

ESTUDO DO MÉTODO UNIFORMIZAÇÃO DE TEMAS (UNITOT) E ANÁLISE DA CORRELAÇÃO ENTRE ÁREAS ESTIMADAS UTILIZANDO-SE DADOS DO LANDSAT E FOTOGRAFIAS AÉREAS

M.A. Moreira, S.C. Chen e A.M. Lima

Instituto de Pesquisas Espaciais

Conselho Nacional de Desenvolvimento Científico e Tecnológico

Caixa Postal 515, 12200 - São José dos Campos, SP, Brasil

RESUMO

Os objetivos deste trabalho são: (a) selecionar para uma filtragem espacial os parâmetros (limiar e peso) a serem utilizados no método de pós-processamento Uniformização de Temas - opção UNITOT, disponível no "Sistema Interativo Image-100", bem como estudar o seu efeito sobre os resultados da classificação automática da cultura do trigo, obtidos utilizando-se a combinação das abordagens não-supervisionadas (média-K) e supervisionada (MAXVER), denominada Híbrida; (b) variando-se o tamanho dos segmentos amostrais, analisar a correlação existente entre as áreas com cultura do trigo estimadas pelo "Sistema Image-100", utilizando-se informações dos 4 canais do MSS do LANDSAT, e as mesmas áreas obtidas por fotografias aéreas. Foi selecionada no município de Cruz Alta, Estado do Rio Grande do Sul, uma área de estudo de aproximadamente 800 km². Os resultados alcançados foram: a classificação obtida pela opção UNITOT (combinação limiar 2 - peso 2) foi a que apresentou maior classificação correta e menor erro de inclusão. Quando comparados os resultados obtidos pela opção UNITOT (combinação limiar 2 - peso 2) com aqueles obtidos sem o seu emprego, os valores de t calculados foram significativos aos níveis de 5% e 1% para a classificação correta e erro de inclusão, respectivamente. Todos os coeficientes de correlação calculados para os diversos tamanhos dos segmentos amostrais tiveram valores altos ($r \geq 0,96$). De acordo com este estudo, conclui-se que o uso do método UNITOT contribuiu para melhorar a precisão de mapeamento, pois apresentou uma classificação correta mais alta e um erro de inclusão mais baixo; a área estimada pelo Image-100 através de dados do LANDSAT possui alta correlação com as mesmas áreas estimadas por fotografias aéreas para os tamanhos dos segmentos estudados.

ABSTRACT

The objectives of this study are: (a) selecting the optimal threshold values for spatial filtering using the post-processing-option UNITOT of the Image-100 system and verifying their effects after wheat is classified by a hybrid classifier; (b) evaluating the correlation between wheat area estimates obtained by Image-100 system using 4 channels information of LANDSAT data and that obtained from the corresponding aerial photographs for several segments sizes. Results of the test area (approximately 800 km²) in Cruz Alta, Rio Grande do Sul, show that using spatial filtering with threshold value 2 after classification presented the highest percentage of correct classification and the lowest percentage of commission error. Comparing the classification results obtained using spatial filtering with threshold value of 2 to that without using filtering, the classification improvements were statistically significant at 5 and 1% for correct classification and commission error respectively. All the calculated correlation coefficients for various sizes of sampling segments were equal to or larger than 0.96. This study concludes that the utilization of UNITOT improved the mapping accuracy by increasing the percentage of correct classification and decreasing the percentage of commission error. The area estimates using Image-100 system and LANDSAT data correlated highly with that obtained by aerial photographs for all the segment sizes studied.

1. INTRODUÇÃO

Durante muito tempo empregou-se a classificação automática, usando-se dados digitalizados do LANDSAT, para verificar a potencialidade do "Sistema Interativo Image-100 (I-100)" de identificar e avaliar áreas agrícolas. Para isto, pequenas áreas agrícolas completamente aerofotografadas, que serviam de verdade terrestre, eram utilizadas na aquisição das estatísticas de treinamento durante a fase de classificação automática; posteriormente estas mesmas áreas serviriam, também, para avaliar o desempenho desta classificação.

