



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

**INPE-5944-TDI/570**

**EXATIDÃO NO PROCESSO DE MAPEAMENTO TEMÁTICO DA  
VEGETAÇÃO DE UMA ÁREA DE MATA ATLÂNTICA  
NO ESTADO DE SÃO PAULO, A PARTIR DE  
IMAGENS TM-LANDSAT**

Elaine Cristina Cardoso Fidalgo

Dissertação de Mestrado em Sensoriamento Remoto, orientada pela Dra. Thelma Krug e pelo Dr. João Roberto dos Santos, aprovada em 24 de agosto de 1995.

528.711.7:630(816.1)

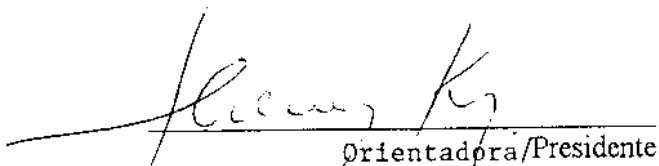
FIDALGO, E.C.C.

Exatidão no processo de mapeamento temático da vegetação de uma área de Mata Atlântica no Estado de São Paulo, a partir de imagens TM-Landsat / E. C.C. Fidalgo. - São José dos campos: INPE, 1995. 186p. - (INPE-5944-TDI/570).

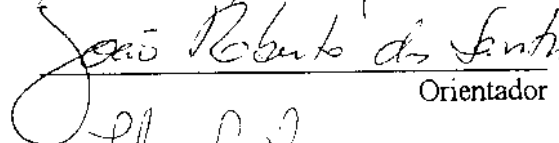
1. Mapeamento temático. 2. Vegetação. 3. Mata Atlântica. I. Título

Aprovada pela Banca Examinadora em  
cumprimento a requisito exigido para a  
obtenção do Título de Mestre em  
Sensoriamento Remoto

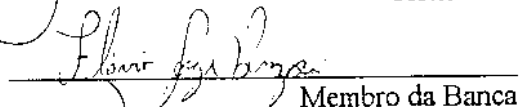
Dra. Thelma Krug

  
Orientadora/Presidente

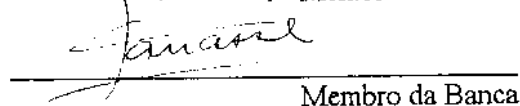
Dr. João Roberto dos Santos

  
Orientador

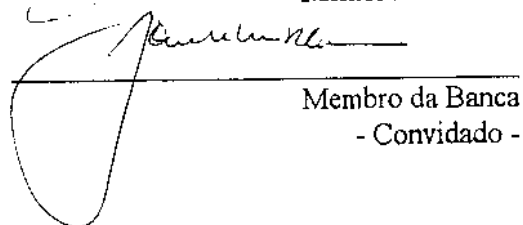
Dr. Flávio Jorge Ponzoni

  
Membro da Banca

Dra. Corina da Costa Freitas Yanasse

  
Membro da Banca

Dr. Jansle Vieira da Rocha

  
Membro da Banca  
- Convidado -

Candidata: Elaine Cristina Cardoso Fidalgo

São José dos Campos, 24 de agosto de 1995



### **Del rigor en la ciencia**

...En aquel Imperio, el Arte de la Cartografía logró tal perfección que el mapa de una sola Provincia ocupaba toda una Ciudad, y el mapa del imperio, toda una Provincia. Com el tiempo, esos Mapas Desmesurados no satisficieron y los Colegios de Cartógrafos levantaron un Mapa del Imperio, que tenía el tamaño del Imperio y coincidía puntualmente con él. Menos Adictas al Estudio de la Cartografía, las Generaciones Sigüientes entendieron que ese dilatado Mapa era Inútil y no sin Impiedad lo entregaron a las Inclemencias del Sol y de los Inviernos. En los desiertos del Oeste perduran despedazadas Ruinas del Mapa, habitadas por Animales y por Mendigos; en todo el País no hay otra reliquia de las Disciplinas Geográficas.

Suárez Miranda, *Viajes de varones prudentes*.

libro cuarto, cap. XLV, Lérida, 1658.

(Conto do livro *Narraciones* de Jorge Luiz Borges).



Este trabalho é dedicado às pessoas que colaboraram direta ou indiretamente com a sua realização, especialmente ao companheiro de todos os momentos, João, e à minha família.





## AGRADECIMENTOS

O desenvolvimento deste trabalho contou com a preciosa colaboração de pessoas e entidades, às quais a autora expressa seus agradecimentos.

À Dra. Thelma Krug, orientadora deste trabalho, pelas sugestões e críticas imprescindíveis à sua realização, e pelo apoio e incentivo durante o curso de mestrado.

Ao Dr. João Roberto dos Santos, pela orientação, críticas e sugestões, especialmente nos temas relacionados à vegetação.

Ao Dr. Flávio Ponzoni, pelo incentivo ao ingresso no curso de mestrado e pelo apoio significativo ao trabalho, especialmente na elaboração de uma referência para a análise de exatidão de classificação, em conjunto com Dra. Evelyn M. L. de Moraes Novo, a quem os agradecimentos também são devidos.

A Camilo Daleles Rennó pela interpretação e elaboração de um dos mapas temáticos.

A Renê A. Novaes Jr. pela introdução dos mapas temáticos no SGI.

À Coordenadoria de Proteção dos Recursos Naturais (CPRN), ao Departamento Estadual de Recursos Naturais (DEPRN) da Secretaria do Meio Ambiente do Estado de São Paulo, e à Fundação de Ciência, Aplicações e

Tecnologia Espaciais (FUNCATE) pela liberação da autora, dos trabalhos que vinha desenvolvendo nestes órgãos, para o ingresso no curso de mestrado.

A Zeni Goes Brandino pelas informações e auxílio na seleção do material utilizado.

Ao Instituto Geográfico e Cartográfico (IGC) da Coordenadoria de Planejamento Regional da Secretaria de Planejamento e Gestão do Estado de São Paulo pelo empréstimo das fotografias aéreas da região da Baixada Santista.

Ao Engenheiro Agrônomo Eduardo Pereira Lustosa e ao Engenheiro Florestal Evandro José A. Patton pelo auxílio no desenvolvimento do trabalho de campo.

A Anibal Evaristo Fernandes e Moacir Godoy Júnior pelos programas desenvolvidos especialmente para o desenvolvimento deste trabalho.

A Joaquim Godoi Filho pelo desenho das cartas de vegetação.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de pesquisa.

Aos alunos do Curso de Mestrado em Sensoriamento Remoto, Camilo Daleles Rennó e José Carlos de Souza Leite, pelas contribuições dadas durante a

participação do grupo de estudos sobre o tema "exatidão de mapeamento" orientado pela Dra. Thelma Krug.

Aos amigos do curso de mestrado por tudo que aprendemos juntos.

Aos funcionários do INPE que colaboraram direta ou indiretamente com este trabalho, especialmente a Maria Etelvina R. D. Arbex, aos funcionários da Biblioteca e da Divisão de Sensoriamento Remoto.



## RESUMO

O presente trabalho foi realizado visando auxiliar os usuários de produtos de sensoriamento remoto no processo de seleção de métodos de análise da exatidão de mapeamentos propostos na literatura. Seu objetivo geral é a aplicação de métodos estatísticos para estimar a exatidão no processo de mapeamento temático e a análise de algumas fontes de erro que podem interferir nessa exatidão. Para tanto, foi selecionada uma área de Mata Atlântica na região da Baixada Santista, Estado de São Paulo, e realizado o mapeamento de sua vegetação utilizando imagens TM-LANDSAT. As fontes de erro foram analisadas a partir dos resultados da classificação realizada por três diferentes intérpretes, e da introdução dos mapas temáticos resultantes em um sistema de informações geográficas através da digitalização automática e da digitalização manual realizada por dois diferentes operadores. O delinamento amostral utilizado, baseado na amostragem estratificada sistemática não alinhada, e a coleta de dados de referência, feita predominantemente a partir de fotografias aéreas, mostraram-se adequados às análises realizadas. Os métodos aplicados à estimativa da exatidão (matrizes de erro originais, matrizes de erro normalizadas, exatidão mínima, coeficiente Kappa e teste de hipóteses) apresentaram coerência nos resultados, exceto no caso das matrizes normalizadas. As fontes de erro analisadas não influíram de forma significativa nos valores de exatidão dos produtos elaborados.



## SUMÁRIO

	<u>Pág.</u>
LISTA DE FIGURAS .....	xix
LISTA DE TABELAS .....	xxi
<u>CAPÍTULO 1 - INTRODUÇÃO</u> .....	1
<u>CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA</u> .....	7
2.1 - A vegetação da Mata Atlântica .....	7
2.1.1 - A Floresta Atlântica .....	7
2.1.2 - O manguezal .....	13
2.1.3 - A restinga .....	14
2.2 - Mapeamento da vegetação .....	18
2.3 - Sistema de informações geográficas como fer- ramenta para a criação de uma base de dados georreferenciados .....	26
2.3.1 - Digitalização manual .....	29
2.3.2 - Digitalização automática .....	29
2.4 - Exatidão de classificação dos dados de senso- riamento remoto .....	31
2.4.1 - Técnicas para a estimativa da exatidão de classificação .....	33
2.4.1.1 - Técnicas descritivas .....	34
2.4.1.2 - Técnicas analíticas .....	37
2.4.1.2.1 - Técnicas analíticas para comparação entre diferentes produtos da classi- ficação .....	56
2.4.2 - Tipos de amostragem .....	61
2.4.3 - Tamanho da amostra .....	66

2.4.4 - Dados de referência .....	71
<u>CAPÍTULO 3 - MATERIAL E MÉTODOS</u> .....	73
3.1 - Área de estudo .....	73
3.2 - Estabelecimento da legenda temática .....	79
3.3 - Interpretação visual das imagens .....	85
3.4 - Introdução das cartas temáticas em um sistema de informações geográficas .....	87
3.4.1 - Introdução dos dados através da digitaliza- ção manual .....	87
3.4.2 - Introdução dos dados através da digitaliza- ção automática .....	88
3.5 - Estabelecimento dos dados de referência para a análise de exatidão .....	89
3.5.1 - Delineamento amostral e tamanho da amostra ..	90
3.5.2 - Interpretação visual de fotografias aéreas ..	94
3.5.3 - Levantamento de campo .....	97
3.6 - Análise de exatidão de classificação .....	98
3.7 - Análise comparativa entre os produtos .....	102
<u>CAPÍTULO 4 - RESULTADOS E DISCUSSÃO</u> .....	107
4.1 - Produtos gerados .....	107
4.2 - Estabelecimento dos dados de referência .....	107
4.2.1 - Delineamento amostral selecionado .....	108
4.2.2 - Classificação das unidades amostrais .....	110
4.3 - Análise da exatidão de classificação dos pro- dutos .....	111
4.4 - Análise comparativa entre os produtos .....	132
<u>CAPÍTULO 5 - CONCLUSÕES E RECOMENDAÇÕES</u> .....	139



REFERÊNCIAS BIBLIOGRÁFICAS .....	145
APÊNDICE A - LISTAGEM DAS CARTAS E FOTOGRAFIAS AÉREAS ONDE FORAM LOCALIZADAS AS UNI- DADES AMOSTRAIS .....	159
APÊNDICE B - RELAÇÃO DAS CARTAS PLANIALTIMÉTRICAS DO PROJETO BAIXADA SANTISTA .....	163
APÊNDICE C - PROGRAMA DE NORMALIZAÇÃO DE MATRIZES ...	165
APÊNDICE D - CARTA DE VEGETAÇÃO - SANTOS - INTÉR- PRETE I .....	167
APÊNDICE E - CARTA DE VEGETAÇÃO - SANTOS - INTÉR- PRETE II .....	169



## LISTA DE FIGURAS

	<u>Pág.</u>
2.1 - Esquema do sistema de sensoriamento remoto ....	21
2.2 - Espectro de reflectância da folha do tabaco ...	22
2.3 - Distribuição amostral de unidades corretamente classificadas, identificando o risco do consumidor ( <i>A</i> ) e o risco do produtor ( <i>B</i> ) .....	45
2.4 - Distribuição amostral de unidades corretamente classificadas, identificando o risco do consumidor ( <i>A</i> ) e o risco do produtor ( <i>B</i> ), para uma amostra grande .....	46
2.5 - Representação da distribuição de $\hat{p}$ para amostras de tamanho <i>n</i> , mostrando a relação entre $p_0$ , <i>d</i> e $\alpha$ , numa distribuição normal, considerando $p=p_0$ .....	68
3.1 - Esquema de trabalho .....	74
3.2 - Esquema ilustrativo da amostragem estratificada sistemática não alinhada .....	93
3.3 - Detalhamento dos métodos utilizados para a análise da exatidão de classificação de cada produto .....	99
4.1 - Intervalos de confiança de 95% para a exatidão global, por produto .....	114
4.2 - Intervalos de confiança de 95% para o coeficiente Kappa, por produto .....	128
4.3 - Intervalos de confiança de 95% para a exatidão global dos produtos: INT II, VE2, VE5 .....	135



## LISTA DE TABELAS

	<u>Pág.</u>
2.1 - Bandas espectrais do sensor "Thematic Mapper" .	19
2.2 - Exemplo de matriz de erro .....	35
2.3 - Dados para medida de concordância de uma única categoria .....	54
3.1 - Planos de informação do projeto FLO referentes à vegetação .....	88
3.2 - Produtos comparados para a análise da influência de algumas fontes de erro na exatidão ...	103
4.1 - Matrizes de erro, por produto .....	112
4.2 - Estimativas da exatidão global (matrizes originais e normalizadas) e da exatidão mínima ( $p_l$ e $p'_l$ ), por produto .....	113
4.3 - Estimativas das exatidões do produtor e do consumidor, por classe e por produto .....	114
4.4 - Erros de omissão, por classe e por produto ....	115
4.5 - Erros de inclusão, por classe e por produto ...	116
4.6 - Exemplo de matriz de erro com os valores da exatidão do produtor e dos erros de omissão - intérprete I .....	118
4.7 - Matrizes de erro normalizadas, por produto ....	119
4.8 - Estimativas da exatidão, por classe e por produto, a partir das matrizes normalizadas ...	120
4.9 - Estimativas do coeficiente Kappa, por produto .	127
4.10 - Correlação entre os termos da Equação (2.20) e a variância estimada do coeficiente Kappa estimado .....	129

4.11 - Estimativas do coeficiente Kappa condicional ( $K_c$ ) e Kappa ( $K_c'$ ), por classe e por produto ...	130
4.12 - Hipóteses nulas, valores $\chi^2$ calculados e conclusões obtidas para testar a igualdade entre as exatidões globais dos produtos .....	133
4.13 - Valores $r$ calculados para testar a igualdade entre pares de valores do coeficiente Kappa dos produtos .....	135
4.14 - Hipóteses nulas, valores $\chi^2$ calculados e conclusões obtidas para testar a igualdade entre os valores do coeficiente Kappa dos produtos .....	138

## CAPÍTULO 1

### INTRODUÇÃO

Com a intensificação do uso de técnicas de sensoriamento remoto para a elaboração de mapas temáticos, aumentou a necessidade de estabelecer critérios mínimos de qualidade para os produtos gerados. Kalensky (1985) sugeriu que uma análise rigorosa sobre a exatidão dos produtos fosse incluída em cada projeto de sensoriamento remoto, em particular naqueles que envolvessem o mapeamento temático. Segundo o autor, mapas sem referências à exatidão são somente aproximações gráficas que não consideram a qualidade dos produtos.

Mesmo conscientes da importância de associar uma estimativa da exatidão aos mapeamentos temáticos, os usuários de produtos de sensoriamento remoto podem se deparar com uma dificuldade, qual seja, a diversidade de propostas metodológicas na literatura para esse fim.

O presente trabalho foi elaborado visando auxiliar no processo de seleção de métodos propostos na literatura. Dentro deste contexto, seu objetivo geral é a aplicação de métodos estatísticos para estimar a exatidão no processo de mapeamento temático e a análise de algumas fontes de erro que podem interferir nessa exatidão. Essas fontes de erro foram analisadas nos seguintes processos:

- na classificação visual da imagem, a partir dos resultados obtidos por três diferentes intérpretes;

- na entrada dos dados em um sistema de informações geográficas, através da digitalização manual e automática; e
- na digitalização manual, a partir dos resultados obtidos por dois diferentes operadores.

Dentre os procedimentos considerados na seleção do método estatístico e que compõem os objetivos específicos deste trabalho, citam-se os seguintes:

- uma ampla revisão bibliográfica sobre os métodos propostos na literatura e a padronização da notação empregada;
- a seleção de um tipo de amostragem e um tamanho de amostra baseados na literatura e a avaliação de sua adequação aos métodos usados para a análise de exatidão;
- a apresentação de um método para a coleta de dados de referência;
- a avaliação de métodos estatísticos propostos na literatura para a estimativa da exatidão de mapeamentos temáticos e para a comparação entre diferentes produtos da classificação.

Para o desenvolvimento deste trabalho selecionou-se uma área da Floresta Atlântica. A seleção dessa área se justificou por tratar-se a Floresta Atlântica de uma das mais comprometidas florestas tropicais do mundo. O processo de devastação que a atinge é antigo, ocorrendo



desde a época do descobrimento do Brasil (Costa e Corrêa, 1992).

Esse processo acarretou uma drástica redução da cobertura vegetal dessa floresta e de seus ecossistemas associados. Da vegetação original que se estendia do Rio Grande do Norte até o Rio Grande do Sul, pouco resta preservado. Os melhores e mais preservados remanescentes da Floresta Atlântica estão situados no Sudeste do Brasil, particularmente nos Estados de São Paulo e Paraná, graças à topografia íngreme da Serra do Mar, que impediu o uso dessas áreas para atividades agrícolas (Leitão Filho, 1993).

As exigências crescentes da sociedade por medidas de proteção e a acelerada devastação desses remanescentes sensibilizaram governantes e legisladores. Assim, foram criadas inúmeras unidades de conservação na região de ocorrência de Floresta Atlântica e, em outubro de 1994, a Nova Constituição Brasileira foi promulgada declarando a Mata Atlântica e a Serra do Mar como patrimônios nacionais (Costa e Corrêa, 1992).

Na Mata Atlântica encontra-se a região da Baixada Santista, cuja problemática ambiental, causada principalmente pela poluição provocada pelo pólo petroquímico de Cubatão, desencadeou algumas medidas de proteção e de recuperação pelos governos estadual e dos municípios da região.

Para uma ação efetiva de conservação é fundamental o conhecimento da extensão e da distribuição dos remanescentes da vegetação. Uma importante ferramenta para esses estudos são as imagens do sensor TM-LANDSAT, utilizadas no reconhecimento de padrões de cobertura vegetal para inúmeras aplicações.

Mapeamentos dos remanescentes da Mata Atlântica a partir da interpretação visual de imagens do TM foram realizados pela Secretaria do Meio Ambiente do Estado de São Paulo (Hahn et al., 1990; Covre et al., 1992) e pela Fundação SOS Mata Atlântica e INPE (1993).

No presente trabalho, a análise de exatidão foi realizada na carta Santos do mapeamento de vegetação natural do Estado de São Paulo, feito pela Secretaria do Meio Ambiente (São Paulo, 1989). Este material foi escolhido por ser de grande importância à proteção do meio ambiente na região e no Estado, sendo empregado principalmente na fiscalização e no licenciamento dos recursos naturais da região pelo governo estadual.

A análise da exatidão de produtos como esse é fundamental para assegurar a confiabilidade dos trabalhos que neles se baseiam.

O presente trabalho é apresentado em cinco capítulos. O Capítulo 2 apresenta uma descrição da vegetação da Mata Atlântica e as considerações teóricas que subsidiaram a escolha da metodologia adotada. O Capítulo 3 descreve a área de estudo e os procedimentos metodológicos

utilizados. Os Capítulos 4 e 5 apresentam os resultados, e as conclusões e recomendações, respectivamente.



## CAPÍTULO 2

### FUNDAMENTAÇÃO TEÓRICA

Neste capítulo apresenta-se uma descrição da vegetação da Mata Atlântica e tecem-se as considerações teóricas necessárias para subsidiar a escolha da metodologia adotada.

#### 2.1 - A VEGETAÇÃO DA MATA ATLÂNTICA

O termo Mata Atlântica refere-se genericamente aos ecossistemas da área costeira, tendo como formação predominante a Floresta Ombrófila Densa ou Floresta Pluvial Tropical (IBGE, 1992), também denominada Floresta Atlântica, e os ecossistemas associados, como os manguezais e as restingas.

##### 2.1.1 - A FLORESTA ATLÂNTICA

A Floresta Atlântica, presente nas áreas costeiras, reunida a outra fisionomia ecológica localizada na Amazônia, compõem a Floresta Ombrófila Densa. Esse tipo de vegetação foi subdividido em cinco formações, ordenadas segundo uma hierarquia topográfica que reflete fisionomias diferentes associadas às variações ecotípicas resultantes de ambientes distintos. Dessa subdivisão resultaram as seguintes formações:

- Floresta Ombrófila Densa Aluvial: também denominada "floresta ciliar", não varia topograficamente e apresenta sempre ambientes repetitivos nos terraços aluviais dos flúvios;

- Floresta Ombrófila Densa das Terras Baixas nas altitudes entre 5 e 50 metros quando situada entre 16° e 24° de latitude Sul e entre 5 e 30 metros quando situada entre 24° e 32°;
- Floresta Ombrófila Densa Submontana: localizada nas altitudes entre 50 e 500 metros quando situada entre 16° e 24° de latitude Sul e entre 30 e 400 metros quando situada entre 24° e 32°;
- Floresta Ombrófila Densa Montana: localizada nas altitudes entre 500 e 1500 metros quando situada entre 16° e 24° de latitude Sul e entre 400 e 1000 metros quando situada entre 24° e 32°; e
- Floresta Ombrófila Densa Alto-Montana: situada acima dos limites estabelecidos para a formação Montana (IBGE, 1992).

A Floresta Atlântica é composta por formações que possuem flora e estruturas próprias, as quais têm-se mostrado muito complexas, chegando mesmo a apresentar riqueza e diversidade maiores que algumas florestas amazônicas (Klein, 1980; citado por Mantovani et al., 1990; Silva e Leitão Filho, 1982; citados por Mantovani et al., 1990).

A Floresta Atlântica, estruturada em diversas sinusias ou estratos lenhosos (árvores, arvoretas e arbustos) bastante desenvolvidos, seguidos por um estrato herbáceo bastante diversificado, é constituída por uma composição florística muito heterogênea que, aliada à uma

profusão de epífitas e lianas, ostenta uma pujança tipicamente de floresta tropical (Klein, 1990).

Segundo Leitão Filho (1993), a Floresta Atlântica é muito pouco conhecida sob o ponto de vista florístico, existindo divergências entre estudiosos quanto à sua conceituação e à determinação dos seus limites reais. Para o autor, ao longo do domínio atlântico ocorrem vários tipos de florestas. Em termos florísticos, o chamado domínio atlântico abrigava e ainda abriga floras diversas quanto a suas composições e origens, mas, em termos fisionômicos, havia uma certa continuidade de cobertura, bastante descaracterizada atualmente em função da pesada perturbação antrópica.

Ao estudarem a vegetação da Serra do Mar em Salesópolis, Estado de São Paulo, Mantovani et al. (1990) observaram que as variações locais, a distância do oceano, a influência de floras diversas e os estágios sucessionais são fatores que contribuem para a complexidade das formações encontradas na encosta atlântica. Nesse trabalho os autores identificaram quatro padrões florestais: a mata de topo de morros, a mata de fundo de vale, a mata de encosta voltada para o oceano Atlântico e a mata de encosta voltada para o vale do Paraíba.

A mata de topo de morros é uma formação condicionada pela presença de solos rasos (litosolos) e que apresenta um dossel contínuo. Pode ser contígua ao campo onde ocorrem afloramentos rochosos, e com o qual pode ter espécies comuns, representando uma sere sucessional

posterior àquela campestre. Geralmente apresenta em seu interior populações densas de Bromeliaceas terrícolas e, em muitas áreas, de espécies de *Chusquea* sp. (Gramineae).

A mata de fundo de vale situa-se sobre solos profundos, relacionados com a rede hidrográfica. As espécies arbóreas do dossel atingem 20 metros de altura, havendo indivíduos emergentes maiores que 30 metros. De modo geral, encontram-se em áreas de difícil acesso, em vales profundos ou grotas formadas entre escarpas íngremes. Possui um subosque com grande riqueza florística e que apresenta muitos indivíduos de *Euterpe edulis* (Arecaceae). É riquíssima em epífitas, estabelecidas notadamente sobre indivíduos dos estratos mais elevados.

A mata de encosta voltada para o oceano Atlântico é um tipo de formação que apresenta variações estruturais muito grandes, pois é dependente da cota em que se encontra, da declividade e das influências das massas de ar vindas do oceano (Klein, 1980, citado por Mantovani et al., 1990). Pode ter características marcadamente ecotonais entre as matas de topo de morros e de fundo de vale. Por encontrar-se em declives, o dossel permite a penetração de luz difusa, favorecendo a riqueza em epífitas e em espécies de subosque.

A mata de encosta voltada para o vale do Paraíba contém algumas espécies decíduas, mas assemelha-se estruturalmente à formação anterior.



Mantovani et al. (1990) observaram que os processos sucessionais na região têm formas bastante diferentes dependendo do local, da intensidade e da extensão da perturbação ocorrida, mas apresentam padrões que são estabelecidos por diferentes espécies que exercem funções semelhantes nas formações sucessionais.

As áreas com vegetação secundária caracterizam-se, em geral, por uma pequena diversidade de espécies e homogeneidade fisionômica em cada estágio sucessional ou seral. É comum encontrar algumas espécies dominantes em cada sere, que lhe conferem o aspecto homogêneo.

Embora essas áreas ocupadas pela vegetação secundária sejam típicas pelo seu aspecto fisionômico, encontram-se agrupamentos de espécies distintos, característicos de cada zona de vegetação. Dessa forma, existem espécies mais características das formações de média e alta encosta e outras que têm ampla distribuição, ocorrendo em todas as zonas ou faixas de vegetação.

As seres sucessionais são distintas nos casos de iniciarem-se após a derrubada pura e simples da vegetação primária, após o cultivo sucessivo durante um ou mais anos e posterior abandono, ou no caso da retirada de solo e de deslizamentos naturais ou provocados.

No caso da sere sucessional após a derrubada da mata primária, a regeneração de arvoretas ou árvores de médio porte é intensa, formando-se rapidamente uma

cobertura vegetal arbórea densa, com representantes das famílias Compositae, Melastomataceae e Solanaceae, principalmente. Essa formação, denominada de mata pioneira, evolui para um estágio secundário inicial e depois para o estágio secundário tardio, muito difícil de distinguir da verdadeira mata primária.

No caso de áreas de cultivo abandonadas, a sere sucessional é mais lenta, iniciando-se com a ocupação do terreno por espécies ruderais e arbustos de pequeno porte. Essa primeira etapa é normalmente denominada de pioneira ou carrascal. Havendo fonte de sementes próxima, esse estágio pioneiro evolui para os estágios posteriores de sucessão num processo mais lento do que no primeiro caso descrito.

Nas áreas de retirada de solo e cortes de estradas, o processo sucessional é ainda mais demorado, sendo o terreno ocupado primeiramente por musgos, líquens e pteridófitas para, posteriormente, instalarem-se espécies arbustivas e, depois, arbóreas.

Nas áreas de início de regeneração encontram-se principalmente ervas ou arbustos, como *Rubus rosaefolius* (Rosaceae), *Capsicum frutescens*, *Solanum swartzianum*, *Solanum americanum* (Solanaceae), *Croton lundianus* (Euphorbiaceae), *Piper gaudichaudianum*, *Potomorphe umbellata* (Piperaceae), *Hedyosmium brasiliensis* (Chloranthaceae) e *Baccharis genistelloides* (Compositae).

Nos estágios seguintes, encontram-se espécies com indivíduos arborescentes ou árvores de pequeno porte. As espécies mais comuns são as várias *Tibouchina* sp (Melastomataceae), *Solanum granulosum-leprosum* (Solanaceae), *Miconia cabucu*, *Miconia theaezans* (Melastomataceae), *Trema micrantha* (Ulmaceae), *Cecropia pachystachia*, *Cecropia glaziou* (Cecropiaceae), *Casearia sylvestris* (Flacourtiaceae), *Croton floribundus*, *Croton urucurana* (Euphorbiaceae), *Urera baccifera* (Urticaceae), *Cytharexylum myrianthum* (Verbenaceae) e *Mimosa scabrella* (Leguminosae), entre outras.

#### 2.1.2 - O MANGUEZAL

Manguezal é a denominação dada às formações vegetais pioneiras com influência fluviomarinha, de porte arbóreo (IBGE, 1992). Essas formações, que se estendem ao longo da zona costeira, estão expostas aos processos transicionais do ambiente marinho, estuarino e lagunar, com alternância de inundações derivadas da atuação das marés.

O ambiente resultante da mistura de águas continentais e marinhas pela ação das correntes geradas por propagação das marés é essencial na composição das unidades de manguezais. Os movimentos hidráulicos e a propagação das correntes, dificultada pelas raízes e troncos, formam o ambiente apropriado para a precipitação do material em suspensão transportado pelas águas estuarinas, com a formação de um substrato de composição predominantemente orgânica (Herz, 1991).

De características fortemente marcadas pela salinização do substrato, as formações de manguezais abrigam espécies vegetais altamente especializadas, de morfologia singular e fisiologia compatibilizada à alternância do ingresso das águas estuarinas, envolvendo seus troncos e sistema radicular por períodos de duração variável. Essas espécies pertencem aos gêneros *Rhizophora*, *Avicennia* e *Laguncularia*. Pode-se observar ainda espécies facultativas dos gêneros *Hibiscus*, *Acrostichum* e *Spartina* (Herz, 1991).

A predominância de alguma das espécies típicas de manguezais em função da localização dessas áreas foi observada por Kirizawa et al. (1992), na Ilha Comprida, Estado de São Paulo. Na região lagunar, com alto teor salino, predominam indivíduos de *Rhizophora mangle*. Nos mangues interiores, próximos à foz de rios e nas bordas de canais de drenagens, a espécie predominante é *Laguncularia racemosa*. Em locais mais secos do mangue, em áreas de transição para a restinga, as espécies mais freqüentes são *Avicennia schaueriana*, *Hibiscus tiliaceus* e *Acrostichum aureum*.

### 2.1.3 - A RESTINGA

Variações nos níveis dos oceanos ocorridas na época holocena, entre sete e dois mil anos atrás, ocasionaram a formação, ao longo do que é hoje a costa brasileira, de vastas planícies sedimentares arenosas. Em consequência das transgressões e regressões marinhas elas tomaram a forma de uma sucessão de cordões litorâneos

arenosos. Após a deposição da areia pelo mar, as tempestades, as correntes litorâneas e os ventos passaram a modelar uma topografia complexa e diversificada, que pode assumir a forma de altas barreiras que bloqueiam a foz dos rios ou separam lagunas do mar, de dunas móveis de altura variável ou ainda de planícies de cordões arenosos e relevo pouco acidentado (Araújo e Lacerda, 1987).

Esses depósitos arenosos são, em geral, cobertos por comunidades vegetais características e também muito diversas, como campos ralos de gramíneas, matas fechadas ou brejos com densa vegetação aquática. A esse conjunto de formações geomorfológicas e às diferentes comunidades biológicas que as ocupam dá-se o nome genérico de restingas (Araújo e Lacerda, 1987).

