

ALAN DE BRITO

**ACURÁCIA E PRECISÃO DO MAPEAMENTO DA FLORA NATIVA E
DOS REFLORESTAMENTOS DE MINAS GERAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Ciências Florestais, área de concentração em Manejo de Florestas de Produção, para a obtenção do título de "Mestre".

Orientador
Prof. Dr José Marcio de Mello

**LAVRAS
MINAS GERAIS – BRASIL
2008**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Brito, Alan de.

Acurácia e precisão do mapeamento da flora nativa e dos
reflorestamentos de Minas Gerais / Alan de Brito. – Lavras : UFLA,
2009.

110 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2009.

Orientador: José Marcio de Mello.

Bibliografia.

1. Acurácia. 2. Precisão. 3. Mapeamento temático. 4.
Amostragem. 5. Mapeamento da flora nativa e dos reflorestamentos
de Minas Gerais. I. Universidade Federal de Lavras. II. Título.

CDD – 634.9285

ALAN DE BRITO

**ACURÁCIA E PRECISÃO DO MAPEAMENTO DA FLORA NATIVA E
DOS REFLORESTAMENTOS DE MINAS GERAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Ciências Florestais, área de concentração em Manejo de Florestas de Produção, para a obtenção do título de "Mestre".

APROVADA em 23 de Setembro de 2008

Prof. Dr José Roberto Soares Scolforo UFLA

Prof. Dr Luís Marcelo Carvalho de Tavares UFLA

Prof. Dr João Roberto dos Santos INPE

Prof. Dr José Marcio de Mello
UFLA
(Orientador)

**LAVRAS
MINAS GERAIS – BRASIL**

“Detesto, de saída, quem é capaz de marchar em formação com prazer ao som de uma banda. Nasceu com cérebro por engano; bastava-lhe a medula espinhal.”

Albert Einstein

Aos meus amados pais
Aos queridos irmãos e sobrinhos
À minha afetuosa família
À família Itaipava e agregados
Aos verdadeiros amigos e discípulos
Ao inesquecível Paulo Jarbas
E à natureza

DEDICO

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ,
pela concessão da bolsa.

Ao Governo do Estado de Minas Gerais por meio da Secretaria de Estado de
Meio Ambiente e Desenvolvimento Sustentável (SEMAD) e do Instituto
Estadual de Florestas (IEF) pelos recursos disponibilizados mediante o Projeto
“Inventário Florestal de Minas Gerais”.

À Universidade Federal de Lavras, ao Departamento de Ciências Florestais e ao
Programa de Pós Graduação em Engenharia Florestal pela formação acadêmica.
Ao Professor e orientador José Marcio de Mello, que se esforçou ilimitadamente
para a conclusão deste trabalho.

Ao Professor José Roberto Soares Scolforo por todas as oportunidades.

Aos Professores Luís Marcelo Tavares de Carvalho e Fausto Weimar Acerbi
Junior pela co-orientação.

À sempre mestra e amiga, Maria Zélia Ferreira.

Aos colegas de mestrado.

Ao Mauro Sérgio Rangel que, em mais de cinco meses de convivência intensiva,
sempre se mostrou um ótimo companheiro de viagem e um grande amigo.

Aos estagiários Priscila e Willian.

Aos queridos amigos que sempre fizeram tudo para que minhas lágrimas se
transformassem em sorrisos. Em especial à Eduarda Duda, Eveline, Gabriela,
Marcela Capetão, Adriano Mini, Mayesse, Gú Badaró, Suzana, Renato Miranda,
Guto dentre tantos.

Aos funcionários, professores e estagiários do Laboratório de Estudos e Projetos
em Manejo Florestal.

SUMÁRIO

| | |
|--|-----|
| RESUMO GERAL..... | i |
| ABSTRACT..... | iii |
| CAPÍTULO 1..... | 1 |
| 1 INTRODUÇÃO GERAL..... | 2 |
| 2 REFERENCIAL TEÓRICO..... | 5 |
| 2.1 Padrões de amostragem..... | 5 |
| 2.1.1 Tipo de unidade amostral..... | 6 |
| 2.1.2 Procedimento de amostragem..... | 7 |
| 2.2 Avaliação da acurácia temática..... | 14 |
| 2.2.1 Matriz de confusão e Coeficientes de Concordância..... | 16 |
| 2.2.1.1 Coeficientes de Concordância Total..... | 18 |
| 2.2.1.2 Coeficientes de Concordância para classes individuais..... | 20 |
| 2.3 Determinação da Precisão da Classificação Temática..... | 21 |
| 2.3.1 Tamanho da amostra..... | 21 |
| 2.3.2 Estimadores..... | 22 |
| 3 REFERÊNCIAS BIBLIOGRÁFICAS..... | 25 |
| CAPÍTULO 2..... | 29 |
| AMOSTRAGEM PARA DETERMINAÇÃO DA ACURÁCIA E PRECISÃO DO MAPEAMENTO DA FLORA NATIVA E DOS REFLORESTAMENTOS DE MINAS GERAIS..... | 29 |
| RESUMO..... | 30 |
| ABSTRACT..... | 32 |
| 1 INTRODUÇÃO..... | 34 |
| 2 MATERIAL E MÉTODOS..... | 37 |
| 2.1 Caracterização geral da área..... | 37 |
| 2.2 Coleta de dados..... | 38 |
| 2.3 Avaliação da acurácia..... | 42 |
| 2.4 Determinação da Precisão..... | 43 |
| 3 RESULTADOS E DISCUSSÃO..... | 44 |
| 3.1 Avaliação da acurácia..... | 44 |
| 3.2 Amostragem Casual Simples - Estimação de Proporções..... | 49 |
| 4 CONCLUSÕES..... | 51 |
| 5 REFERÊNCIAS BIBLIOGRÁFICAS..... | 53 |
| CAPÍTULO 3..... | 55 |
| ACURÁCIA E PRECISÃO DO MAPEAMENTO DA FLORA NATIVA E DOS REFLORESTAMENTOS DE MINAS GERAIS POR CLASSE TEMÁTICA E BACIA HIDROGRÁFICA..... | 55 |
| RESUMO..... | 56 |
| ABSTRACT..... | 58 |
| 1 INTRODUÇÃO..... | 60 |
| 2 MATERIAL E MÉTODOS..... | 61 |
| 2.1 Caracterização geral da área..... | 61 |
| 2.2 Coleta de dados..... | 62 |

| | |
|---|----|
| 2.3 Avaliação da acurácia | 66 |
| 2.4 Determinação da Precisão | 67 |
| 3 RESULTADOS E DISCUSSÃO | 68 |
| 3.1 Avaliação por Classe Temática | 68 |
| 3.2 Avaliação da Acurácia e Precisão por Bacia Hidrográfica | 79 |
| 3.2.2 Avaliação da acurácia | 84 |
| 4 CONCLUSÕES | 91 |
| 5 REFERÊNCIAS BIBLIOGRÁFICAS | 93 |
| CONSIDERAÇÕES GERAIS | 95 |
| ANEXOS | 98 |

RESUMO GERAL

BRITO, Alan de. **Acurácia e Precisão do Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais**. Lavras: UFLA 2008. 110p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Lavras, Lavras, MG.*

O presente trabalho foi realizado com o objetivo geral de avaliar a acurácia e a precisão de mapeamentos temáticos. Especificamente objetivou-se avaliar a acurácia e precisão da classificação temática do Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais. Foram coletadas informações de referência de campo em todo o Estado, utilizando o procedimento de amostragem aleatória simples com proporção, contando com o apoio do Método dos Itinerários (Belart et al., 2001). Desse modo, as unidades amostrais - polígonos determinados pelos pontos - foram tomadas aleatoriamente ao longo de rodovias, pavimentadas ou não. Para tal, foram selecionadas rodovias obrigatórias, de deslocamento e adicionais. O tamanho da amostra de referência foi calculado de acordo com a fórmula de dimensionamento do tamanho da amostra (Meyer, 1984). No Capítulo 2, o número de polígonos amostrados por sub-bacia obedeceu à relação entre a área de cada classe presente em determinada sub-bacia em relação à área total da classe no Estado. No Capítulo 3, a mesma amostra foi analisada tanto por classes temáticas quanto por bacias hidrográficas. Com base na obtenção dos dados de referência de campo, foram comparados aos do Mapeamento e a eles atribuídos os valores 0 ou 1 para fracasso ou acerto, respectivamente. Posteriormente foram confeccionadas matrizes de confusão e calculados coeficientes de concordância total e para as classes que a compunham. Realizou-se, também, a determinação da precisão dos polígonos corretamente classificados, além de criar intervalos de confiança para esta variável. O intervalo de confiança para a proporção real de polígonos corretamente classificados pelo Mapeamento foi de $86,67\% \leq \text{Proporção Real} \leq 87,87\%$, sendo considerado satisfatório. O valor apresentado pelo coeficiente de exatidão global foi de 87,27%, enquanto o Kappa de 81,61% e o Tau 86,29%. Apenas o Kappa apresentou valor inferior ao da exatidão fixada *a priori* (85%). As classes temáticas que apresentaram maior confusão na classificação foram Eucalipto e Floresta Ombrófila. Entre as bacias, as que apresentaram maior confusão entre as classes que compuseram o Mapeamento foram as do Rio Piracicaba/Jaguari, Pardo, Jucuruçu e Jequitinhonha. A proporcionalização obtida com a distribuição das unidades amostrais, de acordo com a área que as

* Comitê de Orientação: José Marcio de Mello - UFLA (Orientador), Fausto Weimar Acerbi Júnior - UFLA, Luís Marcelo Tavares de Carvalho -UFLA (Co-orientadores).

classes temáticas ocupavam nas sub-bacias, garantiu que todas as classes apresentassem um número mínimo de unidades para atender o erro pré-estabelecido e criar intervalos de confiança para a proporção de polígonos corretamente classificados. A utilização das rodovias como base para se tomar a amostra foi eficiente, uma vez que a malha viária da área em questão possuía grande extensão e boa distribuição espacial, favorecendo a distribuição das unidades amostrais em toda a área estudada.

Palavras-chave: acurácia, mapeamento temático, inventário florestal.

ABSTRACT

BRITO, Alan de. **Accuracy and Precision of the Thematic Mapping of Native Flora and Reforestation of Minas Gerais**. Lavras: UFLA 2009. 110p. Dissertation (Master of Forest Science) – Federal University of Lavras, Lavras, MG.*

This study aimed to evaluate the overall accuracy and precision of thematic maps. Specifically aimed to evaluate the accuracy and precision of the classification of Thematic Mapping of Native Flora and Forestry of Minas Gerais. Data were collected from the field around the reference state, using simple random sampling procedure in proportion with the support of the itineraries Method (Belart et al., 2001). Thus, the sampling units - polygons determined by points - were taken randomly along the roads, paved or not. To this end, selected roads were mandatory, and additional travel. The size of the reference sample was calculated according to the formula for scaling the size of the sample (Meyer, 1984). In Chapter 2, the number of polygons sampled by sub-basin followed the relationship between the area of each class in this particular sub-basin in relation to the total area of class in the state. In Chapter 3, the same sample was analyzed by both classes as a thematic basins. Upon obtaining the reference data in the field, they were compared to the mapping, and they assigned the values 0 or 1 for failure or success, respectively. Subsequently plates were made of confusion and calculated correlation coefficients for total and for classes that comprised. Was also carried out to determine the accuracy of correctly classified polygons, and generate confidence intervals for this variable. The confidence interval for the proportion of actual polygons correctly classified by mapping was 86.67% Proportion $\leq \leq$ 87.87% Real and is considered satisfactory. The figures given by the coefficient of overall accuracy was 87.27% while 81.61% of the Kappa Tau and 86.29%. Only the Kappa value showed less than the accuracy set *a priori* (85%). The classes had themes that were more confusion in the classification and Eucalyptus Forest Rain. Among the basins, which showed more confusion between the classes that were of composed Mapping Rio Piracicaba / Jaguari, Pardo, and Jucuruçú Jequitinhonha. The proporcionalização obtained with the distribution of sample units according to the area that held the classes in thematic sub-basins, ensured that all classes have a minimum number of units to meet the pre-set error and generate confidence intervals for the proportion of correctly classified polygons.

* Steering Committee: José Marcio de Mello - UFLA (Advisors), Fausto Weimar Acerbi Júnior - UFLA, Luís Marcelo Tavares de Carvalho -UFLA (Co-Advisors).

The use of roads as a basis for taking the sample was efficient since the road network of the area had great length and good spatial distribution, favoring the distribution of sampling units across the area.

Keywords: accuracy, thematic mapping, forest inventory.

CAPÍTULO 1

1 INTRODUÇÃO GERAL

As técnicas de mapeamento da cobertura do solo de grandes áreas destacam-se entre as principais aplicações do uso do sensoriamento remoto (Lingnau, 1998). Este tipo de mapeamento geralmente divide os tipos de cobertura em classes específicas. Além disso, visam a 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-se imprescindível a determinação da acurácia de mapas temáticos (Congalton & Green, 1999).

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

A avaliação da acurácia por meio de censo torna-se impraticável uma vez que seria necessário que se conhecessem, na realidade, todos os *pixels* (ou conjunto de *pixels*, ou polígonos) da área de estudo. Assim, é necessário lançar mão de procedimentos de amostragem para definir a exatidão e a precisão da classificação. Com base em uma amostra é possível inferir sobre a população (Cochran, 1965; Scolforo e Mello, 1997). O conhecimento da cobertura verdadeira do solo implicaria em gastos excessivos de tempo e recursos financeiros, tornando-se inviável.

A utilização de conceitos de teoria da amostragem para a avaliação da acurácia e precisão de mapas temáticos é de fundamental importância, tanto no que tange às suas aplicações práticas quanto técnicas e políticas. Na prática, essa teoria está relacionada com um tamanho de amostra representativa e confiável

da população, reduzindo os gastos de tempo e recursos. No que se refere às técnicas e políticas, isso 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 florestais, principalmente no Brasil, é a escolha correta do procedimento de amostragem para este fim. O procedimento é diretamente responsável pela determinação das medidas de acurácia e precisão do mapeamento e define o custo de sua determinação. Para realizar uma avaliação estatisticamente rigorosa, o procedimento de amostragem deve satisfazer o protocolo de amostragem probabilística e as estimativas deverão ser estatisticamente consistentes, além de obter estimadores mais precisos com menor custo possível (Péllico-Neto & Brena, 1998).

Diante desta necessidade, o presente trabalho foi realizado com o objetivo geral de avaliar uma metodologia de amostragem para estimativa da acurácia e da determinação da precisão de mapeamentos temáticos de grandes áreas, que seja viável economicamente e com alto rigor estatístico. Especificamente, objetivou-se obter os índices de acuracidade da classificação temática do Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais no ano 2006, realizado pela Universidade Federal de Lavras, em parceria com o Governo Estadual de Minas Gerais por meio da Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável (SEMAD) e do Instituto Estadual de Florestas (IEF).

O trabalho foi dividido em três capítulos. O Capítulo 1 refere-se à “Introdução Geral” e “Referencial Teórico” acerca dos temas abordados nos demais capítulos.

No Capítulo 2, intitulado “*Amostragem para Determinação da Acurácia Temática do “Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais”*”, foi conduzido com o objetivo de determinar a acurácia e precisão da classificação temática do “Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais”. Além disso, procurou-se desenvolver/aperfeiçoar um sistema de amostragem dessas medidas, quando a classificação temática foi realizada para grandes áreas.

No Capítulo 3, “*Acurácia e Precisão do Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais por classe temática e bacia hidrográfica*”, o objetivo foi determinar a precisão da classificação dos estratos definidos pelas classes temáticas presentes no mapeamento, criando intervalos de confiança para a proporção de polígonos corretamente classificados. Além disso, objetivou-se determinar a acurácia e a precisão da classificação para os estratos definidos pelas bacias hidrográficas, presentes no “Mapeamento da Flora Nativa e Reflorestamentos de Minas Gerais”.

2 REFERENCIAL TEÓRICO

2.1 Padrões de amostragem

Quando se trabalha com grandes áreas e/ou recursos financeiros escassos, o custo torna-se um aspecto da avaliação da acurácia bastante relevante no planejamento e afeta todo o projeto de avaliação da acurácia de mapeamentos, uma vez que as visitas a campo, para obtenção dos pontos de verdade terrestre, são necessárias para essa avaliação, em grande número e dispendiosas. Uma avaliação mais eficiente quanto aos custos pode tornar a avaliação mais praticável, ainda mantendo o rigor estatístico. Para isso torna-se necessário construir uma estratégia de amostragem eficiente empregando conceitos e métodos tradicionais de amostragem científica. (Stehman, 2001).

Brites et al.(1996) definem padrão de amostragem como a forma como as amostras serão tomadas em campo, em termos espaciais. Os mesmos autores afirmam que os índices de exatidão podem ser superestimados dependendo da estratégia de amostragem empregada na determinação dos dados de referência. Quando se trabalha com dados de sensoriamento remoto, este padrão pode ser responsável pela incorporação ou não de alta correlação espacial entre as unidades amostrais.

Como o estudo de procedimentos de amostragem, para realizar comparações de acurácia em práticas de sensoriamento remoto, freqüentemente, é realizado somente em um único local de estudo a habilidade de generalizar os resultados dessas comparações é limitada. Estudos de comparação de acurácia de mapas podem ser definidos para fornecerem generalizações mais fortes, incorporando princípios experimentais do procedimento, assim como a replicação e identificação de uma unidade experimental apropriada para a aplicação (Stehman, 2006).

Segundo Stehman et al.(2000), a inferência amostral é o processo de generalização de dados amostrais para fazer conclusões sobre uma determinada população. O padrão de amostragem é um sistema comumente invocado quando técnicas de amostragem são utilizadas na avaliação da acurácia temática de mapeamentos.

Ao se considerar uma amostra, a acurácia é apresentada pelo desvio padrão da amostra do valor real da população. Já a precisão poderia ser relacionada ao desvio padrão da média da amostra. Sendo assim, de acordo com Antunes e Lingnau (1997), seria aceitável que uma amostra possuísse uma elevada precisão, ou seja, um pequeno desvio em relação à média e, ao mesmo tempo, por causa da falta de exatidão, um grau de acurácia inferior.

Belart et al. (2001) avaliam que, para determinar uma dentre as múltiplas opções de métodos de controle da qualidade de classificações, obtidas por imagens de satélite, deve-se levar em conta o custo em tempo, esforço e financeiro. Os mesmos autores colocam alguns aspectos para obter as medidas de controle de qualidade de um mapeamento, sendo eles: a escolha da unidade amostral mais adequada; o tamanho da amostra a ser tomada; como distribuir estas amostras em campo.

2.1.1 Tipo de unidade amostral

Para a escolha do tipo de unidade amostral existem diversas recomendações, em função do tipo de trabalho, da finalidade, da escala e dos meios e informações auxiliares (Belart et al., 2001). *Pixels* individuais, agrupamentos de *pixels (cluster)*, polígonos ou ainda agrupamentos de polígonos podem ser utilizados como tipo de unidade amostral.

A utilização de *pixels* individuais é recomendada, quando se deseja um incremento no detalhamento da avaliação, aumentando assim a possibilidade de ocorrência de erros além da confiabilidade da avaliação (Aronoff, 1989, citado por Belart et al., 2001). Janssen (1994) também recomenda a utilização de *pixels*

quando este tipo de unidade amostral tenha sido utilizado na classificação da imagem.

A escolha de polígonos como unidade amostral é recomendada por Jensen (1986) em caso em que existam dificuldades de acesso às áreas nas quais será distribuída a amostra.

Gong & Howart (1990), citados por Brites et al. (1998) indicam a amostragem *pixel a pixel* como uma forma de reduzir os efeitos da correlação espacial e garantir que todas as categorias de interesse sejam adequadamente amostradas (Dicks & Lo, 1990).

A amostragem por pontos, consiste em um método de amostragem de uma área geográfica pela seleção de pontos em seu interior, especificamente pela escolha casual ou sistemática destes pontos em um mapa, fotografia aérea ou no campo. Esses pontos servirão, primeiramente, de indicadores dos locais em que os dados serão coletados.

2.1.2 Procedimento de amostragem

Os elementos básicos de uma estratégia de amostragem estatisticamente rigorosa estão compreendidos na especificação da probabilidade do procedimento de amostragem, além de serem acompanhados por estimativas consistentes seguindo os princípios de estimativas de Horvitz-Thompson (Stehman, 2001).

Stehman (2006) também considera que alguns critérios adicionais podem ser seguidos para a escolha de procedimentos de amostragem e, conseqüente determinação da acurácia, sendo eles: precisão adequada para estimadores-chave; eficácia dos custos; simplicidade apropriada para a implementação e análise.

Entre as possibilidades de procedimento de distribuição das unidades amostrais, existem opções desde as mais simples até as mais complexas. Entre as mais simples pode-se citar a amostragem casual simples, muito utilizada por sua

facilidade de aplicação estatística (Belart et al., 2001). A essa também se podem estabelecer pontos de controle sistemáticos, possibilitando maior flexibilidade às análises.

Diversos autores afirmam que diferentes tipos de procedimentos de amostragem podem ser combinados para avaliar a acurácia de mapeamentos temáticos (Zhu et al., 2000; Yang et al., 2001; Stehman et al., 2003).

Uma amostra casual simples pode ser obtida, selecionando aleatoriamente n unidades amostrais e os locais representados por estas unidades, que são visitados para obter a cobertura local, ou seja, a referência de campo. Usando esses dados de campo é possível determinar a proporção de unidades corretamente classificadas em relação ao número total de unidades visitadas (Stehman, 1996).

