

Estudo do comportamento dos índices de Exatidão Global, Kappa e Tau, comumente usados para avaliar a classificação de imagens do sensoriamento remoto

Geíza Coutinho Figueiredo
Carlos Antonio Oliveira Vieira

Universidade Federal de Viçosa – UFV
Departamento de Engenharia Civil
Avenida P. H. Rolfs s/n - Campus UFV - 36570-000 - Viçosa - MG, Brasil
geizacf@hotmail.com
carlos.vieira@ufv.br

Abstract: Evaluation of the accuracy plays an important role in the process data analysis of remote sensing products. The evaluation of the accuracy can be obtained through derived measures from a confusion matrix. This article describes some of these measures, such as Overall Accuracy, Kappa and Tau indexes, which are commonly used to evaluate the classification of remote sensing images. A comparative study about the behavior of these indexes was accomplished and a preliminary discussion is presented. Results show that despite Kappa and Tau indexes are computed following different formulations and assumptions, the present consistently very similar values, which indicates that both measures could indiscriminately be used for remotely sensed accuracy assessment.

Palavras-chave: classification images, thematic accuracy, index Kappa, index Tau, accuracy global, classificação de imagens, precisão temática, índice Kappa, índice Tau, exatidão global.

1. Introdução

A exatidão de um mapa indica a proximidade de uma determinada medida ao seu valor real, logo, a confiabilidade de um mapa está vinculada a sua exatidão. Neste contexto, é necessário realizar algum procedimento estatístico, no produto de uma classificação de imagens digitais, para determinar a acurácia ou exatidão desta classificação (Bernardes, 2006).

No processo de análise dos dados do sensoriamento remoto, um passo fundamental é a avaliação da precisão temática. Os usuários necessitam saber quão confiáveis são os dados provenientes dos mapas temáticos, derivados da classificação de um produto do sensoriamento remoto, e através da matriz de confusão é possível derivar medidas e consequentemente verificar erros oriundos do processo de atribuição dos pixels a determinadas classes (Vieira, 2000).

A avaliação da acurácia pode ser obtida por meio de coeficientes de concordância derivados da matriz de confusão, sendo que estes podem ser expressos como concordância total ou para classes individuais. Congalton (1991) relata que o uso do coeficiente Kappa (K) é satisfatório na avaliação da precisão de uma classificação temática, pelo fato de levar em consideração toda a matriz de confusão no seu cálculo, inclusive os elementos de fora da diagonal principal, os quais representam as discordâncias na classificação, diferentemente da exatidão global, por exemplo, que utiliza somente os elementos diagonais (concordância real).

Com relação ao coeficiente Kappa, Foody (1992) observou que o grau de concordância por chance poderia estar sendo superestimado, pelo fato de incluir também a concordância real, e por causa disso a magnitude de Kappa não refletiria a concordância presente na classificação, apenas descontada a casualidade. Na tentativa de corrigir essa deficiência no cálculo do índice Kappa, Ma e Redmond (1995) propuseram um outro índice para a medição da precisão da classificação, o índice Tau (T).

Apesar de esses índices serem amplamente usados pela comunidade científica, não existem estudos que indiquem como esses índices se comportam, quando da variação do número de classes informacionais e/ou do número de padrões de validação - considerando que

esses índices levam em consideração ou o número total de classes ou o número total de padrões de validação; ou mesmo se houver discrepância entre os valores amostrados, para validação, entre as classes informacionais.

Este trabalho tem como objetivo fazer uma descrição dos coeficientes de concordância que são comumente usados para avaliar a classificação de imagens do sensoriamento remoto e também observar o comportamento dos mesmos variando o número de classes informacionais na classificação, variando o número total de pixels e também quando ocorre discrepância entre o número de amostras entre as classes informacionais. Uma motivação para observar esses comportamentos, é para estudar se as diferenças entre esses índices de exatidão são constantes ou não dentro dos níveis do desempenho da classificação, e o que isso implica na escolha de qual índice utilizar.

