

# Comparação entre medidas de acurácia de classificação para imagens do satélite ALOS

Eliana Pantaleão  
Graziela Balda Scofield

Instituto Nacional de Pesquisas Espaciais - INPE  
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil  
{graziela, elianap}@dpi.inpe.br

**Abstract.** The accuracy of land cover maps is of high importance to remote sensing users and researchers. Several ways to measure accuracy are available, but there is no agreement about which one is the best. ALOS satellite data are available since October 2006, and some works evaluating the use of ALOS images in land cover recognition were already published. Most of them used *kappa* and *global accuracy* to assess the classification results. This work aims to compare the performance of 16 accuracy measures. The measures are grouped into ten global map measures and six categories measures, and they are compared within each group. The value of each measure is calculated for 28 confusion matrices. In order to perform the comparison, we calculate *PC* (agreement percentage) and *PD* (disagreement percentage) and compute the  $\tau$  coefficient. The confusion matrices used in this work were obtained by the classification of an image from PLR sensor, from ALOS satellite, and are from Floresta Nacional de Tapajós, Pará. The chosen classifiers were SVM (Support Vector Machine) and Polarimetric. The result showed that several measures are equivalent, in the sense that they agree in the choice of the better map, even though the absolute values are different.

**Palavras-chave:** medidas de acurácia, processamento de imagens, classificação de imagens, radar.

## 1. Introdução

A avaliação da acurácia de mapas de cobertura da terra, utilizando dados de satélite, tem sido intensamente estudada pelos pesquisadores em sensoriamento remoto. Várias medidas foram desenvolvidas (Cihlar et al., 2003; Muchoney e Strahler, 2002; e Stehman e Czaplewski, 1998) e testadas para diferentes conjuntos de dados. Os dados do satélite ALOS foram disponibilizados em outubro de 2006. Alguns trabalhos de classificação já foram publicados mostrando o desempenho através do uso do *kappa* e a *acurácia global* na separação de vários usos da terra, como solo exposto, floresta primária e agricultura, por exemplo (Lim et al., 2007; Rahman e Sumantyo, 2008). Mas um estudo mais aprofundado de comparação de várias medidas de acurácia com dados ALOS ainda não foi desenvolvido.

Este trabalho tem como objetivo comparar o desempenho de várias medidas de acurácia divididas por categoria de mapas e classes, utilizando-se a metodologia proposta por Liu et al. (2007). Os dados utilizados são do sensor PLR, produto 1.1, do satélite ALOS, na área da Floresta Nacional de Tapajós, ao sul de Santarém, estado do Pará. As matrizes de confusão utilizadas foram obtidas a partir dos classificadores SVM (Cortes e Vapnik, 1995) e do polarimétrico (Correia, 1998).

## 2. Metodologia de Trabalho

A maioria dos métodos quantitativos utiliza a matriz de confusão derivada dos conjuntos de dados de classificação e referência como ponto inicial da maioria das medidas de acurácia (Story e Colgalton, 1986; e Foody, 2002). A matriz de confusão, apresentada na Tabela 1, apresenta uma visão tabulada da acurácia do mapa de classificação além de permitir o cálculo de medidas específicas tais como a *acurácia global*, *acurácia do usuário* e *acurácia do produtor* (Congalton, 1991). A *acurácia global* é uma das medidas mais simples, sendo calculada pela soma do total dos pixels corretamente classificados dividida pelo número total de pixels da matriz de confusão. As acurácias das categorias individuais são calculadas através da *acurácia do produtor* que indica a probabilidade de um pixel de referência ter sido

corretamente classificado e da *acurácia do usuário* que calcula a probabilidade de um pixel classificado no mapa representar a categoria no solo (Congalton, 1991).

