

## Rede Bayesiana aplicada ao monitoramento da Moratória da Soja

Marcio Pupin Mello<sup>1</sup>  
Marcos Adami<sup>1</sup>  
Bernardo Friedrich Theodor Rudorff<sup>1</sup>  
Daniel Alves Aguiar<sup>1</sup>

<sup>1</sup>Instituto Nacional de Pesquisas Espaciais – INPE  
Divisão de Sensoriamento Remoto – DSR  
Av. dos Astronautas, 1758 - Jd. Granja - São José dos Campos - SP, 12227-010, Brasil  
{mello, adami, bernardo, daniel}@dsr.inpe.br

**Abstract.** The Soy Moratorium is intended to reduce deforestation in the Brazilian Amazon. Its monitoring is supported by remote sensing satellite images to identify soybean crop in deforested areas after 24 July 2006. The image classification procedure currently used to identify soybean crop is based on conventional methods that can be improved by modeling a Bayesian Network (BN) using artificial intelligence techniques. Input variables were: MODIS/EVI values of minimum, maximum and amplitude during soybean crop season; SRTM elevation data; year of deforestation; distance from rivers; and distance from roads. Reference data for training and accuracy assessment were obtained through oblique aerial photographs and field work. The probability functions in the BN were modeled based on the number of pixels within discrete intervals. The result was encouraging with overall classification accuracy greater than 80%. Moreover, the probability image is an important decision tool to reduce the number of deforested fields to be surveyed and visited during field work. The R algorithm used to define the BN model is available at <http://www.dsr.inpe.br/~mello>.

**Palavras-chave:** remote sensing, soy moratorium, artificial intelligence, deforestation, Brazilian Amazon, sensoriamento remoto, moratória da soja, inteligência artificial, desflorestamento, Amazônia brasileira.

### 1. Introdução

O impacto negativo sobre o meio ambiente provocado pela conversão do uso e cobertura da terra para atender as crescentes necessidades de consumo humano tem ocasionado modificações nos ciclos biogeoquímicos do sistema terrestre. Minimizar as consequências dessas alterações para a manutenção da vida no planeta atribuiu um novo enfoque aos estudos das mudanças de uso e cobertura da terra (Foley et al., 2005). Isto faz com que a conversão de florestas para o uso agropecuário torne-se um dos principais temas no debate atual das mudanças globais (Galford et al., 2010). Nesse contexto, a Amazônia tem recebido crescente atenção devido à elevada emissão de carbono e à perda de biodiversidade como consequência do desflorestamento (Cardille e Foley, 2003; Alves, 2008). O mapeamento e o monitoramento do desflorestamento na Amazônia Legal brasileira têm sido realizados desde 1988 pelo Instituto Nacional de Pesquisas Espaciais (INPE, 2010). Esse mapeamento evidencia a dinâmica do desflorestamento associada ao crescimento econômico nacional e à demanda global por alimentos (Fearnside e Laurance, 2003; Alencar et al., 2004).

As elevadas taxas de desflorestamento da Amazônia legal e a sua relação com a sojicultura (Brown et al., 2005) foram evidenciadas na campanha “Eating up the Amazon” (Greenpeace, 2006). Nesta campanha, organizações da sociedade civil realizaram um movimento em prol do fim do desflorestamento no bioma Amazônia do Brasil ganhando o apoio das empresas ligadas à *commodity* da soja. Assim, em julho de 2006, a Associação Brasileira das Indústrias de Óleos Vegetais (ABIOVE) e a Associação Nacional dos Exportadores de Cereais (ANEC), em conjunto com organizações da sociedade civil, anunciaram a Moratória da Soja sobre áreas desflorestadas com duração inicial de dois anos. Este acordo preconizou que as empresas associadas à ABIOVE e à ANEC não comprariam soja produzida em áreas desflorestadas a partir de 24 de julho de 2006 no bioma Amazônia.

O acordo da Moratória da Soja já foi renovado por três vezes e desde julho de 2009 vem utilizando imagens de satélites de sensoriamento remoto para auxiliar no monitoramento da

área desflorestada para identificação de possíveis cultivos de soja (Rudorff et al., *submetido*). Contudo, o atual processo utilizada abordagem booleana e poderia ser incrementado valendo-se de técnicas de inteligência artificial.

