

## Validação do Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais

José Márcio de Mello

Alan de Brito

Luis Marcelo Tavares de Carvalho

Fausto Weimar Acerbi Junior

Eduarda Martiniano de Oliveira Silveira

Universidade Federal de Lavras - UFLA

Caixa Postal 3037 - 37200-000 - Lavras - MG, Brasil

{mello, passarinho, fausto}@ufla.br, alandebrito@hotmail.com, dudalavras@hotmail.com

**Abstract.** This paper describes the methodology used for accessing the accuracy of vegetation maps from Minas Gerais, Brazil. Its main objective was to define suitable accuracy measures which characterize maps of vegetation cover to be used within the framework of the statewide forest inventory. Sampling size was defined based on a desired a priori accuracy of 85%. It was implemented a simple random sampling design. A road transportation database was used as the reference for sampling selection. Standard error matrices were produced to derive the accuracy measures global accuracy, Kappa and Tau. Confidence intervals were determined based on random sampling with proportions. The results showed that the minimum sampling size to achieve the desired accuracy was of 196 sample points. Using the road network, 11989 sample points were visited in the field. Values found for Global accuracy, Kappa, and Tau were 87.29%, 81.74%, and 86.31%, respectively. The confidence interval for the precision of the vegetation map of Minas Gerais is between 84.84% and 85.15%. It is concluded that simple random sampling was adequate to allocate the sample points as the a priori accuracy was lower than the actual accuracy and the later is within the short confidence interval.

**Palavras-chave:** Vegetation mapping, accuracy assessment, confidence interval, mapeamento da vegetação, verificação da acurácia, de confiança.

### 1. Introdução

Os trabalhos de mapeamento da cobertura do solo de grandes extensões territoriais destacam-se entre as principais aplicações das técnicas de sensoriamento remoto. Este tipo de mapeamento, geralmente divide os tipos de cobertura em classes temáticas específicas. Além disso, o mapeamento visa representar, com certa precisão, a cobertura real encontrada em campo. Embora as técnicas mais sofisticadas sejam aplicadas, algumas limitações ainda são encontradas. Dentre elas a presença de nuvens, a resolução espacial dos sensores, erros de retificação, erro posicional e temático, além dos erros provenientes da classificação digital de imagens, tornando imprescindível a determinação da acurácia dos mapas temáticos gerados a partir do processamento de imagens de sensoriamento remoto.

A acurácia temática refere à precisão da classificação das diferentes categorias que compõem um mapa. Ao trabalhar-se utilizando mapas provenientes de dados de sensoriamento remoto, deve-se avaliar a precisão dos mesmos. Para esta avaliação são usados dados de referência coletados em campo (Richards, 1986).

A utilização dos conceitos de teoria da amostragem para a avaliação da acurácia de mapas temáticos é de fundamental importância nessa avaliação, tanto no que tange suas aplicações práticas, quanto às técnicas e políticas. Na prática está relacionada com um tamanho de amostra representativa e confiável da população, reduzindo os gastos de tempo e recursos. Já no que se refere às técnicas e políticas, isto se faz verdadeiro na medida em que constitui uma ferramenta de refinamento da qualidade dos mapas que, além de proporcionar o planejamento de diretrizes para uma política de conservação realista, fornecem informações aplicáveis em diferentes campos do conhecimento.

Uma necessidade pertinente ao tema, pouco estudada pelos técnicos da área florestal é a escolha correta do procedimento de amostragem para este fim. Este é diretamente responsável

pela determinação das medidas de acurácia do mapeamento e define o custo de sua determinação. Uma avaliação de acurácia estatisticamente rigorosa é aquela que o procedimento de amostragem satisfaz o protocolo de amostragem probabilística e as estimativas são estatisticamente consistentes, além de obter estimadores mais precisos com menor custo possível.

O presente trabalho teve como objetivo determinar a acurácia e a precisão da classificação temática do “Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais”.

## 2. Material e Métodos

### 2.1 Caracterização geral da área

A área de estudo compreendeu todo Estado de Minas Gerais (Figura 1). Procedeu-se a avaliação da acurácia a partir do mapa gerado pela classificação temática realizada para o Estado (Scolforo & Carvalho, 2006). Nesta foram utilizadas 35 cenas do satélite Landsat, coletadas através dos sensores TM e ETM+, retratando três épocas do ano (primavera, verão e inverno).