Como de modo geral, as culturas ocupam grande extensão de nosso território, muitas variáveis (tais como época de plantio, tipos de solo, concentração da cultura, sistema de manejo, etc.) devem ser consideradas, quando se pensa em estender as estatísticas de treinamento adquiridas de uma pequena área para identificar e estimar culturas em grandes áreas. Neste caso, poder-se-iam inferir erros, de amplitudes bastante significativos, acarretados pela não-representatividade da resposta espectral das culturas. Por outro lado, a cobertura aerofotográfica de grandes áreas agrícolas, com a finalidade de obter todos os padrões de respostas espectrais para o treinamento do sistema Image-100, se não impossível, é tecnicamente impraticável, devido ao elevado custo e tempo requeridos para análise.

Por isto, tem-se feito pesquisas para desenvolver uma metodologia que permita estimar áreas cultivadas através de métodos de amostragens, como a regressão e/ou proporção.

Para Wigton e Bormann (1977), a amostragem é uma técnica básica para coletar dados estatísticos de agricultura para um rápido e compreensivo sistema de informações exatas, através da tomada de amostras representativas de somente uma pequena parte da área total.

O sucesso no uso do método de regressão como estimador de áreas agrícolas depende, antes de tudo, da boa correlação entre áreas estimadas no I-100 pela classificação automática e dados de verdade terrestre. Conforme afirmam Hanuschack et alii (1979), quanto maior o coeficiente de correlação, menor a variância da estimativa. Desta forma, quando γ^2 tende à unidade (1), o limite da variância tende a zero.

Vários estudos vêm sendo realizados pela equipe de Estatísticas Agrícolas (EAGRI), do Instituto de Pesquisas Espaciais (INPE), com o objetivo de testar o método de regressão para estimar áreas com a cultura do trigo.

Em trabalhos anteriores, Chen et alii (1981) compararam o desempenho da classificação do trigo usando diferentes classificadores, disponíveis no Image-100, com diversas

abordagens, na aquisição de áreas de treinamentos. Posteriormente Lima et alii (1982) fizeram estudos no sentido de avaliar dois métodos de aquisição de estatísticas de treinamento para a classificação automática, a fim de obter maior exatidão da área estimada com a cultura do trigo.

Este trabalho tem como objetivo:

- a) Estudar a viabilidade da aplicação de pós-processamento na classificação de padrões e o seu efeito sobre a precisão da classificação e do mapeamento da cultura do trigo, obtida pela combinação das abordagens supervisionada e não-supervisionada, denominada Híbrida (M-2). O método de pós-processamento escolhido foi o método Uniformização de Temas - opção UNITDT (seção 3).
- b) Variando-se o tamanho dos segmentos amostrais, analisar correlação existente entre as áreas estimadas com cultura do trigo por meio do Sistema Interativo Image-100, através dos dados do LANDSAT, e as mesmas áreas obtidas por fotografias aéreas

2. ÁREA DE ESTUDO E AQUISIÇÃO DE DADOS

Como este trabalho se compõem de mais uma etapa para alcançar um objetivo maior, que é o de testar a regressão para estimar áreas cultivadas com a cultura do trigo, a seleção da área de estudo constituiu uma importante parte para esta pesquisa. Para este estudo utilizou-se uma área de 800 km², na região de Cruz Alta, localizada a noroeste do estado do Rio Grande do Sul, entre os paralelos 28° 35' a 28° 45'S e 54° 45' a 54° 00'W.

Basicamente dois fatores contribuíram diretamente na escolha da área de estudo neste município: grande concentração da cultura do trigo nesta região e representatividade da cultura a nível estadual.

2.1 - LEVANTAMENTO AÉREO:

Em 2 de setembro de 1979, realizou-se sobre a área de estudo um sobrevôo utilizando-se câmera RC-10 e filme AEROCROME do tipo 2443 da KODAK. Foi utilizado o recobrimento de 30% tanto no sentido lateral quanto no longitudinal. O produto obtido na escala aproximada de 1:20.000 foi plotado em mapa fotográfico na escala de 1:100.000, para verificar se houve falhas no recobrimento aéreo.

