ainda de acordo com MA e REDMOND (1995), existem pelo menos três razões para justificar o uso do coeficiente Tau na verificação da exatidão das classificações baseadas em dados de sensores remotos:

a) Conceitualmente, o coeficiente Tau é mais fácil de entender e interpretar do que o Kappa. A concordância por chance (P_c), usada para calcular Tau, é uma medida direta do assinalamento casual; em razão disso, Tau provê uma clara indicação de como a classificação se comporta quando comparada com a atribuição casual dos pixels às categorias ou classes. Com Kappa, por outro lado, a concordância por chance (ou casual) pode resultar do julgamento do analista adicionado à casualidade, de tal forma que este coeficiente ou índice superestima a concordância por chance (P_c) (Brennan e Prediger, 1981, citados por MA e REDMOND, 1995).

b) Tau e sua estimativa de variância são relativamente simples de serem calculados. Em razão de a concordância por chance, no cálculo de Tau, ser independente da matriz de erros, ela pode ser conhecida antes da construção da matriz.

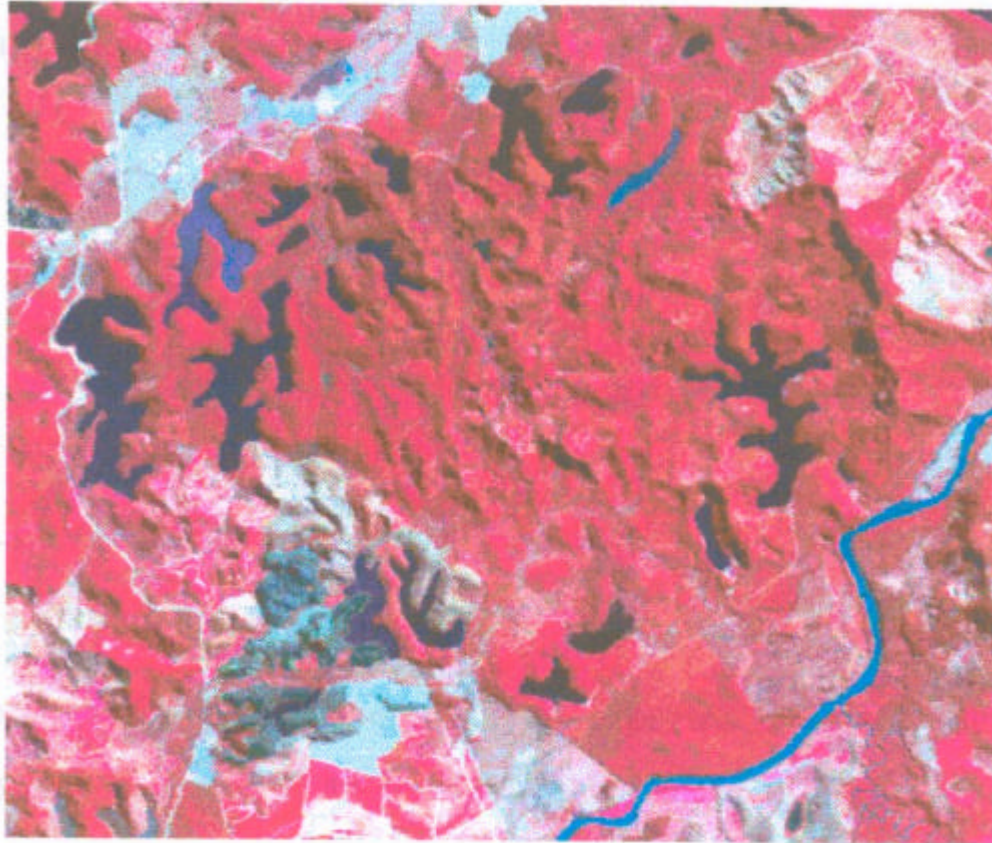
c) Diferentemente de Kappa, Tau toma em consideração o número de categorias ou classes presentes na classificação.

3. MATERIAL E MÉTODOS

3.1. *Localização da Área de Estudo*

A área de estudo está localizada no Vale do Rio Doce, Estado de Minas Gerais, compreendida entre as coordenadas UTM 745.595 N, 7.797.446 E, 758.315 N e 7.808.096 E. Esta área compreende uma superfície de aproximadamente 13.500 ha, incluindo plantações de *Eucalyptus*, florestas naturais, pastagens, lagos, rios e áreas de solo exposto. De acordo com GOLFARI (1975), a região possui terreno relativamente rugoso, com uma elevação média de 350 m, variando entre 150 e 650 m. As terras são de propriedade da Companhia Agrícola e Florestal Santa Bárbara, que utiliza suas florestas cultivadas para a produção de carvão. A escolha desta área se deveu à disponibilidade dos dados de inventário florestal, da imagem orbital, e, principalmente, do recobrimento fotogramétrico realizado na mesma época da obtenção da imagem, além do fato de ela se localizar em região de fácil acesso para as verificações de campo que fossem necessárias. Na Figura 3 tem-se uma visão geral da área.

Composição Colorida das Bandas TM4, TM3 e TM2, RGB



Composição colorida das bandas TM4, TM e TM2, obtida a partir de imagem orbital adquirida pelo satélite LANDSAT 5, órbita 2\7, ponto 74, em 15/07/1989.

Figura 3 - Visão geral da área de estudo situada ao sul do Parque Estadual do Rio Doce, município de Dionísio, Estado de Minas Gerais.

4.2. *Dados Utilizados*

Durante o desenrolar deste estudo, várias foram as fontes de dados utilizadas, descritas a seguir.

3.2.1. *Imagem Orbital*

Foi utilizada uma imagem obtida pelo sensor TM (Thematic Mapper), a bordo do satélite Landsat 5, órbita 217 e ponto 74 (quadrante noroeste), em sua passagem de 15 de julho de 1989, contemplando as seguintes bandas espectrais:

Banda	λ (μm)	Res.(m)
1	0,45 - 0,52	30
2	0,52 - 0,60	30
3	0,63 - 0,69	30
4	0,76 - 0,90	30
5	1,55 - 1,75	30
7	2,08 - 2,35	30

A banda 6 (10,4 - 12,5 *mm*) não foi utilizada por representar a porção do infravermelho emitida no espectro das radiações eletromagnéticas e possuir resolução espacial da ordem de 120 m.

3.2.2. *Fotografias Aéreas*

Fotografias em preto e branco, na escala de 1:20.000, foram obtidas sobre toda a área de estudo, em outubro de 1989, três meses após a obtenção da imagem orbital, tornando-se, dessa forma, excelente recurso adicional para a seleção das áreas de treinamento e teste para o processo de classificação da imagem orbital, bem como para a obtenção de dados adicionais não constantes do inventário.

3.2.3. Mapas

Foram utilizados mapas topográficos na escala de 1:100.000 e mapas discriminando os limites dos talhões na escala de 1:25.000, sendo os últimos preparados pela empresa proprietária dos plantios.

3.2.4. Dados de Campo

Os dados de inventário dos talhões presentes no mapa foram preparados pela empresa e colocados à disposição para toda a área em estudo. Estes dados incluem espécies de *Eucalyptus*, classes de idade, altura média das árvores, DAP médio, área basal por ha, densidade (número de indivíduos por ha) e volume (m³/ha com e sem casca).

3.3. Sistemas Computacionais e Equipamentos Utilizados

Para a execução da pesquisa foi utilizado o sistema IDRISI versão 1.0, para Windows, a planilha de cálculos EXCEL e o editor gráfico PHOTOSTYLER, utilizando-se um microcomputador compatível com o IBM PC-At com processador 80486 DX2, 66 Mhz.

3.3.1. Periféricos

- Impressora Laser;
- Unidade de disco flexível 3,5";
- Mesa digitalizadora, formatoA0; e
- Impressora em cores, jato de tinta, com resolução de 300 dpi.

3.3.2. Equipamentos Especiais

Estereoscópio de Espelhos para interpretação das fotografias aéreas.

3.4. Metodologia

3.4.1. A Coleta das Amostras de Treinamento

As amostras que foram utilizadas no treinamento do classificador foram coletadas de forma blocada, a partir da interpretação das fotografias aéreas, com o auxílio de um estereoscópio de espelhos e de oculares de aumento acopladas a ele. Do próprio conhecimento da região e pela interpretação das fotos aéreas, em estudo previamente conduzido, verificou-se que as seguintes categorias informacionais seriam capazes de expressar a variabilidade presente na área de estudo:

- Rebrota sobre Solo Vegetado (RSV)

Categoria caracterizada pela rebrota do eucalipto sobre solo coberto por vegetação essencialmente herbácea-arbustiva.

- Rebrota sobre Solo Parcialmente Exposto (RSPE)

Caracterizada pela rebrota do eucalipto sobre solo que apresenta partes totalmente expostas, sem qualquer tipo de vegetação.

- Rebrota sobre Solo Exposto (RSE)

Caracterizada pela rebrota do eucalipto sobre solo totalmente exposto.

- Eucalipto com Fechamento de Copa (EUFECH)

Caracterizada por plantios de eucaliptos, já em idade adulta, com fechamento de copa, sem que haja exposição do solo.

- Eucalipto sem Fechamento de Copa (EUABER)

Caracterizada por plantios de eucaliptos, já em idade adulta, com falhas impedindo que ocorra o fechamento de copa e permitindo a exposição do solo, vegetado ou não.

- Sombras (SOMBRAS)

Categoria constituída por áreas da imagem que se encontravam sob sombra, impossibilitando a correta identificação dos pixels.

- Herbácea-Arbustiva (HERB_ARB)

Categoria constituída por vegetação de natureza herbácea-arbustiva.

- Solo Exposto (SOLO_EXP)

Categoria constituída por solo descoberto, isto é, desprovido de cobertura vegetal.

- Água Não-Turva (AGUALIMP)

Água com pouca ou nenhuma turbidez.

- Água Turva (AGUATURV)

Água com muito material em suspensão, acarretando turbidez.

- Nativas (NATIVAS)

Vegetação arbórea, nativa da região.

3.4.2. *O Algoritmo de Classificação*

Todas as classificações levadas a efeito no estudo foram produzidas aplicando-se um algoritmo de máxima verossimilhança, provido pelo sistema IDRISI (programa MAXVER), assumindo-se a equiprobabilidade de ocorrência das categorias listadas anteriormente. Este algoritmo exige como entrada a definição de probabilidades *a priori* e os nomes dos arquivos de assinaturas a serem usados no processamento. A não-especificação de probabilidades para nenhuma das classes implica equiprobabilidade.

3.4.3. *As Combinações de Bandas e Transformações Especiais*

O algoritmo de máxima verossimilhança foi aplicado às seguintes combinações de bandas, gerando as classificações à direita:

COMBINAÇÃO	CLASSIFICAÇÃO
TM3, TM4, TM5	MAX345
TM3, TM4, TM7	MAX347
TM2, TM3, TM4	MAX234

Da mesma forma ele foi aplicado às componentes “verdor”, “brilho” e “umidade” geradas a partir da aplicação dos coeficientes mostrados a seguir às bandas reflexivas da imagem, de acordo com JENSEN (1986):

	1	2	3	4	5	7
BRILHO	0,33183	0,33121	0,55177	0,42514	0,48087	0,25252
VERDOR	-0,24717	-0,16263	-0,40639	0,85468	0,05493	-0,11749
UMIDADE	0,13929	0,22490	0,40359	0,25178	-0,70133	-0,45732

A classificação gerada a partir destas componentes recebeu o nome de MAXTC.

Uma última transformação utilizada foi a resultante da análise de componentes principais, cujas matrizes que caracterizam os dados estão explicitadas no Quadro 3. As três primeiras componentes principais foram escolhidas por contemplarem 99,15% da variância presente em toda a imagem, ao considerar todas as bandas reflexivas. A classificação resultante da utilização destas componentes recebeu o nome de MAXPC123.

3.4.4. A Amostragem sobre os Dados de Referência

A amostragem sobre os dados de referência foram conduzidas de acordo com as estratégias blocada, sistemática estratificada, sistemática e casualizada.

Quadro 3 - Matrizes que caracterizam os dados utilizados e gerados na análise de componentes principais

MATRIZ DE VARIÂNCIA / COVARIÂNCIA						
	TM1	TM2	TM3	TM4	TM5	TM7
TM1	14,52					
TM2	13,18	14,53				
TM3	26,31	27,46	57,53			
TM4	12,18	16,22	24,95	249,19		
TM5	66,15	65,69	136,65	217,74	573,25	
TM7	26,99	26,36	56,44	45,38	199,40	82,02
MATRIZ DE CORRELAÇÕES						
	TM1	TM2	TM3	TM4	TM5	TM7
TM1	1					
TM2	0,907366	1				
TM3	0,910336	0,9496	1			
TM4	0,202554	0,2695	0,2084	1	,	
TM5	0,725144	0,7196	0,7525	0,5761	1	
TM7	0,782301	0,7634	0,8216	0,3174	0,9196	1
MATRIZ DE COMPONENTES						
COMP.	C 1	C 2	C 3	C 4	C 5	C 6
% DA VAR.	79,40	17,21	2,55	0,55	0,18	0,12
AUTOVAL.	786,86	170,51	25,31	5,40	1,75	1,22
autovet. 1	0,968600	-0,122568	0,325654	-0,050684	0,589832	0,720468
autovet. 2	0,098622	-0,102753	0,405759	-0,049737	0,578549	-0,691290
autovet. 3	0,200706	-0,259187	0,749669	-0,107172	-0,563161	0,043579
autovet. 4	0,380754	0,885938	0,214049	0,153230	-0,014309	0,025273
autovet. 5	0,844579	-0,219127	-0,346475	-0,343693	0,004338	-0,021947
autovet. 6	0,286918	-0,272650	-0,037982	0,917533	0,002179	-0,005026

No caso da amostragem blocada, foram selecionados, sobre a imagem, polígonos representativos das 11 categorias informacionais em concordância com a interpretação das fotos aéreas. Estes polígonos foram delineados sem que pixels que já tivessem sido amostrados na etapa de treinamento fossem contemplados. Para que isso fosse conseguido, foi sobreposto o arquivo vetorial, contendo os polígonos de treinamento, à imagem sobre a qual seriam coletados os novos polígonos que serviriam para a verificação da exatidão das classificações. Os polígonos coletados, no formato vetorial, foram então transformados para o formato matricial ou *raster*, gerando um novo arquivo.

Para as amostragens não-blocadas, o procedimento foi bastante semelhante. Utilizando o algoritmo denominado SAMPLE, do sistema IDRISI, foram gerados arquivos vetoriais, um para cada amostragem não-blocada, contendo, inicialmente, 1.200 pontos cada. Este número de pontos foi determinado empiricamente, objetivando contemplar todas as categorias presentes na imagem, especialmente aquelas que ocorrem com pouca frequência, ocupando pequenas áreas. De acordo com JENSEN (1986), e como explicado no item 2.4.1, o número mínimo de pontos pode ser determinado pela relação:

$$N = [4(p)(q\sim)] / E^2$$

em que

p = percentual de exatidão esperado;

q \sim = 100 - p; e

E = erro permissível.