Richards (1986) afirma que a amostragem aleatória simples não considera o peso de cada uma das classes, ou seja, classes grandes tendem a ser representadas por um número maior de pontos de amostragem que as classes menores, prejudicando a avaliação dessas últimas. Porém, avalia que, para solucionar esse problema, deve-se considerar a adoção de uma amostragem aleatória estratificada, onde o usuário deve, então, proceder, em primeiro lugar, a uma divisão da imagem em grupos de extratos, correspondentes às próprias classes reais e proceder à amostragem aleatória dentro de cada estrato.

Janssen (1994), citado por Belart et al. (2001), recomenda a utilização da amostragem aleatória simples sempre que exista defasagem temporal entre a tomada dos dados de imagem e a sua classificação.

Congalton & Green (1999) recomenda a utilização do procedimento de amostragem casual simples antes de realizar a classificação e, após realizada, a utilização da amostragem aleatória estratificada.

A amostragem estratificada é um procedimento que, segundo Stehman (1996), pode produzir modestos ganhos em precisão por um pequeno aumento

dos recursos financeiros. Uma vantagem da utilização da amostragem estratificada é a precisão da estimativa da acurácia de classes específicas, pois as classes definem os estratos, e o procedimento de amostragem é escolhido para aumentar a precisão da estimativa da acurácia ou erros de comissão (Stehman, 1996).

Um tipo de estratificação comumente utilizada na avaliação da acurácia de mapeamentos temáticos é a estratificação geográfica. Dentre os objetivos principais da escolha dessa estratificação pode-se citar o de obter estimativas de acurácia para regiões geográficas “chave” (uma unidade administrativa ou uma unidade ecológica) e o de obter uma amostra espacial bem distribuída. Uma desvantagem da escolha dessas duas maneiras de estratificação é que os recursos podem não ser suficientes para obter uma amostra de tamanho adequado para estimar precisamente a acurácia em cada estrato desejado (Edwards et al., 1998).

Ao relacionar custo e um procedimento de amostragem eficaz, o autor cita os procedimentos de amostragem estratificada e por conglomerado como estruturas básicas para avaliação da acurácia. A primeira motivação, quando se pensa na amostragem estratificada, é a precisão da estimativa da acurácia das classes específicas. Quando se implementa esse tipo de amostragem, as classes definem os estratos, e o desenho de amostragem é definido para realçar a precisão da estimativa da acurácia ou erros de comissão (Stehman, 1996).

A amostragem estratificada produz eficientes estimativas da precisão. Os estratos geralmente correspondem às classes das áreas mapeadas, escolhidas com o objetivo de estimar a acurácia do usuário. Outra opção é a estratificação baseada em ecorregiões ou regiões geográficas de interesse. Com a utilização da estratificação pode-se assegurar o mínimo de amostras em cada estrato para atingir estimativas mais precisas de classes raras ou classes menores, quando comparadas às obtidas pela amostragem aleatória simples e sistemática (Stehman, 2001).

Os custos da avaliação podem ser reduzidos, quando se utiliza esse tipo de amostragem, uma vez que se pode estabelecer uma distância máxima de estradas ou áreas de fácil acesso para realizar a escolha das unidades amostrais (Edwards et al., 1998).

Outro procedimento de amostragem amplamente utilizado em avaliações da qualidade de mapeamentos temáticos é o por conglomerados ou *cluster*. Esse procedimento, utilizando uma unidade primária de amostragem que é comumente adotada para redução de custos de locomoção, concentra os esforços de amostragem em regiões específicas (Stehman, 2001). Essa concentração possibilita que mais locais de referência sejam amostrados, e o aumento do número de unidades amostradas pode compensar a perda de informação por unidade devido à correlação intra-conglomerado. Dessa forma, pode-se estimar a precisão sem aumentar o custo (Stehman, 1996).

Sendo assim, a amostragem por conglomerado é o meio pelo qual se obtém controle espacial da distribuição das unidades amostrais sobre a amostra e pode ocorrer em duas escalas: regional e local. O controle espacial regional refere-se a limitar a distribuição espacial das unidades amostrais em macro escala. O controle espacial local consiste em distribuir distintas unidades amostrais espacialmente próximas, provocando pequenos esforços adicionais além dos necessários para amostrar uma única unidade (Stehman, 2004).

Ambos os controles, regional e local, podem ser empregados no mesmo procedimento, sendo a combinação mais efetiva um procedimento multiestágio. Neste procedimento seria realizado o controle regional através da amostragem por conglomerado em dois estágios e o controle local realizado por meio da amostragem por conglomerado em um estágio (Stehman, 2006).

Traçando um paralelo entre o procedimento de amostragem por conglomerados e o de amostragem aleatória simples, Edwards et al.(1998) considera que, em geral, amostras em conglomerados contêm menos

informações por unidade de área que a amostragem aleatória simples. Entretanto, é normalmente mais fácil ou menos dispendioso obter amostras em conglomerados. Moisen et al. (1994), citados por Edwards et al.(1998), foram os primeiros a levantar o questionamento entre os custos de unidade amostral e o índice de informação para mapas de cobertura de grandes áreas, concluindo que a informação retornada era significativa para os custos de dados coletados por conglomerado quando comparados ao de amostragem aleatória simples.

As regiões primariamente amostradas por conglomerados podem ser definidas como propriedades, municípios, regiões estaduais. Stehman (2001) exemplifica definindo que a escolha de municípios, como a unidade primária da amostragem, cria uma opção de procedimento em que não é necessário amostrar elementos em todos os municípios do Estado, diminuindo o número de incursões a campo.

Ainda segundo o autor acima citado, uma desvantagem potencial desse procedimento é a aglomeração espacial dos erros de classificação. Esta desvantagem pode gerar informações por unidade amostral menores que as obtidas com a utilização de outros procedimentos.

Entretanto, a adoção de dois estágios de amostragem por conglomerado pode reduzir esta desvantagem. O segundo estágio consiste na tomada de uma subamostra de elementos dentro de uma unidade primária de amostragem. A restrição espacial da amostra dentro da unidade primária de amostragem reduz custos de locomoção quando são realizadas incursões a campo. Quando a obtenção de dados de referência é realizada, com base em fotografias aéreas ou videografia digital, este segundo estágio reduz a carga de trabalho para obtenção destas para apenas áreas cobertas pela primeira fase amostral (Stehman, 2001).

Outro tipo de procedimento utilizado em um estudo de controle de qualidade de classificações de imagens por Belart et al. (2001) foi denominado de Método dos Itinerários.

Este método consiste basicamente na coleta da amostra, ao longo de uma rede rodoviária ou de qualquer terreno, em que possa circular, tomando as observações a uma distância predefinida em suas margens. Esses mesmos autores compararam este procedimento com amostragem aleatória simples, aleatório restritivo, aleatório estratificado por classes, sistemático, sistemático não alinhado e conglomerados.

Além desta comparação, para todos os procedimentos, atribuíram aos dados coletados, dois tipos de comportamento. O primeiro foi que os dados apresentavam distribuição binomial e, o segundo, multinomial. Estes tipos de comportamento definiram o número de unidades amostrais necessárias. Como unidade amostral, definiram o *pixel*, baseados na simplicidade de implantação e eleição das unidades, na variabilidade da área amostrada e na unidade natural de trabalho.

Ao final do estudo concluíram que, quando a aproximação multinomial é empregada, o número de amostras obtido é muito elevado e, no seu estudo, não apontaram diferenças significativas entre os procedimentos de amostragem testados. Consideram, no entanto, que o estratificado seja a opção mais correta, uma vez que se podem ajustar as unidades amostrais à proporção de cada classe, em relação a todas as classes, além de garantir que todas estas estejam presentes na amostra. O método dos itinerários foi o único que apresentou diferenças significativas quando comparados com os valores considerados como verdadeiros. Segundo os autores, este fato ocorreu, pois, a distribuição das unidades amostrais ocorre, limitadamente, ao longo das vias de comunicação. Além disso, os valores obtidos do coeficiente Kappa foram os menores encontrados, levando a considerar que a qualidade da classificação tenha sido subestimada. Todavia, afirmam que este método poderia ser operativo se as classes de interesse se dispusessem, majoritariamente, nas margens das vias amostradas.

Brites et al.(1996), ao avaliar diferentes procedimentos de amostragem para avaliação da acurácia, concluiu que a amostragem blocada é o único procedimento inadequado. Isso ocorre em virtude do problema de correlação espacial e cria uma superestimativa da exata classificação. Em seu estudo, não encontrou diferença significativa no que se refere à tendência na estimativa da exatidão para os procedimentos de amostragem aleatória, aleatória estratificada, e amostragem sistemática.

O procedimento de amostragem, utilizado ao avaliar a exatidão do Mapa Nacional dos Dados de Cobertura da Terra (NLCD) dos Estados Unidos, ilustra como a amostragem estratificada e por conglomerado podem ser combinadas para conseguir a eficácia de custos e estimativas precisas de classes específicas. O procedimento do NLCD utilizou dez avaliações regionais baseadas nas regiões administrativas federais da *Environmental Protection Agency's* (A U. S.). Dentro de uma única região, a avaliação do NLCD foi projetada para fornecer o controle espacial regional e estratificação por classe de cobertura. Uma amostra em primeiro estágio de unidades amostrais preliminares foi selecionada com taxa da amostragem de, aproximadamente, 2,0%. A estratificação por classe de cobertura foi executada no segundo estágio do procedimento. As classes de cobertura mapeadas foram usadas para estratificar todos os *pixels* encontrados na amostra em primeiro estágio (UAPs). Uma amostra aleatória simples de *pixels* de cada estrato foi então selecionada, com 100 *pixels* por classe. Esse procedimento mostrou-se eficaz para assegurar que todas as classes de cobertura, incluindo as classes raras, estivessem representadas adequadamente de modo que as estimativas de exatidão do usuário fossem razoavelmente precisas, sendo que a característica de conglomerado implementado, para alcançar o controle regional, reduziu os custos consideravelmente.

Em outro exemplo, a vegetação em Utah, nos Estados Unidos, foi modelada por três ecorregiões, cada uma com diferentes algoritmos modeladores e avaliação separada requerida para cada um. Para a avaliação da acurácia, cada ecorregião foi dividida por estradas, definindo um corredor de 1 km de largura considerando os estratos de estradas de asfalto e de estradas de terra. Nestes estratos foram aleatorizados os pontos amostrais. Em função dos confinamentos financeiros e da duração das estações do ano, procurou-se um procedimento que maximizasse a quantidade de informação que uma equipe de campo com duas pessoas pudesse coletar em 4,5 meses. O trabalho preliminar determinou que, em um dia de averiguação, o grupo poderia coletar informações em 20 pontos amostrais, dez deles aglomerados no estrato estrada de asfalto e dez distribuídos aleatoriamente no estrato das estradas de terra. O excessivo tempo de viagem entre os pontos foi um fator que impediu que mais pontos fossem coletados. Todos os pontos amostrais foram aleatoriamente gerados no laboratório antes do trabalho de campo. Os pontos amostrais foram selecionados aleatoriamente por meio de uma grade. (Edwards et al., 1998).

2.2 Avaliação da acurácia temática

Existem diversas definições do termo acurácia, geralmente confundido com o termo precisão. Lobão et al. (2005) definem que a acurácia está relacionada à veracidade de obtenção de dados em um intervalo de probabilidade especificado, enquanto que precisão relaciona-se à repetição dos dados próximos a um valor sendo, portanto, uma grandeza que mede a dispersão dos dados. Assim, consideram que o processo de avaliação da acurácia transcende o processo de mensurar a precisão dos dados.

Já a avaliação da acurácia de mapas temáticos é um tipo específico de comparação entre mapas. Esta avaliação é freqüentemente aplicada em problemas de classificação por sensoriamento remoto. Para a comparação entre mapas, quando se realiza esta avaliação, um mapa representa a saída da

classificação, ou seja, o próprio mapeamento e, o outro representa a referência de campo ou condição “referência” (Stehman, 2006).

Na avaliação da acurácia temática de um mapa podem ser considerados vários objetivos. Um objetivo bastante utilizado é o de descrever os erros da classificação e confeccionar uma matriz de erros ou de confusão. Outro também importante objetivo é o de descrever os padrões espaciais dos erros da classificação, avaliando os efeitos das condições ambientais na acurácia de mapeamentos. Neste, pode-se, ainda, comparar diferentes técnicas de mapeamento ou esquemas de classificação, ou mesmo selecionar o melhor mapa da cobertura do solo para uma dada região (Stehman et al., 2000).

O conceito de acurácia de mapeamentos pode ter dois enfoques, sendo o primeiro posicional e o segundo temático. O primeiro trata da própria retificação geométrica da imagem, e seus processos de controle são basicamente os de cartografia. Já o enfoque temático refere-se, não aos aspectos posicionais dos dados espaciais e, sim, aos atributos. Enfim entende-se que a acurácia temática de um mapeamento refere-se à correspondência entre a classe atribuída a um *pixel* (ou outros tipos de unidade amostral) e a verdadeira classe a que este pertence, verificada em campo (Janssen, 1994, citado por Belart et al., 2001).

Para Machado Júnior et al. (2004), a acurácia temática reflete a fidelidade dos dados descritivos, com uma avaliação geral da identificação de entidades e atribuição de valores de atributo no conjunto de dados. A avaliação desta fidelidade é obtida fundamentalmente através de funções estatísticas que medem a concordância e a discordância dos atributos existentes em relação àqueles considerados verdadeiros.

A avaliação da acurácia é considerada a etapa final na avaliação da qualidade de um mapa temático de cobertura da terra. O protocolo padrão de avaliação da acurácia é o de comparar o mapa de cobertura com os pontos de referência em locais amostrados, em que o ponto de referência é assumido como

correto.

Segundo Stehman (2001), a avaliação da acurácia em mapeamentos temáticos segue os conceitos e teorias da estatística clássica, podendo, assim, utilizar a inferência estatística para generalizar ou extrapolar os resultados da amostra para toda a população do mapa temático da região mapeada.

Pela teoria clássica, há, então, que o valor médio de uma dada característica, em uma região amostrada, é igual ao valor da esperança desta característica em qualquer ponto dentro da região, com um erro de estimativa correspondente à variância dos dados amostrados (Cochran, 1977; Trangmar et al., 1987).

2.2.1 Matriz de confusão e Coeficientes de Concordância

A matriz de erro ou de confusão (TABELA 1) é a ferramenta mais utilizada em avaliações de acurácia de mapeamentos, além dos coeficientes dela derivados. Esta apresenta de forma sintetizada os resultados da classificação digital em relação aos dados de referência.

TABELA 1. Esquema de uma matriz de confusão

| | | Referência de campo | | | | | |
|------------|-------|------------------------|-----------|-----------|-----------|-----------|------------------------|
| Classes | | 1 | 2 | 3 | 4 | 5 | Total |
| Mapeamento | 1 | $P_{1,1}$ | | | | $P_{1,5}$ | $\sum_{j=1}^5 P_{j,i}$ |
| | 2 | | $P_{2,2}$ | | | | |
| | 3 | | | $P_{3,3}$ | | | |
| | 4 | | | | $P_{4,4}$ | | |
| | 5 | $P_{5,1}$ | | | | $P_{5,5}$ | |
| | Total | $\sum_{i=1}^5 P_{i,j}$ | | | | | $\sum_{i=1}^5 P$ |

Spiegel (1993) afirma que cada frequência observada na matriz corresponde, também, a uma frequência esperada a qual 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.

2.2.1.1 Coeficientes de Concordância Total

Dentre os coeficientes de concordância total, tem-se: o de Exatidão Global, proposto por Hellden & Stern (1980), o Kappa, proposto por Cohen (1960) e o Tau, proposto por Klecka (1980).

A Exatidão Global pode ser estabelecida com base na elaboração da matriz de confusão e avalia a precisão do classificador. De acordo com Watzlawick et al. (2002), esse valor representa a relação entre o número de pontos classificados corretamente e o número total de pontos amostrais utilizados. 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 está distribuída por meio das classes individuais.

$$P_0 = \frac{\sum_{i=1}^K x_{ii}}{N}, \quad (2)$$

onde:

P_0 = Coeficiente de Exatidão Global,

x_{ii} = elementos da diagonal principal da matriz de confusão;

N = número total de pontos da matriz de confusão.

Já o Kappa é um coeficiente de concordância para escalas nominais que mede 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 descreve, ainda, que todos os elementos da matriz de erros, diferentemente dos que assinalam apenas aqueles que se situam na

diagonal principal da mesma, estimando assim a soma da coluna e linha marginais (Cohen, 1960).

É expresso pela seguinte equação (Congalton, 1991; Ma & Redmond, 1995):

$$K = \frac{P_o - P_{co}}{1 - P_{co}}, \quad (3)$$

onde:

K = coeficiente de concordância Kappa;

$$P_o = \frac{\sum_{i=1}^M n_{ii}}{N}, \text{ representando a proporção de pontos de verdade terrestre}$$

concordantes;

$$P_{co} = \frac{\sum_{i=1}^M n_{i+} n_{+i}}{N^2}, \text{ representando a proporção de pontos atribuídos à}$$

classe ao acaso;

N = número total de pontos da matriz de confusão;

n = elemento da matriz de confusão;

n_{ii} = elementos da diagonal principal da matriz de confusão;

n_{i+} = somatório da linha para dada classe temática;

n_{+i} = somatório da coluna para dada classe temática.

Já o coeficiente de concordância Tau, utiliza a probabilidade *a priori*, visto que a probabilidade esperada é obtida antes da elaboração da matriz de confusão. Considera o método de classificação digital adotado na imagem, avaliando se a classificação digital foi efetuada com base na mesma probabilidade para todas as classes temáticas ou não (Klecka, 1980). É expresso pela equação:

$$\tau = \frac{P_o - P_{ce}}{1 - P_{ce}}, \quad (4)$$

onde:

τ = coeficiente de concordância Tau;

P_o = proporção de pontos de verdade terrestre concordantes;

P_{ce} = concordância esperada, sendo expresso pela relação $1/M$, em que M é o número de classes do mapeamento.

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 melhoria em relação à 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.

2.2.1.2 Coeficientes de Concordância para classes individuais

Outro meio de obter a acurácia é através dos coeficientes de concordância para classes individuais, sendo eles: Acurácia do Produtor e Acurácia do Usuário.

Acurácia do Produtor refere-se aos erros de exclusão ou omissão,, não sendo atribuídos à determinada classe os erros de classificação das demais classes. Sendo assim, a acurácia do produtor refere-se à probabilidade de determinada classe ter sido corretamente classificada de acordo com os pontos de referência de campo (Antunes & Lingnau, 2003). Pode variar entre 0 e +1 e é expresso por:

$$Ap_i = \frac{x_{ii}}{x_{i+}}, \quad (5)$$

onde:

x_{ii} = elementos da diagonal principal da matriz de confusão;

x_{i+} = somatório da linha para dada classe temática.

Já a acurácia do usuário refere-se aos erros de inclusão ou comissão (Congalton, 1991). Os de inclusão relacionam-se a todos os pontos atribuídos a uma determinada classe, oriundos de erros de classificação das outras classes, indicando, assim, a probabilidade de uma classe classificada representar a mesma em campo. Este índice também pode assumir valores entre 0 e +1 e expresso por (Antunes & Lingnau, 2003):

$$Au_i = \frac{x_{ii}}{x_{+i}}, \quad (6)$$

onde:

x_{ii} = elementos da diagonal principal da matriz de confusão;

x_{+i} = somatório da coluna para dada classe temática.

2.3 Determinação da Precisão da Classificação Temática

2.3.1 Tamanho da amostra

Belart et al. (2001) afirmam que a determinação do tamanho da amostra, para avaliar a acurácia temática e a precisão de um mapeamento, pode ocorrer de duas maneiras. A primeira e mais utilizada é obtida através do comportamento binominal dos dados aplicados a populações finitas e, a segunda, considerando um comportamento multinomial desses dados.

O tamanho da amostra de referência para a determinação da acurácia de mapeamentos temáticos pode ser obtido mediante a fixação da acurácia desejada e não de uma amostra piloto. Com isso, comete-se um erro de aproximação no dimensionamento da amostra necessária para representar corretamente o

parâmetro desejado (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 afasta muito da exatidão estimada pela amostragem piloto. Esse procedimento apresenta maior custo quando comparado ao dimensionamento por intermédio da exatidão fixada *a priori* (Costa & Brites, 2004).

Neste tipo de avaliação, em que os dados de Referência de campo coletados assumem uma distribuição binomial, a fórmula empregada para o cálculo do tamanho da amostra de referência é (Meyer, 1984; Jensen, 1986; Richards, 1986):

$$n = \frac{Z^2 \cdot p \cdot q}{E^2}, \quad (7)$$

em que:

n = tamanho da amostra de referência para cada classe temática;

Z = ponto crítico para um limite de confiança de inclusão para o parâmetro exatidão real, assumindo $Z \sim N(0,1)$;

p = exatidão desejada;

$q = 1-p$;

E = erro de amostragem admitido.

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 .

2.3.2 Estimadores

Existem inúmeras situações cujo objetivo de dado levantamento é estimar a proporção ou percentual de acertos, seguido do cálculo da precisão desta estimativa de determinada variável, na qual as observações se referem a uma variável discreta.

A teoria da amostragem em proporções baseia-se na ocorrência de uma série de Bernoulli, a qual caracteriza a distribuição binomial. A característica dessa distribuição é o controle da variável avaliada por zero ou um, respectivamente, para errado e certo (Ferreira, 2005).

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, esses serão categorizados por zero se não estiverem classificados corretamente ou por um se forem corretamente classificados. Assim, se Y_i fragmentos estiverem classificados corretamente, então, a proporção populacional será:

$$P = \frac{\sum_{i=1}^N Y_i}{N}$$

(8)

$$\bar{Y} = P = \frac{\sum_{i=1}^N Y_i}{N} \quad (9)$$

Os estimadores da Amostragem Casual Simples com Proporção estão apresentados na sequência.