Foi feita a fotointerpretação empregando-se o método convencional de interpretação visual, segundo uma legenda pré-determinada. O resultado obtido pela fotointerpretação serviu de verdade terrestre para a interpretação automática dos dados digitais do LANDSAT.

2.2 - AQUISIÇÃO DOS DADOS DO LANDSAT

A época ideal para a aquisição dos dados do LANDSAT é final de setembro ou início de outubro para aquela região porque, nesta época, o trigo encontra-se em estado de maturação, apresentando uma coloração amarelo ouro, diferente das outras culturas (predominantemente pastagem) que ainda permanecem verdes, conforme descrevem Chen et alii (1981). Entretanto, como os dados do LANDSAT do dia 22 de setembro apresentaram 100% de cobertura de nuvens, tornou-se impossível a sua utilização. Para este estudo foram utilizados dados do LANDSAT gravados em fitas compatíveis com o computador (CCTs), referentes à órbita 220, ponto 32 da passagem do dia 4 de setembro de 1979. Nesta época, a maioria das áreas com cultura de trigo encontrava-se em estágio de emborrachamento.

3. METODOLOGIA

Para verificar a viabilidade do método UNITOT, bem como estudar os efeitos do tamanho do segmento amostral sobre a correlação entre as áreas estimadas através dos dados do LANDSAT e de fotografias aéreas, foram utilizados os resultados da classificação obtidos pelo método híbrido (M-2).

3.1 - MÉTODO HÍBRIDO (M-2)

Este método utiliza, inicialmente, um algoritmo não-supervisionado média K ("clustering"), cuja amostragem é feita sistematicamente, obtendo-se assim as classes espectrais. Posteriormente, áreas de treinamentos que contêm "pixels" puros foram adquiridas com base nos resultados do "clustering" para obtenção do pacote estatístico que foi usado pelo classificador MAXVER. (Lima et alii, 1982).

3.2 - UNIFORMIZAÇÃO DE TEMAS - OPÇÃO UNITOT

O UNITOT é um dos métodos de pós-processamento que atua no sentido de melhorar os resultados da classificação automática, minimizando os erros de omissão e inclusão; consequentemente, aumenta a exatidão da classificação automática.

O método consiste em uma filtragem espacial, que analisa o "pixel" central de uma matriz 3 x 3 "pixels". Para a análise dois parâmetros devem ser considerados: limiar (L) e peso (P). Estes são pre-determinados pelo analista.

Após determinar a frequência de ocorrência de cada classe no interior da matriz, em que a classe do "pixel" central é contada P vezes, o valor da frequência da classe de maior ocorrência é comparada com o limiar L. Se o valor do limiar for menor, o "pixel" central será substituído pela classe de maior ocorrência dentro da matriz. Se o valor do limiar L for maior, a classe de "pixel" central permanece sem alteração (Dutra, 1982).

Durante a classificação automática de culturas pode existir uma certa grandeza de erro de inclusão causado pela semelhança de respostas espectrais de diferentes alvos. Este tipo de erro pode ser minimizado utilizando-se um pós-processamento (no caso o UNITOT) que se baseia numa filtragem espacial.

Além de estudar os efeitos do uso do método UNITOT sobre os resultados da classificação automática, procurou-se também determinar a melhor combinação limiar-peso. Utilizaram-se as seguintes combinações: (1, 2); (2, 1); (2, 2); (2, 3); (3, 2); (3, 4) e (3, 5).

3.3 - PROCEDIMENTO

Uma vez delimitada a área de estudo, com o auxílio do programa traço, procedeu-se à sua classificação, utilizando-se para tal o método híbrido (M-2). Os resultados desta classificação foram impressos em forma de mapa alfanumérico "printout" na escala de 1:20.000, o qual contém a distribuição espacial do trigo classificado. Pela sobreposição ao "printout" de um mosaico confeccionado a partir da interpretação visual das fotos aéreas, fez-se a delimitação dos talhões de trigo. Este mapa, obtido sem a utilização do método UNITOT, foi denominado *mapa base*.