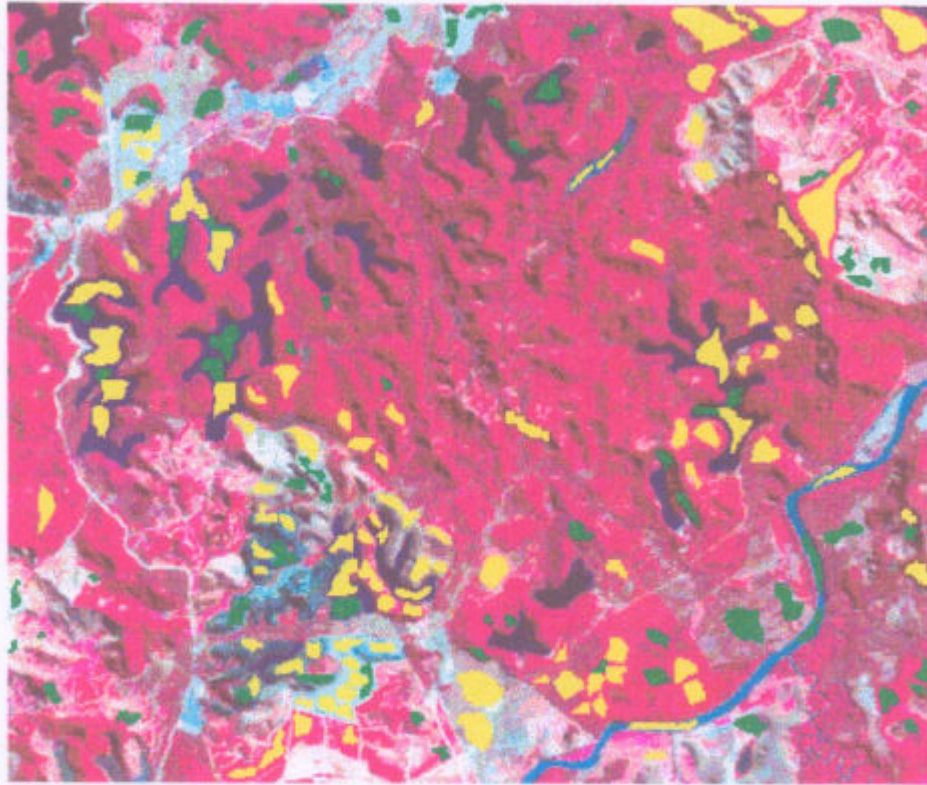
No presente caso foi considerado um percentual de exatidão de 85% e um erro admissível de 5%, o que implica, após a aplicação da relação citada, um número mínimo de 204 pixels a serem amostrados.

À semelhança do que ocorreu com a amostragem blocada, para as outras amostragens também foram desconsiderados aqueles pixels que foram tomados para o treinamento do classificador, o que justifica o fato de o

número de pixels amostrados em cada procedimento não-blocado de amostragem ser inferior a 1.200. As Figuras 4, 5, 6 e 7 ilustram cada uma das estratégias amostrais utilizadas. Os pontos amostrais sobre cada uma das imagens são os 1.200 pontos inicialmente lançados, ou seja, antes de desconsiderar aqueles que haviam sido contemplados pelas amostras de treinamento. Para este número de pontos isto implicaria, mantido o percentual de exatidão esperado, a redução do erro permissível para 2,06%.

De posse dos arquivos vetoriais contendo os pontos a serem amostrados, eles foram sobrepostos a diferentes composições de bandas que permitissem a melhor observação dos diversos tipos de cobertura vegetal. Dessa forma, ora foi utilizada a composição em falsa cor dos canais TM4, TM3 e TM2, ora a imagem resultante da primeira componente principal e outras. Cada ponto presente sobre a imagem orbital foi localizado nas fotografias e, com o auxílio de um estereoscópio de espelhos, foi verificada a que categoria informacional o pixel sob aquele ponto pertenceria. Para que a correspondência imagem/fotografia fosse a mais exata possível, lançou-se mão de operações de ampliação sobre a imagem e do uso de oculares de aumento com o estereoscópio, bem como de um reticulado auxiliar que subdividia a imagem em 16 subimagens, para que o analista não se perdesse na tarefa de identificar cada um dos pontos. Dessa forma, a cada ponto que era identificado, digitalizava-se um outro coincidente com o que havia sido sobreposto anteriormente, com o atributo referente à categoria informacional a que ele pertencesse. Em resumo, o arquivo contendo os pontos amostrais foi sobreposto à imagem apenas para servir de orientação locacional na coleta dos pontos que comporiam o arquivo de referência. Essa seqüência foi repetida para todos os procedimentos não-blocados de amostragem, gerando novos arquivos vetoriais que, em seguida, foram submetidos a um processo que os

Amostragem Blocada



Os polígonos em amarelo foram utilizados para o treinamento do classificador de máxima verossimi-lhança, enquanto aqueles em verde foram tomados para a verificação da exatidão.

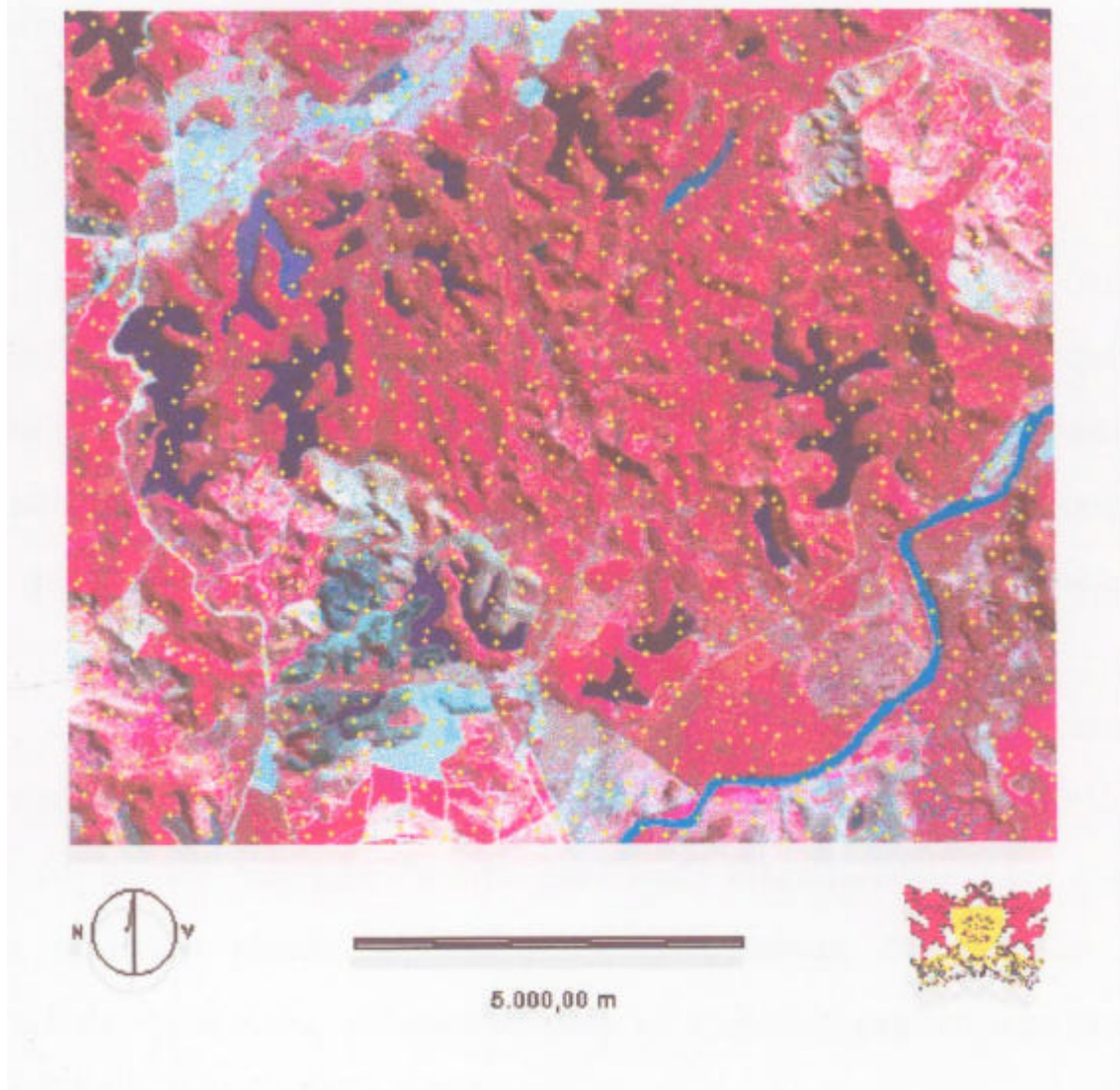
Figura 4 - Polígonos utilizados na amostragem blocada.



Os pixels que coincidiram, locacionalmente, com aqueles tomados para o treinamento do classificador foram desconsiderados.

Figura 5 - Disposição inicial dos pixels, sobre a imagem, para servirem de base à amostragem casualizada.

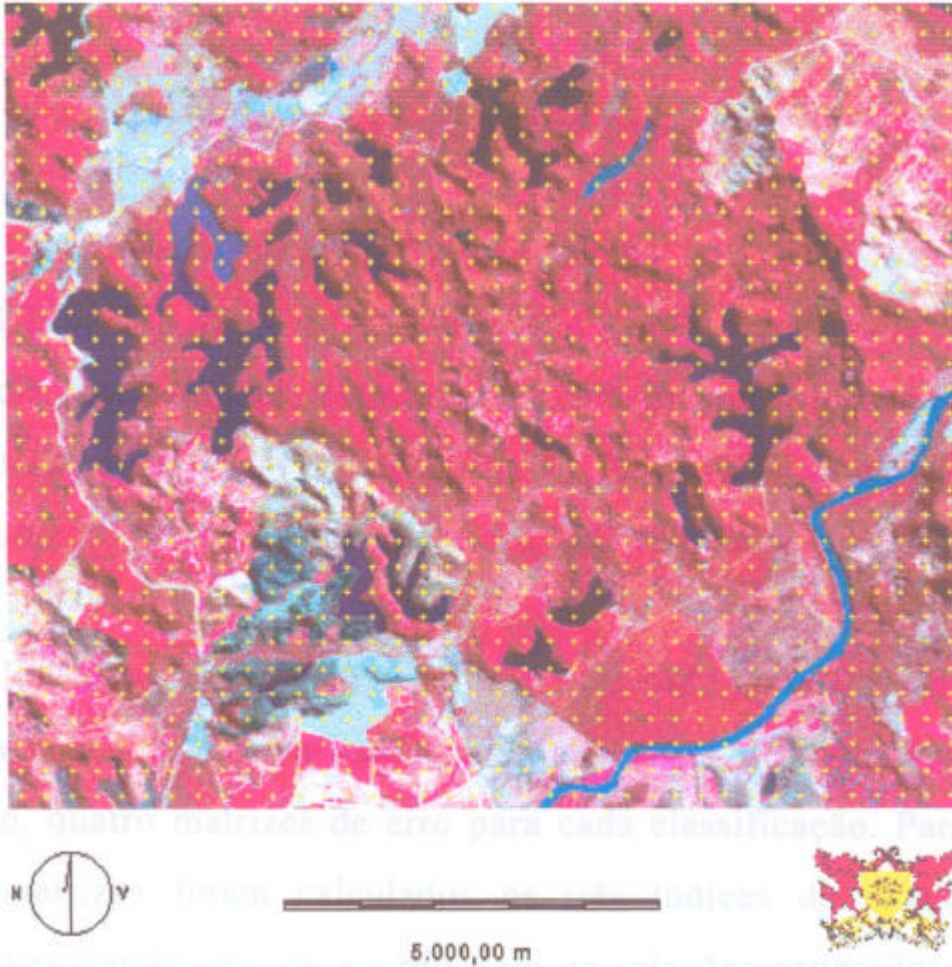
Amostragem Sistemática Estratificada



Os pixels que coincidiram, locacionalmente, com aqueles tomados para o treinamento do classificador foram desconsiderados.

Figura 6 - Disposição inicial dos pixels, sobre a imagem, para servirem de base à amostragem sistemática estratificada.

Amostragem Sistemizada



Os pixels que coincidiram, locacionalmente, com aqueles tomados para o treinamento do classificador foram desconsiderados.

Figura 7 - Disposição inicial dos pixels, sobre a imagem, para servirem de base à amostragem sistemizada.

transformaram em arquivos matriciais. Estes arquivos foram, então, submetidos a uma operação de álgebra cartográfica, na qual cada um deles foi multiplicado por um arquivo booleano em que os pixels no interior dos polígonos tomados como amostras de treinamento receberam o valor zero, e os demais, o valor um. Assim, os arquivos de referência passaram a conter unicamente pixels que não faziam parte das amostras de treinamento, estando prontos para serem usados nas verificações de exatidão.

3.4.5 O Cálculo dos Índices e de Suas Variâncias

Uma vez estabelecidas as cinco classificações (MAXTC, MAXPC123, MAX345, MAX347 e MAX234) e os quatro arquivos de referência, um para cada estratégia de amostragem (blocada, casualizada, sistemática estratificada e sistematizada), foi possível gerar as matrizes de erros que possibilitariam o cálculo dos índices de exatidão (Global, Kappa e Tau).

Foi gerada uma matriz de erro para cada classificação comparada com cada um dos quatro arquivos de referência, resultando, ao final do processo, quatro matrizes de erro para cada classificação. Para cada uma destas matrizes foram calculados os três índices de exatidão e suas respectivas variâncias, de acordo com as relações específicas para cada índice (MA e REDMOND, 1995).

O índice Exatidão Global, bem como sua variância, são determinados pelas seguintes relações:

$$G = \frac{\sum_{i=1}^M n_{ii}}{N}$$

$$s_G^2 = \frac{P_0(1 - P_0)}{N}$$

De forma semelhante, para o índice Kappa, tem-se:

$$K = \frac{P_0 - P_C}{1 - P_C}$$

$$s_k^2 \approx \frac{P_0(1 - P_0)}{N(1 - P_c)^2}.$$

Para o índice Tau, tem-se:

$$T = \frac{P_0 - 1/M}{1 - 1/M}$$

$$s_T^2 = \frac{P_0(1 - P_0)}{N(1 - 1/M)^2}.$$

As relações para P_0 e P_c podem ser explicitadas por:

$$P_0 = \frac{\sum_{i=1}^M n_{ii}}{N} \quad \mathbf{e}$$

$$P_c = \frac{\sum_{i=1}^M n_{i+} n_{+i}}{N^2}$$

em que

N = número total de pixels contemplados pela matriz de erros;

n = elemento da matriz de erros;

M = número de categorias informacionais presentes na matriz de erros;

$i+$ = total da linha para uma dada categoria informacional; e

$+i$ = total da coluna para esta mesma categoria informacional.

No caso específico de Kappa, foram calculados também os valores para cada categoria informacional, além do valor relativo à matriz de erros inteira.

De posse dos índices e de suas respectivas variâncias, foi possível efetuar comparações entre as classificações e o índices.

3.4.6. *As Comparações dos Índices de Exatidão*

As comparações foram feitas a partir dos valores absolutos dos índices, com base nos valores de Z a 99% de probabilidade e, graficamente, com a finalidade de melhor visualizar o comportamento dos três índices em função da variação das estratégias de amostragem e classificações.

Para testar a significância da diferença entre quaisquer dois índices de exatidão (C_1 e C_2), o teste estatístico utilizado foi o de Z, especificado a seguir, como preconizam MA e REDMOND (1995):

$$Z = \frac{C_2 - C_1}{\sqrt{\sigma_{K_2}^2 + \sigma_{K_1}^2}}$$

Essa relação pode ser verbalmente expressa como: o valor de Z será igual ao resultado da razão obtida pela diferença entre os índices pela raiz quadrada da soma das variâncias destes mesmos índices.