Mello et al. (2010) apresentam uma alternativa que pode aprimorar o método de identificação de cultivo de soja, atualmente utilizado por Rudorff et al. (*submetido*), utilizando inteligência artificial baseada em Redes Bayesianas (RB). RBs são grafos acíclicos direcionados, ou seja, elas representam relações causais entre variáveis aleatórias em modelos probabilísticos. A modelagem matemática do teorema de Bayes propõe o cálculo da probabilidade baseado tanto no conhecimento *a priori* quanto na condicionalidade. Em outras palavras, ele conecta a inferência racional (probabilidade *a posteriori*) à subjetividade (probabilidade *a priori*) e a experiência empírica (probabilidades condicionais) (Pena, 2006). Assim, o teorema de Bayes liga a razão humana ao universo físico. Nesse contexto, o objetivo do presente trabalho foi modelar uma Rede Bayesiana para identificar áreas de soja no contexto da Moratória da Soja.

## 2. Materiais e Métodos

A área de estudo compreende os municípios de Cláudia, Feliz Natal, Gaúcha do Norte, Nova Ubiratã, Paranatinga, Santa Carmen, União do Sul e Vera (Figura 1). Esses oito municípios estão localizados no Estado do Mato Grosso à nordeste da capital Cuiabá e possuem, conjuntamente, uma área de aproximadamente 5,63 milhões de ha. Essa região está completamente compreendida no *tile* h12v10 do *grid* de produtos do sensor MODIS (*MODerate resolution Imaging Spectroradiometer*).

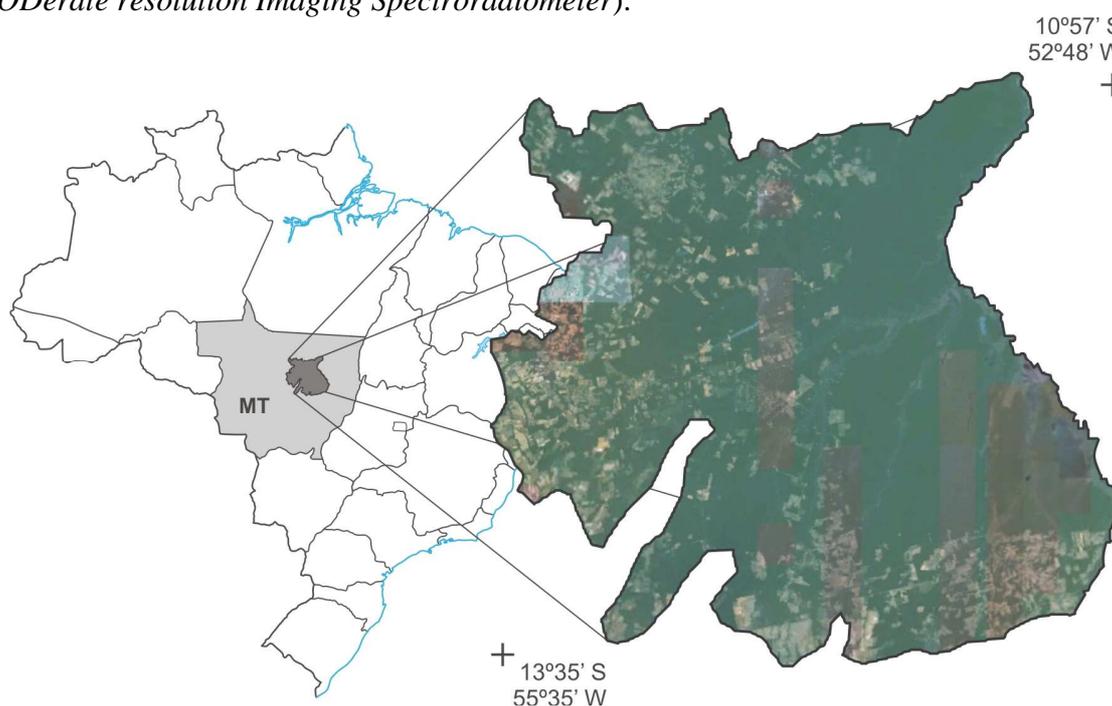


Figura 1. Localização da área de estudo (mosaico de imagens, FONTE: Google Earth).

O calendário agrícola da soja tem início com o plantio em outubro-novembro e termina com a colheita em janeiro-fevereiro, sendo que o máximo desenvolvimento vegetativo da cultura ocorre em dezembro. Para adequar-se a esse calendário, uma série temporal de imagens EVI (*Enhanced Vegetation Index*)/MODIS (produto MOD13 – composições de 8 dias) foi processada e filtrada, utilizando *wavelets* (Daubechies, 1993; Freitas e Shimabukuro, 2008) para minimização dos efeitos de ruídos. A série temporal das imagens EVI/MODIS corresponde ao período do cultivo da soja representando a dinâmica do manejo agrícola expressa na variação temporal do EVI, como apontado por Rizzi et al. (2006).

