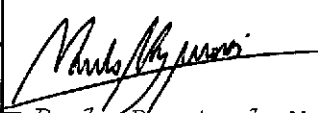
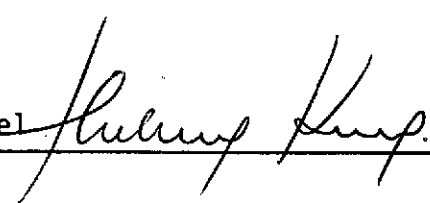
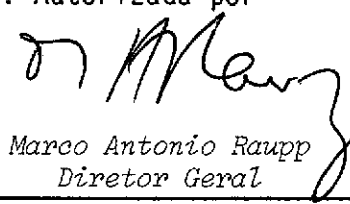


1. Publicação nº <i>INPE-4102-RPE/534</i>	2. Versão	3. Data <i>Jan. 1.986</i>	5. Distribuição <input type="checkbox"/> Interna <input checked="" type="checkbox"/> Externa <input type="checkbox"/> Restrita
4. Origem <i>DSR</i>	Programa <i>Agricultura</i>		
6. Palavras chaves - selecionadas pelo(s) autor(es) <i>SAFRAS AMOSTRAGEM ESTIMATIVA DE ÁREA</i>			
7. C.D.U.: <i>528.711.7:632.164</i>			
8. Título <i>INPE-4102-RPE/534</i>  <i>ESTIMATIVA DE SAFRAS AGRÍCOLAS UTILIZANDO DADOS COLETADOS POR SATÉLITES DE SENSORIAMENTO RE MOTO E DADOS TERRESTRES, ATRAVÉS DE AMOS TRAS DE SUBSTRATOS GEOGRÁFICOS</i>		10. Páginas: <i>51</i>	11. Última página: <i>42</i>
9. Autoria <i>Thelma Krug Corina da Costa Freitas Yanasse</i>		12. Revisada por  <i>Paulo Renato de Moraes</i>	
Assinatura responsável 		13. Autorizada por  <i>Marco Antonio Raupp Diretor Geral</i>	
14. Resumo/Notas  <i>Este trabalho apresenta uma metodologia estatística para esti mativa de áreas cultivadas com diversas culturas utilizando dados coletados pelo satélite LANDSAT e dados terrestres coletados por enumeradores de cam po. Subestratos geográficos construídos dentro dos distritos de análise são utilizados para melhorar a precisão das estimativas desejadas. São propos tos aqui estimadores por expansão direta, a serem utilizados sobre dados coletados somente em campo, e um estimador de regressão, que combina os re sultados da classificação dos dados LANDSAT com os dados terrestres sobre as áreas onde se encontram disponíveis estas duas fontes de informação.</i>			
15. Observações			

### ABSTRACT

*This paper presents a statistical methodology for estimating areas cultivated with several cultures using information gathered by the LANDSAT satellite, as well as ground data collected by trained enumerators. Paper strata constructed within analysis district are used to improve the precision of the desired estimates. Direct expansion estimators to be used over ground data and a regression estimator combining the results of LANDSAT classification with survey data on those areas where these two sources of data are available are proposed.*

## SUMÁRIO

	<u>Pág.</u>
LISTA DE FIGURAS .....	<i>v</i>
LISTA DE TABELAS .....	<i>vii</i>
<u>INTRODUÇÃO</u> .....	1
1 - Divisão da área de estudo em estratos de uso do solo .....	3
2 - Construção das unidades primárias de amostragem (UPAs) dentro de cada estrato de uso do solo .....	7
3 - Construção dos distritos de análise .....	10
4 - Atribuição de um determinado número de segmentos a cada unidade primária de amostragem .....	12
5 - Construção dos substratos geográficos .....	12
6 - Seleção de um determinado número de segmentos em cada substrato geográfico .....	18
7 - Construção dos segmentos nas unidades primárias de amostragem .....	21
8 - Coleta de dados nos segmentos amostrados .....	25
9 - Tratamento dos segmentos amostrados .....	26
10- Utilização do estimador adequado para a geração das estatísticas desejadas .....	27
<u>AVALIAÇÃO DO ESTIMADOR DE REGRESSÃO <math>\hat{Y}_C</math> (tot)</u> .....	39
<u>COMENTÁRIOS FINAIS</u> .....	39

## LISTA DE FIGURAS

	<u>Pág.</u>
1 - Diferenciação entre blocos de estrato e estratos.....	4
2 - Ciclo de repetição orbital do satélite TM-LANDSAT .....	11
3 - Distribuição espacial de culturas dentro de um estrato .....	13
4 - Numeração das UPAS nos municípios e pseudomunicípios .....	15
5 - Exemplo hipotético de ordenação de municípios .....	16
6 - Variação típica dentro de uma UPA .....	24

LISTA DE TABELAS

	<u>Pág.</u>
1 - Construção dos substratos geográficos .....	19
2 - Segmentos amostrados nos 13 substratos geográficos, os quais dão origem a 15 réplicas, ou 15 amostras interpene- trantes .....	20

## INTRODUÇÃO

A obtenção de estimadores de áreas agrícolas através do uso de informações orbitais não é uma abordagem nova para o problema; já vem sendo utilizada há vários anos, principalmente nos Estados Unidos.

Vários resultados das pesquisas já desenvolvidas foram levados em consideração neste trabalho, como, por exemplo, o fato de o estimador de regressão, que usa dados orbitais como valores de uma variável auxiliar, constituir, em geral, um estimador mais preciso do que os que não utilizam essa informação.

Um delineamento amostral que seleciona amostras de substratos geográficos tem sido utilizado nos E.U.A.; entretanto, nos trabalhos pesquisados as informações dos segmentos amostrados provêm somente de uma coleta de campo.

Este trabalho é original, pois apresenta pela primeira vez o conceito de substratos geográficos no contexto de utilização de dados obtidos por imagens de satélite como variável auxiliar no estimador de regressão.

A presente proposta, possível de ser utilizada para a geração de estimativas de safras agrícolas para uma dada área de estudo, considera uma amostragem probabilística de segmentos, que constituem parcelas de terra sobre as quais se coletam as informações pertinentes ao estudo desejado.

Para atender os requisitos necessários ao delineamento amostral a ser aplicado nesta metodologia (construção e seleção das unidades amostrais e procedimento de estimação), alguns procedimentos devem ser efetuados, os quais serão identificados e descritos mais detalhadamente a seguir.

- 1) divisão da área de estudo em estratos de uso do solo;
- 2) construção das unidades primárias de amostragem (UPAs) dentro de cada estrato de uso do solo;
- 3) construção dos distritos de análise;
- 4) atribuição de um determinado número de segmentos a cada unidade primária de amostragem;
- 5) construção dos subestratos geográficos;
- 6) seleção de um determinado número de segmentos em cada subestrato geográfico;
- 7) construção dos segmentos nas unidades primárias de amostragem;
- 8) coleta de dados nos segmentos amostrados;
- 9) tratamento dos segmentos amostrados;
- 10) utilização do estimador adequado para a geração das estimativas desejadas.

Convém ressaltar que o passo 9 pode ser realizado paralelamente à execução do passo 8.

Neste ponto, é importante salientar também que a construção do painel amostral (caracterização das unidades amostrais, que estará perfeitamente configurada após a execução do passo 5) deve ser feita de forma a garantir sua utilização em múltiplos estudos. Isto se deve ao fato de que este elemento do delineamento amostral é o que mais requer esforço e tempo; deve portanto, ser utilizado por um período de tempo considerável antes de ser refeito ou revisto.

A seguir apresenta-se uma descrição mais detalhada dos passos identificados acima, os quais caracterizam a metodologia a ser proposta neste estudo.

## 1 - DIVISÃO DA ÁREA DE ESTUDO EM ESTRATOS DE USO DO SOLO

A estratificação da região de estudo em estratos de uso do solo será realizada para garantir uma maior homogeneidade entre as unidades amostrais dentro de cada estrato específico, o que aumenta a precisão das estimativas a serem geradas. Para proceder à estratificação, são necessários mapas detalhados (topográficos e de uso do solo), imagens de satélite e fotografias aéreas. As imagens TM-LANDSAT serão utilizadas no processo de estratificação a fim de aumentar a eficiência da amostragem. Elas são muito úteis por oferecer uma visão sinótica de uma grande área em uma só foto. Caso a mesma área fosse coberta por fotografias aéreas, milhares delas seriam necessárias. Entretanto, é bom ter em mente que as imagens de satélite nem sempre são suficientes para garantir uma estratificação detalhada da área de estudo, porque não se podem garantir imagens sem cobertura de nuvens e a diferenciação visual das diversas categorias de uso do solo, principalmente aquelas que apresentam pequena separabilidade de albedos.

A área em estudo é então dividida em blocos de estrato, que são áreas contíguas de uso homogêneo da terra. Todos os blocos de estrato de mesma categoria de uso do solo constituem os chamados estratos. A Figura 1 mostra a diferenciação entre blocos de estrato e estratos de uso do solo.

Quando se pretende construir um painel amostral para uma dada área de estudo (um Estado, por exemplo), devem-se avaliar dados sobre as suas características agrícolas para determinar quais estratos de uso do solo devem ser definidos. Em geral, num projeto específico de estimativa de safras agrícolas, apresentam-se os seguintes tipos de estratos:

- a) área agrícola intensivamente cultivada;
- b) área agrícola esparsamente cultivada;
- c) área agrícola urbanizada;
- d) pastagens;



- e) áreas não-agrícolas;
- f) água.

