

# Efeitos da Estratégia de Amostragem na Exatidão Reportada pelo Índice Kappa na Classificação de Imagens Orbitais

Ricardo Seixas Brites<sup>1</sup>  
Vicente Paulo Soares<sup>1</sup>  
Carlos Antonio Álvares Soares Ribeiro<sup>1</sup>

<sup>1</sup>Universidade Federal de Viçosa / Departamento de Engenharia Florestal  
Av. P. H. Rolfs s/n, 36571.000 Viçosa, Minas Gerais, Brasil  
{brites, vicente}@brufv.bitnet.br

**Abstract.** This paper describes the effects of using different sampling approaches on the generation of a reference file to access the accuracy of maps generated by classifications of orbital imagery. The results show that the blocked sampling largely overestimates the Kappa index when compared with non-blocked procedures.

**Keywords:** Remote Sensing, Sampling, Accuracy, Kappa

## 1. Introdução

De acordo com o que é discutido por EBERT e LYONS (1983), amostragem é um processo cuidadoso de inspeção de uma pequena parte representativa de uma área no que se refere às características ou parâmetros nos quais se está interessado e que podem, então, serem projetados para uma área maior. De maneira semelhante KELLY (1970) define amostragem como sendo a seleção de uma parte (amostra) de um agregado que representa todo o agregado (população). De acordo com os conceitos acima é possível se perceber a importância da representatividade da amostra. Se de antemão é sabido ser homogênea a população a ser amostrada, o processo de amostragem não necessita de grandes sofisticacões vêz que qualquer amostra selecionada representará fidedignamente a população. Por outro lado, situação normalmente encontrada em sensoriamento remoto, se a variabilidade da população é alta este processo reveste-se de toda uma série de técnicas e precauções para que as amostras a serem coletadas possam representar o universo da população com o mínimo possível de erro.

A coleta dos dados de referência, ou seja aqueles que representam a situação real de campo à época da obtenção da imagem, é uma parte essencial de qualquer projeto de classificação e mapeamento envolvendo dados obtidos por meio de sensoriamento remoto. Esses dados são usados para verificar a exatidão da classificação, para descobrir confusão entre classes e como entrada no processo de refinamento da classificação (CONGALTON e BIGING, 1992).

A questão relativa à forma como são realizadas estas amostragens sobre a imagem orbital, vem de há muito sendo razão de discussão entre os especialistas da área de sensoriamento remoto. O que se faz corriqueiramente é coletar estas amostras em regiões da imagem onde se tem certeza de que a feição que ali se encontra corresponde àquela que se tem em mente. Este tipo de abordagem, largamente utilizada, peca pela falta de aleatoriedade do procedimento. Este método viola o requisito de independência entre as amostras e, em consequência, faz com que os dados coletados para representar as diversas classes sejam menos representativos (GONG e HOWARTH, 1990).

Como expresso por WARREN et al. (1990), muito pouco tem sido escrito no que se refere à representatividade das amostras de campo em relação à interpretação das imagens orbitais. A extração casual de amostras no campo é um importante fator na confiabilidade estatística dos dados resultantes da interpretação. Todavia, em virtude de os diferentes tipos de cobertura de solo não ocorrerem de forma casual na natureza, uma simples amostragem casualizada pode não representar adequadamente todas as categorias de cobertura de solo espectralmente discerníveis. A estratificação da amostragem pode melhorar as chances de que todas as categorias sejam adequadamente representadas. De forma semelhante, CARD (1980), citado por CONGALTON (1984), comenta que não vêm sendo feitos estudos comparativos entre amostragens sistemáticas em comparação com amostragens casualizadas. Na verdade, em virtude dos altos custos envolvidos nos processos de coleta de dados e do tempo a ser dispendido neste tipo de tarefa,

não tem sido feitas comparações diretas entre quaisquer esquemas de amostragem usadas na verificação da exatidão dos mapas derivados dos dados obtidos por sensores remotos (CONGALTON, 1992).

O que se depreende do parágrafo anterior pode ser extrapolado eliminando-se, por razões de custo e outras, as visitas a campo e, ao invés delas, utilizar-se de recobrimentos aerofotogramétricos desde que os mesmos tenham sido realizados em escala compatível com o objetivo do trabalho e em época suficientemente próxima àquela da obtenção da imagem orbital a ser utilizada, em concordância com o que afirmam CONGALTON e BIGING (1992). De acordo com eles, os dados de referência são frequentemente coletados a partir de fotointerpretação de fotografias aéreas, em grande escala ou sobrevoando-se a área em um helicóptero ou outra aeronave, tomando-se notas durante o percurso.