No modelo da RB (Figura 2), foram utilizadas sete variáveis de entrada, sendo:

( $X_1$ ) *EVImín* – imagem contendo o menor valor da série temporal de EVI filtrada;

( $X_2$ ) *EVImax* – imagem contendo o maior valor da série temporal de EVI filtrada;

( $X_3$ ) *EVlamp* – imagem resultante da diferença entre *EVImax* e *EVImín*;

( $X_4$ ) *hidro* – distância até o curso d'água mais próximo;

( $X_5$ ) *estradas* – distância até a estrada mais próxima<sup>1</sup>;

( $X_6$ ) *srtm* – modelo digital de elevação derivado dos dados SRTM (*Shuttle Radar Topography Mission*) (Rabus et al., 2003), disponíveis gratuitamente no site da EMBRAPA Monitoramento por Satélite;

( $X_7$ ) *ano* – ano em que o desflorestamento foi identificado pelo PRODES, dado este disponibilizado gratuitamente pelo INPE.

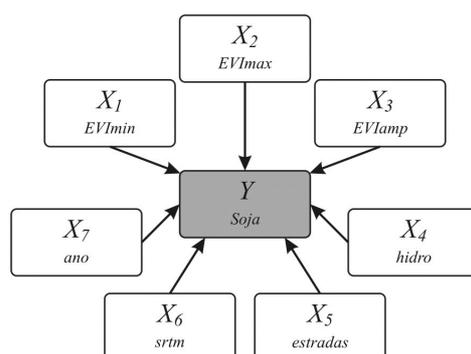


Figura 2. Representação gráfica da Rede Bayesiana utilizada.

A variável denotada por  $Y$  representa a ocorrência do plantio de soja, assumindo valor 1 (um) no caso de *presença* e 0 (zero) no caso de *ausência*. O modelo formal da RB, como proposto por Mello et al. (2010), é dado por

$$P(Y | X_1, X_2, \dots, X_7) = \frac{P(X_1, X_2, \dots, X_7 | Y) \cdot P(Y)}{P(X_1, X_2, \dots, X_7)} \quad (1)$$

O mapa de referência com as classes temáticas *soja* e *outros* (não soja), foi utilizado nas fases de treinamento e validação. Este mapa foi construído com base em fotos aéreas panorâmicas e dados de campo (Figura 3). É importante ressaltar que os polígonos mapeados como desflorestamento pelo PRODES, do ano 2007 em diante (posterior à Moratória da Soja), foram utilizados como máscara no procedimento. Todas as imagens foram trabalhadas na resolução espacial de 250 m, adequando-se à resolução do sensor MODIS.



Figura 3. Exemplo de fotografias aéreas panorâmicas utilizadas para identificar o uso e cobertura da terra. As fotografias panorâmicas podem ser visualizadas, junto com outras imagens de sensores orbitais, em [http://www.abiove.com.br/english/ss\\_relatoriouso09\\_us.asp](http://www.abiove.com.br/english/ss_relatoriouso09_us.asp).

<sup>1</sup>Tanto os vetores de estradas quanto os vetores de cursos d'água foram disponibilizados pela SEMA – Secretaria de Estado do Meio Ambiente (Mato-Grosso: <http://www.sema.mt.gov.br>).

As funções de probabilidade utilizadas pela RB foram discretizadas em intervalos e seus valores definidos por contagem de *pixel*, como descrito em Mello et al. (2010). Para aplicação da RB utilizou-se o algoritmo implementado, no software R (R Development Core Team, 2010)<sup>2</sup>, por Mello et al. (2011), disponível gratuitamente em <http://www.dsr.inpe.br/~mello>.

Os limites dos intervalos utilizados na discretização das funções de probabilidades relacionadas a cada variável de entrada na RB foram definidos empiricamente, e de maneira interativa com o algoritmo. Após definidos os intervalos, a RB foi treinada e posteriormente utilizada para gerar uma imagem de probabilidade (PI: *Probability Image*). Nessa PI, cada *pixel* possui um valor entre 0 e 1 (0 e 100%) representando a probabilidade da presença de soja ( $Y=1$ ), dados os valores observados conjuntamente para as variáveis de entrada da rede.

Essa abordagem possui algumas vantagens sobre as abordagens convencionalmente empregadas (*booleanas*). Entre elas: i) a possibilidade de estabelecer grau de confiança, baseado em interação de probabilidades segundo Bayes; ii) a possibilidade de considerar áreas com o plantio presente, mesmo quando uma ou outra variável de entrada não é suficiente para evidenciar a presença da cultura; e iii) a possibilidade de gerar um mapa temático utilizando apenas uma operação de fatiamento. Nesse caso, existe uma gama de possibilidades implementadas para escolher o valor do fatiamento, com critérios baseados em índices de exatidão global, *kappa*, omissão e inclusão, sensibilidade e especificidade, além do critério de área.