2. Matriz de confusão (ou matriz de erro)

O método padrão para avaliação da precisão temática atualmente tem sido índices derivados da matriz de confusão. A matriz de confusão fornece a base para descrever a precisão da classificação e caracterizar os erros, ajudando a refinar a classificação. De uma matriz de confusão podem ser derivadas várias medidas de precisão da classificação, sendo a exatidão global uma das mais conhecidas (Foody, 2002).

A matriz de confusão é formada por um arranjo quadrado de números dispostos em linhas e colunas que expressam o número de unidades de amostras de uma categoria particular relativa – inferida por um classificador (ou regra de decisão), comparado com a categoria atual verificada no campo (Congalton, 1991). Normalmente abaixo das colunas representa-se o conjunto de dados de referência que é comparado com os dados do produto da classificação que são representados ao longo das linhas. Os elementos da diagonal principal (em negrito) indicam o nível de acerto, ou concordância, entre os dois conjuntos de dados. A **Tabela 1** mostra a representação de uma matriz de confusão.

Tabela 1 – Representação matemática de uma matriz de confusão.

Classificação	Dados de referência			Total nas linhas n_{i+}
	1	2	c	
1	x_{11}	x_{12}	x_{1c}	x_{1+}
2	x_{21}	x_{22}	x_{2c}	x_{2+}
c	x_{c1}	x_{c2}	x_{cc}	x_{c+}
Total nas colunas n_{+i}	x_{+1}	x_{+2}	x_{+c}	n

Fonte: Adaptada de Bernardes (2006).

3. Medidas e índices derivados da matriz de confusão

As medidas derivadas da matriz de confusão são: a exatidão global, precisão de classe individual, precisão de produtor, precisão de usuário e índice Kappa, entre outros. A exatidão global é calculada dividindo a soma da diagonal principal da matriz de erros x_{ii} , pelo número total de amostras coletadas n , ou seja:

$$G = \frac{\sum_{i=1}^c x_{ii}}{n} \quad (1)$$

A distribuição da precisão ao longo das categorias individuais não é apresentada na precisão global, entretanto a precisão de uma categoria individual é obtida através da divisão do número total de amostras classificadas corretamente naquela categoria pelo número total de amostras daquela categoria. Congalton e Green (1999) descrevem os cálculos associados com estas medidas.

A precisão de produtor e de usuário são maneiras de representar a precisão de uma categoria ou classe individualmente. A precisão de produtor refere-se às amostras que não foram classificadas corretamente como pertencendo àquela categoria sendo omitidas de sua categoria correta. E a precisão de usuário indica a probabilidade que um *pixel* classificado na imagem de fato representa aquela categoria no campo.

A análise de Kappa é uma técnica multivariada discreta usada na avaliação da precisão temática e utiliza todos os elementos da matriz de confusão no seu cálculo. O coeficiente Kappa (K) é uma medida da concordância real (indicado pelos elementos diagonais da matriz de confusão) menos a concordância por chance (indicado pelo produto total da linha e coluna, que não inclui entradas não reconhecidas), ou seja, é uma medida do quanto à classificação está de acordo com os dados de referência. O coeficiente Kappa pode ser calculado através da seguinte equação:

$$K = \frac{n \sum_{i=1}^c x_{ii} - \sum_{i=1}^c x_{i+} x_{+i}}{n^2 - \sum_{i=1}^c x_{i+} x_{+i}} \quad (2)$$

Onde K é uma estimativa do coeficiente Kappa; x_{ii} é o valor na linha i e coluna i ; x_{i+} é a soma da linha i e x_{+i} é a soma da coluna i da matriz de confusão; n é o número total de amostras e c o número total de classes.

Embora o coeficiente Kappa seja muito utilizado na avaliação da exatidão de mapeamento, não existe uma fundamentação teórica para recomendar quais os níveis mínimos aceitáveis deste coeficiente numa classificação. Entretanto, a **Tabela 2** apresenta níveis de desempenho da classificação para o valor de Kappa obtido, normalmente aceitos pela comunidade científica.

Tabela 2 – Índice Kappa e o correspondente desempenho da classificação.

