

# Cálculo de tamanho de amostra para análise de acurácia em mapeamentos temáticos

## Sample Size Calculation For Accuracy Analysis In Thematic Mappings.

Marilia Ribeiro Zanetti<sup>1</sup>  
Vanessa Silva Pugliero<sup>2</sup>  
Mariela Brito de Almeida<sup>3</sup>  
Eduardo Delgado Assad<sup>4</sup>

**Resumo** – A análise da acurácia entre mapeamentos de fontes diversas é a metodologia mais difundida para estimar taxas de acerto e erro em produtos cartográficos. Essa metodologia consiste na comparação de um determinado elemento mapeado em uma classe em relação com um resultado ideal. O número e a alocação de pixels para a aferição do produto gerado com os dados de referência são definidos por técnicas estatísticas. A proposta neste trabalho é calcular a quantidade de pontos amostrais necessários para realizar a acurácia e verificar a sua confiabilidade em relação ao mapeamento.

Termos para indexação: acurácia, geoprocessamento, mapa temático.

**Abstract** – Calculation of sample size for analysis of accuracy in thematic mappings. The accuracy analysis among mappings of several sources is the most widespread methodology to estimate hit and error rates in cartographic products. This methodology consists in the comparison of a certain element mapped in a class in relation to an ideal result. The number and the allocation of pixels for measurement the generated product with the reference data are defined by statistical techniques. The purpose of this review is to calculate the number of sampling points required to perform the accuracy and to verify its reliability in relation to the mapping.

Index terms: accuracy, geoprocessing, thematic map.

---

1 Especialista em Sistemas de Gestão da Qualidade, pós-graduada, bolsista DTI-B do CNPq

2 Especialista em Geoprocessamento, pós-graduada, bolsista DTI-B do CNPq

3 Mestranda em Geografia, pós-graduanda, bolsista nível DTI-B do CNPq

4 Cientista agrário, doutor, pesquisador da Embrapa Informática Agropecuária, Campinas, SP

## Introdução

A acurácia é o grau de proximidade de uma estimativa com seu parâmetro ou valor verdadeiro (Mikhail; Ackermann, 1976). Essa é a principal forma de verificar as diferenças entre mapeamentos, pois além de dizer qual a taxa de acerto geral, a análise de acurácia também revela estimativas das taxas de acerto, de erro de omissão e de erro de introdução em cada classe mapeada.

Na classificação tradicional (rígida), considera-se que todo elemento (pixel ou polígono) está associado a uma única classe temática. O erro surge sempre que esta associação à imagem classificada ou o mapa temático diverge da verdade que representa um resultado ideal. A referência pode ser obtida a partir de: dados pré-existentes, dados de campo e/ou imagens com de maior resolução espacial. Independentemente da base adotada como verdade, ela nunca será perfeita (Congalton; Green, 2009).

De acordo com a International Organization for Standardization (2013) existem diferentes tipos de acurácia, sendo a acurácia analisada no presente estudo a acurácia temática:

- 1) acurácia posicional é definida como a precisão da posição de feições dentro de um sistema de referência espacial.
- 2) acurácia temática é definida como a correta classificação entre as classes e seus atributos em relação a um universo de discussão.
- 3) acurácia temporal é definida como a qualidade dos atributos temporais e relações temporais de feições.

Segundo Barbetta (2004), o cálculo do tamanho de amostra é um problema complexo e envolve a heterogeneidade da população estudada e os tipos de parâmetros que se deseja estimar, como proporções, médias, entre outros. Além disso, a questão de recursos financeiros e tempo também deve ser levada em consideração.

As estimativas de acurácia em mapeamento são baseadas em dados de referência que consiste em interpretação visual de bases cartográficas por meio de padrões de localização, tonalidade e cor, tamanho, forma, textura, padrão, sombra, altura e profundidade, volume e declividade, aspecto, situação e associação para comparação com o mapeamento gerado (Jensen, 2009). O número e a alocação de pixels para a aferição do produto gerado com os dados de referência são determinados por técnicas de amostragem estatísticas, sendo o método adotado nesse projeto o de amostragem estratificada. Tal método foi escolhido devido à população estudada possuir estratos com características e quantidades diferentes dessa forma, adotar tal técnica permite que a proporção de cada estrato seja repetida na amostra. Além disso, técnicas que apliquem o fator de proporcionalidade tem como vantagem a facilidade na realização do trabalho em campo, implicando em uma significativa economia de recursos (Bracarense, 2012).

No caso do projeto TerraClass que realizou o mapeamento do bioma Cerrado, foi feita a amostragem aleatória e a avaliação do resultado do mapeamento efetuado, para isso foi realizada a definição dos estratos da classe de uso do solo utilizando-se a função multinomial e a distribuição da quantidade de amostras por estratos foi realizada pelo método de alocação ótima de Neyman (TerraClass, 2015).

O objetivo desse estudo é calcular a quantidade de pontos necessários para fazer a acurácia dos biomas brasileiros mapeados pelo projeto Mapeamento Anual da Cobertura e Uso do Solo do Brasil (MapBiomas), esse projeto é uma iniciativa que envolve uma rede colaborativa com especialistas nos biomas, usos da terra, sensoriamento remoto, SIG e ciência da computação utilizando processamento em nuvem e classificadores automatizados desenvolvidos e operados a partir da plataforma Google Earth Engine para gerar uma série histórica de mapas anuais de cobertura e uso da terra do Brasil (Projeto Mapbiomas, 2018).