### 1.1 Os padrões de amostragem

Entenda-se por padrão de amostragem, a maneira, em termos espaciais, em que as amostras serão tomadas sobre uma dada área. Em sensoriamento remoto, estes padrões são de fundamental importância, uma vez que a depender deles poder-se-á incorrer na incorporação de amostras com alta correlação espacial.

Como afirmado por CAMPBELL (1987), a correlação espacial é a tendência das medidas em um dado local geográfico se assemelharem àquelas tomadas em outros locais que se encontram na vizinhança imediata do primeiro. Este conceito é importante em amostragem porque, idealmente, é desejável que cada observação forneça informações independentes. Se as observações estiverem pouco espaçadas entre si, elas estarão mensurando informações espacialmente correlacionadas e, em virtude disso, fornecerão menos informações do que seria desejável. A respeito disso, GONG e HOWART (1990) recomendam a amostragem pixel a pixel como uma forma de minimizar os efeitos da correlação espacial. O que se pretende, em última análise, é que o procedimento de amostragem a ser utilizado deverá minimizar os efeitos da correlação espacial bem como assegurar que todas as categorias de interesse sejam adequadamente amostradas (DICKS e LO, 1990). A amostragem por pontos, como preconizam KENDALL e BUCKLAND (1972) citados por LUND (1981), consiste simplesmente de um método de amostragem de uma área geográfica pela seleção de pontos em seu interior, mais especificamente pela escolha casual ou sistemática destes pontos em um mapa, fotografia aérea ou no campo. Estes pontos servirão, primariamente, de indicadores dos locais onde os dados serão coletados. O mesmo autor ainda comenta

que é frequente o não entendimento do que, exatamente, seja a amostragem por pontos. Segundo ele, a falta de entendimento se daria por três razões: o uso na literatura da mesma terminologia para designar diferentes métodos de amostragem; aversão à amostragem por pontos quando ela é, erroneamente, associada a um tipo particular de levantamento ou intensidade de amostragem; e, mais comumente, o não reconhecimento da contribuição da amostragem por pontos quando ela está sendo aplicada.

#### 1.1.1 A Amostragem Totalmente Casualizada

Neste caso específico todos os indivíduos possuem igual chance de serem selecionados. Desta forma, uma vez determinado o número de indivíduos a serem amostrados, procede-se ao sorteio dos mesmos seja por geração automática de números casualizados, seja pela utilização de tabelas apropriadas ou qualquer outro método que assegure a casualização do procedimento..

O principal problema com relação a esta abordagem está apenas à questão da uniformidade na distribuição das amostras por toda a extensão da área de interesse. Não há nada que garanta esta uniformidade fazendo com que, na maioria das vezes, partes da área fiquem subamostradas e outras superamostradas. No caso de imagens orbitais isto pode não representar um grande problema desde que de antemão se saiba que a correlação espacial é bastante grande. Caso contrário os resultados obtidos poderão ser de qualidade duvidosa.

#### 1.1.2 A Amostragem Sistemática

A maioria dos autores de textos em estatística não recomenda este tipo de amostragem quando se pretende casualização (LOETSCH e HALLER, 1964). Como afirma CAMPBELL (1987), em virtude de a seleção do ponto de partida condicionar o posicionamento de todos os outros pontos amostrais, os dados derivados deste tipo de amostragem não cumprem as exigências da estatística de inferência para a seleção casualizada de pontos, fazendo com que os indivíduos não possuam a mesma chance de serem incluídos na amostra.