### 3. Resultados e Discussão

As relações probabilísticas entre as variáveis foram consideradas na definição dos intervalos utilizados para discretizar as funções de probabilidade. Estas relações podem ser evidenciadas por meio de histogramas que apresentam as variáveis de entrada em conjunto com a presença de soja ( $Y=1$ ) como mostra a Figura 4.

Inspecionado a Figura 4 é possível fazer algumas inferências a respeito da interação entre as variáveis de entrada e a variável alvo (soja), como por exemplo: i) no caso do *EVI<sub>min</sub>*, nota-se que é bem mais provável encontrar valores de EVI acima de 0,35 na imagem. Entretanto, a probabilidade associada à presença de soja onde o *EVI<sub>min</sub>* é maior ou igual a 0,35 é de apenas 2,3%, contra 16,2% onde o *EVI<sub>min</sub>* é menor que 0,35; ii) já para a variável *EVI<sub>max</sub>*, em geral, a presença de soja é mais provável quando ocorrem valores mais altos de *EVI<sub>max</sub>*; iii) a variável *EVI<sub>amp</sub>*, por sua vez, apresenta valores mais prováveis da presença de soja associados a grandes amplitudes, evidenciando que, quanto maior o “degrau” no valor do EVI entre os períodos de máximo e mínimo maior será a correlação com a presença de soja (Rizzi et al., 2006); iv) no caso da variável hidrografia (*hidro*), existe uma tendência de que o plantio da cultura se dê a uma distância mínima dos cursos d’água<sup>3</sup>; v) para a variável *estradas* a função de probabilidade mostrou que a maior probabilidade de ocorrência de soja situa-se a uma distância, estabelecida empiricamente, de 3,25 km; vi) quanto à variável *srtm*, as áreas de menor altitude, associadas às planícies, apresentam maior probabilidade da presença de soja; e por fim vii) a variável *ano* mostra que desflorestamentos mais antigos possuem maior probabilidade da presença de soja que desflorestamentos recentes. Foi constatado que as áreas com desflorestamento identificados em 2007 pelo PRODES tem probabilidade associada à presença de soja de 8,5%, contra 2,9% para os desflorestamentos identificados em 2008 e apenas 0,45% para os polígonos de 2009.

<sup>2</sup>A biblioteca *rgdal* (Keitt et al., 2010), que não é parte do pacote básico do R, foi utilizada para manipulação (entrada e saída) das imagens em formato GeoTiff no R.

<sup>3</sup>Em geral essa distância é preconizada pelas Áreas de Preservação Permanente (APP). É importante lembrar que as imagens possuem resolução espacial de 250m e, portanto, APPs com distâncias menores que essa apresentaram-se com mistura espectral, sendo impossível defini-las com melhor acurácia.