Tabela 1. Matriz de confusão com  $p_{ij}$  representando a proporção da categoria  $i$  da cobertura da terra da área mapeada e da categoria  $j$  da cobertura da terra de referência.

| Classificada | Referência |          |     |          |          |
|--------------|------------|----------|-----|----------|----------|
|              | 1          | 2        | ... | $m$      | Total    |
| 1            | $p_{11}$   | $p_{12}$ | ... | $p_{1m}$ | $p_{1+}$ |
| 2            | $p_{21}$   | $p_{22}$ | ... | $p_{2m}$ | $p_{2+}$ |
| ...          | ...        | ...      | ... | ...      | ...      |
| $m$          | $p_{m1}$   | $p_{m1}$ | ... | $p_{mm}$ | $p_{m+}$ |
| <i>Total</i> | $p_{+1}$   | $p_{+2}$ | ... | $p_{+m}$ |          |

A estatística *kappa*, introduzida por Cohen (1960), é uma medida largamente utilizada da acurácia de classificação em vários estudos (Lee et al., 2001; Lim et al., 2007). O *kappa* é uma proporção de acerto depois da eliminação do acerto por acaso (Rosenfield e Fitzpatrick-Lins, 1986). Vários estudos recomendam o cálculo do *kappa*. Foody (1992) sugeriu que esta medida superestima o acerto por acaso e recomendou duas medidas alternativas que são o *kappa modificado* de Brennan e Prediger (1981) e a estatística *alfa* de Aickin (1990). Ma e Redmond (1995) propuseram o coeficiente  $\tau$ , no qual o *kappa* é um caso especial, e determinaram que sua medida de acurácia de classificação é melhor do que *kappa* e *acurácia global*. A estatística *kappa com peso* (Noesset, 1996) tem sido proposta também como uma medida global de acerto na qual a importância de diferentes omissões aos objetivos do usuário pode ser incorporada na medida da acurácia, mas Stehman (1997) afirmou que não se pode recomendá-la como uma medida de acurácia útil, já que embutindo o peso dentro do contexto da estrutura do *kappa*, resulta em uma medida que sofre da mesma definição e problemas de interpretação inerentes ao *kappa*.

O *kappa condicional*, desenvolvido por Rosenfield e Fitzpatrick-Lins (1986), e *kappa condicional modificado* (Stehman, 1997), têm sido apresentados como um meio adicional de incorporação de acerto por acaso hipotético na avaliação da acurácia por classe.

Medidas de informação, incluindo *informação mútua média* e suas formas normalizadas, foram usadas na avaliação da acurácia da classificação por Finn (1993), Forbes (1995) e Kew (1996).

Não existe concordância em um padrão para a avaliação da acurácia da classificação e a comunidade de sensoriamento remoto não adota universalmente medidas que são frequentemente recomendadas (Foody, 2002). Cada medida de acurácia pode ser mais relevante que outras para um objetivo particular desde que as diferentes medidas incorporem informações diferentes sobre a matriz de erro. O uso de medidas de acurácia diferentes pode resultar em interpretações e conclusões diferentes e possivelmente conflitantes (Stehman, 1997 e Foody, 2002).

Enquanto Rosenfield e Fitzpatrick-Lins (1996) e Stehman (1997) descreveram as principais características e relações entre algumas medidas de acurácia comuns, Liu et al (2007) descreveram a consistência entre as várias medidas existentes quando elas estão separadas em acurácia por categoria e mapa.

Baseando-se na metodologia de Liu et al. (2007), foram escolhidas as medidas de acurácia utilizadas neste estudo, relativas à matriz de confusão. As medidas de acurácia estão separadas por categoria de classe e mapa, como pode ser visto nas tabelas 2 e 3, respectivamente. Cada medida de acurácia por categoria varia de acordo com uma determinada faixa de valores, assim:  $ua$  e  $pa$  variam de zero até 1,  $kcu$  e  $kcp$  entre infinito negativo até zero e  $kcmu$  e  $kcmp$  variam entre  $-1$  a 1. As faixas de valores para as medidas de

categoria por mapa são as seguintes: para *ag*, *amu*, *amp*, *maup*, *iamH*, *iamS* e *acmp* estão compreendidas entre 0 e 1, *kappa* e *kappa<sub>m</sub>* são menores ou iguais a 1 e *imm* está entre 0 e infinito positivo.