- Média aritmética:

$$\bar{y} = p = \frac{\sum_{i=1}^n y_i}{n} \quad (10)$$

visto que \bar{y} é uma estimativa não tendenciosa da média da população e p uma proporção não tendenciosa de P .

- Variância:

$$s_y^2 = \frac{n \cdot p \cdot q}{n - 1} \quad (11)$$

- Variância da média da proporção:

$$s_{\frac{p}{p}}^2 = \frac{p \cdot q}{n - 1} \cdot \left(\frac{N - n}{N} \right), \quad (12)$$

- Erro padrão da média da proporção:

$$s_{\frac{p}{p}} = \sqrt{\frac{p \cdot q}{n - 1} \cdot \left(\frac{N - n}{N} \right)}, \quad (13)$$

visto que o fator de correção para população finita pode ser negligenciado se a razão $\frac{n}{N}$ for pequena ou menor que 7 % (Cochran, 1963).

- Precisão do levantamento:

$$E = p \cdot z \quad (14)$$

- Intervalo de confiança:

Cochran (1965), Péllico-Netto & Brena (1998) e Scolforo & Carvalho (2006) definem o intervalo de confiança como a determinação dos limites inferior e superior, no qual o valor verdadeiro da característica de interesse deve variar segundo um nível de significância previamente estabelecido. O intervalo de confiança é obtido pela expressão:

$$p - z \cdot s_{\frac{p}{p}} \leq P \leq p + z \cdot s_{\frac{p}{p}}.$$

No presente estudo a característica de interesse é a proporção de pontos corretamente classificados pelo mapeamento digital das classes temáticas presentes no “Mapeamento da Flora e dos Reflorestamentos de Minas Gerais”.

3 REFERÊNCIAS BIBLIOGRÁFICAS

ANTUNES, A. F. B.; LINGNAU, C. Uso dos índices de acurácia para avaliação de mapas temáticos obtidos por meios de classificação digital. In: CONGRESSO E FEIRA PARA USUÁRIO DE GEOPROCESSAMENTO, 3., 1997, Curitiba. **Anais...** Curitiba: Sagres, 1997. 1 CD-ROM.

BELART, P.; PINILLA, C.; ARIZA, F. J. **Control de calidad de clasificaciones de imágenes por el método de los itinerarios**. Jaén: Escuela Politécnica Superior Universidad de Jaén, 2002. 223 p.

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. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 8., 1996, Salvador. **Anais...** Salvador: INPE, 1996. p. 823-829.

COCHRAN, W. G. **Sampling techniques**. 3. ed. New York: J. Wiley, 1977. 448 p.

COCHRAN, W. G. **Técnicas de amostragem**. New York: J. Wiley, 1965. 428 p.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Measurment**, New York, v. 20, n. 1, p. 37-46, 1960.

CONGALTON, R. A review of assessing the accuracy of classification remotely sensed data. **Remote Sensing of Environment**, Washington, v. 37, n. 1, p. 35-36, Jan. 1991.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. Boca Raton: Lewis, 1999. 137 p.

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**, Rio de Janeiro, v. 56, n. 2, p. 151-155, ago. 2004.

COSTA, T. C. C.; BRITES, R. S. Tamanho da amostra de referencia para verificação de exatidão em imagens orbitais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 9., 1998, Santos. **Anais...** Santos: INPE, 1998. 1 CD-ROM.

DICKS, S. E.; LO, T. H. C. Evaluation of thematic map accuracy in a land-use and land-cover mapping program. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 55, n. 4, p. 475-478, Apr. 1990.

EDWARDS, T. C.; MOISEN, G. G.; CUTLER, D. R. Assessing map accuracy in a ecoregion-scale cover-map. **Remote Sensing Environment**, Washington, v. 63, n. 1, p. 73-83, Jan. 1998.

FERREIRA, D. F. **Estatística básica**. Lavras: UFLA, 2005. 644 p.

HELLDEN, U.; STERN, M. Evaluation of landsat imagery and digital data for monitoring desertification indicators in Tunisia. In: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT, 14., 1980, San Jose. **Proceedings...** Ann Arbor: Environmental Research Institute of Michigan, 1980. p. 1601-1611.

JANSSEN, L. F.; LUCAS, F. J. M.; WEL, F. J. M. van der. Accuracy assessment of satellite derived land-cover data: a review. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 60, n. 4, p. 419-426, Apr. 1994.

JENSEN, J. R. **Introductory digital image processing**. New Jersey: Prentice-Hall, 1986. 379 p.

KLECKA, W. R. **Discriminant analysis**. Beverly Hills: SAGE, 1980. 71 p.

LINGNAU, C. Potencialidades de imagens de satélite no monitoramento florestal. In: SEMINÁRIO DE ATUALIZAÇÃO EM SENSORIAMENTO REMOTO E SIG APLICADO A ENGENHARIA FLORESTAL, 3., 1998, Curitiba. **Anais...** Curitiba: FUPEF, 1998. v. 1, p. 45-50.

LINGNAU, C.; ANTUNES, A. F. B. Avaliação da acurácia de mapas temáticos oriundos de classificação digital de imagens. In: SIMPÓSIO BRASILEIRO DE GEOPROCESSAMENTO, 4., 1997, São Paulo. **Anais...** São Paulo: USP, 1997. 1 CD-ROM.

LINGNAU, C.; ANTUNES, A. F. B. Object oriented analysis and semantic network for high resolution image classification. **Boletim de Ciências Geodésicas**, Curitiba, v. 9, n. 2, p. 233-243, 2003.

LOBÃO, J. S. B.; ROCHA, W. J. S. F.; CHAVES, J. M.; NOLASCO, M. C. Uso da imagem de satélite para a discriminação de alvos urbanos e não-urbanos no município de Feira de Santana. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11., 2003, Belo Horizonte. **Anais...** Belo Horizonte: INPE, 2003. p. 1867-1874.

MA, Z.; REDMOND, R. Tau coefficients for accuracy assessment of classification of remote sensing data. **Photogrammetric Engineering and Remote Sensing**, Falls Church, v. 61, n. 4, p. 435-439, 1995.

MACHADO JÚNIOR, D. M.; SANTO, M. A. D.; LOCH, C. Considerações acerca de trabalhos em áreas de divisas de fusos UTM. In: CONGRESSO BRASILEIRO DE CADASTRO TÉCNICO MULTIFINALITÁRIO, 1., 2004, Florianópolis. **Anais...** Florianópolis: UFSC, 2004. p. 1-14.

MEYER, P. L. **Probabilidade**: aplicações à estatística. 2. ed. Rio de Janeiro: LTC, 1984. 426 p.

PÉLLICO NETTO, S.; BRENA, D. A. **Inventário florestal**. Curitiba: UFPR, 1997. 316 p.

PÉLLICO NETTO, S.; BRENA, D. A. **Inventário florestal**. Curitiba: UFPR, 1998. 245 p.

RICHARDS, J. A. **Remote sensing digital image analysis**: an introduction. Berlin: Springer-Verlag, 1986. 281 p.

SCOLFORO, J. R. S.; CARVALHO, L. M. T. **Mapeamento e inventário da flora nativa e dos reflorestamentos de Minas Gerais**. Lavras: UFLA, 2006. 288 p.

SCOLFORO, J. R. S.; MELLO, J. M. de. **Inventário florestal**. Lavras: UFLA/FAEPE, 1997. 344 p. (Texto Acadêmico).

SPIEGEL, M. R. **Estatística**. São Paulo: Makron Books, 1993. 643 p.

STEHMAN, S. V. Design, analysis, and inference for studies comparing thematic accuracy of classified remotely sensed data: a special case of map comparison. **Journal of Geographical Systems**, Berlin, v. 8, n. 2, p. 209-226, July 2006.

STEHMAN, S. V. Estimating the Kappa coefficient and its variance under stratified random sampling. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 62, n. 4, p. 401-402, Apr. 1996.

STEHMAN, S. V. Statistical rigor and practical utility in thematic map accuracy assessment. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 67, n. 6, p. 727-734, June 2001.

STEHMAN, S. V.; WICKHAM, J.; SMITH, J.; YANG, L. Thematic accuracy of the 1992 national land-cover data for the eastern United States: statistical methodology and regional results. **Remote Sensing of Environment**, Washington, v. 86, n. 4, p. 500-516, Aug. 2003.

STEHMAN, S. V.; WICKHAM, J. D.; YANG, L.; SMITH, J. H. Assessing the accuracy of large-area land cover maps: experiences from the multi-resolution land-cover characteristics (MRLC) project. In: INTERNATIONAL SYMPOSIUM ON SPATIAL ACCURACY ASSESSMENT IN NATURAL RESOURCES AND ENVIRONMENTAL SCIENCES, 4., 2000, Amsterdam. **Proceedings...** Amsterdam: The Netherlands, 2000. p. 601-608.

TRANGMAR, B. B.; YOST, R. S.; WADE, M. K.; UEHARA, G.; SUDJADI, M. Spatial variation of soil properties and rice yield on recently cleared land. **Soil Science Society of America Journal**, Madison, v. 51, n. 3, p. 668-674, Mar. 1987.

WATZLAWICK, L. F.; KIRCHNER, F. F.; SANQUETTA, C. R.; SCHUMACHER, M. V. Fixação de carbono em floresta ombrófila mista em diferentes estágios de regeneração. In: _____. **As florestas e o carbono**. Curitiba: UFPR, 2002. cap. 8, p. 153-173.

YANG, L.; STEHMAN, S. V.; SMITH, J. H.; CUTLER, D. R. Thematic accuracy of MRLC land-cover for the eastern United States. **Remote Sensing Environment**, Washington, v. 76, n. 3, p. 418-422, June 2001.

ZHU, Z.; YANG, L.; STEHMAN, S. V.; CZAPLEWSKI, R. L. Accuracy assessment for the U.S. geological survey regional land-cover mapping program: New York and New Jersey region. **Photogrammetric Engineering and Remote Sensing**, Falls Church, v. 66, n. 12, p. 1425-1435, Dec. 2000.

CAPÍTULO 2

AMOSTRAGEM PARA DETERMINAÇÃO DA ACURÁCIA E PRECISÃO DO MAPEAMENTO DA FLORA NATIVA E DOS REFLORESTAMENTOS DE MINAS GERAIS

RESUMO

BRITO, Alan de. **Amostragem para determinação da acurácia e precisão do Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais.** Lavras: UFLA 2009. 110p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Lavras, Lavras, MG.*

O presente trabalho foi realizado com o objetivo de avaliar uma metodologia de amostragem para a determinação da acurácia e da precisão da classificação temática do “Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais”. A área de estudo compreendeu todo o Estado. Procedeu-se à avaliação da acurácia tomando-se por base o mapa determinado pela classificação temática realizada para o Estado. As informações de referência foram obtidas mediante incursões terrestres, realizadas para retratar amostras das diversas formações vegetais da área. O procedimento de amostragem utilizado foi o de Amostragem Casual Simples (ACS) com proporção e estratificada. A unidade de controle do número de pontos de cada fisionomia foi a sub-bacia. A distribuição das unidades amostrais nas sub-bacias ocorreu de forma proporcional à área de cada classe temática presente nessas em relação à área total da classe presente no Estado. A escolha de polígonos - definidos por pontos amostrais aleatorizados - como unidade amostral, teve como base as dificuldades de acesso às áreas onde foi distribuída a amostra. O tamanho da amostra de referência por meio da exatidão fixada *a priori* (Meyer, 1984; Jensen, 1986; Richards, 1986), foi calculado considerando a fórmula de dimensionamento da amostra de referência. O procedimento de amostragem casual simples com proporção e estratificado contou ainda com o apoio do método dos itinerários utilizado por Belart et al. (2001). 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, mostraram-se adequadas, uma vez que a acurácia *a priori* foi inferior à determinada *a posteriori* para dois dos três coeficientes testados (Exatidão Global e Tau) e está contida no intervalo de confiança para a precisão real de todo o “Mapeamento da Flora e dos Reflorestamentos de Minas Gerais”. Baseando-se nos resultados, conclui-se que a definição de um sistema de amostragem detalhado é fundamental para o entendimento e determinação da acurácia e precisão da classificação com rigor estatístico. Além disso, a proporcionalização, obtida com a distribuição das unidades amostrais, de acordo com a área que as classes temáticas ocupam em determinada unidade de controle de amostragem, garante que todas as classes apresentem um número mínimo de

* Comitê de Orientação: José Marcio de Mello - UFLA (Orientador), Fausto Weimar Acerbi Júnior - UFLA, Luís Marcelo Tavares de Carvalho -UFLA (Co-orientadores).

unidades para atender erros pré-estabelecidos e causar intervalos de confiança para a proporção de unidades corretamente classificadas. Em virtude dos altos custos com a obtenção da amostra de referência, a escolha do procedimento de amostragem correto torna-se fundamental em qualquer processo de determinação da qualidade de mapeamentos de grandes áreas. A utilização das rodovias como base para se tomar a amostra é eficiente quando a malha viária da área em questão possui grandes extensões e boa distribuição espacial. Recomendam-se estudos, para obtenção de uma rede de colaboradores para realização de tomadas de dados de referência em campo, por unidades de controle definida que, apoiada por um protocolo de identificação das classes em campo, possam auxiliar na redução significativa dos custos da avaliação da acurácia e precisão de mapeamentos temáticos de grandes áreas.

Palavras-chave: amostragem, mapeamento temático, inventário florestal.

ABSTRACT

BRITO, Alan de. **Sampling to determine the accuracy and precision of the mapping of Native Flora and Reforestation of Minas Gerais**. Lavras: UFLA 2009. 110p. Dissertation (Master of Forest Science) – Federal University of Lavras, Lavras, MG.*

The objective of this study was to evaluate a method of sampling to determine the accuracy and precision of the thematic classification of "Mapping of Native Flora and Reforestation of Minas Gerais". The study included every state. This was the assessment of accuracy from the map generated by the thematic classification performed for the State. The referral information was obtained through land incursions, made to portray the various plant formation samples from the area. The Casual Simple Sampling (CSS) procedure was used, with proportion and stratification. The control unit of the number of points of each physiognomy was the sub-basin. The distribution of sampling units in the sub-basins occurred in proportion to the area of each class in this subject in relation to the total area of this class in the state. The choice of polygons - defined by randomized sampling points - as a sampling unit, was based on the difficulties of access to areas where the sample was distributed. The reference sample size set by *a priori* accuracy (Meyer, 1984; Jensen, 1986; Richards, 1986), was calculated from the formula for scaling of the reference sample. The sampling procedure with simple random and stratified ratio also had the support of the Itinerary Method used by Belart et al. (2001). The use of simple random sampling in proportion to the allocation of the sampling points and, using the accuracy given *a priori* to calculate the size of the sample, were appropriate, since the accuracy was less than the *a priori* determined *a posteriori* for two of the three factors tested (Global Accuracy and Tau), and is contained in the confidence interval for the accuracy of all the real " Mapping of Native Flora and Reforestation of Minas Gerais". From the results, it appears that the definition of a detailed sampling is fundamental to understanding and determining the accuracy and precision of classification, with statistical rigor. Furthermore, the proportioning obtained with the distribution of sample units according to the area that hold classes in a certain thematic unity of control sampling, ensures that all classes have a minimum number of units to meet pre-established errors and generate confidence intervals for the proportion of correctly classified units. Due to the high costs of obtaining the sample, the correct choice of the sampling procedure becomes essential in any process of determining the quality of mapping large areas. The use of roads as a basis for

* Steering Committee: José Marcio de Mello - UFLA (Advisor), Fausto Weimar Acerbi Júnior - UFLA, Luís Marcelo Tavares de Carvalho -UFLA (Co-Advisors).

taking the sample is efficient when the road network of the area is large and has good spatial distribution. Studies to obtain a network of collaborators to conduct in field reference data collection are recommended, defined by the control units which, supported by a protocol for the identification of classes in the field, can significantly help reduce the accuracy and precision assessment costs of thematic maps of large areas.

Keywords: sampling, thematic mapping, forest inventory.

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. O sensoriamento remoto permite a aquisição e interpretação de informações sobre um determinado objeto na superfície da Terra, baseando-se em um vantajoso ponto de vista: a visão “de cima”. Em inglês é muito comum a utilização da expressão *bird's eyes view* para qualificar as imagens de sensoriamento remoto. Outro fator importante diz respeito à alta correlação existente entre as variações presentes nas imagens de sensoriamento e as variações ao longo da superfície da Terra, fazendo do sensoriamento remoto uma excelente base para a produção de mapas de uso e cobertura do solo.

Este tipo de mapeamento, geralmente, divide os tipos de cobertura em classes temáticas específicas (Lingnau, 1998). Além disso, o mapeamento visa a representar, com certa precisão, a cobertura real encontrada em campo. Embora as técnicas mais sofisticadas sejam aplicadas, algumas limitações ainda são encontradas. Algumas delas, conforme citado por Lingnau & Antunes (2003) e Costa & Brites (2004), são: a aproximação espectral das classes, a presença de nuvens, a resolução espacial dos sensores, erros de retificação, erros posicional e temático, além dos erros provenientes da classificação digital de imagens. Dessa forma, é imprescindível a determinação da acurácia e precisão dos mapas temáticos provocados tomando-se como base o processamento de imagens de sensoriamento remoto.

Assim pode ser evidenciado na afirmativa de Congalton e Green (1999): *“As decisões de manejo dos recursos naturais (incluindo a decisão de não fazer nada) requer mapas, e efetivas decisões requer mapas de conhecida acuracidade”*.

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

A determinação da acurácia e da precisão é considerada a etapa final na avaliação da qualidade de um mapa temático. É realizada mediante a comparação entre o mapa de cobertura com os pontos de referência em locais amostrados, em que o ponto de referência é assumido como correto (Stehman, 2006). A sua determinação por meio de censo torna-se impraticável em grandes extensões territoriais, uma vez que seria necessário verificar todos os *pixels* (ou conjunto de *pixels*) da área de estudo. Portanto, o procedimento mais utilizado consiste na definição de amostras para fazer inferências sobre toda a população (Cochran, 1965). E, para que esta avaliação seja estatisticamente rigorosa, o procedimento de amostragem deve satisfazer o protocolo de amostragem probabilística e as estimativas devem ser estatisticamente consistentes, obtendo assim estimadores mais precisos com menor custo possível (Péllico-Neto & Brena, 1998).

A utilização de conceitos de teoria da amostragem para a avaliação da acurácia e precisão de mapas temáticos é de fundamental importância, tanto no que tange às suas aplicações práticas, quanto às técnicas e políticas. A prática está relacionada com um tamanho de amostra representativa e confiável da população, reduzindo os gastos de tempo e recursos (Scolforo e Mello, 1997). Já no que se refere às técnicas e políticas, isso 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 de todo o sistema de amostragem para este fim. O procedimento de amostragem, especificamente, é diretamente responsável pela determinação das medidas de acurácia e precisão do mapeamento e os custos envolvidos neste processo (Stehman, 2001).

O presente trabalho foi realizado com o objetivo de 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 o Estado de Minas Gerais, tendo seu território 58638072,69 hectares (FIGURA 1). Procedeu-se à avaliação da acurácia considerando o mapa causado pela classificação temática realizada para o Estado (Scolforo & Carvalho, 2006). Foram utilizadas 105 cenas do satélite Landsat, coletadas por meio dos sensores TM e ETM+, retratando três épocas do ano (primavera, verão e inverno).

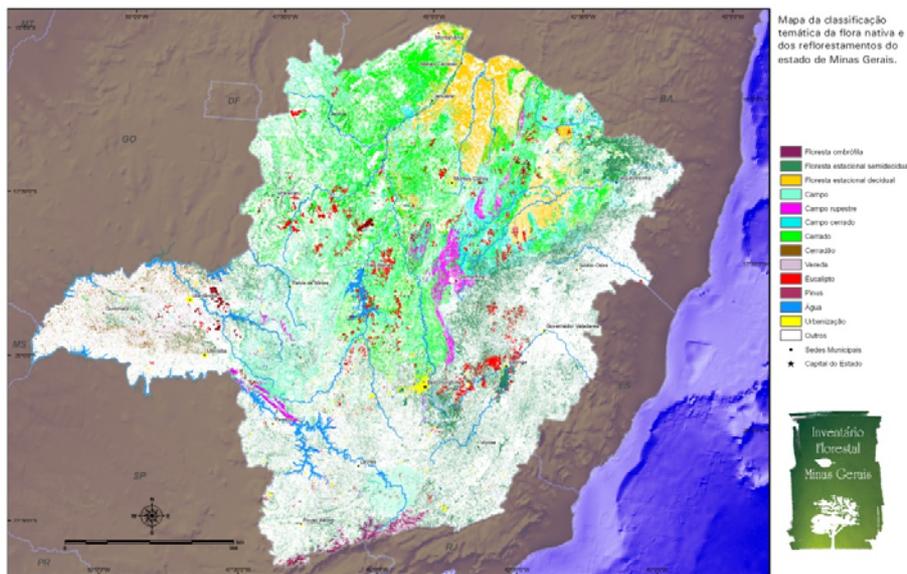


FIGURA 1. Mapa da classificação temática da flora nativa e dos reflorestamentos do estado de Minas Gerais

O método de classificação foi o supervisionado e o algoritmo classificador utilizado foi o de árvores de decisão (Carvalho et al., 2004). Após a

classificação das imagens, classes com pastagem, culturas agrícolas e solo exposto foram agrupadas em uma única classe chamada “outros”. As demais classes compreendidas foram: floresta ombrófila, floresta estacional semidecidual, floresta estacional decidual, cerradão, cerrado, campo cerrado, campo, veredas, eucalipto e Pinus. A área de cada classe temática, em porcentagem, presentes no Estado estão discriminadas na TABELA 1, conforme Scolforo & Carvalho (2006).

TABELA 1. Área ocupada, em hectare e porcentagem, por classe temática

| FISIONOMIA | 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 | 0,6 |
| 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 por meio 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, assegurando, dessa forma, o princípio de independência dos dados.