As restingas brasileiras ainda são pouco conhecidas devido à falta de estudos florísticos, estruturais e ambientais, dificultando a delimitação global das suas diversas formações vegetais (Araújo et al., 1984, citados por Kirizawa et al., 1992).

Kirizawa et al. (1992), citaram a grande dificuldade em delimitar as formações vegetais de restinga, em especial as de transição, em função da complexidade desse ecossistema e do uso variado da terminologia restinga nos campos de botânica, geologia e ecologia. Face a isto, os autores adotaram o conceito amplo dado por Araújo et al. (1984, citados por Kirizawa et al., 1992), em que o termo restinga refere-se ao ecossistema adjacente ao oceano, encontrado em planícies arenosas.

Segundo o IBGE (1992), as restingas são formações pioneiras com influência marinha podendo apresentar porte herbáceo, arbustivo ou arbóreo.

Ao descrever a vegetação da Ilha Comprida, Estado de São Paulo, Kirizawa et al. (1992) dividiram a vegetação de restinga em quatro formações: a formação pioneira de dunas, o escrube de restinga, os brejos de restinga e a mata de restinga.

A formação pioneira de dunas é constituída por psamo-halófitas (vegetação que habita meios arenosos e salinos) ou psamófilas herbáceas (vegetação que apresenta preferência por solo arenoso), periodicamente atingidas por borrifos de água salgada ou pela maré, em especial em épocas de ressaca.

Em seus aspectos fisionômicos e florísticos, a vegetação pioneira de dunas da região estudada assemelha-se à descrita por outros autores para as demais regiões do litoral paulista. Dentre as espécies encontradas, foram citadas *Blutaparon portulacoides*, *Hidrocotyle bonariensis*, *Ipomoea pes-caprae*, *Remirea maritima*, *Spartina alterniflora*, *Acycarpha spathulata* e *Polygala cyparissias*.

O escrube de restinga é uma formação vegetal que, em função do local de ocorrência, pode apresentar porte herbáceo-arbustivo até arbustivo-arbóreo com indivíduos lenhosos de caules tortuosos ou ramificados desde a base, não ultrapassando 7 metros de altura. A vegetação de porte arbustivo-arbóreo se encontra

estabelecida sobre sedimentos arenosos já consolidados, cuja camada superficial, de coloração mais escura, apresenta na superfície pequena deposição de matéria orgânica ainda não decomposta. Algumas vezes é coberta por um denso "tapete" de Bromeliáceas. Compõem o estrato arbóreo, entre outras espécies, *Gomidesia gaudichaudiana*, *Ocotea pulchella*, *Myrcia rostrata*, *Andira anthelmia*, *Ilex theezans*, *Erythroxylum ambiguum*, *Psidium cattleyanum* e *Clusia criuva*.

Os brejos de restinga referem-se às áreas permanente ou periodicamente saturadas ou cobertas por água, localizadas nas depressões. Nesse ambiente de rica produtividade primária e solos hidromórficos ocorrem Gramíneas, Ciperáceas, Juncáceas e macrófitas aquáticas. Em locais mais elevados, com solo menos úmido e mais arejado, a vegetação pode ser constituída por exemplares herbáceo-arbustivos e alguns exemplares arbóreos como palmeiras dos gêneros *Attalea* e *Syagrus* e a Bignoniacea *Tabebuia cassinoides*.

A mata de restinga é encontrada em superfícies arenosas já consolidadas, de coloração fortemente escura e rica em matéria orgânica e húmus na camada superficial. Apresenta árvores de até 20 metros de altura. Podem-se distinguir dois estratos arbóreos: o superior, com cerca de 10 metros de altura, onde ocorrem *Rapanea ferruginea*, *Andira fraxinifolia*, *Matayba elaeagnoides*, *Ocotea pulchella*, e o inferior, com altura média de 5 metros, onde se observam as espécies *Gomidesia gaudichaudiana*, *Endlicheria paniculata* e

*Nectandra grandiflora*, entre outras. Em locais onde a camada de matéria orgânica é mais densa (aproximadamente 20 centímetros) essa formação apresenta árvores de maior porte nos dois estratos e observa-se um aumento no número e na diversidade das epífitas.

## 2.2 - MAPEAMENTO DA VEGETAÇÃO

O conhecimento das características espectrais de dosséis e de sua reflectância nas regiões do espectro referentes às bandas nas quais atuam os sensores, auxilia na identificação e na delimitação das áreas de vegetação a partir das imagens geradas por esses sensores. O resultado desse processo pode ser um mapa de vegetação. A proposta do mapeamento é armazenar informações sobre a vegetação de uma forma simbólica, em um formato que permita recuperá-las com facilidade, exatidão e em pouco tempo (Mulder, 1988).

O sensor "Thematic Mapper" (TM), atualmente operando a bordo do satélite LANDSAT 5, tem sido muito utilizado no mapeamento de vegetação. Seus dados são adquiridos em 7 bandas espectrais, cujas faixas espectrais (em micrômetros,  $\mu\text{m}$ ) são apresentadas na Tabela 2.1. Os detectores das faixas 1 a 5 e 7 possuem um campo de visada instantâneo (IFOV) de 42,5 microrradianos, o que corresponde a uma área de 900 metros quadrados no terreno (30 x 30 metros), em função da altitude em que se encontra o satélite. O IFOV dos detectores da banda 6 é igual a 170 microrradianos (120 x 120 metros no terreno) (USGS, 1984).



Segundo Goel (1988), a radiação que incide num dossel é espalhada e refletida, tendo sua direção e composição espectral alteradas de maneira complexa pela vegetação. Para o propósito do sensoriamento remoto da vegetação, parte da radiação alterada e refletida é interceptada e medida por um sensor. Considerando-se que as causas das alterações podem ser devidas aos parâmetros da vegetação ou aos fatores ambientais, pode-se relacionar as medidas de radiação interceptada por um sensor com algumas propriedades da vegetação.

TABELA 2.1 - BANDAS ESPECTRAIS DO SENSOR "THEMATIC MAPPER"

BANDA	FAIXA ESPECTRAL ( $\mu\text{m}$ )
1	0,45 - 0,52
2	0,52 - 0,60
3	0,63 - 0,69
4	0,76 - 0,90
5	1,55 - 1,75
6	10,4 - 12,5
7	2,08 - 2,35

FONTE: Modificada de USGS (1984), p.4-1.

Goel (1988) descreveu o sensoriamento remoto da vegetação como um sistema constituído por cinco subsistemas:

- a) a fonte de radiação, normalmente o sol, definida por um conjunto de propriedades ou parâmetros que incluem a intensidade radiante espectral e sua localização;

- b) a atmosfera, caracterizada por um conjunto de propriedades ou parâmetros que incluem as propriedades seletivas de absorção e espalhamento dos aerossóis, vapor d'água e ozônio;
- c) o dossel da vegetação, caracterizado por um conjunto de propriedades ou parâmetros. Esses incluem os parâmetros ópticos (reflectância e transmitância) e estruturais (formas geométricas e posições) dos componentes vegetais (folhas, galhos, troncos, etc.), a geometria de plantio e parâmetros ambientais como umidade, velocidade do vento e precipitação;
- d) o solo ou a superfície, caracterizado por um conjunto de propriedades ou parâmetros como reflectância e absortância, rugosidade da superfície, textura e umidade; e
- e) o detector, caracterizado por um conjunto de propriedades ou parâmetros, que indica sua resolução espectral, abertura, calibração e posição.

O esquema do sistema de sensoriamento remoto pode ser visto na Figura 2.1.

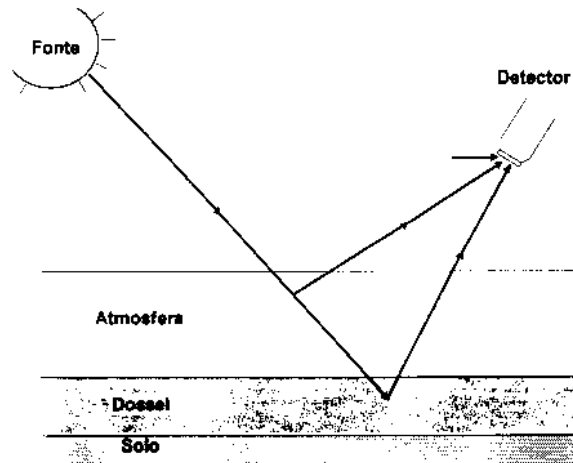


Fig. 2.1 - Esquema do sistema de sensoriamento remoto.

FONTE: Modificada de Goel (1988), p. 3.

Segundo Gausman (1985), a resposta das folhas domina a maior parte dos sinais que atingem os sensores. Embora outras partes das plantas, como troncos, galhos e flores influam no sinal, suas contribuições são menores.

O espectro de reflectância típico de uma folha verde é apresentado na Figura 2.2.

A curva mostra a reflectância percentual da energia incidente em função do comprimento de onda, num intervalo espectral de 0,4 a 2,7  $\mu\text{m}$ . Na faixa visível do espectro, entre 0,4 a 0,7  $\mu\text{m}$ , a reflectância da folha é baixa, apresentando valores em torno de 10%, e um pico na região do verde, aproximadamente em 0,55  $\mu\text{m}$ . A reflectância eleva-se a cerca de 50% no infravermelho, entre 0,7 a 1,3  $\mu\text{m}$  e, a partir desse ponto, decai até apresentar valores muito baixos no comprimento de onda igual a 2,7  $\mu\text{m}$ .

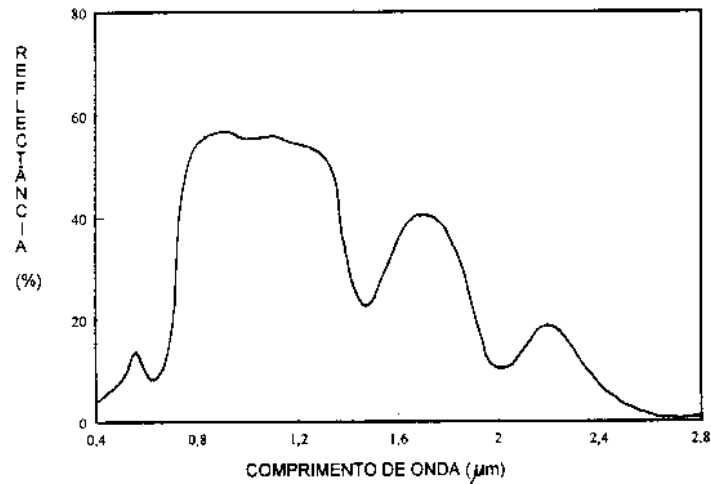


Fig. 2.2 - Espectro de reflectância da folha do tabaco.

FONTE: Knipling (1970), p. 155.

Conhecendo-se as características espectrais de dosséis e sua reflectância nas regiões do espectro referentes às bandas do sensor TM pode-se realizar a interpretação dos dados das imagens desse sensor e o posterior mapeamento das feições de interesse.

A interpretação visual é ainda hoje um método bastante utilizado para o mapeamento da vegetação de extensas áreas a partir de produtos de sensoriamento remoto. Como exemplos, tem-se: o projeto PRODES, que monitora o desflorestamento na Amazônia (Tardin et al., 1980; Batista et al., 1994), o projeto Olho Verde, que monitora a vegetação no Estado de São Paulo (Hahn et al., 1990; Covre et al., 1992) e o trabalho desenvolvido pela Fundação SOS Mata Atlântica e INPE (1993), que monitora a vegetação da Mata Atlântica.

Segundo Hernandez Filho (1988), o processo de interpretação de dados de produtos de sensoriamento

remoto consiste basicamente na identificação de alvos e na determinação de seus significados.

Os procedimentos fotointerpretativos, sejam de imagens orbitais ou de fotografias aéreas, envolvem três fases distintas: a fotoleitura, a fotoanálise e a fotointerpretação. A fotoleitura consiste na identificação das feições ou objetos de interesse sobre as imagens. A fotoanálise consiste no estudo do relacionamento entre as feições ou objetos das imagens e no ordenamento das partes que apresentam propriedades comuns. A fotointerpretação é o processo que utiliza o raciocínio lógico, dedutivo ou indutivo, para compreender os princípios e meios que deram origem à associação e ao ordenamento das feições ou objetos identificados (Veneziani, 1988).

Os procedimentos da interpretação visual de imagens orbitais levam em consideração os elementos clássicos de fotointerpretação convencional, os quais o analista procura utilizar para a caracterização dos alvos (Hernandez Filho, 1988). Segundo Garcia (1982) e Estes et al. (1983), esses elementos são: tonalidade/cor, tamanho, forma, textura fotográfica, padrão, sombra, localização e associação de evidências.

Os tons de cinza de uma imagem são provenientes do registro de variações da radiação eletromagnética do terreno, detectadas por um sensor.

A atribuição de cor às imagens acrescenta uma nova dimensão ao processo fotointerpretativo, uma vez

que o olho humano é capaz de distinguir um maior número de combinações entre cores do que de tons de cinza (Garcia, 1982).

Outro elemento da fotointerpretação, o tamanho dos objetos, auxilia na identificação e medição de objetos de interesse, porém varia em função da escala da imagem.

A forma está relacionada à geometria do objeto analisado. Ela pode ser o único elemento utilizado na caracterização de objetos quando esses apresentam formas específicas e características na posição de visada do sensor. Segundo Pereira et al. (1989), existem basicamente duas classes de formas, as regulares e as irregulares. As regulares são indicadoras de aspectos artificiais, implantados pelo homem, como reflorestamentos e áreas agrícolas. As formas irregulares, por outro lado, são indicadoras de aspectos naturais e ocorrem na ausência de formas geométricas definidas.

A textura fotográfica é dada pela repetição de tons de cinza de grupos de objetos que são muito pequenos para serem vistos individualmente (Garcia, 1982). Em imagens orbitais, a textura permite diferenciar, em alguns casos, áreas de reflorestamento de áreas de floresta natural. As áreas de reflorestamento, por possuírem uma vegetação homogênea no tocante ao espaçamento e altura dos indivíduos, geralmente apresentam textura mais lisa que as áreas de floresta natural (Novo, 1989).

O padrão refere-se ao arranjo espacial dos objetos. A repetição de certas formas estabelece padrões que ajudam a identificar objetos. Por exemplo, em certas áreas de cultura, o traçado retilíneo e paralelo de canais de irrigação constitui um padrão facilmente identificável (Pereira et al., 1989).

A sombra corresponde à ausência ou ao pequeno fluxo de radiação eletromagnética, proveniente do objeto, que atingiu o detector. Ela pode realçar as formas dos objetos, mas também pode ocultar detalhes, dificultando a fotointerpretação. Reflorestamentos ou áreas agrícolas compostas por plantas altas, como cana-de-açúcar, podem ser identificados, em alguns casos, pela sombra ao longo das bordas dos campos de plantio (Pereira et al., 1989).

A localização é dada pela posição dos objetos no relevo e auxilia tanto na identificação desses como no conhecimento da região (Garcia, 1982). Alguns tipos de vegetação, por exemplo, são característicos de alguns sítios geográficos como pântanos e bancos de areia. A partir da identificação direta do sítio geográfico nas imagens é possível deduzir o tipo de vegetação provável na área (Pereira et al., 1989).

A associação de evidências, denominada "relação de aspectos" por Pereira et al. (1989), consiste na correlação de fenômenos físicos, naturais e antrópicos. Por exemplo, pode-se inferir as características de um solo a partir do tipo de vegetação que nele se encontra.

Hernandez Filho (1988) enfatizou que os dados cartográficos e bibliográficos, e o conhecimento do analista sobre o tema abordado e a área de estudo influem no grau de detalhes e na confiabilidade dos resultados. O autor ressaltou que uma base cartográfica é fundamental para a identificação de referências geográficas que podem auxiliar no trabalho de campo e são apresentadas na edição dos resultados.

### 2.3 - SISTEMA DE INFORMAÇÕES GEOGRÁFICAS COMO FERRAMENTA PARA A CRIAÇÃO DE UMA BASE DE DADOS GEORREFERENCIADOS

A característica fundamental de um sistema de geoprocessamento ou sistema de informações geográficas (SIG) é sua faculdade de armazenar, recuperar e analisar informações georreferenciadas num ambiente computacional.

Segundo Câmara (1993), as principais características dos sistemas de informações geográficas são referentes à possibilidade de:

- integrar, numa única base de dados, as informações espaciais provenientes de dados cartográficos, dados de censo e de cadastro urbano e rural, imagens de satélite, redes e Modelos Numéricos de Terreno;
- combinar as várias informações através de algoritmos de manipulação, para gerar mapeamentos derivados; e



- consultar, recuperar, visualizar e desenhar o conteúdo da base de dados geográficos.

Os dados geográficos descrevem objetos do mundo real em termos de sua posição com respeito a um sistema de coordenadas conhecido, seus atributos e suas interrelações espaciais (relações topológicas) (Burrough, 1986). Desse modo, segundo Felgueiras e Câmara (1993), um SIG lida tanto com dados espaciais de localização (atributos gráficos), quanto com dados não espaciais (atributos não gráficos de entidades geográficas).

Os dados geográficos são referenciados à sua localização na superfície da Terra pelo uso de um sistema de coordenadas que pode ser local, no caso do estudo de uma área limitada, ou uma grade aceita nacional ou internacionalmente, como o Sistema de Coordenadas Universal Transversa de Mercator (UTM) (Burrough, 1986).

Os atributos não espaciais associados referem-se às propriedades de uma entidade espacial que são manuseadas num SIG. Uma mesma entidade espacial pode ter vários atributos associados. Os atributos podem ser armazenados e processados separadamente dos dados espaciais. Para se estabelecer uma ligação entre os dados espaciais e não espaciais utiliza-se um identificador, o qual é atribuído a cada tipo de dado (Burrough, 1986).

Segundo Burrough (1986), a entrada dos dados é a operação básica para a formação de um banco de dados.

Essa operação envolve a codificação e a introdução desses no banco, sendo descrita em três estágios:

- a entrada dos dados espaciais,
- a entrada dos atributos não espaciais associados, e
- o estabelecimento de uma ligação entre os dados espaciais e não espaciais.

Dentre os métodos de entrada de dados no SIG, tem-se a digitalização, que pode ser:

- manual, realizada através de mesa digitalizadora, ou
- automática, realizada através de "scanner", (Scarim e Teixeira, 1994).

Os erros que podem surgir da entrada de dados podem ser reunidos em 6 grupos:

- a) dados espaciais incompletos ou duplicados,
- b) dados espaciais com localização errada,
- c) dados espaciais com escala errada,
- d) dados espaciais distorcidos,
- e) dados espaciais ligados a dados não espaciais errados, e
- f) dados não espaciais incompletos (Burrough, 1986).

No caso da digitalização, esses erros podem ocorrer por descuido do operador durante a operação.

Segundo Aronoff (1989), o processo de digitalização consome tempo e é propenso a erros, sendo que sua eficiência depende da qualidade do programa e da habilidade do operador.

Segundo Burrough (1986), o erro citado no item 4 acima pode ainda ocorrer devido a distorções geométricas nos dados originais, por exemplo, as distorções geométricas que ocorrem no processo de aquisição das imagens orbitais ou aquelas devidas à instabilidade dos materiais utilizados na confecção dos mapas.

#### 2.3.1 - DIGITALIZAÇÃO MANUAL

A digitalização manual se baseia no uso de um digitalizador que codifica as coordenadas X e Y dos pontos, linhas, áreas ou grades de células desenhados. Um digitalizador é um equipamento periférico eletrônico ou eletromagnético, que consiste em uma placa ou mesa sobre a qual o mapa ou o documento pode ser colocado. Nesse caso, as coordenadas de cada ponto da superfície do digitalizador são enviadas ao computador por um "mouse" ou um caneta magnética (Burrough, 1986).

#### 2.3.2 - DIGITALIZAÇÃO AUTOMÁTICA

A digitalização automática ou rasterização é a discretização do mapa em unidades retangulares homogêneas ou pixels através do uso do "scanner" (Scarim e Teixeira, 1994).

Para seu uso, é feito um novo desenho do mapa eliminando detalhes indesejáveis, como símbolos e textos, ou apenas realçando informações que serão vetorizadas (Scarim e Teixeira, 1994). Em seguida é feita a rasterização do mapa, e o produto resultante, no formato "raster", é vetorizado e georreferenciado para uso em um SIG (Imagem, 1995).

Segundo Scarim e Teixeira (1994), os métodos de vetorização podem ser: manual, semi-automático ou automático.

O método manual consiste basicamente em seguir cada feição "raster" com o cursor e escolher os pontos que a modelem melhor na estrutura vetorial.

A vetorização semi-automática, também conhecida por vetorização supervisionada, conjuga o método automático com o manual de maneira interativa com o operador. O operador identifica manualmente o ponto inicial de uma determinada linha "raster" a ser vetorizada. Em seguida, a linha vai sendo rastreada e vetorizada automaticamente pelo sistema através de técnicas de processamento digital de imagens. Quando um obstáculo é encontrado (como um cruzamento de linha ou um texto sobre a linha) o processo é automaticamente interrompido. Nesses casos, o operador precisa intervir para reposicionar o cursor sobre a linha.

A vetorização automática usa técnicas de processamento digital de imagens para rastrear pixels

acesos (pretos) de uma imagem "raster" binária e convertê-los para a estrutura vetorial. O processamento é feito segundo regras preestabelecidas, sem a ajuda do operador. A principal vantagem do método é a redução do tempo gasto para a vetorização.

Segundo Scarim e Teixeira (1994), o tempo gasto na vetorização automática pode variar entre 20 e 25% do total gasto na digitalização manual.

#### 2.4 - EXATIDÃO DE CLASSIFICAÇÃO DOS DADOS DE SENSORIAMENTO REMOTO

O emprego de diferentes técnicas para o reconhecimento e a delimitação de áreas de cobertura vegetal a partir de dados de sensoriamento remoto pode apresentar diferentes resultados. Muitas vezes a avaliação dessas técnicas é subjetiva e a qualidade dos produtos gerados é desconhecida.

Face à necessidade de estabelecer critérios de qualidade para os produtos gerados a partir do uso de técnicas de sensoriamento remoto, alguns autores sugeriram níveis mínimos para a aceitação desses produtos. Anderson (1971), por exemplo, propôs que a exatidão dos produtos esteja compreendida entre 85 e 90%, enquanto Aronoff (1982) propôs a associação de um nível mínimo de exatidão a um nível máximo de confiança (por exemplo, 85% de exatidão a um nível de confiança de 95%). Watson et al. (1992) sugeriram um nível mínimo de exatidão igual a 95%.

Diversas técnicas para gerar uma estimativa da exatidão de classificação têm sido propostas. É importante que os pesquisadores e usuários dos dados de sensoriamento remoto conheçam essas técnicas, bem como os fatores que influem nos valores de exatidão.

Skidmore e Turner (1992) citaram dois tipos de erros nos mapeamentos temáticos: o de atributos e o de localização. O primeiro ocorre quando um atributo ou nome de categoria é designado erroneamente. O segundo é um erro associado à localização das feições como pontos, linhas ou polígonos. Na realidade os dois erros ocorrem conjuntamente e contribuem para o erro global, sendo difícil isolá-los.

Nem sempre esses dois tipos de erros são diferenciados nas técnicas utilizadas para estimar a exatidão de classificação. Entretanto, métodos distintos para a análise de cada um deles são comuns na literatura como, por exemplo, os propostos por Hord e Brooner (1976) e Skidmore e Turner (1992).

As técnicas abordadas no presente estudo baseiam-se principalmente nos erros de atributos, sendo empregadas para a estimativa da exatidão de classificação de um produto ou para a comparação entre diferentes produtos.

Os itens 2.4.1.1 e 2.4.1.2 do presente trabalho apresentam técnicas descritivas e analíticas para a estimativa da exatidão de classificação. Dentre as técnicas analíticas, incluem-se algumas específicas para a

comparação entre diferentes produtos da classificação (item 2.4.1.2.1). Nos itens 2.4.2 e 2.4.3 são descritos tipos de amostragem e discutidos tamanhos de amostra, respectivamente, bem como os critérios para a escolha do esquema amostral. Por último, o item 2.4.4, apresenta algumas considerações sobre os dados usados como referência nas análises de exatidão.

#### 2.4.1 - TÉCNICAS PARA A ESTIMATIVA DA EXATIDÃO DE CLASSIFICAÇÃO

Segundo Congalton (1991), até o início da década de 80, a exatidão de classificação era simplesmente expressa como um número, empregando-se, muitas vezes, a exatidão denominada genérica ("non-site-specific accuracy").

Esse tipo de exatidão, conforme Mead e Szajgin (1982), é calculado através da razão entre a área total de cada categoria classificada e a área da categoria correspondente, determinada por medidas de campo ou resultantes da fotointerpretação. Essa medida se baseia somente na área total, desconsiderando a sua localização. Congalton (1991) ressalta que a exatidão genérica pode resultar em valores elevados e espúrios de exatidão, principalmente quando ocorre, no balanço geral, a compensação de erros entre categorias.

Outro problema relatado pelo autor é o uso de amostras de treinamento para testar o resultado da classificação. Nesses casos a exatidão também é

superestimada, pois ela não é calculada a partir de uma amostra independente.

Uma vez identificados esses problemas, foram sugeridos novos métodos objetivando a estimativa de uma exatidão específica ("site-specific accuracy"), a partir de amostras independentes daquelas usadas para treinamento.

Nesses métodos considera-se a natureza espacial dos dados, podendo-se utilizar polígonos, grades ou pontos de referência para a localização dos dois conjuntos de unidades amostrais que são comparadas: as de referência (muitas vezes denominadas "verdade terrestre") e as de classificação (Mead e Szajgin, 1982).

Segundo Congalton (1991), as técnicas usadas atualmente para a estimativa da exatidão de classificação podem ser divididas em dois tipos: descritivas e analíticas.

#### 2.4.1.1 - TÉCNICAS DESCRITIVAS

A forma mais utilizada para representar a exatidão de classificação de um mapa é através da percentagem da área do mapa corretamente classificada, quando comparada aos dados de referência ou "verdade terrestre" (Story e Congalton, 1986). O meio mais utilizado para expressar o resultado da comparação entre os dados do mapa e da referência é a matriz de erro, cujo uso é recomendado por diversos pesquisadores (Congalton, 1991).



Uma matriz de erro, também denominada matriz de confusão ou tabela de contingência, é uma matriz quadrada de números que expressam a quantidade de unidades amostrais - isto é, pixels, conjunto de pixels ou polígonos - associada a uma dada categoria durante o processo de classificação e à categoria real a que elas pertencem (Story e Congalton, 1986; Congalton, 1991). Um exemplo de matriz de erro é dado na Tabela 2.2.

TABELA 2.2 - EXEMPLO DE MATRIZ DE ERRO

DADOS CLASSIFICADOS	DADOS DE REFERÊNCIA				TOTAL DA LINHA
	A	B	C	D	
A	65	4	22	24	115
B	6	81	5	8	100
C	0	11	85	19	115
D	4	7	3	90	104
TOTAL DA COLUNA	75	103	115	141	434

FONTE: Modificada de Congalton (1991), p. 36.

A matriz de erro pode ser usada como base para uma série de técnicas estatísticas descritivas e analíticas para estimar a exatidão de classificação. A estatística descritiva mais simples é a exatidão global, dada pela razão entre o número de pixels classificados corretamente (equivalente à soma dos elementos da diagonal principal da matriz) e o número total de pixels considerado (Congalton, 1991).

A matriz de erro permite ainda avaliar o desempenho da classificação para cada categoria investigada. Ela favorece a visualização dos resultados da

classificação e expressa a relação entre os dois tipos de erro associados ao sistema de classificação: os de inclusão e os de omissão (Card, 1982).

Os erros de inclusão correspondem ao número de pixels classificados em uma dada categoria, porém não pertencentes a essa categoria. Os erros de omissão correspondem ao número de pixels pertencentes a uma dada categoria, porém classificados como pertencentes às demais categorias.

A exatidão por categoria pode ser calculada de duas formas. Uma delas é através da razão entre o número de unidades amostrais corretamente classificadas e o número de unidades que pertencem realmente àquela categoria, sendo uma medida associada aos erros de omissão. A outra corresponde à razão entre o número de unidades amostrais corretamente classificadas e o número total de unidades que foram classificadas naquela categoria, sendo associada aos erros de inclusão. A primeira recebe a denominação de "exatidão do produtor" e a segunda "exatidão do consumidor" (Story e Congalton, 1986; Congalton, 1991).

A exatidão de classificação de um mapa, quando calculada a partir do número de pixels, conjunto de pixels ou segmentos, não considera a área de cada categoria isoladamente. Assim, para se estimar a exatidão do mapa é necessário introduzir um peso à cada categoria em função da sua área (Fitzpatrick-Lins, 1981; Rosenfield et al., 1982; Kalensky, 1985). Esse peso equivale à razão entre a área ocupada por cada categoria e a área total do mapa. Prisley

e Smith (1987) propuseram uma ponderação em função da importância que as categorias têm, face ao objetivo do mapeamento.

#### 2.4.1.2 - TÉCNICAS ANALÍTICAS

Segundo Prisley e Smith (1987), a hipótese de que a distribuição dos erros na matriz é representativa dos tipos de erros existentes em toda a área classificada permite que sejam feitas inferências sobre a probabilidade dos pixels serem atribuídos às categorias. Portanto, em adição às técnicas descritivas, uma matriz de erro pode originar uma série de técnicas estatísticas analíticas (Congalton, 1991).

Hord e Brooner (1976) afirmaram que qualquer estimativa de exatidão baseada em amostragem deve estar associada a intervalos de confiança, os quais dependem do número de unidades amostrais selecionadas no mapa. Os autores propuseram como valor de exatidão da classificação o limite inferior de um intervalo de confiança de 95%. Para esse cálculo, os autores basearam-se na distribuição binomial dos erros de classificação, aproximando-a a uma distribuição normal.

Rosenfield et al. (1982) salientaram que o número de erros ou acertos de classificação apresentam uma distribuição binomial. Face a isto, propuseram que qualquer estimativa de exatidão tivesse como base essa distribuição. No caso de uma aproximação normal à distribuição binomial,

para grandes amostras, eles sugeriram a utilização do fator de correção da continuidade.

Fleiss (1981) citou que a aproximação normal à distribuição binomial pode ser feita quando:

$$np \geq 5 \text{ e } nq \geq 5$$

onde:

$n$  = tamanho da amostra;

$p$  = proporção da população que apresenta determinada característica; e

$$q = 1 - p \text{ } ^1 . \quad (2.1)$$

Nesse caso, o intervalo de confiança de  $(1-\alpha)100\%$  para  $p$  é dado por:

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} - \frac{1}{2n} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} + \frac{1}{2n} \quad (2.2)$$

onde:

$\hat{p}$  = proporção de unidades amostrais que apresentam uma determinada característica;

---

<sup>1</sup> As notações nas equações apresentadas no presente trabalho não seguem, necessariamente, as apresentadas pelos autores. Algumas alterações foram feitas de forma a padronizá-las com vistas a facilitar a compreensão do texto.

$z_{\alpha/2}$  = valor na função de distribuição acumulada da normal padrão relativo a uma área igual a  $1-\alpha/2$ ; e

$\frac{1}{2n}$  = fator de correção da continuidade para a aproximação da normal à binomial.