Uma vez classificada a área de estudo, usando-se o método híbrido (M-2), aplicou-se o método UNITOT aos resultados dessa classificação, variando-se a combinação limiar-peso. Os pontos classificados como trigo, para cada combinação, foram impressos em forma de "printout", na escala de 1:20.000. Obtidos os "printouts", fez-se a comparação de cada um pela sobreposição deles ao mapa base. Esta comparação foi feita visualmente, sobre mesa de luz, observando-se os erros de omissão e inclusão. Com base nesta comparação, fez-se a seleção da melhor combinação limiar-peso.

Para comparar quantitativamente o efeito da melhor combinação sobre os resultados da classificação, fez-se um estudo estatístico, selecionando-se - tanto no "printout" adquirido pela melhor combinação quanto no mapa base

se - cinco áreas testes ($\cong 40 \text{ km}^2$ cada) com diferentes densidades de trigo. Para este estudo, aplicou-se o teste t aos dados da classificação correta* (CC), dos erros de inclusão** (EI) e das áreas estimadas.

Utilizando-se a fórmula abaixo, foram determinadas as diferenças relativas (DR) das áreas estimadas das cinco áreas testes e para toda área de estudo:

$$DR\% = \left(\frac{X}{Y} - 1 \right) \times 100,$$

onde:

X → áreas estimadas pelo I-100;

Y → áreas estimadas por fotografias aéreas.

3.3 - ESTUDO DA CORRELAÇÃO

Nesta etapa, procurou-se verificar a influência do tamanho do segmento amostral sobre a correlação existente entre as áreas de trigo estimadas pelo I-100 e as mesmas áreas obtidas por fotografias aéreas. Para este estudo utilizaram-se os dados do "printout" adquirido na etapa anterior, usando-se o método híbrido (M-2) e o programa UNITOT com a melhor combinação limiar-peso.

Uma vez delimitadas sobre o printout as áreas de trigo, este foi dividido em 72 unidades básicas ($\cong 3,16 \text{ km} \times 3,16 \text{ km}$ cada). Utilizaram-se para este estudo segmentos com áreas aproximadas de 10, 20, 30, 40 e 60 km^2 . Para áreas acima de 10 km^2 os segmentos foram formados pela soma de 2, 3, 4 e 6 unidades básicas, a saber:

- a) Para os segmentos com áreas de 20 e 30 km^2 , tomaram-se unidades adjacentes nos sentidos horizontal, vertical e diagonal da área de estudo;
- b) para segmentos com área de 40 km^2 , tomaram-se unidades adjacentes nos sentidos horizontal, vertical e áreas de 2 x 2 unidades básicas;
- c) para os segmentos com área de 60 km^2 , usaram-se 3 x 2 ou 2 x 3 unidades básicas.

A Figura 1 mostra esquematicamente os sentidos em que foram tomadas as unidades na formação dos segmentos.

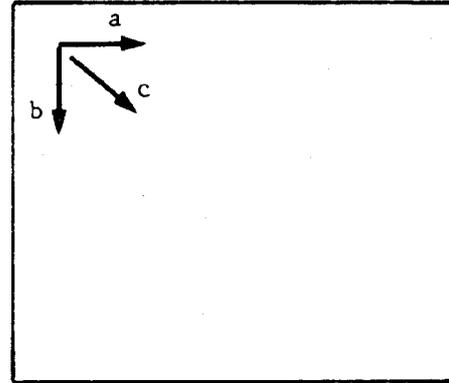


Fig. 1 - Esquema demonstrativo da formação dos segmentos.

- a - horizontal
- b - vertical
- c - diagonal

Devido à restrição do tempo de computação, as áreas estimadas pelo I-100 e por fotografias aéreas, que são os dados necessários para o cálculo dos coeficientes de correlação, foram adquiridas da seguinte maneira:

- 1) Sobre o "printout" obtido, fez-se a contagem manual dos pontos classificados corretamente como trigo (a); não-trigo classificado como trigo (b); os pontos de trigo que não foram classificados (c).
- 2) As áreas estimadas pelo I-100, em cada segmento, foram adquiridas somando-se a e b.
- 3) As áreas estimadas por fotografias aéreas, em cada segmento, foram obtidas somando-se a e c.
- 4) Desta forma, várias tabelas foram geradas de acordo com o tamanho do segmento e com o sentido considerado. Estes dados foram submetidos à análise de correlação.

