



Ministério da
Ciência e Tecnologia



INPE-15661-TDI/1437

ESTATÍSTICAS AGRÍCOLAS POR AMOSTRAGEM AUXILIADAS PELO SENSORIAMENTO REMOTO

Alfredo José Barreto Luiz

Tese de Doutorado do Curso de Pós-Graduação em Sensoriamento Remoto,
orientada pelo Dr. José Carlos Neves Epiphânio, aprovada em 3 de dezembro de
2003.

Registro do documento original:

<<http://urlib.net/sid.inpe.br/jeferson/2004/02.04.16.08>>

INPE
São José dos Campos
2009

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3945-6911/6923

Fax: (012) 3945-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO:

Presidente:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Haroldo Fraga de Campos Velho - Centro de Tecnologias Especiais (CTE)

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Jefferson Andrade Ancelmo - Serviço de Informação e Documentação (SID)

Simone A. Del-Ducca Barbedo - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Marilúcia Santos Melo Cid - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Viveca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



Ministério da
Ciência e Tecnologia



INPE-15661-TDI/1437

ESTATÍSTICAS AGRÍCOLAS POR AMOSTRAGEM AUXILIADAS PELO SENSORIAMENTO REMOTO

Alfredo José Barreto Luiz

Tese de Doutorado do Curso de Pós-Graduação em Sensoriamento Remoto,
orientada pelo Dr. José Carlos Neves Epiphânio, aprovada em 3 de dezembro de
2003.

Registro do documento original:

<<http://urlib.net/sid.inpe.br/jeferson/2004/02.04.16.08>>

INPE
São José dos Campos
2009

Dados Internacionais de Catalogação na Publicação (CIP)

Luiz, Alfredo José Barreto.

L969e Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto / Alfredo José Barreto Luiz. – São José dos Campos : INPE, 2009.

112 p. ; (INPE-15661-TDI/1437)

Tese (Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2003.

Orientador : Dr. José Carlos Neves Epiphânio.

1. Imagens de satélite. 2. Sistema de Informação Geográfica. 3. Estatísticas agrícolas. 4. LANDSAT. 5. Agricultura. 6. Cana-de-açúcar. 7. Soja. 8. Área plantada. 9. Amostragem. I.Título.

CDU 631:519:22:528

Copyright © 2009 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, reprográfico, de microfilmagem ou outros, sem a permissão escrita do INPE, com exceção de qualquer material fornecido especificamente com o propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2009 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, microfilming, or otherwise, without written permission from INPE, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

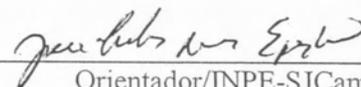
Aprovado pela Banca Examinadora em cumprimento a requisito exigido para a obtenção do Título de **Doutor** em **Sensoriamento Remoto**.

Dr. Antônio Roberto Formaggio



Presidente/INPE-SJCampos/SP

Dr. José Carlos Neves Epiphânio



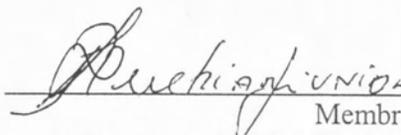
Orientador/INPE-SJCampos/SP

Dr^a Corina da Costa Freitas



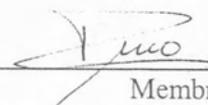
Membro da Banca/INPE-SJCampos/SP

Dr. Ariovaldo Luchiari Junior



Membro da Banca
Convidado EMBRAPA, Jaguariúna/SP

Dr. Francisco Alberto Pino



Membro da Banca
Convidado IEA, São Paulo/SP

Candidato (a): Alfredo José Barreto Luiz

São José dos Campos, 3 de dezembro de 2003.

“O ser individual é imprevisível, porém as reações das multidões podem ser tratadas estatisticamente. Quanto maior a multidão, tanto maior a precisão que se pode conseguir”.

ISAAC ASIMOV, no livro *Fundação*, escrito entre 1942 e 1949.

“Inexiste no mundo coisa mais bem distribuída que o bom senso, visto que cada indivíduo acredita ser tão bem provido dele que mesmo os mais difíceis de satisfazer em qualquer outro aspecto não costumam desejar possuí-lo mais do que já possuem. E é improvável que todos se enganem a esse respeito; mas isso é antes uma prova de que o poder de julgar de forma correta e discernir entre o verdadeiro e o falso, que é justamente o que é denominado bom senso ou razão, é igual em todos os homens; e, assim sendo, de que a diversidade de nossas opiniões não se origina do fato de serem alguns mais racionais que outros, mas apenas de dirigirmos nossos pensamentos por caminhos diferentes e não considerarmos as mesmas coisas. Pois é insuficiente ter o espírito bom, o mais importante é aplicá-lo bem. As maiores almas são capazes dos maiores vícios, como também das maiores virtudes, e os que só andam muito devagar podem alcançar bem mais, se continuarem sempre pelo caminho reto, do que aqueles que correm e dele se afastam”.

RENÉ DESCARTES, na primeira parte do *Discurso do Método*, 1637.

*À minha família,
MARCIA, CAROLINA,
JAIME, MAYARA E MATHEUS.*

AGRADECIMENTOS

Ao meu Orientador, Dr. José Carlos Neves Epiphanyo que, através de inúmeras conversas que seriam infundáveis não fosse o telefone que nos lembrava de voltar para casa, mais do que me ensinar, me permitiu aprender. Espero sinceramente corresponder à confiança que ele depositou em mim.

Às minhas irmãs, Bea, Bete, Lena e Gina, e à minha tia Vina, por me sustentarem financeiramente durante o curso de graduação, de 1981 a 1985, sem o qual eu não teria chegado a esse doutorado.

À Salette Gürtler, que me auxiliou com sua competência e eficiência.

Aos colegas Júlio César Oliveira e José Marinaldo Gleriani, companheiros até no barro e na poeira dos trabalhos de campo.

Ao Dr. Antônio Roberto Formaggio, que me incentivou a transformar idéias em publicações, pela forma amistosa e confortante com que sempre me recebeu.

Aos amigos Rogério Costa Campos e Stélio Soares Tavares Júnior, pelos papos na cantina no fim de tarde, que ajudavam a desembaralhar as idéias.

Às colegas Ieda Del'Arco Sanches, Helen da Costa Gurgel, Mariana Abrantes Giannotti e ao Professor Nelson Jesus Ferreira, pela agradável parceria que resultou em alguns trabalhos publicados.

À incansável Etel e sua fiel escudeira Angelucci, por toda a trabalhadeira que lhes dei e pelo carinho, além do profissionalismo, com que sempre me trataram.

À Cristina e à Sabrina, pela atenção com que me atendiam e, especialmente, pelo alto astral permanente.

A todos os professores, pesquisadores, funcionários e colegas do INPE, pelo privilégio da convivência nesses últimos quatro anos.

Agradeço também à Fapesp, que financiou parte dos trabalhos de campo, através do Projeto nº 99/09485-8; ao INPE, por oferecer essa maravilhosa oportunidade de me tornar doutor em uma área tecnológica de ponta como o Sensoriamento Remoto, sem precisar sair do Brasil; e à Embrapa, empresa da qual me orgulho em fazer parte, por investir tanto na formação dos seus pesquisadores.

Finalmente, um agradecimento muito especial é devido ao meu amigo Clayton Campanhola e à Psicóloga Rosana Pires da Silva, que, cada um a seu modo, desempenharam papéis determinantes para o término deste trabalho. Aos dois, o meu muito obrigado.

RESUMO

A utilização correta da estatística em qualquer área de aplicação depende, primordialmente, do grau de conhecimento da natureza dos dados empregados e do nível de clareza dos objetivos estabelecidos. O sensoriamento remoto oferece um conjunto muito particular de dados, com características que precisam ser levadas em consideração na escolha dos métodos estatísticos a serem utilizados na sua análise. A aplicação destes dados em agricultura, em especial quando visa à delimitação e à quantificação de área das culturas agrícolas, define objetivos bem específicos que devem influir na determinação das técnicas de análise estatística. A união teoricamente embasada do sensoriamento remoto e da estatística, de forma adequada à estimativa de área plantada, resulta em um avanço na eficiência do uso dos dados de sensores orbitais para fins agrícolas. Nesse sentido, este documento apresenta: 1) um método para preparação e utilização de imagens obtidas por satélite para serem usadas em levantamentos agrícolas por amostragem; 2) a forma de cálculo das estimativas objetivas de área plantada e suas respectivas variâncias; 3) um estudo de caso que consiste na estimativa da área plantada com as principais culturas agrícolas, em âmbito municipal, através do uso de dados obtidos por sensoriamento remoto; e 4) a expansão do método, por simulação, para a escala estadual e nacional.

AGRICULTURAL STATISTICS BY SAMPLING WITH THE AID OF REMOTE SENSING

ABSTRACT

The proper utilization of statistics in any application area depends primarily on the degree of knowledge of the nature of employed data, as well as on the level of clarity of established goals. Remote sensing offers a rather particular set of data that bear characteristics that must be taken into consideration when choosing the statistical methods to be used in analysis. Application of those data in agriculture will, particularly when the aim is the delimitation and quantification of crop areas, define quite specific goals that should influence on the selection of statistical analysis techniques. The theoretically-based alliance between remote sensing and statistics in a way that is adequate to estimate crop areas, results in a step ahead in the efficient use of data from orbital sensors for agriculture aims. In this direction, this document brings: 1) a method to prepare and use satellite images in agricultural surveys by sampling; 2) the way to calculate objective estimates of crop area and their respective variances; 3) a case study consisting in the estimate of total planted area in a municipality, considering all major crops, with the use of data from remote sensing; and 4) the up scaling of method, by simulation, to regional and national level.

SUMÁRIO

Pág.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

PREFÁCIO

CAPÍTULO 1 - INTRODUÇÃO.....	23
1.1 - Contextualização e Histórico	23
1.2 - Objetivos	27
1.3 - Pressupostos	28
1.4 - Estrutura	29
CAPÍTULO 2 - FUNDAMENTOS TEÓRICOS	31
2.1 - Estimativas Agrícolas e Relações com o Sensoriamento Remoto	31
2.2 - Culturas Agrícolas como Alvo do Sensoriamento Remoto	33
2.3 - Análise Estatística	38
2.3.1 - Opção pela Amostragem	40
2.3.2 - Amostragem Aleatória Simples	42
2.3.3 - Construção de uma Amostra Aleatória	45
2.3.3.1 - Hipóteses	46
2.3.3.2 - Demonstração	47
2.3.4 - Estratificação	48
2.3.4.1 - Estratificação dentro de um Município	49
2.3.4.2 - Estratificação entre Municípios	52
CAPÍTULO 3 - MATERIAL E MÉTODOS	57
3.1 - Área de Estudo	57
3.2 - Construção da Referência	66
3.3 - Método de Amostragem por Pontos no Município	71
3.4 - Estratificação no Município	77
3.5 - Estratificação em um Conjunto de Municípios	80
CAPÍTULO 4 - RESULTADOS E DISCUSSÃO	85
4.1 - Dados de Referência	85
4.2 - Estimativas por Amostragem Aleatória Simples em Ipuã	86
4.3 - Estimativas por Amostragem Aleatória Estratificada em Ipuã	89
4.4 - Estimativa por Amostragem Aleatória Estratificada Simulada	91

CAPÍTULO 5 - CONCLUSÕES	95
REFERÊNCIAS BIBLIOGRÁFICAS	97
APÊNDICE A	105
APÊNDICE B	109

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 - Curva que representa a relação entre o tamanho da amostra (n), e o coeficiente de variação esperado (CV %), dada a probabilidade $p = 0,3468$ de ocorrência da cultura na área.	44
3.1 - Localização e classes de solo do município de Ipuã – SP	57
3.2 - Local de coleta dos 13 pontos de referência usados no registro das imagens.	58
3.3 - Frequência de ND na banda 1 (a) e sua função de crescimento (b).....	60
3.4 - Imagem RGB 453 do município de Ipuã – SP, de 14 de agosto de 2001.....	64
3.5 - Imagem RGB 453 do município de Ipuã – SP, de 05 de janeiro de 2002.	65
3.6 - Plano de informação vetorial contendo os limites dos segmentos identificados em Ipuã – SP, em agosto de 2001.	68
3.7 - Mapa temático do município de Ipuã – SP, de 14 de agosto de 2001, com as classes de uso do solo identificadas.....	69
3.8 - Mapa temático do município de Ipuã – SP, de 05 de janeiro de 2002, com as classes de uso do solo identificadas.....	70
3.9 - Exemplo de distribuição de 100 pontos, escolhidos para compor uma amostra aleatória, em Ipuã – SP.....	74
3.10 - Exemplo de distribuição de 200 pontos, escolhidos para compor uma amostra aleatória, em Ipuã – SP.....	75
3.11 - Exemplo de distribuição de 300 pontos, escolhidos para compor uma amostra aleatória, em Ipuã – SP.....	76
3.12 - Exemplo de distribuição de 200 pontos, escolhidos para compor uma amostra aleatória estratificada, com 114 pontos no estrato I e 86 no estrato II, em Ipuã – SP.....	78
3.13 - Exemplo de distribuição de 200 pontos, escolhidos para compor uma amostra aleatória estratificada, com 86 pontos no estrato I e 114 no estrato II, em Ipuã – SP.....	79

4.1- Área estimada para a cultura da soja, em amostras de 100 (a), 200 (b) e 300 (c) pontos, distribuídos aleatoriamente no município de Ipuã, SP, em janeiro de 2002.....	88
--	----

LISTA DE TABELAS

	<u>Pág.</u>
2.1 - Área plantada, em hectares, das três principais culturas agrícolas nos municípios abrangidos pelo Escritório de Desenvolvimento Regional (EDR) de Orlândia - SP, e participação percentual das três na área agrícola total dos municípios na safra 1995/1996 (Pino et al., 1997).	36
2.2 - Área média, em hectares, das unidades de produção ocupadas com as três principais culturas agrícolas nos municípios abrangidos pelo Escritório de Desenvolvimento Regional (EDR) de Orlândia - SP na safra 1995/1996 (Pino et al., 1997).	37
3.1 - Valores do ângulo de elevação solar, pixel escuro, reflectância e tipo de atmosfera para as imagens utilizadas nos levantamentos.	61
4.1 - Área ocupada com diferentes classes de uso do solo, em hectares, no município de Ipuã - SP, em quatro datas.	85
4.2 - Área ocupada com diferentes classes de uso do solo, em hectares, estimada com três tamanhos de amostra (média de dez repetições), no município de Ipuã - SP, em duas datas.	87
4.3 - Comparação entre as estimativas oficiais, as obtidas por amostragem e a referência, para as principais culturas no município de Ipuã, na safra 2001/2002.	89
4.4 - Área plantada e CV de cana-de-açúcar e soja, em Ipuã, estimados por AAE.	91
4.5 - Amostra estratificada de municípios paulistas, com área plantada e área municipal em hectares e proporção p da área ocupada com a cana-de-açúcar, dados de 2001.	92
4.6 - Dados de simulação do sorteio de uma AAS de tamanho 100 em cada município.	93
4.7 - Número total de municípios N_h e tamanho da amostra n_h , área com cana-de-açúcar (em hectares), variância, desvio padrão e CV (em %), por estrato.	94
A.1 - Quantidade produzida, em toneladas, de todas as culturas registradas pelo IBGE/PAM, para o município de Ipuã - SP, entre 1990 e 2001.	106

A.2 - Área colhida, em hectares, de todas as culturas registradas pelo IBGE/PAM, para o município de Ipuã – SP, entre os anos de 1990 e 2001.	107
A.3 - Valor não corrigido da produção, em moeda da época*, de todas as culturas registradas pelo IBGE/PAM, para o município de Ipuã – SP, entre os anos de 1990 e 2001.....	108

LISTA DE SIGLAS E ABREVIATURAS

AA	– Amostra Aleatória
AAE	– Amostragem Aleatória Estratificada
AAS	– Amostragem Aleatória Simples
AGRIT	– Progetti Di Innovazione Nell'ambito Delle Statistiche Agricole
CAP	– Community Agricultural Policy
CV	– Coeficiente de Variação
DSR	– Divisão de Sensoriamento Remoto
EAGGF	– European Agricultural Guidance and Guarantee Fund
EAGRI	– projeto Estatísticas Agrícolas
EC	– European Community
EDR	– Escritório de Desenvolvimento Regional
ETM+	– Enhanced Thematic Mapper Plus
f.d.p.	– função de densidade de probabilidade
GPS	– Global Positioning System
IBGE	– Instituto Brasileiro de Geografia e Estatística
ICMS	– Imposto sobre a Circulação de Mercadorias e Serviços
i.i.d.	– independente e identicamente distribuída
INPE	– Instituto Nacional de Pesquisas Espaciais
MARS	– Monitoring Agriculture with Remote Sensing
ND	– Número Digital
PAM	– Produção Agrícola Municipal
PI	– Plano de Informação
PREVS	– Pesquisa Objetiva de Previsão de Safras
SPE	– Subtração do Pixel Escuro
SIAG	– Sistema de Informações Agropecuárias
SIDRA	– Sistema IBGE de Recuperação Automática
SIG	– Sistema de Informação Geográfica

v.a.

– variável aleatória

WRS

– Worldwide Reference System

PREFÁCIO

As duas citações escolhidas para iniciar esse trabalho representam, além de duas fortes influências sobre minha formação intelectual na figura de seus autores, duas importantes crenças que me acompanham já há algum tempo. Uma delas é a crença na precisão das estimativas estatísticas, crença que não é baseada na fé e sim na aceitação das demonstrações matemáticas da Lei Forte dos Grandes Números e do Teorema Central do Limite (Souza, 1998). A outra é a da inteira confiança no bom senso, coisa típica de agrônomo, que me leva sempre a procurar um caminho mais simples, mais prático, mais reto, para que, mesmo andando devagar, consiga chegar longe. São esses os dois pilares dessa tese. A confiança no poder da estatística e a intenção, que se pretende boa, de fazer as coisas de maneira simples.

Desde as primeiras aulas do curso de doutorado em Sensoriamento Remoto deparei-me com um conjunto de conhecimentos totalmente novos para mim, e que julgo terem um lugar estratégico e de destaque no desenvolvimento tecnológico nacional. Ao descobrir os mistérios das radiações eletromagnéticas, suas interações com os corpos e sua captura pelos sensores, percebi as imensas possibilidades de aplicação desse conhecimento, na forma de inovações tecnológicas, na minha área original de interesse profissional, a pesquisa agropecuária.

É evidente para mim hoje que, em um País com mais de 850.000.000 ha de área, distribuídos por dois hemisférios e quatro fusos-horários, com uma diversidade ambiental enorme e uma vocação agrícola variada, um instrumento que permita o monitoramento continuado, periódico e à distância de todo o território nacional, tem de ser desenvolvido, estudado, aplicado, incentivado e intensamente utilizado. Esse instrumento é o Sensoriamento Remoto Orbital. Na mesma linha de raciocínio, admito que, mesmo com o uso das imagens de satélite, é forçosa a utilização de métodos estatísticos de amostragem que permitam a obtenção de informações objetivas e acuradas com custos mínimos. Entre os principais usuários desta combinação de instrumento e método estará sempre, com certeza, a agropecuária, juntamente com as atividades de monitoramento ambiental, tão ligadas às atividades agrícolas.

CAPÍTULO 1

INTRODUÇÃO

1.1 Contextualização e Histórico

Atualmente, cerca de 800 milhões de pessoas no mundo não se alimentam de forma satisfatória, sendo que destas, 55 milhões estão na América Latina (Anda, 2003). Entre os motivos para esse quadro desolador estão: a má distribuição de renda e de terras, a má qualidade nutricional de alguns tipos de alimentos tradicionalmente utilizados e, ainda, a falta de estabilidade na produção e distribuição dos produtos agrícolas em determinadas regiões.

Diante desse cenário, o acesso oportuno à informação agrícola confiável tem se tornado cada dia mais importante para os processos de tomada de decisão nacionais e internacionais. O estabelecimento e o desenvolvimento de programas de levantamento de dados do setor é um componente importante para qualquer sistema de informação agrícola. Métodos de levantamento agrícola que considerem o crescente aumento da disponibilidade de imagens de sensoriamento remoto, bem como outros avanços tecnológicos, por exemplo, Sistemas de Posicionamento Global (GPS) e Sistemas de Informações Geográficas (SIGs), podem vir a se tornar as formas mais práticas para um país produzir dados básicos sobre as principais mercadorias agrícolas (FAO, 1996). Gao (2002) afirma que o advento da tecnologia GPS tem não só aumentado a facilidade e versatilidade de aquisição de dados espaciais, como tem também diversificado os enfoques pelos quais ela pode ser integrada com o sensoriamento remoto e os SIGs. Segundo esse último autor, essa integração tem sido extensivamente estudada para o mapeamento de terras agrícolas, tendo-se mostrado especialmente valiosa nas aplicações que requerem dados obtidos em tempo real ou quase em tempo real.

O uso do sensoriamento remoto, associado aos SIGs, além de facilitar a obtenção de estimativas de área plantada e, futuramente, da produção das principais culturas, permite o acesso a uma outra informação essencial em muitas aplicações, especialmente as ambientais, qual seja, a localização das áreas cultivadas (Epiphanyo, 1997). O cruzamento, via mapas e através dos SIGs, destas informações de uso atual do solo agrícola com dados de declividade,

tipo de solo, pluviométricos, entre outros, permitirá uma enorme gama de aplicações, alimentando modelos matemáticos de crescimento de plantas, disponibilidade de água no solo, dinâmica de nutrientes e de agrotóxicos, etc. Esta será também uma poderosa ferramenta para que o Estado, as Organizações não Governamentais, Agentes Financeiros, etc., possam fiscalizar o cumprimento de leis de uso do solo e dos recursos hídricos, elaborar zoneamentos agroambientais, ordenar a aplicação de recursos, e assim por diante.

Enquanto a utilização dos SIGs já se faz presente em diversos campos das aplicações agrícolas (Assad e Sano, 1998), apesar do seu potencial inquestionável, até o momento pode-se afirmar que o único espaço dentro dos levantamentos agrícolas onde o sensoriamento remoto realmente se estabeleceu, foi o da construção de painéis amostrais por área e, mesmo assim, surpreendentemente apenas um país, os EUA, o incorporou definitivamente, e de forma operacional, ao seu programa de estatísticas agrícolas oficiais (FAO, 1998).

Na Europa, para atender às necessidades da Política Agrícola Comum (CAP), da Comunidade Européia (EC), foi proposto o projeto denominado Monitoramento Agrícola por Sensoriamento Remoto (MARS) que teve como primeiro objetivo distinguir, identificar e medir a área das culturas agrícolas mais importantes, e em segundo lugar estimar a produção a tempo de permitir a tomada de decisões (Klersy, 1992). O MARS tornou-se o suporte técnico do Fundo Europeu de Direcionamento e Garantia da Agricultura (EAGGF), que subsidia os agricultores, conforme as normas da CAP, para tentar reduzir a super produção de certas culturas dentro da EC. Como o subsídio é pago em função da área plantada declarada pelo agricultor, é fácil entender a importância de um sistema de checagem confiável, rápido e barato. A principal diferença entre o método usado no MARS e o utilizado nos EUA é que os contornos das unidades básicas de amostragem não são definidos por limites físicos no terreno e sim traçados artificialmente na forma de quadrados de 50×50 km ou círculos de 25 km de raio, de maneira a serem compatíveis com uma imagem individual do SPOT (Terres et al., 1995). Individualmente, alguns países europeus também estão tentando incorporar a tecnologia do sensoriamento remoto aos seus levantamentos agrícolas, principalmente devido à pressão da EC, através do EAGGF, por informações padronizadas. Eles geralmente baseiam-se nos métodos do projeto MARS, como a Itália (Meyer-Roux e King, 1992), a

Espanha (FAO, 1998), a Grécia (Tsiligirides, 1998), Portugal, Bélgica, Checoslováquia e Romênia (MARS, 1993).

No caso específico da Itália, o Ministério da Política Agrícola e Florestal, com a colaboração técnica do Consórcio Italiano para o Sensoriamento Remoto na Agricultura (Consórcio ITA), promove já há 13 anos um programa denominado AGRIT, com a finalidade de produzir as estatísticas agrícolas (área e previsão de rendimento) de algumas culturas agrícolas consideradas de interesse estratégico nacional, através do emprego da tecnologia de satélites aplicada à agricultura, integrada à metodologia de levantamentos de campo e auxiliada por um modelo agrometeorológico de previsão de rendimento (modelo SAM). Todo ano o Consórcio ITA fornece estas estimativas através da publicação de boletins específicos decendiais, a partir do mês de abril. A partir de 2000 esses dados passaram a estar disponíveis também por via eletrônica no próprio sítio do Consórcio (<http://www.itacon.it>). Entre outras informações, os boletins apresentam a estimativa de área plantada (em escala regional) para as culturas de milho, girassol, soja, beterraba açucareira, tomate e fumo (PREVISIONI, 2000).

No Brasil, compete à União organizar e manter os serviços oficiais de estatística, inclusive as estatísticas agrícolas, cuja realização é atribuída, por delegação, ao Instituto Brasileiro de Geografia e Estatística (IBGE). O IBGE tem mantido tanto um levantamento sistemático da produção agrícola, com dados obtidos de forma subjetiva, através de consulta a especialistas, por município, como um censo agropecuário, de periodicidade variável, com informação colhida, através de entrevistas, por estabelecimento rural (Collares et al., 1993).

A deficiência destas formas de coleta de dados, embora continue sendo apontada em trabalhos recentes (Ippoliti-Ramilo, 1998; Pino, 1999b), já é reconhecida há muito mais tempo. Ainda nos anos setentas, Batista et al. (1978b), referindo-se à necessidade brasileira de conhecimento da produção agrícola anual, considerava que a metodologia utilizada nas previsões de safra não satisfazia às exigências devido à sua subjetividade e à magnitude dos erros.

A incorporação de avanços tecnológicos alcançados em diversas áreas do conhecimento aos procedimentos utilizados nos levantamentos das estatísticas agrícolas nacionais pode vir a se configurar como um poderoso impulsionador no sentido de aprimorar, em objetividade,

rapidez e confiabilidade, as estimativas oficiais da safra brasileira. Nesse sentido, é importante lembrar que, já no planejamento do programa Landsat, no final dos anos sessentas e início dos anos setentas, o Brasil foi um dos países escolhidos para possuir e operar uma estação terrestre de recebimento dos sinais do satélite e um centro de processamento e distribuição de imagens, e o Instituto Nacional de Pesquisas Espaciais (INPE) foi o órgão encarregado de construir e administrar essas instalações. Com isso, desde o lançamento do Landsat 1 o Brasil tem acesso aos dados de sensoriamento remoto através do INPE. Com uma equipe onde havia membros com formação em ciências agrárias e estatística, conscientes das deficiências dos levantamentos agrícolas nacionais, e que estavam em contato estreito com o que ocorria nos EUA, logo surgiram na Divisão de Sensoriamento Remoto (DSR) do INPE os primeiros trabalhos voltados para a previsão de safras utilizando dados orbitais.

Então, a partir de 1975, o INPE e a Secretaria de Agricultura do Estado de São Paulo passaram a desenvolver o projeto Estatísticas Agrícolas (EAGRI), dentro do qual se realizaram diversos levantamentos parciais, ora abrangendo diversas culturas em um único município, ora avaliando uma única lavoura em todo o Estado. Com o crescimento da experiência do INPE, foram expandidos os trabalhos, com atuações em outros Estados e culturas como café e trigo no Paraná, cana-de-açúcar no Rio de Janeiro, trigo no Rio Grande do Sul (Tardin et al., 1976; Batista et al., 1978a e 1978b; Mendonça et al., 1979; Assunção e Duarte, 1980; Mendonça, 1981; Mendonça et al., 1981; Costa et al., 1983). O amadurecimento das idéias, fruto da convivência com trabalhos de cunho prático e operacional, permitiu inclusive a proposição de avanços metodológicos originais, como no trabalho de Krug e Yanasse (1986).