Baseada nesses pressupostos, Fitzpatrick-Lins (1981) propôs, como estimativa da verdadeira exatidão, o limite inferior do intervalo de confiança gerado a partir da exatidão amostral. Esse limite, ao nível de significância  $\alpha$ , é dado por:

$$p_L = \hat{p} - \left\{ z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{1}{2n}} \right\} \quad (2.3)$$

onde:

$p_L$  = limite inferior do intervalo de confiança de  $(1-\alpha)100\%$  construído para a exatidão do mapa;

$\hat{p}$  = exatidão amostral, calculada através da razão entre o número de unidades amostrais classificadas corretamente na amostra ( $r$ ) e o número total de unidades amostradas ( $n$ ), isto é:

$$\hat{p} = \frac{r}{n}. \quad (2.4)$$

Rosenfield et al. (1982) e Fitzpatrick-Lins (1981) salientaram que o uso da Equação (2.3) só se aplica quando a amostragem é representativa de cada categoria no mapa, em função de sua área. Caso contrário, os autores

sugeriram a ponderação da área dessas categorias através da seguinte equação:

$$p_L = \hat{p}_p - \left\{ z_{\alpha/2} \sqrt{\sigma_{\hat{p}}^2 + \frac{1}{2n}} \right\} \quad (2.5)$$

onde:

$\hat{p}_p$  = exatidão amostral ponderada, dada por:

$$\hat{p}_p = \sum_{h=1}^f w_h \hat{p}_h \quad (2.6)$$

onde:

$f$  = número de categorias;

$w_h$  = proporção da área ocupada pela categoria  $h$  no mapa;

$\hat{p}_h$  = exatidão amostral para a categoria  $h$ , calculada através da razão entre o número de unidades amostrais classificadas corretamente na categoria  $h$  ( $r_h$ ), e o número de unidades amostradas nessa categoria ( $n_h$ ); e

$\sigma_{\hat{p}}^2$  = variância da exatidão amostral, dada por:

$$\sigma_{\hat{p}}^2 = \sum_{h=1}^f \frac{w_h^2 p_h (1 - p_h)}{n_h} \quad (2.7)$$

e estimada por:

$$\hat{\sigma}_{\hat{p}}^2 = \sum_{h=1}^f \frac{w_h^2 \hat{p}_h (1 - \hat{p}_h)}{n_h} \quad (2.8)$$

Alguns autores sugeriram a aplicação do teste de hipóteses para decidir se um mapa, ou uma categoria, tem uma exatidão mínima desejada,  $p_0$  (Ginevan, 1979; Rosenfield et al., 1982; Aronoff, 1982).

Rosenfield et al. (1982) descreveram um método baseado no teste de hipóteses e na distribuição binomial acumulada para concluir se uma dada categoria apresenta o valor de exatidão desejado para o número de unidades amostrais utilizado. Para aceitar a hipótese de que a categoria possui a exatidão mínima desejada ( $p_0$ ), a um nível de significância  $\alpha$ , o número de erros de classificação deve ser menor que um valor crítico ( $c$ ), que é menor que o número mínimo de unidades amostrais que devem ser corretamente classificadas na amostra ( $r_L$ ).

As hipóteses nula ( $H_0$ ) e alternativa ( $H_1$ ) do teste são dadas por:

$$H_0: p \geq p_0$$

$$H_1: p < p_0$$

Seja  $R_L$  o número de pontos corretamente classificados na amostra. A um nível de significância  $\alpha$ ,  $c_1$  é definido de forma que a probabilidade:

$$P(R_L \leq c_1 | p = p_0) \leq \alpha \quad (2.9)$$

seja satisfeita, ou seja:

$$\sum_{r_L=0}^{c_1} \binom{n}{r_L} p^{r_L} (1-p)^{n-r_L} \leq \alpha . \quad (2.10)$$

$H_0$  é rejeitada ao nível de significância  $\alpha$ , se  $R_L \leq c_1$ . Logo, a probabilidade de rejeitar a hipótese nula,  $H_0$ , quando ela é verdadeira, não é maior que  $\alpha$ .

Segundo os autores, há ainda outra maneira de se aplicar o teste:

$$H_0: p < p_0$$

$$H_1: p \geq p_0$$

Nesse caso, o valor crítico  $c_2$  se refere ao menor valor inteiro tal que a probabilidade:

$$P(R_L \geq c_2 | p = p_0) \leq \alpha \quad (2.11)$$

seja satisfeita, ou seja:

$$\sum_{r_L=c_2}^n \binom{n}{r_L} p^{r_L} (1-p)^{n-r_L} \leq \alpha . \quad (2.12)$$

Rejeita-se  $H_0$  se  $R_L \geq c_2$ . Assim, a probabilidade de aceitar  $H_1$  quando  $H_0$  é verdadeira, não é maior que  $\alpha$ .

Essas duas formulações do teste de hipóteses resultam em dois valores críticos ( $c$ ) para a aceitação da exatidão mínima desejada para uma dada categoria, quando o nível de significância adotado é igual nos dois testes.

Aronoff (1982) apresentou o teste identificando as duas probabilidades de erro citadas por Rosenfield et al. (1982) em uma única formulação. O autor associou dois tipos de risco à probabilidade de ocorrência



dos dois erros possíveis na condução de um teste de hipóteses: o "risco do consumidor" e o "risco do produtor". Aplicando-se o teste para todo o mapa, o primeiro risco corresponde à probabilidade de aceitar um mapa com exatidão menor que a mínima desejada ( $p_0$ ), e o segundo à probabilidade de rejeitar um mapa de exatidão  $p_U$ , sendo  $p_U > p_0$ .

Nesse caso o teste também se baseia na distribuição binomial acumulada. As hipóteses são dadas por:

$H_0: p < p_0$  (o mapa apresenta exatidão menor que a mínima desejada)

$H_1: p \geq p_0$  (o mapa apresenta exatidão maior ou igual à mínima desejada).

O autor ilustrou a teoria estatística do teste de exatidão de classificação a partir de um exemplo hipotético descrito a seguir.

Considere um mapa de exatidão desconhecida, de onde se seleciona aleatoriamente uma amostra de  $n$  unidades amostrais. Conhecendo-se a verdadeira classe a que elas pertencem, pode-se calcular a proporção de unidades amostrais corretamente classificadas. Se esse processo for repetido várias vezes, pode-se gerar a distribuição amostral e apresentá-la num gráfico cujas coordenadas indicam a proporção de unidades corretamente classificadas ou a exatidão da amostra, e a freqüência ou a probabilidade de se obter esse valor. Um modelo estatístico de

distribuição binomial é a representação adequada da distribuição das proporções amostrais, usando um tamanho de amostra  $n$  em um mapa de exatidão  $p_0$ . Isto é ilustrado na Figura 2.3, onde a distribuição normal foi usada para facilitar a visualização.

Para um teste com nível de confiança de  $(1-\alpha)100\%$ ,  $p_0$  é a exatidão mínima desejada e  $H_0$  é aceita se a exatidão amostral ( $\hat{p}$ ) for menor que o valor  $p_T$ . O valor  $p_T$  é calculado de forma que a probabilidade de se obter um valor amostral tão elevado ou maior que  $p_T$  é  $\alpha$ , supondo que o mapa tem uma exatidão  $p_0$  (Figura 2.3). Se a exatidão amostral ( $\hat{p}$ ) for maior ou igual a  $p_T$ , então a exatidão do mapa é considerada maior ou igual a  $p_0$ ,  $H_0$  é rejeitada e o mapa é aceito. O risco do consumidor é dado por  $\alpha$ , que corresponde à área identificada por  $A$  na Figura 2.3.

Para desenvolver o teste deve-se estabelecer previamente a exatidão mínima desejada ( $p_0$ ), o tamanho da amostra ( $n$ ) e o risco do consumidor ( $A$ ). O número máximo de erros de classificação permitido ( $x$ ) é calculado de forma que a probabilidade acumulada de se ter  $x$  ou um número menor de erros de classificação é menor ou igual a  $A$ . Dessa forma, encontra-se  $x$ , tal que:

$$A \geq \sum_{y=0}^x \binom{n}{y} p_0^{n-y} (1-p_0)^y \quad (2.13)$$

onde:

$y$  = número de erros de classificação.

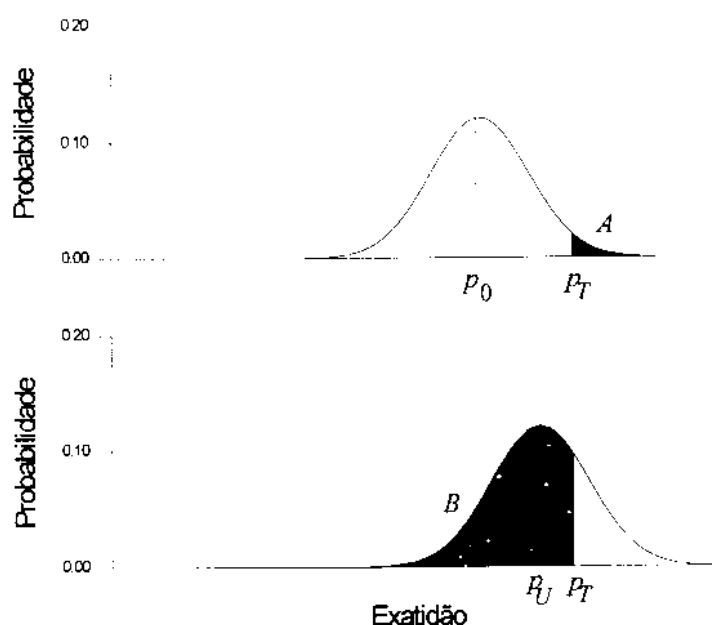


Fig. 2.3 - Distribuição amostral de unidades corretamente classificadas, identificando o risco do consumidor (A) e o risco do produtor (B).

FONTE: Modificado de Aronoff (1982), p. 1300.

Tomando-se a mesma distribuição amostral das mesmas  $n$  unidades amostrais selecionadas aleatoriamente de um mapa que possui exatidão  $p_U$  (sendo  $p_U > p_0$ ), o risco do produtor é a probabilidade de que a proporção de unidades amostrais corretamente classificadas seja menor que  $p_T$ . Esse risco corresponde à área identificada por B na Figura 2.3 e é calculado por:

$$B = \sum_{y=r+1}^n \binom{n}{y} p_U^{n-y} (1-p_U)^y \quad (\text{Aronoff, 1985}). \quad (2.14)$$

O risco do produtor pode ser reduzido através do aumento do risco do consumidor, isto é, da redução de  $p_T$ , ou do aumento do tamanho da amostra.

A Figura 2.4 ilustra o aumento do tamanho da amostra, o qual resulta na redução da variância amostral. Os valores de  $p_0$  e  $p_U$  são iguais aos da Figura 2.3.

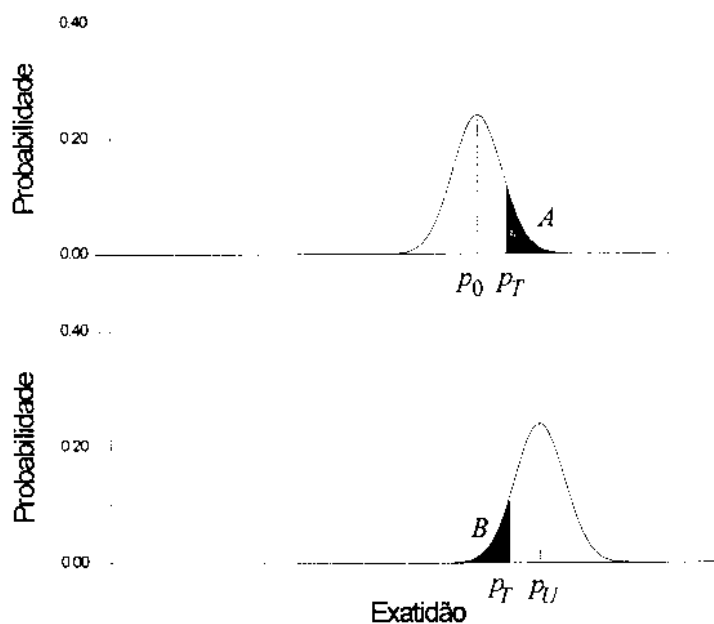


Fig. 2.4 - Distribuição amostral de unidades corretamente classificadas, identificando o risco do consumidor (A) e o risco do produtor (B), para uma amostra grande.

FONTE: Modificado de Aronoff (1982), p. 1301.

A conclusão do teste de hipóteses consiste em afirmar se o mapa possui ou não a exatidão mínima desejada a um nível de significância  $\alpha$ . Porém, um mapa que falha num teste para 85% de exatidão pode ser adequado para um usuário que requer 80%. Baseado nisso, Aronoff (1985)

propôs o cálculo do valor mínimo de exatidão, a ser utilizado como um índice. Selecionado o risco do consumidor ( $A$ ) obtém-se o número de erros de classificação ( $Y$ ) em uma amostra de tamanho  $n$  e encontra-se o valor mínimo de exatidão ( $p'_L$ ) através de:

$$A = \sum_{y=0}^x \binom{n}{y} p'^{n-y} (1-p')^y . \quad (2.15)$$

Segundo Aronoff (1982), esse método fornece uma estimativa geral da exatidão mas não diferencia os erros de omissão e inclusão. Essa distinção é importante quando se quer estimar a exatidão de uma ou mais categorias. Para a análise de exatidão das categorias, separadamente, o autor sugeriu o uso de matrizes de erro, da análise de variância ou de técnicas de análise multivariada.

Congalton et al. (1983) ressaltaram que para se aplicar a análise de variância, os dados devem ser normalmente distribuídos. Os dados de classificação são discretos e geralmente apresentam distribuição multinomial, ou seja, são binomialmente distribuídos em cada uma das várias categorias. Nesse caso, para se aplicar a análise de variância seria necessário transformar os dados. Outro pressuposto da análise de variância, o de que as categorias são independentes, dificilmente se aplica a dados de sensoriamento remoto.

Congalton et al. (1983) sugeriram o uso de técnicas analíticas baseadas na análise multivariada

discreta, afirmando que essas não requerem a transformação dos dados e não pressupõem a independência desses. A técnica sugerida por esses autores, para dados de sensoriamento remoto, denomina-se Kappa e foi proposta em Cohen (1960).

Segundo Cohen (1960), para o uso do coeficiente Kappa, supõe-se as seguintes condições:

- que as unidades amostrais sejam independentes<sup>2</sup> ;
- que as categorias sejam independentes, mutuamente exclusivas e exaustivas<sup>3</sup> ; e
- que os intérpretes operem independentemente.

Sob a suposição da normalidade, a independência das unidades amostrais é caracterizada pela ausência de correlação entre elas. Congalton (1988a), ao analisar a correlação entre pixels de uma imagem binária, representando erros e acertos de classificação, observou

---

<sup>2</sup> Seja A um evento (seleção de uma unidade amostral) que ocorre com uma probabilidade  $P(A)$  e B um evento (seleção de outra unidade amostral) que ocorre com probabilidade  $P(B)$ . Se A e B são independentes, ou seja, a ocorrência de um não afeta a ocorrência do outro, então  $P(A \cap B) = P(A) \cdot P(B)$ .

<sup>3</sup> Sejam dois eventos  $E_i$  e  $E_j$ . Se  $E_i \cap E_j = \emptyset$ , ou seja, a interseção entre eles for vazia, então esses eventos são mutuamente exclusivos. Se os eventos  $E_1, \dots, E_k$ , além de mutuamente exclusivos, forem também exaustivos, então  $E_1 \cup E_2 \cup \dots \cup E_k = S$ , ou seja, a união desses eventos corresponde ao espaço amostral (Brownlee, 1965, p. 1-86).

que a correlação diminuiu com o aumento da distância entre os pixels.

É importante notar que a pressuposição da independência das categorias também é requerida para a aplicação do Kappa, eliminando uma das supostas vantagens desse método sobre o da análise de variância, citada por Congalton et al. (1983).

Foody (1988) salientou que o pressuposto da independência espacial entre as categorias é inválido quando a geometria de visada é grande. Entretanto, o autor enfatizou que esse efeito do campo de visada não deveria ser significativo para dados advindos de campos de visada instantâneos (IFOV) pequenos, como é o caso do sensor "Multispectral Scanner" (MSS), cujo IFOV, em função da altitude em que se encontra o satélite, corresponde a uma área de 6.400 metros quadrados no terreno (80 x 80 metros).

Com base nessas observações, pressupõe-se a independência das categorias para os dados do sensor TM-LANDSAT, que apresenta um IFOV ainda menor que o MSS.

A outra vantagem apontada por Congalton et al. (1983) para uso do Kappa, qual seja, a ausência da necessidade da transformação dos dados, é confirmada por Brennan e Prediger (1981). Segundo os autores, nenhum tipo particular de distribuição ou aleatoriedade da amostragem é pressuposto para o uso dessa técnica.

O Kappa é um coeficiente de concordância expresso por:

$$K = \frac{p_a - p_c}{1 - p_c} \quad (2.16)$$

onde:

$p_a$  = proporção de concordância entre avaliadores ou, no caso da exatidão de classificação, entre os dados classificados e os dados de referência, dada por:

$$p_a = \sum p_{ii} \quad (2.17)$$

onde:

$p_{ii}$  = proporção de unidades na linha  $i$  e na coluna  $i$ , em relação ao total de unidades; e

$p_c$  = proporção esperada de unidades, se a concordância ocorrer por puro acaso, ou seja:

$$p_c = \sum p_{i+} p_{+i} \quad (2.18)$$

onde:

$p_{i+}$  e  $p_{+i}$  = proporções marginais da linha  $i$  e da coluna  $i$ , respectivamente

A estimativa do Kappa, a estatística  $\hat{K}$ , também é uma medida de exatidão (Congalton, 1991). No caso da análise de exatidão de classificação, tem-se que  $\hat{K}$  é uma medida da concordância geral calculada para cada matriz de erro baseada na diferença entre a concordância real da classificação e a concordância por puro acaso, expressa por:



$$\hat{K} = \frac{n \sum_{i=1}^l x_{ii} - \sum_{i=1}^l (x_{i+} x_{+i})}{n^2 - \sum_{i=1}^l (x_{i+} x_{+i})} \quad (2.19)$$

onde:

$l$  = dimensão da matriz de erro quadrada;

$x_{ii}$  = número de observações na linha  $i$ , coluna  $i$ ;

$x_{i+}$  e  $x_{+i}$  = totais marginais da linha  $i$  e da coluna  $i$ , respectivamente; e

$n$  = número total de observações.

A concordância real corresponde àquela verificada entre os dados da classificação e os dados de referência, sendo indicada pelos elementos da diagonal da matriz de erro. A concordância por puro acaso é a soma dos produtos dos valores marginais das linhas e colunas correspondentes  $(\sum (x_{i+} x_{+i}))$  (Congalton et al., 1983).  $K$  e  $\hat{K}$  são, portanto, coeficientes de concordância que consideram todos os dados em uma matriz de erro.

Segundo Bishop et al. (1975),  $\hat{K}$  é assintoticamente normalmente distribuído e a estimativa da sua variância  $(\hat{\sigma}^2(\hat{K}))$  pode ser utilizada na construção de intervalos de confiança para  $K$ . A expressão para  $\hat{\sigma}^2(\hat{K})$ , segundo esse autor, é dada por:

$$\hat{\sigma}^2(\hat{K}) = \frac{1}{n} \left\{ \frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2)^2}{(1-\theta_2)^4} \right\} \quad (2.20)$$

onde:

$$\theta_1 = \sum_{i=1}^l \frac{x_{ii}}{n}; \quad (2.21)$$

$$\theta_2 = \sum_{i=1}^l \frac{x_{i+} x_{+i}}{n^2}; \quad (2.22)$$

$$\theta_3 = \sum_{i=1}^l \frac{x_{ii}}{n} \left( \frac{x_{i+}}{n} + \frac{x_{+i}}{n} \right); \quad \text{e} \quad (2.23)$$

$$\theta_4 = \sum_{i=1}^l \sum_{j=1}^l \frac{x_{ij}}{n} \left( \frac{x_{i+}}{n} + \frac{x_{+j}}{n} \right)^2. \quad (2.24)$$

O intervalo de  $(1-\alpha)100\%$  de confiança para  $K$  é dado por:

$$\hat{K} \pm z_{\alpha/2} \sqrt{\hat{\sigma}^2(\hat{K})}. \quad (2.25)$$

O coeficiente Kappa trata igualmente todos os tipos de erro, não sendo consideradas a área ou a importância de cada categoria no valor global de concordância. Para uma ponderação desses erros, Rosenfield e Fitzpatrick-Lins (1986) sugeriram o uso do coeficiente Kappa ponderado.

Bishop et al. (1975) propuseram uma medida de concordância condicional ( $K_i$ ) e sua estimativa ( $\hat{K}_i$ ) para determinar o coeficiente de concordância para cada

categoria do mapa. A estimativa da concordância condicional para cada categoria  $i$  ( $i=1,\dots,l$ ) é dada por:

$$\hat{K}_i = \frac{nx_{ii} - x_{i+}x_{+i}}{nx_{i+} - x_{i+}x_{+i}}. \quad (2.26)$$

Rosenfield e Fitzpatrick-Lins (1986) compararam alguns métodos para a estimativa da exatidão de classificação (Kappa e Kappa condicional; método de Turk, 1979; método de Hellden, 1980; e método de Short, 1982; citados por Rosenfield e Fitzpatrick-Lins, 1986). Eles concluíram o trabalho recomendando o uso dos coeficientes Kappa e Kappa condicional para a avaliação da classificação global e de categorias individuais, respectivamente.

É importante observar que a estimativa do coeficiente Kappa condicional ( $\hat{K}_i$ ), segundo formulação proposta por Bishop et al. (1975), é uma medida necessariamente associada aos erros de omissão ou de inclusão. Isto ocorre em função do que representa o termo  $nx_{i+}$  da Equação (2.26), o qual varia segundo a posição dos dados de classificação e de referência em uma matriz de erro. Quando  $x_{i+}$  representa o número de unidades amostrais que pertencem a uma dada categoria,  $\hat{K}_i$  está associada aos erros de omissão. Quando  $x_{+i}$  representa o número de unidades amostrais classificadas em uma dada categoria,  $\hat{K}_i$  está associada aos erros de inclusão. A relação entre  $\hat{K}_i$  e os erros de omissão e inclusão é semelhante à descrita no item 2.4.1.1 para as exatidões do produtor e do consumidor, respectivamente.

Fleiss (1981) propôs a estimativa do Kappa para cada categoria  $i$  ( $\hat{K}_i^*$ ) partindo do modelo de matriz de erro apresentado na Tabela 2.3.

TABELA 2.3 - DADOS PARA MEDIDA DE CONCORDÂNCIA DE UMA ÚNICA CATEGORIA

AVALIADOR B	AVALIADOR A		TOTAL
	CATEGORIA DE INTERESSE	DEMAIS CATEGORIAS	
CATEGORIA DE INTERESSE	$a$	$b$	$p_1$
DEMAIS CATEGORIAS	$c$	$d$	$q_1$
TOTAL	$p_2$	$q_2$	1

FONTE: Modificada de Fleiss (1981), p. 213.

Na Tabela 2.3,  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $p_1$ ,  $p_2$ ,  $q_1$  e  $q_2$  representam proporções.

Considerando:

$$p_{ai} = a + d; \text{ e} \quad (2.27)$$

$$p_{ci} = (p_1 p_2) + (q_1 q_2) \quad (2.28)$$

onde:

$p_{ai}$  = proporção de concordância entre avaliadores para a categoria  $i$ ; e

$p_{ci}$  = proporção esperada de unidades, se a concordância para a categoria  $i$  ocorrer por puro acaso;

tem-se:

$$\hat{K}_i^* = \frac{p_{ai} - p_{ci}}{1 - p_{ci}} = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1} \quad (2.29)$$

Fleiss (1981) mostrou ainda que a estimativa do coeficiente Kappa ( $\hat{K}$ ), considerando as categorias em conjunto, é dada pela razão entre a soma dos valores de  $p_{ai} - p_{ci}$  e a soma dos valores de  $1 - p_{ci}$  para todas as categorias. Assim, considerando-se  $l$  categorias, tem-se:

$$\hat{K} = \frac{p_a - p_c}{1 - p_c} = \frac{\sum_{i=1}^l (p_{ai} - p_{ci})}{\sum_{i=1}^l (1 - p_{ci})} \quad (2.30)$$

onde:

$p_a$  = proporção de concordância entre avaliadores, dada pela Equação (2.17); e

$p_c$  = proporção esperada de unidades, se a concordância ocorrer por puro acaso, dada pela Equação (2.18).

Embora o coeficiente Kappa tenha sido proposto por diversos autores e vem sendo cada vez mais utilizado para análises de exatidão de mapeamento, não há fundamentação teórica para se indicar quais os níveis mínimos desse coeficiente para a aceitação de produtos gerados a partir do uso de técnicas de sensoriamento remoto. Alguns autores (Caldeira, 1994; Damião, 1994; Fitzgerald e Lees, 1994) basearam-se em uma tabela proposta por Landis e Koch (1977) para avaliação de valores obtidos para Kappa. Essa tabela apresenta intervalos de Kappa associados aos critérios de concordância: péssimo, ruim, razoável, bom, muito bom e excelente. As restrições ao uso dessa tabela para os dados de sensoriamento remoto estão

relacionadas à divisão arbitrária dos valores do coeficiente Kappa e ao fato dela ter sido desenvolvida para análise de diagnósticos clínicos.

Outro método para a análise de exatidão de classificação, baseado na análise multivariada discreta, consiste na normalização de matrizes de erro. Isto é feito através de um procedimento denominado "ajuste proporcional iterativo", descrito por Congalton et al. (1981), que consiste em dividir os valores das células da matriz de erro pelos totais das linhas e das colunas, sucessivamente, até que a soma das células de cada linha e coluna seja igual a um. Esse procedimento faz com que o valor de cada célula seja influenciado pelo valor correspondente de sua linha e coluna. Cada célula torna-se uma combinação dos dados de referência e da classificação, sendo representativa dos erros de omissão e inclusão para cada categoria.

A exatidão global para cada matriz normalizada é calculada através da razão entre a soma dos valores da diagonal da matriz e a soma dos totais de suas linhas ou colunas (Congalton, 1991).

#### 2.4.1.2.1 - TÉCNICAS ANALÍTICAS PARA COMPARAÇÃO ENTRE DIFERENTES PRODUTOS DA CLASSIFICAÇÃO

Rosenfield e Melley (1980) sugeriram a aplicação de alguns testes para avaliar diferentes produtos da classificação. Esses testes avaliam se são significativas as diferenças entre as proporções de

concordância obtidas para cada produto. Um dos testes propostos foi a análise de variância de dois fatores. Eles ressaltaram que esse método apresentou melhores resultados do que os obtidos através de outros testes para diferença entre proporções, como o teste t, o teste Wilcoxon de sinal ordenado e o teste de sinal<sup>4</sup>.

Face às restrições apontadas à análise de variância, Congalton et al. (1983) sugeriram alguns métodos que se baseiam na análise multivariada discreta para comparar matrizes.

Um deles é a normalização de matrizes, o qual permite a eliminação do efeito do tamanho da amostra. Com essa técnica pode-se comparar, portanto, matrizes construídas a partir de amostras de tamanhos diferentes.

Outro método sugerido por Congalton et al. (1983) para comparar matrizes consiste em testar a concordância geral entre duas matrizes de erro, a um dado nível de significância  $\alpha$ . A medida de concordância geral é dada pelo coeficiente Kappa estimado ( $\hat{K}$ ), e o teste se baseia na normalidade assintótica do  $\hat{K}$  para grandes amostras. As hipóteses do teste são dadas por:

$$H_0: K_1 = K_2$$

$$H_1: K_1 \neq K_2$$

---

<sup>4</sup> O teste t foi descrito por Brownlee (1965, p. 282-308), e o teste Wilcoxon de sinal ordenado e o teste de sinal, por Hollander e Wolfe (1973, cap. 13).

onde:

$K_1$  e  $K_2$  = valores dos coeficientes Kappa correspondentes às matrizes de erro 1 e 2, a serem comparadas.

Ao nível  $\alpha$  de significância, a hipótese nula é aceita se:

$$\frac{\hat{K}_1 - \hat{K}_2}{\sqrt{\hat{\sigma}^2(\hat{K}_1) + \hat{\sigma}^2(\hat{K}_2)}} \leq z_{\alpha/2} \quad (2.31)$$

onde:

$\hat{K}_1$  e  $\hat{K}_2$  = estimativas de  $K_1$  e  $K_2$ , respectivamente; e

$\hat{\sigma}^2(\hat{K}_1)$  e  $\hat{\sigma}^2(\hat{K}_2)$  = estimativas da variância de  $\hat{K}_1$  e  $\hat{K}_2$ , respectivamente.

Congalton e Mead (1983) aplicaram o teste acima para verificar se havia diferenças significativas entre diferentes interpretações de fotografias aéreas, realizadas por cinco intérpretes. Eles concluíram que não houve diferenças significativas entre as cinco interpretações, ao nível de significância de 0,05.

O mesmo procedimento foi realizado por Dallemand (1987) para verificar se havia diferença significativa entre quatro diferentes classificações de imagens HRV-SPOT e TM-LANDSAT, concluindo-se que não houve diferenças significativas entre os quatro produtos, ao nível de significância de 0,05.



Embora nestes trabalhos os autores tenham generalizado o resultado para todos os produtos comparados, esse teste permite comparar somente duas matrizes de erro, a um nível de significância  $\alpha$ . Ele não pode ser aplicado a uma série de estimativas, referentes a várias matrizes pois, nesse caso, o nível de significância não se mantém igual a  $\alpha$ .

Neter e Wasserman (1974) salientaram que, nos casos em que se deseja comparar um conjunto de estimativas, é preferível usar um teste que indique o nível de significância para o conjunto.

Para comparar  $g$  estimativas de Kappa, quando  $g \geq 2$ , Fleiss (1981) propôs o teste  $\chi^2$ . Para sua aplicação calcula-se:

$$\hat{K}_{global} = \frac{\sum_{m=1}^g \frac{\hat{K}_m}{\hat{\sigma}_m^2(\hat{K}_m)}}{\sum_{m=1}^g \frac{1}{\hat{\sigma}_m^2(\hat{K}_m)}} \quad (2.32)$$

onde:

$\hat{K}_m$  = m-ésima estimativa de  $K$ ; e

$\hat{\sigma}_m^2(\hat{K}_m)$  = variância estimada de  $\hat{K}_m$ , dada pela Equação (2.20).

Para testar a hipótese de que os  $g$  valores de Kappa são iguais, ao nível  $\alpha$  de significância, o valor de:

$$\chi^2 = \sum_{m=1}^g \frac{(\hat{K}_m - \hat{K}_{global})^2}{\hat{\sigma}_m^2(\hat{K}_m)} \quad (2.33)$$

deve ser comparado ao valor tabelado de  $\chi_\alpha^2$ , com  $g-1$  graus de liberdade. A hipótese é rejeitada se o valor calculado for maior que o tabelado.

Dado que os valores de exatidão global dos diferentes produtos são proporções de acertos, pode-se verificar se eles diferem significativamente utilizando um teste para a comparação entre proporções.