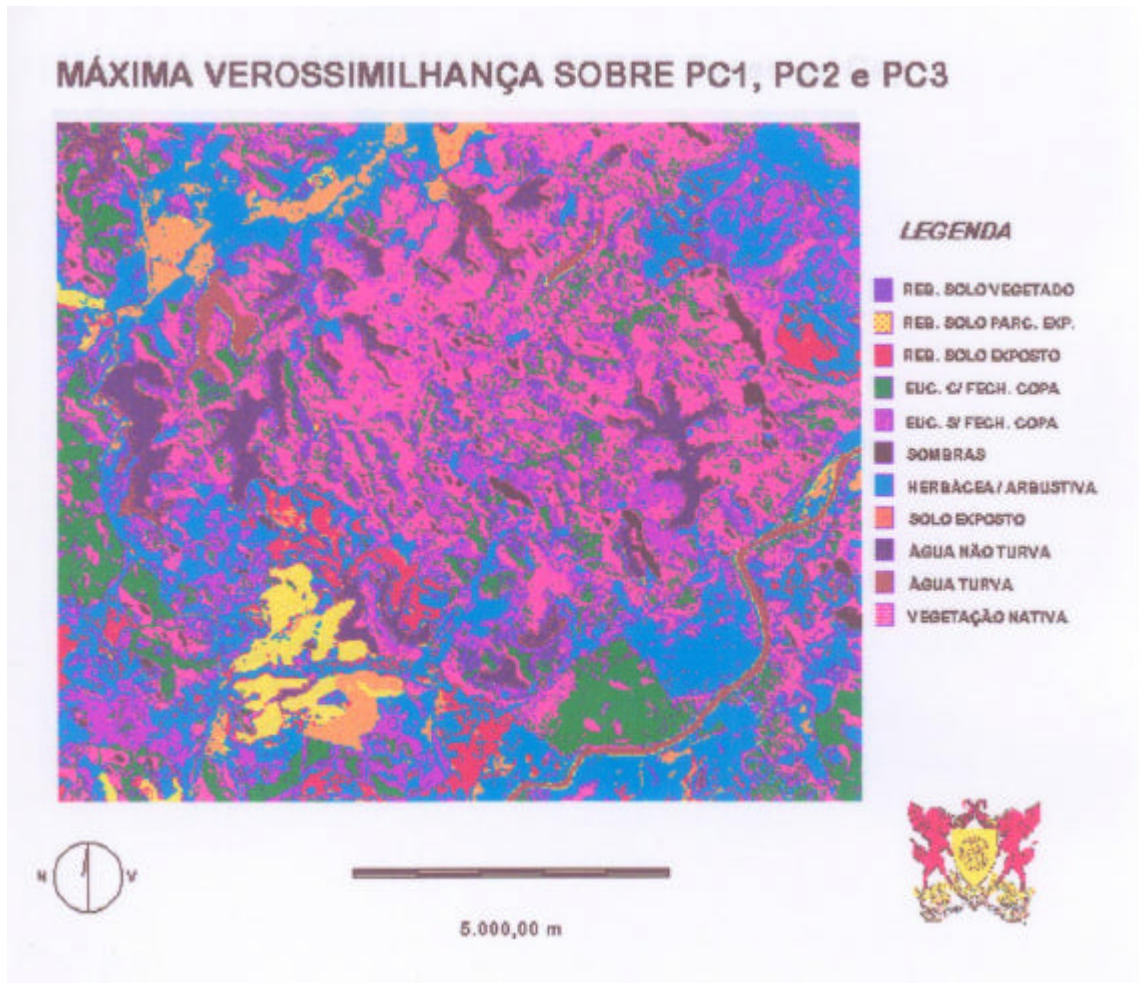
4. RESULTADOS E DISCUSSÃO

São mostrados, a seguir, os resultados alcançados no estudo, bem como a discussão sobre as principais implicações dos mesmos.

4.1. *As Classificações Geradas*

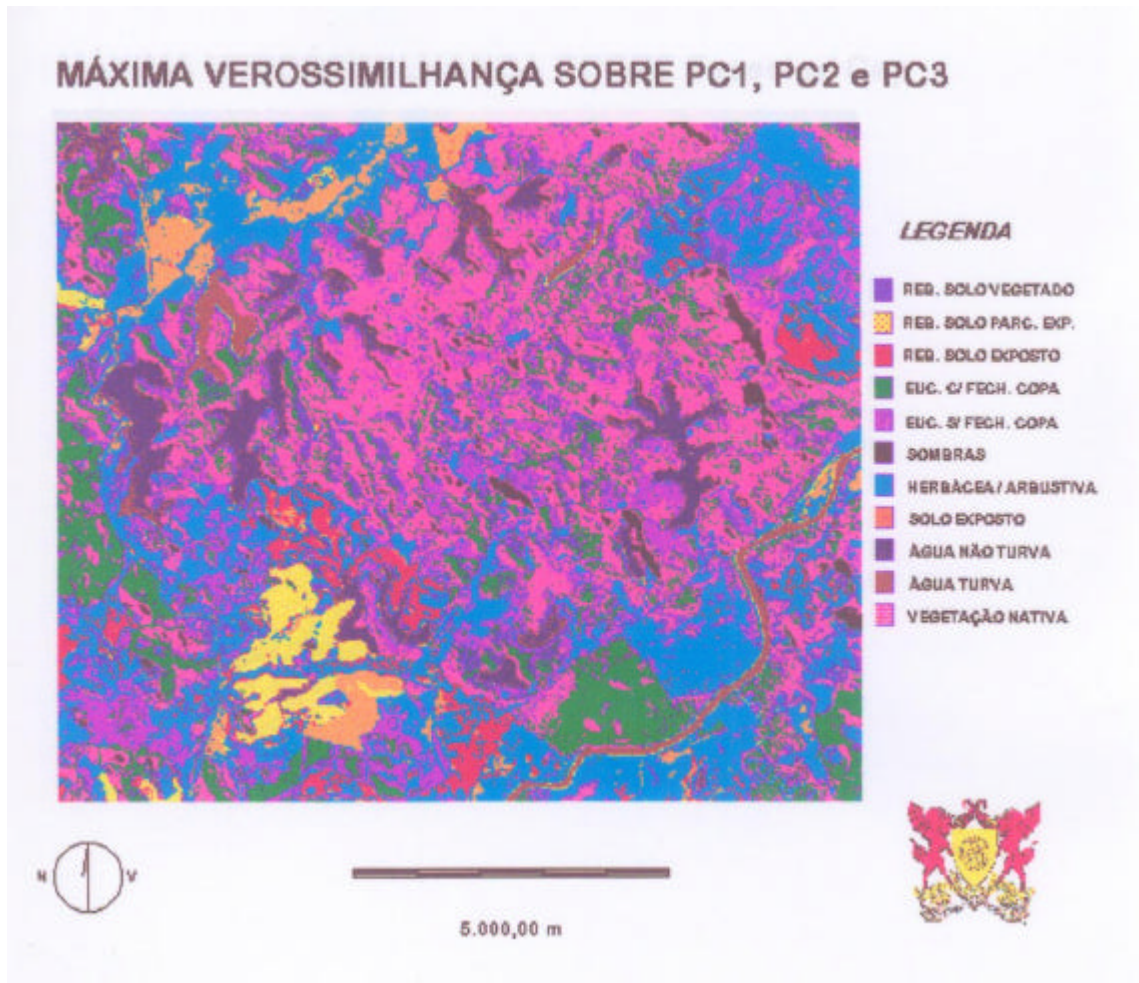
Durante o estudo foram obtidas cinco classificações supervisionadas que utilizaram o mesmo conjunto de dados de treinamento para o classificador de máxima verossimilhança (Figuras 8, 9, 10, 11 e 12). A única diferença entre elas, no que se refere aos dados de entrada, reside nas diferentes combinações de bandas espectrais ou transformações especiais.

Embora não seja, visualmente, muito perceptível a diferença entre as classificações, é possível notar, com base no gráfico da Figura 13, que o maior índice de exatidão (neste caso, Kappa) foi obtido com o uso da transformação *Tasseled Cap*, à exceção da ocasião em que se utilizou a amostragem sistemática para gerar o arquivo de referência. Esta transformação, embora ainda pouco utilizada, vem mais uma vez confirmar a sua grande validade nos estudos que envolvem discriminação e, ou, tipificação de coberturas vegetais.



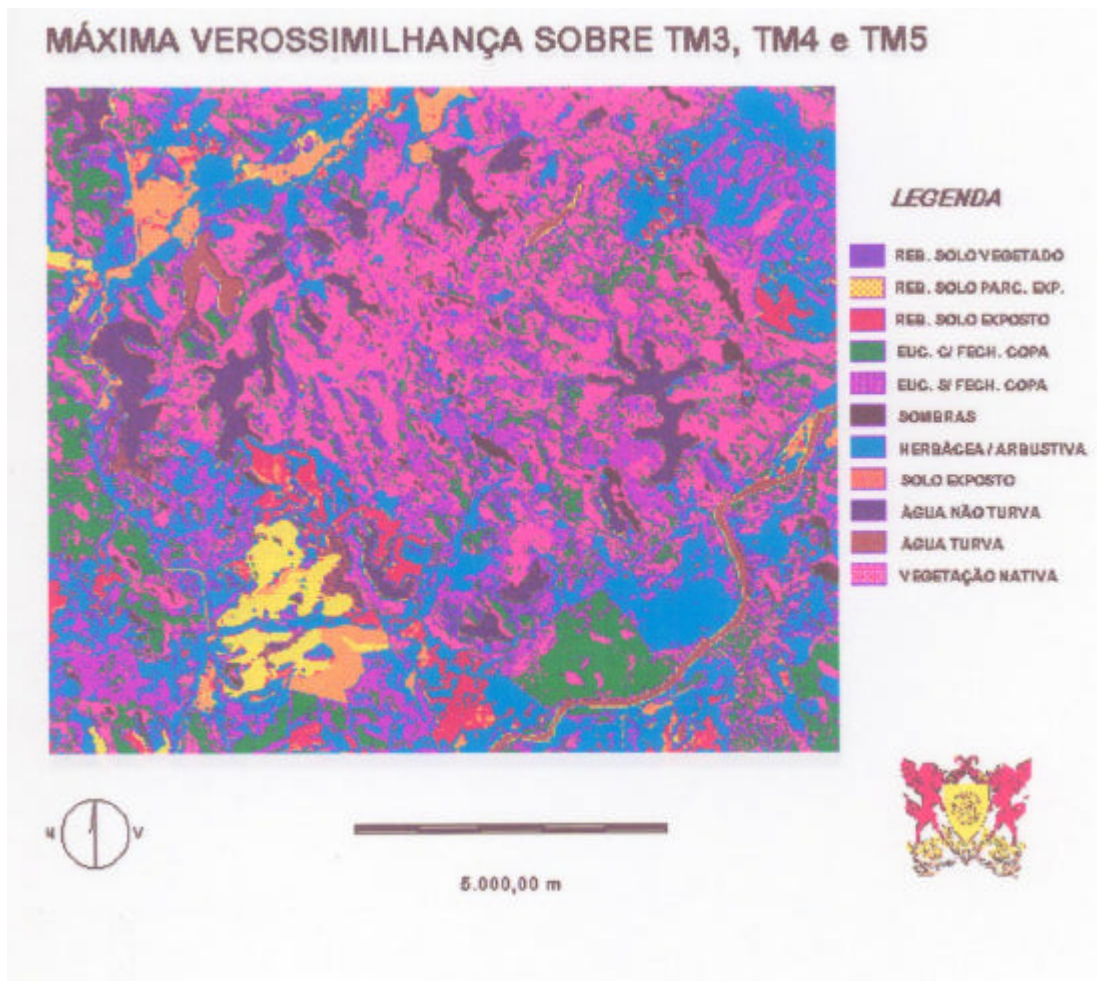
As amostras de treinamento foram coletadas de acordo com o procedimento blocado de amostragem.

Figura 8 - Classificação final resultante da aplicação do algoritmo de máxima verossimilhança sobre as três primeiras componentes principais obtidas a partir das seis bandas reflexivas do sensor TM.



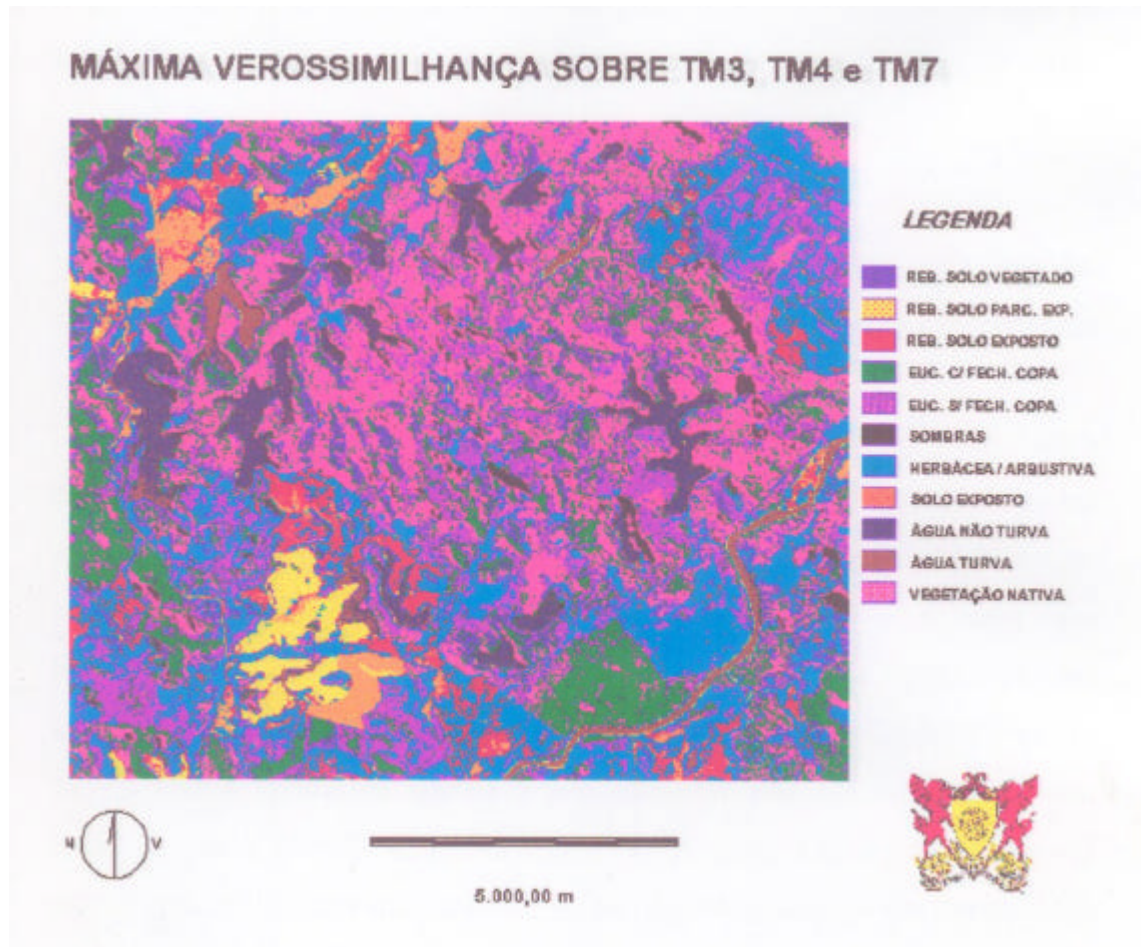
As amostras de treinamento foram coletadas de acordo com o procedimento blocado de amostragem.

Figura 9 - Classificação final resultante da aplicação do algoritmo de máxima verossimilhança sobre as componentes geradas pela transformação *Tasseled Cap*.