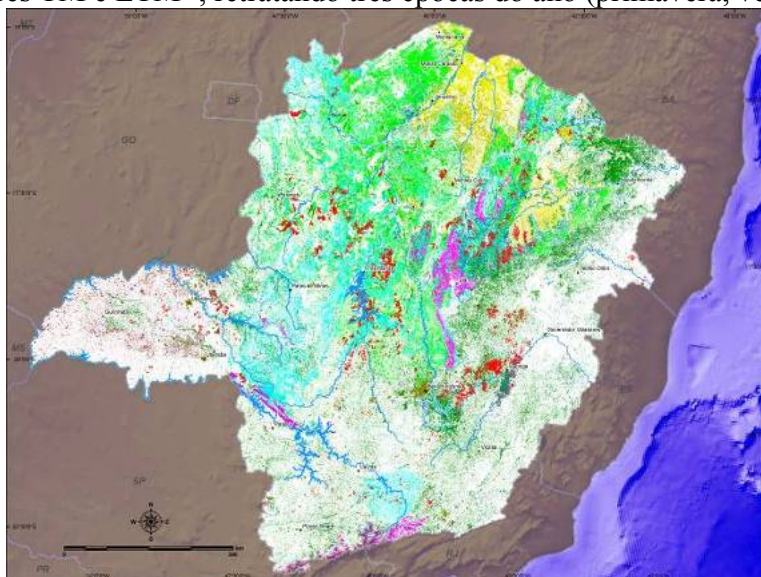


Figura 1. Mapa da classificação temática da flora nativa e dos reflorestamentos do estado de Minas Gerais.

O método de classificação foi o supervisionado e o algoritmo classificador utilizado foi o de árvore de decisão (Carvalho et al., 2005). A área de cada classe temática, em percentagem, presentes no Estado estão discriminadas na Tabela 1.

Tabela 1. Área ocupada, em percentagem, por classe temática.

Fisiomonía	Área (ha)	%
Floresta Ombrófia	224.127	0,38
Floresta Estacional Semidecidual	5.181.382	8,84
Floresta Estacional Decidual	2.035.057	3,47
Campo	3.844.963	6,56
Campo Rupestre	616.009	1,05
Cerrado	5.504.795	9,39
Campo Cerrado	1.489.256	2,54
Cerradão	353.388	0,60
Vereda	406.253	0,69
Eucalipto	1.218.212	2,08
Pinus	143.395	0,24

## 2.2 Coleta de dados

As informações de referência foram obtidas através de incursões terrestres, realizadas para retratar amostras das diversas formações vegetais da área e para averiguar a acurácia do trabalho. Para tal, utilizou-se um aparelho de GPS acoplado a um veículo automotor. Todos os pontos amostrais coletados foram diferentes daqueles utilizados para treinamento do classificador.

O tamanho da amostra de referência para a determinação da acurácia desse mapeamento temático foi obtido através da fixação da acurácia desejada e não de uma amostra piloto. Com isso, comete-se um erro de aproximação no dimensionamento (Costa & Brites, 1998). Porém, os autores concluem que o dimensionamento com a exatidão pré-fixada não provoca grandes tendências no erro de amostragem, quando, a exatidão a priori não se afastada muito da exatidão estimada pela amostragem piloto. Este procedimento apresenta maior custo quando comparado ao dimensionamento através da exatidão fixada a priori (Costa & Brites, 2004).

O tamanho da amostra de referência foi obtido pela fórmula apresentada para variáveis binomiais discretas (Richards, 1986, Jensen, 1986 e Scolfor e Mello, 2006).

Assim, obtém-se o número de pontos amostrais independentes para se obter um intervalo de confiança com erro pré-estabelecido para o parâmetro exatidão real, tendo nível de probabilidade determinado pelo valor de  $Z$ .

O procedimento de amostragem utilizado foi o de amostragem casual simples (ACS), com proporção. A unidade de controle do número de pontos de cada fisionomia foi a sub-bacia.

