



Ministério da
Ciência e Tecnologia



SER-300

**USO DE REDES BAYESIANAS NA DETECÇÃO DA CULTURA DA SOJA EM
PRIMAVERA DO LESTE-MT**

Isaque Daniel Rocha Eberhardt

Monografia escrita na disciplina de Introdução ao Geoprocessamento, orientada pelo
Dr. Antônio Miguel Vieira Monteiro

INPE
São José dos Campos
2013

USO DE REDES BAYESIANAS NA DETECÇÃO DA CULTURA DA SOJA EM PRIMAVERA DO LESTE-MT

Isaque Daniel Rocha Eberhardt ¹

¹Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
isaque@dsr.inpe.br

Abstract: The soybean crop occupies important position in Brazilian agriculture and the Mato Grosso state (MT) is the biggest producer in Brazil. The aim this work is applies a Bayesian network (BN) to map soybean crops in Primavera do Leste/MT region and study the convergence and divergence between the soybean reference map and BN soybean maps. The BN was training using eight context variables, including CEI images for crops seasons 2009/2010, 2010/2011 and 2011/2012, maps of soil, roads, slope, hydrograph and soybean reference map. The BN using 66% of reference map and the remaining context variables for training and 33% of soybean reference map to test classification. Using the probability image was generated two soybean maps adopted ROC curve criteria for define threshold for training and test samples. The soybean maps generated from training and test samples presents specificity 88,8 and 89,40% and sensibility of 90,6 and 90,15%, respectively. These maps has acreage estimate with difference of 1,61 and 1,13% for maps produced using training and test samples, respectively. Thus, the Bayesian network presents great potential to map the soybean area in Primavera do Leste with using a large number of context variables.

Keywords: Spatial analyses, thematic maps, context variables, probability, acreage.

Resumo: A cultura da soja ocupa posição de destaque na agricultura brasileira, tendo no estado do Mato Grosso (MT) o maior produtor. Então, o objetivo deste trabalho foi utilizar uma rede bayesiana (RB) na região de Primavera do Leste/MT para mapear as áreas de soja e analisar as convergências e divergências existentes entre o mapa de referência e os mapas produzidos pela RB. A RB foi treinada utilizando oito variáveis de contexto, entre elas, imagens CEI 2009/2010, 2010/2011 e 2011/2012, mapas de solos, estradas, declividade, hidrografia e um mapa de soja como referência. A RB foi treinada utilizando 66,6% das áreas de soja contidas no mapa de referência bem como as demais variáveis de contexto os demais 33,3% do mapa foram destinados para testar a classificação. A partir da imagem de probabilidade foram gerados dois mapas de soja utilizando a curva ROC como critério para definir o limiar, um tendo como base as amostras de treinamento e outro sobre as amostras de teste. Os mapas de soja foram gerados a partir das amostras de treinamento e teste apresentaram especificidade de 88,8 e 89,40% e de sensibilidade de 90,6 e 90,15%, respectivamente. Estes mapas apresentaram estimativas de área cultivada com diferença de 1,61 e 1,13% para as amostras de treinamento e teste respectivamente. De tal forma, a rede bayesiana demonstrou ótimo potencial de identificação das áreas de soja de Primavera do Leste utilizando um grande número de variáveis de contexto.

Palavras-chave: Análise espacial, mapa temático, variáveis de contexto, probabilidade, estimativa de área.

1 INTRODUÇÃO

A soja apresenta destacada posição entre os cultivos agrícolas mundiais, com projeções de área cultivada superior a 110 milhões de hectares na safra 2013/2014 (USDA, 2013). No Brasil, a soja encontra-se distribuída em praticamente todos os estados, que juntos cultivaram aproximadamente 27 milhões de hectares na safra 2012/2013 (IBGE, 2013). O principal estado produtor brasileiro continua sendo o Mato Grosso (MT) que respondeu por 25% deste total para a mesma safra (CONAB, 2013). Segundo a CONAB (2012) foram produzidas aproximadamente 81 milhões de toneladas de soja para a referida safra. Grande parte desta produção é destinada às exportações, que totalizaram divisas para o país da ordem de 25 bilhões de dólares na safra 2010/2011 (CONAB, 2012), tornando o Brasil o segundo maior produtor e o maior exportador mundial de soja (USDA, 2013).

Em uma cadeia produtiva com estas dimensões existe uma forte competitividade entre os diversos atores envolvidos, que dependem de informações sobre a safra da soja para subsidiar as suas tomadas de decisões. Segundo Pino (1999), de maneira geral, as estatísticas agrícolas disponibilizadas pelos serviços governamentais utilizam em sua produção, informações oriundas de entrevistas/informações e/ou opiniões de agentes ou atores locais (tais como empresas de comercialização ou assistência técnica; IBGE, 2002), este fato pode levar a imprecisões no resultado final da estimativa. Ademais, este procedimento acaba impossibilitando a aplicação de ferramentas matemáticas/estatísticas para a avaliação dos erros, gerando incertezas quanto à qualidade destas estimativas (PINO, 1999). Então, métodos com esta natureza apresentam limitações para captar pequenas variações entre safras, principalmente em regiões de expansão de fronteira agrícola (PINO, 2001). Adicionalmente, levantamentos deste tipo não disponibilizam a localização dos talhões cultivados, o que os torna pouco eficazes na solução de problemas logísticos, de armazenamento e escoamento da produção, por exemplo. Entretanto, o fato mais relevante está vinculado ao tempo de disponibilização das informações, que ocorre em geral em datas demasiadamente tardias para que seja possível a tomada de decisão tendo este tipo de informação como referência.

Alternativamente, alguns trabalhos têm apontado a utilização da classificação digital e/ou fotointerpretação em imagens do tipo Landsat para monitoramentos em escala regional (RIZZI; RUDORFF, 2005; RUDORFF et al., 2010). Entretanto, a utilização destas técnicas depende de imagens adquiridas em períodos chaves do desenvolvimento da cultura em questão e que estas estejam livres de cobertura de nuvens. Ademais, a presença de cobertura de nuvens em regiões tropicais, a exemplo do MT, configura-se como um grande desafio ao longo de toda a safra (ASNER, 2001). Outra limitação é o nível de conhecimento necessário ao fotointérprete sobre os alvos dispostos na região em estudo e o tempo demandado para executar todas as etapas do processo. Desta forma, atendidos estes preceitos é possível gerar uma estimativa agrícola confiável sob ponto de vista de acurácia, contudo para tanto dependerá de uma quantidade de recursos disponíveis para o uso, mas poderá estar disponível demasiadamente tarde para auxiliar a tomada de decisões.

Por outro lado, com o lançamento do sensor *MODerate Resolution Imaging Spectroradiometer* (MODIS) (JUSTICE et al., 1998), ampliaram-se as possibilidades de disponibilização de imagens com menor incidência de nuvens, em razão de sua resolução temporal de 1 a 2 dias, (HUETE et al., 1999; LEEUWEN et. al, 1999; DIDAN; HUETE, 2006). A partir de imagens MODIS é reproduzido o *Enhanced Vegetation Index* (EVI) (HUETE et al., 1997), que têm atingido destaque entre em estudo envolvendo a dinâmica espaço-temporal da vegetação (MULLER et al., 1998; GUSSO et al., 2012; RISSO et al., 2012). Como exemplo, Rizzi et al. (2009) desenvolveram o *Crop Enhancement Index* (CEI), capaz de mapear e estimar a área de soja em grandes regiões ou em nível estadual, utilizando imagens EVI adquiridas em dois períodos específicos da safra da cultura. O primeiro, denominado Mínimo EVI (MinEVI) coincide com o período da entressafra que antecede a semeadura da soja e o segundo, denominado Máximo EVI (MaxEVI), utiliza imagens adquiridas durante as fases de desenvolvimento até o início da senescência da cultura. Esta metodologia foi desenvolvida e testada para as condições do MT e alcançou uma exatidão global do mapeamento de 91%, disponibilizando a identificação e estimativa das áreas de soja sempre a partir de início de março de cada safra (RIZZI et al., 2009). Entretanto, esta

metodologia depende de um limiar fixo de classificação que em última instância segue a lógica booleana determinando que todos os pixels da imagem CEI acima de 0,28 são considerados como sendo pertencentes a classe soja (RIZZI et al., 2009). Uma classificação que utiliza um critério rígido (booleano) pode ser incapaz de classificar pixels que por algum motivo representam o objeto de estudo com valores inferiores aos demais de sua mesma classe, resultando em erros no processo de classificação (ROCCHINI et al., 2013). Adicionalmente, métodos de classificação a semelhança do CEI não possibilitam uma avaliação das incertezas contidas neste processo de classificação (ROCCHINI et al., 2013; COMBER et al., 2012) Ademais, a metodologia CEI considera somente as imagens EVI como fonte de informação, desprezando em sua formulação qualquer outro tipo de dado geográfico.

