

1. Classificação <i>INPE-COM.4/RPE</i>		2. Período	4. Distribuição
3. Palavras Chaves (selecionadas pelo autor) <i>PRECISÃO DE CLASSIFICAÇÃO CULTURAS IMAGENS LANDSAT</i>			interna <input type="checkbox"/> externa <input checked="" type="checkbox"/>
5. Relatório nº <i>INPE-1889-RPE/232</i>	6. Data <i>setembro, 1980</i>	7. Revisado por <i>Armando P. dos Santos</i>	
8. Título e Sub-Título <i>PRECISÃO DE CLASSIFICAÇÃO DO MAPA DE DISTRIBUIÇÃO DE CULTURAS, OBTIDO A PARTIR DE IMAGENS LANDSAT</i>		9. Autorizado por <i>Nelson de Jesus Parada Diretor</i>	
10. Setor <i>DSR/DDP</i>	Código <i>30.241.000</i>	11. Nº de cópias <i>03</i>	
12. Autoria <i>Sherry Chou Chen</i>		14. Nº de páginas <i>07</i>	
13. Assinatura Responsável <i>Sherry Chen</i>		15. Preço	
16. Sumário/Notas <p><i>A precisão de classificação do mapeamento de distribuição de culturas, obtido através de imagens LANDSAT, deve ser demonstrado para alcançar uma aceitação mais ampla entre os usuários. Neste trabalho é apresentada uma abordagem estatística para se verificar a precisão de classificação do mapa de distribuição de culturas. É apresentado, também, um método para avaliar a precisão do mapeamento de cana-de-açúcar do Estado de São Paulo, obtido através de imagens LANDSAT.</i></p>			
17. Observações			

ÍNDICE

ABSTRACT	<i>iv</i>
1. INTRODUÇÃO	1
2. ABORDAGEM ESTATÍSTICA	1
3. METODOLOGIA PARA VERIFICAR A PRECISÃO DE CLASSIFICAÇÃO DO MAPA DE DISTRIBUIÇÃO DE CANA-DE-AÇÚCAR	5
4. SELEÇÃO DE AMOSTRAS E CRITÉRIO PARA VERIFICAÇÃO DE CAMPO	6
BIBLIOGRAFIA	7

ABSTRACT

For a wider acceptance among users of the crop distribution map derived from LANDSAT imagery, the classification accuracy has to be addressed. In this paper a statistical sampling design is presented for verifying the classification accuracy of a crop distribution map. Also a layout to evaluate the accuracy of the sugar-cane map of São Paulo State derived from LANDSAT imagery is described.

1. INTRODUÇÃO

Desde o lançamento do satélite LANDSAT-1 (ERTS-1) em 1972, tem aumentado o interesse para mapeamento de áreas agrícolas, não só para fins de estimativa de área plantada, mas também para o planejamento agrícola.

Entretanto, para uma aceitação mais ampla entre os usuários, o problema da precisão do mapeamento deve ser demonstrado. Geralmente, o mapa de distribuição de culturas envolve uma grande área. Nestas condições, uma precisão de classificação exata só pode ser obtida se for feito um trabalho de campo exaustivo sobre toda a área. Devido a razões práticas (tempo e custo), um procedimento de amostragem é necessário para estimar a precisão de classificação.

Este trabalho apresenta uma abordagem estatística para verificar a precisão de classificação do mapa de distribuição de culturas. Apresenta, também, um método que será utilizado para verificar a precisão de classificação do mapa de cana-de-açúcar do Estado de São Paulo.

2. ABORDAGEM ESTATÍSTICA

A área mapeada pode ser entendida como uma população de pontos infinitos. Cada ponto tem sua classe (atributo) verdadeira e pode ser correta ou incorretamente classificado pelo fotointérprete. A probabilidade (proporção) dos pontos classificados corretamente (P), no mapa, é o que se deseja conhecer. O valor de P somente pode ser obtido através de uma enumeração completa; isto é impraticável, pois consome muito tempo. Entretanto, usando-se um procedimento de amostragem, o valor de P pode ser estimado.

Para quantificar a precisão de classificação, várias amostras são selecionadas aleatoriamente do mapa; suas classes verdadeiras verificadas e designado o valor 1 para amostras classificadas corretamente, e valor zero para aquelas classificadas incorretamente. A média da

amostra \bar{e} :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

onde:

$X = \bar{e}$ uma variável dicotômica (1 ou 0);

$n = \bar{e}$ o número de amostras; e

$\bar{X} = \bar{e}$ a probabilidade de um ponto amostrado ser classificado corretamente, e \bar{e} o mesmo que o estimador p em uma distribuição binomial.