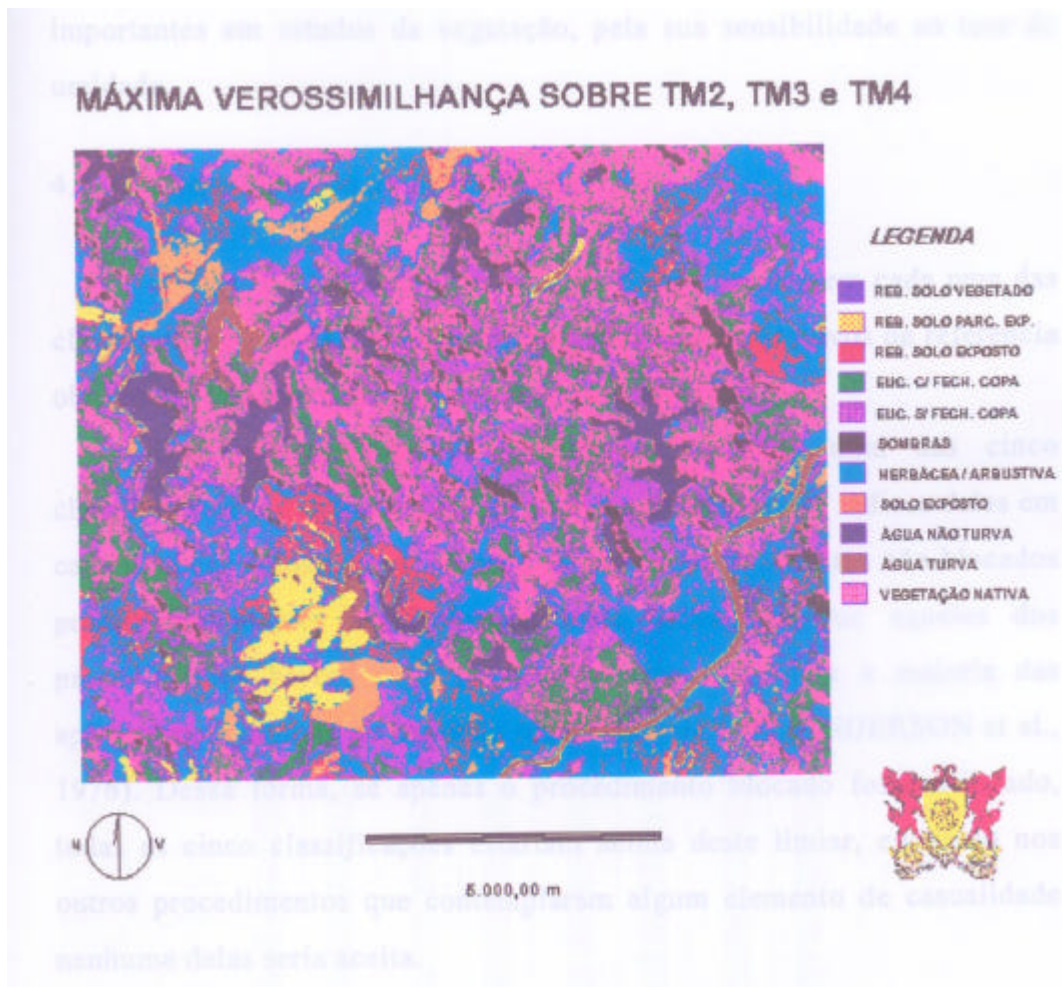
As amostras de treinamento foram coletadas de acordo com o procedimento blocado de amostragem.

Figura 10 - Classificação final resultante da aplicação do algoritmo de máxima verossimilhança sobre as bandas TM3, TM4 e TM5.



As amostras de treinamento foram coletadas de acordo com o procedimento bloqueado de amostragem.

Figura 11 - Classificação final resultante da aplicação do algoritmo de máxima verossimilhança sobre as bandas TM3, TM4 e TM7.



As amostras de treinamento foram coletadas de acordo com o procedimento blocado de amostragem.

Figura 12 - Classificação final resultante da aplicação do algoritmo de máxima verossimilhança sobre as bandas TM2, TM3 e TM4.

Por outro lado, os resultados menos favoráveis foram obtidos a partir do uso da combinação das bandas 2, 3 e 4 do Mapeador Temático, a bordo do satélite Landsat 5. Os baixos índices podem ser justificados pelo fato de esta combinação não contemplar nenhuma banda na região do infravermelho intermediário do espectro das radiações eletromagnéticas, importantes em estudos da vegetação, pela sua sensibilidade ao teor de umidade.

4.2. O Comportamento do Índice Kappa

A Figura 13 ilustra os valores de Kappa obtidos em cada uma das classificações geradas, quando confrontadas com os arquivos de referência obtidos em cada uma das estratégias de amostragem.

Desconsiderando-se o desempenho de cada uma das cinco classificações e atendo-se somente ao comportamento dos índices delas em cada uma das amostragens, percebe-se que os procedimentos não-blocados produziram índices sensivelmente mais baixos do que aqueles dos procedimentos blocados. É importante notar que, para a maioria das aplicações, o mínimo de exatidão requerida é de 85% (ANDERSON et al., 1976). Dessa forma, se apenas o procedimento bloqueado fosse utilizado, todas as cinco classificações estariam acima deste limiar, enquanto nos outros procedimentos que contemplaram algum elemento de casualidade nenhuma delas seria aceita.

Essa marcante diferença, ora descrita, bem mostra o quanto os índices de exatidão podem ser superestimados, dependendo da estratégia de amostragem empregada na geração dos dados de referência. Neste caso específico, esta diferença poderia ser determinante na decisão do usuário em aceitar ou não o produto gerado. É importante notar que, a menos que se mencione a estratégia de amostragem utilizada, o simples índice de exatidão pode não estar representando, com fidedignidade a confiabilidade

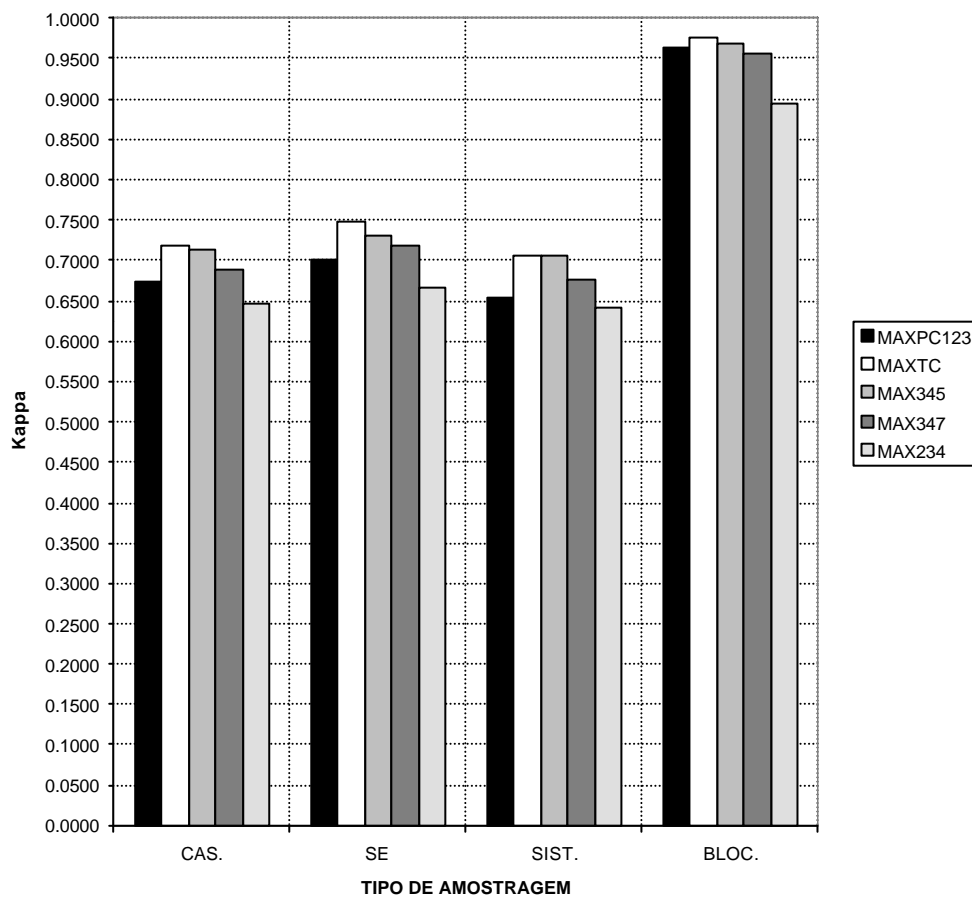


Figura 13 - Comportamento do índice Kappa (K) nas cinco classificações, considerando-se os quatro tipos de amostragem.

esperada do produto, induzindo a ocorrência de erros os mais diversos, principalmente naquelas atividades que, por dependerem de exatidão nas

informações, sofrem efeitos contrários à consecução dos seus objetivos. Além da determinação dos índices Kappa para as classificações de forma geral, ele foi também determinado por categoria informacional. O Quadro 4 contém os valores destes índices para cada uma das classificações, quando comparadas com os arquivos de referência obtidos dentro de cada estratégia de amostragem.

Com a finalidade de propiciar o estabelecimento de comparações entre o procedimento bloqueado de amostragem e os demais, foi calculada a média dos valores de Kappa dos procedimentos não-blocados, para cada uma das categorias espectrais, plotados posteriormente, como pode ser visto pelo gráfico da Figura 14.

O que se pode depreender da Figura 14 é que relativamente poucas categorias informacionais apresentaram índices de exatidão acima do limiar de 85%. Com a repetição da plotagem para os índices obtidos com o uso do arquivo de referência baseado na amostragem bloqueada (Figura 15), essa análise se inverte, passando a apresentar pouquíssimas categorias com índices inferiores a 85%. Mais uma vez é demonstrada a influência da estratégia de amostragem no valor dos índices. Os efeitos de superestimação não se limitam à classificação como um todo, mas, também, se aplicam às categorias consideradas individualmente. Isso pode representar um perigo potencial para aquele que vai utilizar as informações, pois poderia estar considerando valores excessivamente otimistas para quaisquer das categorias constantes da classificação. No caso presente, nenhum dos índices obtidos por meio da amostragem bloqueada, para nenhuma das categorias aqui apresentadas, foi inferior à média dos índices originados dos demais procedimentos, para as mesmas categorias.

Quadro 4 - Médias de Kappa por categoria informacional, considerando-se todas as estratégias de amostragem e todas as classificações

CATEGORIA	MAXPC123				MAXTC				MAX345			
	CAS.	S.E.	SIST.	BLOC.	CAS.	S.E.	SIST.	BLOC.	CAS.	S.E.	SIST.	BLOC.
RSV	0,3883	0,4734	0,4080	0,8640	0,4147	0,4480	0,4650	0,9156	0,3895	0,4049	0,4343	0,8704
RSPE	0,9066	0,9606	0,8589	0,9844	0,9066	0,9605	0,8587	0,9688	0,8754	0,9605	0,8584	0,9572
RSE	0,6718	0,7015	0,7821	0,9730	0,6724	0,6675	0,7471	0,9732	0,7207	0,5575	0,6761	0,9419
EUFECH	0,6956	0,7034	0,6289	0,9817	0,7670	0,7365	0,7056	0,9854	0,7359	0,7346	0,6746	0,9744
EUABER	0,4822	0,4706	0,4393	0,9662	0,6203	0,6420	0,5387	0,9662	0,6820	0,6603	0,6233	0,9614
SOMBRAS	0,8430	0,7446	0,7582	0,9662	0,8027	0,7805	0,7101	0,9578	0,8808	0,7790	0,8057	0,9662
HERB_ARB	0,8381	0,8165	0,7813	0,9434	0,8244	0,8395	0,7793	0,9810	0,8055	0,8137	0,7162	0,9844
SOLO_EXP	0,8904	0,9668	0,7514	0,9829	0,9634	1,0000	0,9374	0,9957	0,9634	0,9336	1,0000	0,9616
AGUALIMP	0,9117	0,9337	0,8793	0,9881	0,9117	0,8675	0,9035	0,9970	0,9705	0,9003	0,9517	0,9911
AGUATURV	0,9462	0,9206	0,8169	1,0000	0,8928	0,8027	0,8900	1,0000	0,7336	0,6868	0,8174	1,0000
NATIVAS	0,7012	0,7514	0,7314	0,9454	0,7569	0,8365	0,7873	0,9590	0,7390	0,8407	0,8285	0,9772

Continua...

Quadro 4, Cont.

CATEGORIA	MAX347				MAX234			
	CAS.	S.E.	SIST.	BLOC.	CAS.	S.E.	SIST.	BLOC.
RSV	0,3778	0,4108	0,4393	0,8189	0,2716	0,2409	0,2647	0,3863
RSPE	0,9066	0,9607	0,8941	0,9844	0,9370	0,9204	0,8935	0,9533
RSE	0,7628	0,7720	0,5586	0,8360	0,6706	0,6648	0,7476	0,9105
EUFECH	0,7392	0,7642	0,7259	0,9598	0,3394	0,4735	0,4040	0,7528
EUABER	0,5937	0,6164	0,5659	0,9710	0,7104	0,6681	0,6143	0,9757
SOMBRAS	0,8814	0,7784	0,8048	0,9662	0,9604	0,8897	1,0000	0,9915
HERB_ARB	0,7081	0,7252	0,6392	0,9790	0,7950	0,7979	0,7511	0,9682
SOLO_EXP	0,8904	0,8677	0,7819	0,9360	0,7090	0,6379	0,6287	0,7482
AGUALIMP	0,9411	0,9666	0,9758	0,9881	1,0000	0,9336	0,9274	0,9941
AGUATURV	0,7863	0,6874	0,8175	1,0000	0,5749	0,6480	0,6366	1,0000
NATIVAS	0,7947	0,8529	0,8026	0,9772	0,7739	0,8264	0,8115	0,9676

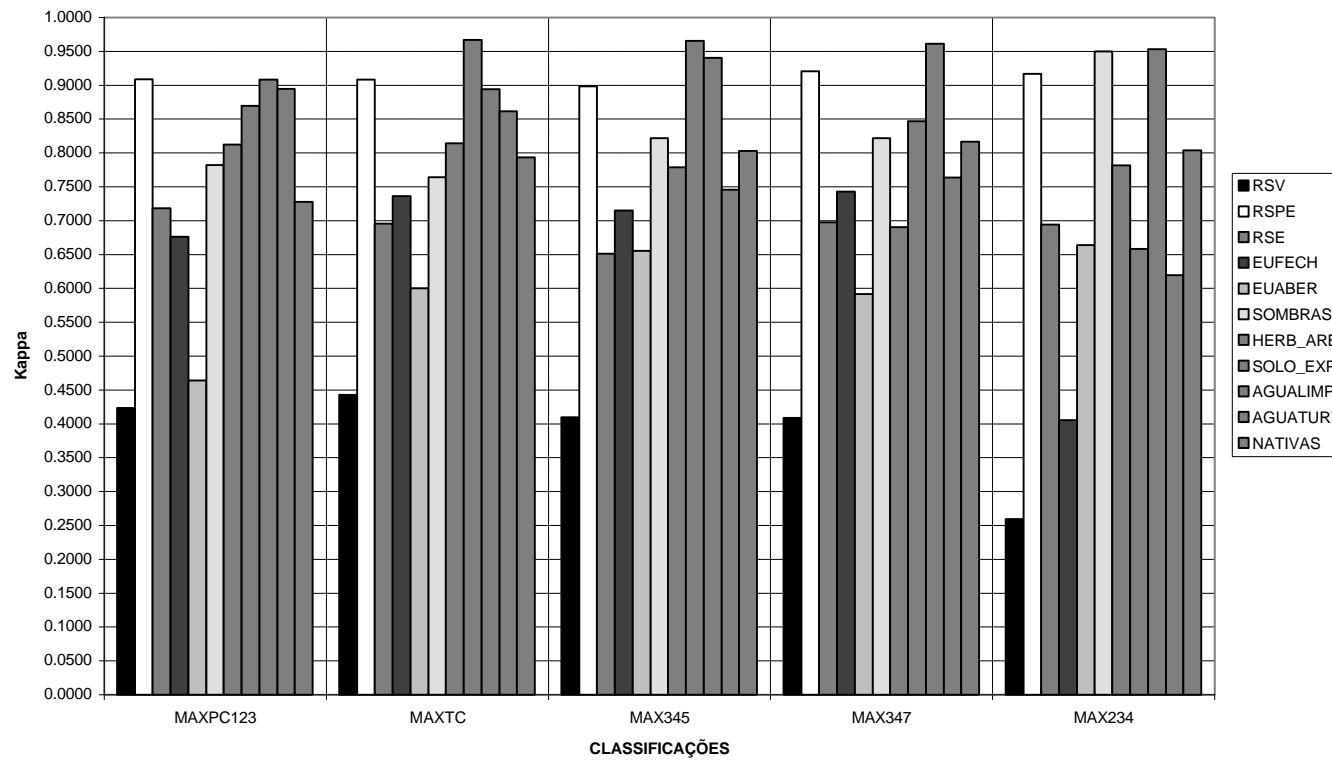


Figura 14 - Plotagem dos valores de Kappa, por categoria informacional, considerando-se todas as amostragens não-blocadas e todas as classificações.