Então, uma maneira de utilizar múltiplas condições no processo de classificação do uso e cobertura do solo pode ser por meio da utilização de um algoritmo de Bayes implementado em por meio de uma Rede Bayesiana (RB) (QIN et al., 2006). Este método utiliza a teoria de probabilidade para estabelecer qual a probabilidade de um dado pixel representar uma determinada classe (MELLO et al., 2010 e 2011). De forma que, por meio de um modelo simplificado de RB Mello et al. (2010) utilizaram diferentes variáveis de contexto para o treinamento do algoritmo, dentre elas, imagens EVI (de mínimo e de amplitude) e um mapa de declividade. Por meio desta técnica Mello et al. (2010) conseguiram que o mapa de soja gerado a partir da imagem de probabilidade contivesse valores de 77,62, 77,56 e 77,5% para sensibilidade, especificidade e acurácia geral, respectivamente. De modo semelhante, Mello et al. (2011) aplicaram RB utilizando como variáveis de contexto imagens de Max e MinEVI, amplitude EVI, mapa de hidrografia, um mapa de estradas, um mapa de declividade, um mapa desflorestamento e um mapa de referencia obtido através de fotografias aéreas. Assim, alcançaram valores de sensibilidade de 73,5%, especificidade de 90,9% e acurácia global de 82,1%. Contudo, nenhum destes resultados foi alcançado utilizando mapas de referência obtido a partir de imagens *Landsat-like*.

Neste contexto, este trabalho tem por objetivo aplicar um algoritmo de Redes Bayesianas a região de Primavera do Leste no estado do MT para mapear as áreas de

soja utilizando variáveis de contexto de diferentes resoluções para treinamento desta rede bayesiana e analisar as convergências e divergências entre o mapa de referência e os mapas produzidos a partir da imagem de probabilidade.

2 MATERIAL E MÉTODOS

2.1. A região de estudo

A região em estudo compreende a porção sul do município de Primavera do Leste no estado do MT, e está delimitada pelas latitudes $15^{\circ} 07' 30''$ e $15^{\circ} 39' 10''$ e as longitudes $53^{\circ} 51' 15''$ e $54^{\circ} 33' 30''$ (Figura 2.1). A safra escolhida para estudo foi a de 2011/2012, por conta da possibilidade de utilização de um mapa contendo a localização geodésica dos talhões de soja na região. Este mapa foi produzido por meio de fotointerpretação de imagens Landsat (EBERHARDT, 2012) para um projeto chamado SusteAgro da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA).

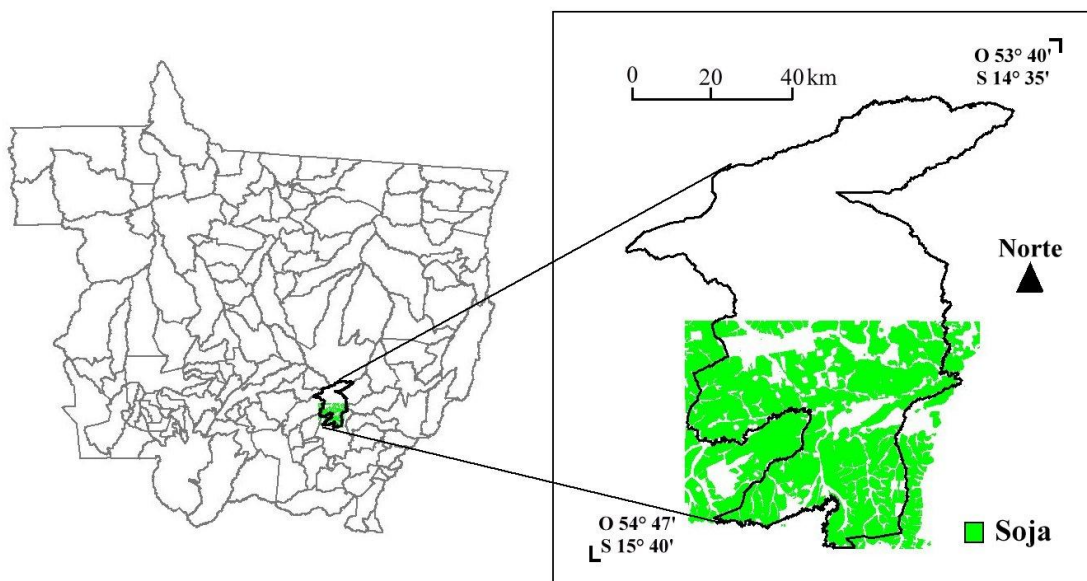


Figura 2.1 - Região de estudo, em evidência as áreas cultivadas com soja no mapa de referência.

2.2. Modelo OMT-G

Com o intuito de definir os processos envolvidos na aplicação de uma rede bayesiana para classificar as áreas de maior probabilidade de ocorrência de soja, a metodologia proposta seguiu a lógica de *Object Modeling Technique for Geographic Applications* (OMT-G). Para a metodologia proposta neste trabalho o modelo OMT-G está demonstrado na Figura 2.2.

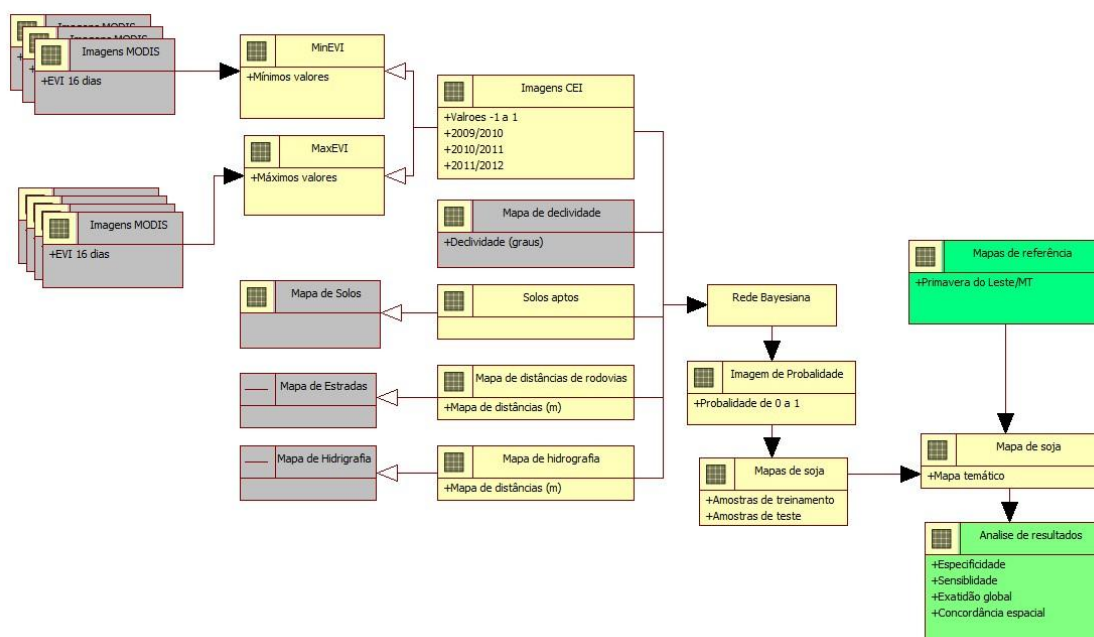


Figura 2.2 – Modelo OMT-G para a classificação das áreas de soja utilizando a técnica bayesiana.

2.3. Geração das variáveis de contexto

O ajuste, formatação e a análise de todas as variáveis de contexto utilizadas para a aplicação da rede bayesiana foi realizada no Sistema de Processamento de Informações Georeferenciadas (SPRING). O SPRING foi concebido sob os conceitos de Geoprocessamento, dentre os quais o banco de dados é a sua unidade mais importante, por agrupar, organizar e tornar manipuláveis os dados geográficos (CÂMARA et al., 2001). Este *software* é Sistema de Informação Geográfica (SIG), produzido e distribuído gratuitamente via rede mundial de computadores pela Divisão de Processamento de Imagens (DPI) do Instituto Nacional de Pesquisas Espaciais (INPE).

Entre as variáveis de contexto foi utilizado um mapa de declividade com resolução de 30x30m do TOPODATA (VALERIANO; ALBUQUERQUE, 2010) disponível em <http://www.dsr.inpe.br/topodata/acesso.php>. Estes mapas foram elaborados através da utilização de técnicas da geoestatística para interpolar utilizando o processo de krigagem (VALERIANO; ROSSETTI, 2010) sobre os dados do *Shuttle Radar Topography Mission* (SRTM) (FARR et al., 2007; RODRIGUEZ et al., 2005) que são originalmente disponibilizados com resolução espacial de 90x90m.

Outra fonte de informações utilizada foi a malha hidrográfica da região de Primavera do Leste, a qual foi obtida junto a base de dados georreferenciados da Agência Nacional das Águas (ANA) no website <http://www.ana.gov.br/bibliotecavirtual/solicitacaoBaseDados.asp>. Este mapa vetorial foi gerado na escala de 1:1.000.000 a partir da delimitação das chamadas ottobacias ou trechos de contribuição tendo por base os modelos digitais de elevação (MDE) (ANA, 2007). Então, com base neste mapa vetorial contendo a hidrografia da área de estudo foi gerado um mapa de distâncias dos cursos de água, de maneira que cada pixel deste mapa contém o valor referente a esta distância em relação aos corpos de água com uma resolução de espacial de 30x30m.