Índice Kappa	Desempenho
< 0	Péssimo
$0 < k \leq 0,2$	Ruim
$0,2 < k \leq 0,4$	Razoável
$0,4 < k \leq 0,6$	Bom
$0,6 < k \leq 0,8$	Muito Bom
$0,8 < k \leq 1,0$	Excelente

Fonte: Fonseca (2000).

Ao redor do valor de Kappa podem ser calculados intervalos de confiança usando a variância da amostra (var) e o fato de que a distribuição estatística do Kappa é normalmente assintótica. Congalton e Green (1999) sugerem meios de testar a significação estatística do Kappa para uma única matriz de confusão, através da variância, a fim de determinar se o nível de acerto da classificação e os dados de referência são significativamente maior que zero. O teste estatístico para testar a significação de uma única matriz de confusão é determinado pela equação a seguir:

$$Z = \frac{k}{\sqrt{var(k)}} \quad (3)$$

Onde Z é unificado e normalmente distribuído e var é a grande variância da amostra do coeficiente Kappa, que pode ser calculado usando o método de *Delta* como segue:

$$\text{var}(k) = \frac{1}{n} \left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right] \quad (4)$$

Onde

$$\theta_1 = \frac{1}{n} \sum_{i=1}^c x_{ii}, \quad \theta_2 = \frac{1}{n^2} \sum_{i=1}^c x_{i+} x_{+i}, \quad \theta_3 = \frac{1}{n^2} \sum_{i=1}^c x_{ii} (x_{i+} + x_{+i}) \quad e$$

$$\theta_4 = \frac{1}{n^3} \sum_{i=1}^c \sum_{j=1}^c x_{ij} (x_{j+} + x_{+j})^2 .$$

Se $Z \geq Z_{\alpha/2}$ a classificação é significativamente melhor que uma distribuição aleatória, onde $\alpha/2$ é o nível de confiança nos dois lados da curva no teste Z e o número de graus de liberdade é assumido ser infinito.

Além dos índices de exatidão global e Kappa, que são comumente usados, Ma e Redmond (1995) introduziram na comunidade do sensoriamento remoto o índice *Tau*, que fornece uma medida quantitativa relativamente precisa e intuitiva sobre a acurácia da classificação. O coeficiente *Tau* (T) é superficialmente similar ao Kappa e pode ser calculado como segue:

$$T = \frac{P_0 - P_r}{1 - P_r} \quad (5)$$

Onde P_0 seria a concordância real e P_r a concordância aleatória expressos da seguinte

forma: $P_0 = \frac{\sum_{i=1}^c x_{ii}}{n}$ e $P_r = \sum_{i=1}^c P_{i+} P_i = \frac{1}{n^2} \sum_{i=1}^c x_{i+} x_{+i}$.

Sendo $P_{i+} = (x_{i+} / n)$ a distribuição marginal dos dados de referência, $P_i = (x_i / n)$ as probabilidades *a priori* para cada classe. Quando as probabilidades *a priori* para as classes forem iguais, ou seja, $x_i = n/c$ temos:

$$T = \frac{P_0 - \frac{1}{c}}{1 - \frac{1}{c}} \quad (6)$$

Segundo Ma e Redmond (1995), devido a concordância aleatória (P_r) ser independente dos elementos da matriz de erros e poder ser calculada antes da classificação, ela é tratada como uma constante. Sendo $\sigma^2(P_r) = 0$ e $\sigma^2(P_0) = \frac{1}{n} P_0(1 - P_0)$, e aplicando a lei de propagação de variâncias, temos a variância de *Tau*:

$$\sigma^2(T) = \frac{P_0(1 - P_0)}{n(1 - P_r)^2} \quad (7)$$

Para testar a significação estatística do índice *Tau*, procede-se da mesma forma descrita para o índice Kappa.