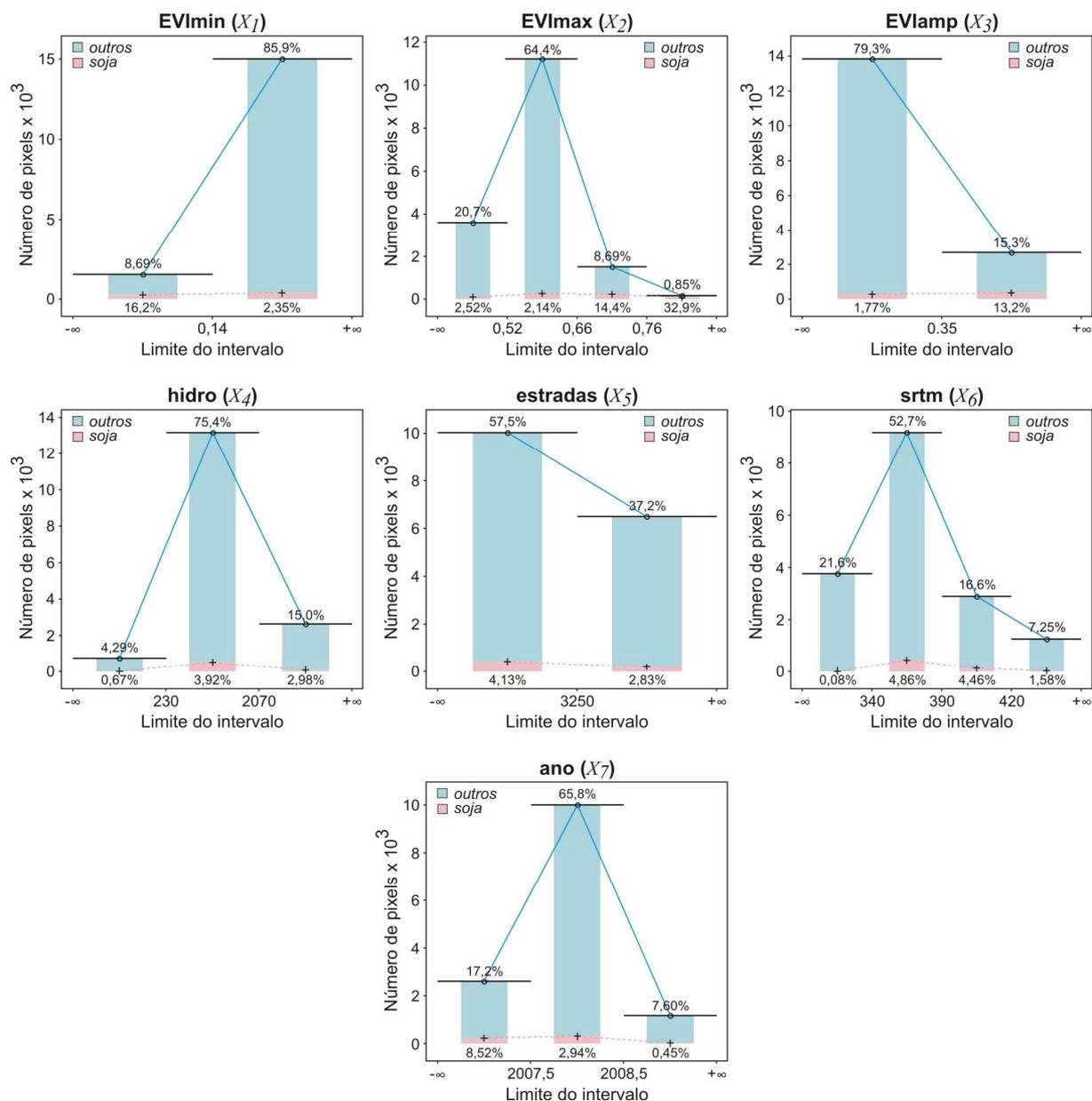


Figura 4. Histograma com intervalos definidos para as variáveis de entrada e suas relações com a presença de soja. A parte rosa das barras representa o número de *pixels* com soja e a parte azul representa o número de *pixels* sem soja. Os valores de porcentagens acima das barras representam a probabilidade do valor do *pixel* estar contido no intervalo considerado<sup>4</sup>; e os valores de porcentagens abaixo das barras representam a probabilidade da presença de soja, dado os limites desse intervalo para a variável de entrada.

Após a definição dos intervalos apresentados na Figura 4 ( $EVlmin=2$ ;  $EVlmax=4$ ;  $EVlamp=2$ ;  $hidro=3$ ;  $estradas=2$ ;  $srtm=4$ ; e  $ano=3$ ) a RB foi treinada com base nos 1152 cenários possíveis, estabelecidos pelas possíveis combinações destes intervalos ( $2*4*2*3*2*4*3=1152$ ). Após a fase de treinamento, o algoritmo calcula a PI, que após a aplicação da máscara dos polígonos do PRODES apresenta-se como na Figura 5.

<sup>4</sup>Pela diversidade e natureza dos dados utilizados, o registro entre as imagens não é perfeito e faz com que, em pontos onde pelo menos uma das variáveis presente dado ausente, o *pixel* não seja contabilizado. Assim, a soma dos valores pode não totalizar 100%.