Contando com uma base digital de rodovias, foram selecionadas aleatoriamente 50% dessas, independentemente de serem rodovias federais ou estaduais, pavimentadas ou não. Nessas rodovias, agora denominadas obrigatórias, foram alocados os pontos amostrais. Uma vez que a demarcação dos pontos dependia da ocorrência das fisionomias ao longo das estradas, considera-se então que, esta também ocorreu de forma aleatória. Outras rodovias, que não as obrigatórias, também foram amostradas, sendo elas as de deslocamento e adicionais (Figura 2).



Figura 2. Mapa das rodovias da sub-bacia GD3 classificadas como obrigatórias, deslocamento e adicionais.

As de deslocamento foram utilizadas para a condução entre uma e outra rodovia obrigatória. Já as adicionais eram amostradas, quando era notada qualquer região com concentração de alguma(s) fisionomia(s) e, nesta não haviam rodovias obrigatórias selecionadas nem de deslocamento. Sendo assim, em cada sub-bacia, foram amostradas, no mínimo, metade das rodovias presentes na base digital do DER.

No campo foram coletadas as seguintes informações:

- Sentido de deslocamento na rodovia;
- Coordenadas geográficas dos pontos amostrais;
- Classe temática ocorrente;
- Lado da rodovia onde o fragmento se encontrava.

### 2.3 Avaliação da acurácia

A ferramenta utilizada neste trabalho para determinar a acurácia do mapa temático foi a matriz de confusão e, os índices dela derivados. A matriz apresenta de forma sintetizada os resultados da classificação digital em relação à imagem referência, onde teremos então uma distribuição binomial.

Spiegel (1993) afirma que cada frequência observada na matriz corresponde também a uma frequência esperada, e esta é calculada sob uma determinada hipótese conforme as regras da probabilidade. Ainda segundo o autor, a diagonal da matriz ( $x_{ii}$ ) apresenta a frequência observada e representa a concordância entre o esperado e observado em cada classe temática, sendo, portanto os pontos classificados corretamente. A coluna que margeia a matriz representa o total de pontos da imagem para cada classe ( $x_{i+}$ ). Por sua vez, a linha marginal representa o número de pontos encontrados para cada classe ( $x_{+i}$ ) em campo.

Lingnau & Antunes (1997) assumem que, tanto coeficientes de concordância total, quanto de concordância para classes individuais podem ser utilizados para avaliar a acurácia de um mapeamento.

Neste trabalho, foram utilizados os coeficientes de concordância total, sendo eles: o de Exatidão Global, proposto por Hellden *et al.* (1980), o Kappa, proposto por Cohen (1960) e o Tau, proposto por Klecka (1980).

A partir da elaboração da matriz de confusão, foi estabelecido o parâmetro exatidão global, que avalia a precisão do classificador. A exatidão global é calculada pela divisão das somas das entradas que formam a diagonal principal da matriz. Estes autores afirmam que a exatidão global representa a precisão de todo o mapeamento, não indicando como a precisão é distribuída através das classes individuais.

Posteriormente foi calculado o coeficiente Kappa. Este é um coeficiente de concordância para escalas nominais que pede a proporção de concordância, depois que a concordância atribuída à casualidade é retirada de consideração (Cohen, 1960). Desta forma, Lingnau & Antunes (1997) consideram-no um coeficiente que determina a concordância esperada a posteriori, uma vez que a concordância esperada é determinada após a elaboração da matriz de erros.

Este coeficiente considera ainda todos os elementos da matriz de erros, diferentemente dos que consideram apenas aqueles que se situam na diagonal principal da mesma, estimando assim a soma da coluna e linha marginais (Cohen, 1960).

As expressões para obtenção do Kappa e do Tau podem ser encontradas em Congalton (1991) e Ma & Redmond (1995).

Tanto o coeficiente Kappa quanto o Tau apresentam valores de -1 a +1, e a interpretação é realizada da seguinte forma (Lingnau & Antunes, 1997):

- se o coeficiente apresentar valor igual a 1, a classificação é totalmente correta, apresentando concordância de 100%;
- caso apresente valor igual a zero, a classificação temática não apresenta melhoramentos em relação a concordância esperada;
- se for menor que zero, significa que não ocorreu discriminação entre as classes, e
- se for maior que zero, a concordância observada foi maior que a esperada.