Semelhantemente, foi aplicado um conjunto de procedimentos para a elaboração de um mapa de distâncias de estradas para a região de estudo. Para tanto, o mapa vetorial contendo a localização das principais estradas que cortam o município de Primavera do Leste foi obtido junto a Agência Nacional de Energia Elétrica (ANEEL), em um sistema chamado Sistema de Informações Georreferenciadas do Setor Elétrico (SIGEL, 2013) disponível no website <http://sigel.aneel.gov.br/>. A partir deste mapa vetorial foi gerado o mapa de distâncias em relação às estradas da região em estudo, com resolução de 30x30m.

Um mapa de solos da região em estudo foi utilizado para definir quais os solos deveriam ser considerados como aptos ao cultivo de soja na região. Para tanto foram identificados os solos da região em estudo e selecionados por meio de uma operação chamada de “consulta por atributos” realizada diretamente no modelo de dados Cadastral banco de dados elaborado no SPRING contendo um mapa de solos do estado

do MT produzido pela Secretaria de Estado de Planejamento e Coordenação Geral do estado do MT (SEPLAN, 2013) com escala de 1:750.000. Para estabelecer quais dos solos presentes na região seriam selecionados foi utilizado como parâmetro as informações as necessidades da cultura no MT (EMBRAPA, 2003). Seguindo estas orientações, foram selecionados como solos aptos ao cultivo de soja os Latossolos Vermelho-Amarelo Podzólicos e Latossolo Vermelho-Escuro.

Em seguida foram geradas as imagens CEI para três safras que serviram de informação espectro-temporal para o treinamento e teste da rede Bayesiana. Esta metodologia CEI (RIZZI et al., 2009) utilizava em sua formatação original imagens EVI adquiridas entre os Dias do Ano (DA) 241 (19 de agosto) e 320 (16 de novembro) para a formulação da imagem de MinEVI. Entretanto, Risso et al. (2012) apontaram vantagens ao antecipar o MinEVI para o período compreendido entre os DA 161 (10 de junho) e 224 (12 de agosto). Este procedimento amplia o potencial de contraste do EVI entre os períodos pré/plantio e de máximo vigor vegetativo das áreas de soja melhorando a discriminação para as demais classes de uso/cobertura do solo, além reduzir a probabilidade de presença de pixels com nuvens nas imagens de MinEVI. De tal forma, para este trabalho foi adotado para o Min o uso de imagens EVI adquiridas entre os DA 161 (10 de junho) e 224 (12 de agosto), bem como, entre 321 (17 de novembro) e 64 (5 de março) para a elaboração do MaxEVI. Em seguida, as imagens de Min e MaxEVI foram aplicadas na equação CEI (Equação 2.1), que retorna a máxima diferença ocorrida entre os períodos, resultando na imagem CEI.

$$CEI = G * \frac{(MaxEVI+S)-(MinEVI+S)}{(MaxEVI+S)+(MinEVI+S)} \quad (2.1)$$

em que, S é o a o coeficiente de realce ($S=10^2$) e o G é o fator de ganho ($G=10^2$). Cabe ressaltar aqui, que as imagens MODIS utilizadas são aquelas do produto MOD13Q1 (HUETE et al., 1999) as quais são composições multitemporais de 16 dias elaboradas imagens de resolução espacial de 250x250m. Para que tornar estas imagens CEI compatíveis com os demais mapas (que tem resolução de 30x30m), foi aplicado um procedimento de reamostragem, executado por meio de uma álgebra de mapas implementada na Linguagem Espacial de Geoprocessamento Algébrico (LEGAL) do

SPRING. Por meio de um script do LEGAL, as imagens CEI foram reamostradas utilizando o interpolador “vizinho mais próximo”, o qual retorna uma imagem sem alterações na distribuição espacial dos valores dos pixels, somente gerando células de menor tamanho.

2.4. Redes Bayesianas

Segundo Doguc e Ramirez-Marquez (2009) uma rede Bayesiana (RB) pode seguir uma abordagem gráfica para determinar a relação entre variáveis. No caso deste trabalho, foi utilizado um algoritmo construído por Mello et al. (2011) implementado na linguagem do *software* R (R Development Core Team, 2013) que utiliza como ponto de partida a definição de um grafo acíclico direcional que descreve, utilizando arcos (setas) as relações de dependência entre pares de variáveis (onde cada variável é representada por um nó; Figura 2.3).

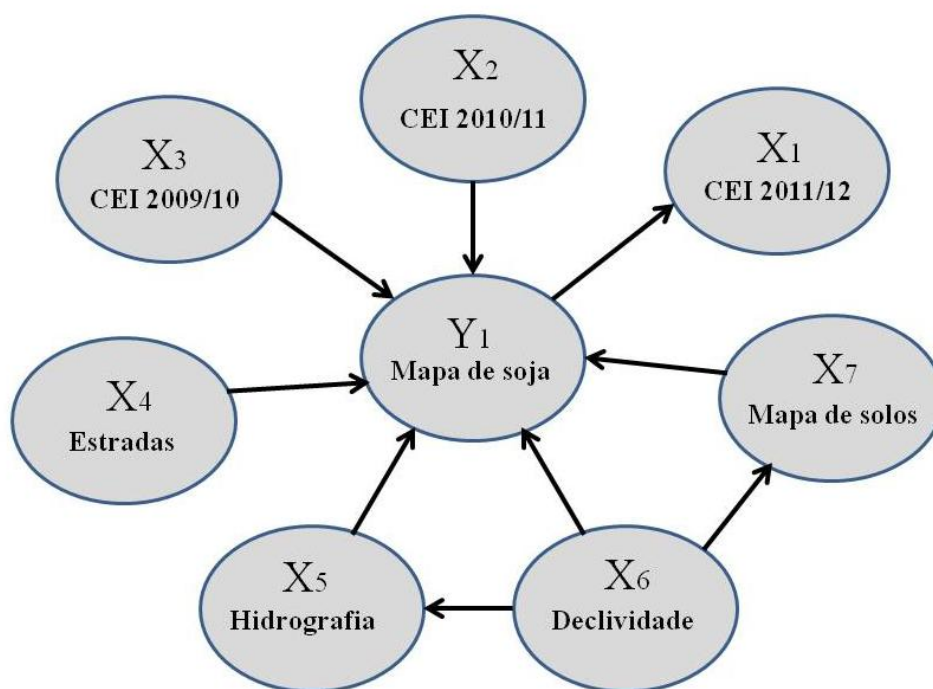


Figura 2.3 - Grafo acíclico direcional para o modelo gráfico de rede Bayesiana definido neste estudo.

Este grafo define a relação probabilística de ocorrência entre as variáveis de contexto e o mapa de treinamento. Deste modo, as variáveis de contexto que influenciam a probabilidade de ocorrência de soja são definidas com “pais” para o mapa de soja, como por exemplo, a imagem CEI 2009/2010 que apresenta a sua seta direcionada para o mapa de soja, evidenciando que na construção do modelo bayesiano para este estudo o mapa CEI 2009/2010 influencia a probabilidade de ocorrência de soja, onde o mapa de soja é “filho” de CEI 2009/2010, e assim sucessivamente para todas as variáveis de contexto (Figura 2.3).

Uma das dificuldades no uso de redes Bayesianas é a definição das funções de probabilidades. Contudo, Mello et al., (2010) propôs a utilização de contagem de pixels em variáveis discretizadas para definir, empiricamente essas funções. A discretização das variáveis é realizada através de um processo iterativo e com base no conhecimento do especialista, de tal forma a definir o número de intervalos e seus limites (Figura 2.4).

A partir dos intervalos definidos a RB passa pelo processo de treinamento, que consiste na definição das funções de probabilidade. Após o treinamento é possível gerar uma Imagem de Probabilidade (IP) onde, em cada pixel, um número entre 0 e 1 representa a probabilidade da presença de soja tendo por nada dados os valores observados para as variáveis de contexto (Figura 2.3), como descrito na Equação 2.2,

$$P(Y | X_1, X_2, X_3, X_4, X_5, X_6, X_7) = \frac{P(X_1, X_2, X_3, X_4, X_5, X_6, X_7 | Y) \cdot P(Y)}{P(X_1, X_2, X_3, X_4, X_5, X_6, X_7)}. \quad (2.2)$$

Y corresponde a ocorrência de soja, podendo assumir 0 para ausência de probabilidade e 1 para a máxima probabilidade de ocorrência de soja. As variáveis de contexto utilizadas foram: (X₁), (X₂) e (X₃) correspondentes às imagens CEI 2011/2012, 2011/2010 e 2009/2010, respectivamente; (X₄) é à distância até a estrada principal mais próxima; (X₅) é à distância até o curso de água mais próximo; (X₆) é a declividade; e (X₇) é o tipo de solo, classificado em adequado ou inadequado para o cultivo da soja.