Com a repercussão destes trabalhos e a capacitação de diversos técnicos brasileiros nos cursos de pós-graduação em Sensoriamento Remoto do INPE, passa a existir uma massa crítica nacional na área, suficiente para que, em 1986, fosse estabelecido um convênio entre o IBGE e o INPE visando à implantação do Sistema de Informações Agropecuárias (SIAG). O SIAG pretendia ser um levantamento em larga escala, baseado em métodos probabilísticos, técnicas de sensoriamento remoto e processamento geográfico das informações. Como em outros países, o trabalho foi inicialmente delineado para apenas parte do território nacional, abrangendo os Estados de São Paulo, Santa Catarina, Paraná e o Distrito Federal. O projeto,

que passou a se denominar Pesquisa Objetiva de Previsão de Safras (PREVS), só chegou a utilizar imagens do satélite Landsat 5 na fase de construção dos estratos, embora se pretendesse vir a utilizá-las como variável auxiliar dos estimadores por regressão. Embora algumas tentativas de desenvolver novas metodologias tenham sido realizadas (Medeiros et al., 1996), ao invés da evolução esperada, tanto em termos metodológicos como territoriais, o projeto até recentemente se mantinha nos moldes originais e era aplicado apenas no Estado do Paraná (FAO, 1998; Ippoliti-Ramilo et al., 1999).

O desafio que ainda se coloca, portanto, é o do estabelecimento de um método que permita associar de maneira definitiva a tecnologia disponibilizada pelo sensoriamento remoto orbital aos procedimentos rotineiramente utilizados nos levantamentos das estatísticas agrícolas no Brasil.

1.2 Objetivos

É objetivo geral desse trabalho definir um método simples e confiável, baseado no uso de imagens de sensoriamento e na amostragem estatística, que permita a localização, a identificação e a quantificação das áreas ocupadas por uma determinada cultura agrícola em escala municipal. Almeja-se ainda que esse método possa se estabelecer como ferramenta auxiliar rotineira nos levantamentos agrícolas brasileiros (inclusive nos oficiais), podendo ser expandido para a escala estadual e nacional.

Para alcançar esse objetivo geral, são estabelecidos os seguintes objetivos específicos:

- a) Desenvolver um roteiro para preparação de imagens de sensoriamento remoto orbital de forma a permitir seu uso em levantamentos agrícolas por amostragem;
- b) Estabelecer uma forma de extrair uma amostra aleatória de pixels de uma imagem, dentre um conjunto de pixels espacialmente delimitado;
- c) Fornecer as fórmulas para o cálculo da estimativa de área plantada por município, bem como da sua variância, a partir dos dados obtidos por amostragem;

- d) Definir uma maneira de se promover a amostragem estratificada, tanto por município como em um conjunto de municípios, incluindo os cálculos das estimativas e respectivas variâncias para o caso estratificado;
- e) Obter, para o município de Ipuã – SP, valores de referência de área plantada com as principais culturas, através de mapeamento completo baseado em imagens de sensoriamento remoto e processamento digital usando o sistema de informações geográficas (SIG) SPRING;
- f) Validar o método de amostragem por pontos no município, comparando as estimativas obtidas com os valores de referência observados em Ipuã; e
- g) Simular valores de estimativas e variâncias, usando o método de amostragem estratificada.

1.3 Pressupostos

O presente trabalho está assentado sobre quatro pressuposições básicas. A primeira delas consiste em supor a existência de um talhão típico para uma determinada combinação de cultura agrícola e região de plantio. Ou seja, considerando-se a área de um ou de poucos municípios vizinhos, existirá uma razoável homogeneidade, em termos de tamanho, época de plantio, cultivares e tratos culturais diversos, entre os diversos talhões de uma mesma cultura agrícola; o que pode ser atribuído ao que alguns autores chamam de tradição agrícola regional (Ortiz et al., 1997). Isto será especialmente verdadeiro nas áreas eminentemente agrícolas, pois as forças de mercado e as políticas agrícolas, que almejam o aumento da produção e da eficiência, induzem uma progressiva intensificação do manejo das terras relativamente planas e férteis, com a mecanização levando ao aumento no tamanho dos talhões, à remoção dos corredores de vegetação natural e ao aumento na aplicação de agroquímicos (Fjellstad e Dramstad, 1999; Hietala-Koivu, 1999).

A segunda pressuposição é que, definidos os limites de um talhão agrícola, tudo o que estiver dentro destes limites pode ser classificado como uma única classe de uso do solo (Epiphanyo et al., 2001). Para as principais culturas agrícolas o plantio já é normalmente feito com uma

única espécie por talhão, mas, mesmo a consorciação de culturas pode ser tratada, caso a caso, como uma única classe de uso do solo. Isto implica em considerar que são puros todos os elementos de cena de cada segmento da imagem, o que concorda com a posição tomada por alguns autores (Smits et al., 1999), mas contraria aqueles que adotam o enfoque dos modelos de mistura ou um enfoque ainda mais nebuloso (Binaghi et al., 1996; Foody, 1999).

Em terceiro lugar, supõe-se que a combinação de imagens obtidas por satélite com o processamento digital de dados espaciais por meio de um SIG e com o uso do aparelho GPS para localização no campo propicia a aplicação do método de amostragem aleatória por pontos, equivalentes aos pixels das imagens, em levantamentos estatísticos para a obtenção de estimativas agrícolas, e que esse método permite gerar estimativas não tendenciosas de área ocupada por diferentes tipos de uso do solo. E ainda que a precisão dessas estimativas é facilmente controlada através da determinação do tamanho da amostra.

Por último, supõe-se que a estratificação do universo amostral aumenta a eficiência da amostragem e pode ser facilitada pelo uso das imagens obtidas pelos sensores a bordo de satélites.

Das assertivas anteriores, as duas primeiras foram confrontadas com dados históricos sobre a agricultura brasileira, e em especial a de São Paulo; dados estes oriundos de censos e outros levantamentos (Pino et al., 1997; IBGE, 2003a e IBGE, 2003b). Já o teste de veracidade das duas últimas consistiu no núcleo de investigação do presente trabalho.

1.4 Estrutura

Para alcançar os objetivos estabelecidos, o presente trabalho apresenta no Capítulo 2, primeiramente um panorama das relações entre as estatísticas agrícolas e o sensoriamento remoto no Brasil e em outros países. A seguir são discutidas as características da cena agrícola que interferem na sua observação como alvo do sensoriamento remoto. Diante dessas características, são considerados os fatores importantes para a realização de uma análise estatística adequada aos dados de sensoriamento remoto, mantendo-se em mente o objetivo de elaborar estimativas de área plantada. Defende-se, então, a opção pela amostragem, em oposição ao mapeamento completo do uso do solo sobre o território em estudo. É dada a base

teórica necessária ao estabelecimento de uma amostra aleatória, descrevendo-se a forma de cálculo das estimativas obtidas a partir da amostragem. As vantagens da estratificação, tanto dentro como entre municípios é também demonstrada, assim como a forma de realizá-la e as fórmulas utilizadas.

No Capítulo 3, é fornecida uma descrição da área de estudo, seguida do relato da forma com que se construíram os dados usados como referência para a verificação do método de amostragem proposto. É nesse ponto que se apresentam também quais foram as imagens utilizadas e os processamentos a que elas foram submetidas. Estabelecidos os dados de referência, é apresentado o método de amostragem por pontos no município, desenvolvido especialmente para se adequar à estimativa de área plantada com o auxílio de imagens de satélite. Na seqüência, é pormenorizado o método que possibilita a estratificação dentro do município, adequado aos casos onde a distribuição espacial das culturas de interesse é heterogênea na área estudada. Para as situações nas quais a heterogeneidade ocorre entre municípios, é detalhado um método de estratificação não espacial.

No Capítulo 4 são apresentados os resultados, em primeiro lugar, do mapeamento realizado nas safras de verão e inverno no município de Ipuã, que servirão como base de comparação das demais estimativas. A seguir são apresentadas as estimativas produzidas pela amostragem aleatória simples de 100, 200 e 300 pontos no município, avaliando o efeito do tamanho da amostra sobre a precisão. Os dados obtidos por amostragem são comparados aos de referência e ambos são contrapostos aos dados oficiais disponíveis, sendo feita a discussão dos mesmos. São então exibidos os valores encontrados quando se procede a estratificação do município e é exposta a influência do critério de estratificação escolhido sobre as estimativas obtidas. Finalmente, uma simulação baseada em dados secundários é levada a efeito para realizar a estratificação em âmbito estadual e nacional.

Ao final, no Capítulo 5, conclui-se o estudo expondo os pontos fortes e fracos dos métodos apresentados. Seguem-se as referências utilizadas para a composição do trabalho.

CAPÍTULO 2

FUNDAMENTOS TEÓRICOS

2.1 Estimativas Agrícolas e Relações com o Sensoriamento Remoto

Apesar da diminuição da importância relativa da agricultura para a economia mundial nas últimas décadas, existe uma crescente necessidade de monitorar o complexo agrícola (Ryerson et al., 1997). Além das preocupações com os aspectos ambientais envolvidos com as atividades agrícolas, esse interesse também é justificado economicamente pois, apesar da redução relativa, os números absolutos ainda são muito expressivos, chegando aos US\$ 90 bilhões anuais só nos Estados Unidos e apenas em valor dos produtos, sem contar toda a cadeia produtiva envolvida no agronegócio (Anderson et al., 1999).

No Brasil, em especial, os produtos agropecuários continuam tendo enorme importância econômica, tendo sido responsáveis pela geração de crescentes saldos positivos no comércio internacional, o que auxiliou no enfrentamento da crise do balanço de pagamentos no período 1992 a 1998. Neste curto intervalo de tempo as exportações de produtos agropecuários passaram de 35,8 para 51,1 bilhões de dólares, respondendo por 38,2% das exportações nacionais no final do milênio (Carvalho, 1999).

As informações sobre as atividades agrícolas atendem aos mais diversos interesses. Geograficamente falando, os usuários podem variar desde o nível global até o local, passando por hemisférico, continental, de blocos econômicos, regionais, nacionais, estaduais, municipais, etc. Segundo Pino (1999b), os países que pretenderem se inserir no novo contexto mundial de economia globalizada e de fortes blocos econômicos precisarão contar com excelentes informações agrícolas, não somente sobre si mesmos, mas sobre os demais países do próprio bloco e fora dele. Deverão contar, acima de tudo, com previsões que sejam eficazes e que tenham rápida percepção das mudanças, por mais sutis que pareçam. Já para Ray et al. (1999), as estimativas de produção, precisas e oportunas, no nível distrital (no caso da Índia, o que provavelmente equivale ao nosso município), são essenciais para as decisões gerenciais relacionadas à economia agrícola de um país. Ainda segundo os mesmos autores,

as estimativas de produção feitas na fase de pré-colheita constituem uma informação essencial para o governo e também para as agroindústrias, no sentido de determinar os fatores econômicos tais como preços, excedentes exportáveis, etc.

Do ponto de vista da inserção no negócio agrícola, os interessados nas informações agrícolas podem ser produtores, exportadores, importadores, indústrias de beneficiamento, consumidores, fornecedores de insumos, investidores, etc. Conforme Luiz (2000), hoje as mercadorias agrícolas negociadas nas bolsas de todo o mundo, como o café, a soja, o cacau, o milho, o trigo, o suco de laranja, entre outras, movimentam anualmente cifras astronômicas (na casa do trilhão de reais) e, por isso mesmo, têm seu preço determinado muito mais pelo “mercado global” do que por qualquer lógica referente ao custo de produção ou à eficiência produtiva locais. Sabe-se que o principal fator a determinar as oscilações nessas bolsas, que sempre trabalham com o mercado futuro, são as expectativas de oferta e demanda mundiais dos produtos. Então, aquela empresa ou nação que conseguir prever com maior acerto a sua safra e a dos seus concorrentes, poderá influir no preço de uma mercadoria estratégica e fazer “bons” negócios.

Uma abordagem também interessante é feita por Pino (1999a), que ressalta a importância da determinação da área cultivada por município a partir do fato de que no Brasil, por força de disposição constitucional, 25% do imposto sobre a circulação de mercadorias e serviços (ICMS) deve ser repartido entre os municípios, e que em diversos Estados brasileiros existem critérios legais que vinculam critérios agrícolas à devolução de parte desse imposto. No caso específico do Estado de São Paulo, 3% do valor do ICMS a ser restituído aos municípios é dividido proporcionalmente à área cultivada em cada um deles. Pode-se então imaginar os potenciais conflitos de interesses que leis deste tipo podem gerar, daí a necessidade de se estabelecerem métodos criteriosos para tal fim.

Fica clara a importância do sensoriamento remoto neste contexto, pois, como a obtenção de informações precisas e em tempo hábil, sobre a extensão, as condições e o potencial de produção das culturas, é de grande importância para a economia de um país (Chen, 1990), e como o custo das tradicionais técnicas de medidas no campo se torna proibitivo quando elas são aplicadas a grandes áreas, o uso de sensores remotos torna-se uma alternativa prática

(Thenkabail et al., 1994) para algumas das etapas deste processo. Para Gurgel et al. (2001), nas últimas décadas, o advento do sensoriamento remoto orbital, tornou possível o monitoramento da evolução espacial e temporal das mudanças na cobertura da superfície terrestre, permitindo ainda que essas mudanças passassem a ser estudadas de forma objetiva.

Outro nicho de aplicação do sensoriamento remoto como auxiliar na elaboração de estatísticas agrícolas está garantido pela demanda que os blocos de países têm por dados comparáveis de todos os seus membros. Segundo Bouma et al. (1998), mesmo na Europa, seja ocidental como oriental, os dados agrícolas raramente são coletados de acordo com categorias uniformes, variando muito em escala, forma de coleta, grau de agregação, terminologia, etc., tanto de um país para outro como de um ano para outro dentro do mesmo país. O sensoriamento remoto orbital é a única ferramenta atualmente disponível que pode ser considerada sem fronteiras, ou seja, não importa o país, a língua, ou o sistema político; os dados de uma imagem podem ser sempre tratados de forma homogênea.

Para tratar esta multiplicidade de dados, sejam eles de diferentes fontes (sensores, dados de campo, etc.), sejam de diferentes datas, torna-se importante levantar o papel dos SIGs. Green (1992) comenta que este tipo de sistema é um instrumento espetacular para a integração de dados e obtenção de melhor precisão nos modelos de representação do mundo real. Através do registro de imagens tornaram-se possíveis as análises de relacionamento entre localização espacial, variação espectral da imagem e variação da cobertura dos solos. A atualização dos dados também fica extremamente facilitada uma vez que, montada a base de dados, fica muito fácil produzir uma cobertura atualizada, obtendo assim um resultado dinâmico e, portanto, mais próximo do real. Um excelente exemplo de integração entre imagens de sensoriamento remoto e dados de campo, utilizando-se um SIG para a criação, alimentação, manutenção e consulta a um banco de dados geográfico, aplicado ao problema de estimativa de área agrícola, pode ser visto em Gürtler (2003).

2.2 Culturas Agrícolas como Alvo do Sensoriamento Remoto

Do ponto de vista do sensoriamento remoto, as atividades agrícolas possuem algumas características que interferem na sua identificação em imagens digitais obtidas por sensores orbitais. Diferentemente dos ambientes naturais, os alvos agrícolas sofrem intensa e

intencional influência antrópica, o que contribui para lhes conferir um aspecto de figuras geométricas relativamente regulares, de grande homogeneidade interna. Isto se deve aos procedimentos de:

- 1) aração e gradagem, que diminuem a irregularidade do terreno;
- 2) calagem e adubação, que suavizam as diferenças nas componentes químicas do solo;
- 3) plantio mecanizado, que confere o caráter geométrico aos “talhões”;
- 4) semeadura em um curto espaço de tempo e uso de sementes selecionadas, que conferem regularidade à cobertura vegetal em cada instante no tempo; e
- 5) controle de pragas, doenças e ervas daninhas, que conferem regularidade e pureza à população de plantas.

Outros fatores de homogeneidade são o calendário agrícola e a vocação regional, que resultam num conjunto limitado e relativamente estável de culturas agrícolas que podem estar presentes em uma imagem específica. Todas estas características contribuem no sentido de facilitar a identificação dos limites dos alvos agrícolas em imagens obtidas por satélites de sensoriamento remoto (Luiz, 2002).

Por outro lado, a resposta ao longo do tempo para um mesmo ponto no terreno, ocupado por uma cultura agrícola, muda drasticamente devido aos fatores sazonais (climáticos) e também aos antrópicos (plantio, colheita, preparo de solo, etc.). Por este motivo, segundo Jensen (2000), o fator temporal é muito importante quando se pretende identificar diferentes tipos de vegetação ou extrair informações biofísicas úteis a partir de dados de sensoriamento remoto. A importância de conhecer e levar em consideração os diferentes ciclos fenológicos das culturas presentes em uma área quando se pretende monitorá-la através de dados obtidos por sensoriamento remoto já era reconhecida desde as primeiras aplicações em previsões agrícolas, conforme relata MacDonald e Hall (1980) em seu clássico artigo.

Segundo Lillesand e Kiefer (1994), quando se consideram os componentes envolvidos no estudo global da oferta e demanda por produtos agrícolas, as aplicações do sensoriamento

remoto são realmente muitas e variadas. O escopo dos problemas físicos, biológicos e tecnológicos que se apresentam à moderna agricultura é extremamente amplo e está intimamente relacionado com os problemas mais gerais de população, energia, qualidade ambiental, clima e condições de tempo. Esses fatores são por sua vez influenciados pelos valores e tradições humanos e pelos sistemas econômico, político e social. Ao se considerar a aplicação direta da interpretação de imagens na área de identificação de culturas e levantamento de áreas, constata-se que a interpretação é baseada na premissa de que culturas específicas podem ser identificadas pelo seu padrão de resposta espectral e pela textura. A identificação de culturas, para ser bem sucedida, requer um conhecimento dos estádios de desenvolvimento de cada lavoura na área a ser levantada. Essa informação é geralmente resumida na forma de um calendário agrícola, que lista os estados de desenvolvimento esperados e a aparência de cada espécie cultivada em uma área ao longo do ano. Por causa das mudanças nas características das culturas durante a estação de crescimento, é geralmente desejável usar imagens obtidas em diversas datas durante o ciclo de crescimento para a identificação da lavoura. Frequentemente, aquelas que parecem muito semelhantes em uma data parecerão muito diferentes em outra data, e podem ser necessárias imagens de muitas datas para obter um padrão único de resposta espectral para cada tipo de cultura. Para essa aplicação, as composições coloridas oferecem vantagens sobre as monocromáticas porque aumentam a informação espectral disponível.

A agricultura, no sentido mais amplo do termo, abrange uma enorme quantidade de atividades, envolvendo várias centenas de espécies vegetais cultivadas. Entretanto, devido às condicionantes ambientais, econômicas e de tradição agrícola, em cada município ou conjunto de municípios vizinhos, apenas algumas espécies são consideradas como as mais importantes. O critério de importância é geralmente definido pela área cultivada ou pelo valor da produção, podendo também ser estabelecido pelo número de empregos gerados ou pelo maior potencial de dano ambiental, mas o resultado é invariavelmente um número muito pequeno de culturas agrícolas compondo o grupo realmente significativo dentro de uma região (Luiz e Epiphanyo, 2001).

Na Tabela 2.1 pode-se notar que em uma região eminentemente agrícola e com agricultura tecnificada, como é o caso da região definida pelos municípios abrangidos pelo Escritório de

Desenvolvimento Regional (EDR) de Orlândia, localizados na região nordeste do Estado de São Paulo, apenas três culturas ocupam, em 10 dos 12 municípios, mais de 80% da área cultivada total.

TABELA 2.1 - Área plantada, em hectares, das três principais culturas agrícolas nos municípios abrangidos pelo Escritório de Desenvolvimento Regional (EDR) de Orlândia - SP, e participação percentual das três na área agrícola total dos municípios na safra 1995/1996 (Pino et al., 1997).

Município	Soja	Milho	Cana-de-açúcar	% do total
Aramina	1.375	328	12.474	84
Buritizal	3.040	3.058	4.464	43
Guará	20.369	9.104	10.514	133
Igarapava	1.913	2.812	16.775	52
Ipuã	36.504	17.124	19.441	161
Ituverava	50.045	26.595	13.615	128
Miguelópolis	68.010	37.760	15.400	178
Morro Agudo	27.131	7.577	98.223	98
Nuporanga	17.679	6.907	18.175	123
Orlândia	2.950	617	16.950	80
Salesópolis	3.706	2.281	19.095	87
São Joaquim da Barra	5.517	2.016	26.814	87

Outra característica que evidencia a especialização de alguns municípios em agricultura é a ocorrência de valores percentuais de área plantada superiores a 100%, conforme apresentado na Tabela 2.1, que são explicados pelo emprego de mais de um cultivo por ano agrícola, prática possibilitada pela existência de muitas áreas irrigadas e pela tradição do plantio de outono ("safrinha") na região. Essa situação é comum também em outras áreas intensamente ocupadas pela agricultura (Xiao et al., 2002).

Devido às características inerentes às atividades agropecuárias, como a ampla dispersão espacial, o grande número de produtores e a variabilidade do tamanho das propriedades, entre outras, a obtenção de informação atualizada sobre este setor da economia é particularmente difícil. Segundo Ippoliti-Ramilo et al. (1999), o caráter multi-espectral, sinóptico, repetitivo e

global dos dados orbitais indica esta tecnologia como uma grande aliada dos sistemas de estimativas agrícolas.

Atualmente, entretanto, apenas algumas culturas agrícolas podem ser efetivamente monitoradas através do sensoriamento remoto. São aquelas que ocupam a maior área, sendo cultivadas mecanicamente e, portanto, em talhões cujas dimensões mínimas geralmente superam uma dezena de hectares (Luiz e Epiphanyo, 2001).

Não por acaso, as espécies vegetais pertencentes a esse grupo geralmente são também aquelas que têm alto interesse econômico ao longo da sua cadeia produtiva. Os dados na Tabela 2.2 oferecem uma amostra da variação do tamanho das unidades de produção na região abrangida pelo EDR de Orlândia - SP, tanto entre os municípios como entre as culturas agrícolas, além de confirmar que, para as principais culturas, o tamanho médio é suficientemente alto para permitir uma utilização eficiente do sensoriamento remoto no seu monitoramento.

TABELA 2.2 - Área média, em hectares, das unidades de produção ocupadas com as três principais culturas agrícolas nos municípios abrangidos pelo Escritório de Desenvolvimento Regional (EDR) de Orlândia - SP na safra 1995/1996 (Pino et al., 1997).

Município	Soja	Milho	Cana-de-açúcar
Aramina	29	12	83
Buritizal	62	28	51
Guará	100	46	122
Igarapava	30	24	51
Ipuã	122	62	237
Ituverava	104	53	130
Miguelópolis	136	80	175
Morro Agudo	116	88	183
Nuporanga	75	35	88
Orlândia	89	21	202
Salesópolis	69	27	99
São Joaquim da Barra	70	40	202

Concordando com as afirmações anteriores, em um trabalho de monitoramento da ocupação agrícola realizado em três municípios no sul do Estado do Maranhão, Sano et al. (1998) encontraram um pequeno número de classes predominantes de uso do solo agrícola (arroz,

soja, milho, caju, pastagem e áreas preparadas não cultivadas) e também ressaltaram que uma das fontes de erro que influenciou nos seus resultados estava relacionada com a dificuldade de interpretação de áreas cultivadas inferiores a 10 hectares.

Outro exemplo expressivo dessa situação ocorre com as culturas da soja e do milho no Brasil. Os 50 municípios maiores produtores destas culturas representam apenas 0,91% do número total de municípios brasileiros (que são 5.507), mas são responsáveis por 14,11% da área plantada e por 20,85% da produção nacional, no caso do milho, e por 36,45% da área e 41,42% da produção, no caso da soja. Isso confirma que existem municípios extremamente especializados nesses cultivos e que a produção é bastante concentrada (Tsunechiro e Freitas, 2001).

No trabalho de Garagorry e Simon (2003), no qual é feito um estudo da evolução, num período recente, do comportamento de sete produtos da agropecuária brasileira, particularmente no que se refere à dinâmica espacial e à concentração, os resultados relacionados com a concentração da produção levam a concluir que, de maneira geral e levando em conta todas as microrregiões para as quais existem registros de um determinado produto, apenas 20% delas são suficientes para se alcançar 75% do volume total.

Todas essas observações permitem afirmar que uma expressiva parcela da produção agropecuária brasileira está concentrada em um número pequeno de municípios, quando comparado ao número total de municípios no país. Além disso, os municípios que são importantes produtores de alguma cultura considerada, em geral apresentarão uma significativa parcela do seu território ocupada pela cultura e de forma espacialmente concentrada.

2.3 Análise Estatística

As técnicas de estatística fornecem um rico e poderoso arsenal de ferramentas para análise de todo tipo de conjuntos de dados. Quando empregadas corretamente, elas podem levar à indução de hipóteses sobre a natureza e o grau de relacionamento entre as variáveis estudadas. O contínuo crescimento da disponibilidade de computadores cada vez mais poderosos, associado ao crescente refinamento dos programas computacionais estatísticos, tem colocado as mais avançadas análises estatísticas ao alcance dos pesquisadores de diversas áreas do

conhecimento, e não só aos estatísticos, como acontecia até pouco tempo (Griffith e Amrhein, 1997; Hair et al., 1998).

Embora o acesso a todo esse poder analítico possa realmente auxiliar na concepção empírica de hipóteses, para evitar seu uso indevido ou inapropriado é necessário adotar um enfoque sistemático na análise dos dados. Neste sentido, é necessário decidir, definir ou registrar, de que forma serão coletadas e medidas as variáveis observadas. Nesta "forma" de medida está incluído se a variável é do tipo categórica (nominal), ordinal (posição) ou numérica (contínua, discreta, intervalar, razão); qual a unidade de medida (nomes, cores, ordem de colocação, gramas, horas, etc.); e qual o elemento básico comum sobre o qual se farão as medidas (pixel, segmento, talhão, fazenda, município, etc.) (Hair et al., 1998).

É preciso também definir como será coletada a amostra com que se vai trabalhar, estabelecendo seu tamanho e um procedimento de escolha que garanta o componente aleatório ou que especifique as correlações em caso de não aleatoriedade, além de identificar a população que essa amostra irá representar (Cochran, 1977).

É desejável ainda que se faça a identificação de dados discrepantes, checando-os para eliminar erros de medição ou de registro. Se não houver fortes indícios de erro de medição ou registro e se o valor, embora extremo, estiver dentro dos limites possíveis para a variável analisada, ele deve ser mantido no conjunto de dados (Hair et al., 1998).

Outro ponto a ressaltar é que, devido às suas inúmeras vantagens em termos de tratamento matemático, a distribuição normal é invariavelmente adotada como base para a maioria dos procedimentos de análise estatística. Na realidade, nem todas as variáveis se ajustam a esta distribuição e, portanto, é necessário estar atento a isso para evitar erros ao se tomarem decisões baseadas em pressupostos inválidos (Neter et al., 1996; Hair et al., 1998).