## Material e Métodos

Adotou-se neste projeto o método de amostragem estratificada, sendo este consistente em calcular a proporção de cada estrato em relação ao total da população e obter uma quantidade de amostras por estrato, seguindo a figura 1:

$$n = \frac{(Z_{\alpha/2})^2 \times p \times q}{e^2}$$

Onde:

n = tamanho da amostra

Z( $\alpha/2$ ) = valor crítico que corresponde ao grau de confiança desejado

p = proporção populacional de indivíduos que pertence à determinada categoria

q = proporção populacional de indivíduos que não pertence à determinada categoria (1- p)

e = margem de erro

Para chegar ao cálculo do tamanho de amostra (n) é também necessário identificar se a população estudada (N) é finita ou infinita, pois caso a população seja considerada finita deve-se realizar um cálculo de fator de correção. Segundo Triola (1999) para considerar uma população finita, o tamanho da amostra (n) deve ser maior ou igual a 5% do tamanho da população (N).

Para o cálculo dessa fórmula, o erro amostral de 2,5% e o intervalo de confiança de 99% foram assumidos, seguindo o mesmo padrão adotado pelo Projeto Canasat (Adami et al., 2012).

Existem seis grandes classes temáticas no mapeamento do uso do solo do MapBiomias: 1) floresta; 2) formações naturais não florestais; 3) uso agropecuário; 4) áreas não vegetadas; 5) corpos d'água; e 6) não observado.

## Resultados

Para se obter os componentes “p” e “q” da equação 1 foi necessário conhecer previamente a área de cada classe. Sendo assim, cada classe ficou com um número de pontos proporcional a sua área, resultando em 5.544 pontos no total para a região do Brasil (Tabela 1).

**Tabela 1.** Número de pontos calculados para cada classe.

Região	Classe	Área (ha)	p	q	Nº Pontos
Brasil	Floresta	532.561.210	0,63	0,37	2.485
	Formações Naturais não florestais	53.724.529	0,06	0,94	627
	Uso agropecuário	243.716.828	0,29	0,71	2.167
	Áreas não vegetadas	4.318.715	0,01	0,99	54
	Corpos D'água	17.205.609	0,02	0,98	210
	Não observado	31.657	0,00	1,00	0
	<b>TOTAL</b>		<b>851.558.548</b>		

## Considerações Finais

Conclui-se que o método de amostragem estratificada pode ser aplicado para realizar o quantitativo de pontos para cálculos de acurácia, pois o número de pontos gerados comportou-se de forma razoável tanto para uma conferência em campo como para uma conferência com outras bases de dados, seja classificações temáticas ou servidores de imagens de satélite.

Um aspecto a ser futuramente discutido deve ser o fato de classes temáticas com uma área proporcional muito baixa resultar em número de pontos nulos. É necessário discutir uma forma de se evitar ou contornar tal situação, pois em alguns casos pode-se prejudicar as análises de acurácia. Sendo assim, sugere-se adotar um número mínimo de pontos por classe a fim de evitar tal situação.

## Referências

ADAMI, M.; MELLO, M. P.; AGUIAR, D. A.; RUDORFF, B. F. T.; SOUZA, A. F. de. A Web platform development to perform thematic accuracy assessment of sugarcane mapping in South-Central Brazil. **Remote Sensing**, v. 4, n.10, p. 3201-3214, 2012.

BARBETTA, P. A. **Estatística aplicada às Ciências Sociais**. 5. ed. Florianópolis: Ed. da UFSC, 2004. 340 p.

BRACARENSE, P. A. **Estatística aplicada às ciências sociais**. Curitiba: IESDE Brasil, 2012. 288 p.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 2nd ed. Boca Raton: CRC Press: Taylor & Francis, 2009. 183 p.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19157:2013**: geographic information: data quality. Geneva, 2013. 146 p.

JENSEN, J. R. **Sensoriamento Remoto do Ambiente**: uma perspectiva em recursos terrestres. Tradução: José Carlos Neves Epiphanyo, Antonio Roberto Formaggio, Athos Ribeiro dos Santos, Bernanrdo Friedrich Theodor Rudorff, Claudia Maria de Almeida, Lênio Soares Galvão. São José dos Campos: Parênteses, 2009. 598 p. il. Tradução de: Remote sensing of the environment: an earth resource perspective, 2nd ed.

MIKHAIL, E.; ACKERMAN, F. **Observations and least squares**. New York: University Press of America, 1976. 497 p.

PROJETO MAPBIOMAS. "**Projeto MapBiomias**. 2018. (Coleção MapBiomias, 3; Série anual de mapas de cobertura e uso de solo do Brasil). Disponível em: <<http://mapbiomas.org>>. Acesso em: 20 ago. 2018.

TERRACLASS. **Mapeamento do uso e cobertura do cerrado**: Projeto Terra Class 2013. Brasília, DF, 2015. 67 p.

TRIOLA, M. F. **Introdução à Estatística**. 7. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1999. 410 p.