O procedimento de amostragem utilizado foi o de Amostragem Casual Simples (ACS) com proporção e estratificada. A adoção dessa estratégia de amostragem atribui a cada classe temática um estrato e, dentro desses estratos,

realiza-se a amostragem casual simples. Assim, o número de estratos será o mesmo de classes, totalizando 14. Este tipo de procedimento foi escolhido em função da garantia de que todas as classes que compõem o Mapeamento fossem amostradas, inclusive as menores. A unidade de controle do número de pontos de cada fisionomia foi a sub-bacia.

A distribuição das unidades amostrais nas sub-bacias ocorreu de forma proporcional à área de cada classe temática nelas presente em relação à área total da classe presente no Estado. Assim, foi realizado o cálculo do número de polígonos a serem amostrados por sub-bacia para cada classe, garantindo que sub-bacias com maior área de determinada classe recebesse maior número de pontos e, bacias com quantidade menor dessa classe, uma menor quantidade de unidades amostrais referentes a ela.

A escolha de polígonos - definidos por pontos amostrais aleatorizados - como unidade amostral, teve como base as dificuldades de acesso às áreas em que foi distribuída a amostra, solução apresentada por Janssen (1994). Além disso, a defasagem temporal existente entre a data da classificação (2006) e a da coleta de dados em campo (2007) foi preponderante para a escolha do polígono como unidade analisada. Foram analisados apenas polígonos com tamanho aproximado de 10 *pixels* (9000m²) e nas anotações de campo foi considerado o tamanho do polígono observado. Os demais polígonos foram descartados no momento da análise dos dados.

O tamanho da amostra de referência por meio da exatidão, fixada *a priori* (Meyer, 1984; Jensen, 1986; Richards, 1986), foi calculado baseando-se na fórmula de dimensionamento da amostra de referência abaixo apresentada.

$$n = \frac{Z^2 \cdot p \cdot q}{E^2}, \quad (1)$$

em que:

n = tamanho da amostra de referência;

Z = ponto crítico para um limite de confiança de inclusão para o parâmetro exatidão real, assumindo $Z \sim N(0,1)$;

p = exatidão desejada;

$q = 1-p$;

E = erro de amostragem admitido.

A exatidão p utilizada foi fixada em 85% e a mínima exigida para mapeamentos temáticos (Jensen, 1986). O erro admissível empregado foi de 5% com probabilidade de 95%.

O procedimento de amostragem casual simples com proporção e estratificado contou, ainda, com o apoio do método dos itinerários, utilizado por Belart et al. (2001), uma vez que a distribuição da amostra ocorreu nas imediações de rodovias.

Assim, contando com uma base digital de rodovias do Departamento de Estradas e Rodagem (DER), foram selecionadas aleatoriamente 50%, 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, essa demarcação, 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 da bacia do Rio Grande classificadas como obrigatórias, deslocamento e adicionais

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

No caso de alguma sub-bacia não possuir nenhuma rodovia da base digital, foi procurada alguma estrada que pudesse ser amostrada. No campo foram coletadas as seguintes informações: sentido de deslocamento na rodovia; coordenadas geográficas dos pontos amostrais; classe temática ocorrente; lado da rodovia em que o fragmento se encontrava. As classes temáticas foram visualmente identificadas em campo.

Quando possível ou necessário, informações adicionais dos fragmentos amostrados também foram coletadas. Essas informações garantiram maior

confiabilidade à amostra, além de formar um banco de dados das classes em todo o Estado: tamanho (grande ou pequeno) e formato do fragmento (estreito ou largo); distância aproximada entre a rodovia e o fragmento, quando este possuía grandes dimensões; idade aproximada e presença de sub-bosque denso com espécies nativas, quando o fragmento em questão pertencia às classes Eucalipto ou Pinus; proximidade a outros fragmentos, com os quais possa ter ocorrido confusão na classificação temática.

Como a classe Floresta Decídua apresenta deciduidade e proximidade espectral à Floresta Estacional Semidecidual, a coleta da amostra para esta foi realizada no período em que a formação se encontrava sem folhas, para evitar confusão no campo.

Foram amostrados aproximadamente 50000km às margens das rodovias obrigatórias, de deslocamento e adicionais. Deste total, aproximadamente 10000km foram amostrados em rodovias não pavimentadas.

2.3 Avaliação da acurácia

Usando os dados de referência de campo, foi possível determinar a proporção de polígonos corretamente classificados em relação ao número total de polígonos visitados.

A ferramenta utilizada neste trabalho para determinar a acurácia do mapa temático foi a matriz de confusão ou matriz de erro e os índices dela derivados, conforme citados por Stehman (1997), Congalton & Green (1999) e Lingnau & Antunes (2003). A matriz de confusão é a forma mais efetiva de determinar a acuracidade de uma classificação temática, uma vez que as acuracidades individuais de cada classe mapeada são representadas por meio de erros de comissão e omissão. O erro de comissão é definido como a inclusão de

uma área em uma categoria (classes temáticas) a qual ela não pertence. O erro de omissão é definido como a exclusão de uma área a qual ela realmente pertence.

Foram utilizados tanto coeficientes de concordância para todo o Mapeamento (Concordância Total) quanto os que avaliam a concordância para as classes mapeadas, individualmente. Dentre os coeficientes de concordância total, utilizou-se o de Exatidão Global, proposto por Hellden & Stern (1980), o Kappa, proposto por Cohen (1960) e o Tau, proposto por Klecka (1980). Para a avaliação das classes individuais, os coeficientes utilizados foram: Acurácia do Produtor e Acurácia do Usuário (Congalton & Green, 1999).

2.4 Determinação da Precisão

A determinação da precisão para o Mapeamento da Flora Nativa e Reflorestamentos de Minas Gerais foi realizada com base no procedimento de amostragem casual simples com proporções. Em cada classe de vegetação foram distribuídas as unidades amostrais aleatoriamente.

A distribuição binomial ou Bernoulli da característica de interesse *proporção de polígonos corretamente classificados* determinou a utilização deste procedimento com a utilização de proporções, uma vez que os dados obtidos em campo podem apresentar apenas dois resultados: 0 quando o polígono está erroneamente classificado e; 1 se o polígono foi corretamente classificado, quando comparado o Mapeamento à referência de campo. Os estimadores deste procedimento podem ser encontrados em Cochran (1965), Péllico-Netto & Brena (1997) e Scolforo & Carvalho (2006).

Foram considerados nas análises efetuadas somente os polígonos que representavam uma classe apenas. Os polígonos que apresentavam transição entre classes como, por exemplo, Campo - Campo Cerrado - Campo Rupestre, foram excluídos da base de dados. Esta ação foi realizada para conferir maior segurança e confiabilidade à amostra.

3 RESULTADOS E DISCUSSÃO

Conforme o valor da exatidão fixada *a priori* ($p=85\%$), a amostra de referência foi de aproximadamente 196 pontos para todas as classes do Mapeamento, para um valor crítico de Z de 1,96.

No presente estudo, a fim de captar ao máximo a variação das classes temáticas presentes no Estado e, aproveitando a malha viável favorável, foram coletados 11942 pontos distribuídos nas 14 classes. O mapa da FIGURA 3 contém a posição espacial de cada ponto no Estado de Minas Gerais.

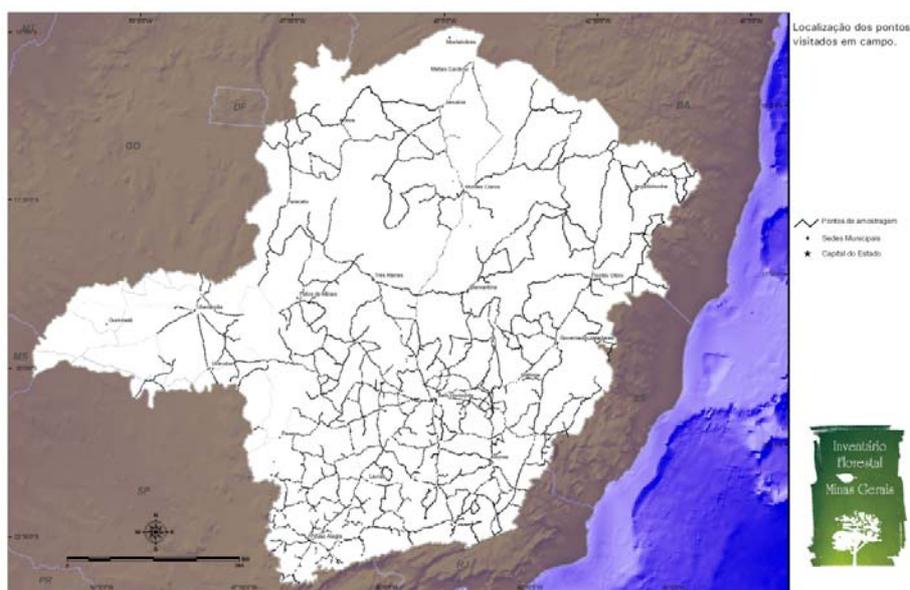


FIGURA 3. Localização dos pontos visitados em campo

3.1 Avaliação da acurácia

Na TABELA 2 está apresentada a matriz de confusão para a avaliação da acurácia do mapeamento temático do Estado de Minas Gerais.

TABELA 2. Matriz de confusão para avaliação da acurácia das classes temáticas: A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | | | | | | | | | | | |
|------------|--------|---------------------|------------|------------|-----------|-----------|------------|------------|------------|-------------|-----------|-------------|-----------|------------|-----------|--------------|
| Mapeamento | Classe | A | C | CC | CR | CO | CT | E | FD | FS | FO | O | P | U | V | Total |
| | | A | 171 | | | | 1 | | | | | | 1 | | | |
| | C | | 109 | 6 | 15 | | 16 | 14 | 5 | 15 | 4 | 61 | | 1 | 3 | 249 |
| | CC | | | 192 | | | 11 | 5 | 38 | 9 | | 14 | | | | 269 |
| | CR | | | 6 | 89 | | 3 | 3 | | 2 | | 1 | | | | 104 |
| | CO | | | | | 64 | | | | | | 4 | | | 3 | 71 |
| | CT | 1 | 1 | 2 | | | 897 | 11 | 34 | 46 | | 31 | | 1 | 13 | 1037 |
| | E | | 1 | | | | 9 | 389 | | 14 | | 9 | 3 | | | 425 |
| | FD | | | 2 | | | 8 | | 539 | | | 10 | | | | 559 |
| | FS | | 1 | 7 | 1 | | 49 | 70 | 26 | 1495 | 53 | 37 | 1 | | 4 | 1744 |
| | FO | | | | | | | 1 | | 2 | 68 | | | | | 71 |
| | O | 22 | 24 | 25 | 38 | 4 | 93 | 211 | 106 | 192 | 24 | 5720 | 7 | 33 | 21 | 6520 |
| | P | | 1 | | | | 1 | 16 | | | | 1 | 58 | | | 77 |
| | U | | | | | | | 1 | | 1 | | 5 | | 572 | | 579 |
| | V | | | | | 1 | | 1 | | 3 | | | | | 59 | 64 |
| | Total | 194 | 137 | 240 | 143 | 70 | 1087 | 722 | 748 | 1779 | 149 | 5894 | 69 | 607 | 103 | 11942 |

Considerando a matriz de confusão, foram determinados os índices de concordância total (FIGURA 4).

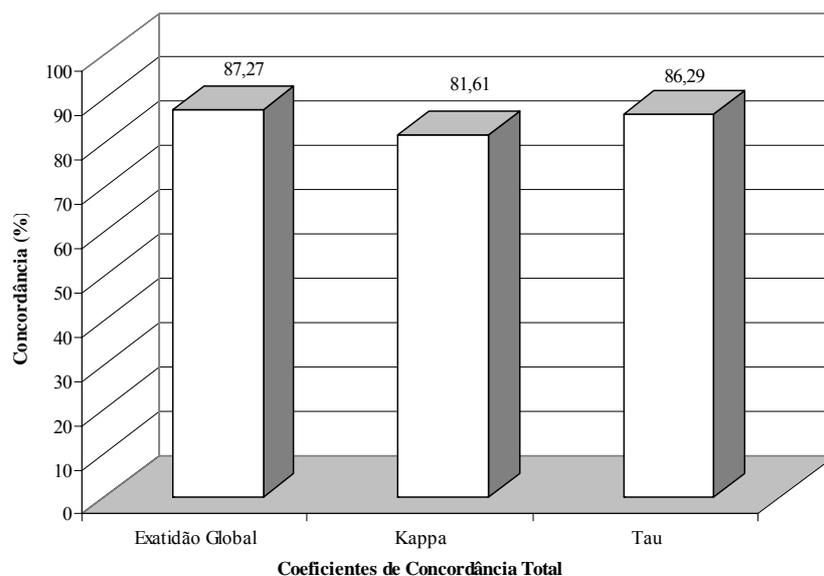


FIGURA 4. Valores dos coeficientes Exatidão Global, Kappa e Tau

Conforme o valor do índice de Exatidão Global (87,27%), afirma-se que a exatidão *a posteriori* foi superior à exatidão *a priori* (85%). Resultado esse que é confirmado pelo coeficiente Tau, portanto, o resultado observado foi melhor que o esperado.

Como esperado, o Kappa (81,61%) para toda a classificação foi menor que o valor de Exatidão Global (87,27%) 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 (Cohen, 1960).

O coeficiente de concordância Tau (86,29%) 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.

A TABELA 3 mostra a Acurácia do Produtor e do Usuário, que medem a concordância para as classes individuais.

TABELA 3. Acurácia do Produtor e do Usuário para as classes individuais

| CLASSE TEMÁTICA | ACURÁCIA (%) | |
|----------------------------------|--------------|---------|
| | PRODUTOR | USUÁRIO |
| Água | 88,14 | 98,84 |
| Campo | 79,56 | 98,79 |
| Campo Cerrado | 80,00 | 96,42 |
| Campo Rupestre | 62,24 | 95,77 |
| Cerradão | 91,43 | 92,19 |
| Cerrado | 82,52 | 91,53 |
| Eucalipto | 53,88 | 90,14 |
| Floresta Decidual | 72,06 | 87,73 |
| Floresta Estacional Semidecidual | 84,04 | 86,50 |
| Floresta Ombrófila | 45,64 | 85,72 |
| Outros | 97,05 | 85,58 |
| Pinus | 84,06 | 75,32 |
| Urbanização | 94,23 | 71,38 |
| Veredas | 57,28 | 43,78 |

Nota-se que a classe Outros foi a que apresentou maior valor de Acurácia do Produtor com 97,05%. Este valor representa 97,05% de todos os polígonos encontrados em campo como Outros estão representados no

Mapeamento como Outros, ou seja, houve um erro de omissão para esta classe de apenas 2,95%. A classe temática que apresentou o menor valor de acuracidade do Produtor foi a Floresta Ombrófila, indicando que apenas 45,64% dos polígonos deste tipo de vegetação existente em campo foram mapeados como Floresta Ombrófila, portanto, omitidos mais de 50% (54,36%) dos polígonos referentes a esta classe. A classe Eucalipto também apresentou um baixo valor de acuracidade do Produtor, com 53,88% de acerto. A confusão ocorrida nessas duas classes pode ter explicação em características do relevo e do estágio de sucessão, respectivamente. A classe Floresta Ombrófila tem sua ocorrência no estado relacionada normalmente a locais de maior inclinação do terreno. Esta inclinação pode provocar presença de sombreamento nos fragmentos imageados e conseqüente confusão na sua classificação. Para a classe Eucalipto, foi notado em campo que, áreas abandonadas em regeneração natural, com sub-bosque formado por espécies nativas, ou ainda áreas em que se realiza talhadia e o dossel é mais denso, podem ter sido os motivos de confusão com outras classes.

Com base na análise dos valores referentes à Acurácia do Usuário, observa-se que a classe Água apresentou o maior valor para este coeficiente (98,84%), seguida da classe Urbanização (98,79%). Analisando a Água, cujo valor representa 98,84% dos polígonos destas classes encontrados no campo verifica-se que estão corretamente classificados no Mapeamento. A classe Campo foi a que apresentou um menor número de polígonos mapeados corretamente quando comparados aos encontrados em campo, com 43,78%, significando um erro de comissão de 56,22%, ou seja, vários polígonos pertencentes a outras classes (principalmente as pastagens inseridas na classe Outros) foram erroneamente mapeadas como campo.

Um fato de significativa importância, porém, é que as duas maiores classes de vegetação, em termos de ocupação do solo no Estado, o Cerrado

(5.504.795ha) e a Floresta estacional Semidecidual (5.181.382ha) apresentaram valores muito bons em termos de acuracidade do Produtor (82,52% para o Cerrado e 84,04% para a Floresta Estacional Semidecidual) e do Usuário (90,14% para o Cerrado e 86,50% para a Floresta Estacional Semidecidual), significando baixos erros de comissão e omissão para estas classes.

3.2 Amostragem Casual Simples - Estimação de Proporções

A análise com estimação por proporções permite criar uma medida de confiança (precisão) quanto à proporção real dos polígonos corretamente classificados. As estatísticas provocadas com os estimadores da ACS com proporção estão apresentadas na TABELA 4.

TABELA 4. Estimadores da Amostragem Casual Simples com proporção

| ESTATÍSTICA | ESTIMATIVA DO PARÂMETRO |
|------------------------|-------------------------|
| Média | 0,8727 |
| Variância | 0,11109 |
| Desvio Padrão | 0,33330 |
| Variância da Média | 0,0000093 |
| Desvio Padrão da Média | 0,00305 |
| Erro (%) | 0,68498 |

O intervalo de confiança com 95% de probabilidade de acerto determinado para a precisão real do “Mapeamento da Flora Nativa e dos Reflorestamentos de Minas Gerais” é:

$$86,67\% \leq \text{Proporção Real} \leq 87,87\%.$$

Tomando-se por base a determinação do intervalo de confiança, afirma-se que a proporção estimada de polígonos corretamente classificados ou concordantes entre referência de campo e o Mapeamento está situada entre 86,67% e 87,87% de todos os polígonos da classificação temática. A proporção real de polígonos classificados corretamente, refere-se à exatidão global real do

mapeamento. Portanto, pelo presente estudo, pode-se afirmar que a precisão da exatidão global real estimada foi inferior a 1% que implica numa estimativa altamente precisa.

4 CONCLUSÕES

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, mostraram-se adequadas, uma vez que a acurácia *a priori* foi inferior à determinada *a posteriori* para dois dos três coeficientes testados (Exatidão Global e Tau) e está contida no intervalo de confiança para a precisão real de todo o “Mapeamento da Flora e dos Reflorestamentos de Minas Gerais”.

A definição de um sistema de amostragem, muitas vezes não detalhados em estudos da qualidade de mapeamentos temáticos, é fundamental para o entendimento e determinação da acurácia e precisão da classificação, com rigor estatístico.

A proporcionalização, obtida com a distribuição das unidades amostrais, de acordo com a área que as classes temáticas ocupam em determinada unidade de controle de amostragem, garante que todas as classes apresentem um número mínimo de unidades para atender erros pré-estabelecidos e criar intervalos de confiança para a proporção de unidades corretamente classificadas. Estes intervalos são estatisticamente mais consistentes na avaliação da precisão de classificações temáticas.

Em virtude dos altos custos, com a obtenção da amostra de referência, a escolha do procedimento de amostragem correto torna-se fundamental em qualquer processo de determinação da qualidade de mapeamentos temáticos de grandes áreas. Estudos envolvendo este tipo de área devem ser realizados com maior frequência e detalhamento para a obtenção de metodologias mais viáveis economicamente.

A utilização das rodovias como base para tomar a amostra é eficiente quando a malha viária da área em questão possui grandes extensões e boa

distribuição espacial, favorecendo a distribuição das unidades amostrais em toda a área de estudo. Este fato foi detectado no presente estudo.

Recomendam-se estudos para obtenção de uma rede de colaboradores para realização de tomadas de dados de referência em campo, por unidades de controle definidas, como bacias, sub-bacias, núcleos regionais ou municipais de órgãos ambientais. Esta rede, apoiada por um protocolo de identificação das classes em campo, pode auxiliar na redução significativa dos custos da avaliação da acurácia e precisão de mapeamentos temáticos de grandes áreas.

5 REFERÊNCIAS BIBLIOGRÁFICAS

- BELART, P.; PINILLA, C.; ARIZA, F. J. **Control de calidad de clasificaciones de imágenes por el método de los itinerarios**. Jaén: Escuela Politécnica Superior Universidad de Jaén, 2002. 223 p.
- CARVALHO, L. M. T.; CLEVERS, J. G. P. W.; SKIDMORE, A. K.; JONG, S. M. de. Selection of imagery data and classifiers for mapping Brazilian semideciduous Atlantic forests. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 5, n. 3, p. 173-186, 2004.
- COCHRAN, W. G. **Técnicas de amostragem**. New York: J.Wiley, 1965. 428 p.
- COHEN, J. A coefficient of agreement for nominal scales. **Educational and Measurement**, New York, v. 20, n. 1, p. 37-46, 1960.
- CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. Boca Raton: Lewis, 1999. 137 p.
- 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**, Rio de Janeiro, v. 56, n. 2, p. 151-155, ago. 2004.
- HELLDEN, U.; STERN, M. Evaluation of landsat imagery and digital data for monitoring desertification indicators in Tunisia. In: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT, 14., 1980, San Jose. **Proceedings...** Ann Arbor: Environmental Research Institute of Michigan, 1980. p. 1601-1611.
- JANSSEN, L. F.; LUCAS, F. J. M.; WEL, F. J. M. van der. Accuracy assessment of satellite derived land-cover data: a review. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 60, n. 4, p. 419-426, Apr. 1994.
- JENSEN, J. R. **Introductory digital image processing**. New Jersey: Prentice-Hall, 1986. 379 p.