Cabe ressaltar que tanto os pixels a serem utilizados no processo de treinamento quanto aqueles que foram destinados ao processo de teste foram selecionados por meio de um processo de escolha aleatória implementada no software R (R Development Core Team,

2013). Assim, de todas as áreas do mapa de referência, 66,6% foi selecionada para compor a mostra de treinamento. Os demais 33,3% foram utilizados como amostras de teste para avaliar a acurácia da do mapa temático gerado a partir da IP. A partir de um processo iterativo realizado pelo algoritmo bayesiano, classificando a imagem de probabilidade e comparando o resultado desta classificação ao mapa de teste, foram definidos os valores de especificidade e sensibilidade.

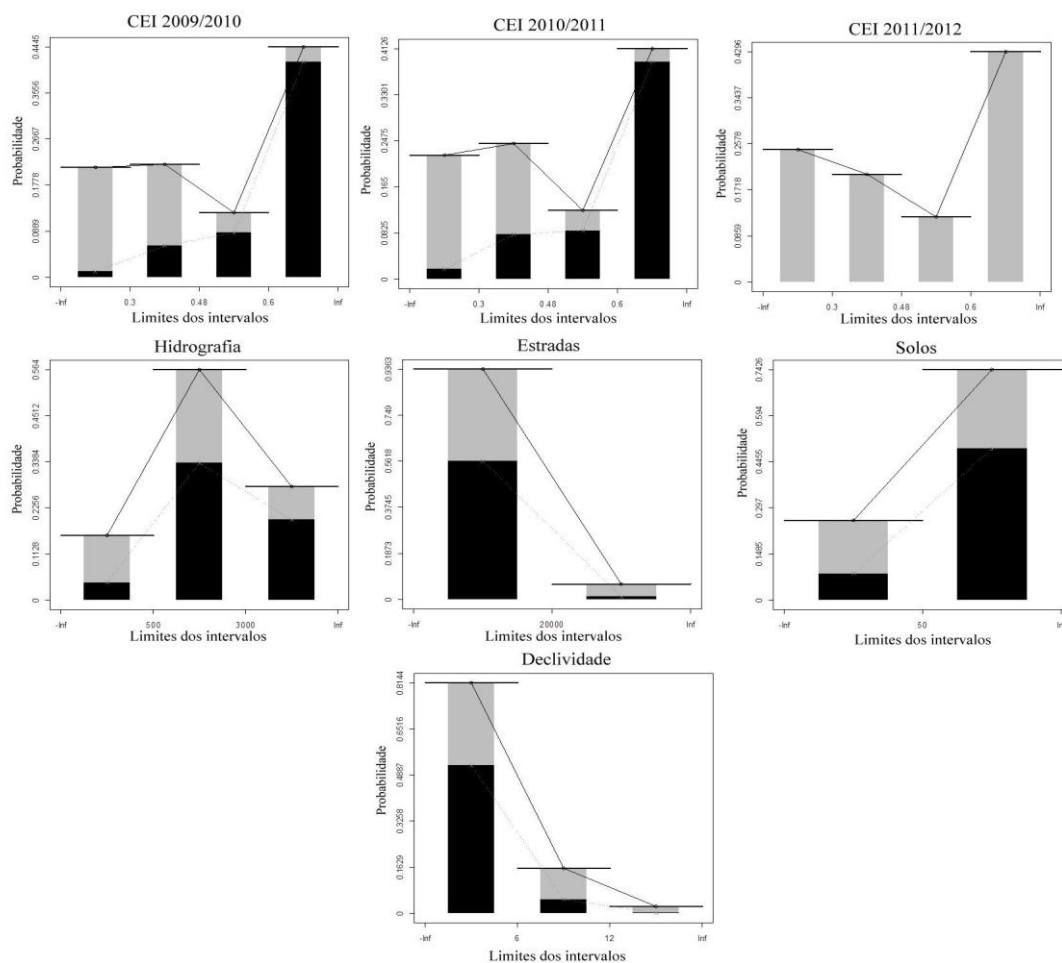


Figura 2.4 - Discretização das variáveis e suas distribuições de probabilidade, conforme rede Bayesiana definida na Figura 2.2. Em cada histograma estão dispostos os valores e o número de classes utilizado para treinar o modelo bayesiano. As barras representam os intervalos definidos na discretização das variáveis e a porção em preto representa a probabilidade da presença de soja dentro do intervalo.

Baseado nos valores de sensibilidade e especificidade de cada mapa temático, gerado a partir da IP com diferentes valores de probabilidade sendo usados como limites para que os pixels fossem rotulados como soja ou não-soja, gerou-se uma curva ROC. Segundo Zweig e Campbell (1993) o mapa temático de maior acurácia seria aquele cujo seu ponto correspondente na curva ROC apresenta a menor distância euclidiana para o ponto de 100% sensibilidade e 100% especificidade. Para este ponto de menor distância euclidiana para os valores de 100% é atribuído o nome de melhor valor de probabilidade, relacionando assim a acurácia da classificação (melhores valores de especificidade e sensibilidade) e o valor de probabilidade no qual será possível alcançar tais valores de acurácia. Então, com base no valor de probabilidade obtido por meio da curva ROC, foram calculados os valores de especificidade, sensibilidade, acurácia global e índice kappa para o mapa temático final gerado a partir da IP.

3 RESULTADOS E DISCUSSÃO

A técnica bayesiana apresenta diversos trabalhos voltados à detecção de padrões por meio do uso de funções de probabilidade (LAMPINEN, J; VEHTARI, 2001; QU et al., 2008; DOGUC; RAMIREZ-MARQUEZ, 2009; AYACHI et al., 2013). Entretanto, aplicação da técnica bayesiana voltada à identificação de áreas agrícolas ou ambientais ainda é incipiente (RITTER; HEPNER, 1990; NASH et al., 2013; QIN et al., 2006; CASTELLETTI; SONCINI-SESSA, 2007; UUSITALO, 2007). Tendo em vista este contexto, a análise dos resultados obtidos neste trabalho possibilitou elucidar algumas questões a respeito da eficácia/acurácia de um produto final da técnica bayesiana.

Primeiramente, a partir da RB utilizada no presente trabalho foi possível estabelecer os diferentes níveis de relação de dependência entre as variáveis de contexto utilizadas para realizar o treinamento da RB. De tal forma, quando analisamos a dependência da variável soja e os valores da imagem CEI 2011/2012, é perceptível que aos maiores os valores de CEI para esta safra estão associados à ocorrência de soja no mapa de referência (Figura 3.1).

Similarmente, quando analisamos a Figura 3.1, fica evidente que as classes de aptidão dos solos da região, os denominados *não aptos* (com valor zero) e os *aptos* (com valores identificados com 100) apresentam uma relação com a declividade. Visto que, ficam

perceptíveis duas porções distintas no gráfico, uma primeira, na qual os pixels que representam os solos classificados como *não aptos* estão correlacionados a diversos valores de declividade (com uma tendência de maior concentração nas menores declividades) e uma segunda porção na qual que os pixels que pertencem a classe de solos *aptos* estão correlacionados em maior número aos menores valores de declividade (Figura 3.1). Quando tratamos da relação entre hidrografia e declividade, fica evidenciada a sua predominante relação com os menores valores de declividade (Figura 3.1). Estas relações influenciam de certa forma, o resultado final da classificação do algoritmo bayesiano.

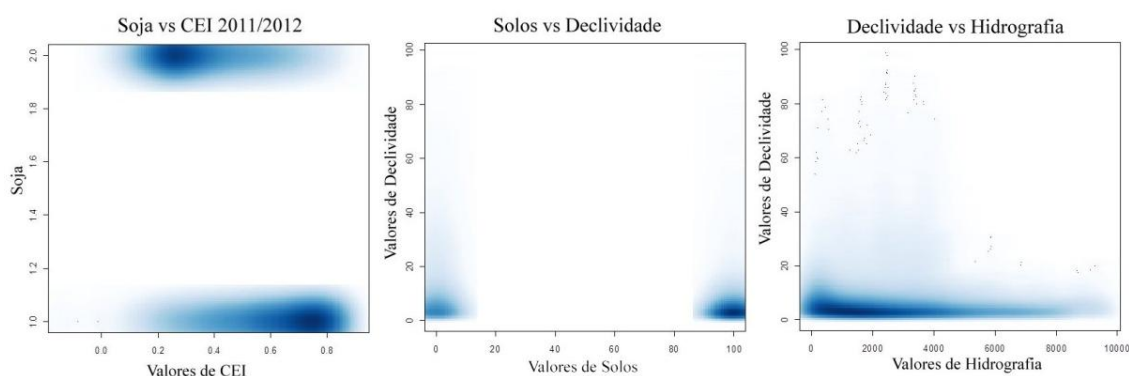


Figura 3.1 - Gráficos de dispersão que evidenciam a relação de dependência entre algumas variáveis de contexto utilizadas no modelo Bayesiano.