Segundo Fleiss (1981), para testar se há diferença significativa entre  $g$  proporções, obtidas a partir de  $g$  amostras independentes, calcula-se:

$$\chi^2 = \frac{1}{\bar{p}\bar{q}} \sum_{m=1}^g n_m (\hat{p}_m - \bar{p})^2 \quad (2.34)$$

onde:

$\hat{p}_m$  = proporção estimada da  $m$ -ésima amostra;

$\bar{p}$  = média dos  $g$  valores de  $\hat{p}_m$ ;

$n_m$  = tamanho da  $m$ -ésima amostra; e

$$\bar{q} = 1 - \bar{p}. \quad (2.35)$$

Se o valor calculado for maior que o valor tabelado de  $\chi_\alpha^2$ , com  $g-1$  graus de liberdade, conclui-se que há diferença significativa entre pelo menos duas proporções, ao nível  $\alpha$  de significância.

#### 2.4.2 - TIPOS DE AMOSTRAGEM

Nos primeiros trabalhos sobre exatidão de classificação, dois tipos de amostragem predominavam: a aleatória simples (Hord e Brooner, 1976; Card, 1982) e a estratificada por categoria (Van Genderen et al., 1978; Hay, 1979; Card, 1982).<sup>5</sup>

No caso da amostragem estratificada por categoria, os cálculos de exatidão podem ser tendenciosos ("biased")<sup>6</sup> devido às diferentes taxas de amostragem empregadas às diversas categorias no mapa requerendo, portanto, a correção dos cálculos (Card, 1982).

Baseados em experimentos, Berry e Baker (1968, citados por Stehman, 1992) afirmaram que, para dados de uso do solo, onde a autocorrelação espacial<sup>7</sup> decai com o aumento da distância, a maior eficiência relativa<sup>8</sup> para

---

<sup>5</sup> A descrição da amostragem aleatória simples e da amostragem estratificada pode ser encontrada no trabalho de Snedecor e Cochran (1967, p. 10-11, 511, 520-527).

<sup>6</sup> Nesse caso, considera-se um estimador não tendencioso ("unbiased") aquele cujo valor médio para todas as amostras possíveis é igual ao parâmetro populacional que está sendo estimado. A média dos erros do estimador, considerando seus sinais, é chamada *tendência* ("bias") desse estimador (Snedecor e Cochran, 1967, p. 45-46, 506).

<sup>7</sup> A definição de autocorrelação espacial pode ser encontrada no trabalho de Ripley (1981, p. 10)

<sup>8</sup> A variância de uma estimativa de um parâmetro deve ser pequena. Se  $\hat{\theta}_1$  e  $\hat{\theta}_2$  são duas estimativas do parâmetro  $\theta$ , com variância  $\sigma_1^2$  e  $\sigma_2^2$ , respectivamente, a eficiência de  $\hat{\theta}_2$  relativa a  $\hat{\theta}_1$  é definida como

a análise de exatidão é obtida com a amostragem sistemática (AS) <sup>9</sup>. Se a forma da função de autocorrelação for desconhecida, uma amostragem estratificada sistemática não alinhada (AESNA) <sup>10</sup> fornece maior eficiência e segurança aos procedimentos de estimativa da exatidão. Esse tipo de amostragem é sugerido por alguns autores para a estimativa da exatidão de classificação (Fitzpatrick-Lins, 1981; Rosenfield et al. 1982), tendo sido apontado como o delineamento amostral menos tendencioso. Ele ainda fornece uma ponderação da área de cada categoria visto que um maior número de unidades amostrais é selecionado nas categorias de maior ocorrência (Fitzpatrick-Lins, 1981).

Um bom desempenho da AESNA não foi observado por Congalton (1988b). O autor analisou cinco tipos de amostragem (amostragem sistemática - AS; AESNA; amostragem aleatória simples - AAS; amostragem estratificada geograficamente - AE; e amostragem por aglomerados <sup>11</sup>) para diferentes tamanhos de amostra, aplicando-os a três diferentes tipos de uso do solo: agricultura, floresta e pastagem. O desempenho dessas foi analisado comparando os

-----  
 $\sigma_1^2/\sigma_2^2$ . Essa razão geralmente é expressa em porcentagem (Alexander, 1961, p. 176-210).

<sup>9</sup> A descrição da AS pode ser encontrada no trabalho de Snedecor e Cochran (1967, p. 519).

<sup>10</sup> A descrição da AESNA pode ser encontrada no trabalho de Congalton (1988b) e no item 3.5.1 do presente trabalho.

<sup>11</sup> A descrição da amostragem por aglomerados pode ser encontrada no trabalho de Snedecor e Cochran (1967, p. 511).

valores amostrais aos populacionais. Os resultados revelaram que a AESNA e a AS superestimaram os parâmetros da população. A AAS teve um bom desempenho em todos os casos. A AE também teve um bom desempenho, podendo ser empregada especialmente quando se deseja que pequenas áreas sejam representadas na amostra. A amostragem por aglomerados teve um desempenho razoável.

Segundo Congalton (1988b), embora os melhores resultados tenham sido obtidos com o uso da AAS, seguida pela AE, a AESNA e a AS forneceram estimativas adequadas dos parâmetros populacionais, quando aplicadas aos dados de floresta.

O autor ressaltou também que a complexidade espacial do ambiente determina o tipo de amostragem apropriado para a criação de matrizes de erro com vistas à estimativa da exatidão.

Stehman (1992) analisou o desempenho de três tipos de amostragem (AAS, AS e AESNA), para diferentes tamanhos de amostra, aplicando-os a oito tipos de população. Ele comparou estimativas da verdadeira proporção de erros e do coeficiente Kappa para amostras muito grandes aos valores estimados obtidos através de diferentes delineamentos amostrais empregados. Os resultados mostraram que os delineamentos sistemáticos são, em geral, mais precisos que a AAS, utilizando, portanto, recursos amostrais mais eficientes. Porém, esses delineamentos devem ser evitados quando houver indícios de forte periodicidade no modelo espacial dos erros, a menos que se tenha

informações disponíveis para se evitar um intervalo amostral desfavorável.

Essa forte periodicidade somente foi observada por Stehman (1992) em uma população criada artificialmente, justamente com o objetivo de avaliar o efeito da periodicidade.

Dois critérios foram utilizados por Stehman (1992) para comparar os delineamentos amostrais e apontar os que apresentaram melhor desempenho: a estimativa não tendenciosa para a proporção de erros ( $\hat{q}$ ) e a eficiência da estimativa, dada pela menor variância amostral ( $\sigma^2(\hat{q})$ ). A variância amostral mede a variabilidade do estimador  $\hat{q}$  sobre o conjunto de todas as amostras da população que podem ser selecionadas, ou seja, ela mede a dispersão da distribuição de  $\hat{q}$ . Ela é um parâmetro e depende do delineamento amostral. Assim, o delineamento amostral que apresentar o menor valor de  $\sigma^2(\hat{q})$ , para uma dada população, deve ser preferido quando outras condições como custo ou praticidade forem iguais.

Stehman (1992) observou que a tendência ("bias") para a estimativa da proporção de erros ( $\hat{q}$ ) foi desprezível para os delineamentos sistemáticos. Já a estimativa da variância requer uma especial consideração. O procedimento comum para a estimativa da variância de  $\hat{q}$  é o cálculo de sua variância estimada ( $\hat{\sigma}^2(\hat{q})$ ), dada por:

$$\hat{\sigma}^2(\hat{q}) = \frac{(N-n)\hat{q}(1-\hat{q})}{n(N-1)} \quad (2.36)$$

onde:

$\frac{N-n}{N-1}$  = fator de correção para o caso do uso de amostragem de tamanho  $n$ , sem reposição, de uma população de tamanho finito  $N$  (Spiegel, 1978, p. 276).<sup>12</sup>

A Equação (2.36) fornece uma estimativa da variância de  $\hat{q}$ , supondo-se uma amostragem aleatória simples. Para amostragens sistemáticas, ela pode superestimar a verdadeira variância, se o delineamento sistemático resultar num ganho de precisão<sup>13</sup> sobre a amostragem aleatória simples, e subestimar a verdadeira variância, se o delineamento sistemático representar uma perda de precisão em relação à amostragem aleatória simples. Outros estimadores da variância de  $\hat{q}$  para delineamentos sistemáticos não foram desenvolvidos para o

---

<sup>12</sup> A fórmula para  $\hat{\sigma}^2(\hat{q})$  foi apresentada por Stehman (1992) com erro. A formulação correta, baseada em Spiegel (1978, p. 223 e 276), é a fornecida no texto acima.

<sup>13</sup> Segundo Stehman (1992), a precisão dos delineamentos sistemáticos foi obtida em relação à AAS calculando-se o "efeito do delineamento", que é dado pela razão entre  $\sigma^2(\hat{q})$  do delineamento sistemático e  $\sigma^2(\hat{q})$  da AAS. Ao se multiplicar o efeito do delineamento sistemático por 1000 tem-se o número de observações requeridas para que um delineamento sistemático apresente a mesma precisão que a AAS com 1000 observações. Se o efeito do delineamento exceder 1, o delineamento sistemático apresenta menor precisão que a AAS.

estabelecimento da exatidão de classificação (Stehman, 1992).

Stehman (1992), utilizou ainda uma amostragem aleatória independente (AAI), a qual baseia-se na eqüiprobabilidade das unidades amostrais, para satisfazer o modelo requerido para o cálculo da variância estimada do coeficiente  $\hat{K}$  ( $\hat{\sigma}^2(\hat{K})$ ). Ao empregar delineamentos sistemáticos, a tendência para o cálculo do  $\hat{K}$  ("bias") observada por Stehman (1992) foi desprezível. Porém, o cálculo de intervalos de confiança para  $\hat{K}$  a partir da estimativa da variância, conforme proposto por Bishop et al. (1975), requer algumas considerações. Nos casos em que houve perda de precisão dos delineamentos sistemáticos em relação à AAI, ( $\hat{\sigma}^2(\hat{K})$ ) subestimou a variância de  $\hat{K}$  e o intervalo de confiança calculado foi menor que o nominal. Quando houve ganho de precisão, ( $\hat{\sigma}^2(\hat{K})$ ) superestimou a variância de  $\hat{K}$  e o intervalo de confiança calculado foi maior que o nominal.

#### 2.4.3 - TAMANHO DA AMOSTRA

O tamanho da amostra é um fator a ser considerado na estimativa da exatidão de classificação. Dado que cada unidade amostral coletada tem um custo, o tamanho da amostra deve ser mínimo, porém suficiente para assegurar que as análises estatísticas sejam válidas.



Ginevan (1979) estabeleceu três critérios para a seleção de um delineamento amostral visando a estimativa da exatidão de classificação:

- deve resultar em uma baixa probabilidade de aceitação de um mapa de baixa exatidão,
- deve resultar em uma alta probabilidade de aceitação de um mapa de alta precisão, e
- deve requerer um número mínimo de amostras como "verdade terrestre".

Diversos autores propuseram a determinação do tamanho adequado da amostra para a verificação da exatidão de classificação baseada na distribuição binomial da proporção de acertos e erros de classificação, na amostra, ou na aproximação para a distribuição normal (Hord e Brooner, 1976; Van Genderen et al., 1978; Ginevan, 1979; Rosenfield et al., 1982; Aronoff, 1982 e 1985).

Rosenfield et al. (1982) apresentaram um método para o cálculo inicial do tamanho da amostra ( $n_0$ ). Nesse caso, ele é dado por:

$$n_0 = \frac{\left(z_{\alpha/2}\right)^2 p_0(1-p_0)}{d^2} \quad (2.37)$$

onde:

$z_{\alpha/2}$  = valor na função de distribuição acumulada da normal padrão relativo a uma área igual a  $1-\alpha/2$ ;

$p_0$  = exatidão de classificação mínima desejada para o mapa;  
e

$d$  = erro máximo de estimativa desejado, ou seja:

$$P(|\hat{p} - p_0| \leq d) = 1 - \alpha \quad (2.38)$$

onde:

$\hat{p}$  = estimativa da exatidão de classificação.

Na Figura 2.5 pode-se observar graficamente a relação entre  $p_0$ ,  $d$  e  $\alpha$  considerando a distribuição normal de amostras de tamanho  $n$ .

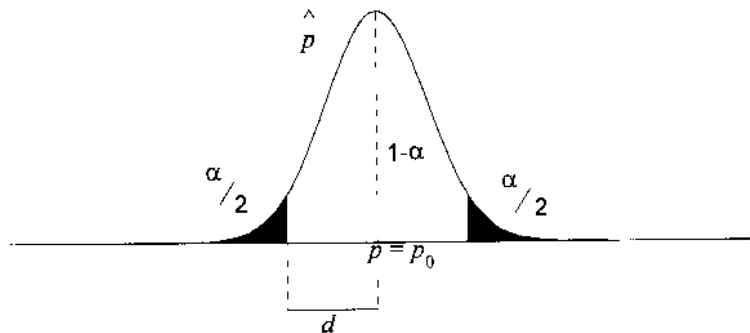


Fig. 2.5 - Representação da distribuição de  $\hat{p}$  para amostras de tamanho  $n$ , mostrando a relação entre  $p_0$ ,  $d$  e  $\alpha$ , numa distribuição normal, considerando  $p = p_0$ .

Rosenfield et al. (1982) observaram que esse cálculo não considera a correção de continuidade para a aproximação normal à distribuição binomial.

Ao se aplicar a correção de continuidade, obtém-se  $n$  através da equação:

$$d = z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n} + \frac{1}{2n}}. \quad (2.39)$$

Rosenfield et al. (1982) propuseram ainda o cálculo do tamanho mínimo da amostra para cada categoria do mapa baseado na probabilidade binomial acumulada. Nesse caso, supõe-se uma estimativa inicial para a exatidão da categoria ( $\hat{p}_0$ ) e definem-se os valores para o erro máximo desejado ( $d$ ) e para o nível de significância desejado ( $\alpha$ ). Considerando a probabilidade expressa na Equação (2.38), estima-se  $n$  a partir da equação:

$$\alpha = \sum_{z=k(n)+1}^n \binom{n}{z} \hat{p}_0^z (1-\hat{p}_0)^{n-z} \quad (2.40)$$

onde:

$k(n)$  = maior número inteiro menor ou igual a  $n(\hat{p}_0 + d)$ .

O tamanho mínimo da amostra é o menor número inteiro tal que  $\alpha$  seja menor ou igual ao valor selecionado.

Congalton (1988b) avaliou o desempenho de diferentes delineamentos amostrais comumente utilizados para a estimativa da exatidão de classificação. A avaliação foi realizada comparando os valores de média e variância de amostras de diversos tamanhos com os valores populacionais correspondentes. Os resultados revelaram que o tamanho da

amostra e o tipo de amostragem influem no desempenho, sendo esses fatores interdependentes.

Utilizando as amostragens aleatória e sistemática, Rocha (1992) analisou a influência do tamanho da amostra na análise da exatidão de classificação. Ele observou que as estimativas obtidas para Kappa e exatidão global não foram afetadas pelo tamanho da amostra, porém a precisão de seus cálculos o foi. A precisão das estimativas, medida através de seus coeficientes de variação<sup>14</sup>, aumentou com o aumento do tamanho das amostras.

Rocha (1992) observou que o coeficiente de variação decresce de forma não linear com o aumento do tamanho da amostra. Dessa forma pode-se determinar um "ponto ótimo", a partir do qual o aumento do tamanho da amostra tem uma influência apenas marginal no coeficiente de variação.

As técnicas propostas para a determinação do tamanho da amostra, baseadas na proporção de amostras corretamente classificadas e num erro tolerável, podem não ser adequadas para o preenchimento de uma matriz de erro. No caso de uma matriz de erro, não se considera simplesmente o que está correto ou incorreto, mas também que categorias estão sendo confundidas (Congalton, 1991).

---

<sup>14</sup> O coeficiente de variação é definido como a razão entre o desvio padrão e a média, sendo frequentemente expresso em porcentagem (Costa Neto, 1977, p. 29). Quanto menor o coeficiente de variação, maior a precisão da estimativa.

Portanto, essas técnicas não são adequadas quando a matriz resultante apresenta um elevado número de células com o valor zero, pois isto dificulta a análise entre categorias.

No estudo desenvolvido por Stehman (1992), amostras de 1% da área apresentaram bons resultados para a estimativa da proporção de erros e do coeficiente Kappa. O autor observou ainda que amostras de tamanho inferior a 60 foram, aparentemente, muito pequenas para satisfazer a condição de normalidade assintótica de  $\hat{K}$ .

#### 2.4.4 - DADOS DE REFERÊNCIA

Para a estimativa da exatidão de classificação devem ser coletados dados de referência ou de campo. A coleta de dados em campo é o método mais recomendado. Porém, ele normalmente tem um custo elevado, devendo ser bem dimensionado para satisfazer as necessidades do estudo e do orçamento (Congalton e Biging, 1992).

Dados de referência podem ser, por exemplo, produtos da interpretação de fotografias aéreas. Em muitos casos supõe-se que a fotointerpretação é 100% correta, o que raramente é válido e pode resultar em erros na avaliação (Congalton, 1991).

Dados de campo podem ser usados para verificar a adequação dos dados de fotointerpretações usados como referência (Mead e Szajgin, 1982). Para tanto pode-se comparar os dados de campo e da fotointerpretação numa matriz de erro utilizando um número reduzido de amostras. Isto limita as inferências estatísticas obtidas a

partir da matriz, porém pode indicar se há problemas de confusão (Congalton e Green, 1993).

Congalton e Green (1993) salientaram que as diferenças entre dados da classificação e dados usados como referência podem ser devidas a erros na coleta, análise e processamento dos dados de referência. Esses erros podem reduzir significativamente o valor da exatidão. Os autores citaram como fontes dessas diferenças, além dos erros de classificação e delimitação dos dados de sensoriamento remoto, os seguintes fatores:

- as diferenças na sobreposição (registro) dos dados de referência e do mapa;
- o erro de delimitação quando os locais escolhidos para a verificação da exatidão são digitalizados erroneamente;
- o erro na entrada dos dados de referência quando esses são introduzidos numa base de dados para a estimativa da exatidão;
- os erros na classificação e delimitação dos dados de referência;
- as mudanças ocorridas na área entre as datas de obtenção dos dados de sensoriamento remoto e de referência; e
- a variação na classificação e delimitação dos dados de referência em função de inconsistências na interpretação, por exemplo, de diferentes tipos de vegetação.

## CAPÍTULO 3

### MATERIAL E MÉTODOS

Neste capítulo descreve-se a área de estudo e os procedimentos metodológicos adotados. Um esquema do trabalho realizado é apresentado na Figura 3.1.

A partir da definição dos objetivos, iniciou-se o trabalho com a seleção da área de estudo (item 3.1) e o estabelecimento da legenda temática (item 3.2). Em seguida, passou-se à interpretação das imagens (item 3.3) e à introdução destas em um sistema de informações geográficas através das digitalizações manual (item 3.4.1) e automática (item 3.4.2). Paralelamente, elaborou-se uma referência (item 3.5) para a análise da exatidão de classificação (item 3.6) e a análise comparativa entre os produtos (item 3.7).

Na Figura 3.1, o termo carta original refere-se às cartas temáticas elaboradas pelos três intérpretes.

#### 3.1 - ÁREA DE ESTUDO

A área de estudo corresponde à área da carta topográfica Santos, escala 1:50000, do IBGE (1984). Ela está situada entre os paralelos 23°45' S e 24°05' S e os meridianos 46°15' W e 46°30' W, totalizando aproximadamente 850 quilômetros quadrados. Parte da área está localizada na Baixada Santista, no litoral do Estado de São Paulo, e parte na região Metropolitana do Estado.

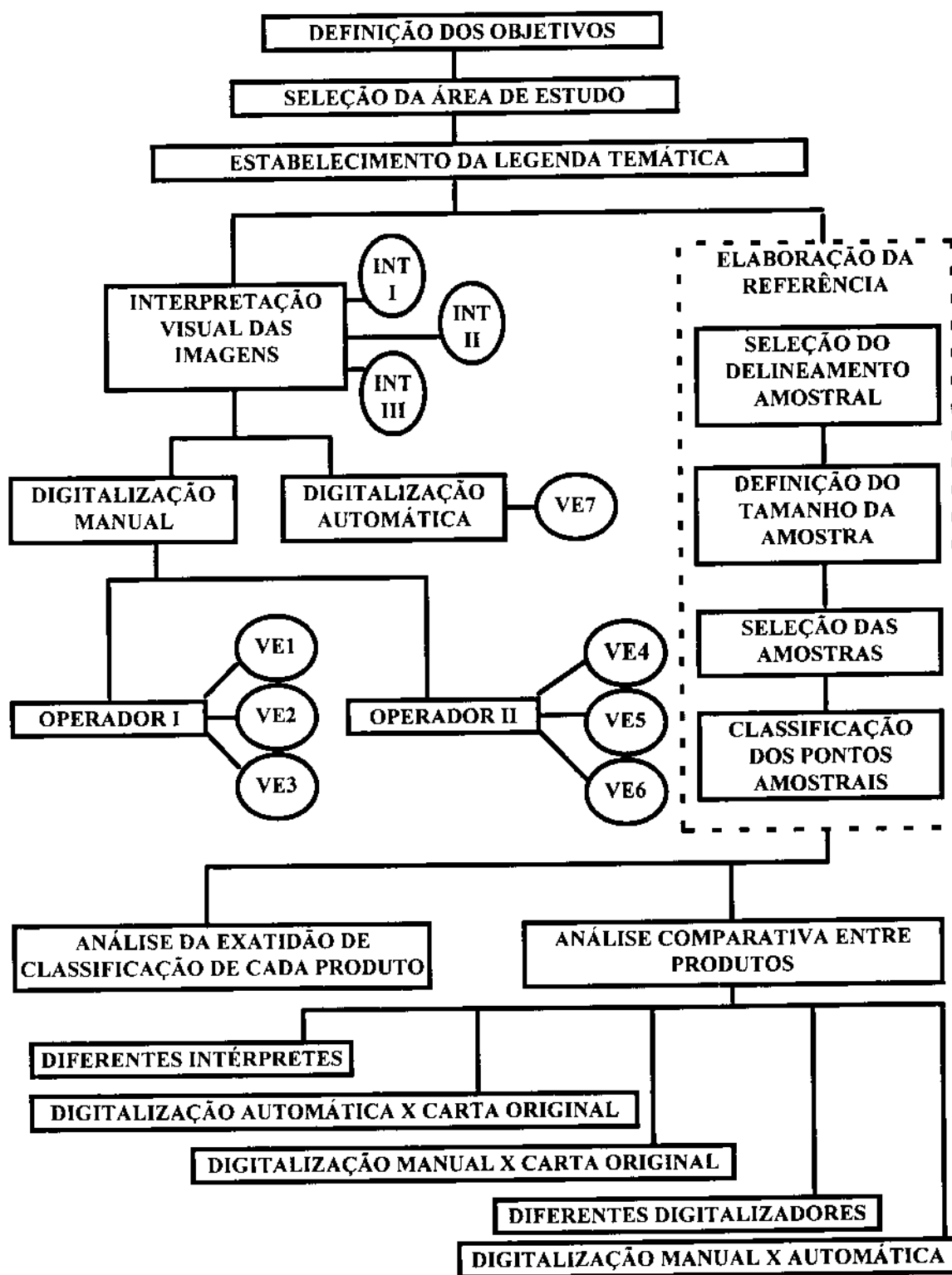


Fig. 3.1 - Esquema de trabalho.



A área abrange parte dos municípios de Ribeirão Pires, Rio Grande da Serra, Mogi das Cruzes, Santo André, São Bernardo do Campo, Santos, São Vicente, Guarujá e Praia Grande, e todo o município de Cubatão. Ao norte da área estão localizadas as cidades de Rio Grande de Serra (parcialmente) e Paranapiacaba, na região central, a cidade de Cubatão, e mais ao sul, Santos, São Vicente, Praia Grande e Guarujá, essas duas últimas parcialmente. A noroeste da área encontram-se parte das represas Billings e Rio Pequeno.

Duas grandes feições geológicas determinam os aspectos principais do material original da região: as escarpas da Serra do Mar, os maciços e morros isolados de um lado, e a planície sedimentar de outro (Queiroz Neto e Küpper, 1964).

Além disso, ao norte da carta acha-se representado um trecho do Planalto Atlântico Paulistano, de terrenos pré-Cambrianos, com altitudes entre 760 e 800 metros ou mais em alguns pontos, e que se individualiza pela presença de um relevo mamelonar, pertencente ao domínio dos "mares de morros" (CETESB, 1985).

As rochas encontradas nas escarpas da Serra, nos morros e nos maciços são principalmente granito e gnaisse, aparecendo também xistos e algumas rochas básicas (Queiroz Neto e Küpper, 1964). Com idades superiores a meio bilhão de anos (pré-Cambriano), as rochas são formadoras do embasamento cristalino (CETESB, 1985).

Esse maciço rochoso apresenta diversas fraturas nas direções nordeste e noroeste, por onde se formaram, através da erosão das águas, os atuais vales, como os dos rios Cubatão e Mogi (IBt/IF/CETESB, 1990).

A Serra do Mar apresenta ressalto topográfico entre 800 a 1000 metros. Suas encostas caracterizam-se pela formação de regolitos relativamente espessos, em decorrência do clima quente e úmido, gerador de intenso intemperismo químico. Mesmo quando cobertas por densa cobertura florestal, a evolução dessas encostas se faz, em função das altas declividades, por processos de remoção de solos rápida - escorregamentos - ou lenta - rastejo - resultando na formação de rampas de colúvio, talus e áreas de exposição de rocha (CETESB, 1985).

Na planície sedimentar distinguem-se as faixas de terrenos mais secos, compreendendo estreitos terraços sedimentares sobrelevados e depósitos coluviais e aluviais - nas baixas encostas de serras e morros - e, também, as faixas de antigas praias e restingas costeiras. As áreas ainda em processo de sedimentação apresentam amplos manguezais - planícies de maré - que se interpõem entre os dois tipos de terrenos planos mais secos citados. Para elas convergem, em forma de pequenos deltas e meandros, os rios oriundos da Serra do Mar, que ali se confundem, freqüentemente, com os canais do complexo estuarino (CETESB, 1985).

A atmosfera da Baixada Santista apresenta, como característica predominante, a média da umidade

relativa do ar elevada ao longo do ano, variando entre 70 e 90%. Isto se deve à posição geográfica e às condições meteorológicas da região. Geograficamente, por ser uma área litorânea próxima ao Oceano Atlântico, sofre constantemente sua marcada influência. Meteorologicamente, pelas massas equatoriais e marítimas que atuam, quase continuamente, na área. A maior quantidade de precipitações pluviométricas ocorre durante os meses de verão - outubro a março - representando 70% da média anual, a qual varia entre 2000 e 3000 milímetros de uma área para outra (CETESB, 1985).

A região da Baixada Santista caracteriza-se por graves problemas ambientais marcados pela forte concentração industrial e rápida urbanização. O aumento da população fixa, assim como a presença expressiva de uma população flutuante, exercem uma grande pressão sobre a estreita faixa de terra enxuta da região, forçando a especulação e o avanço sobre terras de mangues ou das encostas da Serra do Mar (CETESB, 1985).

O trecho do Planalto Atlântico Paulistano encontra-se atualmente recoberto, em sua maior parte, por uma floresta degenerada contendo capoeiras e por "campos antrópicos" resultantes dos desmatamentos que ocorrem principalmente nas áreas próximas às rodovias e sob a rede de alta tensão (CETESB, 1985).

As áreas originais de restinga sofreram os impactos de uma ocupação intensa, ora devastada para a utilização agrícola - bananicultura - e pela exploração de areia, ora dando lugar aos assentamentos urbanos e

loteamentos. Segundo levantamento efetuado por CETESB (1985), algumas áreas estavam sendo repovoadas pela vegetação de restinga, face ao progressivo abandono de áreas ocupadas pela bananicultura.

Os manguezais apresentam deterioração bastante significativa, principalmente nas áreas próximas ao pólo industrial e à rede viária principal (CETESB, 1985).

A poluição oriunda do pólo industrial de Cubatão acelerou a degradação da vegetação. Essa degradação, confirmada por vários trabalhos científicos, é caracterizada pela diminuição do número de árvores e de orquídeas e bromélias epífitas, pelo aumento do número de arbustos e ervas, e pela redução da diversidade das espécies (IBt/IF/CETESB, 1990).

Ao caracterizar a vegetação de dois trechos da Serra do Mar na Baixada Santista, Leitão Filho (1993) concluiu que as diferenças ambientais, fisionômicas e florísticas observadas nessas áreas são resultantes do processo de poluição atmosférica. Os trechos estudados foram próximos ao rio Pilões e ao rio Mogi. Ao longo do vale do rio Mogi, onde a poluição é drástica e ocorre há pelo menos vinte anos, a vegetação se encontra bastante descaracterizada. Em vários pontos existe exposição direta do solo ou de rochas. Em muitos locais a cobertura vegetal está restrita a plantas herbáceas ou arbustivas e as árvores remanescentes raramente formam um dossel contínuo. Na região do vale do rio Pilões observou-se a ocorrência de

floresta secundária em vários estágios sucessionais, resultado de alterações antrópicas como retirada de madeira, extração de palmito, queimadas, etc. Porém, a floresta apresentou alta diversidade e uma dinâmica sucessional dentro dos padrões.

A agricultura é quase nula na região, sendo necessária a importação de alimentos de outros locais para o abastecimento, principalmente nas épocas de maior presença da população flutuante (CETESB, 1985).

### 3.2 - ESTABELECIMENTO DA LEGENDA TEMÁTICA

Com o objetivo de analisar a influência de diferentes intérpretes na exatidão de classificação, foram comparadas cartas temáticas da vegetação elaboradas por três diferentes intérpretes. Dentre estas inclui-se a carta Santos do mapeamento da vegetação natural do Estado de São Paulo, elaborada pela Secretaria do Meio Ambiente.

Para tanto, houve a necessidade de basear a classificação na legenda utilizada pela Secretaria do Meio Ambiente, descrita por Hahn et al. (1990). Segundo os autores, optou-se por utilizar uma classificação fisionômica que, com pequenas variações, foi utilizada por alguns autores (Gonzaga de Campos, 1926; Aroldo Filho, 1950; Kuhlmann, 1956; Serra Filho, 1975; citados por Hahn et al., 1990) e que apresentou como vantagem melhor adaptação à terminologia utilizada na legislação ambiental. Além disso, eles consideraram a experiência de campo dos técnicos envolvidos no trabalho.

A descrição feita pelos autores às classes cujas formações são passíveis de ocorrência na região é dada a seguir.

Mata: formação vegetal complexa, dominada por árvores e com grande riqueza de espécies. Apresenta três estratos distintos. O superior, relativamente pouco denso, é formado por indivíduos de altura superior a 15 metros, de troncos cilíndricos, com engalhecimento médio a alto. O estrato intermediário, de alta densidade, é constituído por indivíduos de 5 a 10 metros, de copas bastante fechadas. O estrato inferior é constituído de ervas e arbustos, além de palmeiras e fetos arborescentes de até 3 metros de altura.

Capoeira: vegetação usualmente secundária, sucessora na derrubada das matas. É constituída principalmente por indivíduos lenhosos de segundo crescimento (que ocorriam na mata derrubada) e por espécies espontâneas, de rápido crescimento que colonizam as áreas desmatadas. Não tem estrutura definida, possui árvores finas, compactamente dispostas e com menor variedade de espécies.