KLECKA, W. R. **Discriminant analysis**. Beverly Hills: SAGE, 1980. 71 p.

LINGNAU, C. Potencialidades de imagens de satélite no monitoramento florestal. In: SEMINÁRIO DE ATUALIZAÇÃO EM SENSORIAMENTO REMOTO E SIG APLICADO A ENGENHARIA FLORESTAL, 3., 1998, Curitiba. **Anais...** Curitiba: FUPEF, 1998. v. 1, p. 45-50.

LINGNAU, C.; ANTUNES, A. F. B. Object oriented analysis and semantic network for high resolution image classification. **Boletim de Ciências Geodésicas**, Curitiba, v. 9, n. 2, p. 233-243, 2003.

MEYER, P. L. **Probabilidade**: aplicações à estatística. 2. ed. Rio de Janeiro: LTC, 1984. 426 p.

PÉLLICO NETTO, S.; BRENA, D. A. **Inventário florestal**. Curitiba: UFPR, 1997. 316 p.

PÉLLICO NETTO, S.; BRENA, D. A. **Inventário florestal**. Curitiba: UFPR, 1998. 245 p.

RICHARDS, J. A. **Remote sensing digital image analysis**: an introduction. Berlin: Springer-Verlag, 1986. 281 p.

SCOLFORO, J. R. S.; CARVALHO, L. M. T. **Mapeamento e inventário da flora nativa e dos reflorestamentos de Minas Gerais**. Lavras: UFLA, 2006. 288 p.

SCOLFORO, J. R. S.; MELLO, J. M. de. **Inventário florestal**. Lavras: UFLA/FAEPE, 1997. 344 p. (Texto Acadêmico).

STEHMAN, S. V. Design, analysis, and inference for studies comparing thematic accuracy of classified remotely sensed data: a special case of map comparison. **Journal of Geographical Systems**, Berlin, v. 8, n. 2, p. 209-226, July 2006.

STEHMAN, S. V. Selecting and interpreting measures of thematic classification accuracy. **Remote Sensing of Environment**, Washington, v. 62, n. 1, p. 77-89, Oct. 1997.

STEHMAN, S. V. Statistical rigor and practical utility in thematic map accuracy assessment. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 67, n. 6, p. 727-734, June 2001.

CAPÍTULO 3

ACURÁCIA E PRECISÃO DO MAPEAMENTO DA FLORA NATIVA E DOS REFLORESTAMENTOS DE MINAS GERAIS POR CLASSE TEMÁTICA E BACIA HIDROGRÁFICA

RESUMO

BRITO, Alan de. **Acurácia e precisão do Mapeamento da Flora Nativa e dos reflorestamentos de Minas Gerais por classe temática e bacia hidrográfica.** Lavras: UFLA 2009. 110p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Lavras, Lavras, MG.*

Neste trabalho objetivou-se determinar a precisão da classificação dos estratos definidos pelas classes temáticas presentes no Mapeamento da Flora Nativa e Reflorestamentos de Minas Gerais, criando intervalos de confiança para a proporção de polígonos corretamente classificados. Além disso, objetivou-se determinar a acurácia e a precisão da classificação para os estratos definidos pelas bacias hidrográficas, presentes no mapeamento. A área de estudo compreendeu todo o Estado. Procedeu-se à avaliação da acurácia em função do mapa provocado pela classificação temática realizada para o Estado. As informações de referência foram obtidas por meio de incursões terrestres, realizadas para retratar amostras das diversas formações vegetais da área. O procedimento de amostragem utilizado foi o de Amostragem Casual Simples (ACS) com proporção e estratificada. A unidade de controle do número de pontos de cada fisionomia foi a sub-bacia. A distribuição das unidades amostrais nas sub-bacias ocorreu de forma proporcional à área de cada classe temática nelas presente em relação à área total da classe presente no Estado. A escolha de polígonos - definidos por pontos amostrais aleatorizados - como unidade amostral, teve como base as dificuldades de acesso às áreas nas quais foi distribuída a amostra. O tamanho da amostra de referência por meio da exatidão fixada *a priori* (Meyer, 1984; Jensen, 1986; Richards, 1986), foi calculado com base na fórmula de dimensionamento da amostra de referência. O procedimento de amostragem casual simples com proporção e estratificado contou, ainda, com o apoio do método dos itinerários utilizado por Belart et al. (2001). Constatou-se, no presente estudo, que algumas classes não apresentaram proporção de polígonos corretamente classificados superior a 85%, resultado esperado quando se avalia a exatidão por classe temática. Para as classes Eucalipto e Floresta Ombrófila, menos do que 50% dos polígonos visitados em campo foram compatíveis com o Mapeamento. Estas classes merecem atenção especial no processo de classificação nas condições em que o trabalho foi desenvolvido. Analisando a precisão da exatidão avaliada para estas classes, detectou-se que ela se apresentou baixa, em função do número de polígonos classificados incorretamente. Foi detectado no estudo que, mesmo a tomada de poucas unidades amostrais criam intervalos de confiança para a proporção de polígonos

* Comitê de Orientação: José Marcio de Mello - UFLA (Orientador), Fausto Weimar Acerbi Júnior - UFLA, Luís Marcelo Tavares de Carvalho -UFLA (Co-orientadores).

corretamente classificados superiores a 85%. Embora exista um déficit no número de unidades amostrais para essas bacias, as poucas tomadas retrataram a variabilidade ocorrente com a variável de interesse, possibilitando assim, intervalos com proporções maiores que 85%, não garantindo, porém, um erro máximo de 10%. Dos intervalos de confiança para as 14 bacias amostradas, quatro deles não englobam a exatidão *a priori* de 85%, em razão dos altos valores de erro nestas bacias. Para todas as bacias hidrográficas, o coeficiente Exatidão Global foi maior que o coeficiente Kappa, verificando-se o fato do último levar em consideração toda a matriz de confusão. Além disso, o Kappa, para todas as bacias amostradas, foi menor quando comparado ao coeficiente Tau. Por sua vez, o Tau determinado foi menor que a Exatidão Global para todas as bacias, apresentando valores mais próximos aos da Exatidão Global quando o comparado com o Kappa.

Palavras-chave: mapeamento temático, classe temática, bacia hidrográfica.

ABSTRACT

BRITO, Alan de. **Accuracy and precision of the mapping of Native Flora and Reforestation of Minas Gerais by class theme and basin.** Lavras: UFLA 2009. 110p. Dissertation (Master of Forest Science) – Federal University of Lavras, Lavras, MG.*

This work aimed to determine the accuracy of the classification of strata defined by classes in the thematic mapping of Native Flora and Reforestation of Minas Gerais, generating confidence intervals for the proportion of correctly classified polygons. Furthermore, it was aimed to determine the accuracy and precision of classification for the strata defined by river basins, in the mapping. The study included every state. This was the assessment of accuracy from the map generated by the thematic classification performed for the State. The referral information was obtained through land incursions, made to portray the various plant formation samples from the area. The sampling procedure used was Casual Simple Sampling (CSS) with proportion and stratification. The control unit of the number of points of each physiognomy was the sub-basin. The distribution of sampling units in the sub-basins occurred in proportion to the area of each class in this subject in relation to the total area of this class in the state. The choice of polygons - defined by randomized sampling points - as a sampling unit, was based on the difficulties of access to areas where the sample was distributed. The reference sample size set by *a priori* accuracy (Meyer, 1984; Jensen, 1986; Richards, 1986), was calculated from the formula for scaling of the reference sample. The sampling procedure with simple random and stratified ratio also had the support of the Itinerary Method used by Belart et al. (2001). It was verified in this study that some classes had no share of correctly classified polygons over 85%, as was expected, to assess the accuracy by class theme. For Eucalyptus and Rain Forest class, less than 50% of polygons visited in the field were consistent with the mapping. These classes deserve special attention in the process of classification under the conditions where the work was developed. Analyzing the accuracy of the estimated accuracy for these classes, it was found that it presented low, in function of the number of polygons misclassified., It was detected in the study that even taking a few sample units generates confidence intervals over 85% for the proportion of correctly classified polygons. However, although there is a deficit in the number of sampling units for these basins, the few taken portray the variability occurrent

* Steering Committee: José Marcio de Mello - UFLA (Advisor), Fausto Weimar Acerbi Júnior - UFLA, Luís Marcelo Tavares de Carvalho -UFLA (Co-Advisors).

with the variable of interest, thus enabling, intervals with ratios greater than 85%, but not guaranteeing a maximum error of 10%. Of the confidence intervals for the 14 basins sampled, four did not include the *a priori* accuracy of 85% due to high error values in these basins. For all watersheds, the Global Accuracy coefficient was greater than the Kappa coefficient, with the fact that the latter take into account the full array of confusion. In addition, the Kappa, for all basins sampled, was lower when compared to the Tau coefficient. In turn, the determined Tau coefficient was less than the Global Accuracy for all basins, showing values close to the Global Accuracy when compared to the Kappa.

Keywords: thematic mapping, class theme, basin.

1 INTRODUÇÃO

Mapeamentos temáticos realizados com base em imagens de satélite são importantes ferramentas para a determinação de políticas públicas ambientais e conservacionistas. Entretanto, existe a necessidade de informações sobre a confiabilidade destes mapeamentos, uma vez que são geralmente tomados como corretos pelos diversos setores da sociedade civil e da administração pública que os utilizam.

Para avaliar esta confiabilidade, diversos tipos de procedimentos de amostragem podem ser utilizados (Stehman, 1996, 2001; Belart et al., 2001). O mais comumente adotado é o de amostragem aleatória simples. Geralmente este procedimento utiliza algum tipo de estratificação, com o objetivo de criar medidas mais confiáveis da acurácia e precisão uma vez que, ao utilizar estratificação, pode-se assegurar o mínimo de unidades amostrais presentes em cada estrato. Entre os tipos de estratificação podem-se mencionar como os mais comumente utilizados os fundamentados em ecorregiões, geralmente as classes temáticas e os por regiões geográficas (Edwards et al., 1998; Stehman, 2001).

Medidas de acurácia e precisão, tanto por classes temáticas quanto por bacias hidrográficas se fazem necessárias uma vez que estes dois tipos de estratos são bases de qualquer planejamento quanto às ações a serem tomadas em determinada área de estudo ou empreendimento ambiental.

O presente trabalho foi realizado com o objetivo de determinar a precisão da classificação dos estratos definidos pelas classes temáticas presentes no mapeamento, criando intervalos de confiança para a proporção de polígonos corretamente classificados. Além disso, foi conduzido, também, com o objetivo de determinar a acurácia e a precisão da classificação para os estratos definidos pelas bacias hidrográficas presentes no “Mapeamento da Flora Nativa e Reflorestamentos de Minas Gerais”.

2 MATERIAL E MÉTODOS

2.1 Caracterização geral da área

A área de estudo compreendeu todo o Estado de Minas Gerais (FIGURA 1). Procedeu-se à avaliação da acurácia e da precisão com base no mapa criado pela classificação temática realizada para o Estado (Scolforo & Carvalho, 2006). Nesta avaliação foram utilizadas 105 cenas do satélite Landsat, coletadas por meio dos sensores TM e ETM+, retratando três épocas do ano (primavera, verão e inverno).

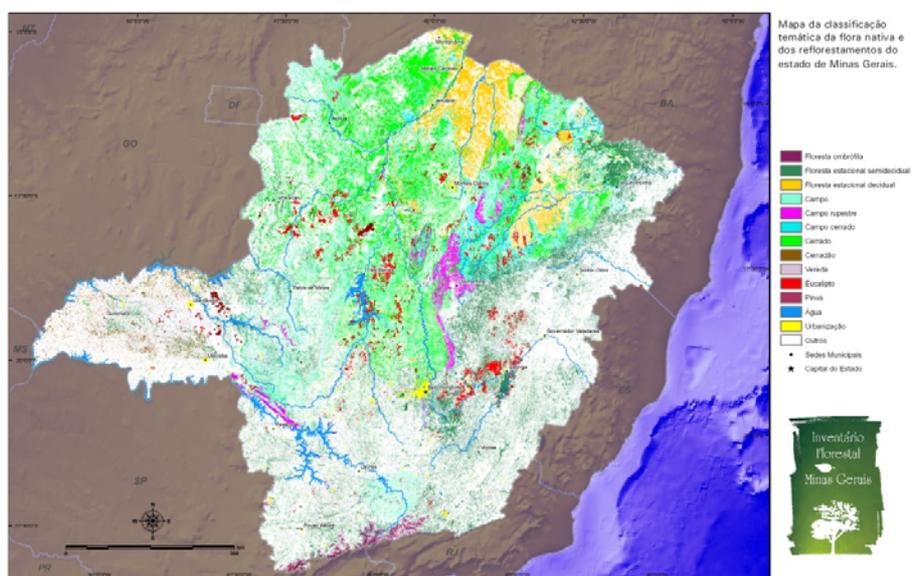


FIGURA 1. Mapa da classificação temática da flora nativa e dos reflorestamentos do estado de Minas Gerais.

O método de classificação foi o supervisionado e o algoritmo classificador utilizado foi o de árvores de decisão (Carvalho et al., 2004).

Após a classificação das imagens, classes como pastagem, culturas agrícolas e solo exposto foram agrupados em uma única classe chamada Outros (O). As demais classes compreendidas foram: Água (A), Campo (C), Campo Cerrado (CC), Campo Rupestre (CR), Cerradão (CO), Cerrado (CT), Eucalipto

(E), Floresta Estacional Decidual (FD), Floresta Estacional Semidecidual (FS), Floresta Ombrófila (FO), Pinus (P), Urbanização (U) e Veredas (V).

2.2 Coleta de dados

O procedimento de amostragem utilizado foi o de Amostragem Casual Simples (ACS) com proporção e estratificada. A adoção desta estratégia de amostragem atribui, primeiramente, a cada classe temática um estrato e, posteriormente, cada bacia hidrográfica um estrato, separadamente. Dentro destes estratos, realiza-se a amostragem casual simples. Assim, o número de estratos será o mesmo de classes temáticas e de bacias hidrográficas. Este tipo de procedimento foi escolhido em função da garantia de que todas as classes que compõem o Mapeamento fossem amostradas, inclusive as menores, além de determinar índices de qualidade para unidades básicas de planejamento de políticas públicas ambientais como as bacias hidrográficas.

A TABELA 1 apresenta a área em hectares e em porcentagem ocupada por cada bacia hidrográfica em relação à área total do Estado.

TABELA 1. Área ocupada, em hectare e porcentagem, por cada bacia hidrográfica

| BACIA HIDROGRÁFICA | ÁREA (ha) | ÁREA (%) |
|--------------------|------------|----------|
| Rio Buranhém | 32400,72 | 0,06 |
| Rio Doce | 7124942,70 | 12,15 |
| Rio Grande | 8605029,33 | 14,67 |
| Rio Itabapoana | 66817,98 | 0,11 |
| Rio Itanhém | 151335,72 | 0,26 |
| Rio Itapemirim | 3309,03 | 0,01 |
| Rio Jequitinhonha | 6562109,88 | 11,19 |
| Rio Jucuruçu | 70933,14 | 0,12 |
| Rio Mucuri | 1474772,22 | 2,52 |
| Rio Pardo | 1271143,71 | 2,17 |

...continua...

TAB. 1, Cont.

| | | |
|------------------------|-------------|--------|
| Rio Piracicaba/Jaguari | 115788,69 | 0,20 |
| Rio Paranaíba | 7059429,45 | 12,04 |
| Rio Paraíba do Sul | 2071837,80 | 3,53 |
| Rio São Francisco | 23464609,92 | 40,02 |
| Rio São Mateus | 563612,40 | 0,96 |
| Total | 58638072,69 | 100,00 |

As informações de referência de campo (referência) foram obtidas pelas incursões terrestres, realizadas para retratar amostras dos diferentes estratos da área e para averiguar a acurácia e precisão deste trabalho. Para tal, utilizou-se um aparelho de GPS acoplado a um veículo automotor. Para assegurar a independência dos dados, todos os pontos amostrais coletados foram diferentes daqueles utilizados para treinamento do classificador.

A distribuição das unidades amostrais nas sub-bacias ocorreu de forma proporcional à área de cada classe temática presente nestas unidades em relação à área total da classe presente no Estado. Assim, foi realizado o cálculo do número de polígonos a serem amostrados por sub-bacia para cada classe, garantindo que sub-bacias com maior área de determinada classe recebesse maior número de pontos desta e, bacias com quantidade menor desta classe, uma menor quantidade de unidades amostrais referentes a ela.

O procedimento de amostragem casual simples com proporção e estratificado contou, ainda, com o apoio do método dos itinerários utilizado por Belart et al. (2001), uma vez que a distribuição da amostra ocorreu nas imediações de rodovias.

Assim, utilizou-se uma base digital de rodovias do Departamento de Estradas e Rodagem (DER), em que foram selecionadas aleatoriamente 50%, 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

classes ao longo das estradas, considera-se, então, que esta demarcação, 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 da bacia do Rio Grande classificadas como obrigatórias, deslocamento e adicionais

As de deslocamento foram utilizadas para a condução entre uma e outra rodovia obrigatória. As adicionais eram amostradas, quando era notada qualquer região com concentração de alguma(s) classe(s) e, nesta não havia rodovias obrigatórias selecionadas nem de deslocamento.

A escolha de polígonos - definidos por pontos aleatorizados - como unidade amostral, teve como base as dificuldades de acesso às áreas em que foi distribuída a amostra, solução apresentada por Janssen et al. (1994). Além disso, a defasagem temporal existente entre a data da classificação (2006) e a da coleta de dados em campo (2007) foi preponderante para a escolha do polígono como

unidade amostral. Foram analisados apenas polígonos com tamanho aproximado de 10 *pixels* (9000m²) e, nas anotações de campo, foi considerado o tamanho do polígono observado. Os demais polígonos foram descartados no momento da análise dos dados.

No presente estudo, foram adotadas como estrato as fisionomias e as bacias hidrográficas. Baseando-se na coleta de dados em todas as áreas do Estado, procedeu-se à separação dos polígonos pertencentes a cada classe temática e, posteriormente, a mesma separação por bacias.

O tamanho da amostra de referência por classe temática foi criado por meio da exatidão fixada *a priori*, calculado pela fórmula apresentada.

$$n = \frac{Z^2 \cdot p \cdot q}{E^2}, \quad (1)$$

onde:

n = tamanho da amostra de referência para cada classe temática;

Z = ponto crítico para um limite de confiança de inclusão para o parâmetro exatidão real, assumindo $Z \sim N(0,1)$;

p = exatidão desejada;

$q = 1-p$;

E = erro de amostragem admitido.

A exatidão p utilizada foi fixada em 85% e a mínima exigida para mapeamentos temáticos (Jensen, 1986). O erro admissível empregado foi de 10% à probabilidade de 95%.

No campo foram coletadas as seguintes informações: sentido de deslocamento na rodovia; coordenadas geográficas dos pontos amostrais; estrato ocorrente; lado da rodovia em que a classe temática se encontrava; tamanho (grande e pequeno) e formato do fragmento (estrito e largo); distância aproximada entre a rodovia e o fragmento, quando este possuía grandes

dimensões e idade aproximada, quando o fragmento em questão pertencia às classes Eucalipto ou Pinus.

Informações adicionais dos fragmentos amostrados também foram coletadas, quando possível ou necessário. Estas informações garantiram maior confiabilidade à amostra, além de formar um banco de dados das classes em todo o Estado, sendo: estágio de sucessão (pioneiras, intermediárias ou clímax) e proximidade a outros fragmentos, com os quais possa ter ocorrido confusão na classificação temática.

Como a classe Floresta Decídua apresenta deciduidade e proximidade espectral à Floresta Estacional Semidecidual, a coleta da amostra para esta foi realizada no período em que a formação se encontrava sem folhas, para evitar confusão no campo.

2.3 Avaliação da acurácia

A determinação da acurácia foi criada para as bacias hidrográficas, usando os dados de referência de campo. Com base nesses dados, foi possível determinar a proporção de polígonos corretamente classificados em relação ao número total de polígonos visitados em cada bacia.

Foi utilizada, então, a matriz de confusão e os índices dela derivados, conforme citados por Stehman (1997), Congalton & Green (1999) e Lingnau & Antunes (2003).

Foram utilizados tanto coeficientes de concordância para todo o Mapeamento (Concordância Total) quanto os que avaliam a concordância para as classes mapeadas, individualmente. Dentre os coeficientes de concordância total, utilizou-se o de Exatidão Global, proposto por Hellden & Stern (1980), o Kappa, proposto por Cohen (1960) e o Tau, proposto por Klecka (1980). Para a avaliação das classes individuais os coeficientes utilizados foram: Acurácia do Produtor e Acurácia do Usuário (Congalton & Green, 1999).

2.4 Determinação da Precisão

Considerando a determinação do número de pontos amostrais necessários para cada estrato (classes temáticas), obtido pela fórmula de dimensionamento do tamanho da amostra de referência (1), com base na distribuição binomial da variável aleatória, foram estimados os parâmetros da proporção de polígonos classificados corretamente, por meio dos estimadores da Amostragem Casual Simples e a média aritmética, variância, variância da média da proporção, erro padrão da média da proporção, precisão do levantamento e intervalo de confiança. Estas estimativas foram criadas, tanto por classe temática, quanto por bacia hidrográfica.

Somente os polígonos que representavam apenas uma classe foram considerados nas análises. Os polígonos que apresentavam transição entre classes como, por exemplo, Campo - Campo Cerrado - Campo Rupestre, foram excluídos da base de dados, conferindo maior segurança e confiabilidade à amostra.

3 RESULTADOS E DISCUSSÃO

3.1 Avaliação por Classe Temática

No presente estudo, as classes temáticas definidas no mapeamento, foram identificadas como estratos. O tamanho da amostra de referência calculado foi de 49 pontos amostrais para cada estrato, para um valor crítico de Z tendendo ao infinito ($Z = 1,96$) e erro admitido de 10%, com proporção mínima de acerto pré-estabelecida de 85% (p).