A partir dos limites dos intervalos definidos no momento de treinamento do algoritmo bayesiano utilizado foi possível estabelecer a contribuição ou peso de cada variável de contexto para a construção das funções de probabilidade estabelecidas para a formação da imagem de probabilidade. No caso da aplicação da RB deste trabalho, as variáveis utilizadas apresentaram contribuições diversas entre si, dependendo de sua característica e das relações estabelecidas entre as variáveis (Figura 2.2). Entre todas as oito variáveis de contexto utilizadas, as imagens CEI foram as que apresentaram o maior peso para as funções de probabilidade (Tabela 3.1). Estes pesos das variáveis de contexto são obtidos por meio de medidas de distância entre a função de distribuição de probabilidade observada e a função de distribuição de probabilidade marginal (estimada). Dentre as

imagens CEI utilizadas a que representa a safra 2011/2012 deteve o maior peso para as funções de probabilidade (Tabela 3.1).

Tabela 3.1 - Pesos das variáveis de contexto na elaboração da imagem de probabilidade.

| Variáveis de contexto | CEI 2012 | CEI 2011 | CEI 2010 | Estradas | Hidrografia | Declividade | Solos |
|------------------------------|-----------------|-----------------|-----------------|-----------------|--------------------|--------------------|--------------|
| Pesos | 1,3215 | 0,5437 | 0,5776 | 0,4747 | 0,4753 | 0,4846 | 0,4827 |

Quanto as variáveis de contexto, outro fato relevante é o de que algumas destas apresentam escala cartográfica bastante inferior às demais, entretanto, estas aparentemente não impactaram de forma prejudicial os resultados, visto que a imagem de probabilidade originada pela RB não apresenta áreas que tenham sofrido restrição ou delimitação dos valores de probabilidade em função destas variáveis de contexto. Ademais, analisando a Figura 3.2, é possível confirmar esta afirmação, já que a imagem de probabilidade não aparenta ter sido influenciada de forma negativa pelo uso do mapa de solos de uma escala cartográfica da ordem de 1:700.000 na construção das funções de probabilidade. Esta condição pode ser demonstrada já que existem áreas que apresentam valores elevados de probabilidade (próximos a 1, áreas em cinza) que não estão contidas pelas áreas delimitadas pelos vetores vermelhos (Figura 3.2) que apontam os solos considerados aptos ao cultivo de soja.

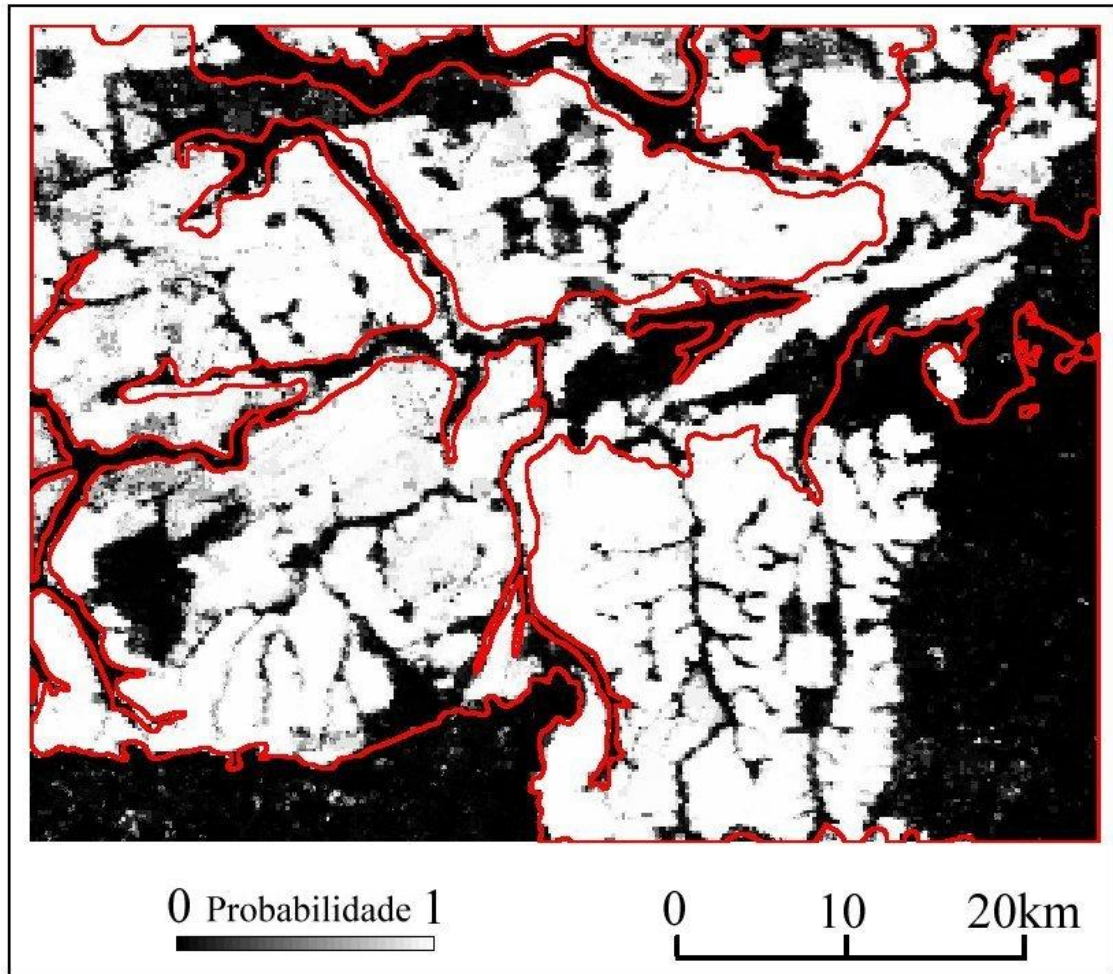


Figura 3.2 - Imagem de probabilidade associada ao mapa de solos considerados aptos ao cultivo de soja na região em estudo. Os limites apontados pelo vetor em vermelho representam os solos aptos para o cultivo de soja na região.

Semelhantemente, ao associar o mapa vetorial que representa as principais estradas da região (produzido em escala de 1:1.000.000) e a imagem de probabilidade não fica perceptível nenhum efeito negativo produzido por esta variável de contexto na imagem de probabilidade (Figura 3.3a). Da mesma forma, a Figura 3.3b, apresenta o mapa vetorial contendo a representação da hidrografia da região e a imagem de probabilidade. Esta representação vetorial também não produziu efeitos perceptíveis nos valores de probabilidade de ocorrência de soja, visto que em toda a região de estudo existem valores de probabilidade diversos mesmo que estejam em diferentes posições em

relação ao vetor da hidrografia. Ou seja, a rede bayesiana foi capaz de produzir elevados valores de probabilidade sem ser influenciada negativamente pelas variáveis de contexto obtidas em menores escalas, demonstrando a capacidade da RB para a elaboração da imagem de probabilidade a partir dos intervalos definidos superou as restrições impostas pelas variáveis de contexto de pior escala cartográfica.

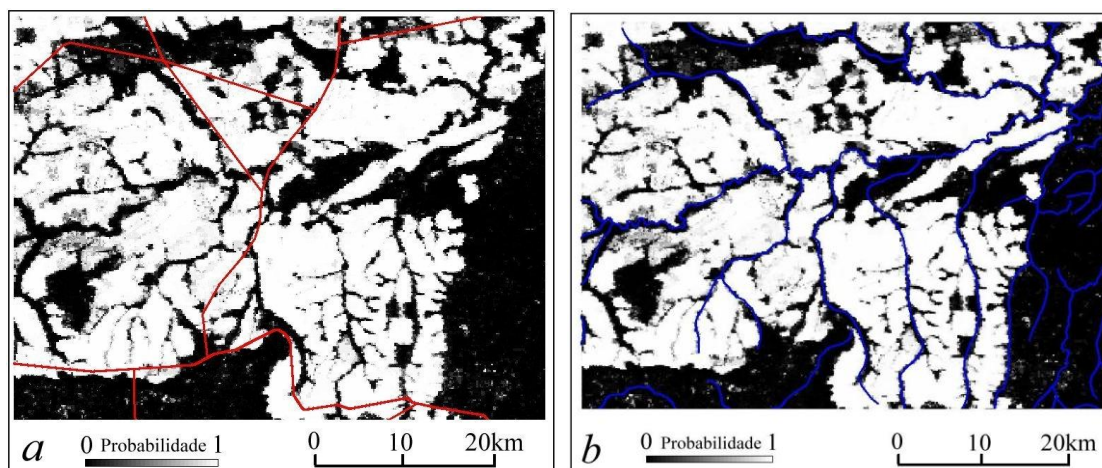


Figura 3.3 - Figura demonstrando o baixo impacto das variáveis de contexto de pior resolução na elaboração da imagem de probabilidade. Na figura 3.3 *a* está apresentada a imagem de probabilidade e o vetor de estradas (vetor em vermelho), a figura *b* contém a imagem de probabilidade e o mapa vetorial representando a hidrografia (vetor em azul).