## 2.4 Amostragem Casual Simples - Estimação de Proporções

A teoria da amostragem em proporções está baseada na ocorrência de uma série de Bernoulli, a qual caracteriza a distribuição binomial. A característica desta distribuição é o controle da variável avaliada por 0 ou 1, respectivamente, para certo e errado.

Considerando então, uma população com  $N$  fragmentos (polígonos) de vegetação nativa definidos pelo mapeamento digital e desejando avaliar, por exemplo, o número de fragmentos classificados corretamente, estes serão categorizados por zero se não estiverem classificados corretamente ou por 1 se forem corretamente classificados.

Os estimadores da Amostragem Casual Simples com Proporção estão listados na seqüência: média aritmética, variância, variância da média da proporção, erro padrão da média da proporção, precisão do levantamento e intervalo de confiança (Scolforo & Mello, 2006).

Maiores detalhes sobre os métodos utilizados neste trabalho podem ser encontrados em Carvalho et al. (2008).

## 3 Resultados e discussão

Foram amostrados aproximadamente 50000km às margens das rodovias obrigatórias, de deslocamento e adicionais.

O tamanho da amostra de referência através da exatidão fixada a priori, foi calculado a partir da fórmula de dimensionamento apresentada por Scolforo & Mello (2006). A exatidão  $p$  utilizada foi fixada em 85%. Segundo Jensen (1986) essa é a mínima requerida para mapeamentos temáticos. O erro admissível empregado foi de 5% à probabilidade de 95%.

Dessa forma o tamanho da amostra de referência calculado foi de aproximadamente 196 pontos amostrais para todas as classes do mapeamento, para um valor crítico de  $Z$  tendendo ao infinito ( $Z = 1,96$ ). Tendo em vista a distribuição aleatória dos pontos, a ausência de tendência é favorecida pelo fato das feições na imagem - utilizadas na classificação - apresentarem geralmente um padrão de correlação espacial (Costa & Brites, 2004).

Foram coletados 11989 pontos (Figura 3), distribuídos nas 14 fisionomias.

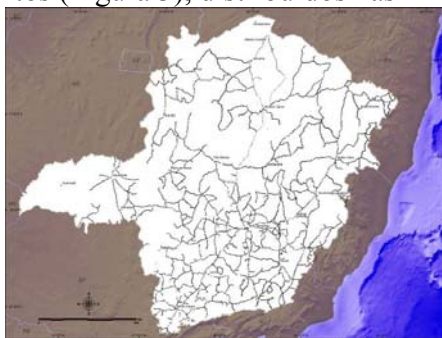


Figura 3. Localização dos pontos visitados em campo.

### 3.1 Avaliação da acurácia

Na Tabela 3 está apresentada a matriz de confusão para a avaliação da acurácia do mapeamento temático do Estado.

Tabela 3. Matriz de confusão para avaliação da acurácia das classes temáticas: FO - Floresta Ombrófila; FS - Floresta Estacional Semidecidual; DC - Floresta Estacional Decidual; C - Campo; CR - Campo Rupestre; CC - Campo Cerrado; CT - Cerrado Sensus Stricto; CO - Cerradão; E - Eucalipto; P - Pinus, U - Urbanização, V - Vereda, O - Outros e A – Água.

Fisionomias	Mapa														Total
	FO	FS	DC	C	CR	CC	CT	CO	E	P	U	V	O	A	
FO	36	53		4									18		111
FS	3	1507		15	2	9	47		14		1	3	204		1805
DC		23	522	4		38	34						100		721
C	1	1		210			1		1	1			25		240
CR		1		15	94								37		147
CC		7	2	11	6	206	2						26		260
CT	1	52	8	14	3	11	861		9	1			92		1052
CO								48				1	4	1	54
E	1	70		14	3	5	11		399	16	1	1	211		732
P		1							3	37			6		47
U				1			1				613		33		648
V		4		3			13	3				60	21		104
O		37	10	60	1	14	31	4	9	1	5		5697	1	5870
A							1						22	175	198
Total	42	1756	542	351	109	283	1002	55	435	56	620	65	6496	177	11989

A partir dessa matriz, foram determinados os índices de concordância total. Conforme o valor do índice de Exatidão Global (87,29%), pode-se afirmar que a exatidão a posteriori foi superior a exatidão a priori (85%). Resultado esse que é confirmado pelo coeficiente Tau.