Tabela 2. Medidas de acurácia por categoria.

| Nome                                    | Fórmula   | Referências                          |
|---|---|--------------------------------------|
| Acurácia do usuário                     | $au_i = p_{ii}/p_{i+}$                            | Story e Congalton(1986)              |
| Acurácia do produtor                    | $ap_i = p_{ii}/p_{+i}$                            | Storye e Congalton(1986)             |
| Kappa condicional (usuário)             | $kcu_i = (au_i - p_{+i})/(1 - p_{+i})$            | Rosenfield e Fitzpatrick-Lins (1986) |
| Kappa condicional (produtor)            | $kcp_i = (ap_i - p_{+i})/(1 - p_{+i})$            | Rosenfield e Fitzpatrick-Lins (1986) |
| Kappa condicional modificado (usuário)  | $kcmu_i = (au_i - \frac{1}{m})/(1 - \frac{1}{m})$ | Stehman (1997)                       |
| Kappa condicional modificado (produtor) | $kcmp_i = (ap_i - \frac{1}{m})/(1 - \frac{1}{m})$ | Stehman (1997)                       |

Tabela 3. Medidas de acurácia por mapas

| Nome  | Fórmula  | Referências  |
|---|--|--|
| Acurácia global   | $ag = \sum_{i=1}^m p_{ii}$   | Story e Congalton (1986)                             |
| Acurácia média pela perspectiva do usuário                          | $amu = \frac{1}{m} \sum_{i=1}^m au_i$  | Fung e LeDrew(1988)                                  |
| Acurácia média pela perspectiva do produtor                         | $amp = \frac{1}{m} \sum_{i=1}^m ap_i$  | Fung e LeDrew(1988)                                  |
| Média das acurácias médias pelas perspectivas do usuário e produtor | $maup = (amu - amp) / 2$   | Fung e LeDrew(1988)                                  |
| Índice de acurácia média de Hellden                                 | $iamH = \frac{1}{m} \sum_{i=1}^m \frac{2p_{ii}}{p_{i+} + p_{+i}}$              | Hellden (1980), Rosenfield e Fitzpatrick-Lins (1986) |
| Índice de acurácia do mapeamento de Short                           | $iamS = \frac{1}{m} \sum_{i=1}^m \frac{p_{ii}}{p_{i+} + p_{+i} - p_{ii}}$      | Short (1982), Rosenfield e Fitzpatrick-Lins (1986)   |
| Acurácia combinada pelas perspectivas do produtor e usuário         | $acpu = (ag + iamH) / 2$   | Fung e LeDrew(1988)                                  |
| Kappa   | $kappa = (ag - \sum_{i=1}^m p_{i+} p_{+i}) / (1 - \sum_{i=1}^m p_{i+} p_{+i})$ | Cohen (1960), Rosenfield and Fitzpatrick-Lins (1986) |
| Kappa modificado  | $kappa_m = (ag - \frac{1}{m}) / (1 - \frac{1}{m})$                             | Cohen (1960), Rosenfield and Fitzpatrick-Lins (1986) |
| Informação mútua média  | $imm = \sum_{i,j=1}^m p_{ij} \log\left(\frac{p_{ij}}{p_{i+} p_{+i}}\right)$    | Finn (1993)  |

Para analisar a consistência entre as medidas estudadas de acurácia, foi utilizado o  $\tau$ , que é a diferença entre  $PC$  e  $PD$ , em que  $PC$  e  $PD$  são as probabilidades de concordância e discordância, respectivamente. Supõe-se que existam  $n$  matrizes de confusão ( $M_1, M_2, M_3, \dots, M_n$ ) e usam-se duas medidas de acurácia ( $a$  e  $b$ ). Assim, tem-se uma série de pares de medidas de acurácia para as  $n$  matrizes de confusão:  $a_1$  e  $b_1$ ;  $a_2$  e  $b_2$ ; e  $a_n$  e  $b_n$ . Para duas matrizes de confusão  $M_i$  e  $M_j$ , têm-se um par concordante se ambas as medidas  $a_i$  e  $b_i$  forem maiores que  $a_j$  e  $b_j$ . Um par é discordante se  $a_i$  e  $b_j$  forem maiores que  $a_j$  e  $b_i$ . Assim, se existem  $n$  matrizes de confusão, o número de pares é calculado por  $T = n(n-1)/2$ , para  $1 \leq i < j \leq n$ . Sejam  $C$  e  $D$  o número de pares concordantes e discordantes, respectivamente.  $PC$  é a razão entre  $C$  e  $T$  e  $PD$  é a razão entre  $D$  e  $T$ . Portanto,  $\tau$  é igual à diferença entre  $PC$  e  $PD$ , ou seja,  $\tau = (C - D)/T$  e a soma de  $PC$  e  $PD$  é igual à soma de  $C$  e  $D$  sobre o número de pares ( $T$ ).