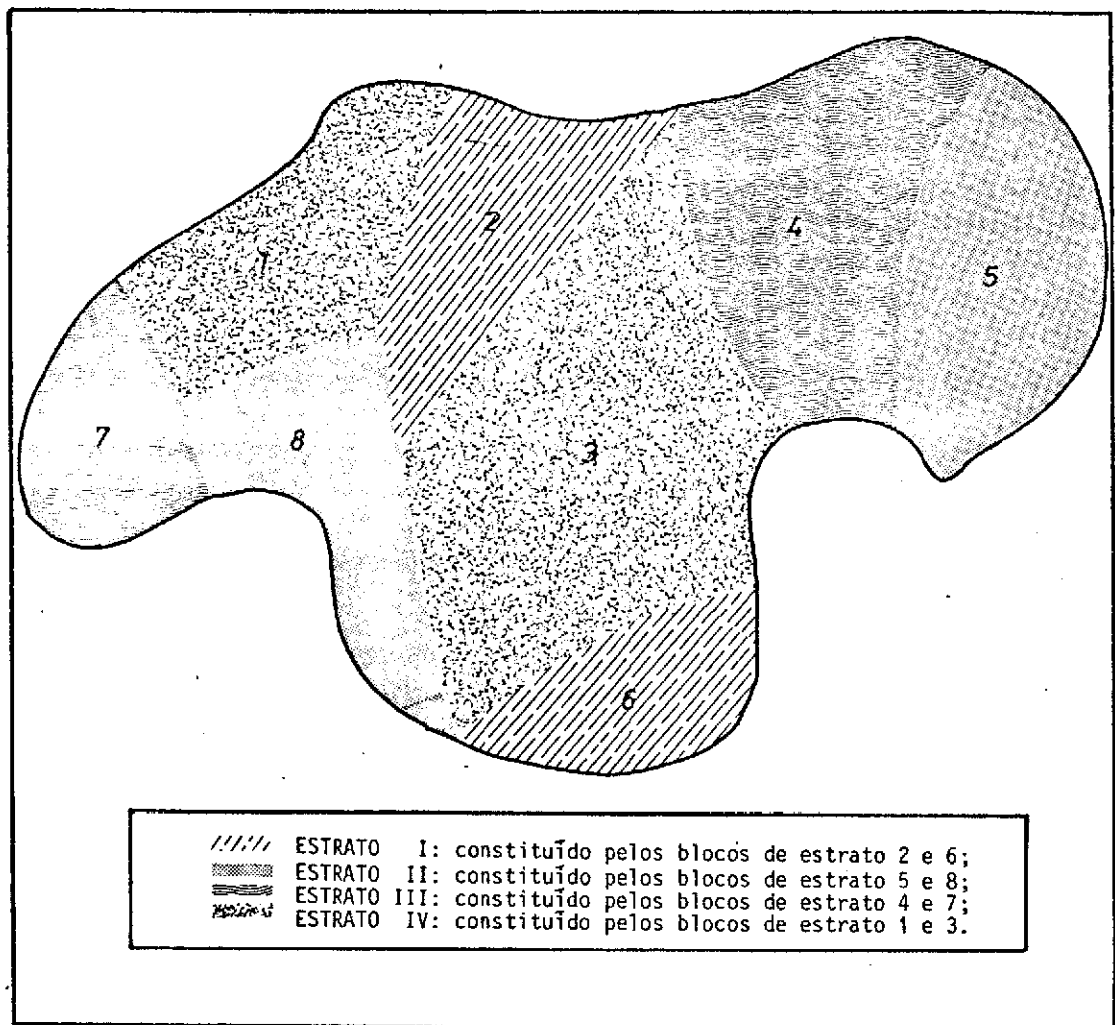


Fig. 1 - Diferenciação entre blocos de estrato e estratos.

Convém ressaltar que a definição dos estratos caracterizados depende, fundamentalmente, das particularidades de cada estudo que se pretende realizar. Assim, poder-se-á pensar em dividir o estrato de área agrícola intensivamente cultivada em dois estratos distintos e independentes, se isto for conveniente para a investigação que está sendo realizada. Desta forma, se o estrato de área agrícola intensivamente cultivada for definido como aquele que apresenta 80% ou mais da área cultivada, pode-se estudar a conveniência de dividi-lo em dois estratos específicos, a saber:

- 1) estrato que apresenta 80% ou mais da área cultivada com grandes talhões;
- 2) estrato que apresenta 80% ou mais da área cultivada com pequenos talhões.

Desta mesma forma, o estrato de área agrícola esparsamente cultivada, se definido como aquele que apresenta de 15% a 49% de área cultivada, pode apresentar estratos distintos para pequenos e grandes talhões.

Segundo Geuder (1984) as definições dos estratos de uso de solo para o Estado de Nebraska em 1983 foram as seguintes:

Código do Estrato	Definição do Estrato
11	Área agrícola: 80% ou mais cultivada.
12	Área agrícola: 50% a 79% cultivada.
20	Área agrícola: 15% a 49% cultivada.
31	Área agrícola urbanizada; núcleo residencial em área agrícola, com mais de 20 residências por milha quadrada.
32	Residencial/comercial, sem área agrícola.
40	Pastagens, menos de 15% de área cultivada.
50	Área não-agrícola (parques, instalações militares etc.).
62	Água, com mais de 1 milha quadrada em área.

Uma adaptação possível das definições acima seria:

CÓDIGO DO ESTRATO	DEFINIÇÃO DO ESTRATO
I	Área agrícola intensivamente cultivada: 80% ou mais da área cultivada com grandes talhões.
II	Área agrícola intensivamente cultivada: 80% ou mais da área cultivada com pequenos talhões.
III	Área agrícola moderadamente cultivada: de 50% a 79% de área cultivada com grandes talhões.
IV	Área agrícola moderadamente cultivada: de 50% a 79% de área cultivada com pequenos talhões.
V	Área agrícola esparsamente cultivada: de 15 a 49% de área cultivada com grandes talhões.
VI	Área agrícola esparsamente cultivada: de 15% a 49% de área cultivada com pequenos talhões.
VII	Pastagens: menos de 15% de área cultivada.
VIII	Área não-agrícola (montanhas, florestas, reservas ecológicas, parques nacionais, etc.).
IX	Água.

Uma vez discutidos e definidos os estratos específicos de uso do solo para uma dada área de estudo, a suas delimitações deverão ser feitas nas imagens orbitais TM-LANDSAT, na escala de 1:250.000 ou 1:100.000; em seguida, esses limites deverão ser passados para um "overlay" e ajustados, utilizando os mapas disponíveis em escala compatível aos limites físicos permanentes e identificáveis.

A vantagem de trabalhar com imagens na escala 1:250.000 é que são necessárias muito menos imagens comparativamente ao número necessário na escala de 1:100.000. Desta forma é possível construir um mosaico da área de estudo e, conseqüentemente, ter uma visão global desta. É claro que isto permite que a estratificação seja feita de maneira muito mais consistente do que o seria caso se trabalhasse com "partes" da área de estudo; este estudo por "partes" quase certamente teria de ser

feito numa área razoavelmente grande (Estado de São Paulo, por exemplo). A vantagem das imagens na escala de 1:100.000 é que elas fornecem mais detalhes físicos do que as imagens na escala de 1:250.000; elas permitem também uma maior diferenciação entre as diversas categorias de uso do solo.

## 2 - CONSTRUÇÃO DAS UNIDADES PRIMÁRIAS DE AMOSTRAGEM (UPAs) DENTRO DE CADA ESTRATO DE USO DO SOLO

Neste passo, os estratos de uso do solo são divididos em unidades primárias de amostragem (UPAs), respeitando os limites dos municípios inseridos na área de estudo, também ajustados aos limites físicos permanentes. As UPAs variam de tamanho, dependendo do estrato de uso do solo, da região de estudo e dos limites físicos permanentes disponíveis.

O objetivo da construção das UPAs é servir de elemento intermediário para a seleção das unidades secundárias (ou segmentos) que constituem, realmente, as unidades sobre as quais se levantarão as características de interesse para o estudo. Cada UPA contém um determinado número de segmentos (entre 6 e 20, segundo Wigton (1984a), os quais, entretanto, não são delineados neste estágio. Os segmentos só serão delineados em algumas UPAs a serem identificadas posteriormente. Vê-se, então, que este procedimento permite economia de tempo e recursos, pelo fato de dispensar que toda a área de estudo seja delineada em segmentos, o que, como mencionado acima, só será feito sobre algumas UPAs específicas.

Uma vez caracterizado que o tamanho da UPA é função do tamanho do segmento, algumas considerações sobre ele serão feitas.

O tamanho do segmento é determinado com base em vários fatores, tais como:

- A disponibilidade de limites físicos permanentes, facilmente identificáveis em campo: quanto menor for o tamanho do segmento, mais difícil fica a sua "amarração" a limites físicos.

- A ineficiência de segmentos muito pequenos para a classificação digital das imagens: como a resolução do satélite TM-LANDSAT é de 30m x 30m (nas bandas refletidas), num segmento muito pequeno haveria um número muito pequeno de "pixels" para ser classificado confiavelmente. Para ter uma idéia, cada "pixel" na imagem TM-LANDSAT sintetiza a informação contida numa área de 900 m<sup>2</sup> (ou 0,0009 km<sup>2</sup>); desta forma, num segmento de 2 km<sup>2</sup>, por exemplo, serão classificados 2222 "pixels".
- A própria convenção adotada na definição de segmento como sendo uma parcela de terra que pode ser coberta pelo enumerador de campo em um único dia de trabalho. Desta forma, entende-se que o tamanho dos segmentos num estrato de área esparsamente cultivada pode ser maior que o tamanho dos segmentos num estrato de área agrícola intensivamente cultivada, porque o número de estabelecimentos agrícolas a ser visitado neste deve ser, muito provavelmente, maior do que naquele.
- A variância entre segmentos, que deve ser minimizada para permitir maior homogeneidade na coleta de informações nos segmentos contidos nas UPAs em estratos específicos.
- O método a ser adotado para coleta de dados nos segmentos selecionados, visto haver três procedimentos básicos disponíveis para tal: método do segmento aberto, método do segmento fechado e método ponderado.

A medida que o tamanho do segmento aumenta, chega-se a um ponto onde a variância amostral aumenta rapidamente. Em geral, uma unidade amostral grande (que contém de 75 a 100 estabelecimentos) é muito ineficiente. Supondo, por exemplo, que um segmento tenha, em média, 100 estabelecimentos e que se vá proceder a uma amostragem de 2.500 estabelecimentos, ter-se-á de amostrar aproximadamente 25 segmentos. Já no caso de o segmento possuir, em média, 5 estabelecimentos, para esta amostragem de 2.500 estabelecimentos serão amostrados, aproximadamente, 500 unidades amostrais. Desta forma, a variância amostral seria bem menor.

A literatura cita que é preferível utilizar segmentos bem pequenos em termos do número de unidades de informação, da forma como são definidas para o estudo. Entretanto, é bom entender que é difícil determinar, na prática, um ótimo universal.

Somente após a realização de um trabalho piloto na área de interesse, ou mesmo de um trabalho completo sobre esta área, haverá condições de avaliar a eficiência dos tamanhos de segmento adotado para cada estrato de uso do solo, bem como de pensar na conveniência de um possível redimensionamento destes segmentos.

Segundo Wigton (1984a), o tamanho das unidades amostrais adotado na Província de Salcedo, República Dominicana, foi conforme o apresentado abaixo:

ESTRATO	TAMANHO DAS UNIDADES AMOSTRAS (SEGMENTOS)- EM KM <sup>2</sup>
Área intensivamente cultivada	2
Área esparsamente cultivada	4
Área não-agrícola	4
Área urbana	0,5

Também segundo Wigton, o número de segmentos em cada UPA deve estar compreendido entre 6 e 20. Assim, o tamanho das UPAs dentro de cada estrato específico de uso do solo ficaria:

ESTRATO	TAMANHO DAS UNIDADES PRIMÁRIAS DE AMOSTRAGEM (UPAs) EM KM <sup>2</sup>
Área agrícola intensivamente cultivada	12 - 40
Área agrícola esparsamente cultivada	24 - 80
Área não-agrícola	24 - 80
Área urbana	3 - 10

Já segundo Geuder (1984), as unidades primárias de amostragem devem conter de 2 a 10 segmentos. Desta forma, considerando que os tamanhos dos segmentos sejam definidos conforme apresentado na tabela de Wigton, na página anterior, o tamanho das UPAs ficaria:

ESTRATO	TAMANHO DAS UNIDADES PRIMÁRIAS DE AMOSTRAGEM (UPAs) EM KM <sup>2</sup>
Área agrícola intensivamente cultivada	4 - 20
Área agrícola esparsamente cultivada	8 - 40
Área não-agrícola	8 - 40
Área urbana	1 - 5

Convém ressaltar que somente a experiência adquirida ao longo da execução deste tipo de trabalho é que levará à determinação dos tamanhos ótimos, tanto para UPAs quanto para segmentos a serem adotados em trabalhos futuros, para cada região de estudo específica.