4. Comparação entre os índices de exatidão Global, Kappa e Tau

Objetivando analisar o comportamento dos índices de Exatidão Global, Kappa e *Tau*, criaram-se várias matrizes de confusão hipotéticas, variando primeiramente o número de classes informacionais para cada classificação (**Figura 1 a 5**); depois variando o total de amostras utilizados para gerar a matriz de confusão (**Figura 6 a 10**); e finalmente, variando o total de pixels e discrepando os valores amostrados, para validação, entre as classes informacionais (**Figura 11 a 15**) para verificar qual seria o comportamento dos três índices dentro dos cinco

níveis de desempenho da classificação. Para cada caso foram geradas 5 matrizes de confusão, de onde foram derivados os índices de exatidão global (G), Kappa (K) e Tau (T), obtidos utilizando as equações (1), (2) e (6), respectivamente. Os resultados são representados nos gráficos, pelas diferenças absolutas entre os índices: ΔGK – diferença entre o índice de Exatidão Global e o Kappa; ΔGT - diferença entre o índice de Exatidão Global e o Tau; e ΔKT - diferença entre o índice de Kappa e o Tau; de onde extraiu-se a média e o desvio-padrão. Chamamos a atenção do leitor para a variação dos valores no eixo Y dos gráficos.

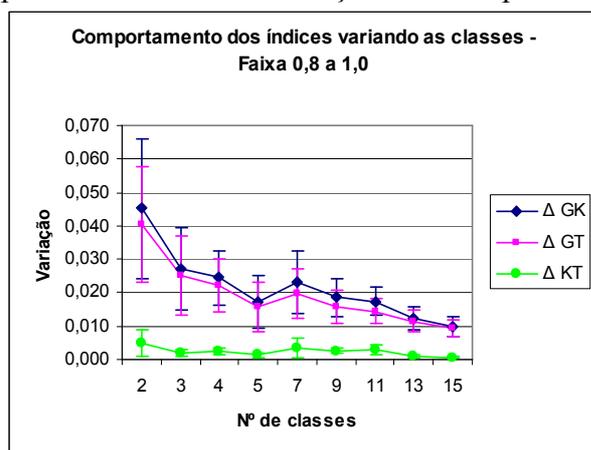


Figura 1 – Variando o número de classes informacionais num nível de desempenho excelente.

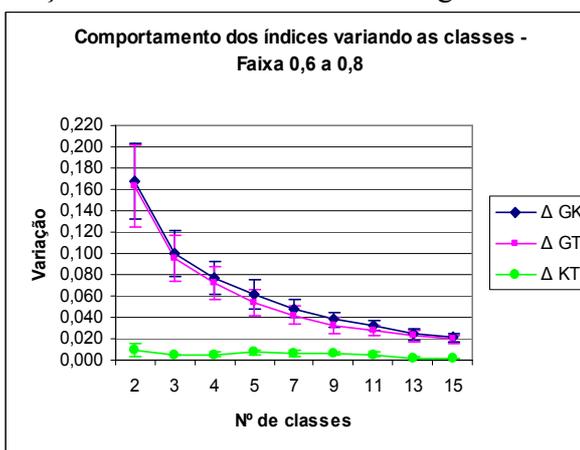


Figura 2 – Variando o número de classes informacionais num nível de desempenho muito bom.

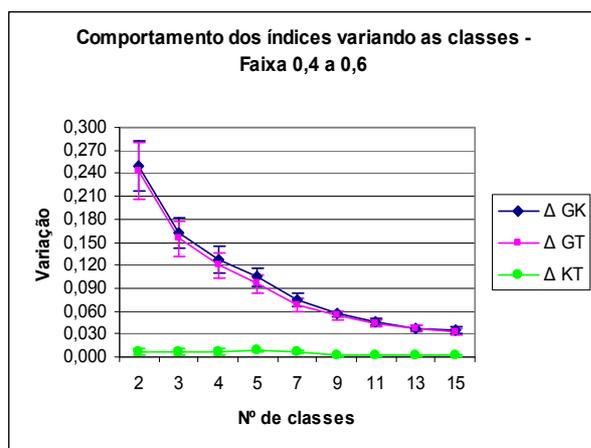


Figura 3 – Variando o número de classes informacionais num nível de desempenho bom.