Para cada matriz de concordância, todas as medidas de acurácia por categoria e mapa foram calculadas (10 para mapas e 6 por categoria). A consistência entre as medidas de mapa foi analisada diretamente, pois só existe um valor de acurácia para cada matriz de confusão e cada medida de acurácia. Portanto, um simples valor de coeficiente foi obtido para todas as matrizes de confusão e para cada par de medidas de acurácia.

A consistência entre as medidas de acurácia por categoria foi analisada da mesma forma, porém as medidas foram comparadas por classe. Para cada dada matriz de confusão, foram calculadas as medidas de acurácia para cada  $m$  categorias, calculado o coeficiente de correlação usando essas  $m$  observações para cada par de acurácia, então calculada a média desses coeficientes de variação sobre todas as matrizes de confusão e obtendo-se um único valor de coeficiente de correlação para cada par de medidas de acurácia.

Depois de calcular os coeficientes de correlação para todos os pares de medidas de acurácia, foi montado um gráfico para visualizar a relação entre as medidas relevantes. Espera-se que as medidas de acurácia com valores maiores de coeficiente de correlação, ou seja, comportando-se similarmente no ranking de acurácia dos mapas ou categorias, estejam próximas uma da outra no gráfico.

### 3. Dados e Área de Estudo

A área de estudo está localizada na Floresta Nacional de Tapajós, no estado do Pará. A Figura 1 mostra a sua localização, assim como das 7 amostras de treinamento e teste utilizadas. As classes estudadas foram floresta primária, secundária, degradada, agricultura de soja tipo 1 e tipo 2, pasto e solo exposto.

A imagem polarimétrica PLR utilizada para a classificação com SVM e classificador polarimétrico foi do dia 23 de abril de 2007. Como o escopo deste trabalho é o desempenho de várias medidas de acurácia, este trabalho fez uso de 28 matrizes de confusão, que foram utilizadas nos trabalhos de Negri et al. (2008).

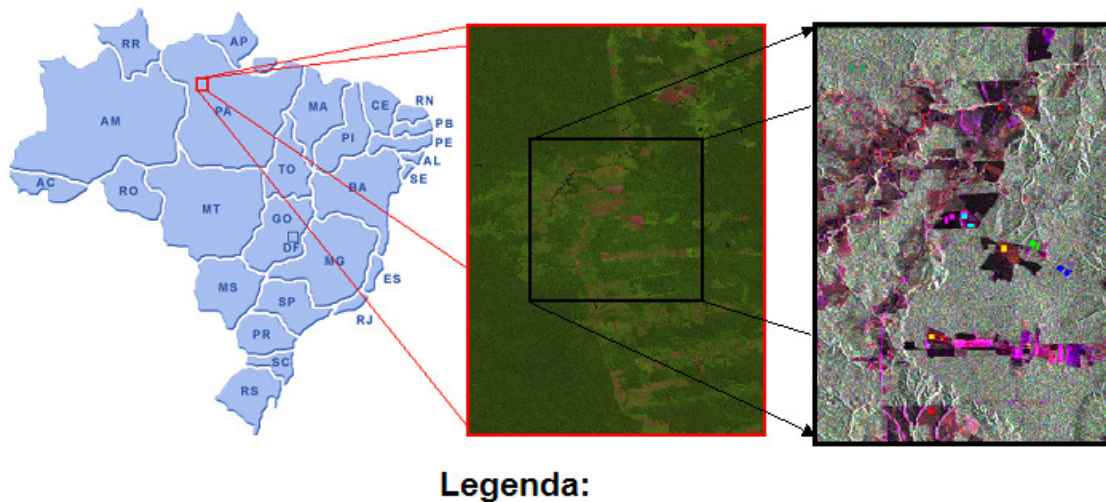


Figura 1. Localização da área de estudo e das classes de treinamento e teste na imagem polarimétrica ALOS do dia 23 de abril de 2007 ((HH)R. (HV)G, (VV)B).