Outra ponderação deve ser feita com relação ao uso de técnicas estatísticas, que é válida em geral e, portanto, também deve ser considerada ao se tratar de dados de sensoriamento remoto. Ela diz respeito aos dados usados na construção das estatísticas ou na elaboração e ajuste de modelos. Eles são a matéria prima do estatístico e, como não poderia deixar de ser, se a matéria prima não for adequada aos objetivos, nenhum produto dela poderá sê-lo. Portanto,

deve ser adotado um enfoque estruturado que garanta uma qualidade homogênea durante todo o processo de aquisição, tratamento e análise dos dados, tanto os de campo como os espectrais, com especial atenção ao estágio de preparação do trabalho de campo, conforme afirmam Mulders et al. (1992).

2.3.1 Opção pela Amostragem

Para atender aos interesses dos agentes políticos e econômicos da agropecuária brasileira, que necessitam de informação a tempo de tomar decisões e intervir na cadeia produtiva, a obtenção de informação sobre área plantada e desenvolvimento das culturas deve ser um processo contínuo, periódico, rápido, nacional, transparente, auditável e de precisão conhecida.

Essas exigências impõem certas restrições aos métodos possíveis de serem adotados. A cobertura nacional, num país de mais de 5.500 municípios e 850.000.000 de hectares, só pode ser obtida através da amostragem, dificilmente pelo método censitário (exaustivo ou de mapeamento).

Apesar de o mapeamento temático ser uma das aplicações mais comuns do sensoriamento remoto, ainda não se atingiu um estágio no qual os mapas de cobertura do terreno possam ser rotineiramente derivados de dados obtidos remotamente. Muitas são as razões de ainda não ter sido concretizado o enorme potencial do sensoriamento remoto como uma fonte de dados de cobertura do solo ou temáticos de maneira geral. Entre elas pode-se destacar a natureza das classes, as resoluções espectrais e radiométricas dos sensores, e os métodos usados nos mapeamentos (Foody, 1999).

A necessidade de informações produzidas com rapidez suficiente para permitir a intervenção ainda dentro da safra agrícola, exige que o tempo gasto para a coleta, processamento e análise dos dados seja o menor possível, o que impõe sérias limitações ao tamanho da amostra.

Num país que possui território em dois hemisférios, a sazonalidade é um fator importantíssimo, só podendo existir um acompanhamento nacional eficiente das safras através

de um processo contínuo de coleta de dados. Isso exige que as unidades amostrais sejam visitadas e analisadas várias vezes ao longo do ano, restringindo ainda mais o número de unidades amostrais possíveis de serem acompanhadas a um custo razoável.

Além da sazonalidade, todas as demais diferenças: regionais, culturais, econômicas, etc., atuam no sentido de tornar a população alvo do levantamento estatístico altamente heterogênea. É obrigatório o uso de alguma forma de estratificação da população para contornar este problema.

Para garantir a transparência e a possibilidade de controle de qualidade dos dados, é necessário que as informações coletadas nas unidades amostrais sejam georreferenciadas e armazenadas em um banco de dados, permitindo que, a qualquer momento, seja feita uma auditoria desses dados. Isso é extremamente facilitado pelo uso conjunto dos dados de sensoriamento remoto na forma de imagens, dos aparelhos de posicionamento global (GPS) e dos Sistemas de Informações Geográficas (SIGs).

Para que o levantamento produza estimativas objetivas e de precisão conhecida é necessário que seja determinado um plano amostral de forma probabilística e que os dados de área plantada sejam obtidos através de “medições” ao invés de por entrevistas ou questionários. A experiência do autor com a aplicação de questionários ou realização de entrevistas com agricultores (Luiz e Silveira, 2000) permite afirmar que quando a pergunta é “qual a área plantada com a cultura X?” as respostas podem ser as mais variadas e com diferentes graus de incerteza, como: “quase meio alqueire”; “14 tarefas”; “plantei dois sacos de semente”; “foi a gleba todinha”; “duas horas no horímetro do trator”; entre outras. Mesmo quando a resposta vem expressa em hectares, quase nunca o valor foi obtido por medição topograficamente correta da área, sendo, na maioria das vezes, estimada pelo agricultor ou técnico responsável, a partir da experiência e conhecimento da propriedade. Nesse sentido, o uso das imagens obtidas por sensoriamento remoto, pela incorporação do conceito do “pixel” da imagem como unidade amostral básica, permite e facilita a realização de uma medição objetiva (Luiz et al., 2002; Luiz e Gürtler, 2003). Já a possibilidade de estabelecer a precisão das estimativas é garantida pelo estudo da natureza das variáveis medidas e pelo estabelecimento do plano amostral adequado (Epiphanyo et al, 2001; Epiphanyo et al., 2002; Luiz e Epiphanyo, 2001).

2.3.2 Amostragem Aleatória Simples

O método da Amostragem Aleatória Simples (AAS) consiste na seleção de uma amostra de n elementos de um total populacional de N , de tal forma que qualquer possível amostra tenha a mesma probabilidade de ser escolhida. No caso de se desejar extrair uma AAS de n pixels de uma imagem digital, enumera-se de 1 até N os elementos de cena que compõem a imagem e, através de um algoritmo de geração de números aleatórios, selecionam-se os n elementos da amostra.

Embora o interesse maior nesse trabalho seja pela área plantada com as culturas agrícolas, como a unidade amostral (o elemento de cena) tem área conhecida e constante em toda a região investigada, a proporção de ocorrência das classes de interesse na amostra equivale à proporção de área amostral ocupada pelas culturas, o que, por expansão, permite estimar as áreas totais no município. Assim, é necessária e suficiente a medição de apenas uma variável categórica em cada unidade amostral. Ou seja, no caso mais simples de uma única cultura (x_j), um elemento poderá ser classificado em apenas duas categorias: U se ele pertence a um segmento ocupado pela cultura x_j ; e \bar{U} se ele pertence a um segmento não ocupado pela cultura x_j . Para facilitar a exposição do método será adotada a seguinte notação:

A = número de elementos da classe U na população (a na amostra)

$P = A/N$ = proporção de elementos da classe U na população ($p = a/n$ na amostra)

$Q = 1-P$ = proporção de elementos da classe \bar{U} na população ($q = 1-p$ na amostra)

$f = n/N$ = fração amostral ($1/f = N/n$ = fator de expansão)

Considerando que a variável é categórica, o processo descrito para a obtenção da amostra é, segundo Johnson e Kotz (1969), a situação clássica que leva naturalmente a uma distribuição discreta do tipo hipergeométrica. Caso N seja suficientemente grande, a distribuição é aproximada para a binomial (Zwillinger e Kokoska, 2000). Neste caso, segundo Cochran (1977), a estimativa de P é dada diretamente por $\hat{P} = p$, já a estimativa de A é obtida pela aplicação do fator de expansão, ou seja, $\hat{A} = a(1/f)$.

Supondo que para cada elemento da população existe uma variável y_i que assume valor 1 se o elemento pertence à classe U , ou valor 0 se o elemento pertence à classe \bar{U} , fica claro que o total para a população (Y) é obtido da **Equação 1**:

$$Y = \sum_{i=1}^N y_i = A \quad (1)$$

e que as médias da população e da amostra podem ser calculadas pelas **equações 2 e 3**, respectivamente:

$$\bar{Y} = \frac{\sum_{i=1}^N y_i}{N} = \frac{A}{N} = P \quad (2)$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{a}{n} = p \quad (3)$$

Conseqüentemente, o problema de obter \hat{A} e \hat{P} pode ser reduzido à simples estimativa do total e da média de uma população onde todo y_i é ou 1 ou 0. Pode-se, então, expressar a variância da população (S^2) e da amostra (s^2), em função de P e p , da seguinte maneira:

$$S^2 = \frac{N}{N-1} PQ \quad (4)$$

$$s^2 = \frac{n}{(n-1)} pq \quad (5)$$

Sob estas condições pode -se afirmar que a proporção amostral $p = a/n$ é uma estimativa não tendenciosa da proporção populacional $P = A/N$. E, na prática, nos casos onde a fração amostral não excede 5% ($f < 0,05$), pode-se considerar que a variância da média amostral (ou seja, a variância de p) tem sua estimativa não tendenciosa dada pela **equação 6**:

$$\hat{s}_p^2 = \frac{pq}{(n-1)} \quad (6)$$

O número estimado total de elementos da classe U na população é representado como $\hat{A} = Np$, e a estimativa não tendenciosa da sua variância pode ser obtida pela **equação 7**:

$$\hat{s}_{Np}^2 = \frac{N^2}{(n-1)} pq \quad (7)$$

Com isso pode-se obter uma boa indicação do tamanho de amostra adequado à identificação de eventos (culturas agrícolas no presente caso) com as mais variadas probabilidades de ocorrência na população. Ou seja, dado um município, a partir de sua área e das dimensões do elemento de cena temos o valor de N e, com base em informações de anos anteriores, obtidas de outros levantamentos, obtém-se um valor para p ; o que permite a construção de uma tabela ou gráfico que represente a relação entre o tamanho da amostra (n) e a variância da estimativa (s_{Np}^2) ou os seus derivados (por exemplo, o coeficiente de variação CV%).

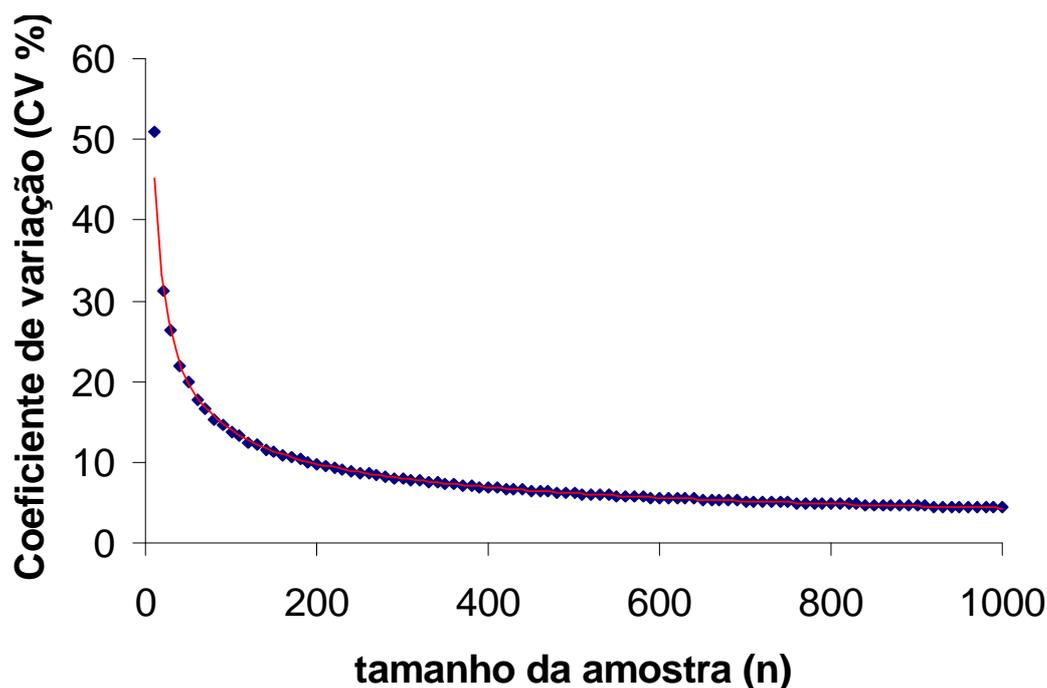


FIGURA 2.1 - Curva que representa a relação entre o tamanho da amostra (n), e o coeficiente de variação esperado (CV %), dada a probabilidade $p = 0,3468$ de ocorrência da cultura na área.

No gráfico da Figura 2.1 está exposto um exemplo no qual empregaram-se os valores correspondentes à área do município de Ipuã ($467.058.750 \text{ m}^2$), a área do elemento de cena

equivalente a $25 \text{ m} \times 25 \text{ m}$ (625 m^2) e o valor estimado pelo IBGE/PAM para a área colhida de soja em 2001 naquele município ($162.000.000 \text{ m}^2$). Nesse caso, $N = 747.294$ e $P = 0,34685144$.

2.3.3 Construção de uma Amostra Aleatória

A construção de uma amostra aleatória não é tarefa trivial e, na verdade, a dificuldade em obtê-la em levantamentos agrícolas vem desafiando os estatísticos envolvidos com esse tema há décadas. Ao considerar o problema de forma tradicional, as unidades básicas de amostragem constituem-se em talhões agrícolas, propriedades, unidades de produção, ou outro tipo de segmento qualquer, mas sempre tendo os limites definidos do ponto de vista da atividade socioeconômica predominante, ou seja, a atividade agropecuária. Isso causa um sério problema para o estabelecimento de amostras aleatórias, pois classicamente, para garantir-se a aleatoriedade, é necessário garantir que todos os componentes da população tenham chance não nula e conhecida de pertencer à amostra. Para isso é necessário que se conheçam todos os elementos de uma população finita ou a distribuição de probabilidade no caso de populações infinitas.

O caso de estimativa de área plantada é caracterizado como pertencente ao primeiro tipo, o das populações finitas. Assim, se a unidade amostral básica escolhida é a propriedade rural, por exemplo, para poder realizar uma amostragem aleatória, será preciso possuir uma lista completa de todas as propriedades existentes na população considerada. Muitos são os entraves encontrados ao se tentar obter essa lista, entre eles podem-se destacar os seguintes: a) a soma da área das propriedades, por erros de medida ou declaração, não é igual à área total da população; b) os registros, devido às diferenças da época em que foram feitos, encontram-se em bases diferentes – cartórios, comarcas, distritos, municípios, estados ou territórios, que foram criados antes ou depois da data do registro do imóvel; c) inexistência pura e simples de registro; d) áreas devolutas; e) áreas de litígio – com dois ou mais registros ou proprietários; f) mudanças rápidas no número, tamanho e posse das propriedades, motivadas por falecimento do proprietário, venda, separação ou dissolução de sociedade.

Para contornar estes obstáculos, uma opção é o uso dos chamados painéis amostrais por área. Embora em tese eles possam contemplar elementos com limites definidos artificialmente

(Luiz, 2002), essa não era a regra até recentemente e os segmentos eram normalmente definidos pelos limites físicos do terreno (Collares et al., 1993).

Apesar de se conhecer há bastante tempo a vantagem que o uso de imagens de sensoriamento remoto poderia trazer aos levantamentos agrícolas por amostragem, somente a partir do final dos anos noventas, com a disponibilização de aparelhos GPS precisos e de baixo custo, assim como o surgimento de programas SIGs amigáveis – e em alguns casos gratuitos, como o SPRING – a utilização do pixel como unidade amostral básica passou a ser possível na prática para levantamentos agrícolas (Luiz et al., 2002). A partir de então, o uso conjunto de imagens de sensoriamento remoto, SIGs e aparelhos GPS, vem crescendo, inclusive nas aplicações agrícolas (Wilkinson, 1996; Pradhan, 2001; Gao, 2002).

Então, existe agora a possibilidade de, utilizando-se imagens de sensoriamento remoto, que fornecem naturalmente elementos com limites artificiais - os pixels - realizar a extração de amostras aleatórias para estimativa da área plantada. Isso facilita muito o planejamento amostral e a análise estatística dos dados coletados. Entretanto, é preciso garantir que o procedimento prático de obtenção da amostra resulta realmente em um processo aleatório, condição fundamental para a aplicação de todos os demais conceitos estatísticos usados nos cálculos das estimativas.

2.3.3.1 Hipóteses

Considerando que cada banda espectral de uma imagem digital qualquer, como as obtidas por sensoriamento remoto orbital, é representada por uma matriz $n \times m$, sendo n o número de colunas e m o número de linhas; sabendo ainda que cada célula dessa matriz compõe um elemento de cena (pixel), que pode ter sua posição no espaço representada por coordenadas planas X e Y, pode-se afirmar que:

- a) se, através de um processo de geração de números aleatórios, forem produzidos t pares ordenados (x_i, y_j) , que determinem a posição de t pixels na imagem, esses pixels constituirão uma Amostra Aleatória (AA) representativa da imagem.

b) se, ao sobrepor um polígono irregular qualquer (p.ex., os limites de um município) sobre a imagem, forem conservados apenas os t pixels ($t \leq k$) cujos centros estejam localizados no interior do polígono, esse novo conjunto de pixels continuará a compor uma AA, representativa da região da imagem correspondente à área do município.

2.3.3.2 Demonstração

Se \mathbf{X} é uma variável aleatória (v.a.) com distribuição univariada uniforme discreta, isso significa que ela pode assumir n valores diferentes, $\{x_1, x_2, \dots, x_n\}$, com igual probabilidade. Entre as propriedades dessa distribuição, a função de densidade de probabilidade (f.d.p.) de \mathbf{X} é:

$$f_{(x)} = \begin{cases} \frac{1}{n} & \forall x_i \text{ se } x_i \in \{x_1, x_2, \dots, x_n\} \\ 0 & \text{caso contrário (c.c.)} \end{cases} \quad (8)$$

Seja \mathbf{Y} outra v.a., independente e identicamente distribuída (i.i.d.) com relação a \mathbf{X} , que pode assumir m valores diferentes, $\{y_1, y_2, \dots, y_m\}$. Então, se \mathbf{Z} é definida pelo par ordenado (\mathbf{X}, \mathbf{Y}) , de tal forma que $z_k = (x_i, y_j)$, onde $1 \leq i \leq n$, $1 \leq j \leq m$, e $1 \leq k \leq (n \times m)$, sendo i, j, k, n e m números inteiros positivos; então, podemos afirmar que \mathbf{Z} tem distribuição uniforme discreta, com a seguinte f.d.p.:

$$f_{(z)} = \begin{cases} \frac{1}{n \times m} & \forall z_k \text{ se } z_k \in \{z_1, z_2, \dots, z_{n \times m}\} \\ 0 & \text{c.c.} \end{cases} \quad (9)$$

Isso é verdade se e somente se \mathbf{X} e \mathbf{Y} forem i.i.d.

Agora, se \mathbf{W} é uma v.a. que pode assumir t valores diferentes, sendo $t < n \times m$, de tal forma que $w_r = z_k$, onde $1 \leq r \leq n \times m$, sendo t e r números inteiros positivos; então, decorre que \mathbf{W} também tem distribuição uniforme discreta, com a seguinte f.d.p.:

$$f_{(w)} = \begin{cases} \frac{1}{t} & \forall w_r \text{ se } \forall r \exists k \Rightarrow w_r = w_k, r \in \mathbb{N}^+, 1 \leq r \leq (n \times m), t < (n \times m) \\ 0 & c.c. \end{cases} \quad (10)$$

Assim sendo, respondendo à hipótese a), se os t pares ordenados retirados da população foram produzidos em uma planilha eletrônica qualquer, na qual as células de duas colunas, de t elementos cada uma, foram preenchidas com valores inteiros positivos por um procedimento de geração de números aleatórios baseado na distribuição uniforme univariada - em uma coluna os números variam de 1 até n (número de colunas da imagem) e na outra de 1 até m (número de linhas da imagem) – ao fazer a equivalência entre esses números e os apontadores das linhas e colunas da imagem em um SIG, terão sido selecionados os t pixels que irão compor uma AA representativa da imagem.

Com relação à hipótese b), exatamente o mesmo argumento é válido, sendo aplicado duas vezes. Explica-se, ao se sobrepor os limites do município sobre a imagem, foi criada uma nova população, constituída por um subconjunto dos pixels da população original, na qual todos os elementos continuam tendo probabilidade idêntica de serem selecionados para a amostra; paralelamente, dos pixels selecionados para compor a amostra da população original o subconjunto dos que se localizam dentro dos limites do município continuam constituindo uma AA.

Demonstra-se assim que a amostra obtida selecionando-se primeiro um conjunto de pixels em uma imagem e, posteriormente, um subconjunto que esteja contido dentro dos limites de um município, constitui uma amostra aleatória simples.

2.3.4 Estratificação

A estratificação é uma técnica comum em amostragem, e existem muitas razões para isso. Entre essas razões pode-se destacar que quando se deseja conhecer o grau de precisão dos dados para certas subdivisões da população é conveniente tratar cada uma delas como uma “população” em separado, ou seja, cada subpopulação equivale a um estrato. Outra forte motivação é a existência de conveniências administrativas; por exemplo, a agência que

conduz a coleta de dados pode ter vários escritórios, cada um com sua área de abrangência pré-definida segundo critérios muitas vezes alheios aos objetivos de um levantamento específico. Outro motivo pode ser percebido no caso de uma amostragem por produtores agrícolas de determinada cultura, quando só se possui uma lista das grandes propriedades, as quais são colocadas em um estrato separado, enquanto as propriedades menores, não listadas, são divididas em outros estratos, nos quais algum tipo de amostragem por área pode ser aplicado (Cochran, 1977; FAO, 1996). A estratificação pode ainda produzir um ganho na precisão das estimativas das características da população completa, pois é possível dividir uma população heterogênea em subpopulações, cada uma das quais sendo internamente homogênea. Isso é sugerido pelo nome “estrato”, com sua implicação de uma divisão em camadas. Se cada estrato é homogêneo, de modo que as medidas variam pouco de uma unidade para outra, uma estimativa precisa da média de cada estrato pode ser obtida de uma pequena amostra naquele estrato. As estimativas dos estratos podem então ser combinadas em uma única estimativa precisa de toda a população (Cochran, 1977).

A estratificação pode ser usada em duas etapas em um programa de levantamento de estatísticas agrícolas como o proposto neste trabalho. Num primeiro momento, a estratificação pode ser feita baseada nas informações obtidas a partir da imagem de sensoriamento remoto, visando dividir o território municipal em áreas internamente homogêneas e diferentes entre si quanto ao uso predominante do solo. Em uma segunda etapa, um conjunto de municípios – pertencentes a um Estado, Região ou mesmo a todo o país – pode ser estratificado de forma a se obterem subpopulações homogêneas quanto a variável de interesse, que no presente trabalho seria a área plantada.

2.3.4.1 Estratificação dentro de um Município

A estratificação dentro do município é essencialmente do tipo espacial, sendo especialmente adequada quando o uso do solo varia muito de uma região para outra dentro dos limites municipais. Vários são os motivos que podem levar a essa situação. O uso agrícola é fortemente afetado pelo clima, tipo de solo e pela topografia; por exemplo, o arroz irrigado por inundação, nas condições de manejo brasileiras, ocorre quase que exclusivamente em planícies hidromórficas; já as grandes culturas com intensa mecanização, ocupam

prioritariamente terrenos com pequena declividade e sobre solos mais férteis; e as culturas perenes e semiperenes podem ocorrer em áreas de relevo mais acidentado, sendo mais condicionadas pelo clima a escolha da sua localização. Outro fator a afetar o uso do solo são os fatores antropomórficos, entre eles estão principalmente a estrutura fundiária e a distribuição de infra-estrutura de transporte e armazenamento dos produtos agrícolas, ficando relativamente fácil de entender que dentro de um mesmo município, uma região de difícil acesso – por ausência de estradas ou pontes – será preterida na ocupação em relação a uma outra região melhor servida nesses aspectos; as culturas adequadas às regiões divididas em minifúndios serão diferentes daquelas ocupadas por latifúndios. É importante observar que, embora geralmente a estrutura fundiária e de infra-estrutura sejam altamente correlacionadas com as variáveis ambientais de solo, relevo e clima, isso nem sempre é válido.

Outro fator determinante do tipo de uso do solo em cada região do município é dado pelo componente legal ou jurídico. O estabelecimento de leis de preservação ambiental pode transformar toda uma área em reserva, abolindo qualquer forma de uso agrícola; por outro lado leis que determinem limites para alguma atividade agrícola podem forçar um rearranjo espacial da produção, como no caso das leis relativas à declividade máxima passível de mecanização ou a que determina a distância mínima em relação a centros urbanos a partir da qual se pode realizar a queima da cana antes da colheita. Para uma melhor compreensão do impacto que esse tipo de restrição pode causar na distribuição espacial das culturas agrícolas uma excelente fonte é o trabalho de Giannotti (2001), que trata dos impactos da lei ambiental que proíbe a queima de cana-de-açúcar na região de Piracicaba (Estado de São Paulo), onde a proibição da queima impulsiona a mecanização da cultura e as variáveis ambientais condicionam a aptidão das terras à mecanização, o que está modificando a disposição espacial da cultura.

Do ponto de vista da estimativa da área plantada, alguns ou todos os fatores citados acima podem agir de forma a promover ou não a concentração espacial de uma determinada cultura dentro de um município. Isso pode variar inclusive de cultura para cultura, podendo haver num mesmo território um exemplo de distribuição homogênea e outro de alta concentração espacial. A opção pela estratificação, portanto, dependerá mais uma vez do objetivo do estudo, devendo ser escolhido o método que privilegie a estimativa da cultura de maior

interesse. Em outras palavras, dependendo da cultura de interesse, em um mesmo município pode-se optar por fazer ou não a estratificação; e mais, mesmo que se decida pela estratificação, a alocação ótima de pontos amostrais pelos estratos também variará em função da cultura de interesse. Por exemplo, se um município é ocupado principalmente pelas culturas da cana-de-açúcar e da soja; e se a distribuição espacial da cana é fortemente influenciada pela distância às Usinas, causando uma maior concentração dessa cultura numa região do município, pode ser o caso de proceder-se à estratificação em duas áreas, uma com maior ocorrência de cana e outra com maior presença da soja. Essa estratificação resultará numa segmentação única do território do município, tanto faz se o interesse principal é pela soja ou pela cana. Mas, o número ótimo de pontos em cada estrato – mantendo-se o número total constante – variará de acordo com o interesse principal do estudo. Com relação à definição dos limites dos estratos, considerando a escala municipal e supondo que o responsável pelo levantamento possua familiaridade com o território estudado, o método de segmentação indicado é o manual, baseado na interpretação visual da imagem (King, 2002).

Feita a divisão da população (área do município dividida em N pixels) em l estratos e, mantendo a notação usada para a amostragem aleatória simples (AAS), observa-se que, na amostragem aleatória estratificada (AAE) os estratos possuem cada um N_h unidades, sendo que o sufixo h define o h -ésimo estrato da população, eles não se sobrepõem e juntos compreendem toda a população, ou seja:

$$N = N_1 + N_2 + \dots + N_h + \dots + N_l = \sum_{h=1}^l N_h \quad (11)$$

Para caracterizar uma AAE é preciso que se tome uma AAS, de forma independente, em cada um dos estratos. Os tamanhos das amostras dentro dos estratos são chamados de $n_1, n_2, \dots, n_h, \dots, n_l$, respectivamente. O tamanho total da amostra continua sendo chamado de n , e é dado pela soma dos l valores de n_h . Para seguir a notação anterior, adota-se:

A_h = número de elementos da classe U no estrato h (a_h na amostra)

$P_h = A_h/N_h$ = proporção de elementos da classe U no estrato h ($p_h = a_h/n_h$ na amostra)

$Q_h = 1 - P_h$ = proporção de elementos da classe \bar{U} no estrato h ($q_h = 1 - p_h$ na amostra)

$f_h = n_h/N_h$ = fração amostral no estrato ($1/f_h = N_h/n_h$ = fator de expansão)

$$W_h = N_h/N = \text{peso do estrato } h$$

Na amostragem estratificada o número de pontos amostrais n_h (ou o tamanho da amostra) em cada estrato é escolhido pelo responsável pela amostragem. Muitos são os critérios que podem ser adotados na determinação desse valor, sendo que geralmente busca-se reduzir a variância da estimativa, mantendo um custo fixo ou reduzido (Cochran, 1977). No caso mais simples, supondo uma função de custo linear, dada a restrição em termos de número total de pontos amostrais, buscar-se-á a melhor alocação de pontos em cada estrato de forma a obter a variância mínima.