4. RESULTADOS

A Figura 2 mostra os resultados da classificação automática obtida pelo método híbrido (M-2), sem a Uniformização de Temas. Pela observação da figura nota-se que, muito embora o método híbrido (M-2) tenha mostrado um melhor desempenho da classificação entres os métodos estudados (Lima et alii, 1982), houve uma certa confusão entre a classe trigo e as classes pastagem e vegetação arbórea esparsa (confusão de menor intensidade). Isto se deve ao fato de que, na época da tomada dos dados

* CC = $\frac{\text{área de trigo classificada corretamente}}{\text{área de trigo no campo}}$

** EI = $\frac{\text{área de não-trigo classificada como trigo}}{\text{área de trigo estimada pelo I-100}}$

do LANDSAT, grande parte da cultura do trigo estava no estágio de emborrachamento. Conseqüentemente, não se pôde obter uma perfeita separabilidade entre as respostas espectrais do trigo e da pastagem, pois ambas ainda permaneciam verdes. Para o caso da vegetação arbórea, notou-se que, na maioria das vezes onde havia confusão com o trigo, a vegetação era muito esparsa, com predominância de um estrato inferior graminóide, que provavelmente causou uma semelhança de respostas espectrais.

A comparação qualitativa entre os "printouts" de cada combinação com o mapa base mostrou que o melhor desempenho do UNITOT foi obtido pelo uso do limiar 2 - peso 2. Estes resultados são ilustrados nas Figuras 2 e 3.

Com relação à análise quantitativa, esta foi efetuada a partir dos dados das Tabelas 1 e 2, obtidos pelo emprego ou não do método UNITOT para as 5 áreas testes. O valor de t calculado foi significativo aos níveis de 5% e 1% para a comparação de classificação correta e erro de inclusão, respectivamente. Desta forma, conclui-se que o uso do método UNITOT melhorou o desempenho da classificação, pois aumentou a classificação correta e diminuiu o erro de inclusão. Quando comparadas as áreas estimadas pelo I-100, empregando-se ou não o método UNITOT, com as mesmas áreas das fotografias aéreas, os valores de t calculados não foram significativos ao nível de 5%. No entanto, verificou-se que o valor de t calculado para as áreas estimadas, sem a utilização do método UNITOT e de fotografias aéreas, foi significativo ao nível de 15%, enquanto usando-se as áreas estimadas através da Uniformização de Temas e de fotografias aéreas, o valor de t calculado foi significativo somente ao nível de 22% de probabilidade. Por outro lado, com a aplicação do método UNITOT, a raiz média quadrática dos erros, diminuiu de 126,18 para 107,02 ha e a diferença relativa mostrou uma redução da área superestimada de 4,5 para 2,7% para as cinco áreas testes; para toda área de estudo a superestimativa passou de 10,9 para 9,7%. Isto leva a crer que, com a utilização do método UNITOT, a área estimada aproximou-se melhor daquela obtida por fotografias aéreas.

Finalmente, com base nos testes realizados, os autores concluíram que a aplicação do método UNITOT oferece, pelo menos, uma estimativa de área tão precisa quanto aquelas obtidas sem a sua utilização. Porém, pelo baixo erro de inclusão e maior classificação correta, este método aumentou a precisão de mapeamento.