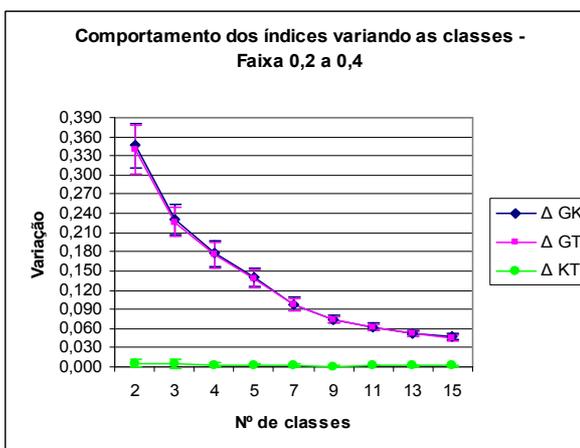


Figura 4 – Variando o número de classes informacionais num nível de desempenho razoável.

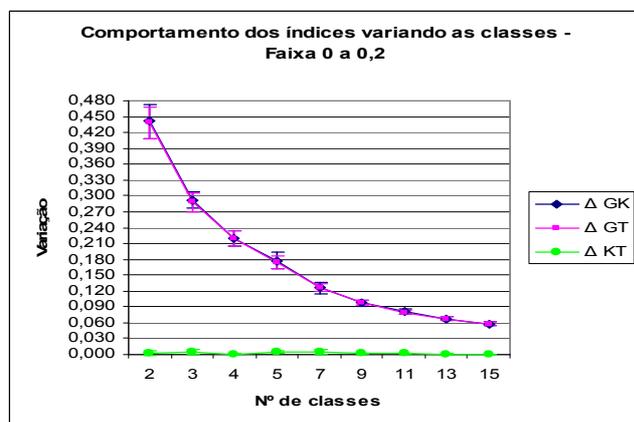


Figura 5 – Variando o número de classes informacionais num nível de desempenho ruim.

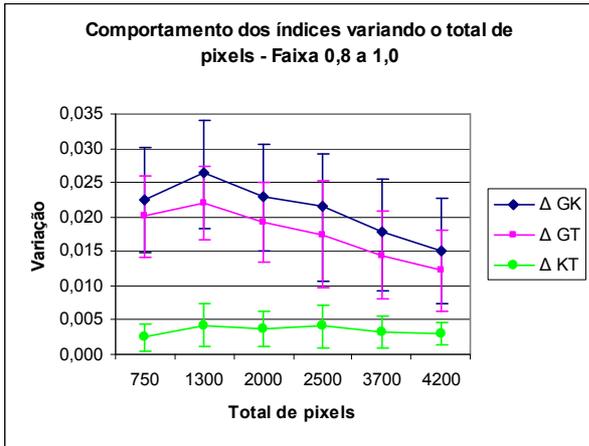


Figura 6 – Variando o total de amostras num nível de desempenho excelente.

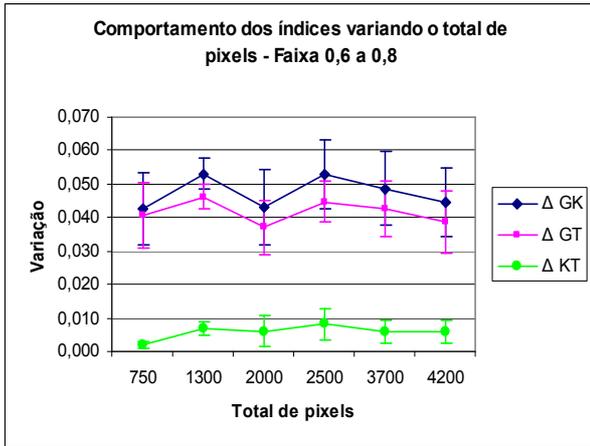


Figura 7 – Variando o total de amostras num nível de desempenho muito bom.