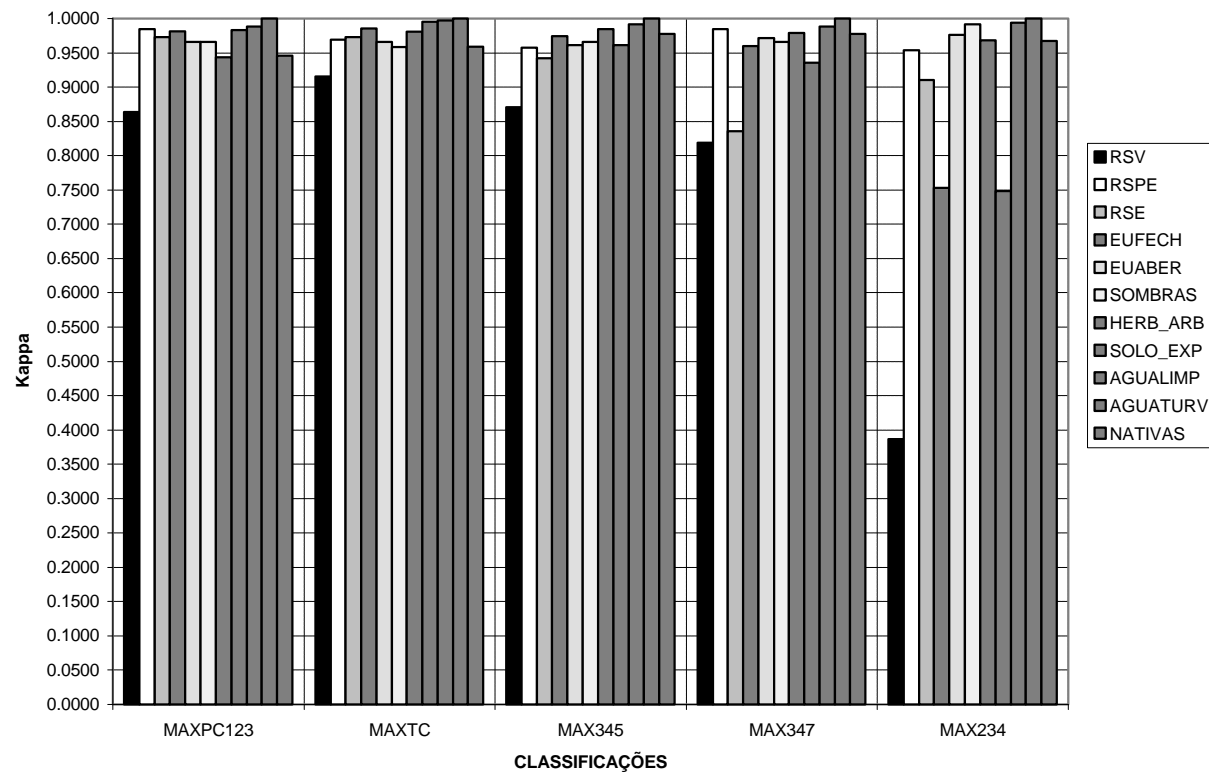


Figura 15 - Plotagem dos valores de Kappa, por categoria informacional, considerando-se somente a amostragem blocada para todas as classificações.

4.3. *Comparações entre os Índices de Exatidão Global, Kappa e Tau*

Um dos objetivos do presente trabalho é comparar três diferentes índices de exatidão para classificações. Dois deles, Global e Kappa, já são comumente usados na área de sensoriamento remoto. O índice Tau foi proposto recentemente à comunidade da área (MA e REDMOND, 1995), tentando corrigir deficiências incorporadas no cálculo do índice Kappa.

Para que as comparações fossem levadas a efeito, foi necessário o cálculo das variâncias para cada índice, em cada classificação, tomando-se todos os quatro arquivos de referência, gerados a partir das diferentes estratégias de amostragem. Os Quadros 5 e 6 apresentam os índices calculados e as variâncias, respectivamente.

Para que se pudesse proporcionar uma visão geral e simultânea do comportamento dos três índices, os seus valores foram plotados (Figura 16). Da análise desta Figura, percebe-se que os três índices apresentam curvas extremamente semelhantes, apenas deslocando-se umas em relação às outras. A posição das curvas pode ser explicada pela forma com que cada índice contempla a concordância casual em sua formulação. Assim, o índice Global (G) apresenta os maiores valores pelo fato de considerar apenas a diagonal principal da matriz de erros, ignorando a ocorrência de tal concordância. Por outro lado, o índice Kappa (K), como já explicado, ao calcular a concordância casual, inclui nos cálculos os elementos da diagonal principal, fazendo com que esta concordância seja superestimada, reduzindo o valor do índice. Na formulação do índice Tau, o valor da concordância casual é estabelecido *a priori*, evitando as falhas em que incorrem G e K. A colocação da curva representando T, entre G e K, indica um bom balanceamento na consideração das concordâncias real e casual, tomando-se em consideração os dois extremos que representam as outras duas curvas, ou seja, de total desconsideração da casualidade e sua superestimação. Tal fato é um indicador de que talvez seja ele o índice que com mais fidedignidade represente a exatidão dos processos de classificação.

Quadro 5 - Índices por classificação e por amostragem

Classificação	Casualizada			Sist. Est.			Sistemática			Blocada		
	G	K	T	G	K	T	G	K	T	G	K	T
MAXPC123	0,7247	0,6727	0,6972	0,7493	0,7020	0,7243	0,7087	0,6545	0,6701	0,9685	0,9642	0,9654
MAXTC	0,7622	0,7174	0,7384	0,7878	0,7469	0,7666	0,7515	0,7058	0,7267	0,9795	0,9767	0,9775
MAX345	0,7587	0,7136	0,7346	0,7747	0,7315	0,7522	0,7524	0,7070	0,7276	0,9719	0,9681	0,9691
MAX347	0,7369	0,6885	0,7106	0,7624	0,7185	0,7386	0,7249	0,6757	0,6794	0,9603	0,9547	0,9563
MAX234	0,7012	0,6455	0,6713	0,7188	0,6648	0,6907	0,6967	0,6402	0,6620	0,9063	0,8930	0,8969

Quadro 6 - Variâncias por índices, classificação e amostragens

Classificação	Casualizada			Sist. Est.		
	G	K	T	G	K	T
MAXPC123	0,00017400	0,00024576	0,00021030	0,00016400	0,00023196	0,00019850
MAXTC	0,00015800	0,00022292	0,00019100	0,00014600	0,00020763	0,00017670
MAX345	0,00015900	0,00022473	0,00019300	0,00015200	0,00021647	0,00018450
MAX347	0,00016900	0,00023681	0,00020440	0,00015800	0,00022217	0,00019140
MAX234	0,00018300	0,00025690	0,00021640	0,00017700	0,00025072	0,00021360

Classificação	Sistemática			Blocada		
	G	K	T	G	K	T
MAXPC123	0,00017600	0,00024881	0,00021410	0,00000962	0,00001243	0,00001186
MAXTC	0,00016000	0,00022424	0,00019360	0,00000633	0,00000804	0,00000766
MAX345	0,00016000	0,00022344	0,00019320	0,00000861	0,00001115	0,00001042
MAX347	0,00017100	0,00023746	0,00020680	0,00001202	0,00001561	0,00001455
MAX234	0,00018100	0,00025469	0,00021910	0,00002678	0,00003492	0,00003240

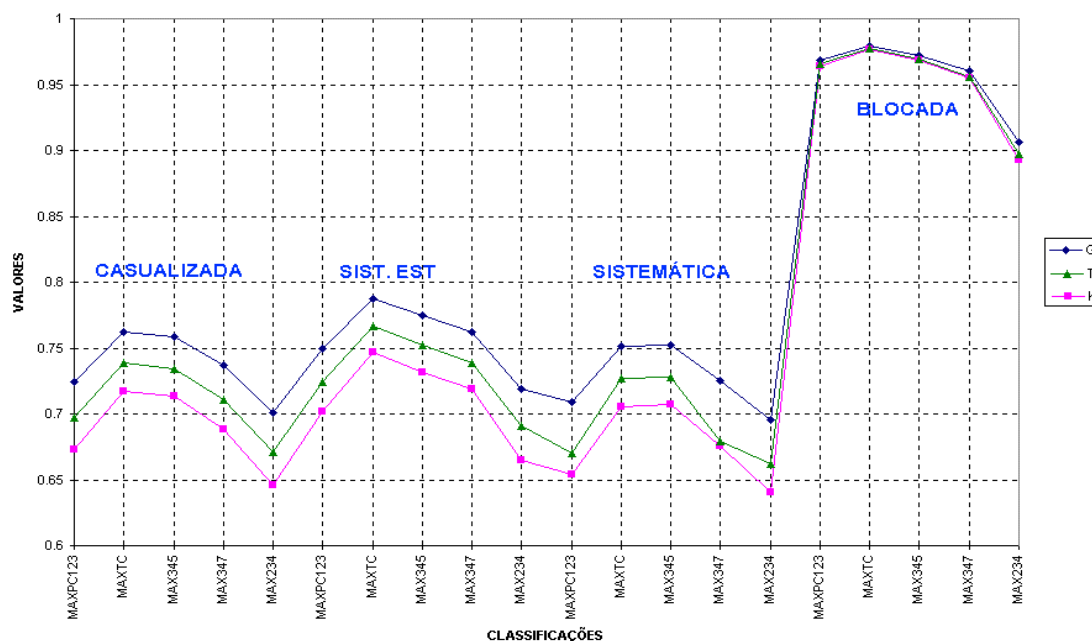


Figura 16 - Comportamento dos índices de exatidão Global (G), Kappa (K) e Tau (T), considerando-se todas as classificações e todas as estratégias de amostragem.

No que se refere às três curvas (Figura 16), na porção relativa ao procedimento bloqueado de amostragem, depende-se da análise visual que as diferenças entre os valores dos três índices são muito pequenas, pouco importando o índice que se eleja para verificar a exatidão da classificação. Quando se consideram os procedimentos não-blocados, as diferenças são bem mais representativas, reportando valores de exatidão bastante distintos, a depender do índice que seja escolhido.

A comparação dos valores dos índices obtidos a partir do uso de estratégias não-blocadas e bloqueada mostra ser o tipo de amostragem o principal fator que concorre para a superestimação dos índices. Considerando-se cada um dos índices, no caso das amostragens não-blocadas, para cada classificação, é notável a diferença entre seus valores absolutos, porém, à exceção da amostragem sistemática, as diferenças entre os índices de mesma

natureza, para quaisquer duas classificações, são bastante semelhantes para qualquer índice que se tome. Esse aspecto evidencia que qualquer dos três índices utilizados deverá ter comportamento semelhante no que se refere ao estabelecimento de diferenças entre as classificações, como será demonstrado a seguir.

4.4. *A Significância das Diferenças*

Com o objetivo de se mensurar a significância das diferenças entre classificações por estratégia de amostragem dos dados de referência e por estratégia de amostragem considerando-se individualmente cada classificação, foi empregado o teste estatístico Z a 99% de probabilidade, de acordo com o que já foi descrito na metodologia.

Os Quadros 7 e 8 apresentam essas comparações, utilizando todas as possíveis combinações de classificações e amostragens. No primeiro deles, foram comparadas entre si todas as classificações geradas dentro de cada procedimento de amostragem. Nota-se, a partir de sua observação, que a sensibilidade do teste às diferenças nas classificações aumenta para as amostragens sistemática e blocada, nesta ordem, o que parece indicar o efeito da interferência do analista no processo amostral, considerando as amostragens casualizadas como as que apresentam maior liberdade para a seleção dos indivíduos, visto que o analista se restringe a informar o número de pontos a serem amostrados; a amostragem sistemática, com o analista interferindo ao selecionar o ponto de partida para o estabelecimento da malha de pontos e, em virtude disso, violando a independência das observações, por causa do igual espaçamento entre elas; e a amostragem blocada, como sendo desprovida de qualquer casualização.

Quadro 7 - Diferenças, por índices, tomando por base as amostragens

AMOSTRAGEM				AMOSTRAGEM			
Casualizada	G	K	T	Sist. Estratificada	G	K	T
MAXPC123 X MAXTC				MAXPC123 X MAXTC			
MAXPC123 X MAX345				MAXPC123 X MAX345			
MAXPC123 X MAX347				MAXPC123 X MAX347			
MAXPC123 X MAX234				MAXPC123 X MAX234			
MAXTC X MAX345				MAXTC X MAX345			
MAXTC X MAX347				MAXTC X MAX347			
MAXTC X MAX234	X	X	X	MAXTC X MAX234	X	X	X
MAX345 X MAX347				MAX345 X MAX347			
MAX345 X MAX234	X	X	X	MAX345 X MAX234	X	X	X
MAX347 X MAX234				MAX347 X MAX234			

AMOSTRAGEM				AMOSTRAGEM			
Sistemática	G	K	T	Blocada	G	K	T
MAXPC123 X MAXTC			X	MAXPC123 X MAXTC	X	X	X
MAXPC123 X MAX345	X	X	X	MAXPC123 X MAX345			
MAXPC123 X MAX347				MAXPC123 X MAX347			
MAXPC123 X MAX234				MAXPC123 X MAX234	X	X	X
MAXTC X MAX345				MAXTC X MAX345			
MAXTC X MAX347			X	MAXTC X MAX347	X	X	X
MAXTC X MAX234	X	X	X	MAXTC X MAX234	X	X	X
MAX345 X MAX347			X	MAX345 X MAX347	X	X	X
MAX345 X MAX234	X	X	X	MAX345 X MAX234	X	X	X
MAX347 X MAX234				MAX347 X MAX234		X	

As células assinaladas por um “X” significam que a diferença é significativa por Z, a 99% de probabilidade.

Quadro 8 - Diferenças, por índices, tomando por base as classificações

CLASSIFICAÇÕES

MAXPC123	G K T			
CASUAL. X SIST.EST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. EST. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
SIST. EST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		

MAXTC	G K T			
CASUAL. X SIST.EST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. EST. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
SIST. EST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		

MAX345	G K T			
CASUAL. X SIST.EST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. EST. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
SIST. EST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		

CLASSIFICAÇÕES

MAX347	G K T			
CASUAL. X SIST.EST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. EST. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
SIST. EST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		

MAX234	G K T			
CASUAL. X SIST.EST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
CASUAL. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. EST. X SIST.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td><td> </td></tr></table>			
SIST. EST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		
SIST. X BLOC.	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>X</td><td>X</td><td>X</td></tr></table>	X	X	X
X	X	X		

As células assinaladas por um “X” significam que a diferença é significativa por Z, a 99% de probabilidade.

Os índices, por sua vez, à exceção das amostragens sistemática e bloqueada, apresentaram a mesma capacidade de identificar diferenças, ou seja, as classificações que foram diferentes entre si ou foram pelos três índices.

No Quadro 8, a análise foi feita fixando-se cada uma das classificações e comparando os índices de exatidão obtidos a partir de cada estratégia de amostragem aplicada sobre os dados de referência.

Da observação do quadro depreende-se que só ocorreu diferença significativa quando foram comparados procedimentos não-blocados com aquele que utilizou a estratégia bloqueada, o que leva a concluir que não há diferença, no que se refere à discriminação entre classificações, entre as amostragens não bloqueadas. Nesta situação, no caso de se adotar a amostragem pixel a pixel, poderia ser recomendada, por questões de custo e de implementação, a amostragem sistemática, desde que as feições de interesse não se repitam de forma cíclica na paisagem.