### 3 - CONSTRUÇÃO DOS DISTRITOS DE ANÁLISE

Uma consideração importante a ser feita refere-se às áreas cobertas por imagens de datas diferentes. Na construção de um mosaico para uma região de estudo (que pode ser um Estado, por exemplo), faz-se necessário utilizar imagens de datas diferentes, em virtude de o TM-LANDSAT captar informações de uma faixa específica do terreno a cada 16 dias. O intervalo de tempo entre uma órbita sobre uma faixa específica do terreno e sua adjacente é de 7 dias (Figura 2), o que implica que numa área de estudo que envolve várias imagens deve-se esperar que ocorram variações dos seguintes tipos:

- iluminação (que altera a resposta na topografia);
- alterações ambientais (causadas por precipitação, ventos, etc.);
- condições atmosféricas (presença de nuvens, neblina e poluição);
- efeito do ângulo solar;
- estágio fenológico da cultura (para culturas temporárias).

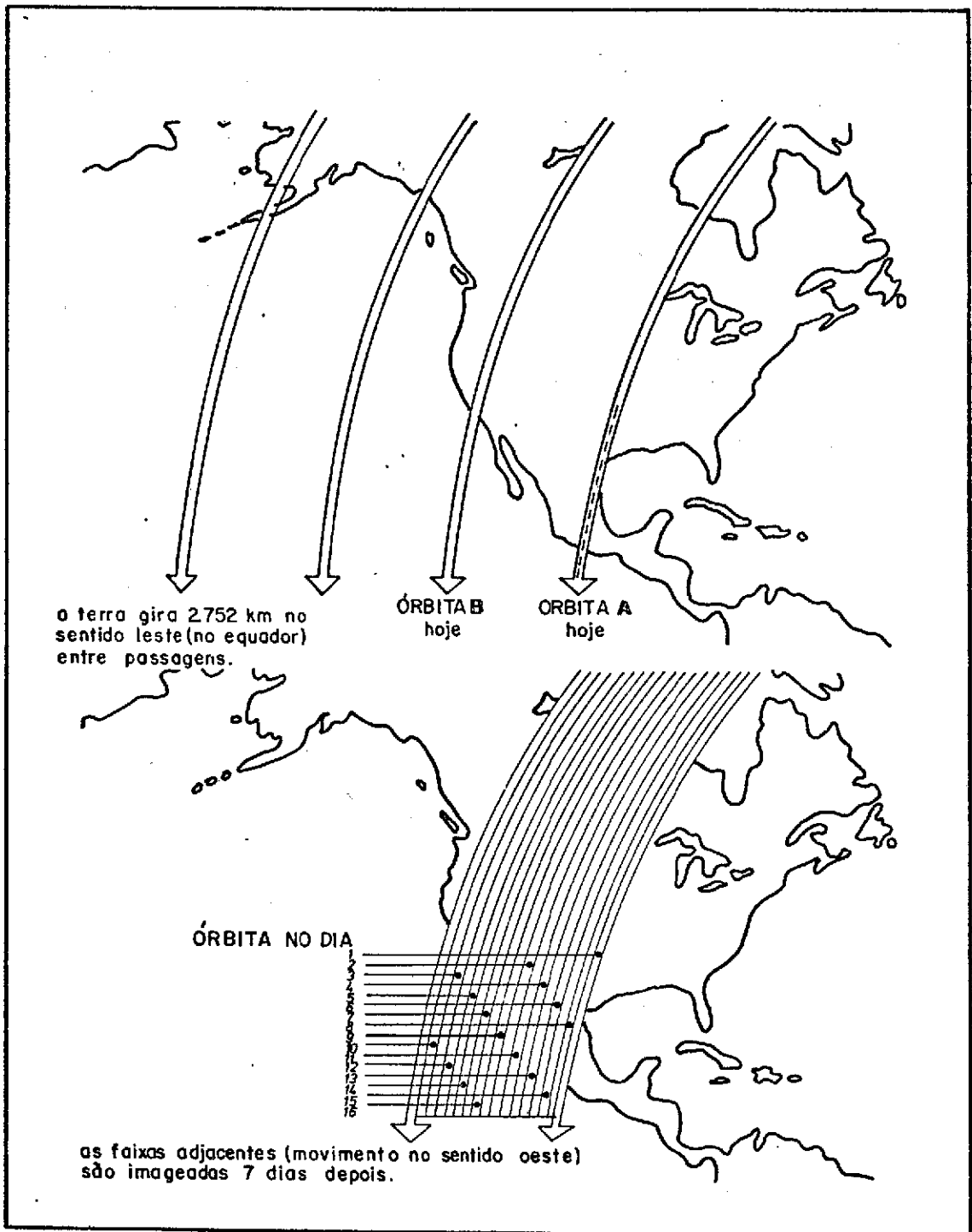


Fig. 2 - Ciclo de repetição orbital do satélite TM-LANDSAT.

FONTE: Irons (1983)



Esses fatores exigem que se trate cada conjunto de imagens de mesma data como uma população distinta. Esses conjuntos distintos dão origem aos chamados distritos de análise, os quais deverão ser construídos de modo que suas bordas obedeçam aos limites municipais (ajustados a limites físicos); isto só não será possível quando um município for cortado por imagens orbitais de datas diferentes. Neste caso, o município em questão será dividido em dois, denominados pseudomunicípios, e as bordas dos distritos de análise, neste caso, obedecerão aos limites das unidades primárias de amostragem que limitem cada um desses pseudomunicípios.

#### 4 - ATRIBUIÇÃO DE UM DETERMINADO NÚMERO DE SEGMENTOS A CADA UNIDADE PRIMÁRIA DE AMOSTRAGEM

Um determinado número de segmentos será designado a cada UPA, dependendo de sua área. Note que os segmentos ainda não são delimitados dentro de cada UPA; apenas o número deles é especificado.

O número de segmentos designados a cada UPA, para um estrato específico de uso do solo, é determinado dividindo a área da UPA pelo tamanho escolhido do segmento neste estrato e aproximando este quociente do inteiro mais próximo.

A área da UPA pode ser calculada através de uma grade de pontos ou da digitação desta num sistema automatizado.

#### 5 - CONSTRUÇÃO DOS SUBSTRATOS GEOGRÁFICOS

A construção dos substratos geográficos (que serão definidos ao longo deste trabalho) surge da preocupação de que uma amostra de UPAs, se selecionada através de um procedimento de amostragem aleatória simples ou proporcional ao número de segmentos contidos na UPA, se agrupe em áreas específicas, em vez de se dispersar ao longo do estrato em estudo, inserido na área de interesse. Esta dispersão ao longo do estrato, além de desejável, é recomendável, visto que se espera ter na amostra uma representatividade da população de onde ela está sendo retirada.

Suponha, por exemplo, um estrato de área intensivamente cultivada, que contém vários tipos de cultura distribuídas ao longo dele, conforme apresentado na Figura 3.

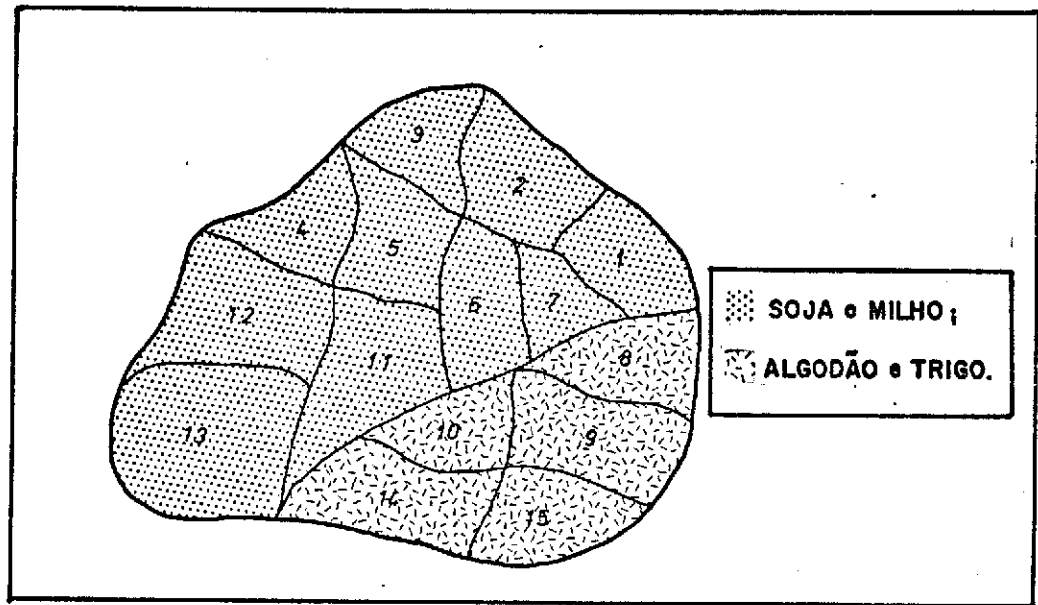


Fig. 3 - Distribuição espacial de culturas dentro de um estrato.

Vê-se pela Figura 3 que, embora o estrato seja de área agrícola intensivamente cultivada, ele poderia ter sido dividido em dois estratos distintos:

- I. área agrícola intensivamente cultivada com soja e milho;
- II. área agrícola intensivamente cultivada com algodão e trigo.

A dificuldade de fazer esta divisão na estratificação está no fato de que muitas vezes não é possível distinguir as diferentes culturas contidas no estrato, principalmente aquelas que apresentam gradientes de albedos próximos. Como na estratificação os estratos são delineados somente através de interpretação visual, o problema acima exposto fica, de certa forma, incontornável.

Se fosse realizada sobre este estrato uma seleção de UPAs utilizando um procedimento de amostragem proporcional ao número

de segmentos nelas contidos, nada impediria que, numa amostragem de 4 unidades primárias (UPAs), as UPAs de números 2, 4, 6 e 12 da Figura 3 fossem as selecionadas.

Caso isto ocorresse, a amostragem não seria representativa desta população, que contém, além de soja e milho (como indica a amostra), também algodão e trigo (que a amostragem não mostra). Note que isto não ocorreria se os estratos tivessem sido definidos como I e II acima, e fossem retiradas amostras independentes tanto de um estrato quanto de outro.

Uma forma de diminuir a probabilidade de ocorrência deste fato é construir os chamados subestratos geográficos, que nada mais são do que agrupamentos que contém um número igual ou aproximadamente igual de unidades amostrais dispostos de forma que apresentem características agrícolas semelhantes.