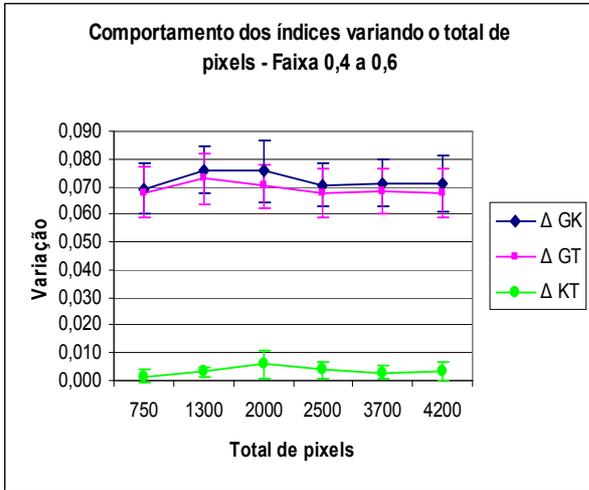


Figura 8 – Variando o total de amostras num nível de desempenho bom.

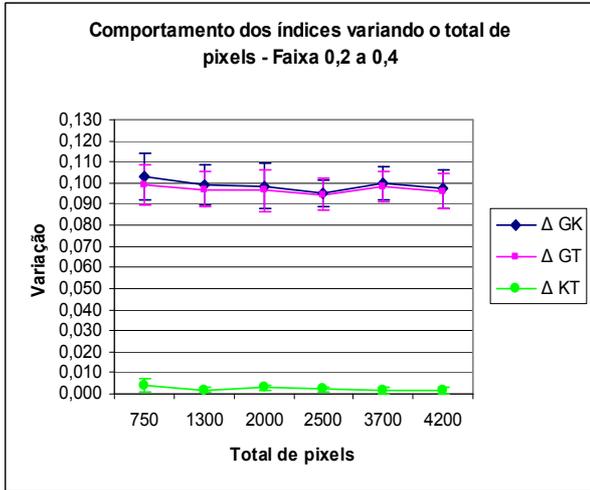


Figura 9 – Variando o total de amostras num nível de desempenho razoável.

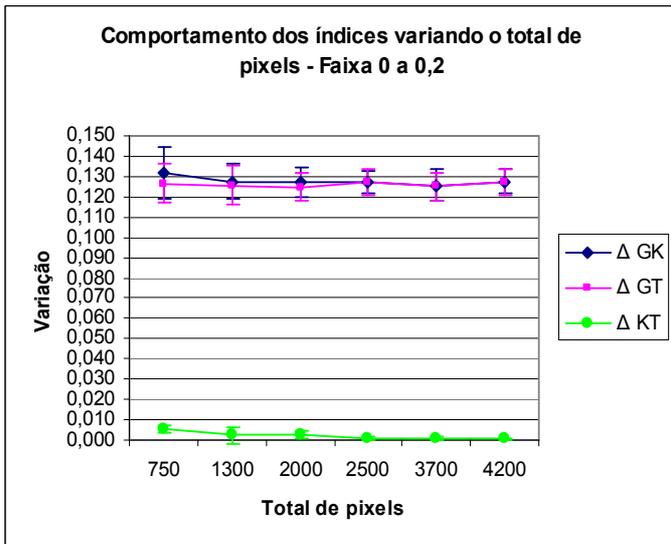


Figura 10 – Variando o total de amostras num nível de desempenho ruim.

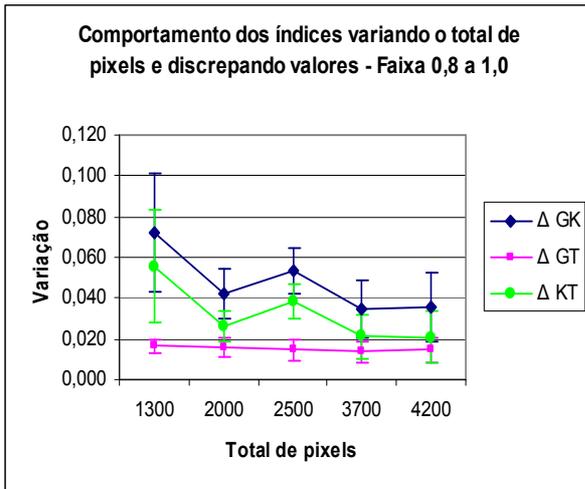


Figura 11 – Variando o total de pixels e discrepando os valores num nível de desempenho excelente.