No que se refere, neste quadro, à comparação entre os três índices, todos se comportaram de forma idêntica, ou seja, apresentaram exatamente a mesma sensibilidade às comparações entre os diferentes tipos de amostragem.

5. RESUMO E CONCLUSÕES

O presente trabalho teve como objetivo geral testar os efeitos de diferentes estratégias de amostragem, aplicadas aos dados de referência, sobre três índices utilizados como mensuradores de exatidão (Exatidão Global, Kappa e Tau) na classificação de imagens orbitais. Como objetivos específicos testou-se o comportamento do índice Tau, comparado-o aos dois índices tradicionalmente utilizados em sensoriamento remoto (Global e Kappa), e compararam-se diferentes combinações de bandas e transformações especiais no sentido de verificar qual delas seria mais adequada para a utilização em classificações de regiões densamente vegetadas.

Foram utilizadas as seis bandas reflexivas do sensor *Thematic Mapper*, a bordo do satélite Landsat 5. A partir dos dados contidos nestas bandas ou nestes canais, foram utilizadas as combinações TM3, TM4 e TM5; TM3, TM4 e TM7; TM2, TM3 e TM4, além de duas transformações especiais (as três primeiras componentes principais e a transformação de *Tasseled Cap*).

Essas combinações e transformações foram submetidas a um algoritmo de classificação por máxima verossimilhança com base em um mesmo conjunto de amostras de treinamento que foram selecionadas por

um procedimento bloqueado de amostragem, tendo como material de referência uma cobertura fotogramétrica da área, obtida aproximadamente na mesma época da obtenção da imagem orbital, além de mapas e dados de inventário florestal.

Para a verificação de exatidão, os dados de referência foram amostrados de acordo com as seguintes estratégias, utilizando-se, aproximadamente, 1.200 pontos em cada uma delas: bloqueada, casualizada, sistemática estratificada e sistemática, excetuando-se aqueles pixels já contemplados quando da obtenção das amostras de treinamento.

Para cada classificação e para cada tipo de amostragem foram obtidos os três índices já mencionados, a serem utilizados nas comparações. Para cada índice calculado, foram determinadas as suas variâncias, a fim de que fossem possibilitadas as comparações feitas pelo teste de Z, a 99% de significância.

As seguintes conclusões puderam ser reunidas a partir dos resultados alcançados neste trabalho:

1) Pelas comparações de cada classificação com todas as combinações possíveis de amostragens, conclui-se que para nenhuma delas houve diferença significativa, pelos três índices, quando foram usadas amostragens não-bloqueadas, o que sugere ser a amostragem bloqueada a principal fonte de superestimação dos índices.

2) O comportamento dos índices G, K e T em todas as classificações e em todas as amostragens é bastante semelhante no que se refere à variação dos mesmos de uma classificação para outra.

3) O comportamento de G, sempre acima de K e T, comprova o efeito de ele não tomar em consideração os erros de omissão e comissão, superestimando a avaliação da exatidão.

4) Em concordância com a literatura, a tendência de K de superestimar a concordância casual é comprovada por seus valores sempre abaixo de G e T.

5) O índice T, ao se situar entre G e K, parece ser o ideal, mostrando tornar possível a correta consideração das concordâncias real e casual, para a situação na qual é assumida a equiprobabilidade de ocorrência das categorias informacionais.

6) A classificação baseada nas bandas TM2, TM3 e TM4 foi a de pior desempenho em todas as amostragens testadas, à exceção da sistematizada.

7) Dentro de uma mesma estratégia de amostragem, à exceção das amostragens sistemática e blocada, os três índices discriminaram as diferenças entre classificações de maneira idêntica.

8) À medida que aumenta o grau de interferência do analista nos procedimentos de amostragem, aumenta a sensibilidade dos índices no que se refere ao estabelecimento de diferenças entre classificações independentes.

9) Não houve diferença significativa entre as estratégias não-blocadas de amostragem, considerados os três índices utilizados.

10) De acordo com a conclusão anterior, parece ser a amostragem sistemática a mais vantajosa, do ponto de vista da facilidade de operacionalização, quando se objetiva melhor quantificar a exatidão dos processos de classificação, desde que as feições de interesse não apresentem caráter de periodicidade na paisagem.

6. RECOMENDAÇÕES

A fim de que este trabalho se transforme em uma linha de pesquisa, explorando outras facetas do problema, são sugeridas algumas idéias para a sua continuidade.

- Repetir o mesmo procedimento com a obtenção das amostras de treinamento nas quatro estratégias de amostragem que foram utilizadas sobre os dados de referência.
- Refazer o estudo sem assumir a equiprobabilidade de ocorrência entre as categorias informacionais.
- Estabelecer procedimentos que permitam a inferência entre o comportamento das estratégias de amostragem e o grau de autocorrelação espacial presentes na imagem.
- Estudar a possibilidade de redução no número de indivíduos amostrados e, ou, outras estratégias de amostragem, a fim de possibilitar a redução de custos na implementação dos processos amostrais não-blocados.

REFERÊNCIAS BIBLIOGRÁFICAS

- AFIFI, A.A., CLARK, V. Computer-aided multivariate analysis. New York: Van Nostrand Reinhold Company, 1990. 505p.
- ANDERSON, J.R., HARDY, E.E., ROACH, J.T., WITMER, R.E. **A land use and land cover classification system for use with remote sensor data**. Washington: U.S.G.S, 1976. 28p. (USGS Professional Paper, 964).
- BOLSTAD, P.V., LILLESAND, T.M. Rapid maximum likelihood classification. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.57, n.1, p.67-74, Jan. 1991.
- CAMPBELL, J.B. **Introduction to remote sensing**. New York, The Guilford Press, 1987. 551p.
- CHIAO, K. A study on age discrimination of *Criptomeria* with greenness indicators. In: ACSM-ASPRS ANNUAL CONVENTION, 1991, Baltimore. **Technical Papers**. Bethesda, ACSM-ASPRS, 1992. v.3, p.42-51.
- CONGALTON, R.G. **A comparison of five sampling schemes used in assessing the accuracy of land cover/land use maps derived from remotely sensed data**. Blacksburg: Virginia Polytechnic Institute and State University, 1984. 146p. (Tese PhD.).
- 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, v.54, p.593-600, May, 1988.

CONGALTON, R.G., BIGING, G.S. A pilot study evaluating ground reference data collection efforts for use in forest inventory. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.58, n.12, p.1669-1671, Dez., 1992.

CRIST, E.P., CICONE, R.C. A physically-based transformation of Thematic Mapper data - The TM Tasseled Cap. **IEEE Transactions on Geoscience and Remote Sensing**. v.GE-22, n.3, p.256-263, maio, 1984.

DAVIS, J.C. **Statistics and data analysis in geology**. New York. John Wiley & Sons, 1986. 646p.

DEFEO, N., SAVAGE, K., COCHAND, J. Using perspective viewing for better visualization of forest damage. In: ACSM-ASPRS FALL CONVENTION, 1990, Anaheim. **Technical Papers**. Bethesda, ACSM-ASPRS, 1990. p.B-44-B-48.

DICKS, S.E., LO, T.H.C., Evaluation of thematic map accuracy in a land-use and land-cover mapping program. **Photogrammetric Engineering and Remote Sensing**. Bethesda, v.56, n.9, p.1247 - 1252. Set., 1990.

DOZIER, J. STRAHLER, A.H. Ground investigations in support of remote sensing. In: American Society of Photogrammetry and Remote Sensing, (2^a Ed.). **Manual of remote sensing**. Falls Church: ASPRS, 1983. v.1., p.959-986.

EBERT, J.A., LYONS, T.R. Archaeology, anthropology, and cultural resources management. In: AMERICAN SOCIETY OF PHOTOGRAMMETRY AND REMOTE SENSING, **Manual of remote sensing**. 2.ed.Falls Church: ASPRS, 1983. v.2, p.1233-1304.

EPIPHANIO, J.C.N., FORMAGGIO, A.R. Relações entre PVI e diferença normalizada e índice de área foliar, cobertura de solo e densidade de clorofila de trigo e de feijão. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 1990, Manaus. **Anais...** São José dos Campos: INPE, 1990. v.3, p.797-805.

- FOODY, G.M. On the compensation for chance agreement in image classification accuracy assessment. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.58, n.10, p.1459-1460, Oct., 1990.
- FOODY, G.M., CAMPBELL, N.A., TRODD, N.M., WOOD, T.F. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.58, n.9, p.1335-1341. Sept., 1992.
- FUNG, T., LEDREW, E. Application of principal components analysis to change detection. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.53, n.12, p.1649-1658, Dec., 1987.
- GATES, D.M. **Physical and physiological properties of plants. Remote sensing with special reference to agriculture and forestry.** Washington, D.C.: National Academy of Sciences, 1970. 424p.
- GOLFARI, L. **Zoneamento ecológico do Estado de Minas Gerais para o reflorestamento.** Belo Horizonte: Centro de Pesquisas do IBDF na Região dos Cerrados, 1975. 65p.
- GONG, P., HOWART, P.J. An assessment of some factors influencing multispectral land-cover classification. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.56, n.5, p.597-603, maio, 1990.
- HARALICK, R.M., FU, K. Pattern recognition and classification. In: AMERICAN SOCIETY OF PHOTOGRAMMETRY AND REMOTE SENSING. **Manual of remote sensing.** 2.ed. Falls Church: 1983. v.1, p.793-805.
- JENSEN, J.R. **Introductory digital image processing.** New Jersey: Prentice-Hall, 1986. 379p.
- KELLY, B.W. Sampling and statistical problems. In: NATIONAL ACADEMY OF SCIENCES. **Remote sensing with special reference to agriculture and forestry**, Washington, DC: 1970. p.324-353.
- KNIPLING, E.B. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation, 1970. **Remote Sensing of Environment**, New York, v.1, p.155-159, 1970.

- LAUVER, C.L., WHISTLER, J.L. A hierarchical classification of Landsat TM imagery to identify natural grassland areas and rare species habitat. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.59, n.5, p.627-634, May, 1993.
- LILLESAND, T.M., KIEFER, R.W. **Remote sensing and image interpretation**. New York: John Wiley & Sons, 1987. 721p.
- LIOU, K.N. **An introduction to atmospheric radiation**. New York. Academic Press, 1980. 392p.
- LOETSCH, F., HALLER, K.E. **Forest inventory**. München. BLV Verlagsgesellschaft, 1964. v.1.
- LOUGHLIN, W.P., Principal component analysis for alteration mapping. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.57, n.9, p.1163-1169, Sept., 1991.
- LUND, H.G. Point sampling - The role in in-place inventories. In: NATIONAL WORKSHOP SOCIETY OF AMERICAN FORESTERS, Orono, 1981. **Proceedings...** Orono: Society of American Foresters, 1982. p.371-374.
- MASSASATI, A.S. Computer modeling of urban area detection and mapping using remote sensing. In: ACSM-ASPRS FALL CONVENTION, 1990, Anaheim. **Technical Papers**. Bethesda, 1990. p.B-87 -B-96. 1990.
- MA, Z., REDMOND, R.L. Tau coefficients for accuracy assessment of classification of remote sensing data. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.61, n.4, p.453-439, Apr., 1995.
- RIBEIRO, C.A.A.S., RIBEIRO, J.C., BRITES, R.S. GAVIÃO - An alternative system for image processing. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 6, 1990, Manaus. **Anais...** São José dos Campos: INPE, 1990. v. 1, p.163-166.
- ROSENFELD, G.H., FITZPATRICK-LINS, K.A. A coefficient of agreement as a measure of thematic classification accuracy. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.52, n.2, 223-227, Feb. 1986.

- SINGH, A., HARRISON, A. Standardized principal components. **International Journal of Remote Sensing**, Hampshire, v.6, n.6, 884-896. 1985.
- STEHMAN, S.V. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.58, n.9, p.1343-1350, 1992.
- SWAIN, P.H., DAVIS, S.M. **Remote sensing: the quantitative approach**. New York: McGraw-Hill, 1978. 396p.
- WARREN, S.D., JOHNSON, M.O., GORAN, W.D., DIERSING, V.E. An automated objective procedure for selecting representative field sample sites. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v.56, n.3, p.333-335, 1990.
- WILLIAMS JR., R.S. Geological applications. In: AMERICAN SOCIETY OF PHOTOGRAMMETRY AND REMOTE SENSING. **Manual of remote sensing**. 2.ed. Falls Church: 1983. v.2, p.1667-1953.
- MASSASATI, A.S. Computer modeling of urban area detection and mapping using remote sensing. In: ACSM-ASPRS FALL CONVENTION, 1990, Anaheim. **Technical Papers**. Bethesda, 1990. p.B-87-B-96.
- YOSHIMURA, M., OSHIMA, T., Monitoring of the vegetation condition by Tasseled Cap transformation - investigation of U.S.A. drought disaster in 1988. In: ACSM-ASPRS ANNUAL CONVENTION, 1991, Baltimore. **Technical Papers**. Bethesda, 1991. v.3, p.506-515.

APÊNDICE

APÊNDICE

Neste apêndice encontram-se as matrizes de erro utilizadas no desenvolvimento deste trabalho, ordenadas por classificação. Dentro de cada classificação estão as matrizes geradas a partir da comparação com os arquivos de referência produzidos por cada estratégia de amostragem. Estas matrizes estão reproduzidas nos moldes de saída do programa ERRMAT do sistema IDRISI, versão 1.0, para Windows.

Os termos ErrorO e ErrorC significam erros de omissão e erros de comissão, respectivamente.