#### 4. Resultados

A Tabela 4 apresenta os resultados do  $\tau$  entre cada par de medidas de acurácia dos mapas. Nota-se que os resultados encontram-se na faixa entre 0,836 e 1, apresentando os menores valores para todas as combinações com a medida *imm*, ou seja, esta é a medida que mais se diferencia das demais. Algumas combinações mostram valores iguais a 1, ou seja, as medidas são totalmente correlacionadas. Este é o caso das medidas *kappa*, *kappa\_m*, *ag* e *amp*.

Tabela 4. Resultado do  $\tau$  entre cada par de medidas de acurácia dos mapas.

|                | <i>amu</i> | <i>amp</i> | <i>maup</i> | <i>iamH</i> | <i>iamS</i> | <i>acpu</i> | <i>kappa</i> | <i>kappa_m</i> | <i>imm</i> |
|----------------|------------|------------|-------------|-------------|-------------|-------------|--------------|----------------|------------|
| <i>ag</i>      | 0,931      | 1,000      | 0,952       | 0,931       | 0,942       | 0,958       | 1,000        | 1,000          | 0,894      |
| <i>amu</i>     |            | 0,931      | 0,979       | 0,905       | 0,926       | 0,921       | 0,931        | 0,931          | 0,857      |
| <i>amp</i>     |            |            | 0,952       | 0,931       | 0,942       | 0,958       | 1,000        | 1,000          | 0,894      |
| <i>maup</i>    |            |            |             | 0,926       | 0,947       | 0,942       | 0,952        | 0,952          | 0,878      |
| <i>iamH</i>    |            |            |             |             | 0,968       | 0,974       | 0,931        | 0,931          | 0,836      |
| <i>iamS</i>    |            |            |             |             |             | 0,974       | 0,942        | 0,942          | 0,857      |
| <i>acpu</i>    |            |            |             |             |             |             | 0,958        | 0,958          | 0,862      |
| <i>kappa</i>   |            |            |             |             |             |             |              | 1,000          | 0,894      |
| <i>kappa_m</i> |            |            |             |             |             |             |              |                | 0,894      |

A tabela 5 mostra que as medidas *au*, *kcu* e *kcmu*; e também *ap*, *kcp* e *kcmp* também apresentaram altos valores de  $\tau$ , ou seja, concordam na grande maioria dos casos. Portanto, essas medidas poderiam ser utilizadas de forma equivalente.

Tabela 5. Resultado do  $\tau$  entre cada par de medidas de acurácia por classes.

|             | <i>ap</i> | <i>kcu</i> | <i>kcp</i> | <i>kcmu</i> | <i>kcmp</i> |
|-------------|-----------|------------|------------|-------------|-------------|
| <i>au</i>   | 0,459     | 0,990      | 0,499      | 0,999       | 0,459       |
| <i>ap</i>   |           | 0,449      | 0,961      | 0,459       | 0,999       |
| <i>kcu</i>  |           |            | 0,489      | 0,989       | 0,449       |
| <i>kcp</i>  |           |            |            | 0,498       | 0,960       |
| <i>kcmu</i> |           |            |            |             | 0,459       |

As Figuras 2 e 3 mostram os gráficos comparativos dos valores das medidas para cada uma das matrizes de confusão utilizadas. Pode-se notar que, apesar de existirem diferenças nos valores absolutos das medidas, em geral elas se comportam de forma bastante semelhante, com uma pequena diferença na medida *imm*.