Padronizando-se p , através da transformação Z , tem-se:

$$Z = \frac{p - P}{\sigma_p} = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} \dots\dots (1)$$

O intervalo de confiança de 95% de Z pode ser calculado, usando-se a tabela de distribuição normal:

$$\text{Pr} (-1,96 < Z < 1,96) = 0,95 \dots\dots (2)$$

Substituindo-se Z pela equação (1), o intervalo de confiança de 95% para P paramétrico pode ser estimado aproximadamente por:

$$\text{Pr} \left(-1,96 < \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} < 1,96 \right) = 0,95$$

$$\left(\frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} \right)^2 < (1,96)^2$$

portanto:

$$n(p^2 - 2pP + P^2) < P(1-P) (1,96)^2$$

o intervalo de confiança será dado por:

$$\frac{1}{1+(1,96)^2/n} \left[p + \frac{(1,96)^2}{2n} \pm 1,96 \sqrt{\frac{p(1-p)}{n} + \frac{(1,96)^2}{4n^2}} \right] \dots (3)$$

A equação (3) é usada para calcular os limites de confiança, entre os quais P está incluído com probabilidade de 0,95 (Chou,1963).

Por exemplo, 50 pontos de amostra foram tirados aleatoriamente do mapa de distribuição de cana-de-açúcar, entre os quais 40 foram classificados corretamente. Então $p = 40/50 = 0,80$ e o intervalo de confiança para P com 0,95 de possibilidade é:

$$50 [0,8^2 - 2(0,8)P + P^2] < 3,8416 P - 3,8416 P^2$$

portanto:

$$0,6896 < P < 0,8876$$

Isto significa que a precisão de classificação de cana-de-açúcar deste mapa é maior que 68,69% e menor que 88,76%, com um risco de erro tipo I (α) de 5%.

Entretanto, para usar este esquema de amostragem, é necessário decidir um número mínimo de amostras (n), tal que ele não dispenda mais tempo e, conseqüentemente, maior custo para a amostragem que o necessário, e também com erros tipo I (α) e II (β) aceitáveis.

Normalmente, é usado nível de 0,05 para os dois tipos de erro. Isto é, tem-se o risco de 5% de rejeitar um mapa com alta precisão (95%*) e aceitar um mapa com baixa precisão (85%*) . Nestas condições, o tamanho ótimo da amostra é 93 (detalhes deste cálculo podem ser encontrados em Ginevan, 1979). Explicando melhor, um mapa impreciso é aquele com uma proporção de classificação correta igual ou menor que 0,85 e com um risco de aceitá-lo de 0,05. Também, por outro lado, se a precisão do mapa é 0,95, o risco de rejeitá-lo é de 0,05. Para verificar-se a precisão do mapa com estas condições, 93 amostras devem ser checadas em campo. O mapa será considerado preciso se não mais que 8 amostras foram classificadas incorretamente.

Para testar a precisão do mapa de uso da terra é usada amostragem aleatória estratificada. O número de amostras, para cada categoria (estrato), é proporcional à sua área, medida pelo planímetro. Entretanto, este procedimento de amostragem é mais difícil de ser aplicado para um mapa de distribuição de culturas, pois a medida pelo planímetro de pequenos campos de cultura é inconveniente. Por isso, a amostragem aleatória simples é usada para testar a precisão de classificação do mapa de distribuição de cultura e poderia ser usado $n = 93$, como uma referência inicial para o procedimento de estimativa. Porém, o número de amostras deve ser decidido, objetivamente, pelo pesquisador, considerando tempo e disponibilidade de recursos humanos e financeiros.

Baseado no número de amostras e na proporção de classificação correta, o intervalo de confiança de 95% pode ser calculado, usando-se a equação (3). Para os cálculos de α e β , a função de densidade de probabilidade binomial é empregada. Com este intervalo de confiança para P paramétrico, o nível de α e β indicará se o mapa é aceito como preciso.