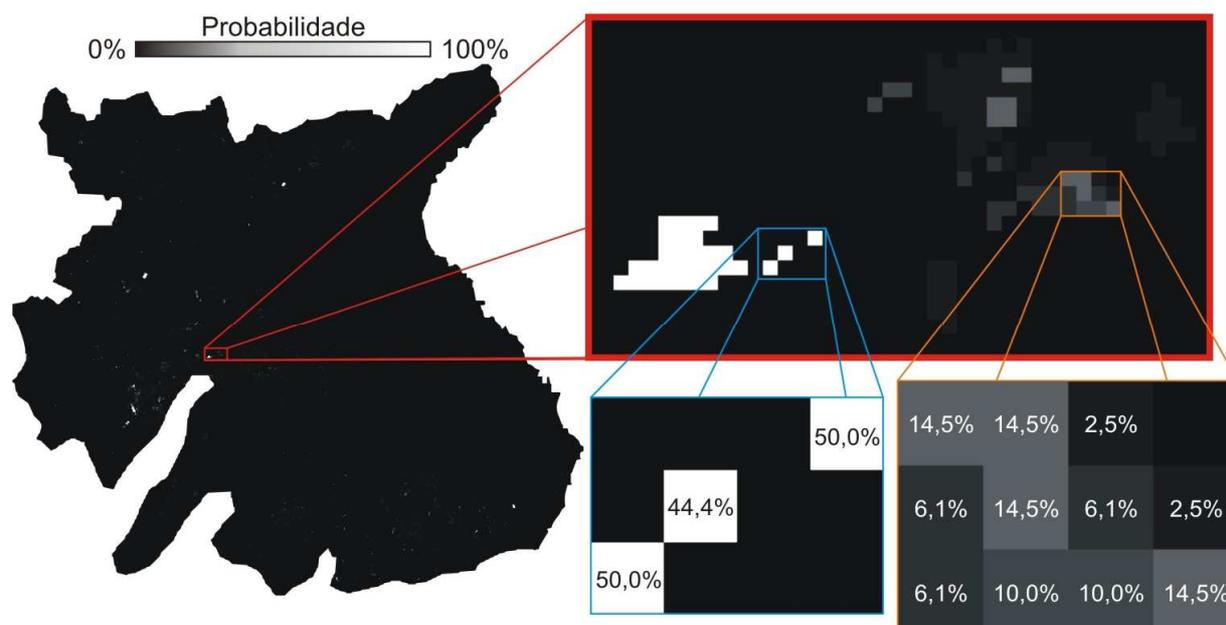


Figura 5. Imagem de Probabilidade (PI) gerada pela Rede Bayesiana.

Cada *pixel* da PI apresenta o valor da probabilidade da presença de soja, dados os valores observados nas variáveis de entrada. Pode-se utilizar a PI, com um simples fatiamento, por exemplo, para gerar um mapa temático da soja.

O valor limite para o fatiamento pode ser inserido manualmente, com o auxílio de gráficos e tabelas apresentados na Figura 7a, ou definidos por meio de critérios baseados em índices de precisão/exatidão. É importante comentar que, dado a grande diferença na proporção das coberturas das classes *soja* e *outros*, foi imprescindível utilizar procedimentos de amostragem. Assim, o algoritmo utilizado nesse trabalho possibilita sortear, com base em uma distribuição estatística multinomial, um tamanho mínimo de amostra por classe, como sugerido por Congalton e Green (2009).

Verificou-se que mesmo com a amostragem não foi possível estabelecer um “bom” valor para o fatiamento devido à grande diferença de proporção de cobertura entre as classes *soja* (3,4%) e *outros* (96,6%) causada pela baixa conversão de áreas desflorestadas, após a moratória, em soja. Mello et al. (2010) também observaram que os valores de probabilidade na PI tendem a apresentar-se abaixo do esperado. Dessa forma, o termo  $P(Y=1)$  no numerador da Equação 1 é muito pequeno (0,003428899), o que faz com que os valores de probabilidades condicionais calculados sejam baixos.

Entretanto, apenas a título de ilustração, foi utilizado o critério de ponto mais próximo do valor 1,1 (100% de sensibilidade e especificidade) na curva ROC para a seleção do melhor valor de probabilidade para o fatiamento (*best Target Probability Value*, ou *best-TPV*). Segundo Zweig e Campbell (1993), esse ponto representa o ponto de melhor acurácia. O valor calculado pelo algoritmo foi de 6%, ou seja, segundo o critério escolhido, o melhor mapa temático para a presença de soja seria conseguido fatiando-se a PI de forma que todo *pixel*, onde a probabilidade da presença de soja fosse maior ou igual a 6%, fosse rotulado como *soja*. Os gráficos exibidos pelo algoritmo para a escolha do ponto são mostrados na Figura 7.