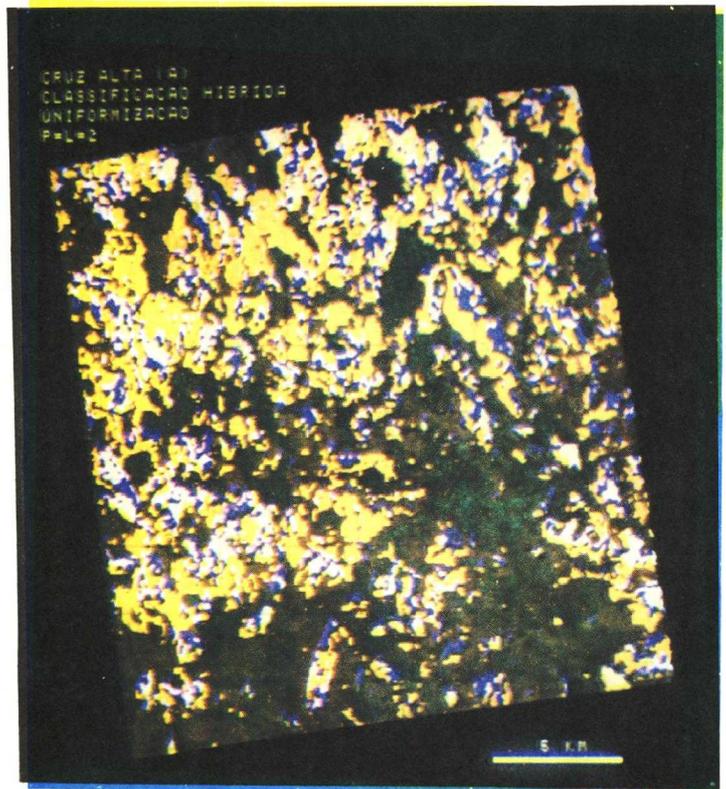


Fig. 2 - Resultados da classificação automática obtidos pelo método híbrido (M-2), sem uniformização de temas.

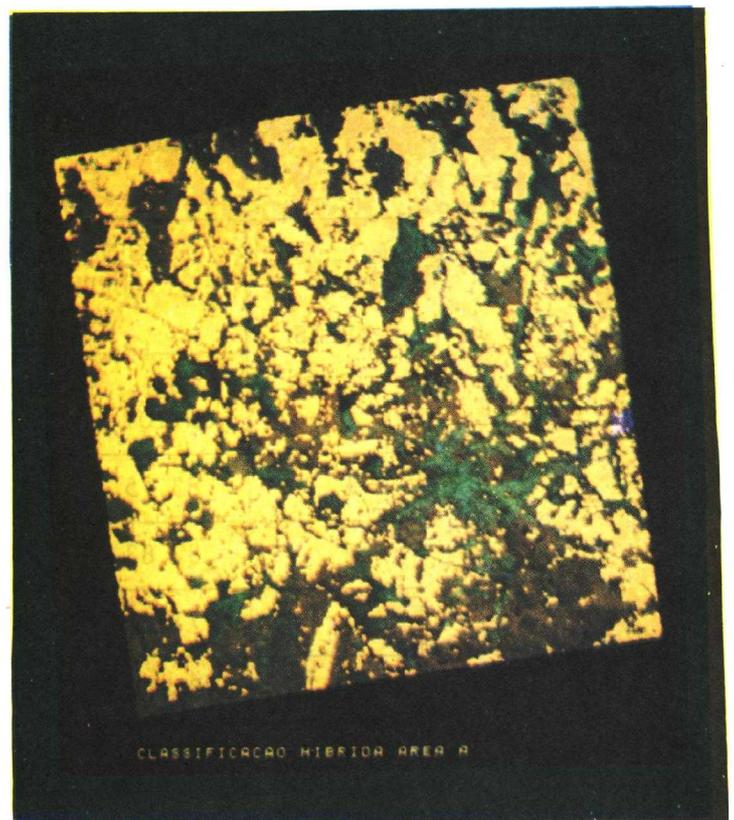


Fig. 3 - Resultados da classificação automática obtidos pelo método híbrido (M-2) com UNITOT (combinação limiar 2 - peso 2).