A seleção das UPAs é feita de forma indireta, ou seja, segmentos dentro de estratos geográficos serão selecionados, o que caracterizará automaticamente as UPAs amostradas. Isto ficará devidamente esclarecido ao longo desta seção.

Em primeiro lugar será esclarecida a forma como os subestratos geográficos serão construídos. Este procedimento envolve dois passos:

- I. numeração, em forma de serpentina, das UPAs dentro dos municípios e pseudomunicípios incluídos nos distritos de análise, sem considerar os estratos de uso do solo. Suponha, por exemplo, dois municípios da área de estudo, onde um deles foi dividido em dois pseudomunicípios, conforme apresentado na Figura 4. Note que a ordenação das UPAs foi feita em cada município ou pseudomunicípio individualmente, identificando-se, após a numeração, o estrato no qual cada uma delas está inserida.

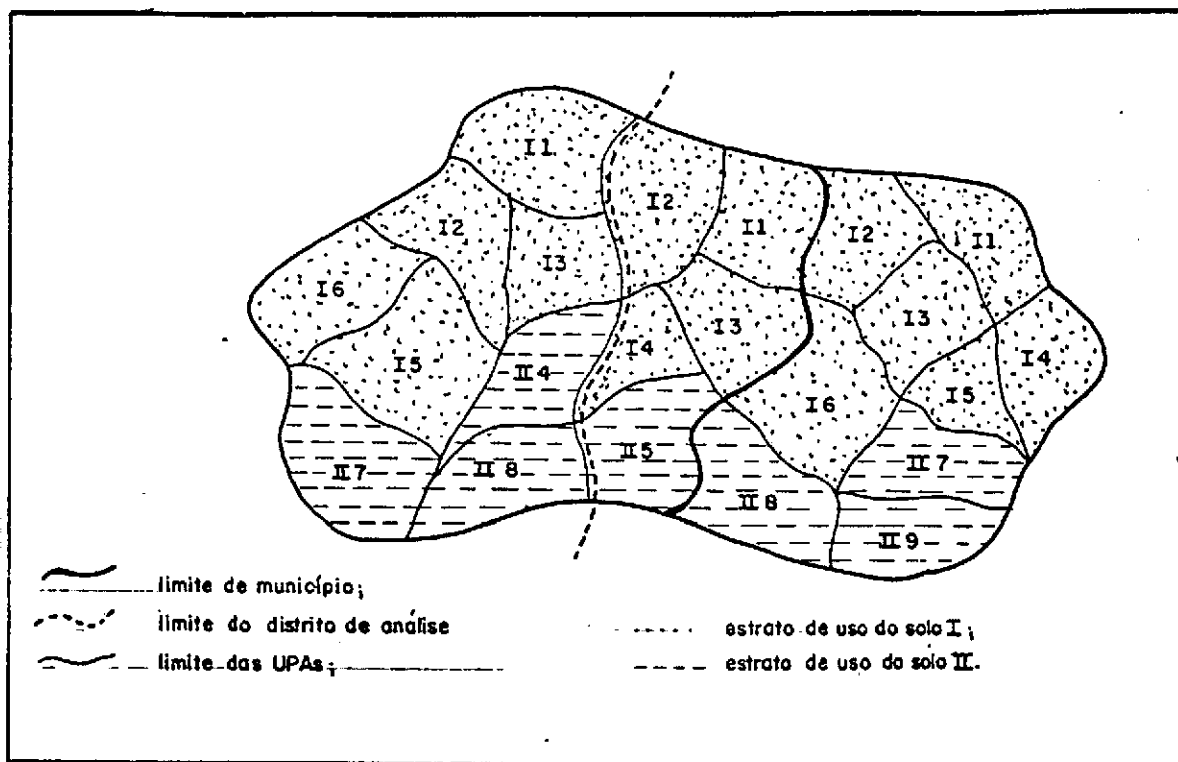


Fig. 4 - Numeração das UPAs nos municípios e pseudomunicípios.

II. ordenação dos municípios e pseudomunicípios em cada distrito de análise, através da utilização de um procedimento de análise de "cluster" (agrupamento) multivariada dos dados agropecuários no município, para identificação daqueles que apresentam aproximadamente as mesmas características agrícolas. É bom ressaltar que as informações nos pseudomunicípios a serem consideradas para análise de "cluster" deverão ser utilizadas na sua totalidade, ou seja, como se o município todo estivesse contido no distrito de análise. A Figura 5 mostra um exemplo de ordenação de municípios.

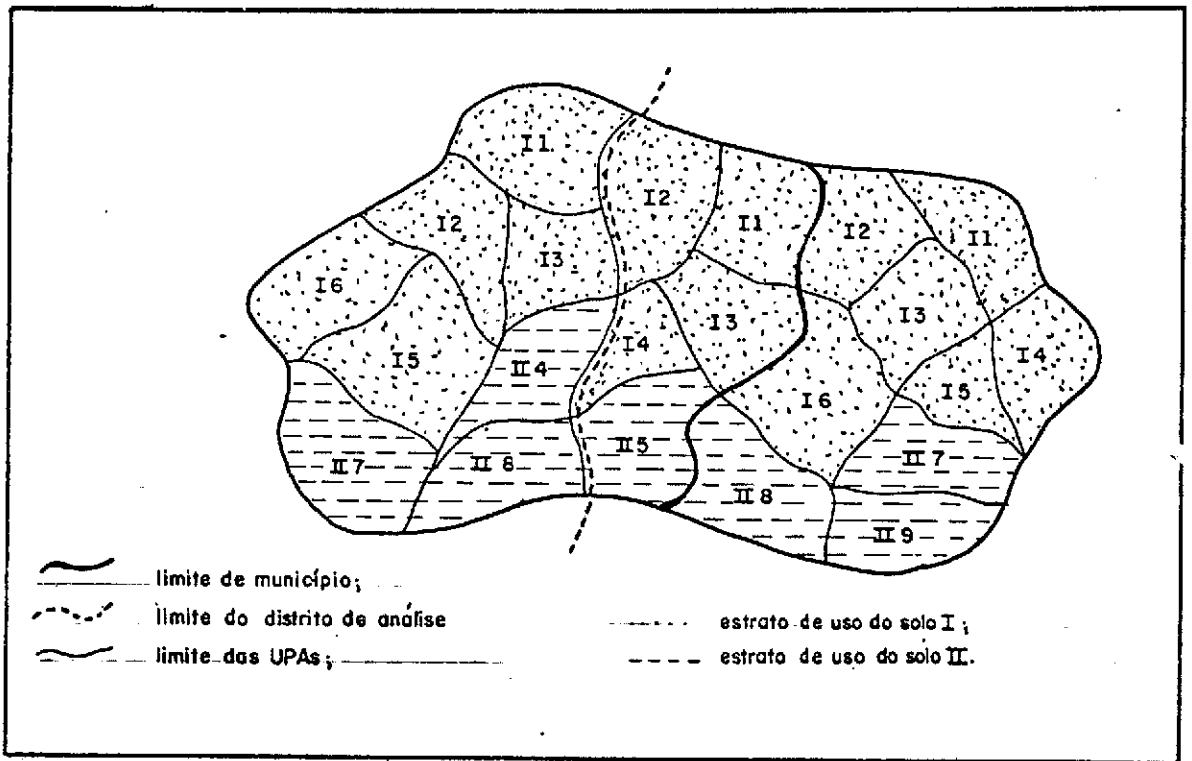


Fig. 4 - Numeração das UPAs nos municípios e pseudomunicípios.

II. ordenação dos municípios e pseudomunicípios em cada distrito de análise, através da utilização de um procedimento de análise de "cluster" (agrupamento) multivariada dos dados agropecuários no município, para identificação daqueles que apresentam aproximadamente as mesmas características agrícolas. É bom ressaltar que as informações nos pseudomunicípios a serem consideradas para análise de "cluster" deverão ser utilizadas na sua totalidade, ou seja, como se o município todo estivesse contido no distrito de análise. A Figura 5 mostra um exemplo de ordenação de municípios.

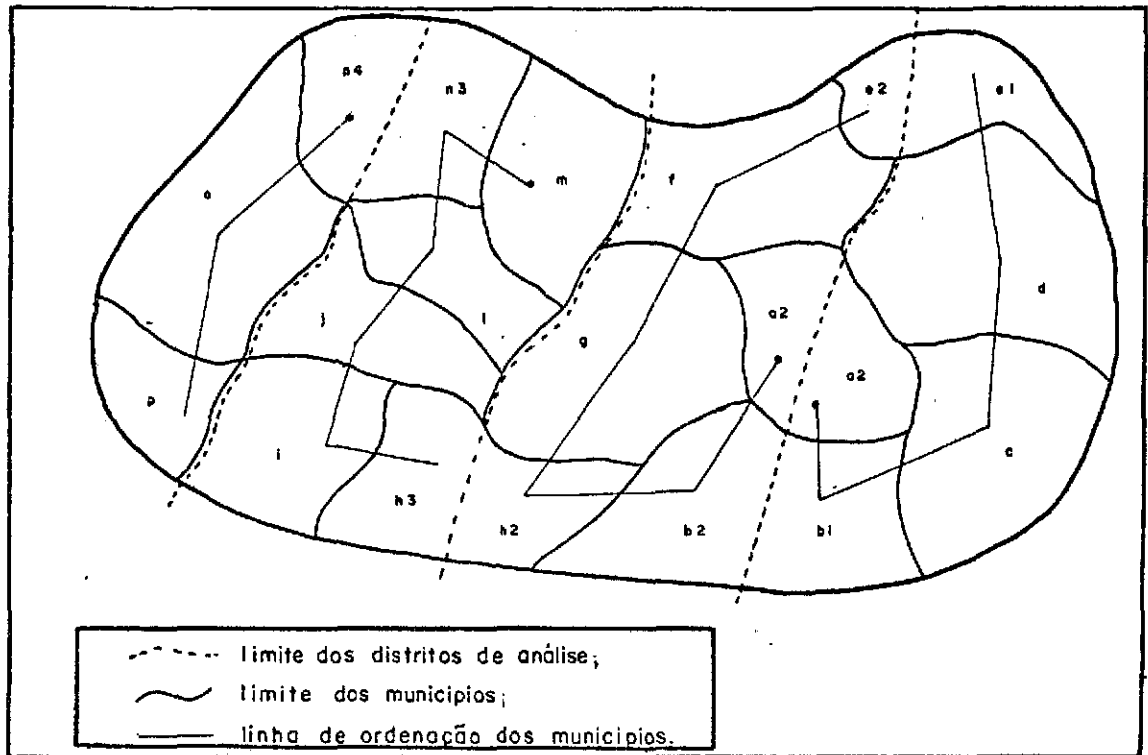


Fig. 5 - Exemplo hipotético de ordenação de municípios.

Muito embora o exemplo apresentado acima mostre uma contiguidade na ordenação dos municípios, isto, na prática, pode não ocorrer. Se, por acaso, os municípios localizados em um lado do distrito de análise apresentarem características similares àqueles situados no seu lado oposto, a ordenação dos municípios poderá ser feita de forma a desconsiderar vários municípios. Para maiores detalhes com relação à ordenação, ver Geuder(1984).