Vegetação Natural de Várzea: vegetação de porte baixo, estrutura bastante variável, cuja característica é suportar inundações periódicas por estar situada nas baixadas que margeiam os rios.

Restinga: formação típica do litoral arenoso, predominantemente arbustiva, com xeromorfismo

acentuado pela pobreza e permeabilidade do solo arenoso e pela ação dissecadora dos ventos.

Mangue: formação típica de litoral lodoso, sob a ação direta das marés. É constituída por um único estrato de porte arbóreo e diversidade muito restrita. As árvores apresentam adaptações compatíveis às especificidades do ambiente, como raízes aéreas ou escoras para a sustentação, pneumatóforos que são estruturas que permitem trocas gasosas das raízes e xeromorfismo acentuado pela grande quantidade de sal, humatos e baixas taxas de oxigênio no solo que dificultam a absorção da água.

Formações Arbóreas com Áreas Menores que a Unidade Mínima de Classificação Temática: formações arbóreas com áreas menores que 4,0 hectares para regiões de relevo suave ondulado e 9,0 hectares para regiões de relevo acidentado, que não puderam ser classificadas dentre os temas anteriormente descritos por não apresentarem um padrão característico de resposta espectral nas imagens TM/LANDSAT e ser altamente dispendioso caracterizá-las individualmente em campo.

A análise prévia da imagem TM utilizada na classificação e da carta Santos, elaborada pela Secretaria do Meio Ambiente, revelou a necessidade de modificar a legenda adotada na carta Santos para as seguintes classes: Mata, Capoeira, Vegetação Natural de Várzea e Formações Arbóreas com Áreas Menores que a Unidade Mínima de Classificação Temática.

As classes Restinga e Mangue foram mantidas. Uma descrição detalhada das formações vegetais pertencentes a estas classes encontra-se nos itens 2.1.2 e 2.1.3.

A diversidade das formações florestais da Mata Atlântica nessa região, em função da localização, estágio de sucessão e outros fatores mencionados anteriormente, dificulta a discriminação das classes Capoeira e Mata descritas por Hahn et al. (1990), até mesmo em levantamentos fitofisionômicos em campo. A discriminação dessas duas classes nas imagens do sensor TM é bastante suscetível a erros pois, além da limitação da resolução espacial, acrescenta-se o fato do relevo acidentado resultar em diferentes condições de iluminação do terreno. Assim, optou-se por englobar as classes Mata e Capoeira em uma única classe, então denominada Mata. Dessa forma, as áreas classificadas como Mata e Capoeira pela Secretaria do Meio Ambiente foram consideradas Mata nas análises realizadas neste estudo.

A classe Mata englobou, portanto, todas as formações de diferentes composições florísticas e características fisionômicas da Floresta Atlântica, descritas no item 2.1.1.

A delimitação das áreas de Restinga e Mata, em muitos casos, não pôde ser feita baseada unicamente na interpretação visual, devido à transição progressiva que elas apresentavam.



Como no método descrito por Covre et al. (1992) não havia referência a esse problema, optou-se por definir um limite topográfico à ocorrência destas classes. A escolha de um fator topográfico como limite à ocorrência da vegetação de Restinga baseou-se na definição de Restinga adotada por Kirizawa et al. (1992) - "ecossistema adjacente ao oceano, encontrado em planícies arenosas" - o que permitiu concluir que esta formação pode ser identificada pela sua localização no relevo. Para tanto, foi selecionada a curva de nível de menor altitude da carta Santos (IBGE, 1984) (20 metros) como o limite máximo à ocorrência dessas planícies arenosas e, portanto, da Restinga.

Na definição dada por Hahn et al. (1990) à classe intitulada Vegetação Natural de Várzea predominaram critérios geomorfológicos. Dessa forma ela engloba diferentes fisionomias vegetais que podem ser inseridas em mais de uma classe da legenda utilizada pela Secretaria do Meio Ambiente. Por exemplo, segundo essa definição, as florestas ciliares da Floresta Atlântica podem ser inseridas nas classes Mata e Vegetação Natural de Várzea. Esse tipo de problema também ocorre com os brejos de restinga, que podem pertencer às classes Restinga e Vegetação Natural de Várzea. Para eliminar essa fonte de erro, que não está relacionada à interpretação visual e sim à definição prévia da legenda, optou-se por englobar a classe Vegetação Natural de Várzea nas demais classes existentes, em função de sua fisionomia.

Havia uma única área classificada como Vegetação Natural de Várzea na carta elaborada pela

Secretaria do Meio Ambiente. Pelo fato dessa área se encontrar no Planalto Atlântico e possuir características espectrais de Mata na imagem TM, ela foi redefinida como Mata no presente estudo.

A classe intitulada Formações Arbóreas com Áreas Menores que a Unidade Mínima de Classificação Temática também apresentou alguns problemas relacionados à sua definição. A carta elaborada pela Secretaria do Meio Ambiente continha algumas áreas menores que a unidade mínima definida na legenda que foram atribuídas a outras classes. Isto pode ter ocorrido devido à dificuldade em se classificar as formações em função do tamanho da área que elas ocupam. Além disso, este fato evidenciou a possibilidade de associar estas áreas às demais classes definidas na legenda. Face a isto, a classe Formações Arbóreas com Áreas Menores que a Unidade Mínima de Classificação Temática foi suprimida da legenda.

Por essa classe não caracterizar um tipo de formação vegetal de importância neste estudo e devido à pequena extensão de sua área na carta elaborada pela Secretaria do Meio Ambiente - 0,5 quilômetros quadrados ou aproximadamente 0,06% da área estudada - ela foi desconsiderada nas análises.

A legenda adotada foi composta, portanto, de três classes: Mata, Restinga e Mangue. Esta legenda se assemelha àquela utilizada pela Fundação SOS Mata Atlântica e INPE (1993) para o estudo da evolução dos remanescentes

da vegetação de Mata Atlântica através de imagens TM, que adotou as classes: Mata Atlântica, Restingas e Mangues.

### 3.3 - INTERPRETAÇÃO VISUAL DAS IMAGENS

Para a análise da possível influência de diferentes intérpretes na exatidão das cartas temáticas elaboradas, utilizou-se três diferentes mapeamentos da mesma área: a carta Santos do mapeamento da vegetação natural do Estado de São Paulo elaborada pela Secretaria do Meio Ambiente, aqui designada por intérprete III, e a interpretação realizada por mais dois intérpretes (intérpretes I e II). Dado que os três intérpretes tinham conhecimento e experiência semelhantes em fotointerpretação, supôs-se que eles apresentavam mesma capacidade fotointerpretativa.

Para esta análise buscou-se assegurar condições semelhantes para a elaboração das cartas temáticas, principalmente em relação ao material e à metodologia empregados.

Devido à impossibilidade de utilizar a mesma cópia das imagens usadas pela Secretaria do Meio Ambiente, os intérpretes I e II utilizaram cópias feitas a partir dos negativos que originaram aquelas imagens.

A imagem utilizada foi uma composição colorida, em papel, na escala 1:50000, das bandas 3, 4 e 5 do sensor TM-LANDSAT, nas cores azul, verde e vermelho, correspondente à cena de 12 de setembro de 1988, órbita 219, ponto 76, quadrante D, subquadrante S. Para a

interpretação de uma parcela ao sul desta área, que não foi abrangida por esta imagem, foi utilizada a mesma composição colorida da órbita 219, ponto 77, quadrante B, subquadrante A, de 8 de junho de 1988.

A metodologia empregada pelos intérpretes I e II assemelhou-se à usada pelo intérprete III, citada por Covre et al. (1992).

Inicialmente foi feita uma fotoleitura parcial da área de estudo da imagem para a seleção de algumas parcelas posteriormente visitadas em campo. O objetivo deste trabalho de campo, realizado em setembro de 1993, foi a identificação de padrões espectrais na imagem associados às formações florestais de interesse. Estas informações subsidiaram a interpretação realizada logo após a realização do trabalho de campo.

Em seguida delimitou-se algumas referências geográficas em papel poliéster estático. Estas referências foram os limites urbanos, as principais estradas, uma parte da hidrografia e o limite oceânico, que foram compiladas da carta topográfica Santos, em escala 1:50000 (IBGE, 1984).

O papel poliéster foi sobreposto à imagem e, para a fotointerpretação de cada parcela da área, as referências geográficas de ambos foram ajustadas. Tal procedimento foi necessário pois a imagem apresentou distorções em relação à carta topográfica.

As áreas de vegetação natural foram delimitadas sobre o papel poliéster, segundo a legenda previamente definida, para toda a área de estudo.

Os critérios adotados pelos intérpretes I e II para a classificação da vegetação foram baseados na legenda utilizada pela Secretaria do Meio Ambiente com as modificações citadas no item 3.2.

### 3.4 - INTRODUÇÃO DAS CARTAS TEMÁTICAS EM UM SISTEMA DE INFORMAÇÕES GEOGRÁFICAS

A introdução das cartas temáticas em um sistema de informações geográficas foi feita através de dois diferentes métodos:

- a digitalização manual e
- a digitalização automática.

O sistema de informações geográficas utilizado foi o SGI, desenvolvido pelo INPE. Esse sistema utiliza ambiente de microcomputadores compatível com a linha IBM-PC, sob sistema operacional DOS (Felgueiras e Câmara, 1993).

#### 3.4.1 - INTRODUÇÃO DOS DADOS ATRAVÉS DA DIGITALIZAÇÃO MANUAL

As cartas temáticas produzidas pelos intérpretes I, II e III foram introduzidas no SGI por dois operadores. Neste caso, foi analisada a possível influência

de diferentes operadores do sistema de informações geográficas na exatidão da carta temática resultante.

A digitalização foi iniciada com a criação de um projeto, denominado FLO, envolvendo a área da carta Santos (IBGE, 1984). A seguir foram digitalizados, por um único operador, dois planos de informação: DREN, contendo os elementos da hidrografia e o limite do oceano, e MOL, contendo o limite da área estudada.

A partir da definição do projeto e da digitalização dos planos DREN e MOL, os operadores I e II digitalizaram as três cartas elaboradas pelos intérpretes I, II e III. Assim, foram criados seis planos de informação referentes à vegetação (Tabela 3.1).

TABELA 3.1 - PLANOS DE INFORMAÇÃO DO PROJETO FLO REFERENTES À VEGETAÇÃO

PLANO DE INFORMAÇÃO	INTERPRETAÇÃO	DIGITALIZAÇÃO	OPERADOR
VE1	INTÉRPRETE I	MANUAL	OPERADOR I
VE2	INTÉRPRETE II		
VE3	INTÉRPRETE III		
VE4	INTÉRPRETE I		OPERADOR II
VE5	INTÉRPRETE II		
VE6	INTÉRPRETE III		
VE7	INTÉRPRETE III	AUTOMÁTICA	

#### 3.4.2 - INTRODUÇÃO DOS DADOS ATRAVÉS DA DIGITALIZAÇÃO AUTOMÁTICA

A carta temática elaborada pelo intérprete III foi introduzida no SGI também através da digitalização automática. Isto foi feito para analisar a possível

influência das diferentes formas de introdução dos dados em um sistema de informações geográficas na exatidão da carta temática resultante.

Inicialmente foram demarcados no papel poliéster somente os limites dos polígonos de vegetação mapeados pelo intérprete III. Através do uso de um "scanner" obteve-se a rasterização deste desenho.

Em seguida foi feita a transformação dos dados do formato "raster" para o formato vetorial. Para tanto utilizou-se um programa que realiza a vetorização automática<sup>15</sup>.

Após a vetorização, foram realizados o georreferenciamento dos dados e o ajuste de linhas automático e manual. Para a associação dos polígonos às classes da carta, foi realizada a rotulação manual dos polígonos (IMAGEM, 1995).

O plano de informação resultante foi denominado VE7 (Tabela 3.1).

### 3.5 - ESTABELECIAMENTO DOS DADOS DE REFERÊNCIA PARA A ANÁLISE DE EXATIDÃO

No item 2.4.4 salientou-se que a análise de exatidão de mapeamentos temáticos requer a adoção de uma referência para a comparação dos mapeamentos produzidos.

---

<sup>15</sup> Este programa foi desenvolvido pela Empresa IMAGEM Sensoriamento Remoto S/C Ltda.

Neste trabalho, a referência foi elaborada a partir de fotografias aéreas de data próxima à das imagens interpretadas.

A proximidade das datas foi necessária para reduzir os erros que podem ocorrer na obtenção dos dados de referência (Congalton e Green, 1993). Nesse caso, procurou-se evitar os erros advindos das mudanças que podem ter ocorrido na área entre as datas de obtenção dos dados de sensoriamento remoto e de referência, face, principalmente, à dinâmica de uso do solo.

A referência foi elaborada segundo o delineamento amostral adotado para as análises de exatidão de classificação efetuadas neste estudo.

### 3.5.1 - DELINEAMENTO AMOSTRAL E TAMANHO DA AMOSTRA

A amostragem utilizada para a análise de exatidão de classificação baseou-se em diversos autores (Fitzpatrick-Lins, 1981; Rosenfield et al., 1982; Stehman, 1992), os quais sugeriram o uso da amostragem estratificada sistemática não alinhada (AESNA). Essa técnica apresenta como vantagem a ponderação da área de cada classe.

Em geral, o uso de amostragens que não apresentam essa ponderação resulta em erros. Nesses casos, para os cálculos de exatidão torna-se necessário se fazer uma ponderação baseada na área classificada de cada categoria, o que somente corresponde à realidade se o mapeamento for 100% correto.



O problema da periodicidade dos erros, citado por Stehman (1992) foi desconsiderado pois espera-se um comportamento mais aleatório dos erros em mapeamentos de recursos naturais, neste caso, da vegetação natural.

A ponderação proposta por Prisley e Smith (1987), baseada na importância de cada classe, não foi aplicada neste estudo por considerar-se todas as classes igualmente importantes.

Uma amostragem de 1% da área, baseada nos resultados obtidos por Stehman (1992), mostrou-se inviável para o presente trabalho. Este tamanho de amostra foi desconsiderado devido ao elevado número de unidades amostrais a serem levantadas nas fotografias aéreas (descritas no item 3.5.2 e no Apêndice A) e à impossibilidade de localizar com precisão estas unidades em uma grade de células muito pequenas.

Optou-se, então, por empregar o cálculo inicial do tamanho da amostra sugerido por Rosenfield et al. (1982), com a inclusão do termo de correção de continuidade (Equação (2.39)).

Como o cálculo requer a definição prévia do intervalo de confiança e da exatidão mínima desejada para a carta, estes foram baseados no que propôs Aronoff (1982). Assim, foi considerado um intervalo de confiança de 95%, ou seja,  $1-\alpha = 0,95$  e  $z_{\alpha/2} = 1,96$ ; e uma exatidão mínima desejada de 85%, ou  $p_0 = 0,85$ .

O erro máximo de estimativa ( $d$ ) foi escolhido dentro do intervalo dos valores propostos por Rosenfield et al. (1982), 0,10, e Fitzpatrick-Lins (1981), 0,02.

Os critérios adotados para a escolha do tamanho de amostra foram:

- a possibilidade de localizar com precisão as unidades amostrais através da seleção de um tamanho de célula apropriado;
- a viabilidade de classificar todas as unidades amostrais nas fotografias aéreas ou em campo; e
- o atendimento à condição de normalidade assintótica de  $\hat{K}$ , escolhendo uma amostra de tamanho maior que 60, conforme os resultados obtidos por Stehman (1992).

A aquisição da amostragem estratificada sistemática não alinhada foi feita conforme descrição de Berry (1962, citado por Congalton, 1988b). Inicialmente a população foi dividida em estratos, e o tamanho de cada estrato foi determinado pelo número de unidades amostrais desejado. Como resultado desta divisão obteve-se uma grade, em que cada célula é um estrato, e o número células é igual ao número de unidades amostrais.

A Figura 3.2 apresenta um esquema desse tipo de amostragem. A seleção do ponto número 1 é aleatória. A coordenada  $x$  do ponto 1 é usada com uma nova coordenada aleatória  $y'$  para localizar o ponto 2. Uma nova coordenada

aleatória  $y''$  com a coordenada  $x$  do ponto 1 define a localização do ponto 3. Isto é feito para toda a linha. Da mesma forma a coordenada aleatória  $y$  do ponto 1 é usada com uma nova coordenada aleatória  $x'$  para localizar o ponto 4. Assim são localizados os pontos da coluna. O ponto 5 é localizado usando a coordenada aleatória  $x'$  do ponto 4 e a coordenada aleatória  $y'$  do ponto 2. Por esse método são localizados os pontos dos demais estratos.

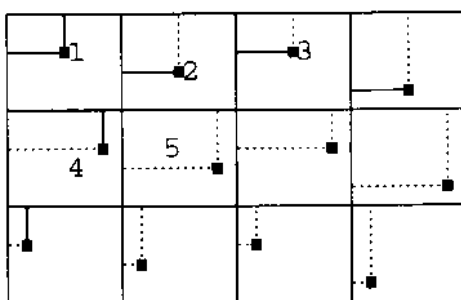


Fig. 3.2 - Esquema ilustrativo da amostragem estratificada sistemática não alinhada.

FONTE: Modificada de Congalton (1988b), p. 595.

O tamanho definido para a amostra possibilitou o uso da grade existente na carta Santos do IBGE (1984), com células de  $16 \text{ cm}^2$ . Esta grade segue as coordenadas UTM, facilitando a localização das unidades amostrais.

A grade foi compilada em papel poliéster estático permitindo a localização das unidades amostrais e posterior sobreposição às cartas temáticas para verificação da classificação de cada unidade.

Em função do limite da carta Santos ser apresentado em coordenadas geodésicas e a grade estar baseada em coordenadas planas (UTM), as células das bordas da carta apresentaram área menor que  $16 \text{ cm}^2$ . Assim, para a localização das unidades amostrais considerou-se como primeira coluna aquela que possuía todas as células de  $16 \text{ cm}^2$ , neste caso, a segunda coluna da carta.

As células menores que  $16 \text{ cm}^2$  foram consideradas para a obtenção das unidades amostrais. Porém, quando a unidade amostral não pôde ser localizada dentro dos limites destas células, ela foi desconsiderada.

A seleção aleatória das coordenadas iniciais foi feita segundo uma tabela de números aleatórios (Spiegel, 1978, p. 496). Para a localização das unidades amostrais na grade, foi considerada como unidade mínima o milímetro. Também foram calculadas as coordenadas UTM correspondentes a estas unidades amostrais.

### 3.5.2 - INTERPRETAÇÃO VISUAL DE FOTOGRAFIAS AÉREAS

As unidades amostrais da grade foram localizadas nas fotografias aéreas para a identificação das classes segundo a legenda utilizada na interpretação das imagens TM.

Embora sujeito a um maior número de erros quando comparado à coleta de dados em campo, o uso de fotografias aéreas para a elaboração da referência tem sido muito comum (Costa, 1990; Fung e LeDrew, 1988; Gong e Howarth, 1990; Marsh et al., 1994; Saxena et al., 1992).

As fotografias utilizadas pertencem ao arquivo do Instituto Geográfico e Cartográfico de São Paulo (IGC). As fotografias, em filme pancromático preto e branco, escala aproximada de 1:35000, referem-se ao levantamento aéreo realizado na Baixada Santista em 1986 e 1987.

Antes de se localizar as unidades amostrais nas fotografias, as mesmas foram localizadas em cartas planialtimétricas do projeto Baixada Santista, em escala 1:10000 (IGC, 1987), as quais foram elaboradas a partir das fotografias aéreas citadas. Este procedimento foi realizado com o objetivo de reduzir os erros devido à delimitação dos dados de referência, conforme ressaltado por Congalton e Green (1993).

O Apêndice A contém a relação das unidades amostrais, das cartas em escala 1:10000 e das fotografias aéreas onde elas foram localizadas. O Apêndice B contém os nomes das cartas utilizadas.

Para a classificação das unidades amostrais nas fotografias aéreas utilizou-se estereoscópio e lupa.

Alguns procedimentos foram adotados para evitar os erros na classificação dos dados de referência citados por Congalton e Green (1993):

- dois intérpretes (diferentes dos que interpretaram as imagens) classificaram, independentemente, as unidades amostrais nas fotografias aéreas, segundo os critérios usados na interpretação da imagem TM;

- a classificação feita pelos dois intérpretes foi feita sem consulta à imagem TM da área de estudo;
- as duas classificações foram comparadas e as unidades que apresentaram discordância foram novamente analisadas, com o objetivo de obter um consenso quanto à classe a ser atribuída;
- para a nova análise, feita em conjunto pelos dois intérpretes, foram utilizadas, em alguns casos, outras fontes de informação sobre a vegetação da área, como a carta do meio ambiente e de sua dinâmica da Baixada Santista (CETESB, 1985), um levantamento sobre os manguezais da região (Herz, 1991) e fotografias em filme infravermelho falsa-cor de 1975 e 1985; e
- as unidades para as quais não houve consenso na classificação, mesmo após a análise conjunta dos intérpretes, foram selecionadas para verificação em campo.

As fotografias em filme infravermelho foram utilizadas para o reconhecimento de algumas áreas em que foi difícil diferenciar padrões de restinga e mangue nas fotografias aéreas em filme pancromático preto e branco. As fotografias foram tomadas em 1975 e 1985, em escala aproximada de 1:25000, e referem-se a levantamentos aéreos realizados pelo INPE.

Em alguns casos, as unidades amostrais caíram no limite entre duas ou mais classes. Para cada uma

destas unidades foi atribuída a classe de maior ocorrência no entorno.

Terminada esta seqüência de procedimentos, as unidades amostrais foram localizadas na imagem TM interpretada para verificar se houve alterações na área que pudessem resultar em erros na elaboração da referência.

Algumas unidades amostrais que incidiam no oceano não foram submetidas a estes procedimentos pois não havia dúvidas quanto à sua classificação.

Uma pequena parcela da área estudada, correspondente a 10 unidades amostrais, não se encontrava nos limites da área das fotografias e cartas do IGC. Para estas unidades utilizou-se o mesmo procedimento, porém a partir de fotografias aéreas do levantamento realizado na região da Grande São Paulo e Baixada Santista contratado pela Empresa Metropolitana de Planejamento da Grande São Paulo (EMPLASA), de junho e julho de 1977, em escala aproximada de 1:40000. A relação destas unidades amostrais e das respectivas fotografias utilizadas encontra-se no Apêndice A.

### 3.5.3 - LEVANTAMENTO DE CAMPO

O levantamento de campo foi realizado em abril de 1995 para verificar as unidades amostrais que não puderam ser classificadas a partir das fotografias aéreas.

Havia um grande intervalo de tempo entre a data da imagem interpretada e o trabalho de campo. Para

minimizar os erros na elaboração da referência devidos a possíveis alterações ocorridas na área, foi comparado o padrão destas unidades amostrais na imagem interpretada ao de uma imagem de mesma composição, mesmo subquadrante, porém mais recente, de 24 de julho de 1993. Procurou-se selecionar uma imagem do mesmo período do ano que a imagem utilizada na interpretação. Devido à constante cobertura de nuvens na região, não foi possível obter uma imagem do mesmo mês (setembro).

A localização das unidades amostrais em campo foi feita com o auxílio das cartas em escala 1:10000. A classificação destas unidades foi baseada nos mesmos critérios usados para a interpretação da imagem TM.

### 3.6 - ANÁLISE DE EXATIDÃO DE CLASSIFICAÇÃO

Os dados de referência foram comparados às cartas temáticas elaboradas pelos intérpretes I, II e III, e às cartas temáticas resultantes da digitalização pelos operadores I e II e da digitalização automática, totalizando dez cartas temáticas, conforme ilustrado na Figura 3.1.

A localização das unidades amostrais nas cartas temáticas elaboradas pelos intérpretes I, II e III foi feita sobrepondo a grade amostral sobre cada carta e anotando as classes a que pertenciam estas unidades. A localização nas cartas digitalizadas foi feita através da introdução no SGI da listagem das coordenadas UTM destas



unidades amostrais para a identificação automática das classes.

Nesta fase foi avaliada a exatidão de cada uma das dez cartas separadamente, segundo alguns métodos propostos na literatura, os quais são apresentados na Figura 3.3.

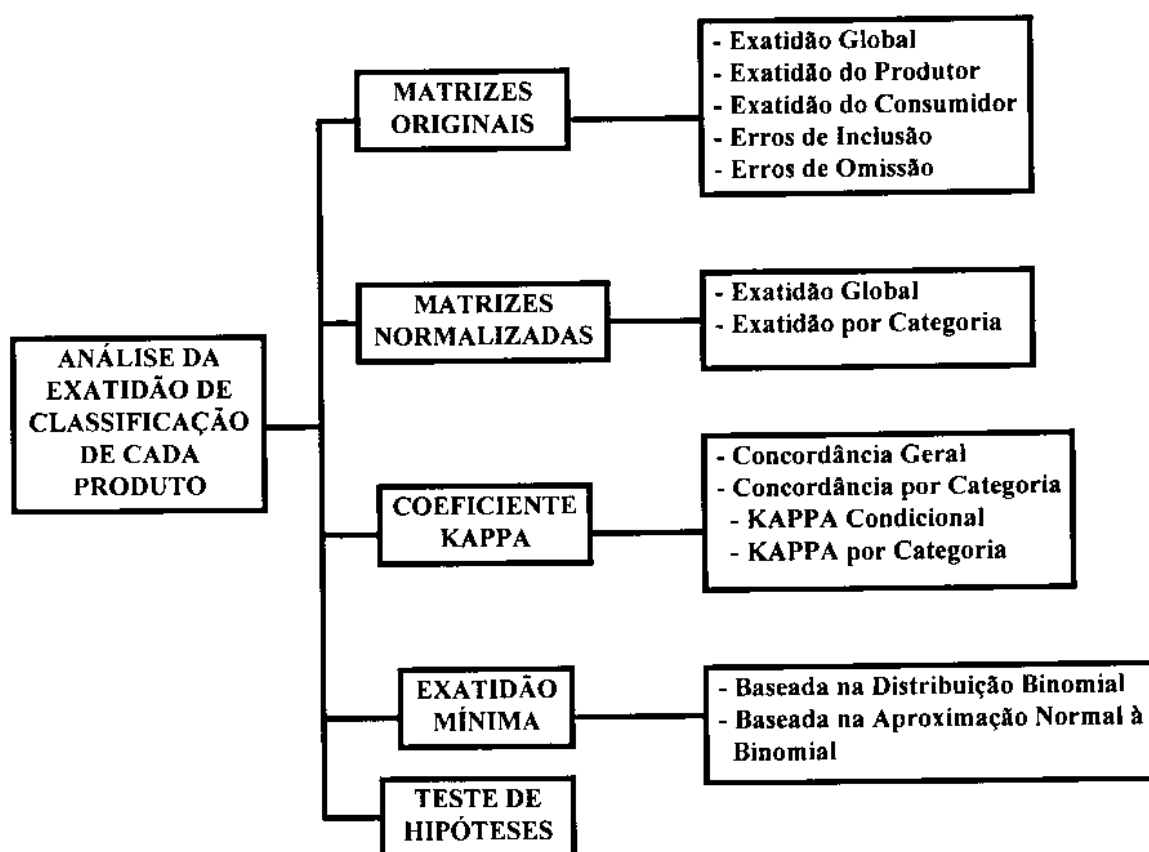


Fig. 3.3 - Detalhamento dos métodos utilizados para a análise da exatidão de classificação de cada produto.

Quatro classes foram consideradas para as análises: Mata, Mangue, Restinga (descritas anteriormente)

e Vazio. Na classe Vazio foram englobadas todas as áreas não classificadas como Mata, Mangue ou Restinga.

Inicialmente, foi elaborada uma matriz de erro de 4 colunas por 4 linhas, referente a cada carta. A análise foi efetuada a partir da aplicação do método descritivo proposto por Story e Congalton (1986). Foram estimadas as exatidões global, do produtor e do consumidor. A matriz permitiu ainda a análise da relação entre os erros de omissão e inclusão.

Calculou-se o intervalo de confiança de 95% para a exatidão global de cada produto, baseado em Fleiss (1981) (Equação (2.2)).

As matrizes de erro foram ainda normalizadas segundo o método descrito por Congalton et al. (1981). Para tanto, foi desenvolvido um programa em linguagem C (anexado no Apêndice C). As estimativas das exatidões global e por classe foram obtidas a partir das matrizes normalizadas.

Foram empregadas duas técnicas analíticas baseadas na distribuição binomial dos erros de classificação e que utilizam o limite inferior do intervalo de confiança como índice de exatidão. Estas técnicas são descritas no item 2.4.1.2, Equações (2.5) e (2.15). Uma delas, utilizada por Fitzpatrick-Lins (1981), considerou a aproximação normal à binomial. Para seu uso considerou-se um intervalo de confiança de 95%. A outra, proposta por Aronoff (1985), utilizou a distribuição binomial acumulada

para a estimativa do índice de exatidão. Para seu uso considerou-se um risco do consumidor de 0,05.

O teste de hipóteses foi empregado para cada um dos produtos. O método empregado foi o proposto por Aronoff (1982).

Considerando a exatidão mínima desejada, previamente definida, de 0,85, as hipóteses testadas foram:

$H_0: p < 0,85$  (a carta apresenta exatidão menor que a mínima desejada)

$H_1: p \geq 0,85$  (a carta apresenta exatidão maior ou igual à mínima desejada).

Para o teste, foi considerado o risco do consumidor igual a 0,05.

O risco do produtor foi calculado segundo a Equação (2.14), supondo dois valores para a exatidão dos produtos ( $p_U$ ): 0,90 e 0,95.

Outra técnica empregada para análise de exatidão a partir da matriz de erro foi a técnica analítica baseada na estimativa do coeficiente de concordância Kappa. Estimativas deste coeficiente foram obtidas como medidas de concordância geral ( $\hat{K}$ ), calculadas segundo a Equação (2.19), e para cada classe ou categoria ( $\hat{K}_i$  e  $\hat{K}_i^*$ ), segundo as Equações (2.26) e (2.29), respectivamente.

Foi ainda construído um intervalo de confiança para  $K$  baseado na estimativa da variância de  $\hat{K}$ , conforme Equação (2.20). Neste caso, foi utilizado um intervalo de confiança de 95%.

### 3.7 - ANÁLISE COMPARATIVA ENTRE OS PRODUTOS

O processo de mapeamento temático pode ser influenciado por algumas fontes de erro. No presente trabalho, essas fontes de erro foram analisadas nos seguintes processos:

- na classificação visual da imagem, a partir dos resultados obtidos por três diferentes intérpretes;
- na entrada dos dados em um sistema de informações geográficas, através da digitalização manual e automática; e
- na digitalização manual, a partir dos resultados obtidos por dois diferentes operadores.

Estas análises foram feitas através da comparação entre os produtos utilizando quatro métodos propostos na literatura.

A Tabela 3.2 apresenta os produtos comparados para cada análise efetuada.

TABELA 3.2 - PRODUTOS COMPARADOS PARA A ANÁLISE DA INFLUÊNCIA DE ALGUMAS FONTES DE ERRO NA EXATIDÃO

FONTES DE ERRO	PRODUTO COMPARADO
DIFERENTES INTÉRPRETES	INT I, INT II, INT III
DIGITALIZAÇÃO MANUAL x CARTA ORIGINAL	INT I, VE1, VE4
	INT II, VE2, VE5
	INT III, VE3, VE6
DIGITALIZAÇÃO AUTOMÁTICA x CARTA ORIGINAL	INT III, VE7
DIFERENTES OPERADORES	VE1, VE4
	VE2, VE5
	VE3, VE6
DIGITALIZAÇÃO MANUAL x AUTOMÁTICA	VE3, VE6, VE7

Na Tabela 3.2, INT I, INT II e INT III representam as cartas produzidas pelos intérpretes I, II e III, respectivamente.