A partir da imagem de probabilidade oriunda da RB foram classificados os pixels de maior probabilidade de pertencerem a classe soja utilizando um limiar definido a partir do valor de probabilidade que detenha a menor distância euclidiana entre as máximas sensibilidade e especificidade, este critério é definido pela curva ROC (ZWEIG; CAMPBELL, 1993). Então, seguindo o algoritmo implementado por Mello et al. (2011) foram gerados uma série de mapas temáticos correlacionando especificidade e sensibilidade até que foi encontrado este valor. No caso das amostras de treinamento retiradas do mapa de referência o valor de probabilidade foi de 0,47 ou 47% (Figura 3.4). Para este valor de probabilidade de 47% a análise do mapa temático de soja apresentou valores de especificidade e sensibilidade de 88,8 e 90,6%, respectivamente.

Junto a estes valores, o limiar de classificação definido pela curva ROC gerou um mapa com um índice kappa de 0,793 com variância para este índice de 0,0000001241941.

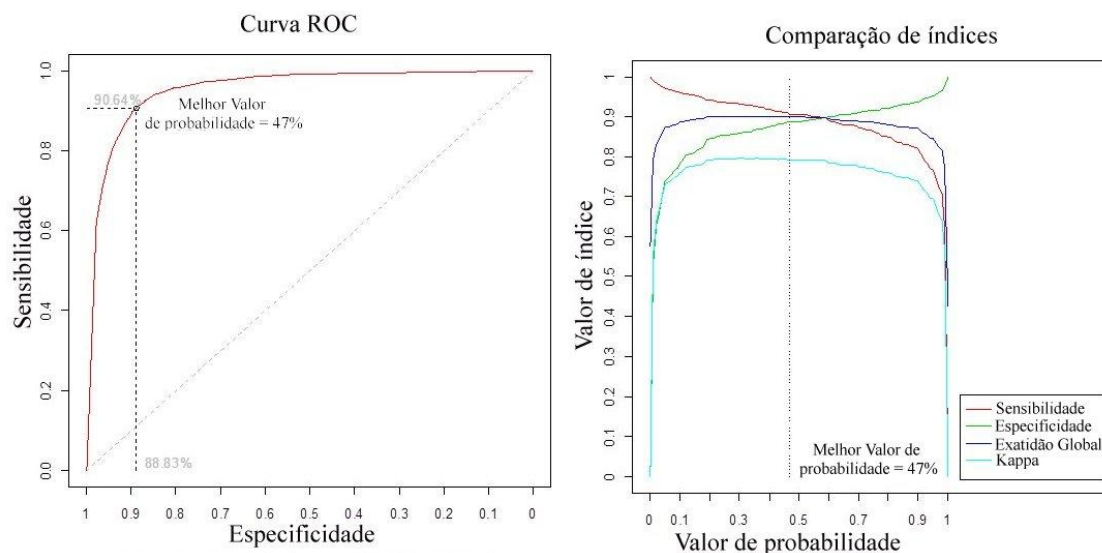


Figura 3.4 - Curva ROC para o resultado da simulação para determinar o valor de probabilidade por meio da menor distância euclidiana entre a melhor condição de sensibilidade e especificidade ao classificar a imagem de probabilidade. A simulação foi realizada com base nas amostras de treinamento.

Quando analisamos o mapa temático para a classe soja elaborado tendo como base as amostras de teste o valor de probabilidade que está associado a melhor relação entre especificidade e sensibilidade apresenta uma pequena diferença de valor apontado para as amostras de treinamento. De modo que, para as amostras de teste o valor de probabilidade identificado é de 0,46 ou 46% (Figura 3.5). Ademais, este mapa temático apresentou valores de especificidade e sensibilidade de 89,40 e 90,15%, respectivamente. Conjuntamente a estes valores, o mesmo mapa temático apresentou um índice kappa de 0,7929327 com uma variância associada de 0,0000002480052.

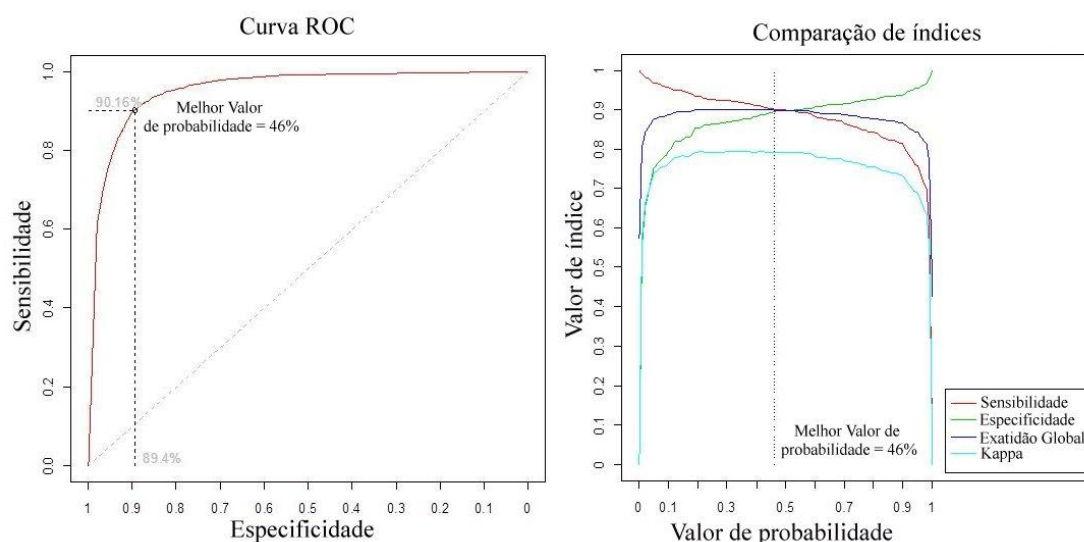


Figura 3.5 - Curva ROC para o resultado da simulação para determinar o valor de probabilidade por meio da menor distância euclidiana entre a melhor condição de sensibilidade e especificidade ao classificar a imagem de probabilidade. A simulação foi realizada com base nas amostras de teste.

Com a finalidade de evidenciar o resultado da classificação implementada utilizando limiares definidos a partir da curva ROC, foi realizada uma álgebra de mapas utilizando um *script* da linguagem LEGAL do SPRING, para gerar um mapa de resultados contendo as concordâncias e discordâncias existentes entre o mapa de referência e os mapas temáticos gerados pela RB utilizando as amostras de treinamento e teste (Figura 3.6). Nestes mapas de resultados estão contidas as quatro classes possíveis para a comparação entre o mapa de referência e os mapas oriundos da RB, dentre elas, estão às classes que representam a concordância entre os mapas, à classe *acerto soja* apresenta os pixels definidos como *soja* no mapa de referência e que foram identificados pela RB como sendo pertencentes à classe soja (pixels em verde). Semelhantemente, a classe *não soja* existente nos mapas de resultados é aquela representa os pixels da classe não soja no mapa de referência e nos mapas produzidos pela RB (áreas em cinza). Quanto às porções em desacordo com o mapa de referência, os mapas de resultados apresentam a classe *inclusão* definidos como os pixels que não pertencem à classe *soja* no mapa de referência e foram identificados como *soja* através da classificação bayesiana (regiões em amarelo). E por fim, a classe definida como *omissão* (porções em vermelho)

corresponde aos pixels que pertencem à classe soja no mapa de referência e não foram identificados como tal pelo algoritmo implementado por Mello et al. (2011).