Os números que identificam as classes têm seu significado de acordo com o quadro abaixo:

NÚMERO	CATEGORIA INFORMACIONAL	SIGLA
1	Rebrota sobre Solo Vegetado	RSV
2	Rebrota sobre Solo Parcialmente Exposto	RSPE
3	Rebrota sobre Solo Exposto	RSE
4	Eucalipto com Fechamento de Copa	EUFECH
5	Eucalipto sem Fechamento de Copa	EUABER
6	Sombras	SOMBRAS
7	Herbácea-Arbustiva	HERB-ARB
8	Solo Exposto	SOLO-EXP
9	Água Não-Turva	AGUALIMP
10	Água Turva	AGUATURV
11	Nativas	NATIVAS

BLOCADA (columns : truth) against MAXPC123 (rows : mapped)

	1	2	3	4	5	
1	141	1	0	3	1	0.0537
2	0	277	0	0	0	0
3	0	1	237	0	0	0.1255
4	8	0	0	298	0	0.0479
5	3	0	0	0	216	0.027
6	0	0	0	0	0	0
7	6	2	5	0	6	0.0292
8	0	0	1	0	0	0.004
9	0	0	0	0	0	0
10	0	0	0	0	0	0.0388
11	4	0	0	2	0	0.026
Total	162	281	243	303	223	
ErrorO	0.1296	0.0142	0.0247	0.0165	0.0314	

	6	7	8	9	10	
1	0	0	0	0	0	0.0537
2	0	0	0	0	0	0
3	0	32	1	0	0	0.1255
4	0	0	0	0	0	0.0479
5	0	1	0	0	0	0.027
6	119	0	0	0	0	0
7	0	732	3	0	0	0.0292
8	0	0	250	0	0	0.004
9	0	0	0	378	0	0
10	4	0	0	4	198	0.0388
11	0	0	0	0	0	0.026
Total	123	765	254	382	198	
ErrorO	0.0325	0.0431	0.0157	0.0105	0	

	11	Total	ErrorC
1	3	149	0.0537
2	0	277	0
3	0	271	0.1255
4	7	313	0.0479
5	2	222	0.027
6	0	119	0
7	0	754	0.0292
8	0	251	0.004
9	0	378	0
10	0	206	0.0388
11	225	231	0.026
Total	237	3171	
ErrorO	0.0506		0.0315

Error Matrix Analysis of
CASUALIZADA (columns : truth) against MAXPC123 (rows : mapped)

	1	2	3	4	5	
1	71	0	0	6	22	0.4083
2	0	30	0	0	0	0.0323
3	0	1	15	0	0	0.5714
4	17	0	0	110	5	0.3168
5	33	0	0	2	103	0.3224
6	0	0	0	0	0	0.0435
7	5	1	7	0	18	0.1377
8	0	1	0	0	0	0.0385
9	0	0	0	0	0	0.0303
10	1	0	0	0	0	0.25
11	30	0	0	31	39	0.348
Total	157	33	22	149	187	
ErrorO	0.5478	0.0909	0.3182	0.2617	0.4492	

	6	7	8	9	10	
1	0	5	0	0	0	0.4083
2	0	0	1	0	0	0.0323
3	0	18	1	0	0	0.5714
4	0	0	0	0	0	0.3168
5	0	7	0	0	0	0.3224
6	22	0	0	0	0	0.0435
7	0	213	1	0	0	0.1377
8	0	0	25	0	0	0.0385
9	0	0	0	32	1	0.0303
10	2	0	0	3	18	0.25
11	2	1	0	0	0	0.348
Total	26	244	28	35	19	
ErrorO	0.1538	0.127	0.1071	0.0857	0.0526	

	11	Total	ErrorC
1	16	120	0.4083
2	0	31	0.0323
3	0	35	0.5714
4	29	161	0.3168
5	7	152	0.3224
6	1	23	0.0435
7	2	247	0.1377
8	0	26	0.0385
9	0	33	0.0303
10	0	24	0.25
11	193	296	0.348
Total	248	1148	
ErrorO	0.2218		0.2753

Error Matrix Analysis of
SIST.EST. (columns : truth) against MAXPC123 (rows : mapped)

	1	2	3	4	5	
1	70	0	0	9	20	0.3913
2	0	25	1	0	0	0.0741
3	1	0	20	0	0	0.5918
4	13	0	0	114	11	0.2919
5	24	0	0	3	89	0.3047
6	1	0	0	0	0	0.125
7	8	0	7	0	20	0.1268
8	0	1	0	0	0	0.0625
9	0	0	0	0	0	0.0645
10	2	0	0	0	0	0.3333
11	14	0	0	27	28	0.2669
Total	133	26	28	153	168	
ErrorO	0.4737	0.0385	0.2857	0.2549	0.4702	

	6	7	8	9	10	
1	0	5	0	0	0	0.3913
2	0	1	0	0	0	0.0741
3	0	27	1	0	0	0.5918
4	0	0	0	0	0	0.2919
5	0	4	0	0	0	0.3047
6	21	0	0	0	0	0.125
7	0	241	0	0	0	0.1268
8	0	1	30	0	0	0.0625
9	0	0	0	29	2	0.0645
10	6	0	0	2	24	0.3333
11	1	1	0	0	0	0.2669
Total	28	280	31	31	26	
ErrorO	0.25	0.1393	0.0323	0.0645	0.0769	

	11	Total	ErrorC
1	11	115	0.3913
2	0	27	0.0741
3	0	49	0.5918
4	23	161	0.2919
5	8	128	0.3047
6	2	24	0.125
7	0	276	0.1268
8	0	32	0.0625
9	0	31	0.0645
10	2	36	0.3333
11	195	266	0.2669
Total	241	1145	
ErrorO	0.1909		0.2507

Error Matrix Analysis of
SISTEMATICA (columns : truth) against MAXPC123 (rows : mapped)

	1	2	3	4	5	
1	66	0	0	6	21	0.4107
2	0	25	0	0	0	0.0385
3	3	0	23	0	0	0.6102
4	16	0	0	101	8	0.3442
5	30	0	0	4	90	0.3478
6	0	0	0	0	0	0.1111
7	8	1	6	0	17	0.1417
8	0	3	0	0	0	0.1379
9	0	0	0	0	0	0.1163
10	0	0	0	0	0	0.2069
11	19	0	0	38	42	0.3377
Total	142	29	29	149	178	
ErrorO	0.5352	0.1379	0.2069	0.3221	0.4944	

	6	7	8	9	10	
1	0	8	0	0	0	0.4107
2	0	0	1	0	0	0.0385
3	0	29	4	0	0	0.6102
4	1	0	0	0	0	0.3442
5	0	6	0	0	0	0.3478
6	16	0	0	0	0	0.1111
7	0	218	3	0	0	0.1417
8	0	1	25	0	0	0.1379
9	0	0	0	38	5	0.1163
10	1	0	0	5	23	0.2069
11	3	1	0	0	0	0.3377
Total	21	263	33	43	28	
ErrorO	0.2381	0.1711	0.2424	0.1163	0.1786	

	11	Total	ErrorC
1	11	112	0.4107
2	0	26	0.0385
3	0	59	0.6102
4	28	154	0.3442
5	8	138	0.3478
6	2	18	0.1111
7	1	254	0.1417
8	0	29	0.1379
9	0	43	0.1163
10	0	29	0.2069
11	202	305	0.3377
Total	252	1167	
ErrorO	0.1984		0.2913

Error Matrix Analysis of
BLOCADA (columns : truth) against MAXTC (rows : mapped)

	1	2	3	4	5	
1	149	0	0	4	1	0.0387
2	0	273	0	0	0	0
3	0	4	237	0	0	0.0558
4	4	1	0	299	0	0.0355
5	5	0	0	0	216	0.0357
6	0	0	0	0	0	0
7	2	2	5	0	6	0.0208
8	0	1	1	0	0	0.0078
9	0	0	0	0	0	0.0026
10	0	0	0	0	0	0.0198
11	2	0	0	0	0	0.013
Total	162	281	243	303	223	
ErrorO	0.0802	0.0285	0.0247	0.0132	0.0314	

	6	7	8	9	10	
1	0	0	0	0	0	0.0387
2	0	0	0	0	0	0
3	0	10	0	0	0	0.0558
4	0	0	0	0	0	0.0355
5	0	1	0	0	0	0.0357
6	118	0	0	0	0	0
7	0	754	1	0	0	0.0208
8	0	0	253	0	0	0.0078
9	1	0	0	381	0	0.0026
10	3	0	0	1	198	0.0198
11	1	0	0	0	0	0.013
Total	123	765	254	382	198	
ErrorO	0.0407	0.0144	0.0039	0.0026	0	

	11	Total	ErrorC
1	1	155	0.0387
2	0	273	0
3	0	251	0.0558
4	6	310	0.0355
5	2	224	0.0357
6	0	118	0
7	0	770	0.0208
8	0	255	0.0078
9	0	382	0.0026
10	0	202	0.0198
11	228	231	0.013
Total	237	3171	
ErrorO	0.0380		0.0205

Error Matrix Analysis of
CASUALIZADA (columns : truth) against MAXTC (rows : mapped)

	1	2	3	4	5	
1	74	0	1	5	15	0.3333
2	0	30	0	0	0	0.0323
3	0	1	15	0	0	0.5455
4	12	1	0	119	0	0.2372
5	35	0	0	3	127	0.2865
6	1	0	0	0	0	0.2759
7	3	1	6	0	15	0.1139
8	0	0	0	0	0	0.0357
9	1	0	0	0	0	0.0303
10	0	0	0	0	0	0.1905
11	31	0	0	22	30	0.3024
Total	157	33	22	149	187	
ErrorO	0.5287	0.0909	0.3182	0.2013	0.3209	

	6	7	8	9	10	
1	0	5	0	0	1	0.3333
2	0	0	1	0	0	0.0323
3	0	17	0	0	0	0.5455
4	1	0	0	0	0	0.2372
5	0	9	0	0	0	0.2865
6	21	0	0	0	1	0.2759
7	0	210	0	0	0	0.1139
8	0	1	27	0	0	0.0357
9	0	0	0	32	0	0.0303
10	1	0	0	3	17	0.1905
11	3	2	0	0	0	0.3024
Total	26	244	28	35	19	
ErrorO	0.1923	0.1393	0.0357	0.0857	0.1053	

	11	Total	ErrorC
1	10	111	0.3333
2	0	31	0.0323
3	0	33	0.5455
4	23	156	0.2372
5	4	178	0.2865
6	6	29	0.2759
7	2	237	0.1139
8	0	28	0.0357
9	0	33	0.0303
10	0	21	0.1905
11	203	291	0.3024
Total	248	1148	
ErrorO	0.1815		0.2378

Error Matrix Analysis of
SIST.EST. (columns : truth) against MAXTC (rows : mapped)

	1	2	3	4	5	
1	66	0	0	9	10	0.34
2	0	25	2	0	0	0.1379
3	1	0	19	0	0	0.5
4	11	0	0	118	5	0.2185
5	26	0	0	5	116	0.2516
6	1	0	0	0	0	0.1852
7	11	0	7	0	15	0.1183
8	0	1	0	0	0	0.0882
9	0	0	0	0	0	0.1
10	1	0	0	0	0	0.2759
11	16	0	0	21	22	0.2271
Total	133	26	28	153	168	
ErrorO	0.5038	0.0385	0.3214	0.2288	0.3095	

	6	7	8	9	10	
1	0	5	0	0	2	0.34
2	0	1	0	0	1	0.1379
3	0	18	0	0	0	0.5
4	0	0	0	0	0	0.2185
5	0	7	0	0	0	0.2516
6	22	0	0	0	0	0.1852
7	0	246	0	0	0	0.1183
8	0	2	31	0	0	0.0882
9	1	0	0	27	2	0.1
10	3	0	0	4	21	0.2759
11	2	1	0	0	0	0.2271
Total	28	280	31	31	26	
ErrorO	0.2143	0.1214	0	0.129	0.1923	

	11	Total	ErrorC
1	8	100	0.34
2	0	29	0.1379
3	0	38	0.5
4	17	151	0.2185
5	1	155	0.2516
6	4	27	0.1852
7	0	279	0.1183
8	0	34	0.0882
9	0	30	0.1
10	0	29	0.2759
11	211	273	0.2271
Total	241	1145	
ErrorO	0.1245		0.2122

Error Matrix Analysis of
SISTEMATICA (columns : truth) against MAXTC (rows : mapped)

	1	2	3	4	5	
1	73	1	0	4	15	0.3178
2	0	25	0	2	0	0.1071
3	1	0	22	0	0	0.5849
4	11	0	0	111	10	0.2885
5	29	0	0	8	107	0.3228
6	0	0	0	0	0	0.1176
7	6	0	4	0	14	0.1033
8	0	3	2	0	0	0.1842
9	0	0	0	0	0	0.0714
10	0	0	0	0	0	0.1667
11	22	0	1	24	32	0.2838
Total	142	29	29	149	178	
ErrorO	0.4859	0.1379	0.2414	0.255	0.3989	

	6	7	8	9	10	
1	0	5	0	0	0	0.3178
2	0	0	1	0	0	0.1071
3	0	29	1	0	0	0.5849
4	1	0	0	0	0	0.2885
5	0	9	0	0	0	0.3228
6	15	0	0	0	0	0.1176
7	0	217	0	0	0	0.1033
8	0	2	31	0	0	0.1842
9	0	0	0	39	3	0.0714
10	1	0	0	4	25	0.1667
11	4	1	0	0	0	0.2838
Total	21	263	33	43	28	
ErrorO	0.2857	0.1749	0.0606	0.093	0.1071	

	11	Total	ErrorC
1	9	107	0.3178
2	0	28	0.1071
3	0	53	0.5849
4	23	156	0.2885
5	5	158	0.3228
6	2	17	0.1176
7	1	242	0.1033
8	0	38	0.1842
9	0	42	0.0714
10	0	30	0.1667
11	212	296	0.2838
Total	252	1167	
ErrorO	0.1587		0.2485

Error Matrix Analysis of
BLOCADA (columns : truth) against MAX345 (rows : mapped)

	1	2	3	4	5	
1	142	0	0	5	2	0.0533
2	0	270	0	0	0	0.011
3	0	9	230	0	0	0.0763
4	7	1	0	296	0	0.0327
5	2	0	0	2	215	0.0487
6	0	0	0	0	0	0.0165
7	3	0	11	0	5	0.0245
8	0	1	2	0	0	0.0121
9	0	0	0	0	0	0.0104
10	0	0	0	0	0	0.005
11	8	0	0	0	1	0.0373
Total	162	281	243	303	223	
ErrorO	0.1235	0.0391	0.0535	0.0231	0.0359	

	6	7	8	9	10	
1	0	0	0	0	0	0.0533
2	0	0	3	0	0	0.011
3	0	4	6	0	0	0.0763
4	0	0	0	0	0	0.0327
5	0	5	0	0	0	0.0487
6	119	0	0	2	0	0.0165
7	0	756	0	0	0	0.0245
8	0	0	245	0	0	0.0121
9	4	0	0	379	0	0.0104
10	0	0	0	1	198	0.005
11	0	0	0	0	0	0.0373
Total	123	765	254	382	198	
ErrorO	0.0325	0.0118	0.0354	0.0079	0	

	11	Total	ErrorC
1	1	150	0.0533
2	0	273	0.011
3	0	249	0.0763
4	2	306	0.0327
5	2	226	0.0487
6	0	121	0.0165
7	0	775	0.0245
8	0	248	0.0121
9	0	383	0.0104
10	0	199	0.005
11	232	241	0.0373
Total	237	3171	
ErrorO	0.0211		0.0281

Error Matrix Analysis of
CASUALIZADA (columns : truth) against MAX345 (rows : mapped)

	1	2	3	4	5	
1	70	0	0	4	13	0.3396
2	0	29	1	0	0	0.0645
3	0	2	16	0	0	0.4074
4	15	1	0	115	3	0.2628
5	31	0	0	8	138	0.3168
6	3	0	0	0	0	0.3784
7	5	1	5	0	10	0.1004
8	0	0	0	0	0	0.069
9	1	0	0	0	0	0.1053
10	0	0	0	0	0	0
11	32	0	0	22	23	0.2867
Total	157	33	22	149	187	
ErrorO	0.5541	0.1212	0.2727	0.2282	0.262	

	6	7	8	9	10	
1	0	8	0	0	0	0.3396
2	0	0	1	0	0	0.0645
3	0	9	0	0	0	0.4074
4	1	0	0	0	0	0.2628
5	0	17	0	0	0	0.3168
6	23	0	0	1	3	0.3784
7	0	206	0	0	0	0.1004
8	0	2	27	0	0	0.069
9	1	0	0	34	2	0.1053
10	0	0	0	0	14	0
11	1	2	0	0	0	0.2867
Total	26	244	28	35	19	
ErrorO	0.1154	0.1557	0.0357	0.0286	0.2632	

	11	Total	ErrorC
1	11	106	0.3396
2	0	31	0.0645
3	0	27	0.4074
4	21	156	0.2628
5	8	202	0.3168
6	7	37	0.3784
7	2	229	0.1004
8	0	29	0.069
9	0	38	0.1053
10	0	14	0
11	199	279	0.2867
Total	248	1148	
ErrorO	0.1976		0.2413

Error Matrix Analysis of
SIST.EST. (columns : truth) against MAX345 (rows : mapped)

	1	2	3	4	5	
1	60	0	0	3	14	0.3258
2	0	25	3	0	0	0.1935
3	1	0	16	0	0	0.5556
4	14	0	0	118	9	0.2532
5	22	0	0	8	119	0.2654
6	4	0	0	0	0	0.3714
7	9	0	9	0	8	0.1011
8	0	1	0	0	0	0.1212
9	0	0	0	0	0	0.1765
10	0	0	0	0	0	0.1
11	23	0	0	24	18	0.2429
Total	133	26	28	153	168	
ErrorO	0.5489	0.0385	0.4286	0.2288	0.2917	