Segundo Cochran (1977), quando a AAE é aplicada a proporções, como é o caso presente, e se deseja estimar a proporção de unidades na população que caem dentro da classe U , a estratificação ideal é atingida ao se colocar no primeiro estrato todas as unidades dessa classe, e nos demais estratos, todas as outras. Como isso geralmente não é possível na prática, especialmente em se tratando de unidades e estratos distribuídos espacialmente, os estratos devem ser construídos de tal forma que a proporção na classe U varie tanto quanto possível de um estrato para outro. Para essa situação, a estimativa da proporção P , apropriada para a AAE, é chamada de \hat{P}_{est} e dada pela **equação 12**:

$$\hat{P}_{est} = \sum_{h=1}^l \frac{N_h \times P_h}{N} = \sum_{h=1}^l (W_h \times p_h) \quad (12)$$

Já a estimativa da variância de P_{est} pode ser calculada com o auxílio da **equação 13**:

$$\hat{S}_{\hat{P}_{est}}^2 = \sum_{h=1}^l W_h^2 \frac{P_h \times q_h}{n_h} \quad (13)$$

2.3.4.2 Estratificação entre Municípios

O método proposto consiste, primeiramente, na criação de estratos compostos de municípios (agora não necessariamente contíguos espacialmente) produtores da cultura de interesse (segundo os dados mais recentes do IBGE). A estratificação pode ser feita por Estado, região ou para todo o país. A variável usada na estratificação é a própria área plantada por

município, fornecida pelo IBGE. Os estratos são formados visando minimizar a variância interna de cada grupo; para isso faz-se uma análise de agrupamento univariada, baseada exclusivamente na variável área plantada. A técnica de análise que mais se ajusta ao objetivo de identificar estratos semelhantes quanto à área plantada é a análise de agrupamentos utilizando a distância euclidiana como medida de proximidade. Para realizá-la é utilizado um algoritmo de agrupamento (clustering) chamado Fastclus, do aplicativo SAS (SAS, 1988), onde a distância entre dois elementos x_1 e x_2 , no caso univariado, é dada simplesmente por:

$$d(x_1, x_2) = \|x_1 - x_2\| \quad (14)$$

Como o número de grupos interfere fortemente na variância, é necessário buscar o número ótimo, com base no compromisso entre número máximo possível de municípios na amostra e variância máxima admissível. É importante lembrar que em cada estrato deverão ser amostrados ao menos três municípios – exceto aqueles estratos que possuam, no total, menos de três municípios, é claro. Essa recomendação visa permitir identificar possíveis dados discrepantes nos municípios componentes da amostra, visto que os dados são secundários e não passíveis de comprovação nessa fase do levantamento. Dessa maneira, com apenas dois municípios, a distância de cada um deles à média amostral será sempre a mesma, e não há critério que permita decidir qual dos dois, na eventualidade de uma discrepância muito acentuada, é o dado contaminado. No caso de pelo menos três municípios, já é possível o estabelecimento de critérios objetivos para a identificação de dados discrepantes.

A melhor combinação entre número de estratos e número de municípios por estrato será aquela que, obedecendo aos critérios anteriores, resulte no menor número total de municípios e, ao mesmo tempo, menor variância total intragrupos.

Definido o número e a composição de cada estrato, é feito um sorteio aleatório dos municípios de cada um deles que irão compor a amostra. Após a escolha dos municípios, em cada um deles é feito então o sorteio aleatório dos pontos (pixels) que deverão ser classificados quanto ao uso, ou via interpretação de imagens, ou através de visita *in loco*, situação na qual as imagens orbitais auxiliarão na localização dos pontos.

A seguir são descritos os passos necessários para o cálculo da estimativa de área e respectiva variância. Supõe-se que será feita uma AAS em cada município e que no grupo de municípios estudados será feita uma AAE.

O primeiro passo é o cálculo da área ocupada por uma cultura de interesse no município. Como essa área será obtida por amostragem, ela é considerada uma estimativa e será definida pelo símbolo $\hat{A}C$, onde o acento circunflexo significa “valor estimado”. Como a proporção de pixels da amostra que pertencem à cultura C , representada por p , é um estimador não tendencioso da proporção existente na população – o que foi garantido pela amostragem aleatória no município – o cálculo de $\hat{A}C$, em um município cuja área total é representada por AM , se reduz a:

$$\hat{A}C = AM \times p \quad (15)$$

A estimativa da variância de $\hat{A}C$ será, por sua vez, simplesmente:

$$\hat{s}_{\hat{A}C}^2 = \frac{AM^2}{(n)} pq \quad (16)$$

Daí segue que o desvio padrão e o coeficiente de variação (CV) podem ser assim calculados:

$$\hat{s}_{\hat{A}C} = AM \sqrt{\frac{pq}{n}} \quad (17)$$

$$CV_{\hat{A}C} = 100 \times \frac{AM}{\hat{A}C} \sqrt{\frac{pq}{n}} \quad (18)$$

Mas, substituindo $\hat{A}C$ pelo seu equivalente, conforme a equação 15, na equação 18, a fórmula do CV fica ainda mais simples:

$$CV_{\hat{A}C} = 100 \times \sqrt{\frac{q}{pn}} \quad (19)$$

O segundo passo consiste no cálculo da área plantada e da sua variância, por estrato, a partir dos valores observados em cada um dos municípios componentes da amostra retirada desse

estrato, considerando que os valores por município também são estimativas obtidas por amostragem. Nesse caso, N_h será o número de municípios no estrato h , e n_h o número de municípios amostrados no mesmo estrato. A estimativa da área ocupada com a cultura de interesse no estrato h será representada por $\hat{A}H$, que é calculada pela equação 20:

$$\hat{A}H = \frac{N_h}{n_h} \times \sum_{i=1}^{n_h} \hat{A}C_i \quad (20)$$

De maneira simples, essa equação significa que basta somar as áreas com as culturas, calculadas para cada município escolhido para compor a amostra no estrato, dividir o somatório pelo número de municípios na amostra (encontrar a média amostral) e multiplicar o resultado pelo número total de municípios no estrato, para se obter a estimativa da área total da cultura no estrato.

Como as amostras em cada município foram obtidas de maneira aleatória e de forma independente, os valores de $\hat{A}C_i$ são i.i.d. e, então, o cálculo da variância dessa nova estimativa é feito pela seguinte equação:

$$s_{\hat{A}H}^2 = \left(\frac{N_h}{n_h} \right)^2 \times \sum_{i=1}^{n_h} s_{\hat{A}C_i}^2 \quad (21)$$

O último passo para a estimativa da área plantada na região estratificada – que pode ser um Estado, um país ou qualquer outro conjunto de municípios – é o mais simples. Chamando a estimativa da área total plantada na região com a cultura de interesse de $\hat{A}R$, ela é obtida pela soma simples das áreas estimadas para cada estrato, conforme representado na equação 22:

$$\hat{A}R = \sum_{h=1}^l \hat{A}H_h \quad (22)$$

Também muito direto é o cálculo da variância dessa estimativa, como pode ser conferido pela equação 23:

$$s_{\hat{A}R}^2 = \sum_{h=1}^l s_{\hat{A}H_h}^2 \quad (23)$$

A aplicação das equações 22 e 23 só é possível porque as estimativas $\hat{A}H$ foram obtidas de tal forma que se comportam como variáveis i.i.d.

O procedimento apresentado permite estimar a área plantada com determinada cultura, em âmbito estadual, regional ou nacional, com reduzido tamanho de amostra, de forma objetiva e com precisão estatística conhecida.

CAPÍTULO 3

MATERIAL E MÉTODOS

3.1 Área de Estudo

A área de estudo localiza-se no nordeste do Estado de São Paulo e refere-se ao município de Ipuã (coordenadas W 47° e S 22°). Este município, que fica sob a jurisdição do EDR (Escritório de Desenvolvimento Regional) de Orlandia, possui uma área de 46.559 ha (IBGE, 2003a) e é caracterizado por ser eminentemente agrícola. Há um predomínio quase total de solos férteis (Latosolos Vermelhos - LV), com pequenas áreas de Gleissolos Háplicos – GX, nas várzeas, como pode ser observado na Figura 3.1 (Oliveira et al., 1999). O relevo é suave ondulado a plano, com altitudes variando de 500 m a 700 m (IBGE, 1972). O clima é considerado pela classificação de Köppen como Cwa, que se caracteriza como mesotérmico tropical de altitude, com chuvas de verão e inverno seco (Koffler, 1996).

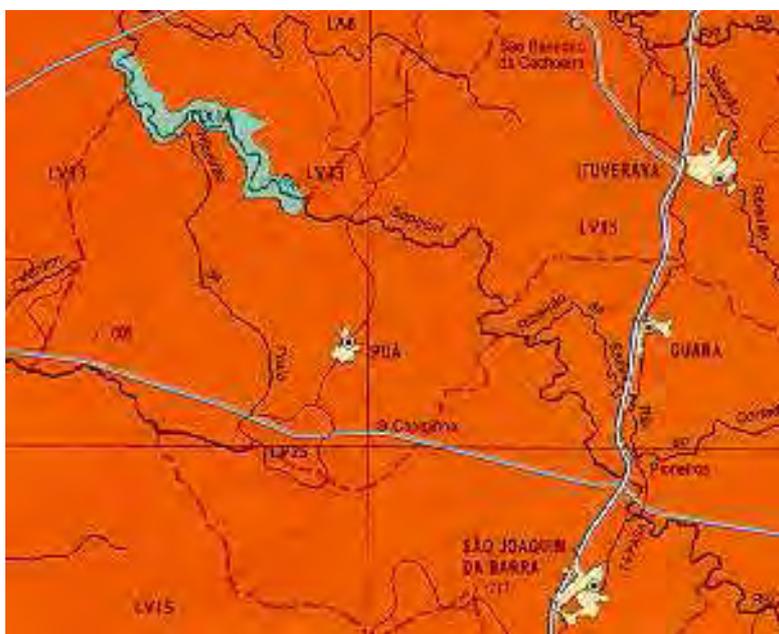


FIGURA 3.1 - Localização e classes de solo do município de Ipuã – SP
FONTE: Mapa Pedológico do Estado de São Paulo, (Oliveira et al., 1999).

Nessa região a agricultura é tecnificada, havendo várias áreas com pivô central. As principais culturas são a soja, a cana-de-açúcar e o milho. É comum a prática de dois ciclos agrícolas por ano, com plantio de milho, sorgo e milheto após o período normal de safra, geralmente chamado de safrinha. Também é possível observar a ocorrência da rotação entre cana-de-açúcar e uma cultura anual por ocasião da reforma do canavial em alguns talhões (Epiphanio et al., 2001). Dados completos sobre a produção agrícola municipal, de 1990 a 2001, podem ser conferidos no Apêndice A.

Para o georreferenciamento das imagens, optou-se pela coleta de pontos de referência *in loco* empregando um aparelho GPS III Plus da marca Garmin. Foram coletados 13 pontos de referência bem distribuídos por toda a área de abrangência do EDR de Orlândia (Figura 3.2).

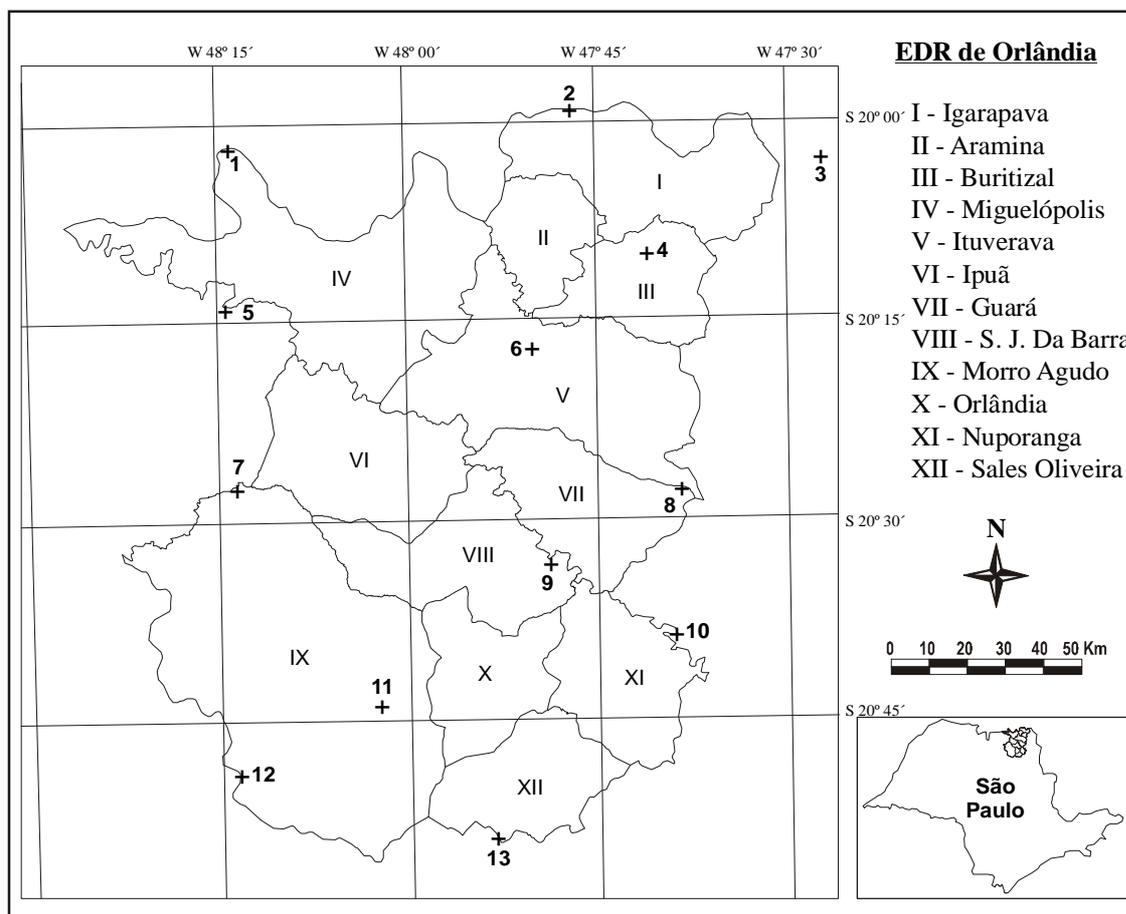


FIGURA 3.2 - Locais de coleta dos 13 pontos de referência usados no registro das imagens.

A preocupação em fazer o registro adequado aos doze municípios do EDR deve-se à intenção de uma futura expansão do trabalho para os demais municípios dela, que se beneficiaria do presente registro. A coleta dos pontos foi realizada nos dias 30 e 31 de outubro de 2001. Utilizou-se uma imagem obtida pelo sensor ETM+, do satélite Landsat 7, de 14/08/2001, órbita 220 e ponto 74 pelo Sistema Landsat de Referência Universal (WRS), para auxiliar na localização e deslocamento pela área e também para pré-selecionar pontos claramente visíveis na imagem. Essa mesma imagem, com um registro anterior feito através de dados secundários obtidos das Cartas do Brasil – Escala 1:50.000, do IBGE, já havia sido utilizada para o mapeamento completo do município, na época do inverno.

Como a intenção era fazer uma estimativa para a safra de verão e uma para o inverno, foi adquirida mais uma imagem ETM+, WRS 220/74, livre de nuvens, de 05/01/2002, a qual foi georreferenciada com o auxílio dos pontos coletados, resultando em um erro de registro de 0,478 pixels, calculado sobre os pontos de controle. O georreferenciamento foi realizado no SPRING, tendo sido empregados quatro pontos para registro (4, 5, 10 e 11), oito para controle (1, 2, 3, 7, 8, 9, 12 e 13) e um foi descartado (6) por dificuldade de localização exata na imagem. A imagem de agosto foi, então, novamente registrada com base no registro da imagem de janeiro, obtendo-se um erro de 0,223 pixels.

As imagens foram corrigidas em relação à atmosfera pelo método de subtração do pixel escuro (SPE) proposto por Chavez (1988). A seleção do pixel escuro foi feita diretamente no histograma de frequência da banda 1 de cada imagem sendo escolhido o de menor ND (número digital) com frequência significativa. A escolha do pixel escuro foi baseada em uma função de crescimento da frequência do ND desenvolvida para esse fim (Gürtler et al., 2003).

A função de crescimento escolhida expressa o aumento percentual relativo (β) da frequência do ND ao passar de um valor para o outro imediatamente superior, sendo calculada da seguinte maneira:

$$\beta_i = 100 \times \frac{F_{i+1} - F_i}{F_i} \quad (24)$$

onde: F_i é a frequência do ND_{*i*}; e F_{i+1} é a frequência do ND_{*i+1*}.

O ND para o qual a função de crescimento atingiu o valor máximo foi o escolhido, pois o custo em termos de perda de informação cresceria muito a partir desse valor. A Figura 3.3 apresenta a frequência, em escala logarítmica, de cada ND (a) e o crescimento relativo (b), para a imagem do dia 05/01/2002, cujo ND escolhido foi o 58. Para a imagem de 14/08/2001, repetindo o mesmo procedimento, o valor encontrado de ND para o pixel escuro foi o 44.

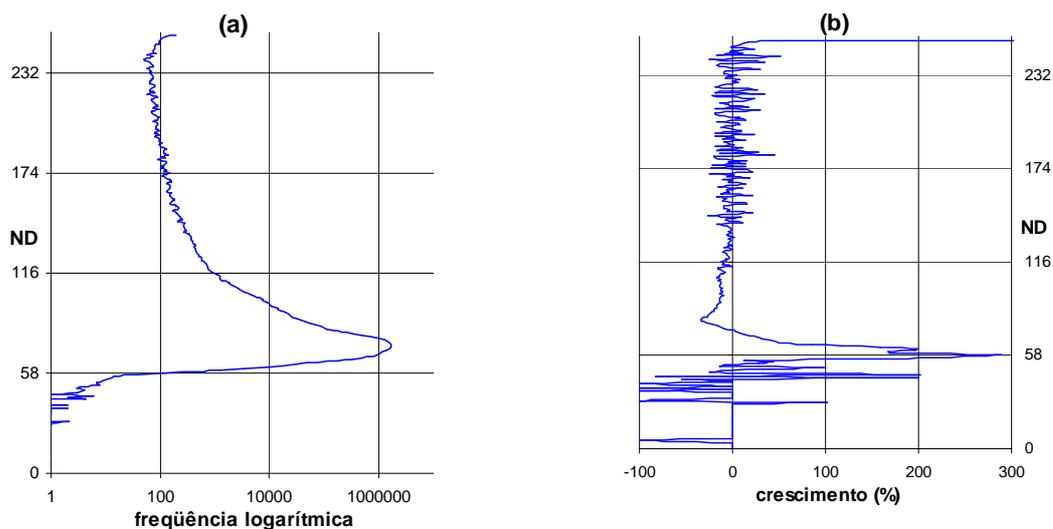


FIGURA 3.3 - Frequência de ND na banda 1 (a) e sua função de crescimento (b).

Em seguida, calculou-se, a partir do ND do pixel escuro, a quantidade de espalhamento para as demais bandas espectrais da imagem, considerando o modelo de espalhamento relativo adequado para o tipo de atmosfera. Esses valores de ND referentes ao espalhamento em cada banda foram subtraídos da imagem referência, o que resultou na imagem corrigida em relação ao espalhamento atmosférico. A Tabela 3.1 apresenta os valores de ND do pixel escuro em cada imagem e a condição da atmosfera segundo a classificação de Chavez (1988).

TABELA 3.1 - Valores do ângulo de elevação solar, pixel escuro, reflectância e tipo de atmosfera para as imagens utilizadas nos levantamentos.

Imagem	Ângulo de elevação solar (em graus)	Pixel escuro na banda 1	Reflectância	Atmosfera	Modelo de espalhamento
05/01/2002	59,18	58	7,0%	Limpa	λ^{-2}
14/08/2001	42,56	44	6,8%	Muito limpa	λ^{-4}

A imagem de inverno, época mais seca nessa região, apresentou um valor mais baixo de ND para o pixel escuro, que corresponde ao modelo de atmosfera muito limpa. Já a imagem do verão, quando há muito mais umidade e presença de maior quantidade de partículas de vapor d'água na atmosfera o valor de ND do pixel escuro foi mais elevado, o que resultou na utilização do modelo de espalhamento para atmosfera limpa.

Escolhidos os valores de pixel escuro para a banda 1, foram estimados os espalhamentos atmosféricos para esta banda. Os valores de espalhamento para as bandas 3, 4 e 5 foram obtidos a partir do espalhamento na banda 1. Foram geradas as equações para correção atmosférica de cada banda para obter os valores de reflectância. Para obter os novos valores digitais da imagem corrigida calculou-se um multiplicador para cada banda. Todo o procedimento foi descrito com mais detalhes em Luiz et al. (2003), Gürtler et al. (2003) e Gürtler (2003).

Como a área do município de Ipuã ocupa aproximadamente 1,5% de uma cena Landsat completa (quase no centro do seu quadrante inferior esquerdo), para diminuir a quantidade de memória necessária para armazenamento dos dados e aumentar a velocidade dos procedimentos eventualmente realizados sobre o conjunto de dados da imagem, como aplicação de contraste, por exemplo, foi definido o menor retângulo envolvente possível que englobasse toda a área do município e criado um Projeto no SPRING no qual a imagem foi registrada.

Sobre a imagem registrada foram digitalizados os limites municipais. Na realidade importaram-se os limites existentes na malha municipal do Brasil, fornecida pelo IBGE e que

acompanha o SPRING, sendo feitos pequenos ajustes com base nas feições na imagem e nas Cartas do Brasil – Escala 1:50.000, também do IBGE.

Para facilitar o uso da imagem durante o trabalho de campo, o material digital precisou ser transferido para a forma impressa. Para tanto, a composição colorida escolhida foi a RGB/453 que, apesar de não guardar relação visual imediata entre as feições na imagem e os alvos na cena real, permite uma boa discriminação entre os principais tipos de usos do solo presentes na área de estudo. Após algum tempo de experiência, adquirida em visitas periódicas ao campo usando imagens nesta composição, o raciocínio que permite a associação de determinado alvo real com sua representação pictórica na imagem torna-se natural e rápido.

Como exemplo, pode-se ressaltar que a soja, em seu pleno vigor vegetativo, por recobrir totalmente o solo com várias camadas de folhas intensamente verdes e planas, absorve fortemente na banda 3 e fracamente tanto na banda 4 como na 5, aparecendo na cor amarelo vivo na imagem (ETM+, WRS 220/74, de 05/01/2002, na figura 3.5). O solo, que nessa região é geralmente escuro e nessa época está quase sempre com algum grau de umidade, ao contrário, reflete mais na banda 3 e muito pouco nas bandas 4 e 5, e, portanto, aparece na imagem na cor azul escuro. A cana, que como a soja apresenta muitas folhas verdes fotossintetizantes, também absorve muito na banda 3 e reflete na banda 4; entretanto, provavelmente devido à sua arquitetura, espaçamento e ciclo (o qual determina a presença concomitante de folhas verdes na superfície e de folhas secas no interior do dossel), não reflete tanto na banda 5 como a soja, aparecendo então na cor vermelha na imagem. Sobre a cana pode-se lembrar ainda que, além da cor, outros elementos de análise facilitam sua identificação, quais sejam o padrão e a forma dos talhões, conferidos pela necessária presença dos carregadores nesta cultura.

Apesar da área do município estudado ser relativamente pequena, para permitir o manuseio mais fácil da imagem impressa e, ao mesmo tempo, produzir um material em cuja escala seja possível a identificação dos talhões agrícolas e das principais características fisionômicas do terreno, optou-se pela confecção de módulos, em tamanho A4, na escala de 1:60.000. O material nessa dimensão e nessa escala é plenamente adaptável a uma prancheta comum e ainda abrange uma área de aproximadamente 18.000 ha, o que garante uma razoável visão sinóptica, sem que se

perca a capacidade de distinção dos principais alvos agrícolas, pois, por exemplo, um talhão de forma quadrada com 9 ha de área apareceria nesse material como um quadrado de 5 mm de lado.

Entretanto, como foram necessários seis módulos nessa escala para cobrir o município, tornou-se também necessária a produção de um módulo síntese, que poderia ser chamado de fotoíndice, numa escala menos detalhada, que reproduzia o conjunto dos seis módulos em uma única prancha A4, permitindo a visualização da área total do município simultaneamente. As figuras 3.4 e 3.5 mostram o módulo síntese das imagens do inverno e do verão, respectivamente.

Para a confecção dos módulos A4 e posterior impressão, foi utilizado o aplicativo SCARTA, que é parte do programa SPRING. O primeiro procedimento executado foi a elaboração do molde de impressão, que consistiu na definição de: tamanho do papel (A4); posição da folha (paisagem); área de impressão (18,7 x 27,0cm); e escala (1:60.000). A seguir foi determinada sobre a imagem a forma de recobrimento dos módulos.

Por causa da escala e do tamanho de papel utilizados, fez-se necessária a confecção de seis módulos para cobrir toda a área de estudo. O módulo inicial foi posicionado no canto superior esquerdo da área de estudo, sendo os demais módulos localizados à sua direita na horizontal, e abaixo, na vertical. Foi deixada uma pequena área de sobreposição entre os módulos.

Com o objetivo de não sobrecarregar a imagem com textos e traços, e possibilitar anotações durante o trabalho de campo, gerou-se um conjunto auxiliar de módulos contento apenas os dados das coordenadas geográficas e os limites do município. Esses módulos, impressos em transparências, foram posteriormente sobrepostos aos seus respectivos módulos de imagem e fixados por meio de fitas adesivas, garantindo a coincidência geográfica através de marcas fiduciais comuns, deixadas para este fim. Os módulos de imagem foram impressos em papel de alta qualidade e específico para impressão de imagens ou fotografias, usando-se uma impressora colorida "jato de tinta" comum, ou seja, não foi utilizado nenhum recurso mais refinado como *plotters* ou técnicas fotográficas na produção do material de campo (Luiz et al., 2002).