Após a realização desses dois passos, passa-se à ordenação da população como um todo, considerando os estratos de uso do solo. Esta ordenação considera inicialmente o resultado do passo II acima descrita (ordenação de municípios e pseudomunicípios, conforme apresentado na coluna 4 do exemplo da Tabela 1); logo após será então considerada a numeração das UPAs dentro desses municípios, resultado do passo I acima descrito (numeração das UPAs, apresentada na coluna 5 da Tabela 1). A renenumeração das UPAs no distrito de análise como um todo, dá origem ao apresentado na coluna 3 da Tabela 1. As áreas das UPAs (coluna 6 da Tabela 1) e a definição do tamanho ótimo de segmento permitem a caracterização do número potencial de segmentos dentro de cada UPA (coluna 7 da

Tabela 1). O número de segmentos é então acumulado, para gerar os valores apresentados na coluna 8 da Tabela 1.

Os subestratos geográficos poderão, neste ponto, ser definidos. O número de unidades amostrais em cada estrato de uso do solo foi determinado para toda a população e para a população ordenada. Após a determinação do tamanho total da amostra a ser tomada, assim como suas alocações ótimas nos estratos de uso de solo, pode-se partir para a determinação do número de subestratos geográficos que serão construídos em cada estrato específico. Esta determinação baseia-se em vários fatores, entre eles o número de segmentos de cada estrato de uso do solo, o grau de homogeneidade entre os segmentos nos estratos, etc. Obviamente espera-se que um estrato de área agrícola cultivada com muitas culturas seja dividido em mais subestratos geográficos do que o seria um estrato que contém um só tipo de cultura, por exemplo.

Note que os subestratos geográficos garantem a distribuição da amostra ao longo dos distritos de análise, para cada estrato de uso do solo, o que lhe possibilita aumentar a precisão das estimativas a serem geradas; este aumento na precisão deve-se ao fato de esses subestratos geográficos serem mais homogêneos que o estrato de uso do solo, o que diminui a quantidade de erro amostral; quanto mais eficiente for o processo de agrupamento das UPAs semelhantes, maior será esta diminuição.

Assim, se uma determinada cultura aparece em poucas áreas pequenas no estrato de uso do solo, e os subestratos geográficos isolam essas áreas, então pode-se esperar um ganho considerável na precisão; entretanto, esses ganhos serão mínimos no caso de a cultura se espalhar naturalmente ao longo do estrato de uso do solo.

Conforme sugerido pela teoria de amostragem estratificada, as variâncias dos estimadores são geralmente reduzidas à medida que o número de subestratos geográficos aumenta. Entretanto Geuder(1983) sugere que um número "moderado" seja tomado, onde "moderado" pode ter significados diferentes nos diferentes tipos de estratos de uso de solo. Por

exemplo, no estrato de área intensivamente cultivada, um número moderado pode ser de 10 a 15 subestratos geográficos, enquanto num estrato de pastagem este número pode ser de 4 a 8.

No exemplo apresentado na Tabela 1, o estrato I de uso do solo possui 24.200 segmentos, dos quais 6132 estão no distrito de análise 1. Suponha que se tenha decidido pela construção de 13 subestratos geográficos neste estrato específico de uso do solo. Desta forma, os primeiros subestratos geográficos contêm número igual de segmentos (1861). Entretanto, isto não ocorre necessariamente no último subestrato geográfico de cada distrito de análise. Por exemplo, o subestrato geográfico número três, que é o último definido no distrito de análise 1, contém 2410 segmentos.

#### 6. SELEÇÃO DE UM DETERMINADO NÚMERO DE SEGMENTOS EM CADA SUBESTRATO GEOGRÁFICO

Uma vez apresentados os subestratos geográficos, parte-se para a seleção aleatória de segmentos dentro de cada um deles.

Assim, se o tamanho da amostra a ser tomada num determinado estrato de uso do solo for 195 e se for considerada conveniente a construção de 13 subestratos geográficos neste estrato, então 15 segmentos de cada subestrato geográfico serão selecionados.

Pode-se definir um conceito que terá uma aplicação muito grande na configuração da amostra a ser tomada nos anos seguintes, na mesma área de estudo. Este é o conceito de réplicas, ou amostras interpenetrantes. A construção dos subestratos geográficos é feita de forma a conter um número aproximadamente igual de segmentos, o que constitui sub-populações de tamanho aproximadamente igual. Uma amostragem de 15 segmentos de cada subestrato geográfico gera 15 amostras de 13 segmentos, cada uma delas podendo ser entendida como uma réplica ou amostra interpenetrante, ou seja, cada amostra interpenetrante contém um elemento de cada subestrato geográfico. Isto fica mais claro na Tabela 2.



TABELA 1

CONSTRUÇÃO DOS SUBSTRATOS GEOGRÁFICOS

(1) DISTRITO DE ANÁLISE	(2) ESTRATO DE USO DO SOLO	(3) REORDENAÇÃO DAS UPAs	(4) ORDENAÇÃO DOS MU- NICÍPIOS OU PSEU- DOMUNICÍPIOS	(5) NÚMERO DA UPA NO MUNICÍPIO OU PSEUDOMUNICÍPIO	(6) ÁREA DA UPA	(7) NÚMERO DE SEGMENTOS NA UPA	(8) SEGMENTO ACUMULADO	(9) SUBSTRATO GEOGRÁFICO
1	I	I.1	a <sub>1</sub>	2	24,3	12	12	1
		I.2	a <sub>1</sub>	3	16,9	8	20	1
		I.3	a <sub>1</sub>	5	12,7	6	26	1
		.	.	.	.	.	.	.
		.	.	.	.	.	.	.
		.	.	.	.	.	.	.
		I.189	c	9	24,8	10	1861	1
		I.189	c	9	24,8	2	1863	2
		.	.	.	.	.	.	.
		I.283	d	3	14,6	7	3722	2
2	I	I.407	a <sub>2</sub>	5	24,9	12	6144	4
		.	.	.	.	.	.	
		.	.	.	.	.	.	
		.	.	.	.	.	.	
		I.406	e <sub>1</sub>	7	24,8	12	6132	3
3	I	I.984	m	7	24,1	12	18500	10
		.	.	.	.	.	.	
		.	.	.	.	.	.	
4	I	.	.	.	.	.	.	
		.	p	11	12,5	6	24200	13

OBSERVAÇÃO: A UPA de nº 9 do município c, que contém 12 segmentos, pertence a dois substratos geográficos distintos: os 10 primeiros segmentos pertencem ao substrato geográfico 1, os 2 restantes, ao substrato geográfico 2.

TABELA 1

CONSTRUÇÃO DOS SUBSTRATOS GEOGRÁFICOS

(1) DISTRITO DE ANÁLISE	(2) ESTRATO DE USO DO SOLO	(3) REORDENAÇÃO DAS UPAs	(4) ORDENAÇÃO DOS MU- NICÍPIOS OU PSEU- DOMUNICÍPIOS	(5) NÚMERO DA UPA NO MUNICÍPIO OU PSEUDOMUNICÍPIO	(6) ÁREA DA UPA	(7) NÚMERO DE SEGMENTOS NA UPA	(8) SEGMENTO ACUMULADO	(9) SUBSTRATO GEOGRÁFICO
1	I	I.1	a <sub>1</sub>	2	24,3	12	12	1
		I.2	a <sub>1</sub>	3	16,9	8	20	1
		I.3	a <sub>1</sub>	5	12,7	6	26	1
		.	.	.	.	.	.	.
		.	.	.	.	.	.	.
		.	.	.	.	.	.	.
		I.189	c	9	24,8	10	1861	1
		I.189	c	9	24,8	2	1853	2
		.	.	.	.	.	.	.
		I.283	d	3	14,6	7	3722	2
2	I	I.407	a <sub>2</sub>	5	24,9	12	6144	4
		.	.	.	.	.	.	
		.	.	.	.	.	.	
		.	.	.	.	.	.	
3	I	I.984	m	7	24,1	12	18500	10
		.	.	.	.	.	.	
		.	.	.	.	.	.	
4	I	.	.	.	.	.	.	.
		.	.	.	.	.	.	
		.	p	11	12,5	6	24200	13

OBSERVAÇÃO: A UPA de nº 9 do município c, que contém 12 segmentos, pertence a dois substratos geográficos distintos: os 10 primeiros segmentos pertencem ao substrato geográfico 1, os 2 restantes, ao substrato geográfico 2.

TABELA 2

SEGMENTOS AMOSTRADOS NOS 13 SUBSTRATOS GEOGRÁFICOS, OS QUAIS DÃO ORIGEM A 15 RÉPLICAS, OU 15 AMOSTRAS INTERPENETRANTES

NÚMERO DO SUBSTRATO GEOGRÁFICO	R E P L I C A Ç Õ E S														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	seg 1261	seg 888	seg 009	seg 447	seg 222	seg 534	seg 476	seg 489	seg 1555	seg 001	seg 1476	seg 456	seg 888	seg 732	seg 490
2	seg 2443	seg 1974	seg 3222	seg 2999	seg 3000	seg 3111	seg 2765	seg 2987	seg 3119	seg 2554	seg 3001	seg 3456	seg 2442	seg 1987	seg 2000
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
13	seg 23441	seg 24199	seg 23876	seg 22999	seg 23333	seg 24111	seg 23001	seg 23045	seg 24167	seg 23876	seg 24355	seg 23444	seg 23132	seg 22987	seg 23030
NÚMERO DA RÉPLICA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Note que cada coluna da Tabela 2 representa uma amostra aleatória independente de 13 segmentos, seguramente espalhados geograficamente na área de estudo.

Uma vantagem das amostras interpenetrantes (ou réplicas) é o fato de que a cada ano pode-se pensar em substituir somente 20% delas antes de proceder à coleta de dados em campo, o que de certa forma evita entrevistar, ano após ano, os mesmos estabelecimentos e impede o cansaço e o desgaste dos entrevistados, o que poderia contribuir para o aumento dos erros não-amostrais.

Assim, no exemplo apresentado nesta seção, na pesquisa a ser aplicada no ano seguinte, apenas 20% das 15 réplicas seriam substituídas, ou seja, apenas 3 delas.

Embora, como foi exposto acima, o número de substratos geográficos seja definido antes de decidir o número de réplicas, pode-se muitas vezes inverter este processo e partir da decisão do número de amostras interpenetrantes e ter como consequência, a definição do número de substratos geográficos a ser formado.

Este procedimento prático pode ser feito pelo fato de que, segundo Geuder (1983), diferentes números de substratos geográficos a serem formados podem ser considerados como números ótimos. O próprio fato de que 20% das amostras serão substituídas cada ano faz com que este tipo de procedimento se torne prático, devendo-se escolher o número de réplicas como múltiplos de 5.