Foram percorridos aproximadamente 50000km ao longo das rodovias amostradas. Em razão da facilidade de coleta de pontos em campo e à grande extensão de rodovias (obrigatórias, de deslocamento e adicionais) percorridas, ocorreu um significativo aumento no número de pontos coletados, acumulando 11942 pontos, distribuídos nas 14 classes (FIGURA 3).

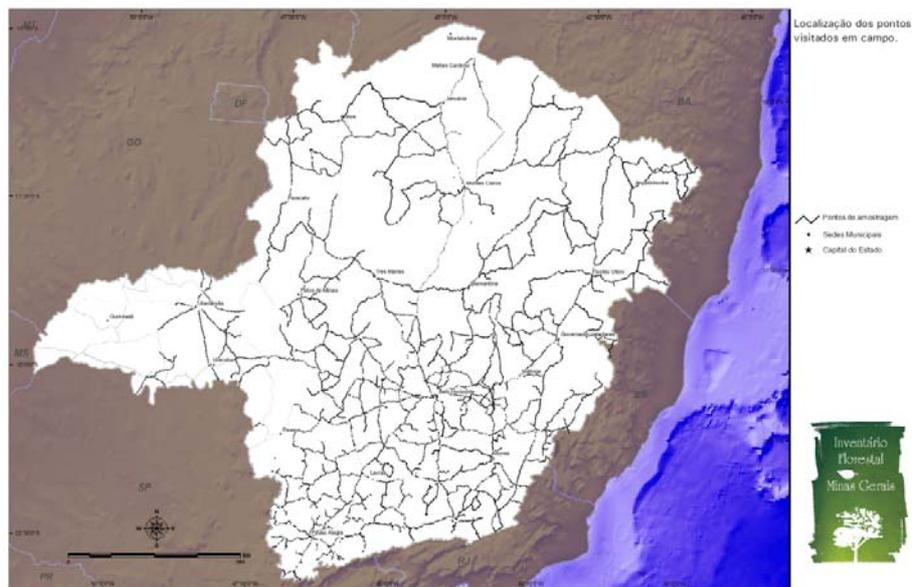


FIGURA 3. Localização dos pontos visitados em campo

O número de polígonos amostrados (n) e corretamente classificados (n*) para as classes temáticas está apresentado na TABELA 6.

TABELA 6. Número de polígonos amostrados (n) e polígonos corretamente classificados (n*)

| CLASSE TEMÁTICA | n | n* |
|----------------------------------|------|------|
| Água | 194 | 171 |
| Campo | 137 | 109 |
| Campo Cerrado | 240 | 192 |
| Campo Rupestre | 143 | 89 |
| Cerradão | 70 | 64 |
| Cerrado | 1087 | 897 |
| Eucalipto | 722 | 389 |
| Floresta Decidual | 748 | 539 |
| Floresta Estacional Semidecidual | 1779 | 1495 |
| Floresta Ombrófila | 149 | 68 |
| Outros | 5894 | 5720 |
| Pinus | 69 | 58 |
| Urbanização | 607 | 572 |
| Veredas | 103 | 59 |

A classe temática com maior número de unidades amostrais foi a classe Outros, seguida pela Floresta Estacional Semidecidual e Cerrado, com 5894, 1779 e 1087 polígonos analisados, respectivamente. Estas três classes são as mais com maior área presente no Estado de Minas Gerais. As classes com menores números de unidades, Pinus e Cerradão, correspondem às com menores áreas em relação à área total do Estado. Esta proporcionalização em relação à ocupação de cada classe nas diferentes sub-bacias (unidades de controle da amostragem) foi resultado da metodologia proposta para este trabalho.

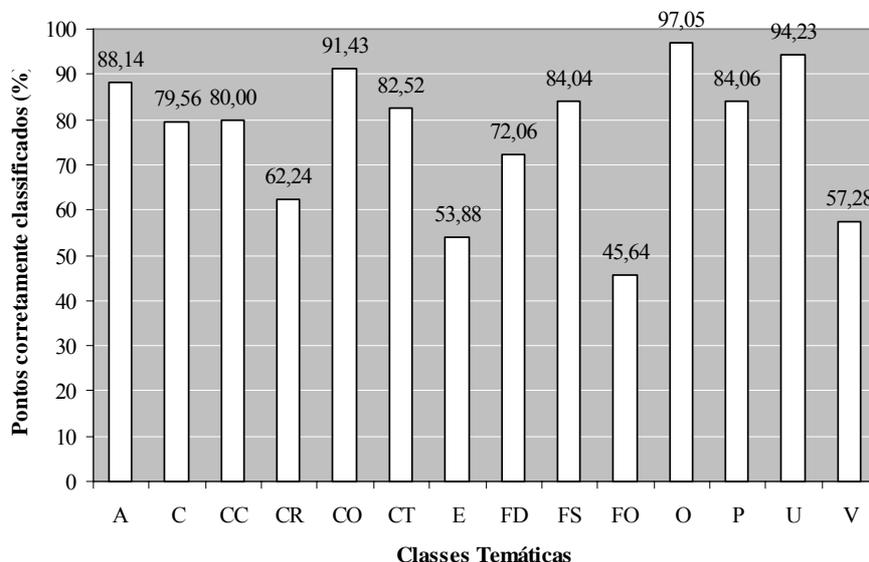


FIGURA 4. Percentual de pontos corretamente classificados correspondentes aos estratos. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

Fundamentando-se na FIGURA 4, identifica-se que o estrato Outros foi o mais corretamente classificado no Mapeamento, apresentando 97,05% dos polígonos analisados corretamente classificados. Este foi seguido pelos estratos Urbanização (94,23%), Cerradão (91,43%) e Água (88,14%).

Abaixo dos 85% de pontos corretamente classificados encontram-se os estratos Pinus (84,06%), Floresta Estacional Semidecidual (84,04%), Cerrado (82,52%), Campo Cerrado (80%), Campo (79,56%), Floresta Decidual (72,06%), Campo Rupestre (62,24%), Veredas (57,28%), Eucalipto (53,88%) e Floresta Ombrófila (45,64%), indicando a necessidade de maiores estudos para realizar futuras classificações destas classes.

A TABELA 7 contém uma matriz de confusão, relacionando a referência de campo com o mapeamento.

TABELA 7. Matriz de confusão. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | | | | | | | | | | | | |
|------------|--------|---------------------|------------|------------|-----------|-----------|------------|------------|------------|-------------|-----------|-------------|-----------|------------|-----|-----------|--------------|
| Mapeamento | Classe | A | C | CC | CR | CO | CT | E | FD | FS | FO | O | P | U | V | Total | |
| | A | 171 | | | | 1 | | | | | | | 1 | | | | 173 |
| | C | | 109 | 6 | 15 | | 16 | 14 | 5 | 15 | 4 | 61 | | 1 | 3 | | 249 |
| | CC | | | 192 | | | 11 | 5 | 38 | 9 | | 14 | | | | | 269 |
| | CR | | | 6 | 89 | | 3 | 3 | | 2 | | 1 | | | | | 104 |
| | CO | | | | | 64 | | | | | | 4 | | | 3 | | 71 |
| | CT | 1 | 1 | 2 | | | 897 | 11 | 34 | 46 | | 31 | | 1 | 13 | | 1037 |
| | E | | 1 | | | | 9 | 389 | | 14 | | 9 | 3 | | | | 425 |
| | FD | | | 2 | | | 8 | | 539 | | | 10 | | | | | 559 |
| | FS | | 1 | 7 | 1 | | 49 | 70 | 26 | 1495 | 53 | 37 | 1 | | 4 | | 1744 |
| | FO | | | | | | | 1 | | 2 | 68 | | | | | | 71 |
| | O | 22 | 24 | 25 | 38 | 4 | 93 | 211 | 106 | 192 | 24 | 5720 | 7 | 33 | 21 | | 6520 |
| | P | | 1 | | | | 1 | 16 | | | | 1 | 58 | | | | 77 |
| | U | | | | | | | 1 | | 1 | | 5 | | 572 | | | 579 |
| | V | | | | | 1 | | 1 | | 3 | | | | | | 59 | 64 |
| | Total | 194 | 137 | 240 | 143 | 70 | 1087 | 722 | 748 | 1779 | 149 | 5894 | 69 | 607 | 103 | | 11942 |

Ao analisar a matriz acima, tem-se que, por exemplo, dos 137 pontos amostrados em campo para a classe Campo, um estava mapeado como Cerrado, um como Eucalipto, um como Floresta Estacional Semidecidual, 24 como Outros e um como Pinus. Esta confusão refere-se à Acurácia do Produtor ou aos erros de exclusão ou omissão. Por outro lado, dos 249 polígonos, apenas 109 estavam classificados corretamente. Outras classes estavam mapeadas como Campo, porém, não correspondiam com o encontrado na realidade, visto que seis polígonos de Campo Cerrado, 15 de Campo Rupestre, 16 de Cerrado, 14 de Eucalipto, cinco de Floresta Estacional Decidual, quinze de Floresta Estacional Semidecidual, quatro de Floresta Ombrófila, 61 de Outros, um de Urbanização e três de Veredas. Estes erros referem-se à Acurácia do Usuário ou de inclusão ou comissão.

Dentre os estratos que receberam os menores números de unidades amostrais destacam-se o Pinus e o Cerradão, com 69 e 70 unidades, respectivamente. Nota-se, com base na TABELA 8 que estas classes contribuem com um percentual pequeno em relação à área total do Estado e, além disso, encontram-se, relativamente às outras classes, concentradas em regiões específicas.

TABELA 8. Área ocupada, em hectare e porcentagem, por classe temática

| FISIOMONIA | ÁREA (ha) | % |
|----------------------------------|-----------|------|
| Floresta Ombrófila | 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 | 0,6 |
| Vereda | 406.253 | 0,69 |
| Eucalipto | 1.218.212 | 2,08 |
| Pinus | 143.395 | 0,24 |

A TABELA 9 contém as estimativas dos parâmetros para a variável polígonos classificados corretamente.

TABELA 9. Parâmetros estatísticos para a variável binomial polígonos corretamente classificados

| Classe Temática | Média | Variância | Desvio Padrão da Média | Erro % |
|----------------------------------|---------|-----------|------------------------|--------|
| Água | 0,88144 | 0,10504 | 0,02327 | 5,17 |
| Campo | 0,79562 | 0,16380 | 0,03458 | 8,52 |
| Campo Cerrado | 0,80000 | 0,16067 | 0,02587 | 6,34 |
| Campo Rupestre | 0,62238 | 0,23668 | 0,04068 | 12,81 |
| Cerradão | 0,91429 | 0,07950 | 0,03370 | 7,22 |
| Cerrado | 0,82521 | 0,14437 | 0,01152 | 2,74 |
| Eucalipto | 0,53878 | 0,24884 | 0,01856 | 6,75 |
| Floresta Decidual | 0,72059 | 0,20161 | 0,01642 | 4,47 |
| Floresta Estacional Semidecidual | 0,84036 | 0,13423 | 0,00869 | 2,03 |
| Floresta Ombrófila | 0,45638 | 0,24977 | 0,04094 | 17,58 |
| Outros | 0,97048 | 0,02865 | 0,00220 | 0,45 |
| Pinus | 0,84058 | 0,13598 | 0,04439 | 10,35 |
| Urbanização | 0,94234 | 0,05443 | 0,00947 | 1,97 |
| Veredas | 0,57282 | 0,24710 | 0,04898 | 16,76 |

A FIGURA 5 apresenta o erro percentual do levantamento relacionado à cada classe temática mapeada. Nota-se que a classe Outros foi a que apresentou o menor erro percentual do levantamento com 0,45%, seguida pela Urbanização com 1,97%. As que apresentaram maiores erros do levantamento foram Floresta Ombrófila (FO) com 17,58%, Veredas (V) com 16,76% e Campo Rupestre (CR) com 12,81%. Com estes resultados podem-se fazer algumas inferências. Os baixos valores de erro das classes Outros e Urbanização estão relacionados com as características peculiares dos atributos da imagem destas coberturas, portanto, já era esperado que este fato ocorresse. Com relação ao alto valor do erro associado ao mapeamento da Floresta Ombrófila, pode-se inferir que esta pequena precisão está relacionada à presença de sombras na imagem, antes de realizar a classificação. Segundo Pimentel et. al (2008), qualquer elemento que se encontra coberto por sombra possui número digital atribuído a seus *pixels* diferentes dos que teriam se não estivessem nesta situação. Como as áreas que contêm a classe Floresta Ombrófila são geralmente áreas montanhosas, sombras podem ter influenciado na classificação.

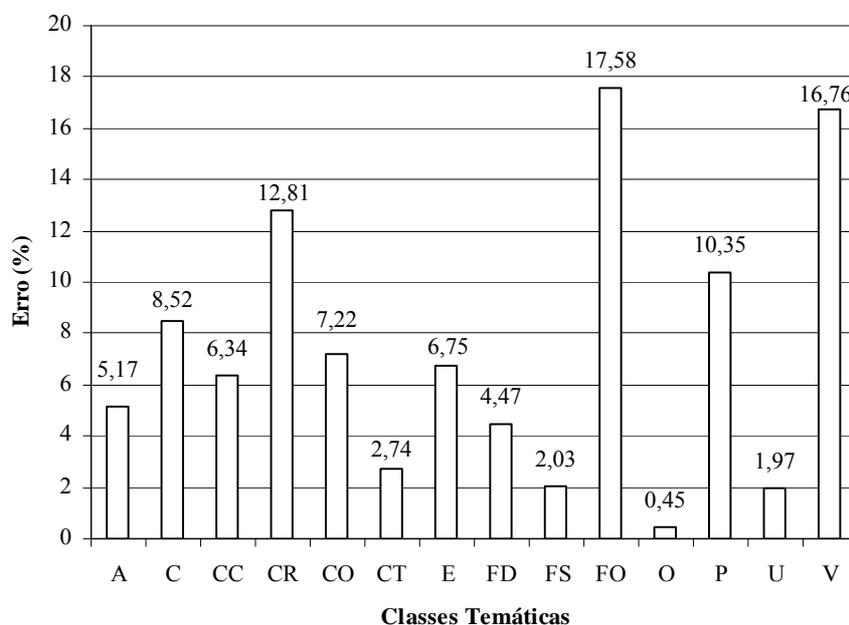


FIGURA 5. Erro percentual do levantamento para o número de polígonos corretamente classificados em cada estrato

Baseando-se no cálculo da precisão para cada classe temática, foi possível criar os intervalos de confiança para o parâmetro proporção real de acertos dos polígonos classificados considerando a imagem (TABELA 10).

TABELA 10. Intervalo de confiança com 95% de probabilidade de acerto para os estratos amostrados

| CLASSE TEMÁTICA | INTERVALO DE CONFIANÇA |
|-----------------|------------------------|
| Água | 83,58 ≤ 88,14 ≤ 92,71 |
| Campo | 72,78 ≤ 79,56 ≤ 86,34 |
| Campo Cerrado | 74,93 ≤ 80,00 ≤ 85,07 |
| Campo Rupestre | 54,26 ≤ 62,23 ≤ 70,21 |
| Cerradão | 84,82 ≤ 91,42 ≤ 98,03 |

...continua...

TAB. 10, Cont.

| | | | |
|----------------------------------|-------|-------------------|-------|
| Cerrado | 80,26 | $\leq 82,52 \leq$ | 84,78 |
| Eucalipto | 50,24 | $\leq 53,88 \leq$ | 57,52 |
| Floresta Decidual | 68,84 | $\leq 72,06 \leq$ | 75,28 |
| Floresta Estacional Semidecidual | 82,33 | $\leq 84,03 \leq$ | 85,74 |
| Floresta Ombrófila | 37,61 | $\leq 45,63 \leq$ | 53,66 |
| Outros | 96,62 | $\leq 97,05 \leq$ | 97,48 |
| Pinus | 75,36 | $\leq 84,06 \leq$ | 92,76 |
| Urbanização | 92,38 | $\leq 94,23 \leq$ | 96,09 |
| Veredas | 47,68 | $\leq 57,28 \leq$ | 66,88 |

Estes intervalos apresentam a localização, em porcentagem, da real proporção de polígonos corretamente classificados para cada classe temática do Mapeamento.

Quanto maior a precisão, mais confiável será a estimativa e mais estreito será o intervalo de confiança. Apresenta-se, por exemplo, a proporção real de polígonos classificados corretamente para a classe Água, que está situada entre 83,58 e 92,71% do número total de polígonos mapeados da classe, com um nível de significância de 95%.

Observa-se que as classes temáticas Outros e Urbanização foram as que apresentaram o menor erro de estimativa e, conseqüentemente, menor foi a amplitude destes intervalos. Classes com erros maiores, como Veredas e Floresta criados para todas as classes, as classes Água, Cerradão, Outros e Urbanização, foram as que contemplaram a proporção real acima de 85%. Este seria um limite aceitável para a classificação temática. Portanto, individualmente, poder-se-á ter proporções reais inferiores a este valor. Segundo Watzlawick et al. (2002), a exatidão para todo mapeamento pode ser alta, não implicando necessariamente alta exatidão para as classes individuais.

O cálculo da precisão é fortemente influenciado pelo número de pontos identificados corretamente no campo, quando comparado com o mapeamento. Cita-se, como exemplo, a classe temática Floresta Ombrófila. Para esta foram

coletadas 149 unidades amostrais. No entanto, apenas 68 estavam corretamente classificadas. Este fato contribuiu para redução de p para esta classe. Conseqüentemente, houve aumento da variância, o que provocou redução na precisão. Retornando à FIGURA 5, observa-se que as classes Pinus, Campo Rupestre, Veredas e Floresta Ombrófila apresentaram este comportamento, provocando erros acima de 10%.

3.2 Avaliação da Acurácia e Precisão por Bacia Hidrográfica

O tamanho da amostra de referência calculado foi de 49 pontos amostrais para cada bacia hidrográfica, para um valor crítico de Z tendendo ao infinito ($Z = 1,96$) e erro pré-estabelecido de 10%, com uma proporção *a priori* de 85%. No entanto, em função dos deslocamentos nas rodovias, foram amostrados 11942 polígonos distribuídos em 14 bacias hidrográficas (TABELA 11).

Tomando-se por base o número de polígonos amostrados (n) e o número de polígonos corretamente classificados pelo Mapeamento (n^*), apresentados na TABELA 11, procedeu-se à determinação da precisão do Mapeamento para cada bacia hidrográfica.

TABELA 11. Área porcentual ocupada pelas bacias hidrográficas no Estado de Minas Gerais, número de pontos amostrados (n) e corretamente classificados (n^*) em cada bacia hidrográfica

| BACIA HIDROGRÁFICA | ÁREA (%) | n | n^* |
|--------------------|----------|------|-------|
| Rio Buranhém | 0,06 | 18 | 18 |
| Rio Itanhém | 0,26 | 98 | 93 |
| Rio Jequitinhonha | 11,19 | 1424 | 1113 |
| Rio Jucuruçu | 0,12 | 33 | 31 |
| Rio Paranaíba | 12,04 | 508 | 467 |
| Rio São Mateus | 0,96 | 128 | 126 |
| Rio Doce | 12,15 | 2054 | 1914 |

...continua...

TAB. 11, Cont.

| | | | |
|------------------------|-------|-------|-------|
| Rio Grande | 14,67 | 1816 | 1470 |
| Rio São Francisco | 40,02 | 4438 | 3908 |
| Rio Itabapoana | 0,11 | 45 | 40 |
| Rio Mucuri | 2,52 | 383 | 356 |
| Rio Paraíba do Sul | 3,53 | 751 | 677 |
| Rio Pardo | 2,17 | 196 | 141 |
| Rio Piracicaba/Jaguari | 0,2 | 50 | 34 |
| Rio Itapemirim | 0,01 | 0 | 0 |
| TOTAL | 100,0 | 11942 | 10388 |

Nota-se que a maior bacia de Minas Gerais, a do Rio São Francisco com 40,02% da área do Estado, foi a que recebeu um maior número de unidades amostrais, com 4438 polígonos amostrados. As bacias menores, como a do Rio Buranhém, Itabapoana e Piracicaba/Jaguari receberam poucas unidades, relativamente às maiores. Na bacia do Rio Itapemirim, a menor do Estado com 0,01% do território, não foi possível coletar nenhuma amostra em virtude de sua pequena dimensão e ausência de estradas trafegáveis. Dos 11942 polígonos visitados, apenas 1520 (12,7%) não possuíam classificação correta no Mapeamento.

A distribuição dos polígonos amostrados em cada bacia hidrográfica, de acordo com a classe temática destes, está apresentada na TABELA 12.