O primeiro método foi proposto por Congalton (1991) e consistiu na comparação direta entre matrizes normalizadas.

O segundo, baseado em Fleiss (1981) (Equação (2.34)), possibilitou a comparação entre os valores de exatidão global dos produtos, aos níveis de significância de 0,01 e 0,05. Diversos testes de hipótese foram elaborados com a finalidade de verificar se houve influência das fontes de erro na exatidão (elencadas na Tabela 3.2).

Primeiramente testou-se se houve diferença significativa entre os valores de exatidão global dos três produtos da interpretação visual, sendo a hipótese nula dada por:

$$H_0: p_1 = p_2 = p_3$$

onde:

$p_1$ ,  $p_2$  e  $p_3$  = exatidões globais dos produtos elaborados pelos intérpretes I, II e III, respectivamente.

Três testes:

$$H_0: p_1 = p_{VE1} = p_{VE4};$$

$$H_0: p_2 = p_{VE2} = p_{VE5}; \text{ e}$$

$$H_0: p_3 = p_{VE3} = p_{VE6}$$

onde:

$p_{VE1}, \dots, p_{VE6}$  referem-se às exatidões globais dos produtos digitalizados manualmente,

foram propostos para testar se houve diferenças significativas entre os resultados obtidos para os produtos, antes e após a digitalização manual.

Outro teste:

$$H_0: p_3 = p_{VE7}$$

onde:

$p_{VE7}$  = exatidão global do produto digitalizado automaticamente,

foi proposto para testar se houve diferença significativa entre os resultados obtidos para a carta temática elaborada

pelo intérprete III, antes e após a sua introdução no SGI através da digitalização automática.

Os três testes seguintes:

$$H_0: p_{VE1} = p_{VE4}$$

$$H_0: p_{VE2} = p_{VE5}$$

$$H_0: p_{VE3} = p_{VE6}$$

foram propostos para testar se houve diferenças significativas entre os resultados da digitalização manual, realizada por dois operadores, das três cartas temáticas elaboradas pelos três intérpretes.

Por último:

$$H_0: p_{VE3} = p_{VE6} = p_{VE7}$$

foi usado para testar se houve diferenças significativas entre a digitalização manual e a digitalização automática da carta temática elaborada pelo intérprete III.

Nestes testes, as hipóteses alternativas representaram a ocorrência de diferença significativa entre pelo menos dois valores de exatidão global, ao nível de significância de 0,01 e 0,05.

O terceiro método, proposto por Congalton et al. (1983) (Equação (2.31)), consistiu em testar a igualdade entre pares de valores do coeficiente de concordância Kappa. Neste caso, foi adotado um nível de significância de 0,05.

O quarto método, proposto por Fleiss (1981) (Equações (2.32) e (2.33)), foi utilizado para testar a igualdade entre dois ou mais valores de Kappa. Para sua aplicação foi considerado um nível de significância de 0,05.

As formulações da hipótese nula, neste caso, analisaram as mesmas fontes de erro que as citadas para o teste entre os valores de exatidão global.



## CAPÍTULO 4

### RESULTADOS E DISCUSSÃO

#### 4.1 - PRODUTOS GERADOS

Três cartas de vegetação resultaram da interpretação visual das imagens TM-LANDSAT: uma elaborada pela Secretaria do Meio Ambiente do Estado de São Paulo e duas elaboradas por dois diferentes intérpretes, denominadas: Carta de Vegetação - Santos - Intérprete I e Carta de Vegetação - Santos - Intérprete II (Apêndices D e E, respectivamente).

As digitalizações manual e automática forneceram mais sete cartas de vegetação. Estas se encontram em arquivos contendo um projeto denominado FLO, que reúne os planos de informação MOL, DREN, VE1, VE2, VE3, VE4, VE5, VE6 e VE7, citados no item 3.4. Estes arquivos estão disponíveis na Biblioteca do Instituto Nacional de Pesquisas Espaciais (INPE).

#### 4.2 - ESTABELECIMENTO DOS DADOS DE REFERÊNCIA

O estabelecimento dos dados de referência foi necessário para se efetuar as análises de exatidão de classificação. A seleção do tipo de amostragem, do tamanho da amostra e o resultado da classificação das unidades amostrais são discutidos a seguir.

#### 4.2.1 - DELINEAMENTO AMOSTRAL SELECIONADO

Neste estudo decidiu-se pela utilização de um esquema de amostragem sistemática não alinhada (AESNA) pois ela apresenta, como vantagem, a ponderação da área de cada classe para o cálculo da exatidão de classificação.

Seguindo os resultados obtidos por Stehman (1992), planejou-se inicialmente uma amostragem compreendendo 1% da área. Porém, os cálculos para a confecção da grade necessária para a AESNA de 1% da área mostraram a inviabilidade de empregar este tamanho de amostra na área de estudo. Para amostrar 1% da área, tendo como unidade amostral a área do terreno correspondente ao pixel (900 m<sup>2</sup>), seriam necessárias 9500 unidades amostrais. No presente estudo, onde a escala de trabalho foi 1:50000, a grade amostral da AESNA seria composta de células de 0,36 cm<sup>2</sup>. Assim, o elevado número de unidades amostrais a serem localizadas e interpretadas nas fotografias aéreas, associado à impossibilidade de localizar com precisão estas unidades em células tão pequenas, inviabilizaram o emprego desse tamanho de amostra.

Estes resultados revelaram que amostragens de 1% da área, tendo como unidade amostral a área do terreno correspondente ao pixel, somente se viabilizaria para escalas de trabalho superiores à utilizada.

O emprego do cálculo inicial do tamanho da amostra considerando a correção da continuidade para a aproximação normal à distribuição binomial, requereu a

adoção de um valor para o erro máximo de estimativa ( $d$ ). Ao se utilizar o valor 0,10 (Rosenfield et al., 1982), obteve-se uma amostra inicial muito pequena, igual a 59. O uso do valor 0,02 (Fitzpatrick-Lins, 1981) resultou em uma amostra inicial muito grande, igual a 1274. Adotou-se, então,  $d=0,05$ , que resultou em um tamanho inicial de 215 unidades amostrais.

Ao se utilizar a metodologia descrita no item 3.5.1 para a elaboração da grade de referência, foram localizadas 218 unidades amostrais.

O método para a determinação do tamanho da amostra por classe proposto por Rosenfield et al. (1982) e citado no item 2.4.3, não foi utilizado por apresentar as seguintes limitações:

- a necessidade de se ter uma estimativa *a priori* da exatidão das classes analisadas, obtida através de um levantamento prévio, o que dificultaria o trabalho e elevaria os custos;
- a impossibilidade de utilizar uma amostragem que apresentasse a ponderação pela área de cada classe, devido à necessidade de se obter um número mínimo de unidades amostrais por classe; e
- não ser a estimativa de exatidão por classe o objetivo principal deste estudo e sim a análise da exatidão global das cartas temáticas.

O tamanho da amostra utilizado possibilitou o preenchimento das matrizes de erro sem resultar em um elevado número de células com valor zero (Tabela 4.1). Este tamanho foi, portanto, compatível com o número de classes definido. Caso o número de classes fosse maior, o número de células com valor zero poderia ser maior e o tamanho da amostra seria insuficiente para o preenchimento adequado da matriz.

#### 4.2.2 - CLASSIFICAÇÃO DAS UNIDADES AMOSTRAIS

No presente trabalho, procurou-se minimizar a ocorrência de erros na elaboração da referência, adotando-se os procedimentos descritos no item 3.5.2. Entretanto, tais procedimentos somente foram possíveis frente à facilidade de acesso a outras fontes de informação sobre a área (fotografias aéreas e levantamentos sobre a vegetação).

Todas as unidades da amostra foram classificadas e utilizadas nas análises de exatidão. Destas, somente três unidades não puderam ser classificadas a partir das fotografias aéreas, por não ter havido concordância entre os diferentes intérpretes. Assim, a classificação destas unidades foi feita a partir do levantamento efetuado em campo.

#### 4.3 - ANÁLISE DA EXATIDÃO DE CLASSIFICAÇÃO DOS PRODUTOS

As matrizes de erro referentes aos dez produtos gerados são apresentadas na Tabela 4.1 <sup>16</sup>.

Nesta tabela e nas que seguem, INT I, INT II e INT III referem-se aos produtos elaborados pelos intérpretes I, II e III, respectivamente. VE1, VE2, VE3, VE4, VE5, VE6 e VE7 são descritos no Capítulo 3 (Tabela 3.1).

Considerando que a AESNA deve representar uma ponderação das áreas das classes de estudo em seus resultados, nota-se, na Tabela 4.1, que as classes Restinga e Mangue foram as de menor ocorrência na área estudada (7,8% e 8,7%, respectivamente), enquanto as classes Mata e Vazio foram as mais frequentes (33,0% e 50,5%, respectivamente).

As estimativas da exatidão global obtidas a partir das matrizes originais e das matrizes normalizadas, e da exatidão mínima ( $p_L$  e  $p'_L$ ), por produto, são apresentadas na Tabela 4.2.

---

<sup>16</sup> A descrição das classes Mata, Restinga e Mangue é dada no item 3.2 e a da classe Vazio, no item 3.6.

TABELA 4.1 - MATRIZES DE ERRO. POR PRODUTO

		DADOS DE REFERÊNCIA				
		MATA	RESTINGA	MANGUE	VAZIO	TOTAIS
I	MATA	62	0	1	3	66
N	RESTINGA	1	6	1	1	9
T	MANGUE	0	3	15	2	20
	VAZIO	9	8	2	104	123
I	TOTAIS	72	17	19	110	218
I	MATA	68	0	0	6	74
N	RESTINGA	0	10	2	2	14
T	MANGUE	0	4	16	1	21
	VAZIO	4	3	1	101	109
II	TOTAIS	72	17	19	110	218
I	MATA	64	2	1	3	70
N	RESTINGA	0	4	0	1	5
T	MANGUE	0	3	15	1	19
	VAZIO	8	8	3	105	124
III	TOTAIS	72	17	19	110	218
	MATA	61	0	1	8	70
V	RESTINGA	0	5	1	1	7
E	MANGUE	0	4	14	3	21
1	VAZIO	11	8	3	98	120
	TOTAIS	72	17	19	110	218
	MATA	66	0	0	8	74
V	RESTINGA	0	9	2	3	14
E	MANGUE	0	4	13	1	18
2	VAZIO	6	4	4	98	112
	TOTAIS	72	17	19	110	218
	MATA	59	2	1	6	68
V	RESTINGA	0	4	0	1	5
E	MANGUE	0	3	10	2	15
3	VAZIO	13	8	8	101	130
	TOTAIS	72	17	19	110	218
	MATA	59	2	1	9	71
V	RESTINGA	2	8	4	5	19
E	MANGUE	0	2	13	6	21
4	VAZIO	11	5	1	90	107
	TOTAIS	72	17	19	110	218
	MATA	65	1	1	12	79
V	RESTINGA	0	5	1	2	8
E	MANGUE	0	7	10	0	17
5	VAZIO	7	4	7	96	114
	TOTAIS	72	17	19	110	218
	MATA	59	2	0	7	68
V	RESTINGA	0	3	0	1	4
E	MANGUE	0	2	9	2	13
6	VAZIO	13	10	10	100	133
	TOTAIS	72	17	19	110	218
	MATA	61	2	1	7	71
V	RESTINGA	0	4	0	0	4
E	MANGUE	0	3	11	3	17
7	VAZIO	11	8	7	100	126
	TOTAIS	72	17	19	110	218

TABELA 4.2 - ESTIMATIVAS DA EXATIDÃO GLOBAL (MATRIZES ORIGINAIS E NORMALIZADAS) E DA EXATIDÃO MÍNIMA ( $p_L$  e  $p'_L$ ), POR PRODUTO

PRODUTO	EXATIDÃO GLOBAL (MATRIZES ORIGINAIS)	EXATIDÃO GLOBAL (MATRIZES NORMALIZADAS)	EXATIDÃO MÍNIMA ( $p_L$ )	EXATIDÃO MÍNIMA ( $p'_L$ )
INT I	0,8578	0,8131	0,8091	0,8129
INT II	0,8945	0,8440	0,8514	0,8538
INT III	0,8624	0,8570	0,8144	0,8180
VE1	0,8165	0,7779	0,7628	0,7679
VE2	0,8532	0,7965	0,8039	0,8078
VE3	0,7982	0,8019	0,7426	0,7482
VE4	0,7798	0,7153	0,7225	0,7286
VE5	0,8073	0,7357	0,7527	0,7580
VE6	0,7844	0,7904	0,7275	0,7335
VE7	0,8073	0,8892	0,7527	0,7580

Observa-se na Tabela 4.2 que a menor estimativa da exatidão global calculada a partir das matrizes originais foi obtida para o produto VE4 (0,7798) e a maior, para INT II (0,8945).

Para o cálculo do intervalo de confiança de 95% para a exatidão global, optou-se por adotar a aproximação normal à distribuição binomial. Neste caso, foram satisfeitas as condições indicadas por Fleiss (1981), citadas no item 2.4.1.2.

Não se utilizou o fator de correção para o uso de amostragem sem reposição, citado no item 2.4.2 (Equação (2.36)). Isto porque, para uma amostra de 218 pixels, retirada de uma população de aproximadamente 990250 pixels, o fator seria próximo de 1, alterando só marginalmente os valores obtidos para os intervalos de confiança.

Os intervalos de confiança de 95% para a exatidão global, por produto, são apresentados na Figura 4.1.

### Exatidão Global

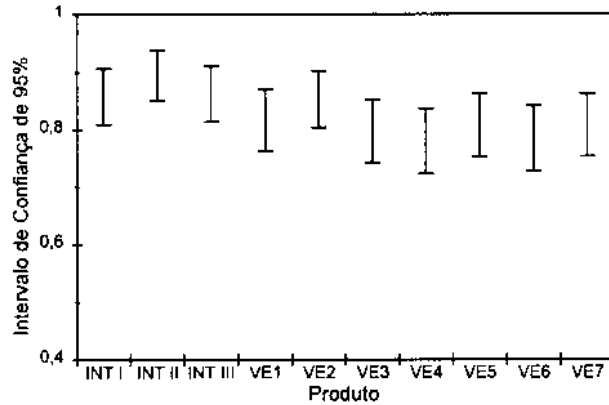


Fig. 4.1 - Intervalos de confiança de 95% para a exatidão global, por produto.

As estimativas das exatidões do produtor e do consumidor para cada classe, por produto, são apresentadas na Tabela 4.3 e os erros de omissão e inclusão, nas Tabelas 4.4 e 4.5, respectivamente.

TABELA 4.3 - ESTIMATIVAS DAS EXATIDÕES DO PRODUTOR E DO CONSUMIDOR, POR CLASSE E POR PRODUTO

PRODUTO	EXATIDÃO DO PRODUTOR				EXATIDÃO DO CONSUMIDOR			
	MATA	RESTINGA	MANGUE	VAZIO	MATA	RESTINGA	MANGUE	VAZIO
INT I	0,8611	0,3529	0,7895	0,9455	0,9394	0,6667	0,7500	0,8455
INT II	0,9444	0,5882	0,8421	0,9182	0,9189	0,7143	0,7619	0,9266
INT III	0,8889	0,2353	0,7895	0,9545	0,9143	0,8000	0,7895	0,8468
VE1	0,8472	0,2941	0,7368	0,8909	0,8714	0,7143	0,6667	0,8167
VE2	0,9167	0,5294	0,6842	0,8909	0,8919	0,6429	0,7222	0,8750
VE3	0,8194	0,2353	0,5263	0,9182	0,8676	0,8000	0,6667	0,7769
VE4	0,8194	0,4706	0,6842	0,8182	0,8310	0,4211	0,6190	0,8411
VE5	0,9028	0,2941	0,5263	0,8727	0,8228	0,6250	0,5882	0,8421
VE6	0,8194	0,1765	0,4737	0,9091	0,8676	0,7500	0,6923	0,7519
VE7	0,8472	0,2353	0,5789	0,9091	0,8592	1,0000	0,6471	0,7937



TABELA 4.4 - ERROS DE OMISSÃO, POR CLASSE E POR PRODUTO

PRODUTO	DADOS DE REFERÊNCIA	DADOS DE CLASSIFICAÇÃO			
		MATA	RESTINGA	MANGUE	VAZIO
I	MATA	---	0,0139	0,0000	0,1250
N	RESTINGA	0,0000	---	0,1765	0,4706
T	MANGUE	0,0526	0,0526	---	0,1053
I	VAZIO	0,0273	0,0091	0,0182	---
I	MATA	---	0,0000	0,0000	0,0556
N	RESTINGA	0,0000	---	0,2353	0,1765
T	MANGUE	0,0000	0,1053	---	0,0526
II	VAZIO	0,0545	0,0182	0,0091	---
I	MATA	---	0,0000	0,0000	0,1111
N	RESTINGA	0,1176	---	0,1765	0,4706
T	MANGUE	0,0526	0,0000	---	0,1579
III	VAZIO	0,0273	0,0091	0,0091	---
V	MATA	---	0,0000	0,0000	0,1528
V	RESTINGA	0,0000	---	0,2353	0,4706
E	MANGUE	0,0526	0,0526	---	0,1579
1	VAZIO	0,0727	0,0091	0,0273	---
V	MATA	---	0,0000	0,0000	0,0833
V	RESTINGA	0,0000	---	0,2353	0,2353
E	MANGUE	0,0000	0,1053	---	0,2105
2	VAZIO	0,0727	0,0273	0,0091	---
V	MATA	---	0,0000	0,0000	0,1806
V	RESTINGA	0,1176	---	0,1765	0,4706
E	MANGUE	0,0526	0,0000	---	0,4211
3	VAZIO	0,0545	0,0091	0,0182	---
V	MATA	---	0,0278	0,0000	0,1528
V	RESTINGA	0,1176	---	0,1176	0,2941
E	MANGUE	0,0526	0,2105	---	0,0526
4	VAZIO	0,0818	0,0455	0,0545	---
V	MATA	---	0,0000	0,0000	0,0972
V	RESTINGA	0,0588	---	0,4118	0,2353
E	MANGUE	0,0526	0,0526	---	0,3684
5	VAZIO	0,1091	0,0182	0,0000	---
V	MATA	---	0,0000	0,0000	0,1806
V	RESTINGA	0,1176	---	0,1176	0,5882
E	MANGUE	0,0000	0,0000	---	0,5263
6	VAZIO	0,0636	0,0091	0,0182	---
V	MATA	---	0,0000	0,0000	0,1528
V	RESTINGA	0,1176	---	0,1765	0,4706
E	MANGUE	0,0526	0,0000	---	0,3684
7	VAZIO	0,0636	0,0000	0,0273	---

TABELA 4.5- ERROS DE INCLUSÃO, POR CLASSE E POR PRODUTO

PRODUTO	DADOS DE REFERÊNCIA	DADOS DE CLASSIFICAÇÃO			
		MATA	RESTINGA	MANGUE	VAZIO
I	MATA	---	0,0000	0,0152	0,0455
N	RESTINGA	0,1111	---	0,1111	0,1111
T	MANGUE	0,0000	0,1500	---	0,1000
I	VAZIO	0,0732	0,0650	0,0163	---
I	MATA	---	0,0000	0,0000	0,0811
N	RESTINGA	0,0000	---	0,1429	0,1429
T	MANGUE	0,0000	0,1905	---	0,0476
II	VAZIO	0,0367	0,0275	0,0092	---
I	MATA	---	0,0286	0,0143	0,0429
N	RESTINGA	0,0000	---	0,0000	0,2000
T	MANGUE	0,0000	0,1579	---	0,0526
III	VAZIO	0,0645	0,0645	0,0242	---
	MATA	---	0,0000	0,0143	0,1143
V	RESTINGA	0,0000	---	0,1429	0,1429
E	MANGUE	0,0000	0,1905	---	0,1429
1	VAZIO	0,0917	0,0667	0,0250	---
	MATA	---	0,0000	0,0000	0,1081
V	RESTINGA	0,0000	---	0,1429	0,2143
E	MANGUE	0,0000	0,2222	---	0,0556
2	VAZIO	0,0536	0,0357	0,0357	---
	MATA	---	0,0294	0,0147	0,0882
V	RESTINGA	0,0000	---	0,0000	0,2000
E	MANGUE	0,0000	0,2000	---	0,1333
3	VAZIO	0,1000	0,0615	0,0615	---
	MATA	---	0,0282	0,0141	0,1268
V	RESTINGA	0,1053	---	0,2105	0,2632
E	MANGUE	0,0000	0,0952	---	0,2857
4	VAZIO	0,1028	0,0467	0,0093	---
	MATA	---	0,0127	0,0127	0,1519
V	RESTINGA	0,0000	---	0,1250	0,2500
E	MANGUE	0,0000	0,4118	---	0,0000
5	VAZIO	0,0614	0,0351	0,0614	---
	MATA	---	0,0294	0,0000	0,1029
V	RESTINGA	0,0000	---	0,0000	0,2500
E	MANGUE	0,0000	0,1538	---	0,1538
6	VAZIO	0,0977	0,0752	0,0752	---
	MATA	---	0,0282	0,0141	0,0986
V	RESTINGA	0,0000	---	0,0000	0,0000
E	MANGUE	0,0000	0,1765	---	0,1765
7	VAZIO	0,0873	0,0635	0,0556	---

Na Tabela 4.3, observa-se que as classes Mata e Vazio apresentaram estimativas mais elevadas das exatidões do produtor e do consumidor que aquelas apresentadas pelas demais classes.

Nota-se que, em alguns casos, as estimativas das exatidões do produtor e do consumidor foram muito diferentes. Por exemplo, a classe Restinga do produto VE7. Enquanto a estimativa da exatidão do produtor foi muito baixa, igual a 0,2353, a da exatidão do consumidor foi alta, igual a 1,0000.

A Tabela 4.4 mostra que a classe mais afetada por erros de omissão foi a Restinga, sendo confundida principalmente com a classe Vazio. Exceções são feitas aos produtos INT II e VE5, onde a classe Restinga foi mais confundida com a classe Mangue, e ao produto VE2 onde ela foi igualmente confundida com as classes Vazio e Mangue.

A Tabela 4.5 mostra que os erros de inclusão foram maiores para as classes Restinga, Mangue e Vazio. Na classificação de Mangue e Vazio observou-se a predominância da inclusão de unidades amostrais da classe Restinga e, na classificação de Restinga, de unidades amostrais da classe Mangue.

Se forem elaboradas matrizes como as do exemplo apresentado na Tabela 4.6, referente ao intérprete I, a soma dos valores das linhas será igual a 1. Nesta matriz, as estimativas da exatidão do produtor foram

acrescidas aos erros de omissão, evidenciando a relação existente entre eles.

TABELA 4.6 - EXEMPLO DE MATRIZ DE ERRO COM OS VALORES DA EXATIDÃO DO PRODUTOR E DOS ERROS DE OMISSÃO - INTÉRPRETE I

PRODUTO	DADOS DE REFERÊNCIA	DADOS DE CLASSIFICAÇÃO				TOTAIS
		MATA	RESTINGA	MANGUE	VAZIO	
I	MATA	0,8611	0,0139	0,0000	0,1250	1,0000
N	RESTINGA	0,0000	0,3529	0,1765	0,4706	1,0000
T	MANGUE	0,0526	0,0526	0,7895	0,1053	1,0000
I	VAZIO	0,0273	0,0091	0,0182	0,9454	1,0000

O mesmo pode ser feito para demonstrar a relação entre as estimativas da exatidão do consumidor e os erros de inclusão.

A Tabela 4.7 apresenta as matrizes de erro normalizadas, por produto.

Na Tabela 4.7, observa-se que a maior frequência de erros ocorreu para a classe Restinga, sendo confundida com as classes Mangue e Vazio. Estes resultados foram consistentes com aqueles obtidos a partir da análise dos erros de omissão e inclusão.

As estimativas das exatidões global e por classe obtidas a partir das matrizes normalizadas são apresentadas nas Tabelas 4.2 e 4.8, respectivamente. As estimativas da exatidão por classe correspondem aos valores das diagonais das matrizes normalizadas.

TABELA 4.7 - MATRIZES DE ERRO NORMALIZADAS, POR PRODUTO

		DADOS DE REFERÊNCIA				TOTAIS
		MATA	RESTINGA	MANGUE	VAZIO	
I	MATA	0,8997	0,0000	0,0388	0,0614	0,9999
N	RESTINGA	0,0504	0,7440	0,1346	0,0710	0,9999
T	MANGUE	0,0000	0,1469	0,7970	0,0561	0,9999
	VAZIO	0,0498	0,1090	0,0296	0,8115	0,9999
I	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
I	MATA	0,9465	0,0000	0,0000	0,0534	0,9999
N	RESTINGA	0,0000	0,7623	0,1730	0,0646	0,9999
T	MANGUE	0,0000	0,1771	0,8040	0,0188	0,9999
	VAZIO	0,0534	0,0605	0,0229	0,8631	0,9999
II	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
I	MATA	0,9156	0,0244	0,0296	0,0302	0,9999
N	RESTINGA	0,0000	0,8290	0,0000	0,1709	0,9999
T	MANGUE	0,0000	0,0746	0,9048	0,0205	0,9999
	VAZIO	0,0843	0,0719	0,0654	0,7783	0,9999
III	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
	MATA	0,8894	0,0000	0,0224	0,0881	0,9999
V	RESTINGA	0,0000	0,7289	0,1817	0,0892	0,9999
E	MANGUE	0,0000	0,1717	0,7493	0,0788	0,9999
1	VAZIO	0,1105	0,0992	0,0464	0,7438	0,9999
	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
	MATA	0,9264	0,0000	0,0000	0,0735	0,9999
V	RESTINGA	0,0000	0,7155	0,1706	0,1137	0,9999
E	MANGUE	0,0000	0,2170	0,7570	0,0259	0,9999
2	VAZIO	0,0735	0,0673	0,0723	0,7867	0,9999
	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
	MATA	0,8811	0,0240	0,0286	0,0662	0,9999
V	RESTINGA	0,0000	0,8128	0,0000	0,1871	0,9999
E	MANGUE	0,0000	0,1045	0,8313	0,0642	0,9999
3	VAZIO	0,1188	0,0587	0,1400	0,6824	0,9999
	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
	MATA	0,8155	0,0735	0,0266	0,0842	0,9999
V	RESTINGA	0,0582	0,6192	0,2240	0,0985	0,9999
E	MANGUE	0,0000	0,1546	0,7272	0,1181	0,9999
4	VAZIO	0,1262	0,1526	0,0221	0,6991	0,9999
	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
	MATA	0,8929	0,0096	0,0145	0,0828	0,9999
V	RESTINGA	0,0000	0,6300	0,1897	0,1802	0,9999
E	MANGUE	0,0000	0,3173	0,6826	0,0000	0,9999
5	VAZIO	0,1070	0,0429	0,1131	0,7369	0,9999
	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
	MATA	0,8901	0,0296	0,0000	0,0803	0,9999
V	RESTINGA	0,0000	0,7945	0,0000	0,2054	0,9999
E	MANGUE	0,0000	0,0930	0,8348	0,0721	0,9999
6	VAZIO	0,1098	0,0828	0,1651	0,6421	0,9999
	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996
	MATA	0,8936	0,0000	0,0225	0,0838	0,9999
V	RESTINGA	0,0000	0,9999	0,0000	0,0000	0,9999
E	MANGUE	0,0000	0,0000	0,8733	0,1265	0,9999
7	VAZIO	0,1063	0,0000	0,1040	0,7896	0,9999
	TOTAIS	0,9999	0,9999	0,9999	0,9999	3,9996

TABELA 4.8 - ESTIMATIVAS DA EXATIDÃO, POR CLASSE E POR PRODUTO, A PARTIR DAS MATRIZES NORMALIZADAS

PRODUTO	MATA	RESTINGA	MANGUE	VAZIO
INT I	0.8997	0.7440	0.7970	0.8115
INT II	0.9465	0.7623	0.8040	0.8631
INT III	0.9156	0.8290	0.9048	0.7784
VE1	0.8894	0.7289	0.7493	0.7438
VE2	0.9264	0.7155	0.7570	0.7867
VE3	0.8811	0.8128	0.8313	0.6824
VE4	0.8155	0.6192	0.7272	0.6991
VE5	0.8929	0.6300	0.6826	0.7369
VE6	0.8901	0.7945	0.8348	0.6421
VE7	0.8936	0.9999	0.8733	0.7896

Na Tabela 4.2, observa-se que a menor estimativa da exatidão global calculada a partir das matrizes normalizadas foi obtida para o produto VE4 (0,7153) e a maior, para VE7 (0,8892).

As estimativas da exatidão global obtidas a partir das matrizes normalizadas diferiram daquelas obtidas a partir das matrizes originais nos seguintes aspectos:

- nos valores mínimos e máximos da estimativa da exatidão [mín,máx]:
  - matrizes originais: [0,7798,0,8945]; e
  - matrizes normalizadas: [0,7153,0,8892]; e
- na ordem decrescente dos produtos segundo sua estimativa da exatidão:
  - matrizes originais: INT II, INT III, INT I, VE2, VE1, VE5 e VE7, VE3, VE6, VE4; e

- matrizes normalizadas: VE7, INT III, INT II, INT I, VE3, VE2, VE6, VE1, VE5, VE4.

Definindo-se  $p_{or}$  e  $p_{nor}$  como as estimativas da exatidão global obtidas a partir das matrizes originais e normalizadas, respectivamente, a diferença percentual:

$$\left| \frac{p_{or} - p_{nor}}{p_{or}} \right| 100 \quad (4.1)$$

ficou compreendida entre 4,73% e 8,88%, com exceção às referentes ao INT III (0,62%) e aos planos digitalizados VE3, VE6 e VE7 (0,47%, 0,77% e 10,14%, respectivamente).

Congalton et al. (1981) também observaram diferenças entre as estimativas da exatidão global obtidas a partir dos dois tipos de matrizes (originais e normalizadas) tendo associado esse fenômeno ao pequeno número de acertos ocorridos para uma das classes dos produtos estudados. Como consequência, o procedimento de normalização resultou em um aumento dos valores na linha e na coluna daquela classe, reduzindo a estimativa da exatidão global a partir da matriz normalizada e aumentando, dessa forma, a diferença entre estimativas geradas.

No caso do intérprete III e dos planos VE3, VE6 e VE7, observou-se um pequeno número de acertos para a classe Restinga. Porém, o procedimento de normalização levou a uma redução na diferença percentual da estimativa da exatidão global para o intérprete III e para os planos VE3 e VE6, e a um aumento desta para o plano VE7. No caso

do intérprete III, a maior estimativa da exatidão global foi obtida a partir da matriz original, enquanto no caso dos planos VE3, VE6 e VE7, as maiores estimativas foram obtidas a partir das matrizes normalizadas.

Congalton et al. (1983) afirmaram que a estimativa da exatidão global obtida a partir das matrizes originais considera apenas os valores da diagonal principal, enquanto que a estimativa obtida a partir das matrizes normalizadas considera também os erros de omissão e inclusão.