Por outro lado, de acordo com STEHMAN (1992), amostras sistemáticas são amostras de igual probabilidade porque cada indivíduo possui a mesma chance de ser incluído na amostra. Para ilustrar o seu ponto de vista, ele exemplifica considerando um simples caso de amostragem sistemática de um universo discreto de sete unidades,  $y_1, y_2, \dots, y_7$ . Se o intervalo para a amostragem sistemática for  $k = 3$ , uma das três possíveis amostras,

amostra 1 =  $y_1, y_4, y_7$

amostra 2 =  $y_2, y_5$

amostra 3 =  $y_3, y_6$   
será selecionada, a depender de que valor a casualização do ponto de partida assumir (1, 2 ou 3, respectivamente). Neste caso a probabilidade de que uma determinada unidade seja incluída na amostra, será a probabilidade de que a amostra contendo a unidade seja selecionada. Em virtude de todas as três amostras possuírem a probabilidade de 1/3 de serem selecionadas, todas as sete unidades terão a mesma probabilidade (1/3) de serem selecionadas. No caso da amostragem sistemática bi-dimensional de pixels, numa malha quadrada, o raciocínio se aplicaria da mesma forma. Com relação à não representatividade quando o intervalo de amostragem está em fase com a periodicidade na população, o mesmo autor postula que é possível que uma determinada amostra, dentre outras, não seja representativa, porém esta asserção poderia ser aplicada a qualquer outro padrão de amostragem.

Os geógrafos defendem ponto de vista semelhante, de acordo com DOZIER e STRAHLER (1983), no qual postulam que no caso de correlação espacial e com um pequeno número de pontos a serem amostrados, a amostragem sistemática será representativa e mais precisa desde que não haja periodicidade nos dados que interagem com o espaçamento sistemático das amostras.

### 1.1.3 A Amostragem Sistemática Estratificada Não Alinhada

Este padrão de amostragem combina propriedades da amostragem sistemática e da amostragem estratificada, propiciando a cobertura de toda a área de interesse e, simultaneamente, preservando a casualidade do procedimento. Como explica CAMPBELL (1987), a área de estudo é dividida em células uniformes, geralmente por meio de um reticulado quadrado. As células deste retículo representam a porção sistemática do processo enquanto que a casualização é proporcionada pela localização das observações dentro de cada parcela. O resultado da localização destas observações dentro da área de estudo possui várias características favoráveis. O uso do reticulado de células como base para a estratificação significa que as observações são distribuídas de maneira a contemplar toda a extensão da área. A componente de casualização não permite o rígido alinhamento, característico da amostragem sistemática pura, introduzindo um elemento de chance que aumenta a probabilidade de que as observações venham a contemplar as categorias presentes na região.

O procedimento para o lançamento dos pontos neste tipo de amostragem se dá, inicialmente, dividindo-se a região a ser amostrada em um número de

subregiões igual ao número de pontos a serem gerados. Na sequência, é gerado casualizadamente 1 ponto em cada uma das subregiões. Percebe-se aí a importância do estabelecimento do número de pontos. Um número muito pequeno irá implicar em uma subregião maior para ser representada por um único ponto.

## 2. Metodologia

### 2.1 O Algoritmo de Classificação

Todas as classificações levadas a efeito no estudo foram produzidas aplicando-se um algoritmo de máxima verossimilhança, provido pelo sistema IDRISI (programa MAXVER), assumindo-se a equiprobabilidade de ocorrência das categorias listadas anteriormente. Este algoritmo exige como entrada a definição de probabilidades *a priori* e os nomes dos arquivos de assinaturas a serem usados no processamento. A não especificação de probabilidades para nenhuma das classes implica na equiprobabilidade.

O algoritmo de máxima verossimilhança foi aplicado às seguintes combinações de bandas, gerando as classificações à sua direita:

COMBINAÇÃO	CLASSIFICAÇÃO
TM3, TM4, TM5	MAX345
TM3, TM4, TM7	MAX347
TM2, TM3, TM4	MAX234

Da mesma forma ele foi aplicado às componentes “verdor”, “brilho” e “umidade”, da transformação *Tasseled Cap* geradas a partir da aplicação dos coeficientes aplicados às bandas reflexivas da imagem, de acordo com JENSEN (1986), gerando a classificação denominada MAXTC.

### 2.2 A Amostragem Sobre os Dados de Referência

As amostragens sobre os dados de referência foram conduzidas de acordo com as estratégias blocada, sistemática estratificada, sistemática e casualizada.

No caso da amostragem blocada foram selecionados, sobre a imagem, polígonos representativos das 11 categorias informacionais em concordância com a interpretação das fotos aéreas. Estes polígonos foram delineados sem que pixels que já houvessem sido amostrados na etapa de treinamento fossem contemplados. Para que isso fosse conseguido foi sobreposto o arquivo vetorial, contendo os polígonos de treinamento, à imagem sobre a qual seriam coletados os novos polígonos que serviriam para a verificação da exatidão das classificações. Os polígonos coletados, no formato vetorial, foram então transformados para o

formato matricial ou “raster”, gerando um novo arquivo.