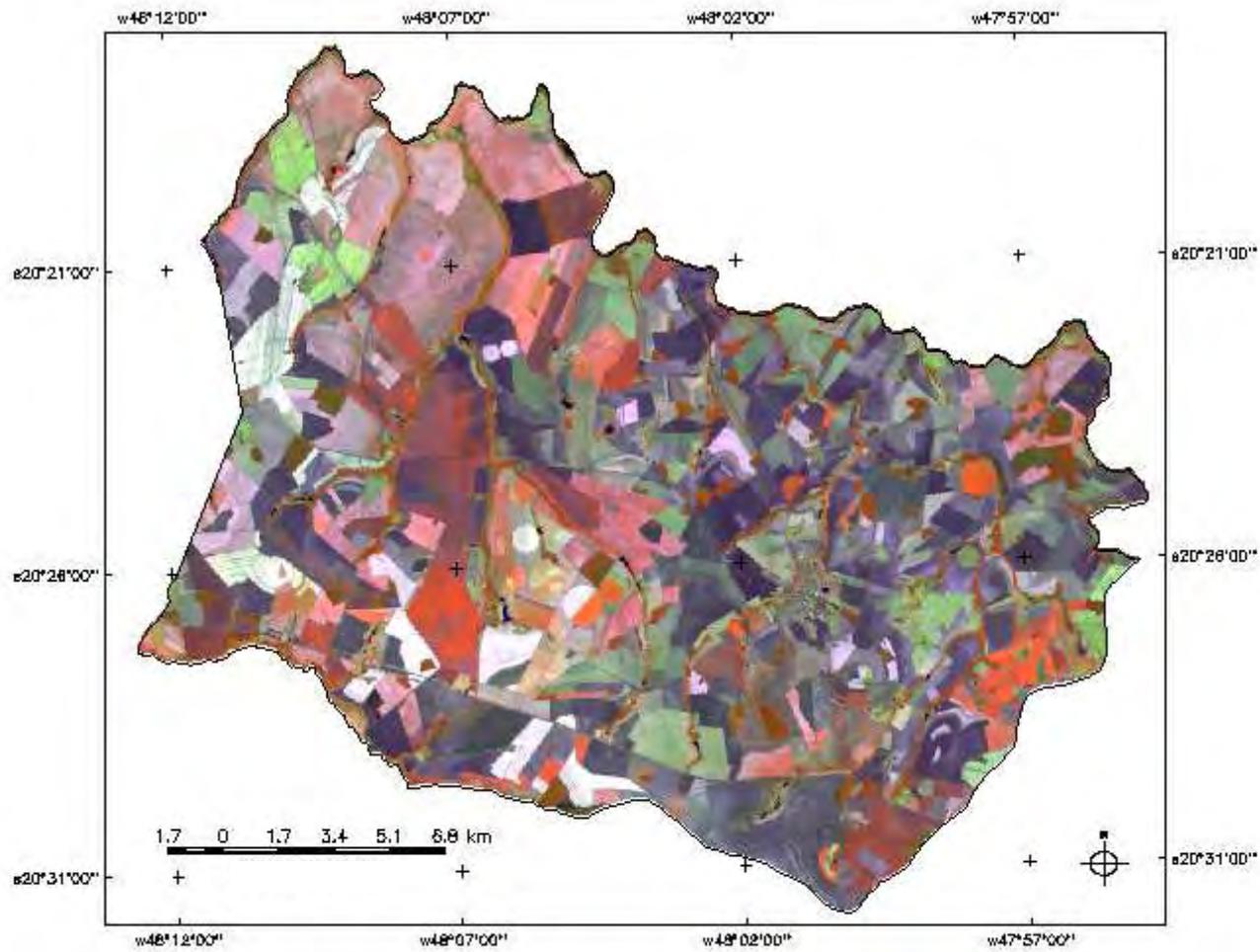


FIGURA 3.4 - Imagem RGB 453 do município de Ipuã – SP, de 14 de agosto de 2001.

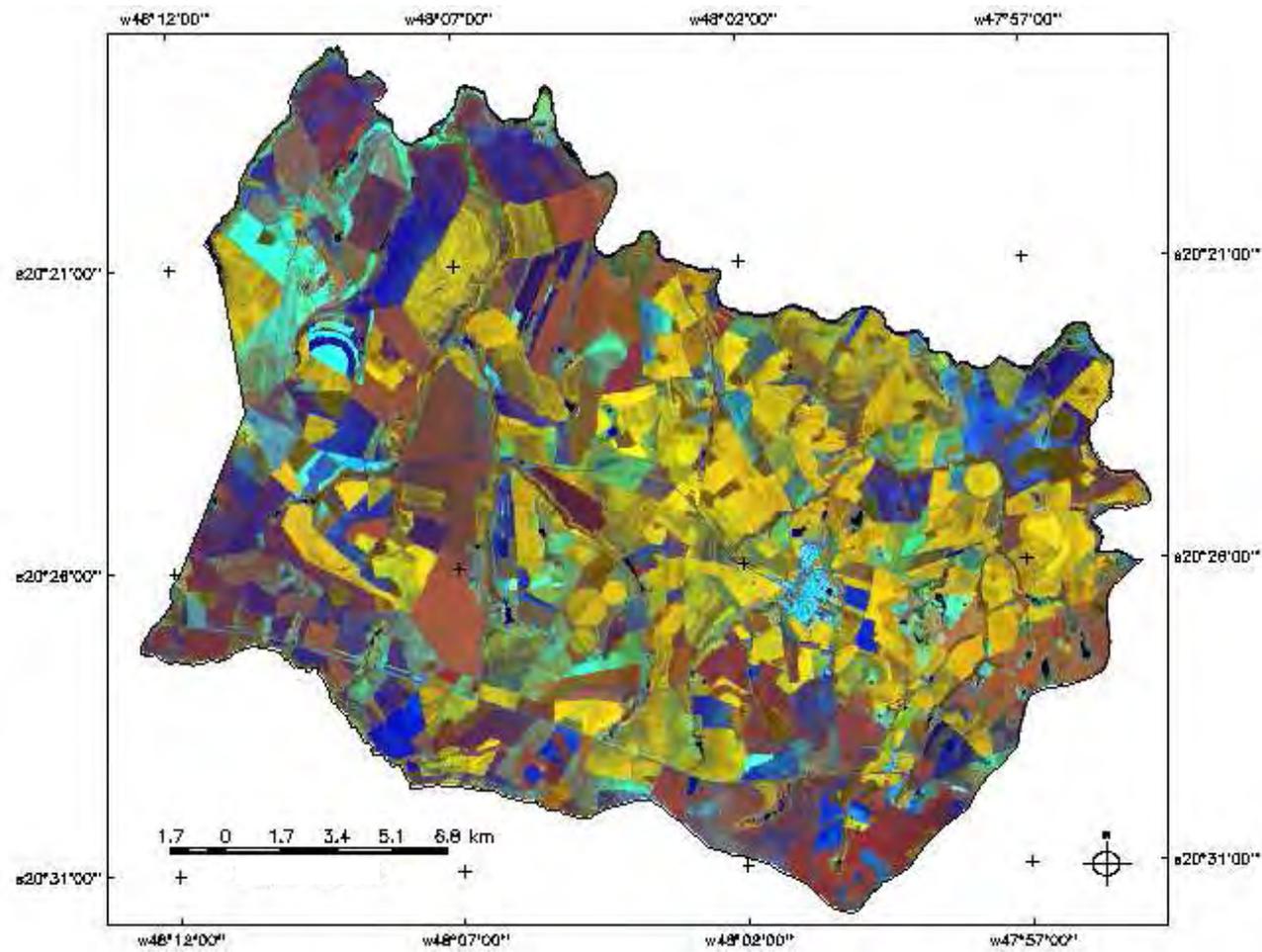


FIGURA 3.5 - Imagem RGB 453 do município de Ipuã – SP, de 05 de janeiro de 2002.

3.2 Construção da Referência

Um problema sempre presente quando se pretende avaliar a qualidade das estimativas produzidas por métodos baseados em amostragem é a obtenção de um valor de referência que possa ser considerado, com bastante confiança, como representativo da realidade. Isso é especialmente verdadeiro quando se trata de levantamentos agrícolas. Ippoliti-Ramilo et al. (2003) que encontraram diferenças de grande magnitude entre estimativas de área plantada por município, obtidas por metodologias diferentes, afirmam que, apesar de não haver dúvidas que existem alguns sérios erros nas estimativas, não é possível dizer qual metodologia é apropriada ou qual está mais próxima da realidade. Os autores indicam que apenas com um intenso trabalho de campo seria possível checar uma área suficientemente expressiva para permitir essa avaliação. Xiao et al. (2002) também encontraram, em âmbito municipal na China, grandes diferenças (de -7% a +92%) entre dados relativos à área cultivada obtidos por sensoriamento remoto e por meio de um censo e, embora afirmem que existe uma tendência, relatada por outros autores, de que os dados dos censos subestimem a área plantada “real”, preconizam cuidados com a comparação entre estimativas obtidas por mais de um método, pois se mesmo os dados do censo não são exatos, a falta de uma referência não permite saber qual estimativa é a mais correta.

Como o objetivo geral desse trabalho foi definir um método que se pudesse estabelecer como ferramenta auxiliar rotineira nos levantamentos agrícolas brasileiros (inclusive nos oficiais), podendo ser expandido para a escala estadual e nacional; e como, para se recomendar uma substituição de procedimento é necessário que se tenha confiança nos resultados do método proposto, era preciso garantir que os valores usados como referência para a comparação das áreas plantadas estimadas fossem os mais próximos possível da realidade.

Para tanto foram realizadas duas campanhas de campo com o auxílio do material preparado conforme descrito anteriormente. Usando as imagens para ajudar no deslocamento pelo território do município, foram registrados sobre as transparências de cada módulo os limites de todos os tipos de uso do solo identificáveis simultaneamente no campo e na imagem, e atribuído a cada segmento um número de identificação. Em planilhas anexas aos módulos eram anotadas as observações relativas a cada segmento identificado. Dessa forma, para cada

imagem, do inverno de 2001 e do verão de 2002, foi produzido um mapa de uso do solo completo do município de Ipuã.

De volta da primeira campanha de campo, foi impressa uma composição RGB 453 na escala 1:40.000 da imagem de 14/08/2003 e os limites registrados nos seis módulos utilizados no campo foram reunidos em uma única folha de papel vegetal, trabalhada sobre a imagem. O desenho completo do município dividido em segmentos foi então capturado para o formato digital usando-se um escaner. O arquivo digital foi importado para o SPRING, passando a compor um Plano de Informação (PI) vetorial. Através das marcas fiduciais colocadas para esse fim, esse PI foi registrado no Projeto que já continha as imagens no SPRING. Na figura 3.6 aparece o desenho do território do município de Ipuã segmentado segundo os diferentes tipos de uso do solo, do ponto de vista do interesse deste trabalho, conforme as classes identificadas durante a campanha de campo e também observadas na imagem.

Sobrepondo-se o PI vetorial dos segmentos à imagem de agosto, foi realizada uma edição matricial, atribuindo-se a cada segmento uma classe de uso, com base nas anotações das planilhas de campo. Ao final desse processo estava pronto um mapa temático que permitia o cálculo automático da área ocupada por cada tipo de uso do solo.

Com o material coletado na segunda campanha, como a localização, tamanho e forma dos talhões variaram pouco entre a imagem de agosto e a de janeiro, para produzir o mapa temático relativo à safra de verão não foi necessário repetir todo o processo de unir os módulos em um único desenho na escala 1:40.000, escanear, importar e registrar no SPRING. Foi preciso apenas duplicar o PI vetorial de agosto, sobrepô-lo à imagem de janeiro e, com base no material de campo, fazer primeiro uma edição vetorial, corrigindo os limites que foram alterados de uma data para outra, seguida de uma edição matricial, atribuindo as classes a cada segmento.

Os mapas temáticos resultantes desses procedimentos são apresentados nas Figuras 3.7 e 3.8, relativas às safras de inverno e verão, respectivamente. As áreas calculadas para as classes identificadas compõem finalmente uma referência com a qual os resultados da amostragem podem ser comparados.

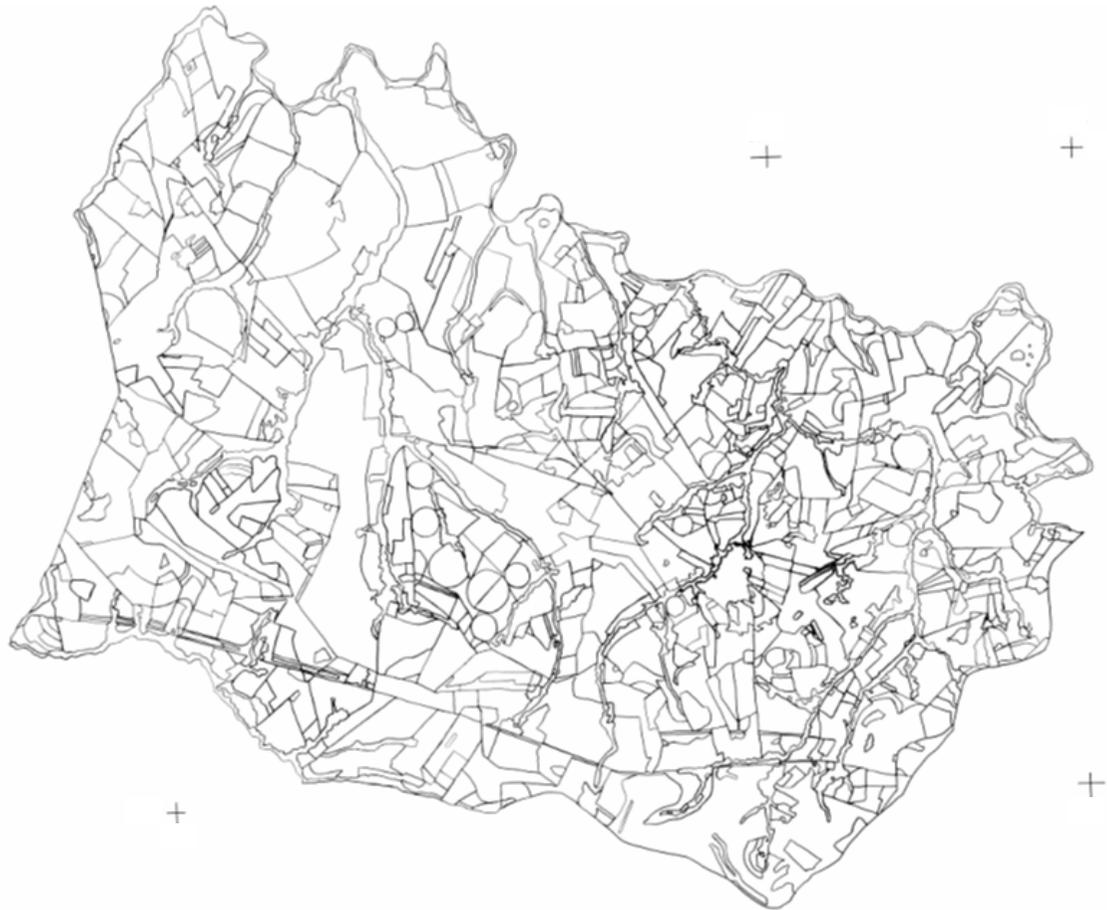


FIGURA 3.6 - Plano de informação vetorial contendo os limites dos segmentos identificados em Ipuã – SP, em agosto de 2001.

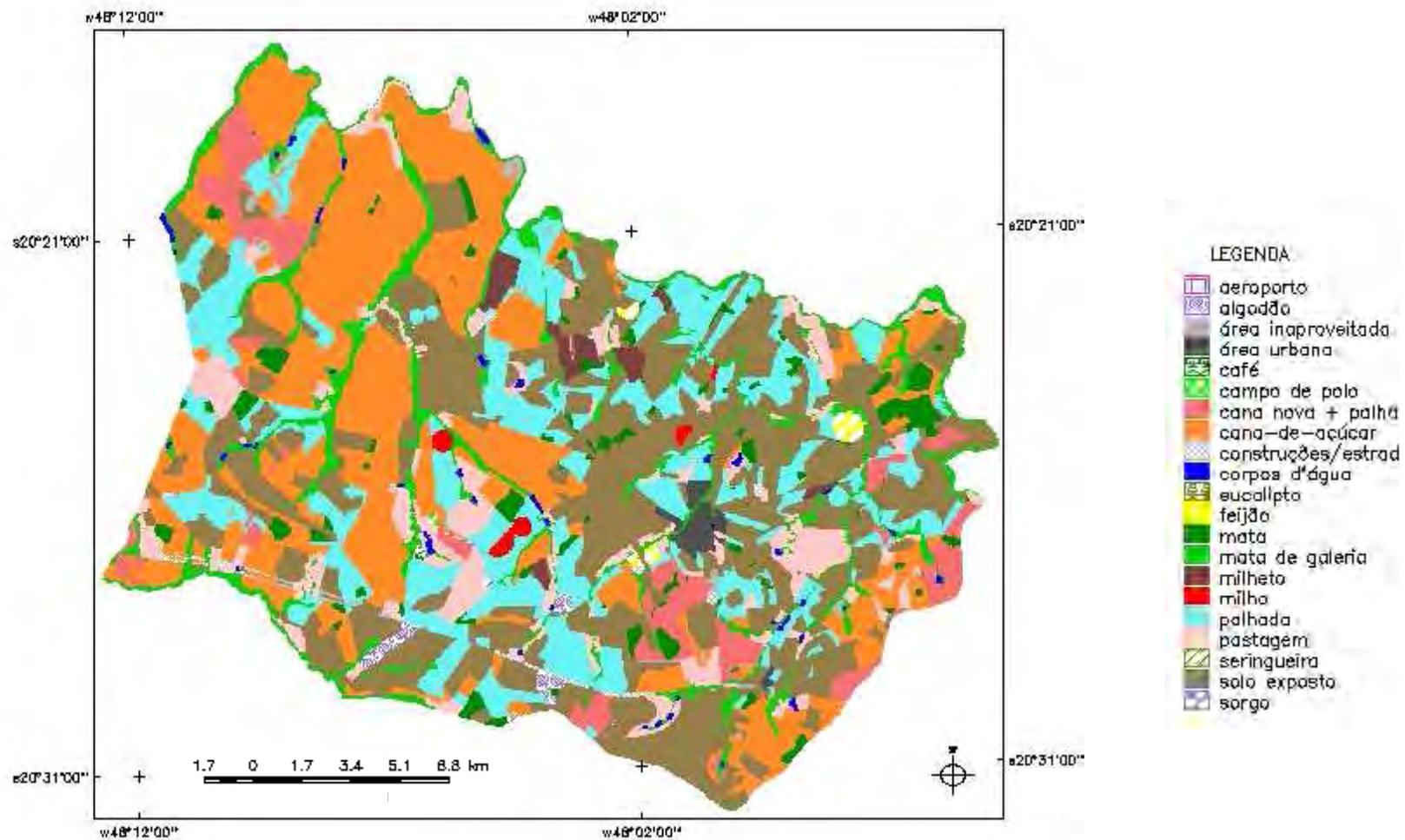


FIGURA 3.7 - Mapa temático do município de Ipuã – SP, de 14 de agosto de 2001, com as classes de uso do solo identificadas.

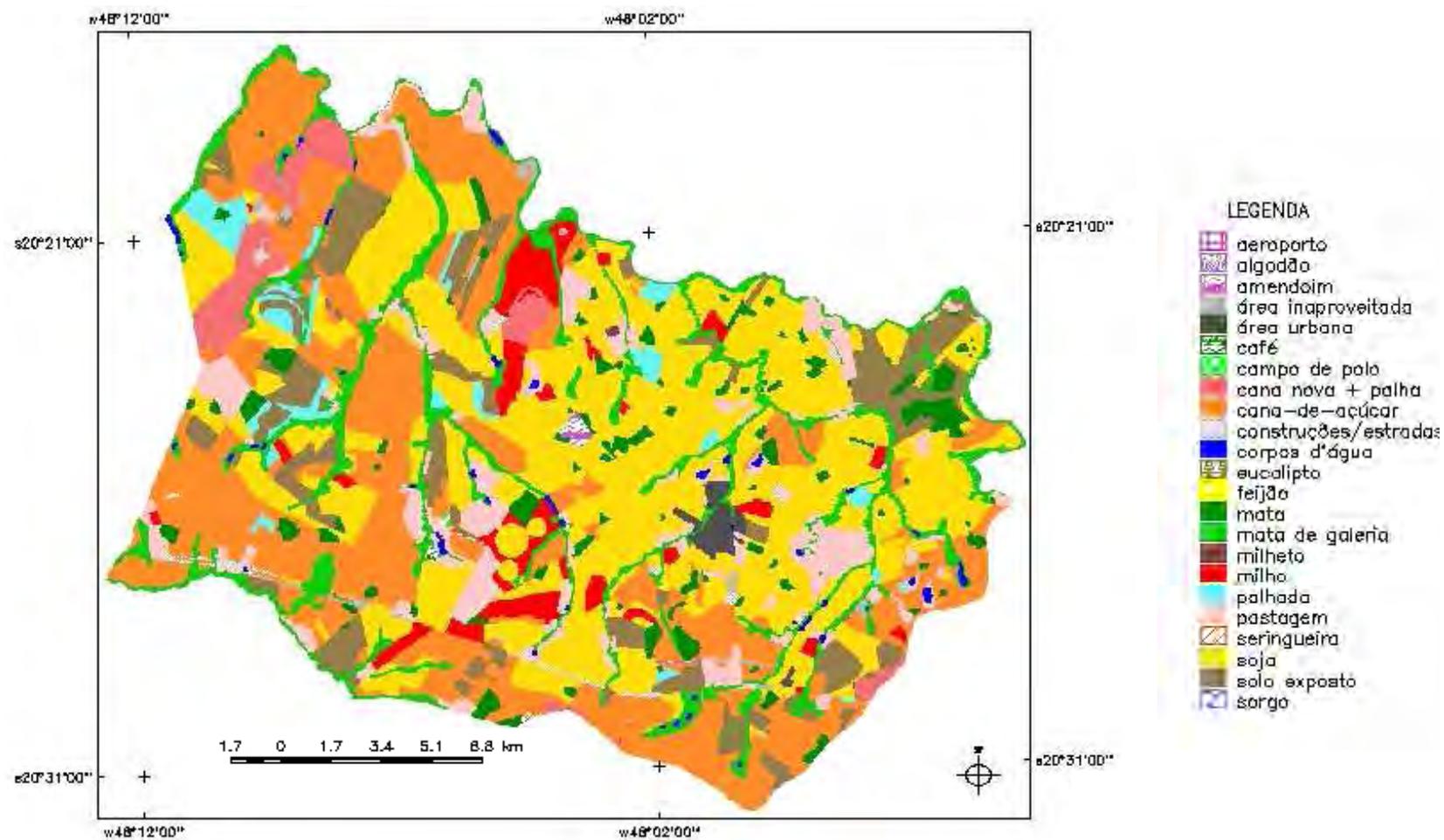


FIGURA 3.8 - Mapa temático do município de Ipuã – SP, de 05 de janeiro de 2002, com as classes de uso do solo identificadas.

3.3 Método de Amostragem por Pontos no Município

Definido teoricamente um método de amostragem auxiliado por imagem de sensoriamento remoto, escolhida uma área de estudo e construída uma referência com a qual comparar os resultados do método proposto, torna-se necessário aplicar o método e obter estimativas geradas pela amostragem para poder compará-las com a referência agora disponível.

O primeiro passo é o sorteio dos pixels para compor as amostras. Para observar o efeito do tamanho da amostra sobre a precisão das estimativas, foram sorteadas amostras de tamanho 100, 200 e 300. Para cada tamanho foram geradas 10 amostras independentes.

Devido às características irregulares das divisas municipais, para distribuir aleatoriamente os pontos amostrais na área do município realizou-se primeiro um sorteio dentro do retângulo envolvente, sobre a imagem armazenada no SPRING, fazendo a equivalência entre ponto e pixel, descartando-se depois os pontos que caíram fora dos limites municipais.

Trabalhando-se sobre a representação matricial da imagem, a posição de cada pixel foi definida pelo binômio linha e coluna. Chamou-se de x o número da coluna e y o valor da linha em uma imagem (note-se que x e y podem ser multiplicados pela dimensão do pixel, transformando sua unidade em metros, ou convertidos em coordenadas de latitude e longitude). O retângulo envolvente era formado por r linhas e s colunas e absolutamente definido pelos pares ordenados das posições de apenas dois elementos de cena, um no canto inferior esquerdo (x_1, y_1) e o outro no canto superior direito (x_2, y_2). Observa-se que $r = (|x_2 - x_1| + 1)$ e que $s = (|y_2 - y_1| + 1)$ e que o número total de pixels do retângulo é igual ao produto entre r e s , que será chamado de T . Para permitir a realização da escolha aleatória dos m candidatos a pontos amostrais entre os T pixels era necessário gerar m pares ordenados (x_i, y_i). Como o gerador utilizado retorna dois números aleatórios z $\{0 \leq z < 1\}$ distribuídos uniformemente, basta aplicar m vezes a transformação: $x_i = (z_1 \times r)$ e $y_i = (z_2 \times s)$.

Esses pontos foram considerados candidatos pois, como nem todo o retângulo era ocupado pela área do município, uma fração destes m elementos estava localizada fora dos limites da região de interesse e não foram amostrados. Apenas n elementos sorteados, cuja localização

estava dentro dos limites determinados, foram considerados pontos amostrais. Como a escolha dos pontos foi aleatória, esperava-se que, em média, a relação entre n e m fosse a mesma existente entre a área do município (A) e a área do retângulo envolvente (R). Desta relação foi possível determinar a melhor estimativa de m que produzisse um determinado valor de n , que foi calculada da seguinte forma: $m = \{1 + \text{INT}[n \times (R/A)]\}$, onde INT significa a parte inteira da expressão entre colchetes.

Considerando a área de Ipuã como 46.837,4 ha (IBGE, 2003a), e a do retângulo envolvente como 86.265 ha; para selecionar amostras de tamanho $n_1 = 100$, $n_2 = 200$ e $n_3 = 300$, seria necessário sortear um total de: $m_1 = \{1 + \text{INT}[100 \times (86.265/46.837,4)]\} = 185$, $m_2 = 369$, e $m_3 = 553$; ou seja, gerados 185 pontos dentro do retângulo envolvente, a esperança matemática era de que 100 deles estivessem localizados dentro dos limites do município, e assim, da mesma forma, para m_2 e m_3 . Para aumentar a probabilidade de conseguir o número desejado de pontos dentro da região de interesse, e como o custo de geração desses números aleatórios é computacionalmente muito baixo, gerou-se o dobro desses pontos, ou seja, 370 para garantir 100, 738 para 200, e 1106 para 300.

Para transformar os pares ordenados sorteados (x_i, y_i) em valores compatíveis com as coordenadas planas (X_i, Y_i) do retângulo envolvente definido no SPRING, no caso da imagem com resolução espacial de 25 metros, utilizou-se a seguinte transformação: $X_i = X_1 + 12,5 + [\text{INT}(x_i \times 25)]$ e $Y_i = Y_1 + 12,5 + [\text{INT}(y_i \times 25)]$, na qual o par (X_1, Y_1) define a posição da extremidade sudoeste do retângulo envolvente, no caso do hemisfério Sul. Esse procedimento garantiu que cada um dos pixels da imagem tivesse a mesma chance de ser escolhido para a amostra, e que a escolha se desse pelo centro do pixel.

As coordenadas dos pontos gerados aleatoriamente, conforme descrito, foram inseridas no projeto do SPRING via importação, na forma de arquivos de pontos em formato ASCII-SPRING, o que gerou um PI para cada conjunto de pontos. Para eliminar os pontos que não estavam dentro do polígono que corresponde ao limite municipal de Ipuã efetuou-se o recorte desses PIs, utilizando uma máscara com o limite. Como os PIs resultantes continham um número de pontos superior ao desejado, realizou-se a exportação dos PIs para possibilitar a

abertura em uma planilha Excel. Os pontos foram, então, classificados em ordem crescente e destas listagens foi selecionado um conjunto com o número de pontos desejado para cada arquivo, ou seja, os primeiros 100, os primeiros 200 e os primeiros 300 pontos. Os demais pontos foram excluídos dos arquivos, que foram novamente importados para o SPRING, concluindo a etapa. O processo, como dito, foi repetido dez vezes para cada tamanho de amostra. As figuras 3.9, 3.10 e 3.11, exemplificam a distribuição dos pontos amostrais pelo território do município de Ipuã, para amostras de tamanho 100, 200 e 300, respectivamente.

Geradas as amostras, os PIs dos pontos sorteados foram cruzados com os PIs temáticos das classes observadas a campo, e o resultado era armazenado em um outro PI da categoria cadastral, onde a cada ponto da amostra correspondia uma classe de uso do solo. A partir da exportação dessas tabelas de dados cadastrais para uma planilha Excel, os dados eram trabalhados para gerar as estimativas. O primeiro procedimento consistiu em uma contagem de pontos por classe, seguida de uma reunião de classes de menor interesse e frequência, de forma a diminuir o número total de classes. As classes consideradas ao final para análise foram: 1) Cana-de-açúcar (cana e cana+palha); 2) Soja; 3) Milho; 4) Outras Culturas (feijão, amendoim, algodão, milheto, sorgo, café, eucalipto, e seringueira); 5) Vegetada (mata, mata de galeria, pastagem e campo de polo); 6) Não Vegetada (aeroporto, área inaproveitada, corpos d'água, área urbana, construções e estradas); 7) Palha; e 8) Solo Exposto.