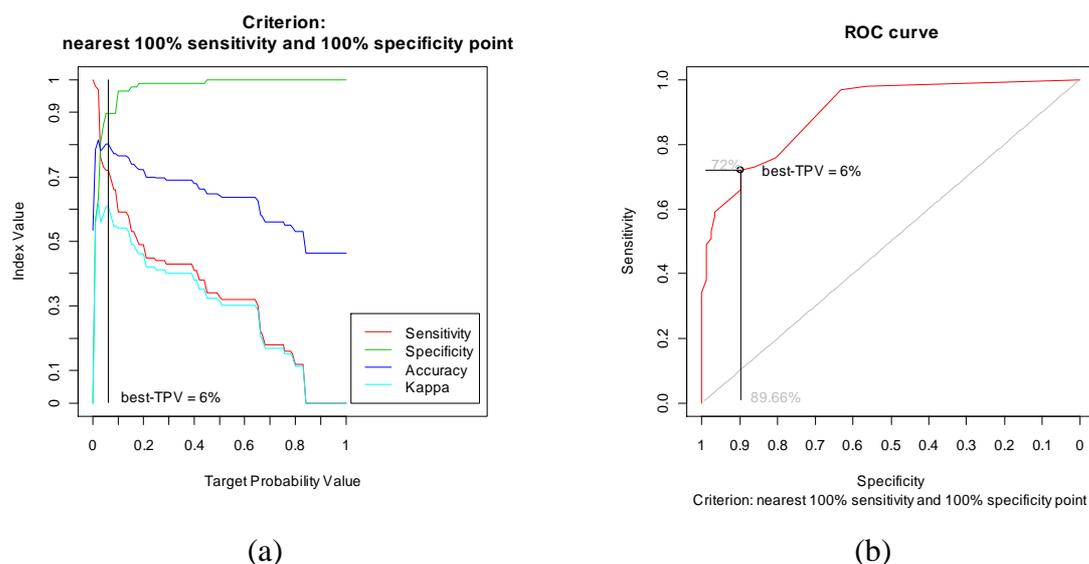


Figura 7. Gráficos de saída do algoritmo ilustrando o melhor valor para o fatiamento, segundo o critério escolhido. Em (a) os valores dos índices sensibilidade, especificidade, exatidão global e *kappa* em função do valor limite do fatiamento para a variável alvo (soja); em (b) a curva ROC e o ponto mais próximo da coordenada [1,1], escolhido como melhor valor para o fatiamento (*best-Target Probability Value* – veja Mello et al. 2010 e Mello et al. 2011).

O mapa temático resultante deste fatiamento, quando avaliado com 100 amostras de cada classe utilizando amostragem aleatória simples, apresentou sensibilidade de 72,0%, especificidade de 89,7%, exatidão global de 80,2% e índice *kappa* de 0,608. Além disso, a área abaixo da curva ROC foi de 90,1% do *plot*, o que significa que a PI apresenta desempenho satisfatório (Hanley e McNeil, 1982).

#### 4. Conclusão

Este trabalho apresentou uma modelagem, baseada em redes Bayesianas para identificar áreas de soja no contexto da Moratória da Soja. Os resultados apresentados confirmam que a metodologia representa um incremento de informação no atual processo de monitoramento associado à Moratória da Soja por meios de dados remotamente sensoriados e Sistemas de Informações Geográficas.

Com a imagem de probabilidade (PI) gerada, foi possível estabelecer critérios de prioridade. Assim, ao invés de simplesmente fatiar para gerar um mapa temático, pode-se utilizar os valores da PI para visualizar os locais onde a probabilidade da presença de soja é maior, permitindo assim otimizar os recursos e a logística de fiscalização.

Atualmente, a implementação feita por Mello et al. (2011) trabalha a abordagem *pixel a pixel*, o que provavelmente diminuiu a eficiência da aplicação, uma vez que a abordagem por objetos, em especial, polígonos, seria mais adequada para o problema estudado. Além disso, os resultados poderiam ter sido melhores se o algoritmo considerasse relações de vizinhança, o que poderia ser conseguido por aperfeiçoamentos na implementação ou mesmo utilizando, como variável de entrada algum dado de correlação espacial.

#### Agradecimentos

Os autores agradecem a todos do Laboratório de Sensoriamento Remoto em Agricultura e Floresta (LAF), do INPE, em especial à M.Sc. Juliana S. Santos pelo banco de dados geográficos cedido, e ao Fernando Yuzo Sato por providenciar e processar a série temporal do EVI/MODIS.