* escolhido objetivamente

3. METODOLOGIA PARA VERIFICAR A PRECISÃO DE CLASSIFICAÇÃO DO MAPA DE DISTRIBUIÇÃO DE CANA-DE-AÇÚCAR.

Utilizando-se imagens LANDSAT multitemporais, na escala 1:250.000 foi obtido um mapa da distribuição espacial da cana-de-açúcar, no Estado de São Paulo, em maio de 1980. Para testar a precisão de classificação, um rol de amostras deveria ser definido por uma grade, cujo tamanho não deveria ser nem tão grande nem tão pequeno.

Uma grade muito grande perderá os talhões pequenos, os quais são mais difíceis de serem interpretados, e, conseqüentemente, haverá uma previsão superestimada. Uma grade muito pequena cria muitas interseções, causando um gasto maior de tempo na amostragem. De acordo com o censo de 1970 (IBGE, 1970), mais de 50% das propriedades de cana-de-açúcar do Estado de São Paulo são maiores que 5 Km². Por isso, uma grade de 1 x 1 cm, correspondente a 6,25 Km², é colocada sobre o mapa para a amostragem.

De acordo com os dados históricos, a área plantada com cana-de-açúcar está concentrada nas DIRAs de Campinas, Ribeirão Preto, parte de Bauru, São José do Rio Preto e Sorocaba. Esta área com concentração de cana contribui, aproximadamente, com 85% da área total plantada, no Estado de São Paulo.

Para uma estimativa significativa da precisão de classificação do mapa de distribuição de cana-de-açúcar, um peso maior deve ser dado para as áreas de concentração de cana. Desta forma, 200 amostras aleatórias serão extraídas da área de concentração de cana, e 20 para o resto da área mapeada (amostragem estratificada não proporcional). O número 20 é o tamanho mínimo de amostra para se verificar se a precisão de classificação é 85% (Gerender, et al., 1977).

Os 220 pontos de amostras serão, então, verificados no campo ou por fotografias aéreas da mesma época do mapeamento, se as fotos estiverem disponíveis e se for possível uma identificação precisa de cana-de-açúcar.

4. SELEÇÃO DE AMOSTRAS E CRITÉRIO PARA VERIFICAÇÃO DE CAMPO

Amostragem aleatória pode ser executada usando-se um sistema de coordenadas espaciais aleatório de uma grade de 1 X 1 cm, obtendo-se 135 colunas e 127 linhas para a área de concentração de cana-de-açúcar. Três dígitos podem ser extraídos de números aleatórios.

O primeiro número menor que 945 será dividido por 135 e o resíduo usado como coluna selecionada; o segundo número deve ser menor que 889 que será dividido por 127 e o resíduo usado como linha selecionada. Assim, o primeiro ponto pode ser localizado. Números aleatórios maiores que 945 são abandonados para a seleção de colunas e maiores que 889 para a seleção de linhas, para garantir que cada coordenada tenha a mesma probabilidade de ser selecionada. Este método é repetido até que 200 pontos sejam localizados, na área de concentração de cana. O mesmo método é aplicado para localizar os 20 pontos na área com baixa concentração de cana-de-açúcar.

As interseções da grade amostradas são pontuais sobre o mapa. O critério do tamanho de área para ser verificado em campo, para cada ponto, merece consideração especial. Num mapa de escala 1:250.000, cada ponto de amostra, assinalado por uma caneta com pena Marsh 0,2 mm de diâmetro, corresponde a uma área de 50 metros de diâmetro. Os pontos amostrados podem ser localizados no campo usando-se rios, estradas, etc., como pontos auxiliares. A fisionomia ao redor do ponto central (50 metros de diâmetro) deverá então, ser observada. Baseado na informação de campo, a precisão de classificação pode ser quantificada, isto é, deve-se dar valor 1 para os pontos classificados corretamente, e valor zero para aqueles classificados incorretamente. Os dados coletados serão analisados, usando-se o método descrito no item 2.

BIBLIOGRAFIA

- CHOU, Y.L. *Statistical analysis with business and economic applications*. New York, Holt, Rinehart and Winston, 1969.
- GENDEREN, J.L. van; LOCK, B.F. Testing land-use map accuracy. *Photogrammetric Engineering and Remote Sensing*, 43(9):1135-1137, Sept. 1977.
- GINEVAN, M.E. Testing land-use map accuracy: another look. *Photogrammetric Engineering and Remote Sensing*, 45(10):1371-1377, Oct. 1979.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Censo Agropecuário São Paulo*, 1970. Rio de Janeiro, 1970. V.3, T.18, P.1.