Para cada uma dessas classes foi estimada a área plantada no município, com base nas amostras com 100, 200 ou 300 pontos. O cálculo da área era feito dividindo-se a frequência observada para a classe pelo tamanho da amostra e multiplicando-se pela área total do município. A razão entre a frequência observada e o tamanho da amostra nada mais é do que a proporção p de ocorrência da classe na amostra, que é um estimador não tendencioso da proporção P da classe na população. Assim, bastava multiplicar o p de cada classe pela área total do município AM para obter a área estimada para aquela classe, isto é:

$$\hat{A} = p \times AM \quad (25)$$

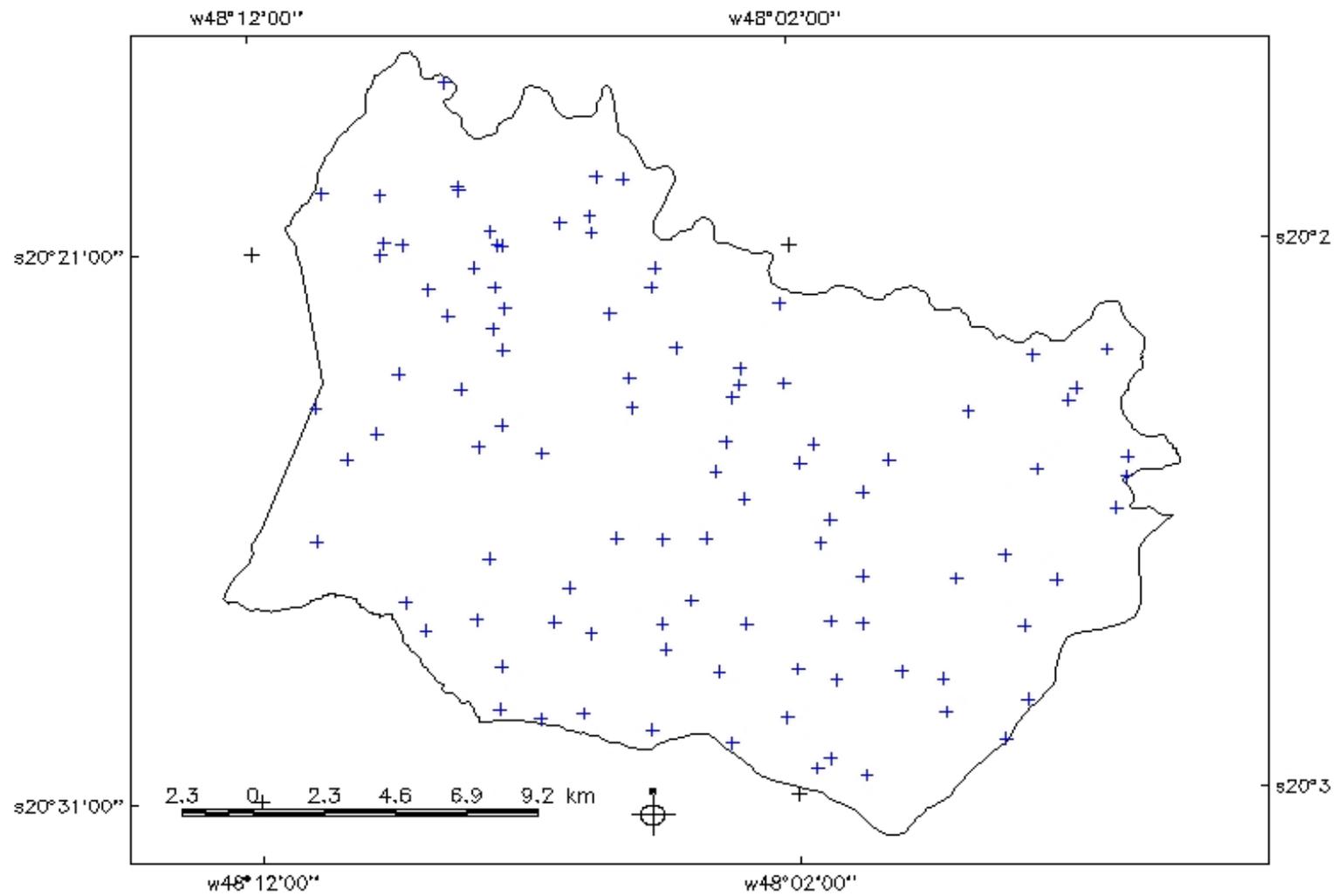


FIGURA 3.9 - Exemplo de distribuição de 100 pontos, escolhidos para compor uma amostra aleatória, em Ipuã – SP.

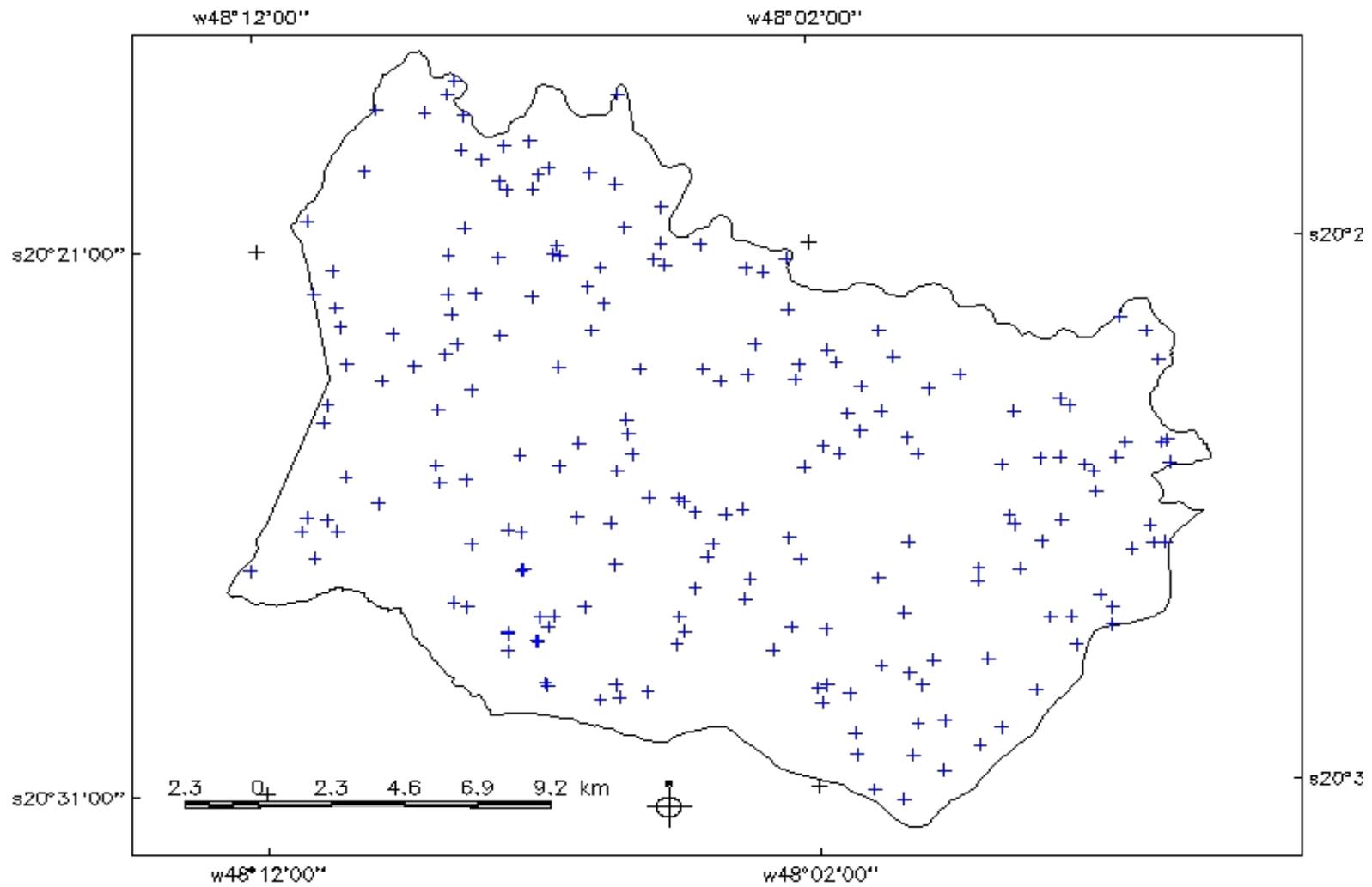


FIGURA 3.10 - Exemplo de distribuição de 200 pontos, escolhidos para compor uma amostra aleatória, em Ipuã – SP.

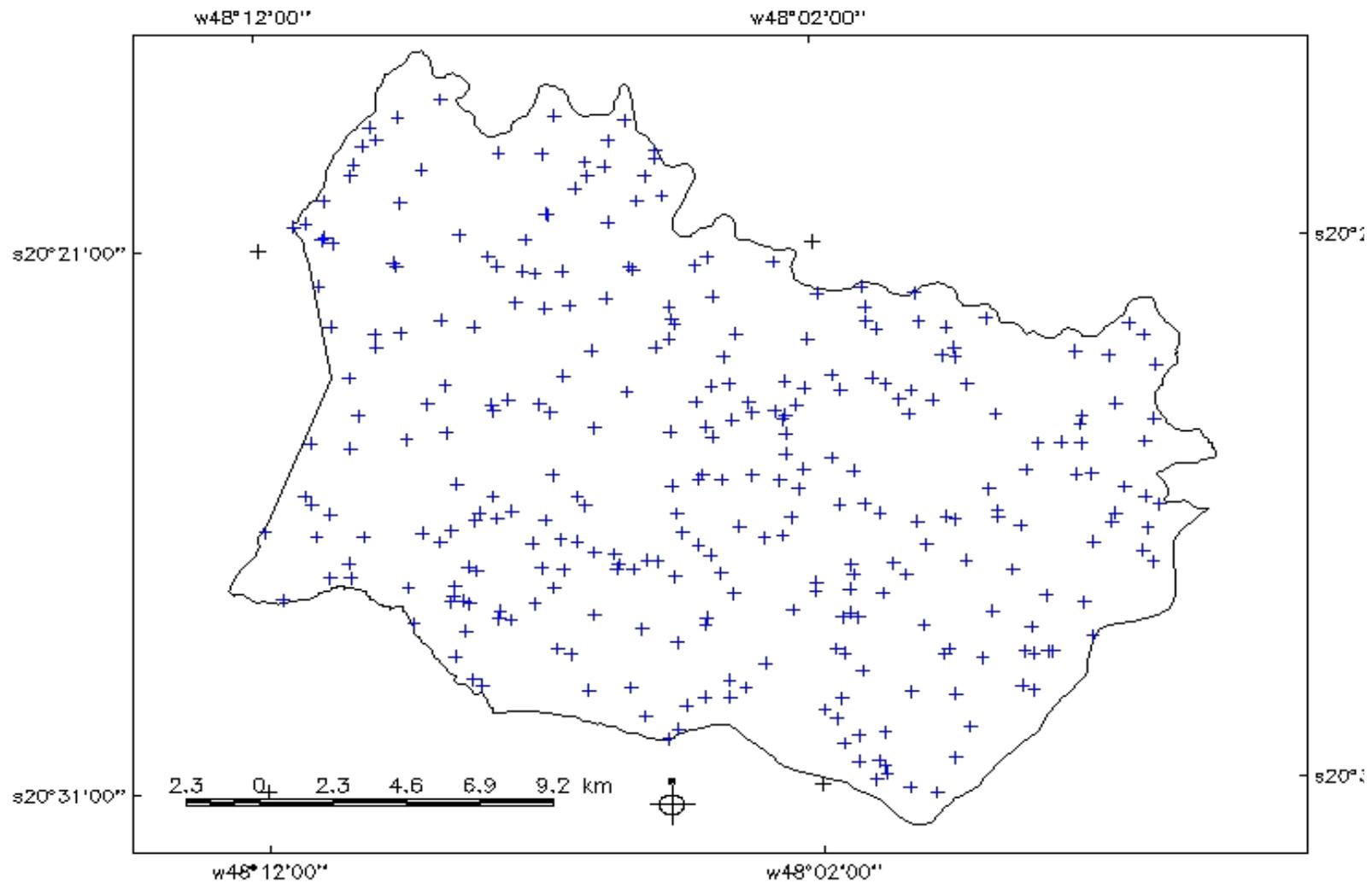


FIGURA 3.11 - Exemplo de distribuição de 300 pontos, escolhidos para compor uma amostra aleatória, em Ipuã – SP.

Já para o cálculo do CV, o procedimento também foi muito simples, bastando multiplicar a proporção p de cada classe pelo seu complemento q , $\{q = 1 - p\}$, dividir pelo tamanho da amostra n e multiplicar pelo quadrado da área do município AM , tudo multiplicado por 100 para ser expresso em porcentagem. Ou seja:

$$CV = 100 \times \frac{pq}{n} \times AM^2 \quad (26)$$

3.4 Estratificação no Município

Visando avaliar o efeito da estratificação dentro do município na precisão das estimativas, o território de Ipuã foi dividido em dois estratos. O estrato I foi estabelecido de maneira a cobrir aproximadamente dois terços da área total (27.858 ha), localizando-se sobre o centro e o leste do município, sempre ao norte da estrada SP 345, que liga a Rodovia SP 330 (Anhanguera) ao município de Guaíra. O estrato II, conseqüentemente, estava situado no lado oeste do município, vizinho à Guaíra, e incluía ainda toda a área ao sul da estrada SP 345, representando o restante um terço da área total (18.979 ha).

A estratificação foi realizada manualmente, tendo sido baseada na interpretação visual da imagem, no conhecimento de campo adquirido ao longo das diversas campanhas realizadas no município. Procurou-se ainda utilizar limites naturais, como córregos, e antrópicos, como a estrada, para o traçado dos limites dos estratos. A principal característica visual que influenciou a definição dos estratos foi a presença das culturas da cana-de-açúcar e da soja. O estrato I é predominantemente ocupado por lavouras de soja, enquanto no estrato II é mais frequente a cana-de-açúcar.

O mesmo procedimento de sorteio de pontos foi realizado nessa etapa, sendo que o número de pontos sorteados foi sempre 200 para o município. Foram feitas duas distribuições entre os estratos, uma supondo um maior interesse na estimativa da área de soja e outra com maior interesse pela cana-de-açúcar. Nos dois casos foram alocados 86 pontos em um estrato e 114 pontos no outro. Esses números resultaram do cálculo da variância esperada, dada pela fórmula 5. Com base nos valores de referência para soja e cana nos estratos I e II, encontrou-se o valor da proporção p de cada cultura nos dois

estratos. A alocação ótima dos pontos amostrais é aquela que produz a menor soma de variâncias para uma cultura nos dois estratos. Nesse caso, apenas por coincidência, a combinação ótima para a soja ocorreu com 114 pontos no estrato I e 86 no estrato II; e para a cana, exatamente o inverso, ou seja, 86 pontos no estrato I e 114 no II.

Um exemplo de sorteio com 114 pontos no estrato I e 86 no estrato II é apresentado na figura 3.12. Essa distribuição favorece, em termos de precisão, a estimativa da área plantada com soja, pois a área com maior número de pontos é justamente a que contribui com a maior parte da soja plantada no município.

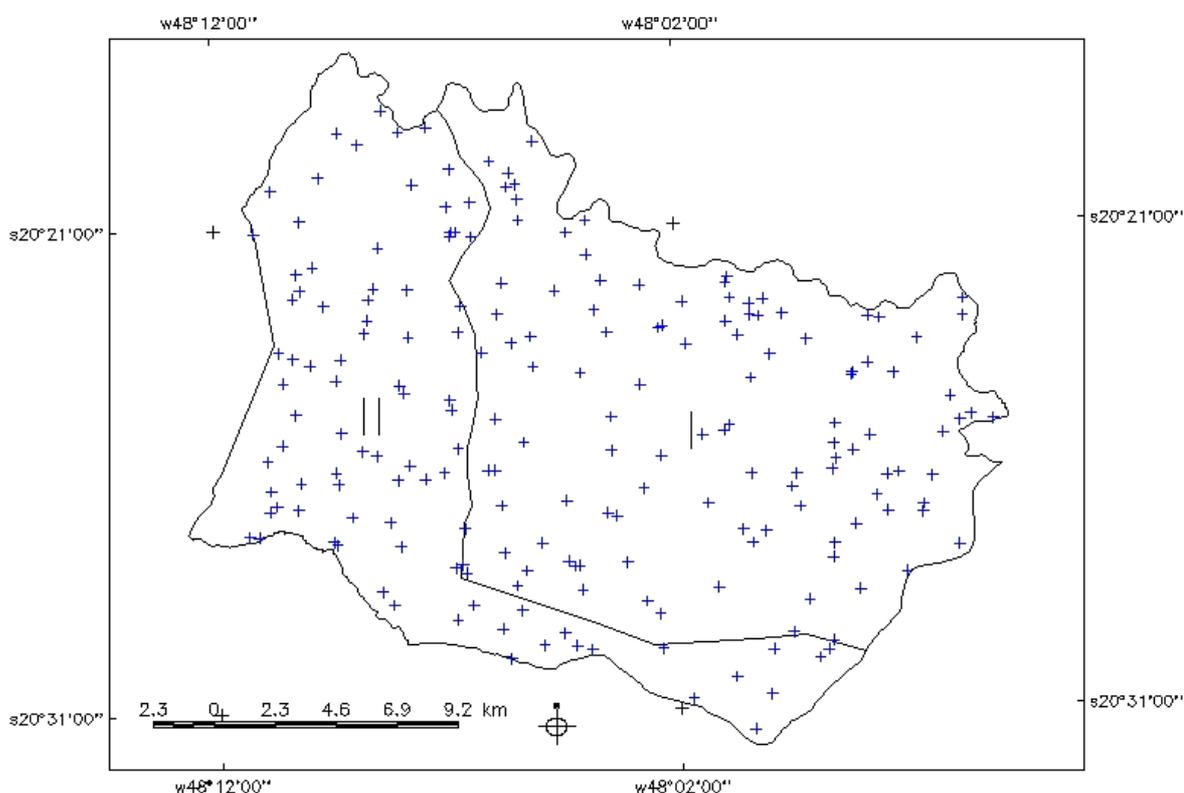


FIGURA 3.12 - Exemplo de distribuição de 200 pontos, escolhidos para compor uma amostra aleatória estratificada, com 114 pontos no estrato I e 86 no estrato II, em Ipuã – SP.

Na figura 3.13 aparece um exemplo oposto ao anterior, ou seja, foram distribuídos 114 pontos no estrato II, preferencialmente ocupado pela cana, e 86 pontos no estrato I, que apresenta maior frequência da cultura da soja. Essa distribuição privilegia a estimativa

da área plantada com cana, pois intensifica a amostragem na área onde ela é predominante.

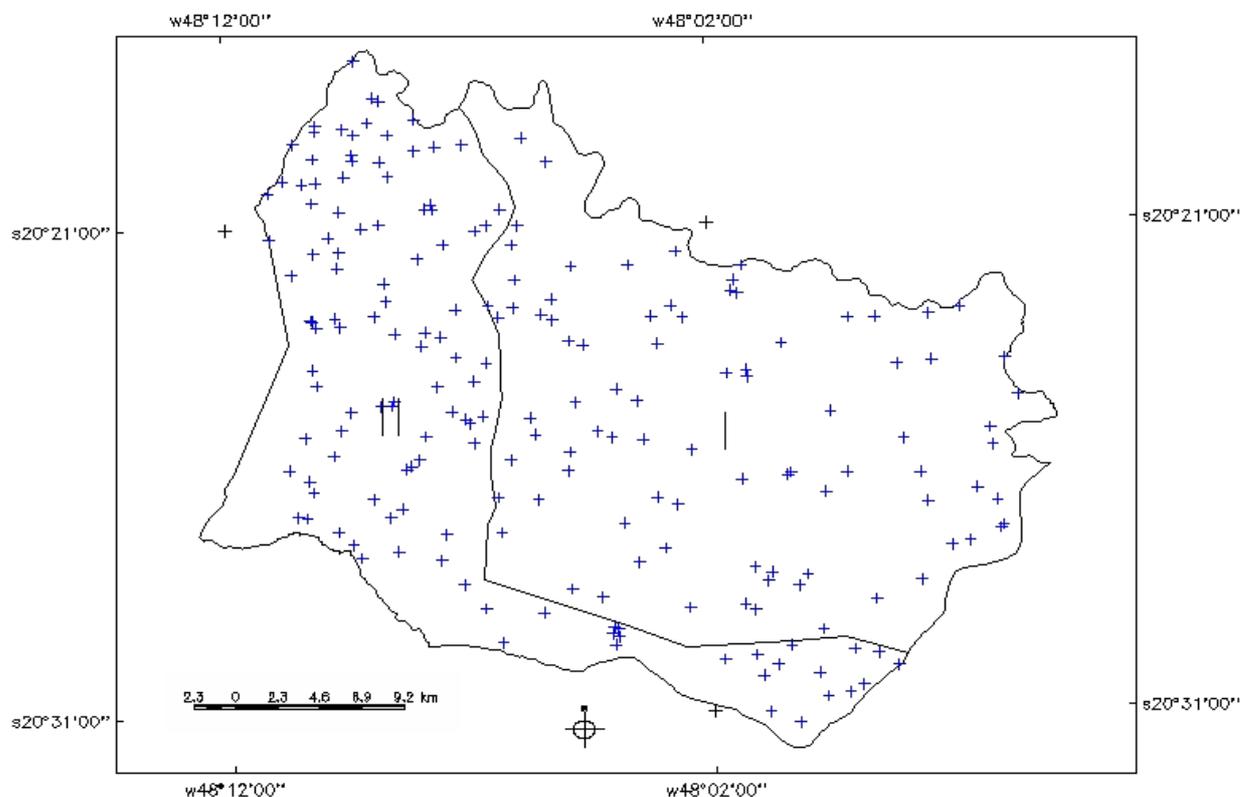


FIGURA 3.13 - Exemplo de distribuição de 200 pontos, escolhidos para compor uma amostra aleatória estratificada, com 86 pontos no estrato I e 114 no estrato II, em Ipuã – SP.

Para proceder ao cálculo da estimativa de área plantada nos estratos e no município, o procedimento foi inicialmente o mesmo que para o caso da AAS. Após obter a frequência de cada classe em cada estrato, a área da cultura estimada no estrato foi obtida simplesmente pela multiplicação da proporção encontrada pela área do estrato AH , isto é:

$$\hat{A}_h = p_h \times AH \quad (27)$$

Para se calcular o coeficiente de variação no estrato CV_h o procedimento também foi o mesmo que no caso da AAS, ou seja:

$$CV_h = 100 \times \sqrt{\frac{q_h}{n_h p_h}} \quad (28)$$

O procedimento sofreu alguma alteração para se calcular a estimativa para o município todo, considerando os estratos. Mesmo assim, para o cálculo da área total do município bastou somar a área estimada para cada estrato, ou seja:

$$\hat{A} = \hat{A}_1 + \hat{A}_2 \quad (29)$$

Já para calcular o CV foi necessário levar em conta o peso de cada estrato, que foi calculado simplesmente dividindo a área de cada estrato pela área total do município, isto é:

$$W_h = \frac{AH}{AM} \quad (30)$$

Dessa maneira, o CV da estimativa para o município, obtido por AAE, foi calculado pela seguinte expressão:

$$CV_{\hat{A}} = 100 \times \frac{\sqrt{W_1^2 \frac{p_1 q_1}{n_1} + W_2^2 \frac{p_2 q_2}{n_2}}}{\hat{p}_{est}} \quad (31)$$

3.5 Estratificação em um Conjunto de Municípios

Para poder expandir o método de amostragem por pontos do âmbito municipal para o regional, estadual ou mesmo nacional, pode-se lançar mão novamente da AAE, dividindo a população de municípios que se tem interesse de estudar em estratos e tomando apenas alguns municípios de cada estrato como amostra. Apesar de se supor que esse procedimento possibilita a estimativa da área plantada de grandes culturas, mesmo em escala nacional, com um pequeno tamanho de amostra, em termos de número de municípios amostrados, não era possível realizar tal levantamento no escopo deste trabalho. Entretanto, para descrever integralmente o método proposto e verificar

como ele se comporta em termos de valores possíveis de confiabilidade dos resultados, foi escolhida uma cultura no Estado de São Paulo, a cana-de-açúcar, e simulada a aplicação do método, utilizando-se para isso dos dados de área plantada fornecidos pelo IBGE. Os dados utilizados foram os relativos à área colhida por município, no ano de 2001, obtidos do sistema SIDRA, provenientes da “Tabela 1612 - Quantidade produzida, Valor da produção, Área plantada e Área colhida da lavoura temporária”, aos quais obteve-se acesso por meio eletrônico (IBGE, 2003b).

Obtidos os dados de área plantada por município, foram criados, com o auxílio do procedimento FASTCLUS do aplicativo SAS (SAS, 1988), agrupamentos baseados na distância euclidiana mínima, variando-se o número de agrupamentos de seis a quatorze. Na realidade, foram obtidos na página do IBGE os dados para diversas outras culturas, e foram feitas as análises de agrupamento para a cultura do café nos Estados de Minas Gerais, São Paulo e Espírito Santo; para a cultura do milho no Estado do Paraná; e para a soja, nos Estados do Mato Grosso, Paraná e Rio Grande do Sul. Para todas essas combinações foram gerados de seis a quatorze estratos. Todo o procedimento descrito para a cana-de-açúcar em São Paulo foi repetido para cada um dos outros casos, mas por motivo de espaço e para não tornar o texto muito repetitivo, apenas os dados dessa cultura serão discutidos. No Apêndice B está o programa, escrito em linguagem SAS, usado para criar os grupos – ou estratos -, que pode ser aplicado para qualquer conjunto de municípios e para qualquer cultura.

O primeiro passo para criar os estratos no grupo dos municípios do Estado de São Paulo foi ordená-los pela variável de interesse – no caso a área colhida – do menor valor para o maior valor. Os municípios onde a cultura não aparecia foram retirados da lista e, no caso de um sistema oficial de estatísticas agrícolas, eles comporiam um estrato com acompanhamento diferenciado, para identificar possíveis mudanças no perfil agrícola ao longo do tempo. Com os 383 municípios onde a cana-de-açúcar aparecia em 2001 fez-se a análise de agrupamento usando o programa em SAS registrado no Apêndice B. A saída do programa cria um identificador para cada município que o associa a um estrato.

Os estratos gerados tinham números diferentes de municípios, mas de cada estrato foi retirada de forma aleatória uma amostra de tamanho três. As exceções foram os estratos com apenas três ou menos municípios. Nesses casos não se retiravam amostras pois todos eram incluídos no estudo.

Compostos os estratos e as suas amostras, em cada município selecionado era simulada a aplicação da amostragem por pontos. Recorreu-se novamente aos valores fornecidos pelo IBGE, tanto de área colhida como da área total do município, e calculou-se para cada um deles o valor da proporção p de ocorrência da cana-de-açúcar. Para simular a aplicação do método bastou multiplicar p pelo tamanho n da amostra escolhida, que nesse caso foi de $n = 100$. Como os pontos são classificados totalmente em uma classe ou em outra, os valores resultantes da multiplicação foram aproximados para o inteiro mais próximo. Dessa maneira, para cada município selecionado obteve-se um número de pontos que, ficticiamente, haviam sido distribuídos aleatoriamente pelo território do município e sua localização tinha coincidido com lavouras de cana-de-açúcar.