Para as amostragens não bloqueadas o procedimento foi bastante semelhante. Utilizando-se o algoritmo denominado SAMPLE, do sistema IDRISI, foram gerados arquivos vetoriais, um para cada amostragem não bloqueada, contendo, inicialmente, 1200 pontos, cada. Este número de pontos foi determinado empiricamente objetivando contemplar todas as categorias presentes na imagem, especialmente aquelas que ocorrem com pouca frequência ocupando pequenas áreas. De acordo com JENSEN (1986), o número mínimo de pontos pode ser determinado pela relação:

$$N = [4(p)(q)] / E^2$$

onde,

p = percentual de exatidão

esperado;

q = 1 - p;

E = erro permissível.

No presente caso foi considerado um percentual de exatidão de 85% e um erro permissível de 5%, o que implica, após a aplicação da relação acima, em um número mínimo de 204 pixels a serem amostrados.

À semelhança do que ocorreu com a amostragem bloqueada, para as outras amostragens também foram desconsiderados aqueles pixels que foram tomados para o treinamento do classificador, o que justifica o número de pixels amostrados em cada amostragem não bloqueada ser inferior a 1200. Os pontos amostrais sobre cada uma das imagens são os 1200 pontos inicialmente lançados, ou seja, antes de se desconsiderar aqueles que haviam sido contemplados pelas amostras de treinamento.

De posse dos arquivos vetoriais contendo os pontos a serem amostrados, eles foram sobrepostos a diferentes composições de bandas que permitissem a melhor observação dos diversos tipos de cobertura vegetal. Desta forma, ora foi utilizada a composição em falsa cor dos canais TM4, TM3 e TM2, ora a imagem resultante da primeira componente principal e outras. Cada ponto presente sobre a imagem orbital foi localizado nas fotografias e, com o auxílio de um estereoscópio de espelhos, verificada a que categoria informacional o pixel sob aquele ponto pertenceria. Para que a correspondência imagem/fotografia fosse a mais exata possível, lançou-se mão de operações de ampliação sobre a imagem e do uso de oculares de aumento com o estereoscópio e, também, de um reticulado auxiliar que subdividia a imagem em 16 sub-imagens para que o analista não se perdesse na tarefa de identificar cada um dos pontos. Desta forma, a cada

ponto que era identificado, digitalizava-se um outro coincidente com o que havia sido sobreposto anteriormente, com o atributo referente à categoria informacional que ele pertencesse. Em resumo, o arquivo contendo os pontos amostrais foi sobreposto à imagem apenas para servir de orientação locacional na coleta dos pontos que comporiam o arquivo de referência. Esta sequência foi repetida para todos os procedimentos não bloqueados de amostragem, gerando novos arquivos vetoriais que, em seguida, foram submetidos a um processo de rasterização gerando, como resultado, arquivos matriciais. Estes arquivos foram então submetidos a uma operação de álgebra cartográfica na qual cada um deles foi multiplicado por um arquivo booleano onde os pixels no interior dos polígonos tomados como amostras de treinamento receberam o valor zero, e os demais valor 1. Desta forma os arquivos de referência passaram a conter unicamente pixels que não faziam parte das amostras de treinamento, estando prontos para serem usados nas verificações de exatidão.

### 2.3 O Cálculo de Kappa

Uma vez estabelecidas as 5 classificações (MAXTC, MAXPC123, MAX345, MAX347 e MAX234) e os quatro arquivos de referência, um para cada estratégia de amostragem (bloqueada, casualizada, sistemática estratificada e sistematizada), foi possível gerar as matrizes de erro que possibilitariam o cálculo de Kappa.