## 5. Referências Bibliográficas

- Alencar, A.; Nepstad, D.; McGrath, D.; Moutinho, P.; Pacheco, P.; Diaz, M. D. C. V.; Soares Filho, B. **Desflorestamento na amazônia: indo além da “emergência crônica”**. 1 ed.; IPAM: Belém, 2004; 89 p.
- Alves, D. S. Taking things public: a contribution to address human dimensions of environmental change. **Philosophical Transactions of the Royal Society B-Biological Sciences**, v. 363, n. 1498, p. 1903-1909, 2008.
- Brown, J. C.; Koeppe, M.; Coles, B.; Price, K. P. Soybean production and conversion of tropical forest in the Brazilian Amazon: The case of Vilhena, Rondonia. **Ambio**, v. 34, n. 6, p. 462-469, 2005.
- Cardille, J.; Foley, J. A., Agricultural land-use change in Brazilian Amazônia between 1980 and 1995: evidence from integrated satellite and census data. **Remote Sensing of Environment**, v. 87, n. 4, p. 551-562, 2003.
- Congalton, R. G.; Green, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 2 ed. Boca Raton, FL, USA: CRC Press, 2009. 183 p.
- Daubechies, I. **Ten lectures on wavelets**. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992. 357 p.
- Fearnside, P. M.; Laurance, W. F. Comment on "Determination of deforestation rates of the world's humid tropical forests". **Science**, v. 299, n. 5609, p. 1015, 2003.
- Foley, J. A.; Defries, R.; Asner, G. P.; Barford, C.; Bonan, G.; Carpenter, S. R.; Chapin, F. S.; Ceo, M. T.; Daily, G. C.; Gibbs, H. K.; Helkowski, J. H.; Holloway, T.; Howard, E. A.; Kucharik, C. J.; Monfreda, C.; Patz, J. A.; Prentice, I. C.; Ramankutty, N.; Snyder, P. K. Global consequences of land use. **Science**, v. 309, n. 5734, p. 570-574, 2005.
- Freitas, R. M.; Shimabukuro, Y. E. Combining wavelets and linear spectral mixture model for MODIS satellite sensor timeseries analysis. **Journal of Computational Interdisciplinary Sciences**, v. 1, n. 1, p. 33-38, 2008.
- Galford, G. L.; Melillo, J. M.; Kicklighter, D. W.; Cronin, T. W.; Cerri, C. E. P.; Mustard, J. F.; Cerri, C. C. Greenhouse gas emissions from alternative futures of deforestation and agricultural management in the southern Amazon. **Proceedings of the National Academy of Sciences**, v. 107, n. 46, p. 19649-19654, 2010.
- Greenpeace, **Eating up the Amazon**; Greenpeace International: 2006. 64 p.
- Hanley, J. A.; McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. **Radiology**, v. 143, n. 1, p. 29-36, 1982.
- INPE. **Monitoramento da Floresta Amazônica Brasileira por Satélite** - Estimativas anuais de desflorestamento desde 1988 até 2009. 2010. Disponível em: <[http://www.obt.inpe.br/prodes/prodes\\_1988\\_2009.htm](http://www.obt.inpe.br/prodes/prodes_1988_2009.htm)>. Acesso em: 02 Ago. 2010.
- Keitt, T. H.; Bivand, R.; Pebesma, E.; Rowlingson, B. **rgdal: bindings for the geospatial data abstraction library**. R package version 0.6-28. 2010.
- Mello, M. P.; Rudorff, B. F. T.; Adami, M.; Aguiar, D. A. An R implementation for Bayesian Networks applied to spatial data. In: Conference on Spatial Statistics (SPAT), 1., 2011, Enschede, The Netherlands. **Proceedings...** Amsterdam: Elsevier, (aceito).
- Mello, M. P.; Rudorff, B. F. T.; Adami, M.; Rizzi, R.; Aguiar, D. A.; Gusso, A.; Foseca, L. M. G. A simplified Bayesian Network to map soybean plantations. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 30., 2010, Honolulu, Hawaii, USA. **Proceedings...** Los Alamitos, CA, USA: IEEE, 2010. p. 351-354.
- Pena, S. D. Bayes: o “cara”!. **Revista Ciência Hoje**, v. 38, n. 228, p. 22-29, 2006.
- R Development Core Team. **R: a language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing. 2010.
- Rabus, B.; Eineder, M.; Roth, A.; Bamler R. The shuttle radar topography mission - a new class of digital elevation models acquired by spaceborne radar. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 57, n. 4, p. 241-262, 2003.
- Rizzi, R., Rudorff, B. F. T.; Shimabukuro, Y. E.; Doraiswamy, P. C. Assessment of MODIS LAI retrievals over soybean crop in Southern Brazil. **International Journal of Remote Sensing**, v. 27, n. 19, p. 4091-4100, 2006.
- Rudorff, B. F. T.; Adami, M.; Aguiar, D. A.; Moreira, M. A.; Mello, M. P.; Fabrini, D.; Amaral, D. F.; Pires, B. M. The Soy Moratorium in the Amazon biome monitored by remote sensing. **Remote Sensing**, (submetido).
- Zweig, M. H.; Campbell, G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. **Clinical Chemistry**, v. 39, n. 8, p. 561-577, 1993.