## 7 - CONSTRUÇÃO DOS SEGMENTOS NAS UNIDADES PRIMÁRIAS DE AMOSTRAGEM

Entende-se que muito embora a seleção dos segmentos a serem amostrados já tenha sido feita até este estágio, eles ainda não estão delineados nas UPAs a que pertencem. Assim, este passo diz respeito à identificação e divisão dos segmentos designados às UPAs selecionadas. Lembre-se de que esta seleção de UPAs é feita indiretamente, através da seleção de segmentos dentro dos substratos geográficos

cos definidos. Por exemplo, vê-se pela Tabela 1 que se o segmento de número 25 for selecionado (como um dos 15 que serão selecionados no subestrato 1), automaticamente a UPA de número 3 localizada no pseudo município  $a_1$  será selecionada.

É claro que será necessário estabelecer a priori o critério a ser adotado para a numeração dos segmentos dentro das UPAs selecionadas. Pode-se pensar na adoção de um sistema de numeração em forma de serpentina, iniciando a numeração no segmento localizado mais acima e mais à direita dentro da UPA, por exemplo.

Deve-se ter em mente, também, que o delineamento dos segmentos dentro das respectivas UPAs nem sempre é uma tarefa fácil. Esses segmentos também deverão estar "amarrados" a limites físicos, identificáveis em campo, visto constituírem elementos sobre os quais efetivamente se coletarão as informações necessárias para o estudo que está sendo realizado.

As fotografias aéreas constituem um material muito importante, pois mostram mais detalhes do que mapas, tais como pequenos caminhos, rede elétrica, córregos, etc., elementos estes que podem ser utilizados na "amarração" do segmento.

Mesmo assim, nem sempre é possível a divisão da UPA selecionada no número de segmentos a ela designado. Na maior parte das vezes, entretanto, isto ocorre devido à qualidade e à quantidade de material disponível, e não à falta real de limites físicos em campo. Entretanto, na impossibilidade de delinear os segmentos atribuídos à UPA, duas soluções são possíveis:

- (a) visitar a UPA e dividi-la no número de segmentos a ela atribuído, sobre a base de detalhes topográficos observados em campo;
- (b) dividir a UPA em tantos segmentos quanto for possível utilizando os materiais disponíveis, para designar a estas parcelas delineadas um determinado número de segmentos, de forma

que o somatório de segmentos alocados em cada parcela seja igual ao número designado à UPA. Verificar, então, a qual parcela de terra o segmento selecionado pertence. Quanto a esta parcela de terra deve-se:

- (i) visitar a parcela de terra, subdividi-la em campo no número de segmentos a ela designado (talvez este trabalho possa ser realizado pelo próprio enumerador de campo bem-treinado, o que evita o deslocamento de pessoal a esta UPA duas vezes: uma para identificação do segmento e outra para coleta de dados sobre ele); ou
- (ii) coletar dados sobre toda a parcela e dividir o total do item de interesse coletado nesta parcela pelo número de segmentos a ela atribuído; desta forma, ter-se-á uma estimativa da média do item investigado, por segmento.

Suponha, por exemplo, que uma UPA deva ser dividida em 5 segmentos. Pode ser possível dividi-la em duas partes (parcelas de terra) e designar três segmentos à primeira e dois à segunda, estabelecendo-se um critério conveniente de numeração para eles dentro das UPAs. Se o segmento selecionado pertencer, por exemplo, à segunda parcela de terra, pode-se ou dividi-la em campo ao visitá-la e coletar informações somente sobre o segmento identificado, ou coletar dados sobre toda a parcela e daí dividir por dois o total do item de interesse obtido nesta parcela. Esta técnica de dividir parcialmente a UPA reduz muito o número necessário de mapas e fotografias aéreas. É importante ressaltar, também, que durante o delineamento dos segmentos é importante especificar detalhes de procedimento que eliminem a possibilidade de viés. Por exemplo, é muito importante que o processo de divisão das UPAs seja feito sem o conhecimento de número do segmento selecionado naquela UPA.

Outro ponto deve ser levantado na construção dos segmentos dentro das UPAs, referente à variação que pode ocorrer na sua delimitação. O que ocorre em termos de variação pode ser devido à variação dentro do segmento e à variação entre os segmentos. Em contraste com o processo de estratificação, onde se busca formar estratos homogêneos que apresentam uma pequena variação *dentro* deles e uma grande variação *entre* eles, os segmentos devem ser construídos de forma a possuir a máxima variação dentro deles e mínima entre eles. Apesar de o tamanho da amostra ser função do tamanho do segmento, isto é, quanto maior o segmento, menor o tamanho da amostra, se os segmentos forem construídos de forma a tornar mínima a variância entre eles, pode-se obter a mesma precisão com uma amostra de tamanho menor, mesmo sem aumentar o tamanho do segmento. Assim, uma vez que a variância amostral é função da variação entre segmentos, um dos objetivos na delimitação deles deve ser tornar tão pequena quanto possível a variação entre eles.

A Figura 6 mostra uma variação típica dentro de uma unidade primária de amostragem.

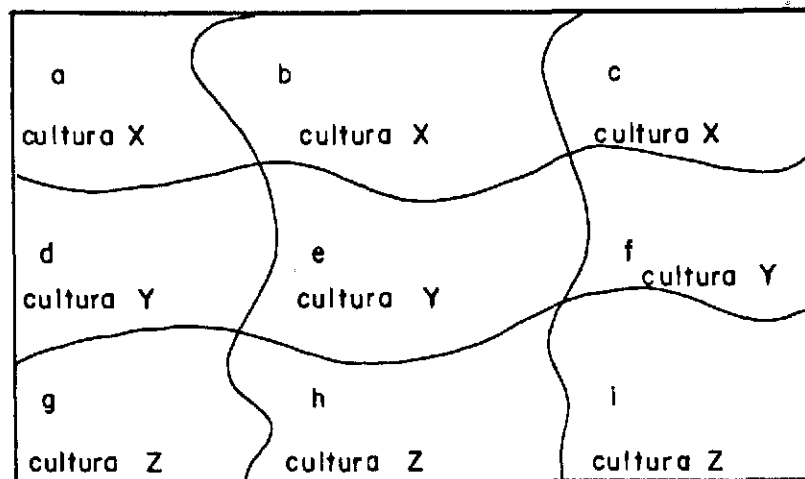


Fig. 6 - Variação típica dentro de uma UPA.

Há duas maneiras de dividir a área em três segmentos. A primeira é colocar a cultura X em um segmento (a, b, c), a cultura Y

no segundo (d, e, f) e a cultura Z no terceiro segmento (g, h, i). Esta será a pior forma possível de divisão, porque maximiza a variação entre os segmentos. A forma mais adequada de subdividir a área em três segmentos é colocar um pouco de cada cultura em cada um deles. Assim, os segmentos (a, d, g), (b, e, h) e (c, f, i) formam uma subdivisão muito mais eficiente estatisticamente, uma vez que, não importando o segmento selecionado, os resultados da coleta serão aproximadamente os mesmos.

## 8 - COLETA DE DADOS NOS SEGMENTOS AMOSTRADOS

A coleta de dados propriamente dita precedem algumas atividades fundamentais descritas sucintamente abaixo:

- (a) Treinamento dos Enumeradores de Campo: para tornar os procedimentos de coleta de dados mais uniformes, faz-se necessário oferecer uma curso de treinamento a todos os enumeradores de campo, não só para capacitá-los a identificar inequivocamente os respondentes ou a preencher os questionários para cada um deles, como também a enfrentar as mais variadas situações que possam ocorrer.
- (b) Organização do Material: Cada enumerador de campo deverá ter consigo:
  - mapa de estradas municipais, com a localização dos segmentos selecionados;
  - fotografia aérea com os limites dos segmentos demarcados;
  - listagem dos estabelecimentos agrícolas que estão total ou parcialmente contidos no segmento selecionando;
  - guia do enumerador, a ser fornecido no curso de treinamento.

Ao visitar cada segmento amostrado, o enumerador de campo deverá:



- (a) Localizar as unidades informantes: dependendo do estimador a ser utilizado (aberto, fechado ou ponderado), as unidades informantes deverão ser identificadas de forma a satisfazer as suas definições, as quais estão caracterizadas no passo 11 deste trabalho.
  
- (b) Identificar os "tracts" dentro do segmento amostrado: os "tracts" são definidos como parcelas de terra dentro do segmento, as quais estão sob uma única administração. Pode ser um estabelecimento agrícola inteiro, uma parte do estabelecimento agrícola, ou uma parte de terra não-agrícola; ou seja, um "tract" é determinado pela definição de estabelecimento agrícola e pelos limites de um segmento. Um estabelecimento agrícola contém um ou mais "tracts".

Idealmente todo segmento amostrado deveria ter uma fotografia aérea correspondente à sua área, na qual deverão ser demarcados os limites dos "tracts"; a área destes deverá ser estimada com o uso de uma grade de pontos sob a fotografia e confrontada com a área declarada pelo entrevistado. Caso os dois valores sejam muito discrepantes, a área em questão deverá ser discutida de forma a chegar a um consenso. A medida da área de cada "tract" deverá ser declarada na própria fotografia. Casose julgue mais conveniente, pode-se em vez de fazer as observações devidas nas próprias fotografias, utilizar um "overlay" sobre elas.

## 9 - TRATAMENTO DOS SEGMENTOS AMOSTRADOS

Cada segmento deve sofrer um tratamento digital que identifica os segmentos amostrados na correspondente imagem LANDSAT. Em seguida, através do uso de algoritmos de análise de "cluster", todos os "pixels" para cada cultura são processados. Várias assinaturas espectrais (subclasses) são geradas pelo algoritmo para cada cultura. A cada uma delas são fornecidos o número de "pixels" classificados como pertencentes à sua classe em cada segmento e o número de "pixels" classificados da mesma classe, em cada subestrato geográfico.

## 10 - UTILIZAÇÃO DO ESTIMADOR ADEQUADO PARA A GERAÇÃO DAS ESTIMATIVAS DE SEJADAS

A metodologia proposta neste trabalho prevê a utilização de dados coletados por satélites de sensoriamento remoto para a geração de estimativas dos totais relativos a itens de interesse, numa determinada área de estudo.

Objetiva-se a utilização desta informação, sempre que possível, como valores de uma variável auxiliar, para melhorar a precisão das estimativas a serem geradas. Entretanto, esta informação orbital nem sempre é disponível para toda a área de estudo. Isto ocorreria devido a problemas de cobertura de nuvens ou devido a características próprias da variável de interesse no estudo; nestes casos, as informações referentes a esta variável serão coletadas somente em campo; assim, os dados são tratados de modo diferente daqueles que possuem informações coletadas pelo satélite. Para os primeiros serão utilizados estimadores por expansão direta (utilizando somente dados coletados em campo) e para os segundos, estimadores de regressão (combinando os dados obtidos em campo com os dados coletados pelo satélite).

A seguir, cada um desses estimadores será descrito com maiores detalhes.

### I - ESTIMADORES POR EXPANSÃO DIRETA

Hã, basicamente, três tipos de estimadores de expansão direta disponíveis num delineamento amostral por áreas, quais sejam: estimador fechado, estimador aberto e estimador ponderado.