Foi gerada uma matriz de erro para cada classificação comparada com cada um dos quatro arquivos de referência resultando, ao final do processo, 4 matrizes de erro para cada classificação. Para cada uma destas matrizes foi calculado o índice Kappa de exatidão, de acordo com o que especificam MA e REDMOND (1995), ou seja:

$$K = \frac{P_0 - P_c}{1 - P_c}$$

As relações para  $P_0$  e  $P_c$  podem ser explicitadas por:

$$P_0 = \frac{\sum_{i=1}^M n_{ii}}{N}, \text{ e}$$

$$P_c = \frac{\sum_{i=1}^M n_{i+} n_{+i}}{N^2}, \text{ onde}$$

$N$  = número total de pixels contemplados pela matriz de erros;

$n$  = elemento da matriz de erros;

$M$  = número de categorias informacionais presentes na matriz de erros;

$n_{ii}$  = elementos da diagonal principal;

$n_{i+}$  = total da linha para uma dada categoria informacional, e

$n_{+i}$  = total da coluna para esta mesma categoria informacional.

### 3. Conclusão

A figura 1 ilustra os valores de Kappa obtidos em cada uma das classificações geradas, quando confrontadas com os arquivos de referência obtidos em cada uma das estratégias de amostragem.

Desconsiderando-se o desempenho de cada uma das cinco classificações e atendo-se somente ao comportamento dos índices delas em cada uma das amostragens, percebe-se que os procedimentos não bloqueados produziram índices sensivelmente mais baixos

do que aqueles dos procedimentos bloqueados. É importante notar que, para a maioria das aplicações, o mínimo de exatidão requerida é de 85%. Desta forma, se apenas o procedimento bloqueado fosse utilizado, todas as cinco classificações estariam acima deste limiar enquanto que nos outros procedimentos que contemplaram algum elemento de casualidade, nenhuma delas seria aceita.

Esta dramática diferença, descrita acima, bem mostra o quanto os índices de exatidão podem ser superestimados a depender da estratégia de amostragem empregada na geração dos dados de referência. Neste caso específico esta diferença poderia ser determinante na decisão do usuário em aceitar ou não o produto gerado. É importante notar que a menos que se mencione a estratégia de amostragem utilizada, o simples índice de exatidão pode não estar, com fidedignidade, representando a confiabilidade esperada do produto, induzindo a ocorrência de erros os mais diversos, principalmente naquelas atividades que por dependerem de exatidão nas informações, a exemplo do planejamento, sofrem efeitos deletérios na consecução dos seus objetivos.