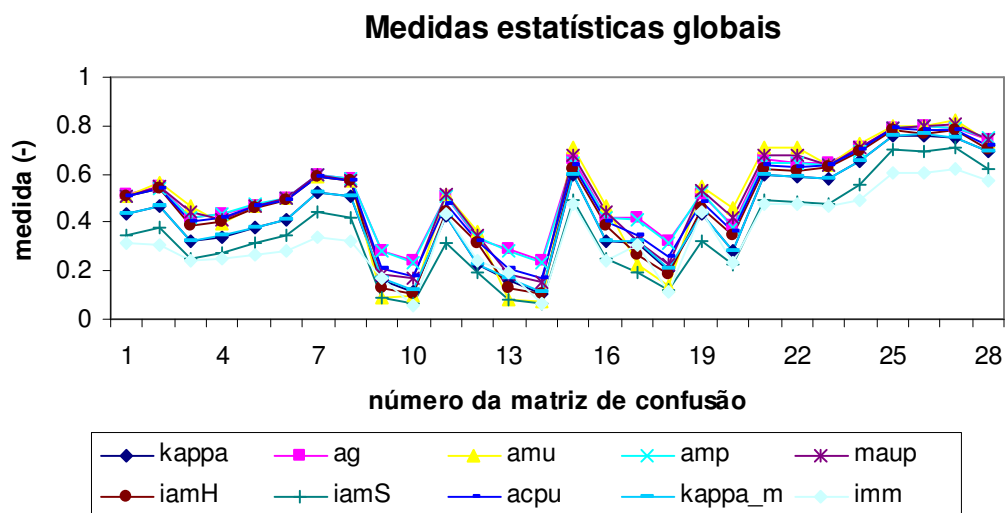


Figura 2. Comparação das medidas estatísticas *ag*, *amu*, *amp*, *maup*, *iamH*, *iamS*, *acpu*, *kappa*, *kappa<sub>m</sub>* e *imm* para as 28 matrizes de confusão utilizadas no estudo.

Na Figura 3, que mostra os valores das medidas apenas para a classe Floresta Secundária, observa-se o comportamento dos dois grupos citados, ou seja, *au*, *kcu* e *kcmu*; e *ap*, *kcp* e *kcmp*. As medidas de cada grupo comportam-se de forma equivalente, com diferenças entre os grupos.

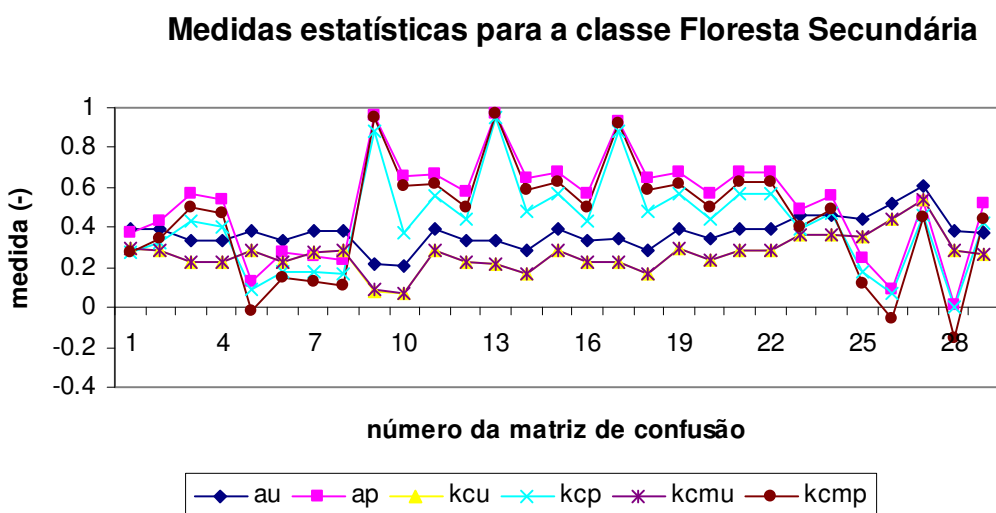


Figura 3. Comparação das medidas estatísticas *au*, *ap*, *kcu*, *kcp*, *kcmu* e *kcmp* para a classe Floresta Secundária.

Em geral, as medidas globais de mapa tiveram comportamento bastante semelhante, enquanto que as medidas por categorias apresentaram maiores diferenças.

## Agradecimentos

Os autores agradecem o suporte financeiro da CAPES e do CNPq (3284469/2006-0).