Como esperado, o Kappa (81,74%) para toda a classificação foi menor que o valor de Exatidão Global (87,29%) e inferior à exatidão fixada a priori. Este fato pode ser explicado pela utilização de todas as células da matriz no cálculo do primeiro, considerando, portanto, os erros de exclusão e inclusão das classes.

O coeficiente de concordância Tau (86,31%) apresentou valor mais próximo ao de Exatidão Global quando comparado ao Kappa. O coeficiente Tau também leva em consideração todos os elementos da matriz de erros, porém, sua determinação é menos rigorosa, pois considera a concordância esperada a mesma para todas as classes temáticas, diferentemente do Kappa que utiliza a proporção de concordância depois que a concordância atribuída à casualidade é retirada de consideração.

Na Tabela 4 está apresentado o número de pontos coletados (PC) em campo de cada classe temática, o número de pontos de concordância com o mapa (PCM) e a diferença entre eles.

Tabela 4. Pontos das classes temáticas amostrados e pontos concordantes.

Classes	PC	PCM	Diferença (%)
Floresta Ombrófila	111	36	67.6
Floresta Estacional Semidecidual	1805	1507	16.5
Floresta Estacional Decidual	721	522	27.6
Campo	240	210	12.5
Campo rupestre	147	94	36.1
Campo Cerrado	260	206	20.8
Cerrado Sensus Stricto	1052	861	18.2
Cerradão	54	48	11.1
Eucalipto	732	399	45.5
Pinus	47	37	21.3
Urbanização	648	613	5.4
Vereda	104	60	42.3
Outros	5870	5697	2.9

Água	198	175	11.6
------	-----	-----	------

A partir dela pode-se notar que a classe que apresentou maior diferença entre o número de pontos coletados em campo e os que concordaram com o mapa foi a FO, seguida de E, V, CR e DC. Para estas classes recomenda-se um estudo mais intenso para determinar os atributos necessários para refinar a aplicação do método de classificação digital. A classe O foi a que apresentou menor diferença, seguidas por U, CO, A e C.

### 3.2 Amostragem Casual Simples - Estimação de Proporções

Analisando os dados de acordo com o procedimento de amostragem casual simples com estimadores de proporção obtiveram-se os resultados apresentados na Tabela 5.

TABELA 5. Estimadores da Amostragem Casual Simples com proporção.

Estimador	Valor
Média	0.87288
Variância	0.12750
Erro Padrão	0.35707
Variância da Média	0.00000
Erro Padrão da Média	0.00124
Precisão	1.05652

O intervalo de confiança determinado para a precisão real do “Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais” é:

$$84,84\% \leq \text{Precisão} \leq 85,15\%$$

A partir da determinação do intervalo de confiança, podemos afirmar que, a proporção verdadeira de polígonos corretamente classificados ou concordantes entre verdade de campo e o Mapeamento está situada entre 84,84% e 85,15% de todos os polígonos da classificação.

### 4 Conclusões

A acurácia do mapeamento de todo o Estado de Minas Gerais foi de 87,29%, 81,74% e 83,31%, determinados pelos coeficientes Exatidão Global, Kappa e Tau, respectivamente.

A utilização da amostragem casual simples com proporção para alocação dos pontos amostrais, assim como, a utilização da acurácia determinada a priori para o cálculo do tamanho da amostra de referência, se mostraram adequadas, uma vez que a acurácia a priori foi inferior à determinada a posteriori e está contida no intervalo de confiança para a precisão real de todo o “Mapeamento da Flora e dos Reflorestamentos de Minas Gerais”.