### COMPORTAMENTO DE Kappa

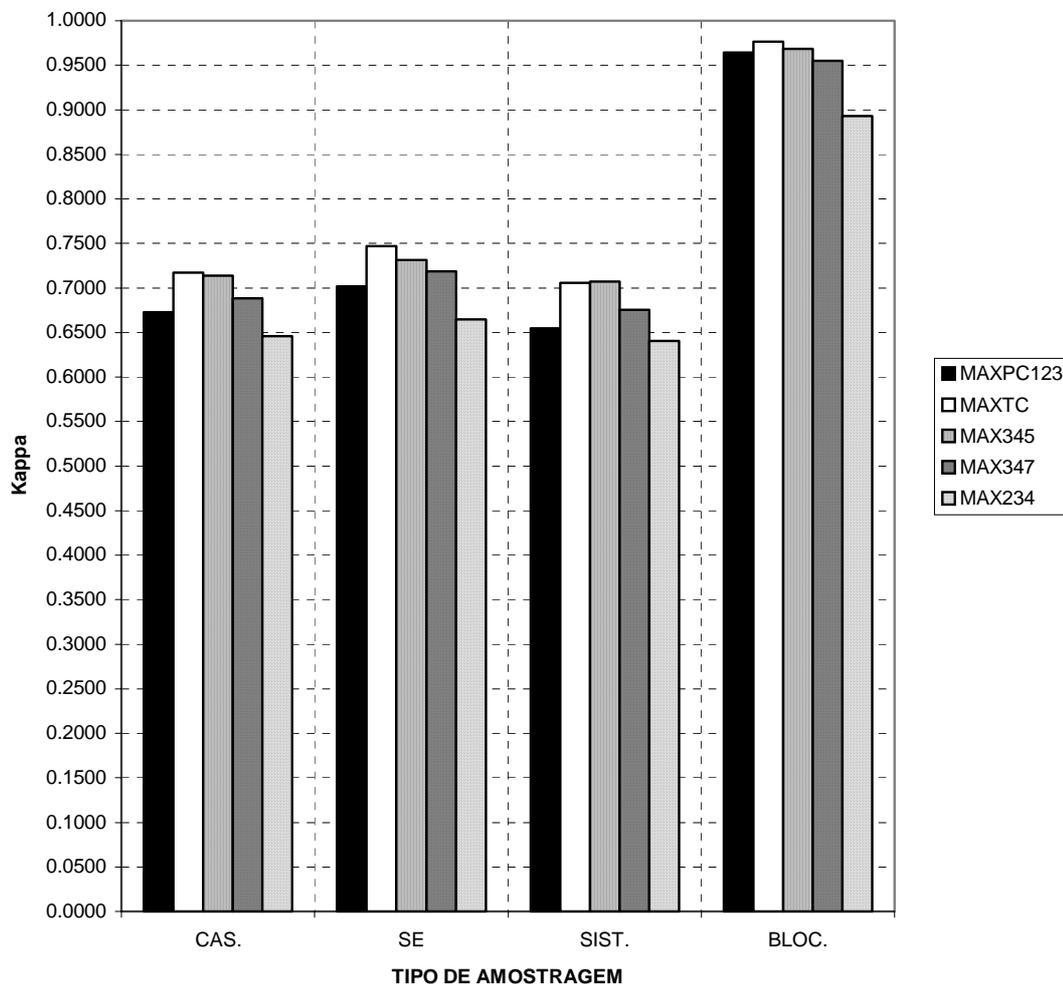


Figura 1 - Comportamento do índice Kappa nas 5 classificações, considerando-se as 4 estratégias de amostragem (casualizada, sistemática estratificada, sistemática e blocada)

#### 4. Referências

CAMPBELL, J. B. *Introduction to remote sensing*. New York, The Guilford Press, 1987. 551p.

CONGALTON, R.G. *A comparison of five sampling schemes used in assessing the accuracy of land cover/land use maps derived from remotely sensed data*. Blacksburg, Virginia Polytechnic Institute and State University. 1984. 146 p. (Tese PhD.)

CONGALTON, R.G. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*. Bethesda, 54, 1669-1671, 1992.

CONGALTON, R.G., BIGING, G.S. A pilot study evaluating ground reference data collection efforts for use in forest inventory. *Photogrammetric Engineering and Remote Sensing*. Bethesda, 58(12), 1669-1671, 1992.

- DICKS, S.E., LO, T.H.C., Evaluation of thematic map accuracy in a land-use and land-cover mapping program. *Photogrammetric Engineering and Remote Sensing*. Bethesda, 56(9), 1247 - 1252, 1990.
- DOZIER, J. STRAHLER, A.H. Ground investigations in support of remote sensing. In: *Manual of remote sensing, 2<sup>nd</sup> edition*, Falls Church, American Society of Photogrammetry and Remote Sensing. v. 1. 1983, p. 959 - 986.
- EBERT, J.A., LYONS, T.R. Archaeology, Anthropology, and cultural resources management. In: *Manual of remote sensing, 2<sup>nd</sup> edition*, Falls Church, American Society of Photogrammetry and Remote Sensing. v. 2. 1983, p. 1233 - 1304.
- GONG, P., HOWART, P.J. An assessment of some factors influencing multispectral land-cover classification. *Photogrammetric Engineering and Remote Sensing*. Bethesda, 56(5), p.597-603, 1990.
- JENSEN, J.R. *Introductory digital image processing*. New Jersey, Prentice-Hall, 1986. 379 p.
- KELLY, B.W. Sampling and statistical problems. In: *Remote sensing with special reference to agriculture and forestry*, Washington, DC, National Academy of Sciences, 1970, p. 324 - 353.
- LOETSCH, F., HALLER, K.E. *Forest inventory*. München, BLV Verlagsgesellschaft. v. 1. 1964. 218p
- LUND, H.G. Point sampling -- The role in in-place inventories. In: *In-Place Resource Inventories: Principles & Practices*, Orono, 1981. *Proceedings of a National Workshop*. Society of American Foresters. 1981. p.371 - 374.
- MA, Z., REDMOND, R.L. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, Bethesda, 61(4), 453 - 439. 1995.
- STEHMAN, S.V. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, Bethesda, 58(9), 1343 - 1350. 1992.
- WARREN, S.D., JOHNSON, M.O., GORAN, W.D., DIERSING, V.E. An automated objective procedure for selecting representative field sample sites. *Photogrammetric Engineering and Remote Sensing*, Bethesda, 56(3), 333-335. 1990.