A partir daí, todo o procedimento de cálculo das estimativas pôde ser feito. Primeiro, foi realizado o cálculo da estimativa da área colhida por município, que usou a mesma equação 25, já apresentada para a AAS, onde $\hat{A} = p \times AM$. O CV também é calculado de forma idêntica ao do CV na AAS, conforme a equação 26, onde $CV = 100 \times (pq/n) \times AM^2$.

Em cada estrato, a área colhida era simplesmente a soma das áreas dos municípios, no caso dos estratos com três elementos ou menos, como mostrado na equação 29. No caso de estratos com pelo menos quatro municípios, era necessário dividir a soma por três e multiplicar pelo número de elementos no estrato, como apresentado na equação 20.

Para a obtenção do CV do estrato já não foi possível obtê-lo diretamente como nos casos anteriores. Foi necessário nesse caso o cálculo da variância do estrato, que era obtido pela aplicação da equação 21 já apresentada. Note-se que nessa equação, quando o número total de municípios no estrato N_h é igual ao número de municípios na amostra n_h , o cálculo da variância do estrato resume-se à soma das variâncias de cada elemento

da população. A partir da variância, foi calculado o desvio padrão e então o CV dos estratos.

Finalmente, para o cálculo da área estimada total ocupada pela cana-de-açúcar no Estado de São Paulo, foram simplesmente somadas as áreas estimadas para cada estrato. O mesmo procedimento foi realizado com as variâncias, que podem ser somadas diretamente, pois as amostras dentro dos estratos e os estratos entre si são independente e identicamente distribuídos.

CAPÍTULO 4

RESULTADOS E DISCUSSÃO

4.1 Dados de Referência

Os valores computados pelo SPRING para as classes identificadas e agrupadas, com base nos dados coletados a campo e digitalizados na forma de mapas temáticos, nas datas referentes ao inverno de 2001 e verão de 2002 são apresentados na tabela 4.1. Esses foram os valores adotados como referência para a avaliação dos resultados obtidos por amostragem, por serem considerados os mais próximos possível da realidade.

TABELA 4.1 - Área ocupada com diferentes classes de uso do solo, em hectares, no município de Ipuã - SP, em quatro datas.

Uso	1996/1997	Fevereiro de 2000	Agosto de 2001	Janeiro de 2002
Cana-de-açúcar	19.441	14.500	15.096	14.711
Soja	36.504	19.500	0	16.155
Milho	17.124	19.500	194	1.546
Outras culturas	967	1.927	817	112
Vegetada	2.898	-	8.906	8.891
Não vegetada	-	-	716	722
Palha	-	-	8.115	970
Solo exposto	-	-	12.994	3.730
Total	76.934	55.427	46.838	46.837

Dados da 1996/1997 segundo Pino et al. (1997), dados de fevereiro de 2000 fornecidos pelo EDR de Orândia e dados de agosto de 2001 e janeiro de 2002 obtidos neste trabalho.

É importante ressaltar que um dado dessa tabela já chama a atenção. Quando foi escolhida essa região para o estudo, entre diversos motivos estava o fato de ser uma região importante para as culturas da soja, da cana-de-açúcar e do milho. Segundo a “estimativa subjetiva de previsão de safras do anos agrícola de 1999/2000” realizada no mês de fevereiro de 2000 pelo Escritório de Desenvolvimento Rural (EDR) de Orândia, a área plantada em primeira planta com milho em Ipuã era de 2.500 ha e a área de milho

safrinha era estimada em 17.500 ha. Nota-se só aí uma discrepância imensa entre os números em um curtíssimo intervalo de tempo, especialmente no caso do milho safrinha.

Possivelmente devido a uma seqüência de anos muito secos no período da safrinha do milho, combinado com a melhora do preço da soja e da cana-de-açúcar, tenham levado um grande número de agricultores a abandonar a intenção de plantar o milho safrinha, tendência que não foi percebida a tempo pela metodologia subjetiva, mas apenas pelo método mais objetivo da medição das áreas.

Outro ponto a comentar é que, como se pode ver, a soma das áreas de todas as classes sempre resulta no mesmo valor, igual à área do município, pelo método baseado no uso de imagens de sensoriamento remoto. Como o método utilizado era praticamente instantâneo – a imagem é obtida em frações de segundo e o trabalho de campo consumiu menos de dez dias úteis na primeira campanha e menos de cinco na segunda – ele realmente “retrata” a situação naquele momento, não sendo suficiente o intervalo de aquisição dos dados para que ocorram mudanças no cenário agrícola do município. Além disso, pelas características já mencionadas da imagem de sensoriamento remoto e do SIG usado no seu processamento, o método não permite que haja falha de cobertura nem superestimação. Pode parecer banal, mas o fato de que a soma das classes obrigatoriamente será igual a área do município a cada levantamento, evita erros grosseiros que, infelizmente, podem acontecer no caso de levantamentos subjetivos e com período de coleta dos dados muito dilatado.

4.2 Estimativas por Amostragem Aleatória Simples em Ipuã

Foram geradas dez amostras tamanho 100, dez tamanho 200 e dez tamanho 300. Cada uma delas teve seus pontos cruzados automaticamente com o mapa temático de uso do solo, produzindo as estimativas de área ocupada com cada uso, com os respectivos desvios. Os dados assim obtidos puderam ser comparados com a área calculada a partir do mapa temático, aqui considerada como referência. A Tabela 4.2 apresenta os valores estimados (na média de dez repetições dos três tamanhos de amostra) para as diversas classes de uso do solo, em agosto de 2001 e janeiro de 2002.

TABELA 4.2 - Área ocupada com diferentes classes de uso do solo, em hectares, estimada com três tamanhos de amostra (média de dez repetições), no município de Ipuã - SP, em duas datas.

Uso	Agosto de 2001					
	Médias (ha)			CVs (%)		
	100	200	300	100	200	300
Cana-de-açúcar	15.316	14.707	15.253	14,35	10,45	10,18
Soja	0	0	0	0,00	0,00	0,00
Milho	94	117	187	223,00	141,30	111,68
Outras culturas	1.265	820	843	60,02	52,97	52,23
Vegetada	8.290	9.086	8.665	21,56	14,41	14,84
Não vegetada	515	890	765	94,84	50,81	54,87
Palha	8.384	8.243	7.838	21,42	15,30	15,77
Solo exposto	12.974	12.974	13.286	16,16	11,42	11,24
Total	46.838	46.837	46.837			
Janeiro de 2002						
Cana-de-açúcar	14.707	14.777	14.832	14,78	10,42	10,39
Soja	17.377	16.674	16.299	13,02	9,51	9,68
Milho	1.639	1.499	1.858	52,51	38,89	34,79
Outras culturas	187	70	140	157,95	182,77	129,14
Vegetada	8.571	8.993	8.446	21,13	14,51	15,08
Não vegetada	468	866	812	99,54	51,52	53,24
Palha	562	937	984	90,74	49,49	48,27
Solo exposto	3.326	3.021	3.466	36,17	26,93	25,01
Total	46.837	46.837	46.837			

Pode-se observar dos dados apresentados que para os tipos de uso que ocupam uma significativa fração da área do município, não há grande ganho com o aumento do tamanho da amostra. Já para os usos pouco frequentes, esse ganho é bastante perceptível.

Na Figura 4.1 está representado o comportamento das estimativas de área de soja, mostrando os valores obtidos em cada uma das 30 amostras, as médias das dez amostras com 100, 200 e 300 pontos, bem como os Limite Superior (LS) e Limite inferior (LI) do intervalo de confiança para essas médias com 95% de probabilidade, todos comparados à “verdade” obtida do mapeamento realizado e usado como referência.

Observa-se que, com o aumento do tamanho da amostra, há uma consistente convergência da média em direção ao valor verdadeiro e uma redução da amplitude de variação das estimativas. Em todos os casos observados as estimativas sempre estiveram dentro do intervalo de confiança de 95%.

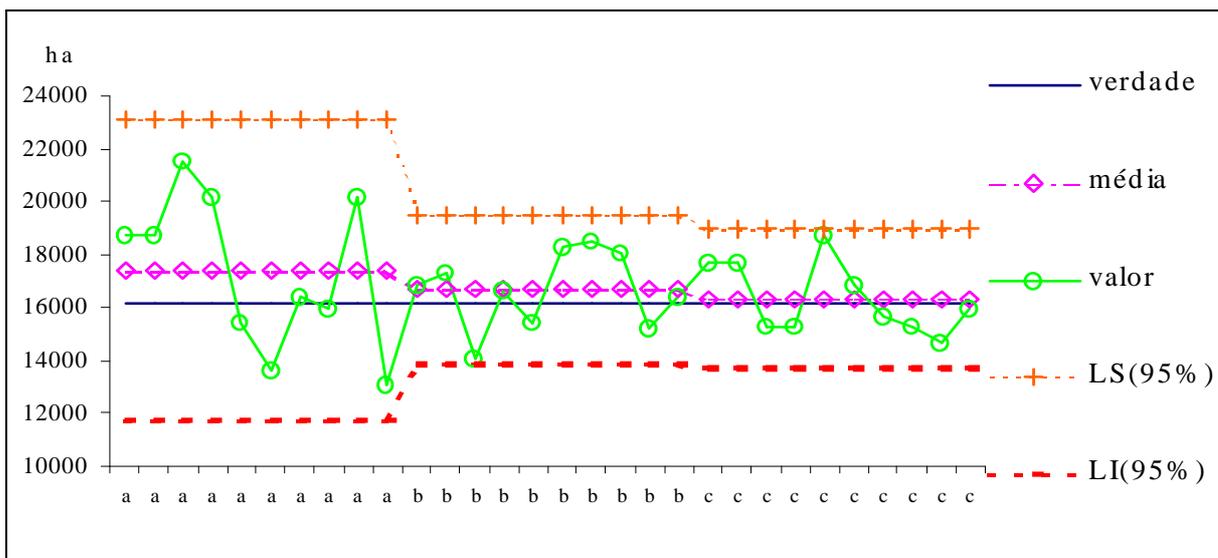


FIGURA 4.1 - Área estimada para a cultura da soja, em amostras de 100 (a), 200 (b) e 300 (c) pontos, distribuídos aleatoriamente no município de Ipuã, SP, em janeiro de 2002.

Para poder comparar melhor os resultados da amostragem frente aos dados de referência, bem como analisá-los em relação aos dados oficiais, e como cada método ou fonte abrange um determinado intervalo de tempo, optou-se pelo seguinte procedimento. Conforme era o objetivo inicial do trabalho, só as três culturas consideradas mais importantes para o município foram consideradas. Tanto para os dados de referência como para os obtidos por amostragem, para cana-de-açúcar foi considerado o maior valor entre as datas de agosto de 2001 e janeiro de 2002. Para o milho, nesses dois casos e foram somados os valores das duas datas; no caso dos dados fornecidos pelo EDR, os valores eram discriminados em milho safrinha e milho primeira safra, e também foram somados. Os dados do IBGE/PAM apresentam um único valor anual para cada cultura e assim foram considerados. Dessa maneira, foi elaborada a Tabela 4.3, que permite uma comparação mais direta das diversas

estimativas. Apenas os resultados das amostras com 100 pontos por município foram escolhidas. Os dados da amostragem refletem a média de dez amostras.

TABELA 4.3 - Comparação entre as estimativas oficiais, as obtidas por amostragem e a referência, para as principais culturas no município de Ipuã, na safra 2001/2002.

Uso	AAS 2001/2002	IBGE/PAM 2001	EDR nov/2001	Ref 2001/2002
Cana-de-açúcar	15.316	24.000	29.000	15.096
Soja	17.377	16.200	4.500	16.155
Milho	1.733	8.500	13.000	1.740

Dados de novembro de 2001 fornecidos pelo EDR de Orlândia.

Como pode se perceber, enquanto que os dados obtidos por amostragem são muito semelhantes aos obtidos pelo mapeamento completo do município, os dados provenientes de levantamentos subjetivos e não probabilísticos ficaram muito distantes do que foi considerado como referência. Além disso, mesmo que não se façam as dez repetições da amostragem, que aqui foram realizadas apenas para efeito de estudo de verificação do método, como os dados obtidos pela amostragem foram oriundos de amostragem probabilística, estão associados a eles um erro de estimativa. Assim, além de ter uma boa noção do valor correto, obtem-se também informação sobre a qualidade da previsão realizada. No caso do milho, por exemplo, além de saber que ele está sendo plantado em uma área reduzida no município, muito menor do que a estimada pelos métodos oficiais, sabe-se que o valor estimado é bem menos confiável, em termos de CV, que as estimativas para soja ou cana-de-açúcar.

4.3 Estimativas por Amostragem Aleatória Estratificada em Ipuã

O município de Ipuã é relativamente homogêneo quanto à sua ocupação agrícola, entretanto, para demonstrar a utilidade da estratificação espacial em âmbito municipal, ele foi dividido em dois estratos, um mais dedicado a soja e outro mais a cana-de-açúcar. Apenas para se verificar o efeito da mudança de objetivo do levantamento na precisão das estimativas, foram executadas duas amostragens. No primeiro caso, chamado de caso A, supondo que o interesse é pela estimativa da soja, colocou-se um maior número de pontos (114) no estrato I, onde a soja está mais presente, deixando os 86 restantes (as amostras sempre foram de tamanho 200, no total), para o estrato II, onde a cana-de-açúcar é predominante. No segundo caso, denominado B, foi feito

exatamente o contrário, supondo maior interesse pela precisão na estimativa da área da cana-de-açúcar, colocou-se 114 pontos no estrato II e apenas 86 no estrato I.

Para ambos os casos foram feitas as estimativas de área e CV para a cana-de-açúcar e para a soja, conforme apresentado na tabela 4.4, onde também aparece o valor de referência, extraído do mapa temático, de cada estrato e para cada cultura, para efeito de comparação.

TABELA 4.4 - Área plantada e CV de cana-de-açúcar e soja, em Ipuã, estimados por AAE.

	Estrato	nh	Cana-de-açúcar		Soja	
			Área (ha)	CV (%)	Área (ha)	CV (%)
Caso A	I	114	3.912	23,2	15.158	8,6
	II	86	9.483	10,8	3.749	21,7
			13.395	10,4	18.907	8,1
Caso B	I	86	4.861	23,5	11.991	12,4
	II	114	10.980	8,0	3.494	19,7
			15.841	9,1	15.485	10,6
Referência	I		4.677		12.991	
	II		10.034		3.164	
			14.711		16.155	

Como pode ser notado na Tabela 4.4, o CV para a estimativa de soja é mais baixo no caso A (8,1%), que foi intencionalmente favorável à essa cultura, enquanto que para a cana-de-açúcar o menor CV (9,1%) ocorreu no caso B, pelo mesmo motivo. Note-se também que embora o CV da estimativa para o município, para soja, tenha sido menor no caso A, dentro do estrato II houve até um aumento do CV, comparado ao caso B. É exatamente assim que age a estratificação, se o interesse é por uma determinada variável, o método vai procurar errar menos onde ela é mais frequente (ou ocupa maior área, nesse caso), mesmo que isso represente errar mais onde ela é rara. O importante é que na estimativa final há um ganho na precisão, desde que o objetivo tenha sido bem estabelecido e a estratificação bem feita.

4.4 Estimativa por Amostragem Aleatória Estratificada Simulada

Como resultado da análise de agrupamentos para formar quatorze estratos entre os municípios do Estado de São Paulo que plantam cana-de-açúcar, realizado com o uso do programa em linguagem SAS apresentado no Anexo B, foi produzida uma lista com os 383 nomes dos municípios, a área municipal, a área plantada com cana-de-açúcar, e um número indicativo do estrato ao qual ele foi atribuído. Dentro de cada estrato com mais de três municípios foram feitos os sorteios, usando números aleatórios, de três municípios para comporem a amostra estratificada. Um exemplo de amostra sorteada é apresentado na Tabela 4.5. Pode-se notar que três estratos continham apenas um município, o que resultou em uma amostra total para o Estado de 36 municípios, ou seja, onze estratos com três municípios cada e três estratos com apenas um.

Com base nesses dados foi simulada a aplicação da AAS por pontos em cada um dos 36 municípios. Dividindo a área plantada pela área do município obtiveram-se os valores da proporção p da cultura. Para simular a amostragem por pontos, simplesmente multiplicou-se esse valor pelo tamanho escolhido da amostra por município, que foi de 100 pontos. Os valores encontrados foram aproximados para o inteiro mais próximo e assumidos como o número de pontos aleatorizados no território do município que caíram em lavouras de cana-de-açúcar. A partir desse número de pontos, todos os cálculos já descritos puderam ser efetuados. Na tabela 4.6 aparecem os valores do número de pontos de cana-de-açúcar por município, assim como a estimativa de área plantada gerada a partir deles, bem como a variância e o CV para cada um.

Continuando os cálculos, foram produzidas as estimativas de área por estrato e para todo o Estado, com os respectivos valores de variância e CV, conforme apresentado na tabela 4.7.

A primeira coisa a notar nestas tabelas, a começar pela Tabela 4.5, é a grande variação que ocorre entre os valores, tanto de área plantada quanto de área municipal entre os estratos. Mas, pode-se observar também, que, com relação à área plantada, a variação dentro dos estratos é bem menor, pois essa foi a variável usada para definir a estratificação.

TABELA 4.5 - Amostra estratificada de municípios paulistas, com área plantada e área municipal em hectares e proporção p da área ocupada com a cana-de-açúcar, dados de 2001.

Município	Estrato	área plantada	área municipal	p
Macaubal	1	170	24.865	0,01
Pracinha	1	280	6.305	0,04
Taquaral	1	1.800	5.421	0,33
Santo Antônio de Posse	2	2.950	15.411	0,19
Echaporã	2	3.560	51.459	0,07
Getulina	2	3.676	67.543	0,05
Itirapina	3	5.447	56.426	0,10
Nova Europa	3	6.500	16.088	0,40
Altair	3	6.900	31.609	0,22
Ariranha	4	9.100	13.311	0,68
Santa Rita do Passa Quatro	4	9.590	75.299	0,13
Mineiros do Tietê	4	11.000	21.189	0,52
Rincão	5	13.000	31.342	0,41
Serra Azul	5	14.163	28.285	0,50
Itapira	5	15.000	51.750	0,29
Viradouro	6	15.500	21.904	0,71
Capivari	6	16.889	32.320	0,52
Boa Esperança do Sul	6	19.000	69.102	0,27
São Carlos	7	20.000	114.092	0,18
Olímpia	7	22.000	80.351	0,27
Penápolis	7	22.000	70.850	0,31
São Manuel	8	23.000	65.104	0,35
Valparaíso	8	23.000	85.876	0,27
Guataporá	8	23.200	41.264	0,56
Araras	9	29.000	64.346	0,45
Pederneiras	9	29.000	72.918	0,40
Dois Córregos	9	30.000	63.256	0,47
Pitangueiras	10	31.600	42.958	0,74
Araraquara	10	32.000	100.597	0,32
Ribeirão Preto	10	32.200	65.037	0,50
Sertãozinho	11	34.000	40.280	0,84
Jaú	12	37.000	68.834	0,54
Paraguaçu Paulista	13	39.000	100.109	0,39
Lençóis Paulista	13	40.000	80.386	0,50
Piracicaba	13	40.000	136.951	0,29
Morro Agudo	14	78.000	138.618	0,56

TABELA 4.6 - Dados de simulação do sorteio de uma AAS de tamanho 100 em cada município.

Município	Estrato	Pontos de cana	Área estimada	CV	Variância estimada
Macaubal	1	1	249	99,5	61.208
Pracinha	1	4	252	49,0	15.264
Taquaral	1	33	1.789	14,2	64.970
Santo Antônio de Posse	2	19	2.928	20,6	365.525
Echaporã	2	7	3.602	36,4	1.723.847
Getulina	2	5	3.377	43,6	2.166.964
Itirapina	3	10	5.643	30,0	2.865.514
Nova Europa	3	40	6.435	12,2	621.146
Altair	3	22	6.954	18,8	1.714.483
Ariranha	4	68	9.052	6,9	385.561
Santa Rita do Passa Quatro	4	13	9.789	25,9	6.412.753
Mineiros do Tietê	4	52	11.018	9,6	1.120.660
Rincão	5	41	12.850	12,0	2.376.265
Serra Azul	5	50	14.142	10,0	2.000.046
Itapira	5	29	15.008	15,6	5.514.216
Viradouro	6	71	15.552	6,4	987.914
Capivari	6	52	16.806	9,6	2.607.245
Boa Esperança do Sul	6	27	18.657	16,4	9.411.614
São Carlos	7	18	20.537	21,3	19.213.069
Olímpia	7	27	21.695	16,4	12.725.303
Penápolis	7	31	21.964	14,9	10.737.186
São Manuel	8	35	22.786	13,6	9.642.687
Valparaíso	8	27	23.186	16,4	14.535.407
Guataporá	8	56	23.108	8,9	4.195.435
Araras	9	45	28.956	11,1	10.247.414
Pederneiras	9	40	29.167	12,2	12.760.848
Dois Córregos	9	47	29.730	10,6	9.967.260
Pitangueiras	10	74	31.789	5,9	3.550.480
Araraquara	10	32	32.191	14,6	22.020.502
Ribeirão Preto	10	50	32.518	10,0	10.574.398
Sertãozinho	11	84	33.835	4,4	2.180.643
Jaú	12	54	37.170	9,2	11.769.386
Paraguaçu Paulista	13	39	39.043	12,5	23.842.081
Lençóis Paulista	13	50	40.193	10,0	16.154.772
Piracicaba	13	29	39.716	15,6	38.617.788
Morro Agudo	14	56	77.626	8,9	47.345.363

Na Tabela 4.6, são notáveis os valores de CV de cada município, sendo que alguns chegam a quase 100%, e que a grande maioria fica acima dos 10%. Isso é importante perceber para que se possa avaliar o benefício obtido pela estratificação quando se observam os valores do CV, agora na Tabela 4.7. Quando a estimativa dos municípios individuais são reunidas para formar a estimativa por estrato já há uma sensível redução nos valores de CV, que agora estão, na maioria, abaixo dos 10%.

TABELA 4.7 - Número total de municípios N_h e tamanho da amostra n_h , área com cana-de-açúcar (em hectares), variância, desvio padrão e CV (em %), por estrato.

Estrato	N_h	n_h	área estrato	variância	desvio	CV
1	159	3	121.354	397.311.046	19.933	16,43
2	67	3	221.265	2.122.965.718	46.076	20,82
3	53	3	336.225	1.623.334.785	40.291	11,98
4	33	3	328.448	958.195.771	30.955	9,42
5	19	3	266.001	396.720.034	19.918	7,49
6	24	3	408.127	832.433.461	28.852	7,07
7	6	3	128.390	170.702.232	13.065	10,18
8	7	3	161.188	154.478.107	12.429	7,71
9	5	3	146.422	91.598.673	9.571	6,54
10	3	3	96.498	36.145.381	6.012	6,23
11	1	1	33.835	2.180.643	1.477	4,36
12	1	1	37.170	11.769.386	3.431	9,23
13	4	3	158.602	139.759.363	11.822	7,45
14	1	1	77.626	47.345.363	6.881	8,86
total			2.521.151	6.984.939.963	83.576	3,31
Valor original IBGE			2.567.178			

Mas, o efeito mais impactante da estratificação se observa quando são calculados a estimativa de área plantada para o Estado e seu respectivo CV, esse último ficando em 3,31% apenas. É importante lembrar que esse é o valor esperado, mantendo-se sem muitas variações a situação retratada pelos dados do IBGE, e com a execução correta do levantamento de campo. Entretanto, é de se destacar que com apenas 36 municípios amostrados e visitando apenas 100 pontos em cada um, é possível alcançar uma estimativa tão precisa.

CAPÍTULO 5

CONCLUSÕES

O uso conjunto de imagens de sensoriamento remoto, sistemas de informação geográfica (SIGs) e de aparelhos de posicionamento global (GPS), associado a um esquema amostral probabilístico bem planejado, e não abrindo mão de dados obtidos a campo, conforme detalhado neste trabalho, permite estimar a área plantada com as principais culturas de um Estado ou do País, de maneira rápida, relativamente econômica, extremamente simples, e com confiabilidade predeterminada.

As imagens de sensoriamento remoto são um fator essencial no estabelecimento do método ora proposto, pois são elas que fornecem, pela própria natureza de sua aquisição, os elementos amostrais básicos, que são os pixels.

Além da utilidade das imagens para a definição do painel amostral e também na localização em campo, ela fornece em diversos casos, quando disponível nas datas adequadas, a possibilidade de se fazer a identificação da classe de uso do solo nos pontos escolhidos para comporem a amostra sem a necessidade de visita a campo, reduzindo sensivelmente o tempo e os custos de um levantamento em larga escala.

A estratificação é um poderoso instrumento no sentido de melhorar a eficiência dos levantamentos estatísticos por amostragem. Novamente, as imagens obtidas por satélites de sensoriamento remoto são grandes auxiliares na definição espacial dos estratos. Aos olhos treinados de técnicos com conhecimento da região de estudo, a interpretação visual de imagens é suficiente e adequada para a finalidade de estratificar uma região com base na predominância de certos usos agrícolas.

O planejamento amostral da forma como está sendo aqui proposto leva em conta e se beneficia de estimativas preexistentes com relação a área plantada por município, tanto para estabelecer estratos não espaciais como para definir, em função do nível de confiabilidade desejado, o melhor tamanho da amostra.

REFERÊNCIAS BIBLIOGRÁFICAS

- Anda, G.G. **Investigación Agrícola y seguridad alimentaria en un mundo globalizado**. Brasília: EMBRAPA 2003. Palestra proferida no Ciclo de Palestras "Alimentação, Agricultura Sustentável e a Importância da Pesquisa para o Brasil", realizado na Sede da Embrapa, em Brasília, DF em 22 de abril de 2003. Disponível em: <<http://www.embrapa.br/acs/palestralimentafao.htm>>. Acesso em: 20 maio 2003.
- Anderson, J.E.; Fischer, R.L.; Deloach, S.R. Remote sensing and precision agriculture: ready for harvest or still maturing? **Photogrammetric Engineering and Remote Sensing**, v.65, n.10, p.1118-1123, 1999.
- Assad, E.D.; Sano, E.E. (eds.). **Sistema de Informações Geográficas: aplicações na Agricultura**. 2. ed. Brasília: Embrapa SPI, 1998. 434p.
- Assunção, G.V.; Duarte, V. **Avaliação de áreas preparadas para plantio (solonú), utilizando-se dados digitalizados do Landsat, através de tratamento automático no I-100**. São José dos Campos: INPE, 1980. 28p. (INPE-1910-RPE/243).
- Batista, G.T.; Mendonça, F.J.; Lee, D.C.L.; Tardin, A.T.; Chen, S.C.; Novaes, R.A. **Uso de dados orbitais para identificação de áreas de cana-de-açúcar**. São José dos Campos: INPE, 1978a. 29p. (INPE-1228-NTE/116).
- Batista, G.T.; Mendonça, F.J.; Lee, D.C.L.; Tardin, A.T.; Chen, S.C.; Novaes, R.A. **Uso de sensores remotos a bordo de satélite e aeronave na identificação e avaliação de áreas de culturas para fins de previsão de safras**. São José dos Campos: INPE, 1978b. 41p. (INPE-1229-NTI/103).
- Binaghi, E.; Brivio, P.A.; Ghezzi, P.; Rampini, A.; Zilioli, E. A hybrid approach to fuzzy land cover mapping. **Pattern Recognition Letters**, v.17, p.1399-1410, 1996.
- Bouma, J.; Varallyay, G.; Batjes, N.H. Principal land use changes anticipated in Europe. **Agriculture Ecosystems & Environment**, v.67, p.103-119, 1998.
- Carvalho, M.A. Comércio agropecuário brasileiro no MERCOSUL. **Informações Econômicas**, v.29, n.6, p.7-22, 1999.
- Chavez Jr., P.S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. **Remote Sensing of Environment**, v.24, p.459-479, 1988.
- Chen, S.C. **Contribuição de dados de satélite no sistema de previsão de safras**. São José dos Campos: INPE, 1990. 7p. (INPE-5087-PRE/1595).
- Cochran, W.G. **Sampling techniques**. 3.ed. New York: John Wiley & Sons, 1977. 428p.