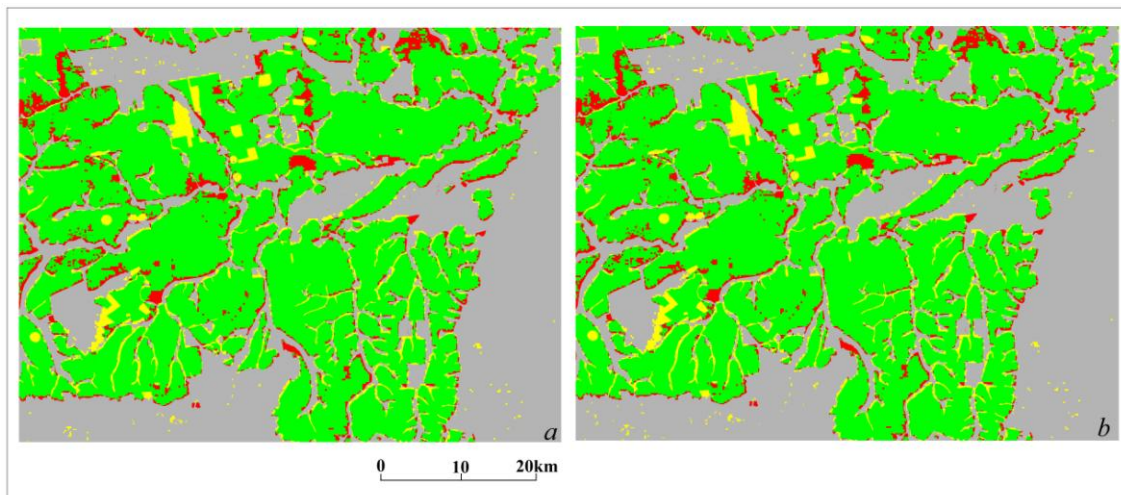


Figura 3.6 - Mapas de resultados originados da comparação entre os mapas gerados pela curva ROC para o conjunto de amostras utilizadas para treinamento e amostra e o mapa de soja de referência. O mapa de resultados *a* representa a comparação pixel-a-pixel realizada entre o mapa resultante da rede bayesiana para as amostras de treinamento do algoritmo comparado ao mapa de soja referencial. O mapa de resultados *b* apresenta a mesma comparação utilizando o mapa de soja para as amostras de teste. As áreas em verde são definidas como *acerto soja*, já que no mapa de referência e nos mapas originados a partir da rede bayesiana apresentam esta mesma classe temática. As porções em vermelho dos mapas de resultados são denominadas *omissão* representam as áreas que são definidas como soja no mapa de referência e não foram assim identificadas nos mapas gerados na rede bayesiana. As áreas em amarelo são denominadas *inclusão*, representam as áreas que não pertencem a classe soja no mapa de referência mas foram incluídas como soja nos mapas oriundos da rede bayesiana. As áreas dos mapas de resultados em cinza são denominadas *não soja*, representando as áreas que não pertencem a classe soja no mapa de referência e foram classificadas com tal por meio da rede bayesiana.

Por meio desta análise é possível perceber que os erros apresentam distribuição espacial variada e que estas em sua grande maioria estão associadas a porções contínuas

identificadas ou não como soja por meio da técnica de rede bayesiana (Figura 3.6). Uma porção importante dos erros de classificação está associada as borda dos talhões de soja definidos no mapa de referência, este fato é resultante da maior influência das imagens CEI para o estabelecimento das funções de probabilidade. Visto que, a resolução espacial das imagens MODIS utilizadas para a formação das imagens CEI é de 250 x 250m contra 30 x 30m no caso das imagens Landsat utilizadas na confecção do mapa de referência. De tal forma, os mapas produzidos através da RB apresentam regiões contínuas com classes de erros (*inclusão* ou *omissão*) nas bordas dos talhões que são frutos da divisão dos pixels de 250 x 250m em pixels de 30 x 30m por meio do interpolador vizinho mais próximo, gerando áreas com valores idênticos de índice CEI na imagem com resolução de 30 x 30m dentro da área de um pixel de tamanho original MODIS. Contudo, visto a necessidade de produzir geocampos de iguais dimensões (tanto em número de linhas e colunas, quanto em resolução) para a aplicação da RB no *software* R, foi necessária a aplicação desta reamostragem utilizando tal interpolador nas imagens CEI.

De tal forma, foi percebida a influência das imagens CEI no resultado final da RB, demonstrando que existe a necessidade de utilizar dados espectro-temporais com maior obtidos com maiores escalas (melhores resoluções espaciais) para aumentar a exatidão do processo de classificação da RB e conseqüentemente, reduzir os erros existentes nos mapas originados a partir da técnica bayesiana. Este fato pode ser percebido ao analisarmos a Figura 3.7.

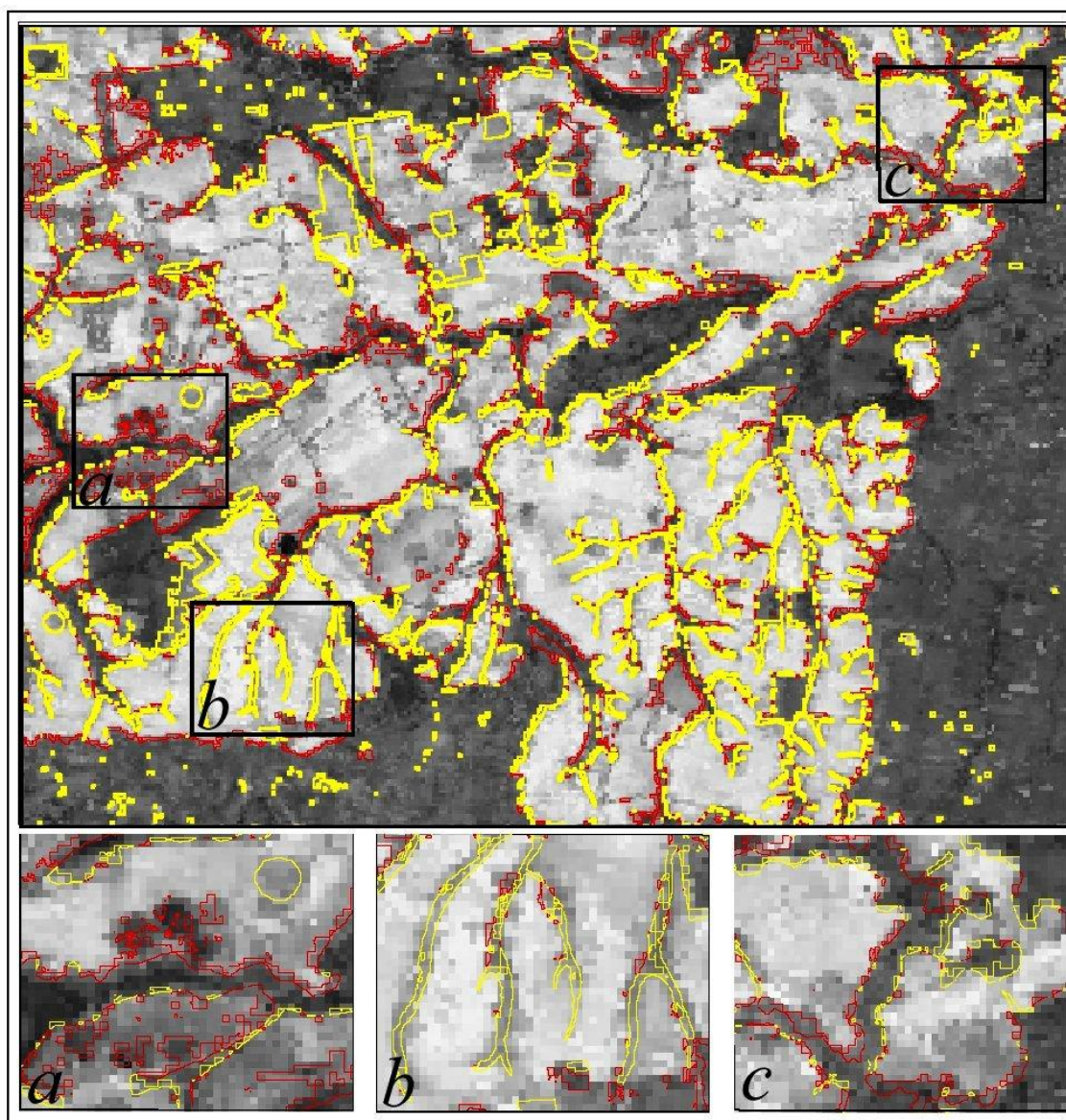


Figura 3.7 – Análise visual dos erros da classificação baseada na técnica bayesiana sobrepostos a imagem CEI 2011/2012. Vetor em amarelo representa os erros de inclusão, os vetores em vermelho representam os erros de omissão na classificação.

As porções contínuas dentro dos talhões nos mapas de resultados apresentam desacordo com o mapa de referência também são erros de *omissão* e *inclusão*. Estes desacordos se originam em especial dos valores das imagens CEI utilizadas no processo de treinamento da RB. Visto que as mesmas porções que apresentam erros de *inclusão* ou *omissão* se repetem para o mapa temático gerado a partir das amostras de treinamento e

de teste (Figura 3.7). Este fato pode estar ligado ao uso das três imagens CEI no processo de treinamento da rede bayesiana já que estas foram as variáveis que apresentaram maior peso na definição das funções de probabilidade utilizadas para a construção da imagem de probabilidade (Tabela 3.1).

Ademais, a partir dos mapas de resultados foram extraídos os valores das classes *soja* e *não soja*, *inclusão* e *omissão*, para que sobre os valores de classe fossem calculados os valores em percentuais de *acerto*, *inclusão* e *omissão* (Tabela 3.2). Esta análise espacial apresentou uma pequena diferença de acerto espacial quando comparados os mapas temáticos produzidos a partir das amostras de treinamento e de teste (tabela 3.2). Similarmente, os erros de *inclusão* e *omissão* apresentaram pequenas diferenças em termos percentuais quando comparados os mesmos erros para o mapa produzido a partir das amostras de treinamento e de teste.