É interessante notar, na Tabela 4.7, que para a classe Restinga do produto VE7, a estimativa da exatidão gerada a partir da matriz normalizada foi próxima a um (0,9999), representando uma elevada concordância. Embora este valor tenha sido compatível com a ausência de erros de inclusão, ele não foi compatível com o elevado número de erros de omissão observados nesta classe, os quais deveriam ser igualmente considerados no procedimento de normalização.

Os aumentos nas estimativas da exatidão obtidas para a classe Restinga, a partir das matrizes normalizadas, foram responsáveis pela elevação dos valores de exatidão global para os quatro produtos (INT III, VE3, VE6 e VE7).

Baseado em sua experiência, Congalton (1991) afirmou que, se a matriz original apresentar um elevado número de zeros nas células fora da diagonal principal, a



estimativa da exatidão global calculada a partir da matriz normalizada tende a ser diferente daquela obtida a partir da matriz original. Isto ocorre porque, no processo de normalização, estes zeros tendem a assumir valores positivos. O processo de normalização tende, então, a reduzir a estimativa da exatidão global devido aos valores positivos fora da diagonal principal. Quando há um pequeno número de células com valor zero fora da diagonal principal, estas duas medidas são mais próximas.

Este fenômeno não foi observado nos resultados do presente trabalho. As células de valor igual a zero continuaram a apresentar valor zero após a normalização.

Estes dados indicaram que o procedimento de normalização de matrizes pode resultar em estimativas de exatidão diferentes das obtidas para as matrizes originais.

Para verificar se estas diferenças são significativas, seria necessário analisar os intervalos de confiança das exatidões obtidas a partir das matrizes normalizadas. Porém, como não existe uma fórmula para o cálculo da variância da exatidão estimada a partir da matriz normalizada, não há como calcular os intervalos de confiança.

As diferenças entre as estimativas da exatidão global obtidas para os dois tipos de matrizes também podem ser atribuídas ao fato de que a soma das linhas e colunas da matriz normalizada deve ser próxima a

um. Assim, classes de maior freqüência têm o mesmo peso na estimativa da exatidão global que classes de pequena freqüência.

As estimativas da exatidão por classe, obtidas a partir das matrizes normalizadas, também diferiram daquelas apresentadas pelas exatidões do produtor e do consumidor. Colocando os produtos em ordem decrescente das estimativas das exatidões por classe, também não foi observada nenhuma relação entre estas três maneiras de estimar a exatidão.

Esse resultado era esperado visto que as estimativas das exatidões do produtor e do consumidor estão relacionadas aos erros de omissão e inclusão, respectivamente. A estimativa da exatidão por classe, obtida a partir da matriz normalizada, não é uma medida associada a algum destes tipos de erro isoladamente.

As estimativas da exatidão mínima, segundo as formulações propostas por Fitzpatrick-Lins (1981) e Aronoff (1985) ( $p_L$  e  $p'_L$ , respectivamente), encontram-se na Tabela 4.2.

Para o cálculo de  $p_L$  não foi utilizado o fator de correção para o uso de amostragem sem reposição, pois ele alteraria apenas marginalmente os valores obtidos.

As estimativas da exatidão mínima obtidas a partir desses dois métodos foram semelhantes. Porém, todos os valores de  $p_L$  foram um pouco inferiores aos correspondentes valores de  $p'_L$ .

Ao colocar os produtos em ordem decrescente dos valores de  $p_L$  e  $p'_L$ , obteve-se a mesma seqüência que a obtida para a exatidão global (INT II, INT III, INT I, VE2, VE1, VE5 e VE7, VE3, VE6, VE4). Isto era esperado, já que estas estimativas referem-se ao valor mínimo do intervalo de confiança de 95% para a exatidão global.

Para testar se os produtos apresentaram a exatidão mínima desejada, escolheu-se o teste de hipóteses proposto por Aronoff (1982) pois este utiliza, em uma única formulação, as duas probabilidades de erro citadas por Rosenfield et al. (1982). Desta forma evitou-se obter dois valores críticos ( $c$ ) para a aceitação dos produtos (item 2.4.1.2).

O teste foi formulado para uma exatidão mínima desejada ( $p_0$ ) de 0,85 e um risco do consumidor ( $\alpha$ ) de 0,05. Nestas condições, o número máximo de erros de classificação permitido ( $x$ ) foi igual a 23. Por este teste, somente a carta elaborada pelo intérprete II, antes de sua introdução no SGI, apresentou a exatidão mínima desejada.

A análise visual desta carta, comparada às cartas elaboradas pelos intérpretes I e III, mostrou que ela continha um maior número de polígonos. Este pode ser um indicador da melhor qualidade deste produto, pois refletiu um maior detalhamento na interpretação visual efetuada.

Para calcular o risco do produtor, associou-se um valor de exatidão  $p_U = 0,90$  à carta elaborada pelo intérprete II. Neste caso, o risco do produtor foi igual a

0,3412. Atribuindo à carta um valor de  $p_0 = 0,95$ , o risco do produtor foi igual a 0,0003.

Os baixos valores encontrados para o risco do produtor indicaram que o tamanho da amostra usado foi adequado.

Para a aplicação do coeficiente de concordância Kappa para estimar as exatidões dos produtos, as condições expressas por Cohen (1960) (item 2.4.1.2) foram satisfeitas.

As unidades amostrais foram consideradas independentes devido à grande distância entre elas. As distâncias foram obtidas convertendo-se as medidas de distância entre unidades amostrais contíguas em número de pixels, sendo que a menor distância calculada correspondeu a 27 pixels.

Alguns testes para verificar a independência das unidades amostrais estabelecem a distância mínima a partir da qual a correlação entre pixels é nula. O uso desses testes não pôde ser viabilizado no presente estudo pois, em geral, se necessita dos valores digitais de níveis de cinza para sua aplicação, os quais não eram disponíveis.

As classes foram consideradas independentes, baseado no que foi exposto por Foody (1988) (item 2.4.1.2).

Para obter classes exclusivas e exaustivas, elas foram definidas de forma a evitar sua sobreposição e a englobar todas as feições contidas na imagem. Para atender

este último quesito, foi criada a classe Vazio, que englobou toda a área da imagem não pertencente às demais classes.

A última condição, a operação independente dos intérpretes, também foi assegurada, na medida em que a interpretação foi realizada isoladamente por cada um deles.

As estimativas do coeficiente Kappa ( $\hat{K}$ ), para os diferentes produtos, são apresentadas na Tabela 4.9, e os respectivos intervalos de confiança de 95%, na Figura 4.2.

**TABELA 4.9 - ESTIMATIVAS DO COEFICIENTE KAPPA. POR PRODUTO**

PRODUTO	COEFICIENTE KAPPA ESTIMADO
INT I	0,7646
INT II	0,8304
INT III	0,7697
VE1	0,6969
VE2	0,7619
VE3	0,6569
VE4	0,6503
VE5	0,6825
VE6	0,6299
VE7	0,6749

Todas as estimativas do coeficiente Kappa foram inferiores às correspondentes estimativas da exatidão global. Porém, a seqüência dos produtos em ordem decrescente das estimativas do Kappa (INT II, INT III, INT I, VE2, VE1, VE5, VE7, VE3, VE4, VE6) foi semelhante àquela referente às da exatidão global (INT II, INT III, INT I, VE2, VE1, VE5 e VE7, VE3, VE6, VE4).

### Coeficiente Kappa

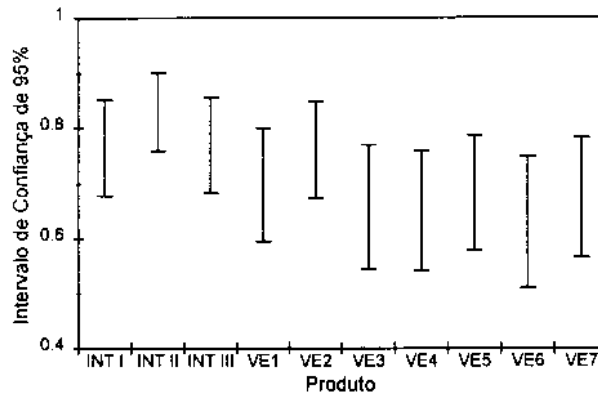


Fig. 4.2 - Intervalos de confiança de 95% para o coeficiente Kappa, por produto.

A Figura 4.2 mostra que as amplitudes dos intervalos de confiança de 95% para o coeficiente Kappa variaram e foram inversamente relacionadas a  $\hat{K}$ . A amplitude se refere à diferença entre os limites máximo e mínimo do intervalo de confiança.

Como a amplitude do intervalo de confiança é função direta da variância do coeficiente Kappa, buscou-se entender quais dos termos da Equação (2.20) apresentaram maior correlação com as estimativas da variância,  $\hat{\sigma}^2(\hat{K})$ . Os coeficientes de correlação<sup>17</sup> são apresentados na Tabela 4.10.

<sup>17</sup> Os coeficientes de correlação foram calculados segundo as equações apresentadas por Sokal e Rohlf (1969, p. 509-510).

TABELA 4.10 - CORRELAÇÃO ENTRE OS TERMOS DA EQUAÇÃO (2.20)  
E A VARIÂNCIA ESTIMADA DO COEFICIENTE KAPPA  
ESTIMADO

TERMOS	CORRELAÇÃO (r)
$\theta_1$	-0,9641
$\theta_2$	0,5389
$\theta_3$	-0,3286
$\theta_4$	0,1179

Os resultados da Tabela 4.10 mostram uma elevada correlação entre o termo  $\theta_1$  (Equação (2.21)), que representa a proporção de concordância (ou estimativa da exatidão global), e  $\hat{\sigma}^2(\hat{K})$ . Isto explica a elevada correlação obtida também entre  $\hat{K}$  e  $\hat{\sigma}^2(\hat{K})$  ( $r = -0,9879$ ), uma vez que a proporção de concordância é um dos componentes do coeficiente Kappa. (Equações (2.16) e (2.17)).

Todas as amplitudes dos intervalos de confiança para o coeficiente Kappa foram maiores que as correspondentes obtidas para a exatidão global. Isto ocorreu porque as estimativas da variância de  $\hat{K}$  ( $\hat{\sigma}^2(\hat{K})$ ) foram maiores que as de  $\hat{p}$  ( $\hat{\sigma}^2_p$ ). Entretanto, isto não implica que  $\hat{p}$  tenha sido um melhor estimador da exatidão de classificação do que  $\hat{K}$ , uma vez que essas estatísticas são obtidas a partir de diferentes concepções: enquanto o cálculo de  $\hat{p}$  considera apenas os elementos da diagonal da matriz de erro, o cálculo de  $\hat{K}$  considera todos os elementos da matriz.

As estimativas do coeficiente Kappa condicional ( $\hat{K}_i$ ) e Kappa ( $\hat{K}_i^*$ ), por classe e por produto, são apresentadas na Tabela 4.11.

TABELA 4.11 - ESTIMATIVAS DO COEFICIENTE KAPPA CONDICIONAL ( $\hat{K}_i$ ) E KAPPA ( $\hat{K}_i^*$ ), POR CLASSE E POR PRODUTO

PRODUTO	KAPPA CONDICIONAL ESTIMADO ( $\hat{K}_i$ )				KAPPA ESTIMADO ( $\hat{K}_i^*$ )			
	MATA	RESTINGA	MANGUE	VAZIO	MATA	RESTINGA	MANGUE	VAZIO
INT I	0,8008	0,3251	0,7682	0,8748	0,8517	0,4308	0,7466	0,7704
INT II	0,9159	0,5600	0,8253	0,8364	0,8970	0,6183	0,7799	0,8440
INT III	0,8363	0,2173	0,7694	0,8946	0,8538	0,3403	0,7694	0,7795
VE1	0,7750	0,2707	0,7088	0,7573	0,7911	0,3889	0,6698	0,6878
VE2	0,8738	0,4971	0,6558	0,7756	0,8558	0,5489	0,6752	0,7614
VE3	0,7376	0,2173	0,4913	0,7973	0,7686	0,3403	0,5539	0,6508
VE4	0,7322	0,4200	0,6505	0,6429	0,7399	0,3946	0,6147	0,6606
VE5	0,8475	0,2672	0,4863	0,7332	0,7875	0,3685	0,5157	0,7063
VE6	0,7376	0,1611	0,4403	0,7668	0,7686	0,2638	0,5292	0,6047
VE7	0,7734	0,2210	0,5433	0,7846	0,7815	0,3620	0,5762	0,6693

A Tabela 4.11 mostra que a classe Restinga apresentou as menores estimativas do coeficiente Kappa condicional e Kappa por categoria, sendo este resultado consistente com aqueles obtidos a partir das matrizes normalizadas e pela análise dos erros de omissão e inclusão.

Para o cálculo de  $\hat{K}_i$ , o termo  $x_{i+}$  da Equação (2.26) referiu-se à soma dos dados de referência para uma dada classe. Como resultado, pôde-se observar que  $\hat{K}_i$  resultou em uma medida associada aos erros de omissão. Embora  $\hat{K}_i$  tenha apresentado valores inferiores aos da exatidão do produtor estimada, a seqüência dos produtos em ordem decrescente dos valores de  $\hat{K}_i$ , por classe, foi



semelhante àquela obtida para as estimativas da exatidão do produtor.

Fung e LeDrew (1988) também observaram a relação entre  $\hat{K}_i$  e a exatidão do produtor ou do consumidor, em função da forma como ele é calculado.

Ao colocar os produtos em ordem decrescente de  $\hat{K}_i^*$ , para cada classe, não se observou semelhanças entre estas seqüências e as obtidas para as estimativas das exatidões por classe: a do produtor, a do consumidor, a obtida a partir das matrizes normalizadas e  $\hat{K}_i$ .

Tanto  $\hat{K}_i^*$  quanto a exatidão por classe obtida a partir da matriz normalizada não são estimativas associadas aos erros de omissão ou inclusão, isoladamente. Porém, não foram observadas semelhanças entre estas estimativas nos resultados obtidos.

O uso de  $\hat{K}_i^*$  como uma estimativa da exatidão por classe apresentou como vantagens:

- não ser uma medida associada aos erros de omissão ou inclusão, isoladamente; e
- a possibilidade de associá-lo ao coeficiente Kappa estimado para toda a matriz, conforme demonstrou Fleiss (1981).

O cálculo proposto por Fleiss (1981) (Equação (2.30)), que estabelece a relação entre  $\hat{K}_i^*$  e  $\hat{K}_i$ ,

foi realizado para os dados do intérprete I, tendo-se verificado a relação entre ambos.

A desvantagem do uso de  $\hat{K}_i^*$  e de  $\hat{K}$  encontra-se na falta de fundamentação teórica para a seleção de um valor mínimo para a aceitação dos produtos. Essa fundamentação deveria estar baseada em experiências acumuladas de estudos sobre a aplicação do coeficiente Kappa a dados de sensoriamento remoto.

#### 4.4 - ANÁLISE COMPARATIVA ENTRE OS PRODUTOS

O primeiro método utilizado para a análise da influência de algumas fontes de erro na exatidão dos produtos foi a comparação entre matrizes normalizadas, conforme proposto por Congalton (1991).

As matrizes normalizadas possibilitaram a visualização das estimativas da exatidão por classe, porém não da exatidão global. Face ainda às divergências entre estimativas de exatidão obtidas a partir destas matrizes e das originais, optou-se por utilizar as estimativas fornecidas pela última.

Utilizando o método proposto por Fleiss (1981), as estimativas de exatidão global dos produtos foram comparadas aos níveis de significância de 0,01 e 0,05.

As hipóteses nulas, os valores calculados para  $\chi^2$  e as conclusões do teste são apresentados na Tabela 4.12. Para as comparações envolvendo três valores de exatidão global, os valores tabelados de  $\chi^2$ , para 2 graus

de liberdade, foram:  $\chi_{0,05}^2 = 5,99$  e  $\chi_{0,01}^2 = 9,2$ . Para as comparações envolvendo dois valores de exatidão global, estes valores, para 1 grau de liberdade, foram:  $\chi_{0,05}^2 = 3,84$  e  $\chi_{0,01}^2 = 6,63$ .

TABELA 4.12 - HIPÓTESES NULAS, VALORES  $\chi^2$  CALCULADOS E CONCLUSÕES OBTIDAS PARA TESTAR A IGUALDADE ENTRE AS EXATIDÕES GLOBAIS DOS PRODUTOS

HIPÓTESE NULA	$\chi^2$ CALCULADO	CONCLUSÃO DO TESTE
$p_1 = p_2 = p_3$	1,5571	NÃO SIGNIFICATIVO
$p_1 = p_{VE1} = p_{VE4}$	4,4583	NÃO SIGNIFICATIVO
$p_2 = p_{VE2} = p_{VE5}$	6,5607	SIGNIFICATIVO AO NÍVEL DE 0,05
$p_3 = p_{VE3} = p_{VE6}$	5,0095	NÃO SIGNIFICATIVO
$p_3 = p_{VE7}$	2,3956	NÃO SIGNIFICATIVO
$p_{VE1} = p_{VE4}$	0,9112	NÃO SIGNIFICATIVO
$p_{VE2} = p_{VE5}$	1,6276	NÃO SIGNIFICATIVO
$p_{VE3} = p_{VE6}$	0,1250	NÃO SIGNIFICATIVO
$p_{VE3} = p_{VE6} = p_{VE7}$	0,3587	NÃO SIGNIFICATIVO

Na Tabela 4.12,  $p_1$ ,  $p_2$  e  $p_3$  correspondem à exatidão global para as cartas temáticas elaboradas pelos intérpretes I, II e III, respectivamente, e  $p_{VE1}, \dots, p_{VE7}$  à exatidão global para as cartas temáticas digitalizadas,  $VE1, \dots, VE7$ .

Os resultados da Tabela 4.12 mostram que não houve diferenças significativas entre os valores de exatidão global para os diferentes intérpretes, para a digitalização automática e a carta original, para os

diferentes operadores do SGI, e para a digitalização manual e a automática.

Ao comparar os resultados da digitalização manual e da carta original, apenas um dos três casos analisados apresentou diferença significativa, ao nível de 0,05 de significância. Este foi o que testou a igualdade entre a exatidão global da carta temática elaborada pelo intérprete II ( $p_2$ ) às das digitalizações manuais desta carta ( $p_{VE2}$  e  $p_{VE5}$ ).

A análise do intervalo de confiança de 95% para a exatidão global destes três produtos (Figura 4.3) permitiu visualizar quais intervalos apresentaram a maior diferença entre si. Esta diferença foi verificada entre o intervalo de confiança para a exatidão global gerado para o produto elaborado pelo intérprete II (INT II) e para a sua digitalização pelo operador II (VE5).

Pelo terceiro método, proposto por Congalton et al. (1983), testou-se a igualdade entre pares de valores do coeficiente Kappa. Os valores  $z$  calculados encontram-se na Tabela 4.13. Os valores tabelados de  $z$  foram:  $z_{0,025} = 1,96$  e  $z_{0,005} = 2,576$ .

### Exatidão Global

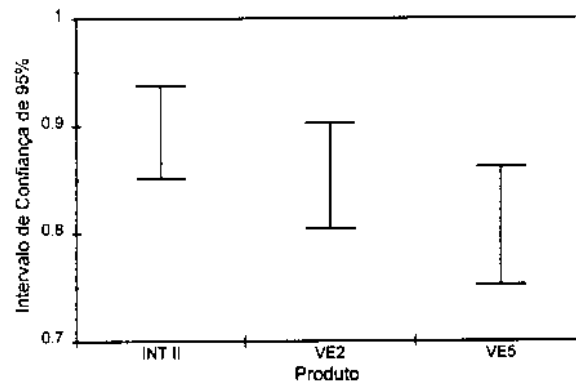


Fig. 4.3 - Intervalos de confiança de 95% para a exatidão global dos produtos: INT II, VE2 e VE5.

TABELA 4.13 - VALORES  $z$  CALCULADOS PARA TESTAR A IGUALDADE ENTRE PARES DE VALORES DO COEFICIENTE KAPPA DOS PRODUTOS

	$K_1$	$K_2$	$K_3$	$K_{VE1}$	$K_{VE2}$	$K_{VE3}$	$K_{VE6}$
$K_2$	1,1476	---	---	---	---	---	---
$K_3$	0,0815	1,0626	---	---	---	---	---
$K_{VE1}$	0,9880	---	---	---	---	---	---
$K_{VE2}$	---	1,1939	---	---	---	---	---
$K_{VE3}$	---	---	1,5593	---	---	---	---
$K_{VE4}$	1,6140	---	---	0,6130	---	---	---
$K_{VE5}$	---	2,2958*	---	---	1,1448	---	---
$K_{VE6}$	---	---	1,8582	---	---	0,3231	---
$K_{VE7}$	---	---	1,3407	---	---	0,2254	0,5466

\* Significativo ao nível de 0,05.

Nas Tabelas 4.13 e 4.14,  $K_1$ ,  $K_2$  e  $K_3$  correspondem aos valores do coeficiente Kappa para as cartas temáticas elaboradas pelos intérpretes I, II e III, respectivamente, e  $K_{VE1}, \dots, K_{VE7}$  aos valores do coeficiente Kappa para as cartas temáticas digitalizadas, VE1, ..., VE7.

Somente foi significativa, ao nível de 0,05 de significância, a diferença entre os valores do coeficiente Kappa da carta temática elaborada pelo intérprete II ( $K_2$ ) e a digitalização manual desta pelo operador II ( $K_{vts}$ ). Este resultado foi compatível com o resultado do teste anterior, que comparou os valores de exatidão global dos produtos.

Estes dois produtos apresentaram, portanto, as maiores diferenças entre valores de medidas de exatidão temática. Em parte, isto pode ter ocorrido em função da maior dificuldade em digitalizar a carta elaborada pelo intérprete II, devido à maior complexidade desta quando comparada às demais. Esta maior complexidade foi atribuída ao maior número de polígonos presente nesta carta.

Conforme discutido no item 2.4.1.2.1, o teste entre pares de valores de  $K$  permite comparar apenas dois valores, ao nível de significância  $\alpha$ . Ao se comparar mais de dois valores de  $K$ , o nível de significância não se mantém igual a  $\alpha$ . Face a essas considerações, este teste não permitiu generalizar as conclusões para as fontes de erro analisadas no presente trabalho.

O quarto método empregado, proposto por Fleiss (1981) permitiu a comparação entre dois ou mais valores de  $K$ .

As hipóteses nulas, os valores calculados para  $\chi^2$  e as conclusões do teste são fornecidos na Tabela 4.14. Para as comparações entre três valores de Kappa, os

valores tabelados de  $\chi^2$ , para 2 graus de liberdade, foram:  $\chi_{0,05}^2 = 5,99$  e  $\chi_{0,01}^2 = 9,2$ . Para as comparações envolvendo dois valores de Kappa, estes valores, para 1 grau de liberdade, foram:  $\chi_{0,05}^2 = 3,84$  e  $\chi_{0,01}^2 = 6,63$ .

As conclusões dos testes de hipóteses expressas na Tabela 4.14 mostram que não houve evidências, ao nível de significância de 0,05, da influência das diferentes fontes de erro investigadas nos valores do coeficiente Kappa.

O teste para comparar a carta temática elaborada pelo intérprete II e os planos VE2 e VE5 não foi significativo, ao nível de 0,05 de significância, para os valores do coeficiente Kappa, embora tenha sido para as exatidões globais.

A predominância de resultados não significativos ao comparar os produtos pode ser devida à simplicidade da legenda utilizada para a elaboração dos mapas temáticos, constituída apenas por quatro classes.

TABELA 4.14 - HIPÓTESES NULAS, VALORES  $\chi^2$  CALCULADOS E CONCLUSÕES OBTIDAS PARA TESTAR A IGUALDADE ENTRE OS VALORES DO COEFICIENTE KAPPA DOS PRODUTOS

HIPÓTESE NULA	$\chi^2$ CALCULADO	CONCLUSÃO DO TESTE
$K_1 = K_2 = K_3$	1,7472	NÃO SIGNIFICATIVO
$K_1 = K_{VE1} = K_{VE4}$	2,7381	NÃO SIGNIFICATIVO
$K_2 = K_{VE2} = K_{VE5}$	5,4316	NÃO SIGNIFICATIVO
$K_3 = K_{VE3} = K_{VE6}$	4,3710	NÃO SIGNIFICATIVO
$K_3 = K_{VE7}$	1,7976	NÃO SIGNIFICATIVO
$K_{VE1} = K_{VE4}$	0,3757	NÃO SIGNIFICATIVO
$K_{VE2} = K_{VE5}$	1,3106	NÃO SIGNIFICATIVO
$K_{VE3} = K_{VE6}$	0,1044	NÃO SIGNIFICATIVO
$K_{VE3} = K_{VE6} = K_{VE7}$	0,2999	NÃO SIGNIFICATIVO

Embora não tenha havido diferenças significativas entre os produtos gerados através da digitalização automática e manual, a automática (VE7) foi mais rápida que a manual (VE3). Enquanto a primeira consumiu 12,5 horas, a segunda durou cerca de 30 horas.



## CAPÍTULO 5

### CONCLUSÕES E RECOMENDAÇÕES

O uso da amostragem estratificada sistemática não alinhada (AESNA) apresentou como vantagem o fato de não exigir, nos cálculos de exatidão de classificação, o uso de ponderações referentes à área de cada classe.

A amostragem de 1% da área, tomando-se como unidade amostral a área do terreno correspondente ao pixel, mostrou-se inviável para a escala do trabalho (1:50000). A grade necessária para a AESNA, de forma a se amostrar 1% da área, seria composta por células muito pequenas, inviabilizando a localização precisa das unidades amostrais.

O tamanho da amostra adotado resultou em uma matriz de erro com baixa frequência de zeros. Este tamanho foi, portanto, compatível com o número de classes utilizado no mapeamento temático. Outra constatação da adequação do tamanho da amostra foram os baixos valores do risco do produtor calculados no teste de hipóteses aplicado para verificar se os produtos apresentaram a exatidão mínima desejada.

O método elaborado para a coleta de dados de referência reduziu bastante o trabalho de campo e possibilitou a determinação da classe de todas as unidades amostrais selecionadas. Porém, este método somente pôde ser viabilizado face à facilidade de acesso a diversas fontes

de informação sobre a área, como fotografias aéreas e levantamentos sobre a vegetação.

Ao comparar os resultados obtidos para a análise da exatidão de classificação através dos métodos utilizados, observou-se que eles foram coerentes, exceto no caso das matrizes normalizadas. Ao colocar os produtos em ordem decrescente de suas estimativas de exatidão, para cada método empregado, a ordem observada para a exatidão global estimada a partir das matrizes normalizadas foi diferente das referentes aos demais métodos. Porém, como não há como calcular os intervalos de confiança para a exatidão global obtida a partir das matrizes normalizadas, não há como afirmar se esta diferença foi significativa.

As estimativas da exatidão mínima calculadas a partir da aproximação normal à binomial e a partir da distribuição binomial apresentaram valores muito próximos. Porém, as últimas foram um pouco superiores para todos os produtos.

As estimativas do coeficiente Kappa para todos os produtos foram inferiores às correspondentes estimativas da exatidão global. A desvantagem do uso do Kappa encontra-se na falta de fundamentação teórica para a seleção de um valor mínimo desse coeficiente para a aceitação dos produtos. A definição de valores mínimos do coeficiente Kappa para a aceitação dos produtos deveria estar baseada em experiências acumuladas de estudos sobre a aplicação desse coeficiente a dados de sensoriamento remoto.

A partir da análise das amplitudes dos intervalos de confiança para o coeficiente Kappa e para a exatidão global e, baseando-se nas estimativas das variâncias do coeficiente Kappa estimado e da exatidão global amostral, pode-se dizer que a exatidão global amostral foi um estimador mais eficiente da exatidão global do que o coeficiente Kappa estimado foi do coeficiente Kappa.

Segundo o teste de hipóteses aplicado, somente a carta temática elaborada pelo intérprete II apresentou a exatidão mínima desejada, ou seja, maior ou igual a 0,85, considerando o risco do consumidor igual a 0,05.

As análises de exatidão por classe mostraram que as estimativas da exatidão do produtor e do coeficiente Kappa condicional foram associadas aos erros de omissão, e a estimativa da exatidão do consumidor, aos erros de inclusão. Os resultados destas análises foram muito diferentes, em função do tipo de erro em que estavam associados. Pôde-se observar, para a mesma classe de um produto, baixas estimativas da exatidão do produtor e do coeficiente Kappa condicional e altas estimativas da exatidão do consumidor.

As estimativas da exatidão por classe obtidas a partir das matrizes normalizadas não foram semelhantes aos resultados de nenhum outro método de análise de exatidão por classe. Embora, em tese, elas devessem refletir uma ponderação dos erros de omissão e

inclusão, como resultado do processo de normalização, isto não foi observado em alguns casos.

O uso da estimativa do coeficiente Kappa, para cada classe, como um estimador da exatidão destas, apresentou como vantagens:

- não ser uma medida associada aos erros de omissão ou inclusão, isoladamente; e
- a possibilidade de associá-la ao coeficiente Kappa estimado para toda a matriz.

O uso do teste para comparar pares de valores do coeficiente Kappa dos produtos apresentou como desvantagem a limitação desta comparação a somente dois valores deste coeficiente. Assim, este teste não pôde ser utilizado para a análise da influência de algumas fontes de erro em que o número de valores do coeficiente Kappa comparados foi maior que dois.

As fontes de erro analisadas não influíram de forma significativa nos valores de exatidão dos produtos elaborados.

Somente foram observadas diferenças significativas, ao nível de 0,05 de significância, em dois casos:

- entre a carta elaborada pelo intérprete II (INT II) e sua digitalização pelo operador II (VE5) ao se aplicar o teste para comparar pares de valores do coeficiente Kappa; e

- entre os produtos: INT II, VE2 (digitalização desta carta pelo operador I) e VE5, ao se aplicar o teste  $\chi^2$  para comparar valores de exatidão global dos produtos.

No segundo caso, a análise dos intervalos de confiança de 95% para a exatidão global dos três produtos mostrou que a maior diferença encontrava-se entre os intervalos de confiança referentes aos produtos INT II e VE5.

A diferença entre os produtos INT II e VE5, que mostrou-se significativa em alguns testes, pode ser devida, em parte, à maior dificuldade em digitalizar a carta elaborada pelo intérprete II. Esta dificuldade está relacionada à maior complexidade desta carta temática quando comparada às demais, sendo esta maior complexidade atribuída ao maior número de polígonos presentes.

Embora não tenham sido observadas diferenças significativas na exatidão dos produtos gerados através da digitalização automática e manual, a automática (plano VE7) consumiu 12,5 horas, enquanto que a manual (plano VE3) utilizou cerca de 30 horas.

Para a determinação do tamanho da amostra e da distância mínima entre unidades amostrais de forma a assegurar a independência entre elas recomenda-se a aplicação de algum teste para verificar a correlação existente entre os níveis de cinza dos pixels separados por diferentes distâncias.