	6	7	8	9	10	
1	0	6	0	0	1	0.3258
2	0	0	1	0	2	0.1935
3	0	19	0	0	0	0.5556
4	0	1	0	0	0	0.2532
5	0	10	0	0	0	0.2654
6	22	0	0	1	3	0.3714
7	0	240	1	0	0	0.1011
8	0	3	29	0	0	0.1212
9	4	0	0	28	2	0.1765
10	0	0	0	2	18	0.1
11	2	1	0	0	0	0.2429
Total	28	280	31	31	26	
ErrorO	0.2143	0.1429	0.0645	0.0968	0.3077	

	11	Total	ErrorC
1	5	89	0.3258
2	0	31	0.1935
3	0	36	0.5556
4	16	158	0.2532
5	3	162	0.2654
6	5	35	0.3714
7	0	267	0.1011
8	0	33	0.1212
9	0	34	0.1765
10	0	20	0.1
11	212	280	0.2429
Total	241	1145	
ErrorO	0.1203		0.2253

Error Matrix Analysis of
SISTEMATICA (columns : truth) against MAX345 (rows : mapped)

	1	2	3	4	5	
1	68	0	0	5	6	0.2609
2	1	25	0	2	0	0.1667
3	0	1	20	0	0	0.5918
4	15	0	0	107	16	0.3141
5	20	0	0	10	121	0.3086
6	2	0	0	0	0	0.2609
7	10	0	6	0	9	0.1135
8	0	3	2	0	0	0.1951
9	0	0	0	0	0	0.0465
10	0	0	0	0	0	0.1154
11	26	0	1	25	26	0.2739
Total	142	29	29	149	178	
ErrorO	0.5211	0.1379	0.3103	0.2819	0.3202	

	6	7	8	9	10	
1	0	5	0	0	0	0.2609
2	0	1	0	0	1	0.1667
3	0	28	0	0	0	0.5918
4	0	0	0	0	0	0.3141
5	0	21	0	0	0	0.3086
6	17	0	0	0	2	0.2609
7	0	203	0	0	0	0.1135
8	0	3	33	0	0	0.1951
9	0	0	0	41	2	0.0465
10	1	0	0	2	23	0.1154
11	3	2	0	0	0	0.2739
Total	21	263	33	43	28	
ErrorO	0.1905	0.2281	0	0.0465	0.1786	

	11	Total	ErrorC
1	8	92	0.2609
2	0	30	0.1667
3	0	49	0.5918
4	18	156	0.3141
5	3	175	0.3086
6	2	23	0.2609
7	1	229	0.1135
8	0	41	0.1951
9	0	43	0.0465
10	0	26	0.1154
11	220	303	0.2739
Total	252	1167	
ErrorO	0.1270		0.2476

Error Matrix Analysis of
 BLOCADA (columns : truth) against MAX347 (rows : mapped)

	1	2	3	4	5	
1	134	0	0	9	0	0.0759
2	0	277	0	0	0	0
3	0	0	206	0	0	0.0925
4	16	0	0	292	0	0.0581
5	0	0	0	2	217	0.0398
6	0	0	0	0	0	0.0165
7	5	2	32	0	5	0.0552
8	0	2	5	0	0	0.0285
9	0	0	0	0	0	0.0105
10	0	0	0	0	0	0.01
11	7	0	0	0	1	0.0333
Total	162	281	243	303	223	
ErrorO	0.1728	0.0142	0.1523	0.0363	0.0269	

	6	7	8	9	10	
1	0	1	0	0	0	0.0759
2	0	0	0	0	0	0
3	0	6	15	0	0	0.0925
4	0	0	0	0	0	0.0581
5	0	5	0	0	0	0.0398
6	119	0	0	2	0	0.0165
7	0	753	0	0	0	0.0552
8	0	0	239	0	0	0.0285
9	4	0	0	378	0	0.0105
10	0	0	0	2	198	0.01
11	0	0	0	0	0	0.0333
Total	123	765	254	382	198	
ErrorO	0.0325	0.0157	0.0591	0.0105	0	

	11	Total	ErrorC
1	1	145	0.0759
2	0	277	0
3	0	227	0.0925
4	2	310	0.0581
5	2	226	0.0398
6	0	121	0.0165
7	0	797	0.0552
8	0	246	0.0285
9	0	382	0.0105
10	0	200	0.01
11	232	240	0.0333
Total	237	3171	
ErrorO	0.0211		0.0397

Error Matrix Analysis of
CASUALIZADA (columns : truth) against MAX347 (rows : mapped)

	1	2	3	4	5	
1	68	0	1	9	11	0.3333
2	0	30	0	0	0	0.0323
3	0	3	17	0	0	0.6458
4	24	0	0	116	12	0.3295
5	26	0	0	5	123	0.3204
6	1	0	0	0	0	0.2581
7	7	0	4	0	15	0.1268
8	0	0	0	0	0	0.0385
9	0	0	0	0	0	0.0571
10	1	0	0	0	0	0.1176
11	30	0	0	19	26	0.2784
Total	157	33	22	149	187	
ErrorO	0.5669	0.0909	0.2273	0.2215	0.3422	

	6	7	8	9	10	
1	0	5	0	0	0	0.3333
2	0	0	1	0	0	0.0323
3	0	26	2	0	0	0.6458
4	0	2	0	0	0	0.3295
5	0	20	0	0	0	0.3204
6	23	0	0	1	3	0.2581
7	0	186	0	0	0	0.1268
8	0	1	25	0	0	0.0385
9	1	0	0	33	1	0.0571
10	0	0	0	1	15	0.1176
11	2	4	0	0	0	0.2784
Total	26	244	28	35	19	
ErrorO	0.1154	0.2377	0.1071	0.0571	0.2105	

	11	Total	ErrorC
1	8	102	0.3333
2	0	31	0.0323
3	0	48	0.6458
4	19	173	0.3295
5	7	181	0.3204
6	3	31	0.2581
7	1	213	0.1268
8	0	26	0.0385
9	0	35	0.0571
10	0	17	0.1176
11	210	291	0.2784
Total	248	1148	
ErrorO	0.1532		0.2631

Error Matrix Analysis of
SIST.EST. (columns : truth) against MAX347 (rows : mapped)

	1	2	3	4	5	
1	61	0	0	7	10	0.3441
2	0	25	0	0	0	0
3	3	0	22	0	0	0.6812
4	14	0	0	122	13	0.2422
5	22	0	0	1	112	0.2533
6	3	0	0	3	1	0.4211
7	11	0	6	0	15	0.127
8	0	1	0	0	0	0.0357
9	0	0	0	0	0	0.2105
10	0	0	0	0	0	0
11	19	0	0	20	17	0.2161
Total	133	26	28	153	168	
ErrorO	0.5414	0.0385	0.2143	0.2026	0.3333	

	6	7	8	9	10	
1	0	7	0	0	1	0.3441
2	0	0	0	0	0	0
3	0	40	4	0	0	0.6812
4	0	0	0	0	0	0.2422
5	0	11	0	0	0	0.2533
6	22	1	0	1	3	0.4211
7	0	220	0	0	0	0.127
8	0	0	27	0	0	0.0357
9	4	0	0	30	4	0.2105
10	0	0	0	0	18	0
11	2	1	0	0	0	0.2161
Total	28	280	31	31	26	
ErrorO	0.2143	0.2143	0.129	0.0323	0.3077	

	11	Total	ErrorC
1	7	93	0.3441
2	0	25	0
3	0	69	0.6812
4	12	161	0.2422
5	4	150	0.2533
6	4	38	0.4211
7	0	252	0.127
8	0	28	0.0357
9	0	38	0.2105
10	0	18	0
11	214	273	0.2161
Total	241	1145	
ErrorO	0.1120		0.2376

Error Matrix Analysis of
SISTEMATICA (columns : truth) against MAX347 (rows : mapped)

	1	2	3	4	5	
1	69	0	0	7	6	0.2887
2	0	26	0	0	0	0.037
3	4	1	17	0	0	0.7671
4	17	0	0	114	14	0.3174
5	17	0	0	5	111	0.2839
6	2	0	0	2	0	0.3929
7	10	0	8	0	16	0.1545
8	0	2	2	0	0	0.1875
9	0	0	0	0	0	0.0455
10	0	0	0	0	0	0.08
11	23	0	2	21	31	0.2809
Total	142	29	29	149	178	
ErrorO	0.5141	0.1034	0.4138	0.2349	0.3764	

	6	7	8	9	10	
1	0	8	0	0	0	0.2887
2	0	0	1	0	0	0.037
3	0	45	6	0	0	0.7671
4	0	2	0	0	1	0.3174
5	0	16	0	0	0	0.2839
6	17	0	0	0	2	0.3929
7	0	186	0	0	0	0.1545
8	0	2	26	0	0	0.1875
9	0	0	0	42	2	0.0455
10	1	0	0	1	23	0.08
11	3	4	0	0	0	0.2809
Total	21	263	33	43	28	
ErrorO	0.1905	0.2928	0.2121	0.0233	0.1786	

	11	Total	ErrorC
1	7	97	0.2887
2	0	27	0.037
3	0	73	0.7671
4	19	167	0.3174
5	6	155	0.2839
6	5	28	0.3929
7	0	220	0.1545
8	0	32	0.1875
9	0	44	0.0455
10	0	25	0.08
11	215	299	0.2809
Total	252	1167	
ErrorO	0.1468		0.2751

Error Matrix Analysis of
BLOCADA (columns : truth) against MAX234 (rows : mapped)

	1	2	3	4	5	
1	65	0	0	7	3	0.1558
2	5	269	0	0	0	0.0182
3	0	8	223	0	0	0.1289
4	75	0	0	228	1	0.2597
5	3	0	0	25	218	0.1174
6	0	0	1	0	0	0.0161
7	9	1	18	0	1	0.0978
8	0	1	1	0	0	0.0102
9	0	0	0	0	0	0
10	0	2	0	0	0	0.0198
11	5	0	0	43	0	0.1756
Total	162	281	243	303	223	
ErrorO	0.5988	0.0427	0.0823	0.2475	0.0224	

	6	7	8	9	10	
1	0	1	0	0	0	0.1558
2	0	0	0	0	0	0.0182
3	0	17	8	0	0	0.1289
4	0	0	0	0	0	0.2597
5	0	0	0	0	0	0.1174
6	122	0	0	0	0	0.0161
7	0	747	52	0	0	0.0978
8	0	0	194	0	0	0.0102
9	0	0	0	380	0	0
10	0	0	0	2	198	0.0198
11	1	0	0	0	0	0.1756
Total	123	765	254	382	198	
ErrorO	0.0081	0.0235	0.2362	0.0052	0	

	11	Total	ErrorC
1	1	77	0.1558
2	0	274	0.0182
3	0	256	0.1289
4	4	308	0.2597
5	1	247	0.1174
6	1	124	0.0161
7	0	828	0.0978
8	0	196	0.0102
9	0	380	0
10	0	202	0.0198
11	230	279	0.1756
Total	237	3171	
ErrorO	0.0295		0.0937

Error Matrix Analysis of
CASUALIZADA (columns : truth) against MAX234 (rows : mapped)

	1	2	3	4	5	
1	53	1	0	15	15	0.4904
2	4	31	1	1	0	0.2791
3	4	0	15	0	0	0.6154
4	40	0	0	63	16	0.5655
5	23	0	0	19	142	0.268
6	0	0	0	3	0	0.2647
7	8	1	6	0	8	0.113
8	0	0	0	0	0	0.0476
9	1	0	0	0	0	0.1026
10	0	0	0	0	0	0
11	24	0	0	48	6	0.2847
Total	157	33	22	149	187	
ErrorO	0.6624	0.0606	0.3182	0.5772	0.2406	

	6	7	8	9	10	
1	0	13	0	0	0	0.4904
2	0	0	1	0	5	0.2791
3	0	16	4	0	0	0.6154
4	0	2	0	0	0	0.5655
5	0	5	0	0	0	0.268
6	25	0	0	0	0	0.2647
7	0	204	3	0	0	0.113
8	0	1	20	0	0	0.0476
9	0	0	0	35	3	0.1026
10	0	0	0	0	11	0
11	1	3	0	0	0	0.2847
Total	26	244	28	35	19	
ErrorO	0.0385	0.1639	0.2857	0	0.4211	

	11	Total	ErrorC
1	7	104	0.4904
2	0	43	0.2791
3	0	39	0.6154
4	24	145	0.5655
5	5	194	0.268
6	6	34	0.2647
7	0	230	0.113
8	0	21	0.0476
9	0	39	0.1026
10	0	11	0
11	206	288	0.2847
Total	248	1148	
ErrorO	0.1694		0.2988

Error Matrix Analysis

SIST.EST. (columns : truth) against MAX234 (rows : mapped)

	1	2	3	4	5	
1	39	0	0	7	17	0.5063
2	4	24	3	1	0	0.3684
3	1	0	19	0	0	0.5957
4	28	1	1	83	18	0.4467
5	24	0	0	21	121	0.3278
6	0	0	0	6	0	0.2424
7	17	0	3	0	11	0.1382
8	0	1	1	0	0	0.1304
9	0	0	0	0	0	0.0938
10	0	0	0	0	0	0.1053
11	20	0	1	35	1	0.223
Total	133	26	28	153	168	
ErrorO	0.7068	0.0769	0.3214	0.4575	0.2798	

	6	7	8	9	10	
1	0	10	0	0	0	0.5063
2	0	0	0	0	6	0.3684
3	0	23	4	0	0	0.5957
4	1	1	0	0	0	0.4467
5	0	7	0	0	0	0.3278
6	25	0	0	0	0	0.2424
7	0	237	7	0	0	0.1382
8	0	1	20	0	0	0.1304
9	0	0	0	29	3	0.0938
10	0	0	0	2	17	0.1053
11	2	1	0	0	0	0.223
Total	28	280	31	31	26	
ErrorO	0.1071	0.1536	0.3548	0.0645	0.3462	

	11	Total	ErrorC
1	6	79	0.5063
2	0	38	0.3684
3	0	47	0.5957
4	17	150	0.4467
6	2	33	0.2424
7	0	275	0.1382
8	0	23	0.1304
9	0	32	0.0938
10	0	19	0.1053
11	209	269	0.223
Total	241	1145	
ErrorO	0.1328		0.2812

Error Matrix Analysis of
SISTEMATICA (columns : truth) against MAX234 (rows : mapped)

	1	2	3	4	5	
1	46	0	0	8	18	0.5106
2	2	26	0	0	0	0.2121
3	1	1	22	0	0	0.5686
4	34	0	0	71	17	0.5
5	20	0	0	25	120	0.337
6	0	0	2	1	0	0.2759
7	11	0	4	0	8	0.1208
8	0	2	0	0	0	0.125
9	0	0	0	1	0	0.1304
10	0	0	0	0	0	0.1
11	28	0	1	43	15	0.2932
Total	142	29	29	149	178	
ErrorO	0.6761	0.1034	0.2414	0.5235	0.3258	

	6	7	8	9	10	
1	0	15	0	0	0	0.5106
2	0	0	0	0	5	0.2121
3	0	21	6	0	0	0.5686
4	0	2	0	0	0	0.5
5	0	10	0	0	0	0.337
6	21	0	0	1	0	0.2759
7	0	211	6	0	0	0.1208
8	0	1	21	0	0	0.125
9	0	0	0	40	5	0.1304
10	0	0	0	2	18	0.1
11	0	3	0	0	0	0.2932
Total	21	263	33	43	28	
ErrorO	0	0.1977	0.3636	0.0698	0.3571	

	11	Total	ErrorC
1	7	94	0.5106
2	0	33	0.2121
3	0	51	0.5686
4	18	142	0.5
5	6	181	0.337
6	4	29	0.2759
7	0	240	0.1208
8	0	24	0.125
9	0	46	0.1304
10	0	20	0.1
11	217	307	0.2932
Total	252	1167	
ErrorO	0.1389		0.3033