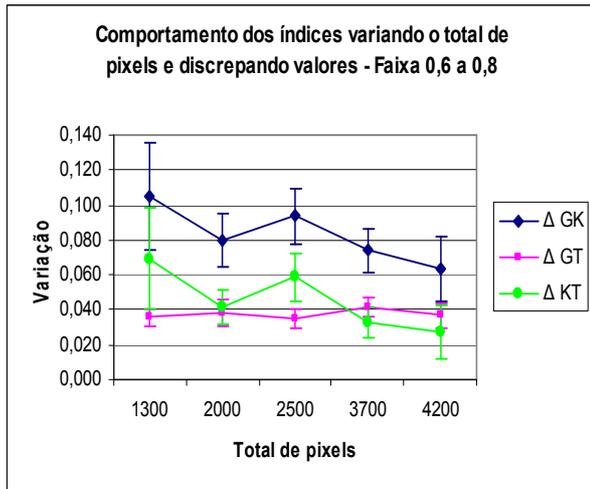


Figura 12 – Variando o total de pixels e discrepando os valores num nível de desempenho muito bom.

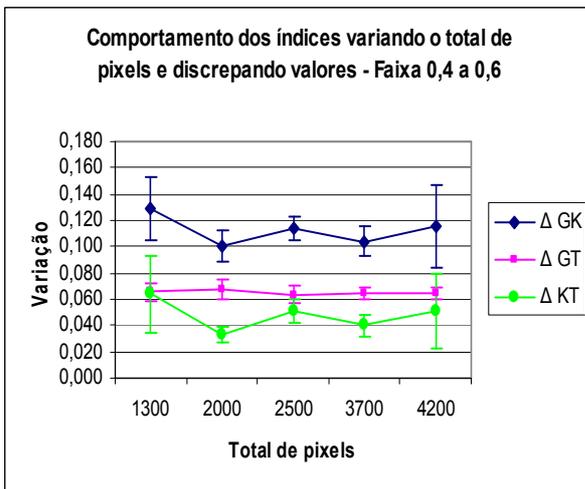


Figura 13 – Variando o total de pixels e discrepando os valores num nível de desempenho bom.

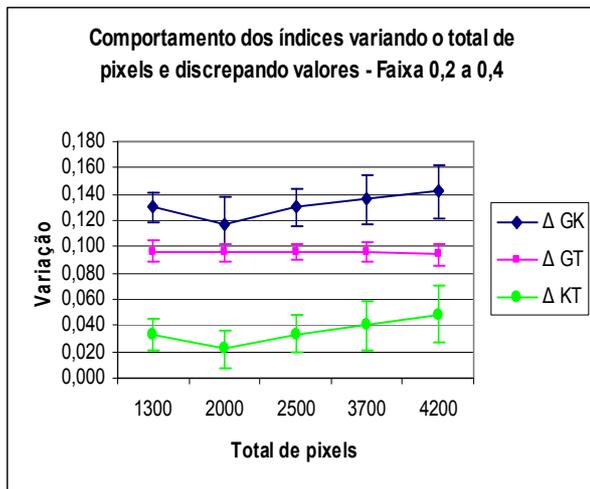


Figura 14 – Variando o total de pixels e discrepando os valores num nível de desempenho razoável.

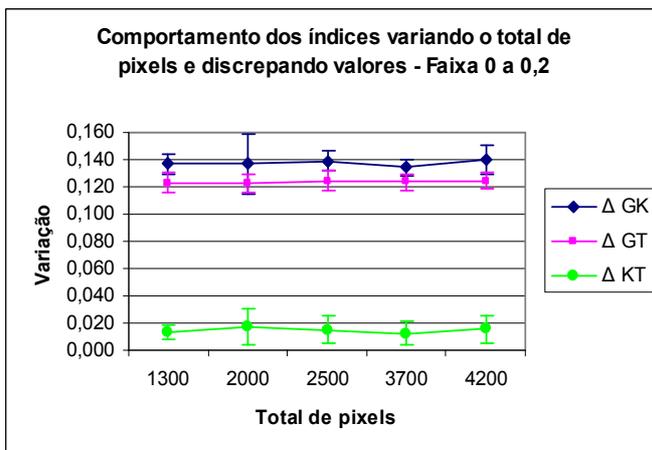


Figura 15 – Variando o total de pixels e discrepando os valores num nível de desempenho ruim.