REFERÊNCIAS BIBLIOGRÁFICAS

- Alexander, H.W. Statistical inference. In: \_\_\_\_\_  
**Elements of mathematical statistics**. New York, John  
 Wiley & Sons, 1961. cap. 4, p. 176-210.
- Anderson, J.R. Land-use classification schemes.  
**Photogrammetric Engineering**, 37(4):379-387, Apr. 1971.
- Araújo, D.S.D.; Lacerda, L.D. A natureza das restingas.  
**Ciência Hoje**, 6(33):42-48, jul. 1987.
- Aronoff, S. Classification accuracy: a user approach.  
**Photogrammetric Engineering and Remote Sensing**,  
 48(8):1299-1307, Aug. 1982.
- \_\_\_\_\_ The minimum accuracy value as an index of  
 classification accuracy. **Photogrammetric Engineering  
 and Remote Sensing**, 51(1):99-111, Jan. 1985.
- \_\_\_\_\_ Data input and output. In: \_\_\_\_\_ **Geographic  
 information systems: a management perspective**. Ottawa,  
 WDL Publications, 1989. cap. 4, p. 103-132.
- Batista, G.T.; Medeiros, J.S.; Mello, E.M.K.; Moreira,  
 J.C.; Bins, L.S. A new approach for deforestation  
 assessment. In: Symposium on Resource and Environmental  
 Monitoring. ISPRS Commission VII, Rio de Janeiro, Sept.  
 26-30, 1994. **International archives of photogrammetry  
 and remote sensing**. São José dos Campos, INPE, 1994, v.  
 30, Part 7a, p. 170-174.

- Bishop, Y.M.M.; Fienberg, S.E.; Holland, P.W. Measures of association and agreement. In: \_\_\_\_\_ **Discrete multivariate analysis: theory and practice.** Cambridge, Mass., MIT, 1975. cap. 11, p. 373-400.
- Brennan, R.L.; Prediger, D.J. Coefficient Kappa: some uses, misuses, and alternatives. **Educational and Psychological Measurement**, 41: 687-699, 1981.
- Brownlee, K.A. **Statistical theory and methodology: in science and engineering.** 2.ed. New York, John Wiley & Sons, 1965.
- Burrough, P.A. **Principles of geographical information systems for land resources assessment.** Oxford, Clarendon Press, 1986. (Monographs on Soil and Resources Survey, 12).
- Caldeira, H.S. **Classificação temática de áreas litorâneas: uma contribuição metodológica.** (Dissertação de Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1994. (INPE-5586-TDI/548).
- Câmara, G. Anatomia de sistemas de informações geográficas: visão atual e perspectivas de evolução. In: Assad, E.D.; Sano, E.E. **Sistema de informações geográficas: aplicações na agricultura.** Planaltina, EMBRAPA-CPAC, 1993. cap. 1, p. 15-37.



Card, D.H. Using known map category marginal frequencies to improve estimates of thematic map accuracy.

**Photogrammetric Engineering and Remote Sensing**, 48(3):431-439, Mar. 1982.

Cohen, J. A coefficient of agreement for nominal scales.

**Educational and Psychological Measurement**, 20(1):37-46, 1960.

Companhia de Tecnologia de Saneamento Ambiental (CETESB).

**Baixada Santista**: carta do meio ambiente e de sua dinâmica. São Paulo, 1985. Acompanha memorial descritivo e carta em escala 1:50000.

Congalton, R.G. Using spatial autocorrelation analysis to

explore the errors in maps generated from remotely sensed data. **Photogrammetric Engineering and Remote Sensing**, 54(5):587-592, May 1988a.

\_\_\_\_\_ 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**, 54(5):593-600, May 1988b.

\_\_\_\_\_ A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, 37(1):35-46, Jul. 1991.

- 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**, 58(12):1669-1671, Dec. 1992.
- Congalton, R.G.; Green, K. A practical look at sources of confusion in error matrix generation. **Photogrammetric Engineering and Remote Sensing**, 59(5):641-644, May 1993.
- Congalton, R.G.; Mead, R.A. A quantitative method to test for consistency and correctness in photointerpretation. **Photogrammetric Engineering and Remote Sensing**, 49(1):69-74, Jan. 1983.
- Congalton, R.G.; Mead, R.A.; Oderwald, R. G.; Heinen, J. **Analysis of forest classification accuracy**. Blacksburg, VA, Virginia Polytechnic Institute and State University, 1981. (NASA-CR-160949).
- Congalton, R.G.; Oderwald, R.G.; Mead, R.A. Assessing LANDSAT classification accuracy using discrete multivariate analysis statistical techniques. **Photogrammetric Engineering and Remote Sensing**, 49(12):1671-1678, Dec. 1983.
- Costa, J.P.O.; Corrêa, F. A reserva da biosfera da Mata Atlântica. **São Paulo em Perspectiva**, 6(1/2):106-111, jan./jun. 1992.
- Costa Neto, P.L.O. Estatística descritiva. In: \_\_\_\_\_ **Estatística**. São Paulo, Edgard Blücher, 1977. cap. 2, p. 5-38.

- Costa, S.M.F. **Avaliação de técnicas de processamento digital de imagens TM-LANDSAT aplicadas à delimitação de áreas urbanas.** (Dissertação de Mestrado em Sensoriamento Remoto) - Instituto de Pesquisas Espaciais, São José dos Campos, 1990. (INPE-5071-TDL/410).
- Covre, M.; Faria, K.; Fidalgo, E.C.C.; Hahn, C.M.; Matheus, D.R. São Paulo state natural vegetation monitoring. In: International Congress for Photogrammetry and Remote Sensing, 17., Washington, Aug. 2-14, 1992. **International archives of photogrammetry and remote sensing.** Washington, ISPRS, 1992, p. 302-306.
- Dallemand, J.F. **Identificação de culturas de inverno por interpretação visual de dados SPOT e LANDSAT TM no Noroeste do Paraná.** (Dissertação de Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1987. (INPE-4334-TDL/290).
- Damião, D.P. **Análise comparativa de imagens multissensores (TM/LANDSAT e SAR/ERS-1) e multitemporais na região de Guaiúra, SP.** (Dissertação de Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1994. (INPE-5581-TDI/544).

- Estes, J.E.; Hajic, E.J.; Tinney, L.R. Fundamentals of image analysis: analysis of visible and thermal infrared data. In: American Society of Photogrammetry. **Manual of remote sensing**. 2.ed. Falls Church, VA, 1983. v. 1, cap. 24, p. 987-1124.
- Felgueiras, C.A.; Câmara, G. Sistema de informações geográficas do INPE. In: Assad, E.D.; Sano, E.E. **Sistema de informações geográficas: aplicações na agricultura**. Planaltina, EMBRAPA-CPAC, 1993. cap. 2, p. 41-59.
- Fitzgerald, R.W.; Lees, B.G. Assessing the classification accuracy of multisource remote sensing data. **Remote Sensing of Environment**, 47(3):362-368, Mar. 1994.
- Fitzpatrick-Lins, K. Comparison of sampling procedures and data analysis for a land-use and land-cover map. **Photogrammetric Engineering and Remote Sensing**, 47(3):343-351, Mar. 1981.
- Fleiss, J.L. **Statistical methods for rates and proportions**. New York, John Wiley & Sons, 1981.
- Foody, G.M. The effects of viewing geometry on image classification. **International Journal of Remote Sensing**, 9(12):1909-1915, Dec. 1988.

Fundação Instituto Brasileiro de Geografia e Estatística (IBGE). Departamento de Recursos Naturais e Estudos Ambientais. **Manual técnico da vegetação brasileira**. Rio de Janeiro, IBGE, 1992. (Série Manuais Técnicos em Geociências, 1).

Fundação SOS Mata Atlântica/Instituto Nacional de Pesquisas Espaciais (INPE). **Evolução dos remanescentes florestais e ecossistemas associados do domínio da Mata Atlântica no período 1985-1990**. Convênio Fundação SOS Mata Atlântica/INPE. São José dos Campos, INPE, 1993.

Fung, T.; LeDrew, E. The determination of optimal threshold levels for change detection using various accuracy indices. **Photogrammetric Engineering and Remote Sensing**, 54(10):1449-1454, Oct. 1988.

Garcia, G.J. **Sensoriamento remoto: princípios e interpretação de imagens**. São Paulo, Nobel, 1982. cap. 2, p. 24-46.

Gausman, H.W. Plant leaf optical properties in visible and near-infrared light. Texas Tech University. **Graduate Studies**, (29):7-78, Feb. 1985.

Ginevan, M.E. Testing land-use map accuracy: another look. **Photogrammetric Engineering and Remote Sensing**, 45(10):1371-1377, Oct. 1979.

- Goel, N.S. Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data. **Remote Sensing Reviews**, 4(1):1-222, 1988.
- Gong, P.; Howarth, P.J. An assessment of some factor influencing multispectral land-cover classification. **Photogrammetric Engineering and Remote Sensing**, 56(5):597-603, May 1990.
- Hahn, C.M.; Schwenck Júnior, P.M.; Mendonça, F.J.; Faria, K.; Rodrigues, R.L.V.; Hernandez Filho, P. Levantamento da vegetação natural do Estado de São Paulo. In: Simpósio Brasileiro de Sensoriamento Remoto, 6., Manaus, 24-29, jun. 1990. **Anais**. São José dos Campos, INPE, 1990, v. 1, p. 117-125. (INPE-8250-PRE/250).
- Hay, A.M. Sampling designs to test land-use map accuracy. **Photogrammetric Engineering and Remote Sensing**, 45(4):529-533, Apr. 1979.
- Hernandez Filho, P. **Metodologia da análise visual de dados de sensoriamento remoto: vegetação**. São José dos Campos, INPE, 1988. (INPE-4696-MD/037).
- Herz, R. **Manguezais do Brasil**. São Paulo, Instituto Oceanográfico/USP, 1991.
- Hollander, M.; Wolfe, D.A. The one-sample location problem. In: \_\_\_\_\_ **Nonparametric statistical methods**. New York, John Wiley & Sons, 1973. cap. 13, p. 26-66.

Hord, R.M.; Brooner, W. Land-use map accuracy criteria. **Photogrammetric Engineering and Remote Sensing**, 42(5):671-677, May 1976.

IMAGEM Sensoriamento Remoto. **O processo de digitalização automática**. São José dos Campos, mar. 1995.  
Comunicação pessoal.

Instituto Brasileiro de Geografia e Estatística (IBGE). **Santos**. Rio de Janeiro, 1984. Carta topográfica; colorida. 1:50000; folha SF 23-Y-D-IV-3, SG 23-V-B-I-I; MI2794/3, MI2815/1.

Instituto de Botânica/Instituto Florestal/Companhia de Tecnologia de Saneamento Ambiental (IBt/IF/CETESB). **A Serra do Mar: degradação e recuperação**. São Paulo, SMA, 1990. (Série Documentos ISSN 0103-264).

Instituto Geográfico e Cartográfico (IGC). **Plano cartográfico do Estado de São Paulo**. São Paulo, 1987. Cartas planialtimétricas; branco e preto. 1:10000.

Kalensky, Z.D. **Accuracy of land use and vegetation mapping from digital remote sensing data**. Rome, Remote Sensing Center, 1985. Presented at the International Training Course on Small Scale Remote Sensing Applications to Assist Rural Development, Feldafing, FRG, 1985.

Knipling, E.B. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. **Remote Sensing of Environment**, 1(3):155-159, 1970.

- Kirizawa, M.; Lopes, E.A.; Pinto, M.M.; Lam, M.; Lopes, M.I.M.S. Vegetação da Ilha Comprida: aspectos fisionômicos e florísticos. **Revista do Instituto Florestal**, 4(parte 2):386-391, , mar. 1992. Apresentado no Congresso Nacional sobre Essências Nativas, 2., São Paulo, 29 mar. - 3 abr., 1992.
- Klein, R.M. Estrutura, composição florística, dinamismo e manejo da "Mata Atlântica" (Floresta Ombrófila Densa) do sul do Brasil. In: Simpósio sobre Ecossistemas da Costa Sul e Sudeste Brasileira, 2., Águas de Lindóia, 6-11, abr. 1990. **Anais**. 1990, v. 1, p. 259-313.
- Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. **Biometrics**, 33:159-174, Mar. 1977.
- Leitão Filho, H.F., org. **Ecologia da Mata Atlântica em Cubatão, SP**. São Paulo, Editora da Universidade Estadual Paulista, Campinas, Editora da Universidade de Campinas, 1993.
- Mantovani, W.; Rodrigues, R.R.; Rossi, L.; Romaniuc-Neto, S.; Catharino, E.L.M.; Cordeiro, I. A vegetação da Serra do Mar em Salesópolis, SP. In: Simpósio sobre Ecossistemas da Costa Sul e Sudeste Brasileira, 2., Águas de Lindóia, 6-11, abr. 1990. **Anais**. 1990, v. 1, p. 348-384.



- Marsh, S.E.; Walsh, J.L.; Sobrevila, C. Evaluation of airborne video data for land-cover classification accuracy assessment in an isolated Brazilian forest. **Remote Sensing of Environment**, 48(1):61-69, Apr. 1994.
- Mead, R.A.; Szajgin, J. Landsat classification accuracy assessment procedures. **Photogrammetric Engineering and Remote Sensing**, 48(1):139-141, Jan. 1982.
- Mulder, N.J. Digital image processing, computer-aided classification and mapping. In: Kùchler, A.W.; Zonneveld, I.S., ed. **Vegetation mapping**. Dordrecht, Kluwer Academic Publishers, 1988. cap. 21, p. 269-316. (Handbook of Vegetation Science, v.10)
- Neter, J.; Wasserman, W. Analysis of factor effects. In: \_\_\_\_\_ **Applied Linear Statistical Models: regression, analysis of variance, and experimental designs**. Homewood, IL, Richard D. Irwin Inc., 1974. cap. 14, p. 458-491.
- Novo, E.M.L.M. Métodos de extração de informações. In: \_\_\_\_\_ **Sensoriamento remoto: princípios e aplicações**. São Paulo, Edgard Blùcher, 1989. cap. 7, p. 205-265.
- Pereira, M.N.; Kurkdjian, M.L.N.O.; Foresti, C. **Cobertura e uso da terra através de sensoriamento remoto**. São José dos Campos, INPE, 1989. (INPE-5032-MD/042).

- Prisley, S.P.; Smith, J.L. Using classification error matrices to improve the accuracy of weighted land-cover models. **Photogrammetric Engineering and Remote Sensing**, 53(9):1259-1263, Sept. 1987.
- Queiroz Neto, J.P.; Küpper, A. Os solos. In: Azevedo, A. **A Baixada Santista: aspectos geográficos**. São Paulo, EDUSP, 1964. v. 1, cap. 3, p. 67-92.
- Ripley, B.D. **Spatial statistic**. New York, John Wiley & Sons, 1981.
- Rocha, J.V. **The influence of ground survey size on accuracy of area estimates from satellite images**. (PhD Thesis) - Silsoe College, England, 1992.
- Rosenfield, G.H.; Fitzpatrick-Lins, K. A coefficient of agreement as a measure of thematic classification accuracy. **Photogrammetric Engineering and Remote Sensing**, 52(2):223-227, Feb. 1986.
- Rosenfield, G.H.; Fitzpatrick-Lins, K.; Ling, H.S. Sampling for thematic map accuracy testing. **Photogrammetric Engineering and Remote Sensing**, 48(1):131-137, Jan. 1982.
- Rosenfield, G.H.; Melley, M.L. Applications of statistics to thematic mapping. **Photogrammetric Engineering and Remote Sensing**, 46(10):1287-1294, Oct. 1980.

São Paulo. Secretaria do Meio Ambiente (SMA), ed.

**Mapeamento de vegetação natural do Estado de São Paulo - Santos.** São Paulo, 1989. carta temática; branco e preto. 1:50000.

Saxena, K.G.; Tiwari, A.K.; Porwal, M.C.; Menon, A.R.R.

Vegetation maps, mapping needs and scope of digital processing of LANDSAT Thematic Mapper data in tropical region of south-west India. **International Journal of Remote Sensing**, 13(11):2017-2037, Nov. 1992.

Scarim, J.L.; Teixeira, A.L.A. Digitalização e conversão

raster/vector de mapas. **Fator GIS**, 2(6):16-21, jul-set. 1994.

Skidmore, A.K.; Turner, B.J. Map accuracy assessment using

line intersect sampling. **Photogrammetric Engineering and Remote Sensing**, 58(10):1453-1457, Oct. 1992.

Snedecor, G.W.; Cochran, W.G. **Statistical methods.** 6.ed.

Ames, Iowa, IA State University Press, 1967.

Sokal, R.R.; Rohlf, F.J. **Biometry.** San Francisco, W. H.

Freeman and Company, 1969.

Spiegel, M.R. **Probabilidade e estatística.** Trad. de

Alfredo Alves de Faria. São Paulo, McGraw-Hill do Brasil, 1978.

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**, 58(9):1343-1350, Sept. 1992.

- Story, M.; Congalton, R.G. Accuracy assessment: a user's perspective. **Photogrammetric Engineering and Remote Sensing**, 52(3):397-399, Mar. 1986.
- Tardin, A.T.; Lee, D.C.L.; Santos, R.J.R.; Assis, O.R.; Barbosa, M.P.S.; Moreira, M.L.; Pereira, M.T.; Silva, D.; Santos Filho, C.P. **Subprojeto Desmatamento**. Convênio IBDF/CNPq/INPE 1979. São José dos Campos, INPE, 1980. (INPE-1649-RPE/103).
- United States Geological Survey (USGS). National Oceanic and Atmospheric Administration (NOAA). **LANDSAT 4 data users handbook**. Alexandria, USGS, 1984.
- Van Genderen, J.L.; Lock, B.F.; Vass, P.A. Remote sensing: statistical testing of thematic map accuracy. **Remote Sensing of Environment**, 7(1):3-14, 1978.
- Veneziani, P. **Interpretação visual de dados**: roteiro de estudos e definições fundamentais. São José dos Campos, 1988. Apostila da disciplina Interpretação Visual de Dados do curso de Mestrado em Sensoriamento Remoto do Instituto Nacional de Pesquisas Espaciais, São José dos Campos.
- Watson, A.I.; Vaughan, R.A.; Powell, M. Classification using the watershed method. **International Journal of Remote Sensing**, 13(10):1881-1890, Jul. 1992.

## APÊNDICE A

## LISTAGEM DAS CARTAS E FOTOGRAFIAS AÉREAS ONDE FORAM LOCALIZADAS AS UNIDADES AMOSTRAIS

PONTO	CARTA*	FAIXA/FOTO**	PONTO	CARTA*	FAIXA/FOTO**
1	---	11/488; 11/489	41	99/108	2/27; 2/28
2	---	12/573; 12/574	42	99/108	2/27; 2/28
3	---	12/573; 12/574	43	99/109	3/27; 3/28
4	---	2/26	44	100/108	4A/3; 4A/4
5	99/108	2/25; 2/26	45	100/108	4A/3; 4A/4
6	99/108	2/26; 2/27	46	101/108	4A/2; 4A/3
7	100/108	2/26	47	101/108	5/25; 5/26
8	100/108	2/26	48	102/108	5/24; 5/25
9	100/108	4A/1; 4A/2	49	102/108	5/24; 5/25
10	101/108	4A/1; 4A/2	50	102/108	5/24; 5/25
11	101/108	4A/1; 4A/2	51	103/108	5/23; 5/24
12	102/108	5/23; 5/24	52	---	---
13	102/108	5/23; 5/24	53	---	---
14	103/108	5/23; 5/24	54	---	2/29; 2/30
15	103/108	5/23; 5/24	55	---	2/29; 2/30
16	---	---	56	98/109	2/28; 2/29
17	---	---	57	98/109	2/28; 2/29
18	---	11/488; 11/489	58	99/109	2/27; 2/28
19	---	11/488	59	99/109	3/27; 3/28
20	---	11/488	60	100/109	4A/4; 4A/5
21	---	2/27; 2/28	61	100/109	4A/4; 4A/5
22	---	2/26; 2/27	62	100/109	4A/4; 4A/5
23	99/108	2/26; 2/27	63	101/109	5/25; 5/26
24	99/108	2/26; 2/27	64	101/109	5/25; 5/26
25	100/108	3/25; 3/26	65	102/109	5/25; 5/26
26	100/108	4A/2; 4A/3	66	102/109	5/25; 5/26
27	100/108	4A/2; 4A/3	67	103/109	5/24
28	101/108	4A/3; 4A/4	68	---	---
29	101/108	5/24; 5/25	69	---	---
30	102/108	5/23; 2/24	70	---	---
31	102/108	5/24; 5/25	71	---	2/30; 2/31
32	103/108	5/23; 5/24	72	---	2/30; 2/31
33	103/108	5/23; 5/24	73	98/109	2/29; 2/30
34	---	---	74	98/109	2/29; 2/30
35	---	---	75	98/109	2/29
36	---	11/488	76	99/109	3/28; 3/29
37	---	11/488	77	99/109	3/28; 3/29
38	---	2/28; 2/29	78	100/109	4A/4; 4A/5
39	---	2/28; 2/29	79	100/109	4A/4; 4A/5
40	---	2/28; 2/29	80	101/109	5/26; 5/27

\* A listagem dos nomes das cartas citadas (IGC, 1987) encontra-se no Apêndice B.

\*\* As fotografias aéreas das faixas 2, 3, 4A, 5, 6 e 7 pertencem ao arquivo do Instituto Geográfico e Cartográfico de São Paulo (IGC). São fotografias, em filme pancromático preto e branco, escala aproximada de 1:35000, do levantamento aéreo realizado na Baixada Santista em 1986 e 1987. As fotografias aéreas das faixas 11 e 12 pertencem ao arquivo do IGC e são do levantamento realizado na região da Grande São Paulo e Baixada Santista contratado pela Empresa Metropolitana de Planejamento da Grande São Paulo (EMPLASA), de junho e julho de 1977, em escala aproximada de 1:40000.

## APÊNDICE A

## CONTINUAÇÃO

PONTO	CARTA*	FAIXA/FOTO**	PONTO	CARTA*	FAIXA/FOTO**
81	101/109	5/26; 5/27	121	---	---
82	101/109	5/26; 5/27	122	---	---
83	102/109	5/25; 5/26	123	97/110	2/32
84	102/109	5/25; 5/26	124	97/110	2/32
85	103/109	6/5; 6/6	125	98/110	2/31; 2/32
86	---	---	126	98/110	3/30; 3/31
87	---	---	127	98/110	3/30; 3/31
88	97/110	2/31; 2/32	128	99/110	4A/7; 4A/8
89	---	2/31; 2/32	129	99/110	4A/7; 4A/8
90	98/109	2/30; 2/31	130	100/110	4A/6; 4A/7
91	98/109	3/29; 3/30	131	100/110	5/29
92	99/109	3/28; 3/29	132	101/110	5/29
93	99/109	3/28; 3/29	133	101/110	5/29
94	100/109	4A/5; 4A/6	134	101/110	5/27
95	100/109	4A/5; 4A/6	135	---	---
96	100/109	4A/5; 4A/6	136	---	---
97	101/109	5/26; 5/27	137	---	---
98	101/109	5/26; 5/27	138	---	---
99	102/109	5/26; 5/27	139	---	---
100	102/109	5/26; 5/27	140	97/110	2/33
101	---	---	141	97/110	2/32
102	---	---	142	98/110	2/32
103	---	---	143	98/110	3/31; 3/32
104	---	---	144	99/110	4A/7; 4A/8
105	---	2/31; 2/32	145	99/110	4A/7; 4A/8
106	97/110	2/31; 2/32	146	99/110	4A/7; 4A/8
107	97/110	2/31; 2/32	147	100/110	5/29; 5/30
108	98/110	2/31; 2/32	148	100/110	5/29; 5/30
109	98/110	3/30; 3/31	149	101/110	5/29
110	99/110	4A/7; 4A/8	150	101/110	5/29
111	99/110	4A/7; 4A/8	151	102/110	5/27; 5/29
112	99/110	4A/7; 4A/8	152	---	---
113	100/110	4A/5; 4A/6	153	---	---
114	100/110	5/27; 5/29	154	---	---
115	101/110	5/27; 5/29	155	---	---
116	101/110	5/27	156	---	---
117	102/110	5/26; 5/27	157	97/111	3/33
118	---	---	158	97/111	3/33
119	---	---	159	98/111	3/32
120	---	---	160	98/111	4A/8; 4A/9

\* A listagem dos nomes das cartas citadas (IGC, 1987) encontra-se no Apêndice B.

\*\* As fotografias aéreas das faixas 2, 3, 4A, 5, 6 e 7 pertencem ao arquivo do Instituto Geográfico e Cartográfico de São Paulo (IGC). São fotografias, em filme pancromático preto e branco, escala aproximada de 1:35000, do levantamento aéreo realizado na Baixada Santista em 1986 e 1987. As fotografias aéreas das faixas 11 e 12 pertencem ao arquivo do IGC e são do levantamento realizado na região da Grande São Paulo e Baixada Santista contratado pela Empresa Metropolitana de Planejamento da Grande São Paulo (EMPLASA), de junho e julho de 1977, em escala aproximada de 1:40000.

## APÊNDICE A

## CONCLUSÃO

PONTO	CARTA*	FAIXA/FOTO**	PONTO	CARTA*	FAIXA/FOTO**
161	99/111	4A/8; 4A/9	190	---	---
162	99/111	4A/8; 4A/9	191	97/111	3/34
163	100/111	5/30	192	97/111	3/34
164	100/110	5/29; 5/30	193	98/111	4A/9
165	100/111	5/29; 5/30	194	98/111	4A/9
166	101/111	5/29; 5/30	195	99/111	4A/9
167	101/111	5/29; 5/30	196	99/111	5/31
168	102/110	6/8; 6/9	197	100/111	5/30; 5/31
169	102/111	6/8; 6/9	198	100/111	5/30; 5/31
170	103/111	7/2; 7/3	199	101/111	5/30; 5/31
171	103/111	7/2; 7/3	200	101/111	5/30; 5/31
172	---	---	201	101/111	6/10; 6/11
173	---	---	202	102/111	6/10; 6/11
174	97/111	3/34	203	102/111	7/5
175	97/111	3/33	204	---	---
176	98/111	3/33	205	---	---
177	98/111	4A/9	206	---	---
178	98/111	4A/9	207	---	---
179	99/111	4A/8; 4A/9	208	97/111	3/34
180	99/111	5/30; 5/31	209	---	11/488; 11/489
181	100/111	5/30; 5/31	210	---	11/488; 11/489
182	100/111	5/30; 5/31	211	98/111	4A/10
183	101/111	5/29; 5/30	212	98/111	4A/10
184	101/111	5/29; 5/30	213	99/108	2/25; 2/26
185	102/111	6/9; 6/10	214	100/111	5/31
186	102/111	6/9; 6/10	215	100/111	5/31
187	102/111	6/9; 6/10	216	102/111	6/10; 6/11
188	103/111	7/2; 7/3	217	102/108	5/24; 5/25
189	---	---	218	---	---

\* A listagem dos nomes das cartas citadas (IGC, 1987) encontra-se no Apêndice B.

\*\* As fotografias aéreas das faixas 2, 3, 4A, 5, 6 e 7 pertencem ao arquivo do Instituto Geográfico e Cartográfico de São Paulo (IGC). São fotografias, em filme pancromático preto e branco, escala aproximada de 1:35000, do levantamento aéreo realizado na Baixada Santista em 1986 e 1987. As fotografias aéreas das faixas 11 e 12 pertencem ao arquivo do IGC e são do levantamento realizado na região da Grande São Paulo e Baixada Santista contratado pela Empresa Metropolitana de Planejamento da Grande São Paulo (EMPLASA), de junho e julho de 1977, em escala aproximada de 1:40000.





APÊNDICE BRELAÇÃO DAS CARTAS PLANIALTIMÉTRICAS DO PROJETO BAIXADA  
SANTISTA (IGC, 1987)

<b>NÚMERO</b>	<b>NOME</b>
97/110	CAMPO GRANDE
97/111	PARANAPIACABA
98/109	SERRA DO POÇO
98/110	RAIZ DA SERRA
98/111	SERRA DO QUILOMBO
99/108	CAMINHO DO MAR
99/109	VILA PARISI
99/110	RIO QUILOMBO
99/111	RIO JURUBATUBA
100/108	RIO ITUTINGA
100/109	CUBATÃO
100/110	ILHAS DOS BAGRES
100/111	MORRO DO GABRIEL
101/108	RIO SANTANA
101/109	SÃO VICENTE II
101/110	SANTOS
101/111	VICENTE DE CARVALHO
102/108	RIO PIAÇABUÇU
102/109	SÃO VICENTE
102/110	ILHA PORCHAT
102/111	GUARUJÁ
103/108	PRAIA GRANDE II
103/109	PRAIA GRANDE
103/111	GUARUJÁ V



APÊNDICE CPROGRAMA PARA NORMALIZAÇÃO DE MATRIZES

Este programa foi desenvolvido em linguagem C, baseado em Congalton et al. (1981).

```
#include <stdio.h>
#include <math.h>
main()
{

float
tab[50][50], rmar[50], cmar[50], fit[50], maxdev, b, d, h, xh;
int  ncount, nummat, ir, ic, i, j, nit, maxit;
FILE  *fp, *fp1;

fp=fopen("matriz.dat", "rt");
fp1=fopen("result.dat", "wt");

ncount = 0;

printf("\n Entre numero de matrizes a serem alteradas : ");
scanf("%d", &nummat);
label2:
printf("\n Entre numero de linhas da matriz : ");
scanf("%d", &ir);
printf("\n Entre numero de colunas da matriz : ");
scanf("%d", &ic);

for(i=0; i<ir; i++)
    for(j=0; j<ic; j++)
        fscanf(fp, "%f", &tab[i][j]);

for(i=0; i<ir; i++)
    rmar[i] = 1.0;

for(i=0; i<ic; i++)
    cmar[i] = 1.0;

printf("\n Entre com o numero maximo de iteracoes : ");
scanf("%d", &maxit);
```

```

printf("\n Entre com desvio maximo aceitavel : ");
scanf("%f",&maxdev);

printf("\n\nMatriz Original :\n ");

for(i=0;i<ir;i++)
{
    for(j=0;j<ic;j++)
        printf("%10.5f",tab[i][j]);
    printf("\n");
}
nit = 0;

for(i=0;i<ir;i++)
{
    fit[i]=0.0001;
    for (j=0;j<ic;j++)
        fit[i]+=tab[i][j];
}
labell:
for (i=0;i<ir;i++)
    for (j=0;j<ic;j++)
        tab[i][j] *= rmar[i]/fit[i];

for (j=0;j<ic;j++)
{
    b=0.0001;
    for (i=0;i<ir;i++)
        b+=tab[i][j];

    for (i=0;i<ir;i++)
    {
        if (tab[i][j] < 1.0e-6) tab[i][j]=0.0;
        tab[i][j]*=cmar[j]/b;
    }
}

nit+=1;
d=0.0;
for (i=0;i<ir;i++)
{
    fit[i]=0.0001;
    for (j=0;j<ic;j++)
        fit[i]+=tab[i][j];
}

```

```

    xh = fit[i] - rmar[i];
    h = xh;
    printf("\n%f",h);
    if(h < 0.0) h= -1.0 * h;
    if (d>=h)continue;
    d=h;
}
if (d<=maxdev)
{
    printf("\nConvergencia apos %d iteracoes",maxit);
    printf("\ncom o desvio maximo de %f",d);

    for (i=0;i<ir;i++)
    {
        for (j=0;j<ic;j++)
            printf("%f",tab[i][j]);
        printf("\n");
    }
}
printf("\n%f",d);
if(nit<=maxit)
    goto label1;
else
{
    printf ("\nNenhuma convergencia apos %5d
iteracoes",nit);
    printf ("\nO desvio maximo corrente e : %10.7f\n",d);
}
ncount+=1;

if (ncount<nummat) goto label2;
fclose (fp);
fclose (fp1);
}

```