## Referências Bibliográficas

- Aickin, M. **Maximum likelihood estimation of agreement in the constant predictive probability model, and its relation to Cohen's kappa**. *Biometrics*, v. 46, 293–302, 1990.
- Brennan, R. L., & Prediger, D. J. **Coefficient kappa: Some uses, misuses, and alternatives**. *Educational and Psychological Measurement*, v.41, 687–699, 1981.
- Cihlar, J., Latifovic, R., Chen, J., Beaubien, J. and Li, Z., **Selecting representative high resolution sample images for land cover studies: Part 1. Methodology**. *Remote Sensing of Environment*, v.71, pp. 26–42. 2000.
- Cohen, J. **A coefficient of agreement of nominal scales**. *Educational and Psychological Measurement*, v.20, p. 37–46. 1960.
- Congalton, R. G. **A review of assessing the accuracy of classifications of remotely sensed data**. *Remote Sensing of Environment*, 37, 35–46. 1991.
- Cortes, C.; Vapnik, V. **Support-Vector Networks**, *Machine Learning*, 20(3):273-297, September 1995.
- Correia, A . H. **Desenvolvimento de classificadores de máxima verossimilhança e ICM para imagens SAR polarimétricas**. São José dos Campos .275 p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Especiais, São José dos Campos, 1999.
- Finn, J. T. **Use of the average mutual information index in evaluating classification error and consistency**. *International Journal of Geographical Information Systems*, v.7, p.349–366. 1993.
- Foody, G. M. **On the compensation for chance agreement in image classification accuracy assessment**. *Photogrammetric Engineering and Remote Sensing*, 58, 1459–1460, 1992.
- Foody, G. M. **Status of land cover classification accuracy assessment**. *Remote Sensing of Environment*, 80, p. 185–201. 2002.
- Forbes, A. D. **Classification algorithm evaluation: Five performance measures based on confusion matrices**. *Journal of Clinical Monitoring*, 11, 189–206, 1995.
- Fung, T.; LeDrew, E. **The determination of optimal threshold levels for change detection using various accuracy indices**. *Photogrammetric Engineering and Remote Sensing*, 54, p.1449–1454. 1988.
- Hellden, U. **A test of Landsat-2 imagery and digital data for thematic mapping illustrated by an environmental study in northern Kenya**. Sweden, Lund University Natural Geography Institute Report, v. 47. 1980.
- Kew, N. R. (1996). Information-theoretic measures for assessment and analysis in image classification. In E. Binaghi, P. A. Brivio, & A. Rampini (Eds.), **Soft computing in remote sensing data analysis** (pp. 173–180). Singapore: World Scientific.
- Lee, J-S; Grunes, M.R. and Pottier, E. **Quantitative comparison of classification capability: fully polarimetric versus dual and single-polarization SAR**. *IEEE Transactions on Geoscience and Remote Sensing*, v.39, n.11, 2343-2351, 2001.
- Lim, H. S.; MatJafri, M. Z.; Abdullah, K. and Saleh, N. M. **High Spatial Resolution Land Cover Mapping Using ALOS Data Over Kedah, Malaysia**. First Joint PI Symposium of ALOS DataNodes for ALOS Science Program, Kyoto, 2007.
- Liu, C ; Frazier, P., Kumar, L. **Comparative assessment of the measures of thematic classification accuracy**. *Remote Sensing of Environment* v. 107 p.606–616. 2007.

- Ma, Z., & Redmond, R. L. (1995). **Tau coefficients for accuracy assessment of classification of remote sensing data**. *Photogrammetric Engineering and Remote Sensing*, 61, 435–439.
- Muchoney, D.M. and Strahler, A.H., 2002. **Pixel- and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data**. *Remote Sensing of Environment* 81, pp. 290–299.
- Negri, R.G **Classificação da cobertura da terra na Amazônia utilizando Imagens Polarimétricas em Banda L e Máquina de Vetores Suporte**. Submetido em: Simpósio Brasileiro de Sensoriamento Remoto. 14. Natal, Rio Grande do Norte, Brasil. 2008.
- Noeset, E. (1996). **Use of the weighted Kappa coefficient in classification error assessment of thematic maps**. *International Journal of Geographical Information Science*, 10, 591–604.
- Rahman, M. M. e Sumantyo, J. T. S. **ALOS PALSAR data for tropical forest interpretation and mapping**. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7. Beijing 2008.
- Rosenfield, G. H., & Fitzpatrick-Lins, K. **A coefficient of agreement as a measure of thematic classification accuracy**. *Photogrammetric Engineering and Remote Sensing*, v.52, p.223–227. 1986.
- Stehman, S. V. **Selecting and interpreting measures of thematic classification accuracy**. *Remote Sensing of Environment*, 62, p. 77–89. 1997.
- Stehman, V.S. and Czaplewski, R.L., 1998. **Design and analysis for thematic map accuracy assessment: fundamental principles**. *Remote Sensing of Environment* 64, pp. 331–344.
- Story, M.; Congalton, R. G. **Accuracy assessment: A user's perspective**. *Photogrammetric Engineering and Remote Sensing*, v.52, p.397–399. 1986.