Analisando as **Figuras 1 a 5** pode-se notar que ΔGK e ΔGT diminuem a medida que aumenta o número de classes, mas o valor da variação aumenta a medida que cai o nível de desempenho da classificação. A variação de ΔKT foi muito pequena em todos os níveis. Das

Figuras 6 a 10 pode-se notar que ΔGK é maior que ΔGT e o valor da variação também aumentou a medida que caiu o nível de desempenho da classificação, sendo que a variação entre ΔKT também foi baixa em todos os níveis. Já nas **Figura 11 a 15** nota-se que ΔGK e ΔGT foram mais discrepantes um em relação ao outro, sendo que ΔGT tende a ser constante; e também o valor da variação aumentou a medida que caiu o nível de desempenho da classificação. A diferença entre ΔKT foi um pouco mais alta quando ocorreu discrepância nos valores amostrados.

5. Considerações finais

Diante desses resultados concluímos que as diferenças entre os três índices não são constantes ao longo dos níveis de desempenho da classificação. No geral, a diferença entre os índices Kappa e Tau foi muito pequena. Embora a Exatidão Global apresente um valor mais alto, os coeficientes de concordância Kappa e Tau são mais consistentes por envolver no valor final todas as células da matriz de confusão.

Foody (1992) mostra que, sem modificações, o coeficiente Kappa superestima a proporção de concordância por chance e subestima a precisão da classificação. A concordância por chance é superestimada pelo fato de incluir no seu cálculo, além da concordância por chance, a concordância real, ou seja, leva em consideração duas vezes o total de acertos. Porém, como os coeficientes de concordância Kappa e Tau apresentam valores aproximados, ao se realizar uma avaliação da acurácia de mapas temáticos ambos podem ser recomendados.

Sugerimos aos pesquisadores interessados, realizar outro teste a fim de verificar se há possibilidades de através de um índice obter o valor de outro. Por exemplo, no nosso estudo poderia se pensar na possibilidade de se estimar a diferença entre os índices global e Tau através da diferença entre os índices global e Kappa, pelo fato de apresentarem comportamento semelhante nos gráficos.

Referências

- Bernardes, T. **Caracterização do ambiente agrícola do Complexo Serra Negra por meio de sensoriamento remoto e sistemas de informação geográfica**. Dissertação (Mestrado). Universidade Federal de Lavras, Minas Gerais, 2006, 119p. Disponível em: <<http://www.epamig.br/geosolos/MaterialSite/Teses/Tiago/Tiagodissertacao.pdf>>. Acesso em: 13 out. 2006.
- Congalton, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 49 n. 12, p. 1671-1678, 1991.
- Congalton, R. G., and Green, K. **Assessing the accuracy of remotely sensed data: Principles and practices**. New York: Lewis Publishers, 1999. 137p.
- Fonseca, L. M. G. **Processamento digital de imagens**. Instituto Nacional de Pesquisas Espaciais (INPE), 2000. 105p.
- Foody, G. M. On the compensation for chance agreement in image classification accuracy assessment. **Photogrammetric Engineering and Remote Sensing**, v. 58, n. 10, p. 1459-1460, 1992.
- Foody, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, v. 80, p. 185– 201, 2002.
- Ma, Z., Redmond, R. L. Tau coefficients for accuracy assessment of classification of remote sensing data. **Photogrammetric Engineering and Remote Sensing**, v. 61, n. 4, p. 435-439, 1995.
- Vieira, C. A. O. **Accuracy of remotely sensing classification of agricultural crops: a comparative study**. 2000. 327p. Thesis (Doctor of Philosophy). University of Nottingham, 2000, p. 128-175.