#### I.1 - ESTIMADOR FECHADO

Para o uso do estimador fechado os dados são coletados sobre todas as parcelas de terra dentro do segmento amostrado ("tracts"), onde se desenvolve alguma atividade agrícola (ou agropecuária). Assim, um estabelecimento agrícola, embora pertencente a um

são proprietário, pode ser operado por várias pessoas, através de arrendamento, usucapião ou empréstimo de parcelas de terra dentro do estabelecimento.

Com o uso de fotografias aéreas que cobrem o segmento amostrado, o enumerador de campo identifica e demarca essas parcelas, coletando então as informações desejadas a cada um de seus respectivos operadores (operadores de terra ou de parcelas de terra).

A notação que se segue refere-se aos elementos das fórmulas utilizadas na obtenção da estimativa da área total para uma dada cultura  $\sigma$ , assim como para a obtenção da variância estimada deste total.

Sejam:

$a$  = índice para representar distrito de análise;  $a = 1, \dots, A$ ;

$i$  = índice para representar estrato de uso do solo;

$i = 1, 2, \dots, s_a$ ;

$j$  = índice para representar subestrato geográfico;

$j = 1, 2, \dots, s_{ai}$ ;

$k$  = índice para representar segmento amostrado;

$k = 1, 2, \dots, n_{aij}$ ;

$l$  = índice para representar estabelecimentos respondentes;

$l = 1, 2, \dots, n_{aijk}$ ;

$A$  = número de distritos de análise dentro da região de estudo;

$s_a$  = número de estratos de uso do solo no  $a$ -ésimo distrito de análise;

$s_{ai}$  = número de subestratos geográficos no  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise;

$n_{aij}$  = número de segmentos amostrados (ou réplicas) no  $j$ -ésimo subestrato geográfico,  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise;

$n_{aijk}$  = número de estabelecimentos respondentes no k-ésimo segmento, j-ésimo substrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise;

$N_{aij}$  = número total de segmentos no j-ésimo substrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise;

$e_{aij} = (n_{aij}/N_{aij})^{-1}$  = fator de expansão ou inverso da probabilidade de que um segmento tem de estar na amostra retirada do j-ésimo substrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise;

$z_{aijkl}$  = valor da característica desejada, relatada pelo l-ésimo operador agrícola no k-ésimo segmento, j-ésimo substrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise;

$\bar{z}_{aij}$  = valor médio da característica desejada, relatada por segmento no j-ésimo substrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise.

Deste modo, a estimativa do total da cultura  $c$  no i-ésimo estrato de uso do solo é dada por:

$$\hat{z}_{c,i} = \sum_{a=1}^A s_{ai} \sum_{j=1}^J e_{aij} \sum_{k=1}^{n_{aij}} z_{aijk} \quad ,$$

onde

$$z_{aijk} = \begin{cases} \sum_{l=1}^{n_{aijk}} z_{aijkl} & \text{se } n_{aijk} > 0 \quad , \\ 0 & \text{se } n_{aijk} = 0 \quad , \end{cases}$$

cuja variância estimada  $\bar{e}$ :

$$v(\bar{Z}_{c,i}) = \sum_{a=1}^A \sum_{j=1}^{s_{ai}} N_{aij} (N_{aij} - n_{aij}) \frac{s_{aij}^2}{n_{aij}},$$

onde

$$s_{aij}^2 = \frac{n_{aij}}{\sum_{k=1}^{n_{aij}} (z_{aijk} - \bar{z}_{aij})^2} / (n_{aij} - 1) \quad (1)$$

O total estimado para a cultura c na área de estudo  $\bar{e}$  é dado por:

$$\bar{Z}_c = \sum_{a=1}^A \sum_{i=1}^{s_a} \sum_{j=1}^{s_{ai}} e_{aij} \frac{n_{aij}}{\sum_{k=1}^{n_{aij}} z_{aijk}} \quad (2)$$

e a variância estimada para este total  $\bar{e}$  dada por:

$$v(\bar{Z}_c) = \sum_{a=1}^A \sum_{i=1}^{s_a} \sum_{j=1}^{s_{ai}} \frac{N_{aij}(N_{aij} - n_{aij})}{n_{aij}} s_{aij}^2 \quad (3)$$

onde  $s_{aij}^2$  foi definido em (1).

## I.2 - ESTIMADOR ABERTO

O estimador aberto requer que os dados sejam coletados nos estabelecimentos cuja sede (definida apropriadamente) se situe no segmento selecionado; esses dados serão coletados sobre toda a área onde se desenvolvem operações agrícolas (ou agropecuárias, dependendo do objetivo do estudo), e não somente nas áreas contidas dentro do segmento.

Uma definição apropriada para sede pode ser aquela que caracteriza como sede todo estabelecimento no qual o operador agrícola efetivamente reside. É claro que há necessidade de definir apropriadamente operador agrícola para permitir que o enumerador de campo não tenha dúvidas quanto à inclusão ou não de determinado estabelecimento, na coleta de dados.

As fórmulas a serem utilizadas para a obtenção da estimativa da área total para uma dada cultura  $c$ , assim como a variância estimada deste total, são as mesmas apresentadas na seção I.1 para o estimador fechado. Deve-se observar, entretanto, que os valores de  $z_{aijkl}$  e  $n_{aijk}$  podem ser diferentes daqueles observados no estimador fechado, uma vez que os dados são coletados diferentemente, dependendo da abordagem a ser utilizada, i.e., de segmento fechado ou aberto.

## I.3 - ESTIMADOR PONDERADO

O uso do estimador ponderado requer que os dados sejam coletados nos estabelecimentos totalmente incluídos no segmento, ou que contenham parcelas de terra, ou "tracts", contidos no segmento amostrado. Com o uso de fotografias aéreas, o enumerador de campo identifica e demarca essas parcelas, coletando então as informações desejadas junto a cada um de seus respectivos operadores. Os dados são coletados sobre todo o estabelecimento ou exploração agrícola e ponderados pela fração de terra do estabelecimento ou pela exploração agrícola que se encontra no segmento.

A estimativa para o total da variável de interesse utilizando o estimador ponderado pode ser expressa pela seguinte fórmula, para uma dada área de estudo:

$$\hat{W}_{c,i} = \frac{A}{s_{ai}} \sum_{j=1}^{n_{aij}} e_{aij} \sum_{k=1}^{n_{aij}} w_{aijk} \quad ,$$

onde

$$w_{aijk} = \begin{cases} \frac{n_{aijk}}{\sum_{l=1}^{n_{aijk}} w_{aijkl}} & \text{se } n_{aijk} > 0 \\ 0 & \text{se } n_{aijk} = 0 \end{cases}$$

e

$$w_{aijkl} = p_{aijkl} \cdot z_{aijkl} \quad ,$$

onde

$p_{aijkl}$  = fração do l-ésimo estabelecimento ou exploração agrícola que se encontra dentro do k-ésimo segmento, j-ésimo subestrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise;

$w_{aijkl}$  = valor ponderado da variável de interesse no l-ésimo estabelecimento que se encontra no k-ésimo segmento, j-ésimo subestrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise.

As outras variáveis são definidas de forma idêntica ao que foi definido para segmento fechado; observa-se novamente que os

valores de  $z_{aijk}$  e  $n_{aijk}$  podem ser diferentes daqueles obtidos em qualquer uma das abordagens anteriores.

O estimador da variância para o total da variável dentro do estrato  $i$  é dado por:

$$v(\bar{W}_{C,i}) = \sum_{a=1}^A \sum_{j=1}^{s_{ai}} N_{aij} (N_{aij} - n_{aij}) \frac{s_{aij}^2}{n_{aij}}$$

onde

$$s_{aij}^2 = \frac{\sum_{k=1}^{n_{aij}} (w_{aijk} - \bar{w}_{aij})^2}{n_{aij} - 1} \quad (4)$$

Assim, a estimativa para o total da variável dentro da área de interesse é:

$$\bar{W}_C = \sum_{a=1}^A \sum_{i=1}^{s_a} \sum_{j=1}^{s_{ai}} \frac{N_{aij}}{n_{aij}} \sum_{k=1}^{n_{aij}} w_{aijk}$$

e a variância estimada do total é dada por:

$$v(\bar{W}_C) = \sum_{a=1}^A \sum_{i=1}^{s_a} \sum_{j=1}^{s_{ai}} \frac{N_{aij} (N_{aij} - n_{aij})}{n_{aij}} s_{aij}^2 ,$$

onde  $s_{aij}^2$  é dada em (4).



## II. ESTIMADORES DE REGRESSÃO - USO DOS DADOS DE SATÉLITE

Para a utilização dos dados de satélite nas estimativas de área cultivada numa dada área de interesse, através de estimadores de regressão, é necessário ter os seguintes limites: dos distritos de análise, dos estratos de uso do solo, de cada um dos subestratos geográficos e dos segmentos amostrados; todos estes limites deverão estar registrados e digitalizados.

A região em estudo é, agora, dividida em dois pós-estratos distintos: um é constituído pelos subestratos geográficos que estão total ou parcialmente cobertos por nuvens; o outro é constituído pelos subestratos geográficos livres de nuvens.

Para o primeiro deles serão obtidos estimadores para os itens de interesse, utilizando somente dados de campo; para tal, um dos estimadores por expansão direta apresentado em 10.I será utilizado. Para o segundo pós-estrato de subestratos geográficos (livre de nuvens), os dados de campo sobre os itens de interesse, em cada segmento selecionado, serão utilizados em combinação com os dados orbitais classificados digitalmente, referentes ao mesmo segmento e aos mesmos itens. Isto é feito através de um estimador de regressão, onde os dados orbitais constituem os valores da variável auxiliar (independente) e os dados coletados em campo, os valores da variável dependente.

Em cada um desses dois pós-estratos, os subestratos geográficos são renumerados.

Um estimador de regressão deverá ser construído para cada item de interesse, em cada um dos subestratos geográficos livres de nuvens definidos em cada estrato de uso do solo, para cada distrito de análise.