TABELA 1

TRIGO ESTIMADO PARA AS ÁREAS TESTES

NÚMERO DE ÁREAS TESTES	DENSIDADE DE TRIGO (%)	ÁREA ESTIMADA (ha)		
		FOTO AÉREA	I-100	
			SEM UNITOT	COM UNITOT
1	56,69	2331,77	2290,26	2286,98
2	46,13	1897,60	1906,26	1852,51
3	39,28	1615,72	1624,08	1600,79
4	22,87	971,05	1199,77	1137,96
5	31,15	1322,50	1481,95	1481,06
SOMA	-	8138,64	8502,32	8359,30
DR	-	-	4,5%	2,7%
R.M.Q.E.	-	-	126,18	107,02

TABELA 2

DADOS DE CLASSIFICAÇÃO CORRETA E ERRO DE
INCLUSÃO PARA AS ÁREAS TESTES

NÚMERO DE ÁREAS TESTES	CLASSIFICAÇÃO CORRETA (CC%)		ERRO DE INCLUSÃO (EI%)	
	SEM UNITOT	COM UNITOT	SEM UNITOT	COM UNITOT
1	83,75	86,05	14,73	12,26
2	82,83	83,93	17,54	14,67
3	77,51	78,93	22,89	20,33
4	88,41	89,08	28,45	23,98
5	87,85	88,30	21,60	21,15

Para estudar a influência do tamanho dos segmentos sobre a correlação existente entre as áreas estimadas pelo I-100, utilizando-se dados do LANDSAT, e as mesmas áreas obtidas por fotografias aéreas, tomaram-se como base os dados obtidos do "printout" usando-se o método híbrido (M-2) e aplicando-se a Uniformização de Temas (combinação limiar 2 - peso 2). A Tabela 3 mostra os coeficientes de correlação para os segmentos estudados.

De acordo com os resultados obtidos, chegou-se à conclusão de que, independente do tamanho dos segmentos e dos sentidos nos quais foram tomados, todos os coeficientes de correlação tiveram valores acima de 0,96, o que mostra que, para os tamanhos estudados, há uma alta correlação entre a área estimada pelo I-100, usando-se dados do LANDSAT, e a área estimada pelas fotografias aéreas.

TABELA 3

VALORES DOS COEFICIENTES DE CORRELAÇÃO ENTRE ÁREAS ESTIMADAS COM TRIGO USANDO-SE DADOS DO LANDSAT E FOTOGRAFIAS AÉREAS

TAMANHO DOS SEGMENTOS (km ²)	NÚMERO DE PARES DE VALORES (n)	COEFICIENTES DE CORRELAÇÃO		
		SENTIDO		
		HORIZONTAL	VERTICAL	DIAGONAL
60	12	0,967	0,982	-
40	18	0,971	0,977	0,98*
30	24	0,973	0,977	0,963
20	36	0,967	0,960	0,960
10	n = 72	r = 0,98		

* Este coeficiente foi calculado utilizando-se segmentos de 2 x 2 unidades básicas.

Uma análise de regressão foi realizada para segmentos com área de 10 km², na qual se observou que a interseção da reta com o eixo do y foi estatisticamente diferente de zero (Figura 4). Neste caso, conforme recomenda Cochran (1963), a estimativa pelo método de

Regressão é mais adequada do que a estimativa obtida pelo método de Razão.

Com base nos resultados desta etapa, será testado um sistema amostral utilizando-se a estimativa de regressão para a cultura de trigo.