TABELA 12. Distribuição dos polígonos amostrados de acordo com a classe temática em cada bacia hidrográfica. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| Bacia Hidrográfica | Classes Temáticas | | | | | | | | | | | | | | Total |
|------------------------|-------------------|-----|-----|-----|----|------|-----|-----|------|-----|------|----|-----|-----|-------|
| | A | C | CC | CR | CO | CT | E | FD | FS | FO | O | P | U | V | |
| Rio Buranhém | - | - | 1 | - | - | - | - | - | 1 | - | 15 | - | 1 | - | 18 |
| Rio Itanhém | - | - | - | - | - | - | 4 | - | 11 | - | 79 | - | 4 | - | 98 |
| Rio Jequitinhonha | 11 | 13 | 103 | 41 | - | 175 | 164 | 382 | 66 | - | 421 | 13 | 35 | - | 1424 |
| Rio Jucuruçu | - | - | - | - | - | - | - | - | 1 | - | 31 | - | 1 | - | 33 |
| Rio Paranaíba | 6 | 7 | - | - | 28 | 9 | 16 | - | 26 | - | 337 | 41 | 14 | 24 | 508 |
| Rio São Mateus | - | - | - | 2 | - | - | - | - | 12 | - | 103 | - | 11 | - | 128 |
| Rio Doce | 31 | 12 | - | 19 | 13 | 9 | 125 | - | 599 | 3 | 1105 | 2 | 119 | 17 | 2054 |
| Rio Grande | 76 | 46 | 17 | 29 | 21 | 47 | 108 | 1 | 412 | 82 | 832 | 2 | 138 | 5 | 1816 |
| Rio São Francisco | 59 | 49 | 91 | 46 | 8 | 776 | 236 | 328 | 440 | - | 2137 | 7 | 204 | 57 | 4438 |
| Rio Itabapoana | - | - | - | - | - | - | 2 | - | 7 | - | 32 | - | 4 | - | 45 |
| Rio Mucuri | 1 | - | 5 | - | - | 1 | 7 | - | 75 | - | 277 | - | 17 | - | 383 |
| Rio Paraíba do Sul | 10 | 2 | - | 5 | - | - | 15 | - | 128 | 47 | 496 | 1 | 47 | - | 751 |
| Rio Pardo | - | 8 | 22 | 1 | - | 67 | 37 | 37 | - | - | 16 | - | 8 | - | 196 |
| Rio Piracicaba/Jaguari | - | - | 1 | - | - | 3 | 8 | - | 1 | 17 | 13 | 3 | 4 | - | 50 |
| Total | 194 | 137 | 240 | 143 | 70 | 1087 | 722 | 748 | 1779 | 149 | 5894 | 69 | 607 | 103 | 11942 |

A FIGURA 6 apresenta o percentual de polígonos corretamente classificados para as bacias amostradas.

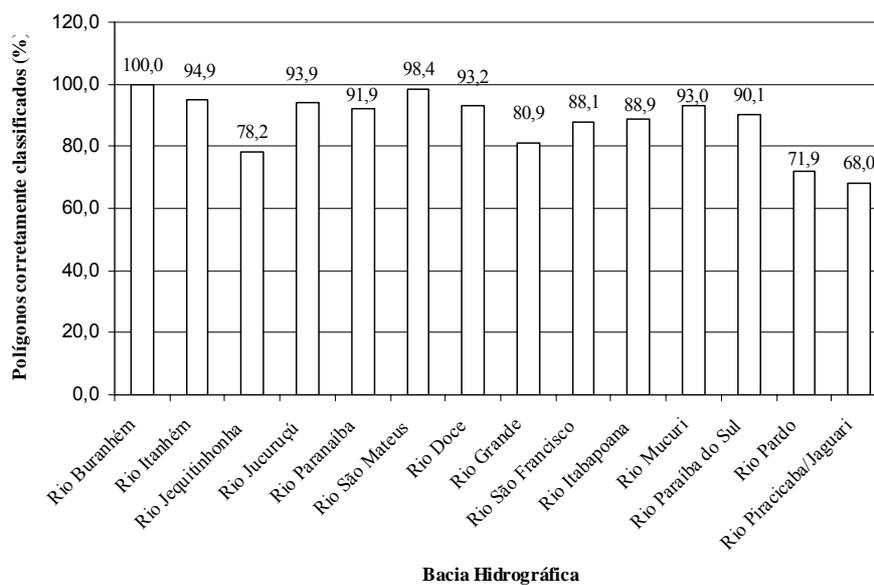


FIGURA 6. Polígonos corretamente classificados por bacia hidrográfica

Dentre as 14 bacias hidrográficas amostradas, a bacia do Rio Buranhém foi a que apresentou a maior quantidade de pontos corretamente classificados pelo Mapeamento (100%).

Detectou-se que 10 das 15 bacias hidrográficas do Estado apresentaram proporção de polígonos corretamente classificados acima de 85%, valor considerado bom pela literatura (Jensen, 1986) para mapeamentos temáticos. Dentre as cinco bacias com proporção de polígonos corretamente classificados inferior a 85%, estão duas bacias significativas em importância e área para o Estado, a bacia do Rio Grande e do Rio Jequitinhonha. Analisando a matriz de

confusão criada para a bacia do Rio Grande, apresentada na TABELA 14A do ANEXO A, observa-se que a classe Pinus não apresentou pontos concordantes entre o mapeamento e os dados de campo. As classes Cerrado e Eucalipto também apresentaram baixos valores de concordância entre o mapa e o campo, campo de Cerrado, apenas 10 estavam corretamente classificados no mapa e, dos 108 de Eucalipto, apenas 23 foram classificados como Eucalipto no mapa. Ao analisar a matriz de confusão (TABELA 2A do ANEXO A), confeccionada para a bacia do Rio Jequitinhonha, as classes Floresta Decidual e Floresta Estacional Semidecidual foram as que apresentaram menores percentuais de concordância, com 60,55% e 64,70%, respectivamente. Estes valores baixos influenciaram a proporção de polígonos corretamente classificados nestas bacias hidrográficas.

As bacias do Rio São Mateus e a do Rio Itanhém também merecem destaque, pois apresentaram 98,4% e 94,9% de polígonos corretamente classificados pelo Mapeamento, respectivamente, sendo considerados altamente satisfatórios.

A TABELA 13 apresenta o intervalo de confiança para a variável polígonos corretamente classificados em cada bacia hidrográfica do Estado de Minas Gerais. Mesmo não tendo amostrado o número mínimo de 49 unidades amostrais, para atingir a proporção de 85% de polígonos corretamente classificados, para as bacias do Rio Jucuruçu e Rio Itapemirim também foram apresentados intervalos de confiança. Embora exista um déficit no número de unidades amostrais para estas bacias, as poucas tomadas retrataram a variabilidade ocorrente com a variável de interesse, possibilitando, assim, intervalos com proporções maiores que 85%. Porém, não se pode garantir um erro máximo de 10%. Para este patamar de erro, seriam necessárias, no mínimo, 49 unidades amostrais. Dos 14 intervalos de confiança, apenas quatro não englobam a exatidão *a priori* de 85%. Observa-se que os valores de erro destas bacias hidrográficas foram superiores a 10%, necessitando, portanto, de um

número maior de unidades amostrais para se ter boa precisão da exatidão real estimada.

TABELA 13. Intervalos de confiança com 95% de probabilidade de acerto para a porcentagem de polígonos corretamente classificados por bacia hidrográfica

| BACIA HIDROGRÁFICA | ERRO (%) | INTERVALO DE CONFIANÇA |
|------------------------|----------|------------------------|
| Rio Buranhém | 0,00 | 100,00 ≤ P ≤ 100,00 |
| Rio Itanhém | 4,61 | 90,52 ≤ P ≤ 99,28 |
| Rio Jequitinhonha | 2,75 | 76,01 ≤ P ≤ 80,31 |
| Rio Jucuruçu | 8,80 | 85,67 ≤ P ≤ 102,21 |
| Rio Paranaíba | 2,58 | 89,56 ≤ P ≤ 94,30 |
| Rio São Mateus | 2,19 | 96,28 ≤ P ≤ 100,59 |
| Rio Doce | 1,17 | 92,09 ≤ P ≤ 94,27 |
| Rio Grande | 2,23 | 79,14 ≤ P ≤ 82,75 |
| Rio São Francisco | 1,08 | 87,10 ≤ P ≤ 89,01 |
| Rio Itabapoana | 10,45 | 79,60 ≤ P ≤ 98,17 |
| Rio Mucuri | 2,76 | 90,38 ≤ P ≤ 95,52 |
| Rio Paraíba do Sul | 2,37 | 88,01 ≤ P ≤ 92,28 |
| Rio Pardo | 8,77 | 65,63 ≤ P ≤ 78,25 |
| Rio Piracicaba/Jaguari | 19,21 | 54,94 ≤ P ≤ 81,06 |

3.2.2 Avaliação da acurácia

Com base nos dados coletados, realizou-se a confecção de matrizes de confusão para as bacias hidrográficas amostradas, a fim de avaliar a acurácia por bacia. Posteriormente realizaram-se os cálculos dos coeficientes de concordância total (Exatidão Global, Kappa e Tau) e dos coeficientes de concordância para as classes temáticas (Acurácia do Produtor e Acurácia do Usuário) presentes em cada bacia hidrográfica do Estado.

Considerando estas matrizes de confusão, extraíram-se os coeficientes de concordância total para cada bacia (TABELA 14). As matrizes de confusão para cada bacia estão apresentadas no ANEXO.

TABELA 14. Coeficientes de concordância total para as bacias hidrográficas amostradas

| BACIA HIDROGRÁFICA | COEFICIENTES DE CONCORDÂNCIA TOTAL (%) | | |
|------------------------|--|--------|--------|
| | EXATIDÃO GLOBAL | KAPPA | TAU |
| Rio Buranhém | 100,00 | 100,00 | 100,00 |
| Rio Itanhém | 94,89 | 84,30 | 93,19 |
| Rio Jequitinhonha | 78,16 | 72,86 | 75,97 |
| Rio Jucuruçu | 93,94 | 64,52 | 90,91 |
| Rio Paranaíba | 91,93 | 84,90 | 91,03 |
| Rio São Mateus | 98,44 | 95,18 | 97,92 |
| Rio Doce | 93,18 | 88,72 | 92,56 |
| Rio Grande | 80,95 | 72,07 | 79,48 |
| Rio São Francisco | 88,74 | 84,06 | 87,80 |
| Rio Itabapoana | 88,89 | 73,44 | 85,19 |
| Rio Mucuri | 92,95 | 83,51 | 91,78 |
| Rio Paraíba do Sul | 90,15 | 80,28 | 88,91 |
| Rio Pardo | 71,94 | 65,60 | 68,43 |
| Rio Piracicaba/Jaguari | 68,00 | 60,53 | 63,43 |

A bacia do Rio Buranhém foi a única que apresentou valores iguais para os três coeficientes, pois todos os dezoito polígonos amostrados em campo correspondiam às classes encontradas no Mapeamento.

Como esperado, para todas as bacias hidrográficas, o coeficiente Exatidão Global foi maior que o coeficiente Kappa. Isto ocorreu em virtude do último levar em consideração toda a matriz de confusão, enquanto que o primeiro considera apenas o número total de pontos amostrados e o número total de pontos corretamente classificados, portanto, menos rigoroso (Lingnau, e. Antunes, 2003).

Além disso, o Kappa, para todas as bacias amostradas, foi menor quando comparado ao coeficiente Tau determinado para elas. Explica-se este fato pela utilização de todas as células da matriz de confusão no cálculo do Kappa, considerando assim os erros de exclusão e inclusão das classes temáticas.

Por sua vez, o Tau determinado foi menor que a Exatidão Global para todas as bacias, apresentando valores mais próximos aos da Exatidão Global quando comparado com o Kappa. Isto se deve ao menor rigor do Tau em relação ao último, pois o primeiro, apesar de também levar em consideração todos os elementos da matriz de confusão, considera a concordância esperada como a mesma para todas as classes temáticas, enquanto o Kappa utiliza a proporção de concordância depois que a concordância atribuída à casualidade é retirada de consideração.

Além dos coeficientes de concordância total, gerou-se a acurácia do usuário e do produtor para as classes temáticas encontradas em cada bacia.

Na TABELA 15, que mostra a acurácia do produtor para as classes presentes nas bacias hidrográficas, pode-se notar que a classe Outros apresentou um alto valor para este coeficiente em todas as bacias analisadas, com valores variando entre 93,55 a 100%. Em todas as 14 bacias foram obtidos valores superiores a 85%. Outra classe que apresentou altos valores foi a Urbanização, com mínimo de 75% e máximo de 100% e, 12, das 14 bacias amostradas, apresentaram valores superiores a 85% para este coeficiente. A classe Eucalipto apresentou valores relativamente pequenos em relação às outras classes e, em nenhuma bacia, apresentou valores mínimos de 85%. Isto se deu por causa do grande percentual de confusão ocorrido entre as classes, indicando que o mapeamento não representou bem os polígonos de Eucalipto encontrados em campo, sendo classificados como outras classes. A classe Veredas, também, não apresentou nenhum valor acerca de 85%, sendo todos menores, variando entre 45,61 e 79,17%, para as quatro bacias em que ocorreu. As demais classes apresentaram valores acima de 85%, em pelo menos uma bacia hidrográfica. É curioso notar a classe Floresta Ombrófila que apresentou o segundo menor valor de acurácia do produtor para a bacia do Rio Grande (20,73%), apresentou também o valor máximo de 100% na bacia do Rio Doce. A bacia do Rio Paraíba

do Sul foi a que apresentou a menor acurácia do produtor e foi atribuída à classe Campo Rupestre. Apenas 20% dos polígonos analisados foram corretamente classificados como Campo Rupestre, de todos os encontrados em campo e identificados como esta classe. Os valores zero encontrados na TABELA 15, referem-se a classes que não apresentaram nenhum polígono em concordância entre a classe encontrada em campo e a mapeada.

TABELA 15. Acurácia do Produtor para as classes temáticas presentes em bacias hidrográficas. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| Bacias Hidrográficas | Acurácia do Produtor (%) | | | | | | | | | | | | | |
|------------------------|--------------------------|--------|--------|--------|-------|--------|-------|--------|--------|--------|--------|--------|--------|-------|
| | A | C | CC | CR | CO | CT | E | FD | FS | FO | O | P | U | V |
| Rio Buranhém | - | - | 100,00 | - | - | - | - | - | 100,00 | - | 100,00 | - | 100,00 | - |
| Rio Itanhém | - | - | - | - | - | - | 25,00 | - | 100,00 | - | 97,47 | - | 100,00 | - |
| Rio Jequitinhonha | 90,91 | 69,23 | 85,44 | 87,80 | - | 70,86 | 67,07 | 63,9 | 66,67 | - | 95,72 | 100,00 | 100,00 | - |
| Rio Jucuruçu | - | - | - | - | - | - | - | - | 100,00 | - | 93,55 | - | 100,00 | - |
| Rio Paranaíba | 100,00 | 100,00 | - | - | 92,86 | 33,33 | 25,00 | - | 100,00 | - | 96,74 | 87,80 | 100,00 | 79,17 |
| Rio São Mateus | - | - | - | 100,00 | - | - | - | - | 100,00 | - | 98,10 | - | 100,00 | - |
| Rio Doce | 93,55 | 75,00 | - | 31,58 | 84,62 | 100,00 | 60,00 | - | 91,49 | 100,00 | 99,46 | 0,00 | 95,80 | 64,71 |
| Rio Grande | 86,84 | 73,91 | 41,18 | 41,38 | 95,24 | 21,28 | 21,30 | 100,00 | 78,40 | 20,73 | 97,72 | 0,00 | 92,03 | 60,00 |
| Rio São Francisco | 88,14 | 85,71 | 78,02 | 69,57 | 87,50 | 88,66 | 69,49 | 83,84 | 80,00 | - | 95,41 | 85,71 | 92,16 | 45,61 |
| Rio Itabapoana | - | - | - | - | - | - | 0,00 | - | 71,43 | - | 96,88 | - | 100,00 | - |
| Rio Mucuri | 100,00 | - | 100,00 | - | - | 100,00 | 0,00 | - | 86,67 | - | 96,75 | - | 94,12 | - |
| Rio Paraíba do Sul | 70,00 | 100,00 | - | 20,00 | - | - | 0,00 | - | 83,59 | 51,06 | 98,59 | 0,00 | 100,00 | - |
| Rio Pardo | - | 75,00 | 86,36 | 100,00 | - | 88,06 | 32,43 | 59,46 | 0,00 | - | 93,75 | - | 87,50 | - |
| Rio Piracicaba/Jaguari | - | - | 100,00 | - | - | 100,00 | 0,00 | - | 100,00 | 58,82 | 100,00 | 100,00 | 75,00 | - |

A TABELA 16 refere-se aos valores de Acurácia do Usuário, criada para as classes presentes em cada uma das 14 bacias hidrográficas analisadas. Com base na análise desta, pode-se observar que a classe Água foi a que apresentou maior acuracidade do usuário, com valores entre 96,30 e 100%, nas 14 bacias analisadas. Para Urbanização também foram encontrados valores acima de 85% em todas as bacias, variando entre 95,92 e 100%. Do mesmo modo, a classe Outros apresentou altos valores de acurácia do produtor, excetuando-se a da bacia do Rio Pardo (31,25%). Ao analisar a matriz de confusão confeccionada para esta bacia, constatou-se que este baixo valor está relacionado ao alto número de polígonos mapeados como Outros e que, na constatação de campo, eram classes diferentes dessas, sendo que 22 polígonos de Eucalipto analisados em campo foram mapeados como Outros. As bacias que apresentaram menores valores de acurácia do usuário para alguma classe nela contida foram as do Rio Piracicaba/Jaguari, Rio Pardo e do Rio São Francisco, com valores de 16,67 31,25%, 33,33% e 36,21%, para as classes Floresta Estacional Semidecidual, Outros, Floresta Estacional Semidecidual e Campo, respectivamente. Igualmente à TABELA 12, os valores zero, encontrados na TABELA 13, referem-se às classes que não apresentaram nenhum polígono em concordância entre a classe encontrada no mapeamento e no campo.

TABELA 16. Acurácia do Usuário para as classes temáticas presentes em bacias hidrográficas. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| Bacias Hidrográficas | Acurácia do Usuário (%) | | | | | | | | | | | | | |
|------------------------|-------------------------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | A | C | CC | CR | CO | CT | E | FD | FS | FO | O | P | U | V |
| Rio Buranhém | - | - | 100,00 | - | - | - | - | - | 100,00 | - | 100,00 | - | 100,00 | - |
| Rio Itanhém | - | - | - | - | - | - | 100,00 | - | 84,62 | - | 96,25 | - | 100,00 | - |
| Rio Jequitinhonha | 100,00 | 56,25 | 67,69 | 80,00 | - | 82,12 | 98,21 | 96,02 | 40,00 | - | 73,41 | 86,67 | 100,00 | - |
| Rio Jucuruçu | - | - | - | - | - | - | - | - | 33,33 | - | 100,00 | - | 100,00 | - |
| Rio Paranaíba | 100,00 | 87,50 | - | - | 89,66 | 100,00 | 100,00 | - | 76,47 | - | 95,04 | 75,00 | 100,00 | 100,00 |
| Rio São Mateus | - | - | - | 50,00 | - | - | - | - | 91,67 | - | 100,00 | - | 100,00 | - |
| Rio Doce | 100,00 | 52,94 | - | 100,00 | 84,62 | 100,00 | 85,23 | - | 96,14 | 100,00 | 92,35 | 0,00 | 99,13 | 78,57 |
| Rio Grande | 100,00 | 43,04 | 100,00 | 70,59 | 90,91 | 83,33 | 85,19 | 100,00 | 80,95 | 93,94 | 79,55 | 0,00 | 99,22 | 100,00 |
| Rio São Francisco | 96,30 | 36,21 | 74,74 | 96,97 | 100,00 | 86,87 | 91,11 | 96,49 | 94,12 | - | 89,55 | 100,00 | 98,43 | 92,86 |
| Rio Itabapoana | - | - | - | - | - | - | 0,00 | - | 83,33 | - | 88,57 | - | 100,00 | - |
| Rio Mucuri | 100,00 | - | 41,67 | - | - | 100,00 | 0,00 | - | 92,86 | - | 94,70 | - | 100,00 | - |
| Rio Paraíba do Sul | 100,00 | 50,00 | - | 100,00 | - | - | 0,00 | - | 76,98 | 100,00 | 92,97 | 0,00 | 95,92 | - |
| Rio Pardo | - | 66,67 | 82,61 | 100,00 | - | 89,39 | 100,00 | 100,00 | 0,00 | - | 31,25 | - | 100,00 | - |
| Rio Piracicaba/Jaguari | - | - | 100,00 | - | - | 100,00 | 0,00 | - | 16,67 | 90,91 | 72,22 | 37,50 | 100,00 | - |

4 CONCLUSÕES

A proporcionalização do número de unidades amostrais por meio da relação entre a área ocupada por cada classe em determinada unidade de controle e a área total ocupada por esta classe em toda a área de avaliação, garante que todas as classes sejam bem representadas pela amostra.

Constatou-se, no presente estudo, que algumas classes não apresentaram proporção de polígonos corretamente classificados superior a 85%. Este é um resultado esperado quando se avalia a exatidão por classe temática.

Para as classes Eucalipto e Floresta Ombrófila, menos do que 50% dos polígonos visitados em campo foram compatíveis com o Mapeamento. Estas classes merecem atenção especial no processo de classificação nas condições em que o trabalho foi desenvolvido. Ao analisar a precisão da exatidão, avaliada para estas classes, detectou-se que ela se apresentou baixa, em função do número de polígonos classificados incorretamente.

Foi detectado no estudo que, mesmo a tomada de poucas unidades amostrais criam intervalos de confiança para a proporção de polígonos corretamente classificados superiores a 85%. Embora exista um déficit no número de unidades amostrais para estas bacias, as poucas tomadas retrataram a variabilidade ocorrente com a variável de interesse, possibilitando, assim, intervalos com proporções maiores que 85%. Não se pode garantir, porém, um erro máximo de 10%.

Foram apresentados intervalos de confiança para as 14 bacias amostradas e quatro deles não englobam a exatidão *a priori* de 85%, em função dos altos valores de erro nestas bacias.

Para todas as bacias hidrográficas, o coeficiente Exatidão Global foi maior que o coeficiente Kappa, em virtude do fato do último levar em consideração toda a matriz de confusão, enquanto que o primeiro considera

apenas o número total de pontos amostrados e o número total de pontos corretamente classificados. Além disso, o Kappa para todas as bacias amostradas foi menor quando comparado ao coeficiente Tau. Por sua vez, o Tau determinado foi menor que a Exatidão Global para todas as bacias, apresentando valores mais próximos aos da Exatidão Global quando o comparado com o Kappa.