### Referências

- Antunes, A. F. B. **Uso do Processamento Digital de Imagens na Segmentação de Tipologias Vegetais na APA de Guaraqueçaba-PR**. Dissertação de Mestrado, Curitiba: Curso de Pós-graduação em Engenharia Florestal, Universidade Federal do Paraná. 1996. 119p.
- Antunes, A. F. B.; Lingnau, C. Uso dos Índices de Acurácia para Avaliação de Mapas temáticos obtidos por meios de classificação digital. In: **III Congresso e Feira para Usuário de geoprocessamento**, 1997, Curitiba, III Congresso e Feira para Usuário de geoprocessamento, Curitiba: Sagres, 1997.
- Bishop, Y. M.; Feinberg, S. E. & Holland, P. W. **Discrete Multivariate Analysis - Theory and Practice**. Cambridge, Mass., MIT press, 357 p., 1975.
- Brites, R. S.; Soares, V. P.; Ribeiro, C. A. A. S. **Efeitos da Estratégia de Amostragem na Exatidão Reportada pelo Índice Kappa na Classificação de Imagens Orbitais**. Anais... VIII SBSR, Salvador, Brasil, 14-19 abril de 1996, INPE, p. 823-829.
- Klecka, W. R. **Discriminant Analysis**. SAGE Publications, Inc., Beverly Hills, California, 71 p., 1980.

- Carvalho, L. M. T. et al. Procedimentos para mapeamento. In: **Mapeamento e inventario da flora nativa e reflorestamentos de minas gerais**, edited by Scolforo, Jose Roberto Soares, and Luis Marcelo Tavares de Carvalho. Vol. 1, 37-57. Lavras: Editora UFLA. 2006.
- Carvalho, L. M. T. et al. 2008. Amostragem para Determinação da Acurácia do Mapeamento. In: **Inventário Florestal de Minas Gerais**. Carvalho, L.M.T e Scolforo, J.R.S (Eds.), Vol. 1, 317-328. Lavras: Editora UFLA.
- Cohen, J. **A Coefficient of Agreement for Nominal Scales. Educational and Measurment**. Vol XX, No 1, p. 37-46, 1960.
- Congalton, R. A Review of Assessing the Accuracy of Classification Remotely Sensed Data. **Remote Sensing of Environment**, USA, No 37, p. 35-36, 1991.
- Costa, T. C. C.; Brites, R. S. A influência do tamanho da amostra de referência na exatidão de classificação de imagens de sensoriamento remoto. **Revista Brasileira de Cartografia**, v.56, n.2, p. 151-155, 2004.
- Costa, T.C.C.; Brites, R. S. **Tamanho da amostra de referência para verificação de exatidão em imagens orbitais**. Anais... IX SBSR, Santos, São Paulo, 11-18 setembro de 1998, INPE. (CD-ROM).
- Jensen, J.R. **Introductory digital image processing**. New Jersey, Prentice-Hall, 1986. 379p.
- Lingnau, C.; Antunes, A. F. B. **Avaliação da acurácia de Mapas temáticos oriundos de classificação Digital de Imagens**. In: IV Simpósio Brasileiro de Geoprocessamento, 1997, São Paulo: USP, 1997.
- MA, Z. & REDMOND, R. Tau Coefficients for Accuracy Assessment of Classification of Remote Sensing Data. **Photogrammetric Engineering and Remote Sensing**, Vol. 61(4), p. 435-439. 1995.
- Richards, J.A. **Remote sensing digital image analysis: an introduction**. Springer-Verlag Berlin Heidelberg, 281p. 1986.
- Scolforo, J.R.S., Mello, J.M. **Inventário Florestal**. Lavras, Editora UFLA, 2006. 561p.
- Scolforo, J.R.S.; Carvalho, L.M.T. **Mapeamento e inventário da flora nativa e dos reflorestamentos de Minas Gerais**. Lavras: UFLA, 2006. 288p.
- Spiegel, M. R. **Estatística**. McGrawhill, 3ª edição, 1993.
- Hellden, U. & Stern, M. Evaluation of Landsat Imagery and Digital Data for Monitoring Desertification Indicators in Tunisia. Proc. 14th. Int. **Symp. on Rem. Sens. of Environ.**, p. 1601-1611, 1980.