Assim, o estimador de regressão para a cultura  $c$ , no  $j$ -ésimo subestrato geográfico,  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise é definido por:

$$\hat{Y}_{aij}(\text{reg}) = N'_{aij} [\bar{y}_{aij} + b_{aij}(\bar{X}_{aij} - \bar{x}_{aij})],$$

onde:

$\bar{X}_{aij}$  = número médio populacional de "pixels" por segmento da cultura  $c$  no  $j$ -ésimo subestrato geográfico,  $i$ -ésimo estrato do uso do solo e  $a$ -ésimo distrito de análise para o pós-estrato livre de nuvens;

$x_{aijk}$  = número total de "pixels" da cultura  $c$  no  $k$ -ésimo segmento amostrado,  $j$ -ésimo subestrato geográfico,  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise para o pós-estrato livre de nuvens;

$N'_{aij}$  = número total de segmentos no  $j$ -ésimo subestrato geográfico,  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise no pós-estrato livre de nuvens;

$n'_{aij}$  = número de segmentos amostrados no  $j$ -ésimo subestrato geográfico,  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise no pós-estrato livre de nuvens;

$\bar{x}_{aij}$  = número médio amostral de "pixels" por segmento da cultura  $c$  no  $j$ -ésimo subestrato geográfico,  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise para o pós-estrato livre de nuvens;

$$\bar{x}_{aij} = \left( \sum_{k=1}^{n'_{aij}} x_{aijk} \right) / n'_{aij};$$

$y_{aijk}$  = área total da cultura  $c$  obtida em campo no  $k$ -ésimo segmento amostrado,  $j$ -ésimo subestrato geográfico,  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise para o pós-estrato livre de nuvens;

$\bar{y}_{aij}$  = área média da cultura c por segmento, utilizando os dados coletados em campo, no j-ésimo subestrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise para o pós-estrato livre de nuvens;

$$= \left( \sum_{k=1}^{n'_{aij}} y_{aijk} \right) / n'_{aij};$$

$b_{aij}$  = estimador de mínimos quadrados do coeficiente de regressão para a cultura c no j-ésimo subestrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise;

$$= \frac{\sum_{k=1}^{n'_{aij}} (y_{aijk} - \bar{y}_{aij})(x_{aijk} - \bar{x}_{aij})}{\sum_{k=1}^{n'_{aij}} (x_{aijk} - \bar{x}_{aij})^2}$$

O próximo passo é a combinação de todos os estimadores para toda a região livre de nuvens. Isto é feito através de:

$$\hat{Y}_c(\text{reg}) = \sum_{a=1}^A \sum_{i=1}^{s_a} \sum_{j=1}^{s'_{ai}} \bar{Y}_{aij}(\text{reg}) \quad ,$$

onde:

$s'_{ai}$  = número de subestratos geográficos livre de nuvens no i-ésimo estrato de uso do solo, a-ésimo distrito de análise;

$s_a$  = número de estratos de uso do solo no a-ésimo distrito de análise, no pós-estrato livre de nuvens;

A = número de distritos de análise na área em estudo.

A variância  $\bar{v}$  é dada por:

$$v(\hat{Y}_c(\text{reg})) = \sum_{a=1}^A \sum_{i=1}^s s_a s_{ai}' \sum_{j=1}^{N'_{aij}} N'_{aij} (N'_{aij} - n'_{aij}) s_{aij}^2 (1 - r_{aij}^2), \quad (5)$$

onde:

$$s_{aij}^2 = \frac{\sum_{k=1}^{n'_{aij}} (y_{aijk} - \bar{y}_{aij})^2}{n'_{aij} - 2}$$

e

$r_{aij}^2$  = estimador do coeficiente de correlação entre os dados de campo e os dados classificados por satélite, para a cultura  $c$ , no  $j$ -ésimo substrato geográfico,  $i$ -ésimo estrato de uso do solo,  $a$ -ésimo distrito de análise;

$$= \frac{\left( \sum_{k=1}^{n'_{aij}} (y_{aijk} - \bar{y}_{aij})(x_{aijk} - \bar{x}_{aij}) \right)^2}{\sum_{k=1}^{n'_{aij}} (y_{aijk} - \bar{y}_{aij})^2 \sum_{k=1}^{n'_{aij}} (x_{aijk} - \bar{x}_{aij})^2}$$

O passo final é unir os estimadores por regressão obtidos no pós-estrato de substratos geográficos livres de nuvens, com os estimadores por expansão direta obtidos no pós-estrato de substratos geográficos total ou parcialmente cobertos com nuvens. Se, para este conjunto de substratos, tiver sido utilizado o estimador de segmento fechado, por exemplo, então o total de cultura  $c$  neste conjunto de substratos é dado por:

$$\bar{Y}_c(\text{fec}) = \sum_{a=1}^A \sum_{i=1}^s s_a s_{ai}' \sum_{j=1}^{N''_{aij}/n''_{aij}} n''_{aij} \sum_{k=1}^{n''_{aij}} y'_{aijk},$$

onde:

$y'_{aijk}$  = área total da cultura c obtida em campo no k-ésimo segmento, j-ésimo subestrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise no pós-estrato coberto com nuvens;

$N'_{aij}$  = número total de segmentos no j-ésimo subestrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise no pós-estrato coberto de nuvens;

$n'_{aij}$  = número de segmentos amostrados no j-ésimo subestrato geográfico, i-ésimo estrato de uso do solo, a-ésimo distrito de análise no pós-estrato com cobertura de nuvens;

$s'_{ai}$  = número de subestratos geográficos no pós-estrato coberto com nuvens.

Neste caso, o total estimado para a cultura c em toda a área de estudo será:

$$\hat{Y}_c(\text{tot}) = \hat{Y}_c(\text{reg}) + \hat{Y}_c(\text{fec}),$$

$$\hat{Y}_c(\text{tot}) = \sum_{a=1}^A \sum_{i=1}^{s_a} \sum_{j=1}^{s'_{ai}} Y'_{aij}(\text{reg}) + \sum_{a=1}^A \sum_{i=1}^{s_a} \sum_{j=1}^{s'_{ai}} \frac{N'_{aij}}{n'_{aij}} \sum_{k=1}^{n'_{aij}} y'_{aijk}$$

a variância  $\hat{Y}_c(\text{tot})$  estimada é dada por:

$$v(\hat{Y}_c(\text{reg})) + v(\hat{Y}_c(\text{fec})), \quad (6)$$

onde  $v(\hat{Y}_c(\text{reg}))$  é dada em (5),

e

$$v(\hat{Y}_c(\text{fec})) = \sum_{a=1}^A \sum_{i=1}^{s_a} \sum_{j=1}^{s'_{ai}} \frac{N'_{aij}(N'_{aij}-n'_{aij})n'_{aij}}{n'_{aij}(n'_{aij}-1)} \sum_{k=1}^{n'_{aij}} (y'_{aijk} - \bar{y}'_{aij})^2$$

### AVALIAÇÃO DO ESTIMADOR DE REGRESSÃO $\hat{Y}_c(\text{tot})$

Os dados do LANDSAT foram usados como informações suplementares para melhorar a precisão da área estimada obtida somente com os dados terrestres. A eficácia do uso do LANDSAT pode ser medida através da eficiência do estimador de regressão, em relação ao estimador de expansão direta (se para todos os dados fosse utilizado somente expansão direta).

Esta eficiência relativa é definida como a razão entre as respectivas variâncias, ou seja:

$$ER = \frac{v(\hat{Z}_c)}{v(\hat{Y}_c(\text{tot}))},$$

onde  $v(\hat{Z}_c)$  é dada em (3) e  $v(\hat{Y}_c(\text{tot}))$  é dada em (6).

Este é o fator pelo qual o tamanho da amostra deve ser aumentado, de modo que o estimador por expansão direta tenha a mesma precisão que o estimador por regressão.

Por exemplo, se o número de segmentos amostrados for 435 e se o estimador de regressão com os dados do LANDSAT e dados terrestres tiver uma ER de 2,0, então ter-se-á de amostrar 870 segmentos para obter a mesma precisão do estimador por expansão direta utilizando somente os dados terrestres.

### COMENTÁRIOS FINAIS

A metodologia exposta, embora possa ser implementada na forma como ora se apresenta, poderá ser simplificada se algumas restrições forem impostas a ela. Pode-se supor, por exemplo, após as devidas considerações, que os coeficientes de regressão dos substratos geográficos dentro de cada distrito de análise e estrato de uso de solo sejam iguais. Neste caso, tem-se que seus estimadores  $b_{aij}$ ,  $j = 1, \dots, s_{ai}$  seriam iguais a  $b_{ai}$  e os estimadores combinados seriam utilizados.

Como os tamanhos de amostra dentro de cada substrato geográfico podem ser pequenos, a estimativa através dos estimadores separados (supondo que os  $b_{aij}$  são diferentes) pode ser tendenciosa. Entretanto, o uso dos estimadores combinados só deverá ser utilizado se se estiver confiante na suposição de igualdade dos coeficientes, já que o seu uso sem o atendimento desta observação produz grande variância. Somente os resul<sub>dos</sub> da implementação da metodologia numa área teste poderão servir de subsíd<sub>io</sub> para o estabelecimento do estimador mais adequado.

## REFERÊNCIAS BIBLIOGRÁFICAS

- COCHRAN, W.G. *Sampling techniques*. New York, NY, John Wiley, 1963.
- DEMING, W.E. *Sample design in business research*. New York, NY, John Wiley, 1960.
- FOOD AND AGRICULTURE ORGANIZATION (FAO) *Estimation of areas in agriculture statistics*, Rome, 1965.
- GEUDER, J. *An evaluation of SRS area sampling frame designs; ordering count units and creating paper strata*. Washington, DC, U.S. Department of Agriculture, Jul. 1983. (Statistical Reporting Service).
- GEUDER, J. *Paper stratification in SRS area sampling frames*. Washington, DC, U.S. Department of Agriculture, Fev. 1984. (Statistical Reporting Service).
- HANSEN, M.H.; HURWITZ, W.N.; MADOW, W.G. *Sample survey methods and theory*, John Wiley, 1953, v.2.
- HOLKO, M.L.; SIGMAN, R.S. The role of LANDSAT data in improving U.S. crop statistics. IN: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT, 18., Paris, France, Oct. 1984. *Proceedings*. Ann Arbor, MI, ERIM, 1984. v.6, p. 307-320.
- HOUSEMAN, E.E. *El muestreo por areas en la agricultura*. Washington, DC, U.S. Department of Agriculture, Washington, DC, Jun. 1976. (Statistical Reporting Service).
- IRONS, J.R. An overview of LANDSAT-4 and the thematic mapper, LANDSAT-4 science characterization early results. In: LANDSAT-4 SCIENCE CHARACTERIZATION EARLY RESULTS SYMPOSIUM, Greenbelt, MD. Feb. 22-24, 1983. *Proceedings*. Greenbelt, MD, NASA, 1983. v.2, p.15-46.
- KISH, L. *Survey sampling*, New York, John Wiley, 1965.
- NEALON, J.P. *Review of the multiple and area frame estimators*, Washington, DC, U.S. Department of Agriculture. (Statistical Reporting Service).



- RAJ, D. *The design of sample surveys*, New Delhi, McGraw Hill, 1972.
- WIGTON, W.H.; HANUSCHAK, G.A.; ALLEN, R.D. Integration of LANDSAT data into the crop estimation Program of USDA's Statistical Reporting Service 1972-1982. *1982 Machine Processing of Remotely Sensed Data Symposium, Purdue University, West Lafayette, IN, July 8, 1982.*
- WIGTON, W. *The production of agricultural statistics using satellite data*. s.l. FAO, 1984a.
- WIGTON, W.; PAUL, C.K. Agricultural statistics in developing countries. IN: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT, 18., Paris, France, Oct. 1984. *Proceedings*. Ann Arbor, MI, ERIM, 1984b, v.1, p. 349-354.