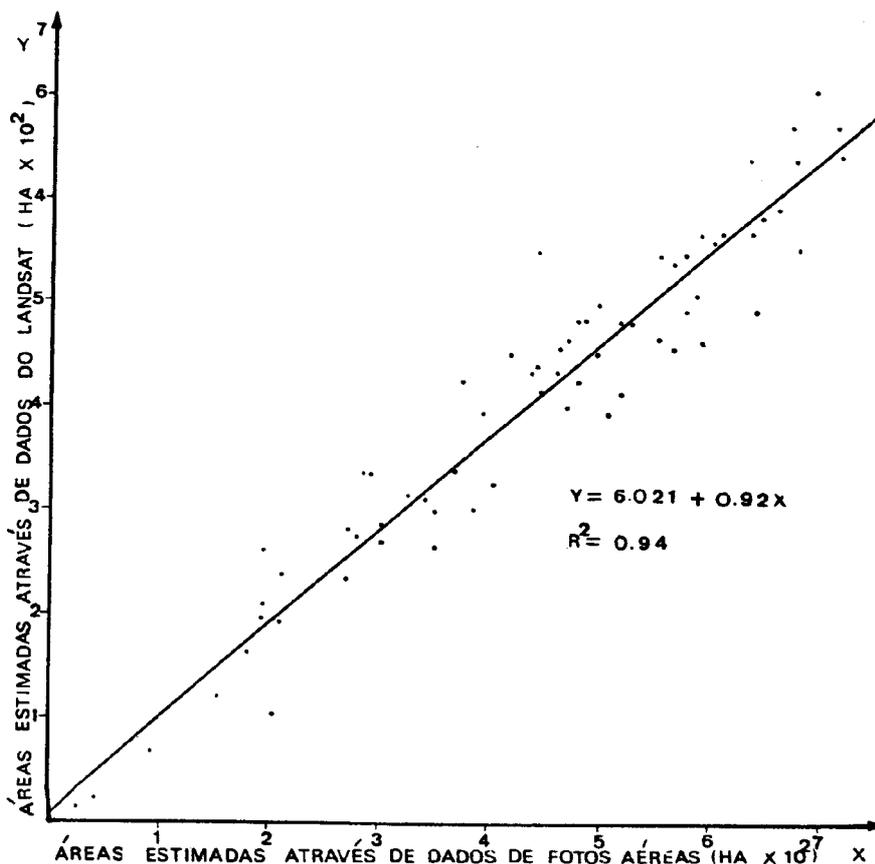


Fig. 4 - Regressão para áreas estimada com trigo obtidas através de dados do LANDSAT e de dados de aeronave para segmentos de 10 km².

5. CONCLUSÕES

Através deste estudo, pôde-se chegar às seguintes conclusões:

- a) Para diminuir os erros de inclusão da classificação automática, usando-se dos dados unitemporais do LANDSAT, a aplicação do método UNITOT melhora a precisão de mapeamento.
- b) De acordo com os resultados obtidos, todos os coeficientes de correlação tiveram valores acima de 0,96, independente do tamanho dos segmentos e dos sentidos em que foram tomados, o que mostra, uma alta correlação entre a área estimada pelo I-100 através de dados do LANDSAT e as mesmas áreas estimadas pelas fotografias aéreas.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- CHEN, S.C.; MOREIRA, M.A.; LIMA, A.M. de *Comparison of wheat classification accuracy using different classifiers of the Image-100 system.* São José dos Campos, INPE, June, 1981. (INPE-2125-RPE/349).
- COCHRAN, W.C. *Técnicas de amostragem.* Rio de Janeiro, Fundo de Cultura, 1965.
- DUTRA, L.V. *Desenvolvimento de alguns métodos de pós-processamento aplicados a imagens digitais multiespectrais.* (Técnicas de uniformização de temas). São José dos Campos, INPE, 1982. (No prelo).
- HANUSCHAK, G.A.; SIGMAN, R.; GRAIG, M.E.; OZGH, M.; LUEBBE, R.C.; COOK, P.W.; KLEWEND, D.D.; MILLER, C.E. *Crop-area estimates from LANDSAT; transition from research and development to timely results.* In: Annual Symposium Machine Processing of Remotely Sensed Data; 5., W. Lafayette, IN, June 27-29, 1979. *Proceedings.* New York, IEEE, c 1979, p. 86-96.
- LIMA, A.M. de; CHEN, S.C.; MOREIRA, M.A. *Avaliação de dois métodos de aquisição de estatísticas de treinamento para classificação automática.* São José dos Campos, INPE, 1982. (No prelo).
- WIGTON, W.; BORMANN, P. *A guide to area sampling from construction utilizing satellite imagery.* Paper presented at the Second International Training Course in Remote Sensing Applications for Agriculture; Crop Statistics and Agricultural Census, Rome 25 April - 13 May 1977.