5 REFERÊNCIAS BIBLIOGRÁFICAS

- BELART, P.; PINILLA, C.; ARIZA, F. J. **Control de calidad de clasificaciones de imágenes por el método de los itinerarios**. Jaén: Escuela Politécnica Superior Universidad de Jaén, 2002. 223 p.
- CARVALHO, L. M. T.; CLEVERS, J. G. P. W.; SKIDMORE, A. K.; JONG, S. M. de. Selection of imagery data and classifiers for mapping Brazilian semideciduous Atlantic forests. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 5, n. 3, p. 173-186, 2004.
- COHEN, J. A coefficient of agreement for nominal scales. **Educational and Measurement**, New York, v. 20, n. 1, p. 37-46, 1960.
- CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. Boca Raton: Lewis, 1999. 137 p.
- EDWARDS, T. C.; MOISEN, G. G.; CUTLER, D. R. Assessing map accuracy in a ecoregion-scale cover-map. **Remote Sensing Environment**, Washington, v. 63, n. 1, p. 73-83, Jan. 1998.
- HELLDEN, U.; STERN, M. Evaluation of landsat imagery and digital data for monitoring desertification indicators in Tunisia. In: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT, 14., 1980, San Jose. **Proceedings...** Ann Arbor: Environmental Research Institute of Michigan, 1980. p. 1601-1611.
- JANSSEN, L. F.; LUCAS, F. J. M.; WEL, F. J. M. van der. Accuracy assessment of satellite derived land-cover data: a review. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 60, n. 4, p. 419-426, Apr. 1994.
- JENSEN, J. R. **Introductory digital image processing**. New Jersey: Prentice-Hall, 1986. 379 p.
- KLECKA, W. R. **Discriminant analysis**. Beverly Hills: SAGE, 1980. 71 p.
- LINGNAU, C.; ANTUNES, A. F. B. Object oriented analysis and semantic network for high resolution image classification. **Boletim de Ciências Geodésicas**, Curitiba, v. 9, n. 2, p. 233-243, 2003.

MEYER, P. L. **Probabilidade**: aplicações à estatística. 2. ed. Rio de Janeiro: LTC, 1984. 426 p.

PIMENTEL, D.; PLANTIER, T.; BICA, V.; NAVARRO, V.; CAETANO, M. Aplicação da detecção remota à actualização da distribuição parcelar do cadastro geométrico da propriedade rústica no concelho do seixal. In: ENCONTRO DE UTILIZADORES DE INFORMAÇÃO GEOGRÁFICA, 10., 2008, Oeiras. **Actas...** Oeiras: ESIG, 2008. In press.

RICHARDS, J. A. **Remote sensing digital image analysis**: an introduction. Berlin: Springer-Verlag, 1986. 281 p.

SCOLFORO, J. R. S.; CARVALHO, L. M. T. **Mapeamento e inventário da flora nativa e dos reflorestamentos de Minas Gerais**. Lavras: UFLA, 2006. 288 p.

STEHMAN, S. V. Estimating the Kappa coefficient and its variance under stratified random sampling. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 62, n. 4, p. 401-402, Apr. 1996.

STEHMAN, S. V. Selecting and interpreting measures of thematic classification accuracy. **Remote Sensing of Environment**, Washington, v. 62, n. 1, p. 77-89, Oct. 1997.

STEHMAN, S. V. Statistical rigor and practical utility in thematic map accuracy assessment. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 67, n. 6, p. 727-734, June 2001.

WATZLAWICK, L. F.; KIRCHNER, F. F.; SANQUETTA, C. R.; SCHUMACHER, M. V. Fixação de carbono em floresta ombrófila mista em diferentes estágios de regeneração. In: _____. **As florestas e o carbono**. Curitiba: UFPR, 2002. cap. 8, p. 153-173.

CONSIDERAÇÕES GERAIS

A proporcionalização obtida com a distribuição das unidades amostrais, de acordo com a área que as classes temáticas ocupam em determinada unidade de controle de amostragem, em relação à área total destas classes, garantiu que todas as classes apresentassem um número mínimo de unidades para atender o erro pré-estabelecido e criar intervalos de confiança para a proporção de unidades corretamente classificadas. Estes intervalos são estatisticamente mais consistentes na avaliação da precisão de classificações temáticas.

Classes temáticas com grande número de unidades amostrais, concordantes entre o mapeamento e a referência de campo, superestimam a precisão e acurácia quando todo o mapeamento é analisado. Deve-se, então, lançar mão da utilização de coeficientes para classes individuais, como a Acurácia do Produtor e do Usuário, para detalhar os erros ocorridos em cada classe temática e apresentar com maior fidedignidade os resultados da classificação temática.

Classes que apresentam características peculiares nos atributos da imagem são classificadas corretamente com maior facilidade. Classes com proximidade espectral ou que, no momento da tomada da imagem apresentavam sombreamento em determinada face do terreno, em virtude da inclinação deste, são afetadas na classificação, apresentando baixos valores de concordância e amplo intervalo de confiança para esta proporção. A presença de sub-bosque com árvores nativas em reflorestamentos, ou ainda a condução destes por regime de talhadia, aumentam a confusão entre classes e, conseqüentemente, a concordância e a amplitude do intervalo de confiança para a variável analisada.

A definição de um sistema de amostragem, muitas vezes não detalhados em estudos da qualidade de mapeamentos temáticos, é fundamental para o

entendimento e determinação da acurácia e precisão da classificação, com rigor estatístico.

Decorrente dos altos custos com a obtenção da amostra de referência, a escolha do procedimento de amostragem correto torna-se fundamental em qualquer processo de determinação da qualidade de mapeamentos temáticos de grandes áreas. Estudos envolvendo este tipo de área devem ser realizados com maior frequência e detalhamento para a obtenção de metodologias mais viáveis economicamente.

A utilização das rodovias como base para se tomar a amostra é eficiente, quando a malha viária da área em questão possui grandes extensões e boa distribuição espacial, favorecendo a distribuição das unidades amostrais em toda a área de estudo.

Recomendam-se estudos para obtenção de uma rede de colaboradores para realização de tomadas de dados de referência em campo, por unidades de controle definidas, como bacias, sub-bacias, núcleos regionais ou municipais de órgãos ambientais. Esta rede, apoiada por um protocolo de identificação das classes em campo, pode auxiliar na redução significativa dos custos da avaliação da acurácia e precisão de mapeamentos temáticos de grandes áreas.

Mesmo dimensionando o tamanho da amostra para cada classe temática por meio de fórmulas adequadas e admitindo proporção de unidades corretamente classificadas *a priori* como 85%, a concordância obtida *a posteriori* entre o mapeamento e a referência de campo influencia fortemente o tamanho da amostra quando se aplica novamente a formulação.

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, mostraram-se adequadas, uma vez que a acurácia *a priori* foi inferior à determinada *a posteriori* para dois dos três coeficientes testados (Exatidão Global e Tau) e está

contida no intervalo de confiança para a precisão real de todo o “Mapeamento da Flora e dos Reflorestamentos de Minas Gerais”.

ANEXOS

| ANEXO A | Página |
|---|--------|
| TABELA 1A. Matriz de confusão para a bacia hidrográfica do Rio Buranhém. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 105 |
| TABELA 2A. Matriz de confusão para a bacia hidrográfica do Rio Jequitinhonha. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 105 |
| TABELA 3A. Matriz de confusão para a bacia hidrográfica do Rio São Mateus. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 106 |
| TABELA 4A. Matriz de confusão para a bacia hidrográfica do Rio Jucuruçu. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 106 |
| TABELA 5A. Matriz de confusão para a bacia hidrográfica do Rio Itanhém. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 107 |

| | |
|---|-----|
| TABELA 6A. Matriz de confusão para a bacia hidrográfica do Rio Itabapoana. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 107 |
| TABELA 7A. Matriz de confusão para a bacia hidrográfica do Rio Mucuri. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 108 |
| TABELA 8A. Matriz de confusão para a bacia hidrográfica do Rio Paraíba do Sul. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 108 |
| TABELA 9A. Matriz de confusão para a bacia hidrográfica do Rio Pardo. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 109 |
| TABELA 10A. Matriz de confusão para a bacia hidrográfica do Rio Piracicaba/Jaguari. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 109 |
| TABELA 11A. Matriz de confusão para a bacia hidrográfica do Rio Doce. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas..... | 110 |

- TABELA 12A. Matriz de confusão para a bacia hidrográfica do Rio São Francisco. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas.....111
- TABELA 13A. Matriz de confusão para a bacia hidrográfica do Rio Paranaíba. Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas.....112
- TABELA 14A. Matriz de confusão para a bacia hidrográfica do Rio Grande. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas.....113

TABELA 1A. Matriz de confusão para a bacia hidrográfica do Rio Buranhém. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | |
|------------|-------|---------------------|----|----|---|----|
| | | Classe | CC | FS | O | U |
| Mapeamento | CC | 1 | | | | 1 |
| | FS | | 1 | | | 1 |
| | O | | | 15 | | 15 |
| | U | | | | 1 | 1 |
| | Total | 1 | 1 | 15 | 1 | 18 |

TABELA 2A. Matriz de confusão para a bacia hidrográfica do Rio Jequitinhonha. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | REFERÊNCIA DE CAMPO | | | | | | | | | | | | |
|------------|-------|---------------------|----|-----|----|-----|-----|-----|----|-----|----|----|----|-------|
| | | CLASSE | A | C | CC | CR | CT | E | FD | FS | O | P | U | TOTAL |
| MAPEAMENTO | A | 10 | | | | | | | | | | | | 10 |
| | C | | 9 | | 1 | 3 | 2 | | | 1 | | | | 16 |
| | CC | | | 88 | | 3 | 4 | 26 | 6 | 3 | | | | 130 |
| | CR | | | 4 | 36 | 1 | 3 | | 1 | | | | | 45 |
| | CT | | 1 | 2 | | 124 | 2 | 9 | 13 | | | | | 151 |
| | E | | 1 | | | | 110 | | | 1 | | | | 112 |
| | FD | | | 2 | | 3 | | 241 | | 5 | | | | 251 |
| | FS | | | 6 | | 25 | 4 | 23 | 44 | 8 | | | | 110 |
| | O | 1 | 1 | 1 | 4 | 16 | 38 | 83 | 2 | 403 | | | | 549 |
| | P | | 1 | | | | 1 | | | | | 13 | | 15 |
| | U | | | | | | | | | | | | 35 | 35 |
| | TOTAL | 11 | 13 | 103 | 41 | 175 | 164 | 382 | 66 | 421 | 13 | 35 | | 1424 |

TABELA 3A. Matriz de confusão para a bacia hidrográfica do Rio São Mateus. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | |
|------------|-------|---------------------|----|-----|----|-----|
| | | Classe | CR | FS | O | U |
| Mapeamento | CR | 1 | | 1 | | 2 |
| | FS | | 11 | 1 | | 12 |
| | O | | | 103 | | 103 |
| | U | | | | 11 | 11 |
| | Total | 1 | 11 | 105 | 11 | 128 |

TABELA 4A. Matriz de confusão para a bacia hidrográfica do Rio Jucuruçu. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | |
|------------|-------|---------------------|----|---|----|
| | | Classe | FS | O | U |
| Mapeamento | FS | 1 | 2 | | 3 |
| | O | | 29 | | 29 |
| | U | | | 1 | 1 |
| | Total | 1 | 31 | 1 | 33 |

TABELA 5A. Matriz de confusão para a bacia hidrográfica do Rio Itanhém. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | |
|------------|-------|---------------------|----|----|---|----|
| | | Classe | E | FS | O | U |
| Mapeamento | E | 1 | | | | 1 |
| | FS | | 11 | 2 | | 13 |
| | O | 3 | | 77 | | 80 |
| | U | | | | 4 | 4 |
| | Total | 4 | 11 | 79 | 4 | 98 |

TABELA 6A. Matriz de confusão para a bacia hidrográfica do Rio Itabapoana. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | |
|------------|-------|---------------------|---|----|---|----|
| | | Classe | E | FS | O | U |
| Mapeamento | E | 0 | | | | 0 |
| | FS | | 5 | 1 | | 6 |
| | O | 2 | 2 | 31 | | 35 |
| | U | | | | 4 | 4 |
| | Total | 2 | 7 | 32 | 4 | 45 |

TABELA 7A. Matriz de confusão para a bacia hidrográfica do Rio Mucuri. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | | | | Total |
|------------|-------|---------------------|---|----|----|----|-----|----|-------|
| | | Classe | A | CC | CT | E | FS | O | |
| Mapeamento | A | 1 | | | | | | | 1 |
| | CC | | 5 | | | 3 | 4 | | 12 |
| | CT | | | 1 | | | | | 1 |
| | E | | | | 0 | | | | 0 |
| | FS | | | | | 65 | 5 | | 70 |
| | O | | | | 7 | 7 | 268 | 1 | 283 |
| | U | | | | | | | 16 | 16 |
| | Total | 1 | 5 | 1 | 7 | 75 | 277 | 17 | 383 |

TABELA 8A. Matriz de confusão para a bacia hidrográfica do Rio Paraíba do Sul. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | | | | | | Total |
|------------|-------|---------------------|---|---|----|-----|----|-----|---|----|-------|
| | | Classe | A | C | CR | E | FS | FO | O | P | |
| Mapeamento | A | 7 | | | | | | | | | 7 |
| | C | | 2 | 1 | | | | 1 | | | 4 |
| | CR | | | 1 | | | | | | | 1 |
| | E | | | | 0 | | | | 1 | | 1 |
| | FS | | | | 8 | 107 | 20 | 4 | | | 139 |
| | FO | | | | | | 24 | | | | 24 |
| | O | 3 | | 3 | 7 | 21 | 3 | 489 | | | 526 |
| | P | | | | | | | | 0 | | 0 |
| | U | | | | | | | 2 | | 47 | 49 |
| | Total | 10 | 2 | 5 | 15 | 128 | 47 | 496 | 1 | 47 | 751 |

TABELA 9A. Matriz de confusão para a bacia hidrográfica do Rio Pardo. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | | | | | | |
|------------|--------|---------------------|----|----|----|----|----|----|----|---|-------|
| Mapeamento | Classe | C | CC | CR | CT | E | FD | FS | O | U | Total |
| | C | 6 | | | | 1 | 1 | | 1 | | 9 |
| | CC | | 19 | | 1 | 1 | 2 | | | | 23 |
| | CR | | | 1 | | | | | | | 1 |
| | CT | | | | 59 | | 7 | | | | 66 |
| | E | | | | | 12 | | | | | 12 |
| | FD | | | | | | 22 | | | | 22 |
| | FS | | | | 5 | 1 | 2 | 0 | | | 8 |
| | O | 2 | 3 | | 2 | 22 | 3 | | 15 | 1 | 48 |
| | U | | | | | | | | | 7 | 7 |
| | Total | 8 | 22 | 1 | 67 | 37 | 37 | 0 | 16 | 8 | 196 |

TABELA 10A. Matriz de confusão para a bacia hidrográfica do Rio Piracicaba/Jaguari. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | | | | | |
|------------|--------|---------------------|----|---|----|----|----|---|---|-------|
| Mapeamento | Classe | CC | CT | E | FS | FO | O | P | U | Total |
| | CC | 1 | | | | | | | | 1 |
| | CT | | 3 | | | | | | | 3 |
| | E | | | 0 | | | | | | 0 |
| | FS | | | 2 | 1 | 3 | | | | 6 |
| | FO | | | 1 | | 10 | | | | 11 |
| | O | | | | | 4 | 13 | | 1 | 18 |
| | P | | | 5 | | | | 3 | | 8 |
| | U | | | | | | | | 3 | 3 |
| | Total | 1 | 3 | 8 | 1 | 17 | 13 | 3 | 4 | 50 |

TABELA 11A. Matriz de confusão para a bacia hidrográfica do Rio Doce. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| Classe | Referência de campo | | | | | | | | | | | | Total |
|--------|---------------------|----|----|----|----|-----|-----|----|------|---|-----|----|-------|
| | A | C | CR | CO | CT | E | FS | FO | O | P | U | V | |
| A | 29 | | | | | | | | | | | | 29 |
| C | | 9 | 6 | | | | | | 2 | | | | 17 |
| CR | | | 6 | | | | | | | | | | 6 |
| CO | | | | 11 | | | | | 1 | | | 1 | 13 |
| CT | | | | | 9 | | | | | | | | 9 |
| E | | | | | | 75 | 10 | | 1 | 2 | | | 88 |
| FS | | | | | | 18 | 548 | | 2 | | | 2 | 570 |
| FO | | | | | | | | 3 | | | | | 3 |
| O | 2 | 3 | 7 | 1 | | 30 | 40 | | 1099 | | 5 | 3 | 1190 |
| P | | | | | | | | | | 0 | | | 0 |
| U | | | | | | 1 | | | | | 114 | | 115 |
| V | | | | 1 | | 1 | 1 | | | | | 11 | 14 |
| Total | 31 | 12 | 19 | 13 | 9 | 125 | 599 | 3 | 1105 | 2 | 119 | 17 | 2054 |

TABELA 12A. Matriz de confusão para a bacia hidrográfica do Rio São Francisco. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Verdade de Campo | | | | | | | | | | | | | |
|------------|----|------------------|----|----|----|-----|-----|-----|-----|------|------|-----|----|------|-------|
| Classe | | A | C | CC | CR | CO | CT | E | FD | FS | O | P | U | V | Total |
| Mapeamento | A | 52 | | | | 1 | | | | | 1 | | | | 54 |
| | C | | 42 | 2 | 3 | | 12 | 3 | 4 | 2 | 44 | | 1 | 3 | 116 |
| | CC | | | 71 | | | 7 | | 10 | | 7 | | | | 95 |
| | CR | | | | 32 | | | | | 1 | | | | | 33 |
| | CO | | | | | 7 | | | | | | | | | 7 |
| | CT | 1 | | | | | 688 | 9 | 18 | 31 | 31 | | 1 | 13 | 792 |
| | E | | | | | | 9 | 164 | | 2 | 5 | | | | 180 |
| | FD | | | | | | 5 | | 275 | | 5 | | | | 285 |
| | FS | | 1 | 1 | | | 11 | 2 | 1 | 352 | 3 | 1 | | 2 | 374 |
| | O | 6 | 6 | 17 | 11 | | 44 | 58 | 20 | 49 | 2039 | | 14 | 13 | 2277 |
| | P | | | | | | | | | | | 6 | | | 6 |
| | U | | | | | | | | | | 1 | 2 | | 188 | 191 |
| | V | | | | | | | | | | 2 | | | | 26 |
| Total | 59 | 49 | 91 | 46 | 8 | 776 | 236 | 328 | 440 | 2137 | 7 | 204 | 57 | 4438 | |

TABELA 13A. Matriz de confusão para a bacia hidrográfica do Rio Paranaíba. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | | | | | | | |
|------------|--------|---------------------|---|----|---|----|----|-----|----|----|----|-------|
| | Classe | A | C | CO | C | E | FS | O | P | U | V | Total |
| Mapeamento | A | 6 | | | | | | | | | | 6 |
| | C | | 7 | | | | | 1 | | | | 8 |
| | CO | | | 26 | | | | 2 | | | 1 | 29 |
| | C | | | | 3 | | | | | | | 3 |
| | E | | | | | 4 | | | | | | 4 |
| | FS | | | | | 1 | 26 | 7 | | | | 34 |
| | O | | | 2 | 5 | 1 | | 326 | 5 | | 4 | 343 |
| | P | | | | 1 | 10 | | 1 | 36 | | | 48 |
| | U | | | | | | | | | 14 | | 14 |
| | V | | | | | | | | | | 19 | 19 |
| | Total | 6 | 7 | 28 | 9 | 16 | 26 | 337 | 41 | 14 | 24 | 508 |

TABELA 14A. Matriz de confusão para a bacia hidrográfica do Rio Grande. A - Água, C - Campo, CC - Campo Cerrado, CR - Campo Rupestre, CO - Cerradão, CT - Cerrado, E - Eucalipto, FD - Floresta Decidual, FS - Floresta Estacional Semidecidual, FO - Floresta Ombrófila, O - Outros, P - Pinus, U - Urbanização e V - Veredas

| | | Referência de campo | | | | | | | | | | | | | | |
|------------|--------|---------------------|----|----|----|----|----|-----|----|-----|----|-----|---|-----|---|-------|
| Mapeamento | Classe | A | C | CC | CR | CO | CT | E | FD | FS | FO | O | P | U | V | Total |
| | | A | 66 | | | | | | | | | | | | | |
| | C | | 34 | 4 | 4 | | 1 | 8 | | 13 | 4 | 11 | | | | 79 |
| | CC | | | 7 | | | | | | | | | | | | 7 |
| | CR | | | 2 | 12 | | 2 | | | | | 1 | | | | 17 |
| | CO | | | | | 20 | | | | | | 1 | | | 1 | 22 |
| | CT | | | | | | 10 | | | 2 | | | | | | 12 |
| | E | | | | | | | 23 | | 2 | | 2 | | | | 27 |
| | FD | | | | | | | | 1 | | | | | | | 1 |
| | FS | | | | 1 | | 8 | 34 | | 323 | 30 | 3 | | | | 399 |
| | FO | | | | | | | | | 2 | 31 | | | | | 33 |
| | O | 10 | 12 | 4 | 12 | 1 | 26 | 43 | | 70 | 17 | 813 | 2 | 11 | 1 | 1022 |
| | P | | | | | | | | | | | | 0 | | | 0 |
| | U | | | | | | | | | | | 1 | | 127 | | 128 |
| | V | | | | | | | | | | | | | | 3 | 3 |
| | Total | 76 | 46 | 17 | 29 | 21 | 47 | 108 | 1 | 412 | 82 | 832 | 2 | 138 | 5 | 1816 |