Collares, J.E.R.; Lauria, C.A.; Carrilho, M.M. Pesquisa de previsão e acompanhamento de safras baseada em painéis de amostras de áreas. In: Simpósio Brasileiro de Sensoriamento Remoto, 7., 1993, Curitiba. **Anais...** São José dos Campos: INPE, v.4, p.450-453, 1993.

Costa, S.R.X.; Paião, L.B.F.C.; Mendonça, F.J.; Shimabukuro, Y.E.; Duarte, V. **Estudos estatísticos preliminares sobre a região canavieira de Campos - RJ, utilizando imagens Landsat e fotografias aéreas.** São José dos Campos: INPE, 1983. 15p. (INPE-2723-NTE/200).

Epiphanyo, J.C.N.; Barros Neto, O.O.; Luiz, A.J.B.; Formaggio, A.R. Sistema de amostragem em imagem como base para estimativa de áreas agrícolas no município de Ipuã-SP. In: Simpósio Brasileiro de Sensoriamento Remoto, 10., 2001, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2001. 1 CD-ROM.

Epiphanyo, J.C.N.; Luiz, A.J.B.; Formaggio, A.R. Estimativa de áreas agrícolas municipais, utilizando sistema de amostragem simples sobre imagens de satélite. **Bragantia**, v.61, n.2, p.187-197, 2002.

Epiphanyo, J.C.N. Monitoreo ambiental y sistemas de información para la evaluación de la relación agricultura-medio ambiente. In: Viglizzo, E. (coord) **Libro verde: elementos para una política agroambiental en el Cono Sur.** Montevideo: IICA – PROCISUR, 1997. Cap.3, p.35-51.

Fjellstad, W.J.; Dramstad, W.E. Pattern of change in two contrasting Norwegian agricultural landscapes. **Landscape and Urban Planning**, v.45, p.177-191, 1999.

Food and Agriculture Organization of the United Nations (FAO). **Multiple frame agricultural surveys: volume 1 current survey based on area and list sampling methods.** Rome: FAO, 1996. 119p. (FAO Statistical Development Series, 7)

_____. **Multiple frame agricultural surveys: volume 2 agricultural survey programmes based on area frame or dual frame (area and list) sample design.** Rome: FAO, 1998. 242p. (FAO Statistical Development Series, 10).

Foody, G.M. The continuum of classification fuzziness in thematic mapping. **Photogrammetric Engineering and Remote Sensing**, v.65, n.4, p.443-451, 1999.

Gao, J. Integration of GPS with remote sensing and GIS: reality and prospect. **Photogrammetric Engineering and Remote Sensing**, v.68, n.5, p.447-453, 2002.

Garagorry, F.L.; Simon, M.F. Métodos computacionais aplicados ao estudo da dinâmica da agricultura. In: Congresso Brasileiro da Sociedade Brasileira de Informática Aplicada à Agropecuária e Agroindústria (SBI-Agro), 4., 2003, Porto Seguro. **Anais...** Lavras: UFLA, 2003. p.106-109.

Giannotti, M.A. **Geotecnologias na análise de impactos sócio-ambientais: o caso da queima de cana-de-açúcar na região de Piracicaba.** São José dos Campos. 147p. (INPE-9821-TDI/864). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, 2001.

Green, K. Spatial imagery and GIS: integrated data for natural resource management. **Journal of Forestry**, v.90, n.11, p. 32-36, 1992.

Griffith, D.A.; Amrhein, C.G. **Multivariate statistical analysis for geographers.** Upper Saddle River: Prentice Hall, 1997. 345p.

Gurgel, H. C.; Ferreira, N. J.; Luiz, A. J. B. Análise da variabilidade espacial e temporal do NDVI sobre o Brasil. In: Simpósio Brasileiro de Sensoriamento Remoto, 10., 2001, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2001. 1 CD-ROM.

Gürtler, S. **Estimativa de área agrícola a partir de sensoriamento remoto e banco de dados de pixels amostrais.** São José dos Campos. 179p. (INPE-9774-TDI/858). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, 2003.

Gürtler, S.; Luiz, A.J.B.; Epiphanyo, J.C.N. Uniformização de imagens Landsat para previsão de safras agrícolas. In: Simpósio Brasileiro de Sensoriamento Remoto, 11., 2003, Belo Horizonte. **Anais...** São José dos Campos: INPE, 2003. p.109-116. 1 CD-ROM.

Hair Jr., J.F.; Anderson, R.E.; Tatham, R.L.; Black, W.C. **Multivariate data analysis.** 5ed. Upper Saddle River: Prentice Hall, 1998. 762p.

Hietala-Koivu, R. Agricultural landscape change: a case study in Yläne, southwest Finland. **Landscape and Urban Planning**, v.46, p.103-108, 1999.

Instituto Brasileiro de Geografia e Estatística (IBGE). **CIDADES.** Disponível em: <<http://www.ibge.gov.br>>. Acesso: em 20 mai. 2003a.

_____. **Ipuã:** Folha SF-22-X-B-III-4. (Rio de Janeiro), 1972. Carta do Brasil – Escala 1:50.000.

_____. **SIDRA.** Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 22 abr. 2003b.

Ippoliti-Ramilo, G.A.; Epiphanyo, J.C.N.; Shimabukuro, Y.E.; Formaggio, A.R. Sensoriamento remoto orbital como meio auxiliar na previsão de safras. **Agricultura em São Paulo**, v.46, n.1, p.89-101, 1999.

Ippoliti-Ramilo, G.A.; Epiphanyo, J.C.N.; Shimabukuro, Y.E. Landsat-5 thematic mapper data for pré-planting crop área evaluation in tropical countries. **International Journal of Remote Sensing**, v.24, n.7, p.1521-1534, 2003.

Ippoliti-Ramilo, G. A. **Imagens TM/Landsat-5 da época de pré-plantio para a previsão da área de culturas de verão**. São José dos Campos. 182 p. (INPE-7116-TDI/668). Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, 1998.

Jensen, J.R. **Remote sensing of the environment: an earth resource perspective**. Upper Saddle River: Prentice Hall, 2000. 544p.

Johnson, N.L.; Kotz, S. **Discrete distributions**. New York: John Wiley & Sons, 1969. 328p.

King, R.B. Land cover mapping principles: a return to interpretation fundamentals. **International Journal of Remote Sensing**, v.23, n.18, p. 3525-3545, 2002.

Klersy, R. The work and role of the Commission of the European Communities. **International Journal of Remote Sensing**, v.13, n.6-7, p.1035-1058, 1992.

Koffler, N.F. **Elementos ambientais de planejamento territorial: apontamentos de aula**. Bauru: UNESP, 1996. 101p.

Krug, T.; Yanasse, C.C.F. **Estimativa de safras agrícolas utilizando dados coletados por satélites de sensoriamento remoto e dados terrestres, através de amostras de substratos geográficos**. São José dos Campos: INPE, 1986. 51p. (INPE-4102-RPE/534).

Lillesand, T.M.; Kiefer, R.W. **Remote sensing and image interpretation**. 3.ed. New York: Wiley, 1994. 750p.

Luiz, A.J.B.; Epiphanyo, J.C.N. Amostragem por pontos em imagens de sensoriamento remoto para estimativa de área plantada por município. In: Simpósio Brasileiro de Sensoriamento Remoto, 10., 2001, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2001. 1 CD-ROM.

Luiz, A.J.B.; Gürtler, S. Aleatorização de pontos no território de um município, usando o SPRING, para a estimativa de área agrícola por amostragem In: Congresso Brasileiro da Sociedade Brasileira de Informática Aplicada à Agropecuária e Agroindústria (SBI-Agro), 4., 2003, Porto Seguro. **Anais...** Lavras: UFLa, 2003. p. 9–12.

Luiz, A.J.B.; Gürtler, S.; Gleriani, J.M.; Epiphanyo, J.C.N.; Campos, R.C. Reflectância a partir do número digital de imagens ETM+. In: Simpósio Brasileiro de Sensoriamento Remoto, 11., 2003, Belo Horizonte. **Anais...** São José dos Campos: INPE, 2003. p.2071-2078. 1 CD-ROM.

Luiz, A.J.B.; Oliveira, J. C.; Epiphanyo, J.C.N.; Formaggio, A. R. Auxílio das imagens de satélite aos levantamentos por amostragem em agricultura. **Agricultura em São Paulo**, v.49, n.1, p.41-54, 2002.

- Luiz, A.J.B. Pesquisa espacial: questão de soberania. **Gazeta Mercantil Planalto Paulista**, ano 2, n.379, p.2, 2000. (01/06/2000).
- Luiz, A.J.B. **Sensoriamento remoto agrícola**. São José dos Campos: INPE, 2002. 98p. (INPE-8858-PUD/57).
- Luiz, A.J.B.; Silveira, M.A. Diagnóstico rápido e dialogado em estudos de desenvolvimento rural sustentável. **Pesquisa Agropecuária Brasileira**, v.35, n.1, p.83-91, 2000.
- MacDonald, R.B.; Hall, F.G. Global crop forecasting. **Science**, v.208, p.670-679, may. 1980.
- Medeiros, A.M.P.; Rudorff, B.F.T.; Shimabukuro, Y.E. Imagens Landsat na estimativa de áreas de cana-de-açúcar, milho e soja. In: Simpósio Brasileiro de Sensoriamento Remoto, 8., 1996, Salvador. **Anais...** São José dos Campos: INPE, 1996. 1 CD-ROM.
- Mendonça, F.J.; Cottrell, D.A.; Tardin, A.T.; Lee, D.C.L.; Shimabukuro, Y.E.; Moreira, M.A.; Lima, A.M.; Maia, F.C.S. **Cultura do trigo: identificação e avaliação de áreas através de dados do Landsat**. São José dos Campos: INPE, 1981. 27p. (INPE-2204-RPE/397).
- Mendonça, F.J.; Lee, D.C.L.; Tardin, A.T.; Chen, S.C.; Novaes, R.A.; Shimabukuro, Y.E. **Resultados significantes do projeto Estatísticas Agrícolas: 1975-1978**. São José dos Campos: INPE, 1979. 29p. (INPE-1609-NTE/155).
- Mendonça, F.J. **Sensoriamento Remoto aplicado à agricultura: princípios básicos, metodologia e aplicações**. São José dos Campos: INPE, 1981. 87p. (INPE-2225-MD/012).
- Meyer-Roux, J.; King, C. Agriculture and Forestry. **International Journal of Remote Sensing**, v.13, n.6-7, p.1329-1341, 1992.
- Monitoring Agriculture with Remote Sensing Techniques (MARS). **Crop area estimation of annual crops through area frame sampling based on segments: results obtained in Europe in 1992**. Varese: Ispra, 1993. 21p.
- Mulders, M.A.; De Bruin, S.; Schuiling, B.P. Structured approach to land cover mapping of the Atlantic zone of Costa Rica using single date TM data. **International Journal of Remote Sensing**, v.13, n.16, p.3017-3033, 1992.
- Neter, J.; Kutner, M.H.; Nachtsheim, C.J.; Wasserman, W. **Applied linear statistical models**. Boston: McGraw-Hill, 1996. 1408p.
- Oliveira, J. B.; Camargo, M. N.; Rossi, M.; Calderano Filho, B. **Mapa pedológico do Estado de São Paulo**. Rio de Janeiro: Embrapa Solos, 1999. Escala 1:500.000.

- Ortiz, M.J.; Formaggio, A.R.; Epiphonio, J.C.N. Classification of croplands through integration of remote sensing, GIS, and historical database. **International Journal of Remote Sensing**, v.18, n.1, p.95-105, 1997.
- Pino, F.A. Área cultivada e ICMS. **Informações Econômicas**, v.29, n.3, p.7-14, 1999a.
- Pino, F.A. Estatísticas agrícolas para o século XXI. **Agricultura em São Paulo**, v.46, n.2, p.71-105, 1999b.
- Pino, F.A. et al. (orgs.) **Levantamento Censitário de Unidades de Produção Agrícola do Estado de São Paulo**. São Paulo: IEA, CATI, SAA, 1997. 4v.
- Pradhan, S. Crop area estimation using GIS, remote sensing and area frame sampling. **JAG: International Journal of Applied Earth Observation and Geoinformation**, v.3, n.1, p.86-92, 2001.
- Previsioni da satellite per alcune importanti colture. **Notiziario Congiunturale Dell'Agricoltura**. Roma, v.9, n.3-4, p. 14, 2000.
- Ray, S.S.; Pokharna, S.S.; Ajai. Cotton yield estimation using agrometeorological model and satellite-derived spectral profile. **International Journal of Remote Sensing**, v.20, n.14, p.2693-2702, 1999.
- Ryerson, R.A.; Curran, P.J.; Stephens, P.R. Agriculture. In: Philipson, W.R. (ed.). **Manual of Photographic Interpretation**. 2ed. Bethesda: ASPRS, 1997. p.285-397.
- Sano, E.E.; Assad, E.D.; Orioli, A.L. Monitoramento da ocupação agrícola. In: Assad, E.D.; Sano, E.E. (eds.) **Sistema de Informações Geográficas: Aplicações na Agricultura**. 2.ed. Brasília: Embrapa SPI, 1998. p.179-190.
- SAS Institute. **SAS procedure guide**, release 6.03. Cary: SAS, 1988. 441p.
- Smits, P.C.; Dellepiane, S.G.; Showengerdt, R.A. Quality assessment of image classification algorithms for land-cover mapping: a review and proposal for a cost-based approach. **International Journal of Remote Sensing**, v.20, n.8, p.1461-1486, 1999.
- Souza, G.S. **Introdução aos modelos de regressão linear e não linear**. Brasília: Embrapa-SPI, 1998. 489p.
- Tardin, A. T.; Palestino, C. V. B.; Sonnenburg, C. R. **Levantamento de áreas ocupadas por café e trigo no noroeste do Paraná por análise automática de dados do satélite Landsat**. São José dos Campos: INPE, 1976. 18p. (INPE-875-NTE/065).
- Terres, J.M.; Delince, J.; Van de Steene, M.; Hawkins, A. The use of remote sensing and GIS capabilities to support the reform of the common agricultural policy of the European Community. **Remote Sensing Reviews**, v.12, p.53-60, 1995.

Thenkabail, P.S.; Ward, A.D.; Lyon, J.G.; Merry, C.J. Thematic mapper vegetation indices for determining soybean and corn growth parameters. **Photogrammetric Engineering and Remote Sensing**, v.60, n.4, p. 437-442, 1994.

Tsiligirides, T.A. Remote sensing as a tool for agricultural statistics: a case study of area frame sampling methodology in Hellas. **Computer and Electronics in Agriculture**, v.20, n.1, p.45-77, 1998.

Tsunechiro, A.; Freitas, B. B. de. Os cinquenta municípios brasileiros maiores produtores de milho e soja. **Informações Econômicas**, v. 31, n. 7, p. 53-58, jul. 2001.

Xiao, X.; Boles, S.; Froking, S.; Salas, W.; Moore, B.; Li, C.; He, L.; Zhao, R. Landscape-scale characterization of cropland in China using VEGETATION sensor data and Landsat TM imagery. **International Journal of Remote Sensing**, v.23, n.18, p.3579-3594, 2002.

Wilkinson, G. G. A review of current issues in the integration of GIS and remote sensing data. **International Journal of Geographical Information Systems**, v.10, n.1, 85-101, 1996.

Zwillinger, D.; Kokoska, S. **CRC Standard probability and statistics tables and formulae**. Boca Raton: Chapman & Hall/CRC, 2000. 554p.

APÊNDICE A

DADOS DA PRODUÇÃO AGRÍCOLA EM IPUÃ DE 1990 A 2001, CONFORME O
IBGE/PAM

TABELA A.1 - Quantidade produzida, em toneladas, de todas as culturas registradas pelo IBGE/PAM, para o município de Ipuã – SP, entre 1990 e 2001.

Ano	Culturas												Total	Cana + Milho + Soja	%
	Algodão	Amen-doim	Arroz	Batata	Cana-de-açúcar	Café	Feijão	Man-dioca	Milho	Soja	Sorgo	Tomate			
1990	226	0	60	600	200.400	59	630	0	63.000	25.000	540	1.400	291.915	288.400	98,8
1991	22	0	105	532	200.400	25	333	0	100.000	35.640	48	17	337.122	336.040	99,7
1992	59	0	105	0	260.400	0	683	0	97.038	35.640	3.312	0	397.237	393.078	99,0
1993	0	0	210	0	483.600	0	238	0	66.000	45.360	2.400	0	597.808	594.960	99,5
1994	0	0	210	450	957.040	0	495	0	50.360	60.000	4.800	0	1.073.355	1.067.400	99,4
1995	0	0	81	200	1.700.000	0	1.188	88	44.959	40.888	792	0	1.788.196	1.785.847	99,9
1996	0	0	99	0	1.627.920	0	0	0	47.600	60.000	3.000	0	1.738.619	1.735.520	99,8
1997	0	0	99	0	1.600.000	0	599	0	55.590	78.705	0	0	1.734.993	1.734.295	100,0
1998	461	63	90	225	1.600.000	0	0	0	47.400	35.640	693	0	1.684.572	1.683.040	99,9
1999	527	0	756	0	1.480.000	0	396	0	55.800	57.500	4.350	0	1.599.329	1.593.300	99,6
2000	527	760	60	0	1.480.000	0	257	0	20.400	46.800	1.851	0	1.550.655	1.547.200	99,8
2001	527	0	54	0	1.920.000	0	388	0	42.000	43.700	1.400	0	2.008.069	2.005.700	99,9

TABELA A.2 - Área colhida, em hectares, de todas as culturas registradas pelo IBGE/PAM, para o município de Ipuã – SP, entre os anos de 1990 e 2001.

Ano	Culturas												Total	Cana + Milho + Soja	%
	Algo- dão	Amen- doim	Arroz	Batata	Cana-de- açúcar	Café	Feijão	Man- dioca	Milho	Soja	Sorgo	Tomate			
1990	184	0	50	20	3.340	66	900	0	25.200	18.940	300	70	49.070	47.480	96,8
1991	15	0	50	24	4.340	66	213	0	45.000	18.000	24	1	67.733	67.340	99,4
1992	24	0	50	0	4.340	0	239	0	42.464	18.000	1.673	0	66.790	64.804	97,0
1993	0	0	100	0	6.045	0	120	0	25.000	18.000	1.000	0	50.265	49.045	97,6
1994	0	0	100	15	11.963	0	250	0	22.889	28.000	1.500	0	64.717	62.852	97,1
1995	0	0	41	10	20.250	0	575	5	17.068	18.006	330	0	56.285	55.324	98,3
1996	0	0	50	0	20.349	0	0	0	21.800	30.000	1.000	0	73.199	72.149	98,6
1997	0	0	50	0	20.000	0	333	0	22.500	31.000	0	0	73.883	73.500	99,5
1998	185	25	50	9	20.000	0	0	0	18.500	18.000	294	0	57.063	56.500	99,0
1999	213	0	258	0	18.500	0	200	0	22.000	20.000	2.900	0	64.071	60.500	94,4
2000	210	348	30	0	14.500	0	170	0	18.500	19.500	1.517	0	54.775	52.500	95,8
2001	213	0	30	0	24.000	0	196	0	8.500	16.200	1.000	0	50.139	48.700	97,1

TABELA A.3 - Valor não corrigido da produção, em moeda da época*, de todas as culturas registradas pelo IBGE/PAM, para o município de Ipuã – SP, entre os anos de 1990 e 2001.

Ano	Culturas												Total	Cana + Milho + Soja	%
	Algodão	Amen-doim	Arroz	Batata	Cana-de-açúcar	Café	Feijão	Man-dioca	Milho	Soja	Sorgo	Toma-te			
1990	6.428	0	1.560	30.498	276.953	3.304	34.335	0	614.250	333.250	5.400	32.200	1.338.178	1.224.453	91,5
1991	4.461	0	11.954	53.200	1.761.866	9.691	88.911	0	6.020.000	3.451.378	2.311	1.584	11.405.356	11.233.244	98,5
1992	149.270	0	139.650	0	19.530.000	0	2.936.900	0	84.093.131	48.470.400	2.599.920	0	157.919.271	152.093.531	96,3
1993	0	0	1.468	0	646.090	0	7.559	0	699.600	680.400	20.400	0	2.055.517	2.026.090	98,6
1994	0	0	45	207	8.613	0	249	0	6.043	10.620	408	0	26.185	25.276	96,5
1995	0	0	15	98	18.700	0	379	18	3.866	5.315	71	0	28.462	27.881	98,0
1996	0	0	21	0	22.791	0	0	0	6.188	15.600	264	0	44.864	44.579	99,4
1997	0	0	22	0	25.600	0	289	0	6.115	21.250	0	0	53.276	52.965	99,4
1998	189	25	29	68	24.000	0	0	0	5.593	7.306	55	0	37.265	36.899	99,0
1999	290	0	287	0	16.280	0	211	0	10.044	15.525	827	0	43.464	41.849	96,3
2000	337	480	16	0	26.403	0	171	0	4.386	13.478	263	0	45.534	44.267	97,2
2001	324	0	19	0	47.482	0	353	0	6.672	15.494	156	0	70.500	69.648	98,8

* como os valores não são corrigidos, não é válida a comparação entre anos, principalmente antes de 1994 devido aos altos valores da inflação anual e pela mudança de moeda; os valores, entretanto, são comparáveis entre as culturas dentro do ano, assim como são comparáveis entre os anos os percentuais que representam o peso relativo da soma das três principais cultura: cana-de-açúcar, milho e soja.

APÊNDICE B

PROGRAMA EM LINGUAGEM SAS PARA DIVIDIR UM CONJUNTO DE MUNICÍPIOS EM ESTRATOS USANDO A DISTÂNCIA EUCLIDIANA MÍNIMA

```
options ls=80 ps=800 nonumber nodate;
title "Análise de agrupamento para Cana em São Paulo";
data canasp;
input num munic $5-30 cana;
cards;
001 Pindamonhangaba      8
002 Cruzeiro             10
003 Itupeva              10
004 Jambeiro             20
.                          .
.                          .
.                          .
378 Jaú                  37000
379 Paraguaçu_Paulista  39000
380 Jaboticabal          40000
381 Lençóis_Paulista    40000
382 Piracicaba           40000
383 Morro_Agudo          78000
;
proc fastclus data=canasp maxc=6 maxiter=99 converge=0
    out=prelim6 cluster=g6 noprint;
    var cana;
    id munic;
run;
proc sort data=prelim6;
    by g6;
proc means mean data=prelim6 noprint;
    var cana;
    by g6;
    output out=estat6 mean=media var=vari n=num;
proc fastclus data=canasp maxc=7 maxiter=99 converge=0
    out=prelim7 cluster=g7 noprint;
    var cana;
    id munic;
run;
proc sort data=prelim7;
    by g7;
proc means mean data=prelim7 noprint;
    var cana;
    by g7;
    output out=estat7 mean=media var=vari n=num;
proc fastclus data=canasp maxc=8 maxiter=99 converge=0
    out=prelim8 cluster=g8 noprint;
    var cana;
    id munic;
```

```

run;
proc sort data=prelim8;
  by g8;
proc means mean data=prelim8 noprint;
  var cana;
  by g8;
  output out=estat8 mean=media var=vari n=num;
proc fastclus data=canasp maxc=9 maxiter=99 converge=0
  out=prelim9 cluster=g9 noprint;
  var cana;
  id munic;
run;
proc sort data=prelim9;
  by g9;
proc means mean data=prelim9 noprint;
  var cana;
  by g9;
  output out=estat9 mean=media var=vari n=num;
proc fastclus data=canasp maxc=10 maxiter=99 converge=0
  out=prelim10 cluster=g10 noprint;
  var cana;
  id munic;
run;
proc sort data=prelim10;
  by g10;
proc means mean data=prelim10 noprint;
  var cana;
  by g10;
  output out=estat10 mean=media var=vari n=num;
proc fastclus data=canasp maxc=11 maxiter=99 converge=0
  out=prelim11 cluster=g11 noprint;
  var cana;
  id munic;
run;
proc sort data=prelim11;
  by g11;
proc means mean data=prelim11 noprint;
  var cana;
  by g11;
  output out=estat11 mean=media var=vari n=num;
proc fastclus data=canasp maxc=12 maxiter=99 converge=0
  out=prelim12 cluster=g12 noprint;
  var cana;
  id munic;
run;
proc sort data=prelim12;
  by g12;
proc means mean data=prelim12 noprint;
  var cana;
  by g12;
  output out=estat12 mean=media var=vari n=num;
proc fastclus data=canasp maxc=13 maxiter=99 converge=0
  out=prelim13 cluster=g13 noprint;
  var cana;
  id munic;

```

```

run;
proc sort data=prelim13;
  by g13;
proc means mean data=prelim13 noprint;
  var cana;
  by g13;
  output out=estat13 mean=media var=vari n=num;
proc fastclus data=canasp maxc=14 maxiter=99 converge=0
  out=prelim14 cluster=g14 noprint;
  var cana;
  id munic;
run;
proc sort data=prelim14;
  by g14;
proc means mean data=prelim14 noprint;
  var cana;
  by g14;
  output out=estat14 mean=media var=vari n=num;
data estat6;
  set estat6;
  NumGrup=6;
  grupo=g6;
data estat7;
  set estat7;
  NumGrup=7;
  grupo=g7;
data estat8;
  set estat8;
  NumGrup=8;
  grupo=g8;
data estat9;
  set estat9;
  NumGrup=9;
  grupo=g9;
data estat10;
  set estat10;
  NumGrup=10;
  grupo=g10;
data estat11;
  set estat11;
  NumGrup=11;
  grupo=g11;
data estat12;
  set estat12;
  NumGrup=12;
  grupo=g12;
data estat13;
  set estat13;
  NumGrup=13;
  grupo=g13;
data estat14;
  set estat14;
  NumGrup=14;
  grupo=g14;
data estat;

```

```

merge estat6 estat7 estat8 estat9 estat10 estat11 estat12 estat13 estat14;
by NumGrup;
proc sort data=prelim6;
  by munic;
proc sort data=prelim7;
  by munic;
proc sort data=prelim8;
  by munic;
proc sort data=prelim9;
  by munic;
proc sort data=prelim10;
  by munic;
proc sort data=prelim11;
  by munic;
proc sort data=prelim12;
  by munic;
proc sort data=prelim13;
  by munic;
proc sort data=prelim14;
  by munic;
data grupos;
  merge prelim6 prelim7 prelim8 prelim9 prelim10 prelim11 prelim12 prelim13 prelim14;
  by munic;
proc print data=estat noobs;
  var grupo media vari num;
  by NumGrup;
proc sort data=grupos;
  by cana;
proc print data=grupos noobs;
  var munic cana g6 g7 g8 g9 g10 g11 g12 g13 g14;
run;

```

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programa de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. São aceitos tanto programas fonte quanto executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.