Tabela 3.2 – Valores da análise espacial obtidos a partir dos mapas de resultados

| | Valores em % | |
|------------------------|--------------|-------------|
| | Teste | Treinamento |
| Acerto espacial | 89,53 | 89,49 |
| Erro Inclusão | 4,91 | 4,72 |
| Erro Omissão | 5,56 | 5,72 |

Outro ponto importante de avaliação para mapas temáticos que representam cultivos agrícolas está ligado ao valor de área cultivada da cultura contida nestes. Então, com o intuito de aferir a acurácia da estimativa de área da técnica bayesiana para este estudo, foram calculados os valores de área de soja dos mapas originados pela bayesiana e comparados ao valor de área de soja do mapa de referência. De tal forma, foi possível constatar que a técnica bayesiana gerou mapas contendo superestimativa de área tanto para aquele produzido a partir das amostras de treinamento como para aquele originado a partir das amostras de teste (Tabela 3.3). Contudo, mesmo demonstrando uma tendência de superestimativa a técnica apresentou valores de estimativa de área similares aos identificados no mapa de referência.

Tabela 3.3 – Valores de estimativa de área para o mapa temático de referência e os mapas temáticos de soja obtidos através da técnica bayesiana para as amostras de treinamento e teste

| | Mapa de referência | Mapa treinamento | Mapa Teste |
|--------------------------------|---------------------------|-------------------------|-------------------|
| Área de soja (há) | 236.703,69 | 232.886,52 | 234.025,74 |
| Diferença absoluta (há) | - | 3.817,17 | 2.677,95 |
| Diferença relativa (%) | - | 1,61 | 1,13 |

5 CONCLUSÕES

A rede bayesiana foi capaz de produzir uma imagem de probabilidade a partir de oito variáveis de contexto sobre a qual foram gerados mapas temáticos de elevada convergência para o mapa de referência. Estes mapas apresentaram valores de acerto espacial do mapeamento temático de 89,53 e 89,49% para as amostras de teste e treinamento, respectivamente. Quanto à estimativa de área a técnica apresentou elevada convergência em relação ao definido no mapa de referência, com diferenças de 1,61 e 1,13% para as amostras de treinamento e teste, respectivamente, demonstrando a tendência de superestimativa de área.

Por outro lado, a técnica bayesiana aplicada neste trabalho foi capaz de identificar a maioria das áreas de soja da região em estudo mesmo utilizando variáveis de contexto geradas em diferentes escalas geográficas. Contudo, as áreas de bordas da maioria dos talhões detiveram menor exatidão espacial na identificação quando comparados às áreas de centro dos mesmos. Dentre estas variáveis de contexto aquelas de maior contribuição foram às imagens CEI, de modo que, estas afetaram diretamente o resultado final da classificação. Por fim, a técnica bayesiana se apresenta com potencial para a detecção de áreas com cultivos agrícolas de modo semiautomático desde que exista um conjunto de variáveis de contexto para abastecer o seu treinamento.

6 AGRADECIMENTOS

Agradeço as contribuições e suporte do Doutorando Márcio Pupin Mello do INPE no desenvolvimento deste trabalho e por disponibilizar o algoritmo bayesiano fundamental a realização desta atividade.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ANA – Agência Nacional das Águas. Manual de Construção da Base Hidrográfica Ottocodificada: fase 1 – construção da base topológica de hidrografia e otobacias conforme a codificação de bacias hidrográficas de Otto Pfafstetter : versão 2.0 de 1/11/2007. Brasília, ANA, SGI, 144 p., 2007.

ASNER, G. P. Cloud cover in Landsat observations of the Brazilian Amazon. **International Journal of Remote Sensing**, vol. 22, no. 18, 3855–3862, 2001.

AYACHI, R.; AMOR, N. B.; BENFERHAT, S. A generic framework for a compilation-based inference in probabilistic and possibilistic networks. **Information Sciences**, article in press, 2013.

CÂMARA, G.; DAVIS, C. MONTEIRO, A. M. V. Introdução ao geoprocessamento **Introdução à Ciência da Geoinformação**. São José dos Campos: INPE, 2001 (INPE-8568-PRE/4312). Disponível em: < <http://www.dpi.inpe.br/gilberto/livro/introd>>. Acesso em: 10 mai. 2013.

CASTELLETTI, A.; SONCINI-SESSA, R. Bayesian networks in water resource modelling and management. Editorial. **Environmental Modelling & Software**, n.22, p.1073-1074, 2007.

CONAB – Companhia Nacional de Abastecimento. Acompanhamento de safra brasileira: grãos, décimo levantamento, julho 2011. Brasília: Conab, 2011. Disponível em:<http://www.conab.gov.br/OlalaCMS/uploads/arquivos/13_05_09_11_56_07_boletim_2_mai_2013.pdf>. Acesso: 30 mai. 2013.

COMBER, A.; FISHER, P.; BRUNSDON, C.; KHMAG, A. Spatial analysis of remote sensing image classification accuracy. **Remote Sensing of Environment**, n.127 p.237-246, 2012.

DIDAN, K.; HUETE, A. R. MODIS Vegetation Index product series Collection 5 change summary. The University of Arizona. Arizona. Jun., 2006. Disponível em: <http://landweb.nascom.nasa.gov/QA_WWW/forPage/MOD13_VI_C5_Changes_Document_06_28_06.pdf>. Acesso em: 30 mai. 2013.

DOGUC, O.; RAMIREZ-MARQUEZ, J. E. A generic method for estimating system reliability using Bayesian networks. **Reliability Engineering and System Safety**, n.94, p.542-550, 2009.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. Tecnologias de Produção de Soja Região Central do Brasil. Boletim eletrônico. 2003. Disponível em:

<<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Soja/SojaCentralBrasil2003/index.htm>>. Acesso em: 20 mai. 2013.

EBERHARDT, I. D. R. Mapa de soja elaborado por meio de fotointerpretação de imagens Landsat da metade Sul do município de Primavera do Leste-MT. Acervo pessoal do autor. 2012.

FARR, T. G.; ROSEN, P. A.; CARO, E.; CRIPPEN, R.; DUREN, R.; HENSLEY, S.; KOBRICK, M.; PALLER, M.; RODRIGUEZ, E.; ROTH, L.; SEAL, D.; SHAFFER, S.; SHIMADA, J.; UMLAND, J.; WERNER, M.; OSKIN, M.; BURBANK, D.; ALSDORF, D. The Shuttle Radar Topography Mission, **Rev. Geophys.**, 45, 2007. Doi:10.1029/2005RG000183.

HUETE, A.; LIU, H. Q.; BATCHILY, K.; LEEUWEN, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote Sensing of Environment**, n. 59, p. 440–451, 1997.

HUETE, A.; JUSTICE, C.; LEEUWEN, W. V. **Modis vegetation index (MOD 13) algorithm theoretical basis document**, 1999. Disponível em: <modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf>. Acesso em: 30 mai. 2013.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Pesquisas agropecuárias**, 2 ed. Rio de Janeiro: IBGE, 2002. 92 p. (Relatórios metodológicos, v. 6). Disponível em: <<http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/PesquisasAgropecuarias2002.pdf>>.

_____. **Levant. Sistem. Prod. Agríc.** Rio de Janeiro, v.26, n.4, p.1-86 abril de 2013. Disponível em: <http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/lspa_201304.pdf>. Acesso em: 30 mai. 2013.

JUSTICE, C. O.; VERMOTE, E.; TOWNSHEND, R. D.; DEFRIES, R.; ROY, D. P.; HALL, D. K.; SALOMONSON, V. V.; PRIVETTE, J. L.; RIGGS, G.; STRAHLER, A.; LUCHT, W.; MYNEMI, R. B.; KNYAZIKHIN, Y.; RUNNING, S. W.; NEMANI, R. R.; WAN, Z.; HUETE, A. R.; LEEUWEN, W. V.; WOLFE, R. E.; GIGLIO, L.; LEEUWEN, W. D.; HUETE, A. R.; LAING, T. W. MODIS Vegetation Index Compositing Approach: A prototype with AVHRR data. **Remote Sensing of Environment**, n. 69, p. 264–280, 1999.

LAMPINEN, J.; VEHTARI, A. Bayesian approach for neural networks - review and case studies. **Neural Networks**, n.14, p.257-274, 2001.

MULLER, J. P.; LEWIS, P.; BARNESLEY, M. J. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land Remote Sensing for Global Change Research. **IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING**, VOL. 36, NO. 4, p.1228-1249, JULY 1998.

MELLO, M. P.; RUDORFF, B. F. T.; ADAMI, M.; RIZZI, R.; AGUIAR, D. A.; GUSSO, A.; FONSECA, L. M. G. A simplified Bayesian network to map soybean plantations. Proceedings of the 2010 IEEE International Geoscience and Remote sensing Symposium (**IGARSS 2010**). Honolulu, HI, USA, IEEE, 2010, p.351-354.

- MELLO, M. P.; RUDORFF, B. F. T.; ADAMI, M.; AGUIAR, D. A. Na R implementation for Bayesian networks applied to spatial data. **Procedia environmental Sciences**, n.7, p.275-280, 2011.
- NASH, D.; WATERS, D.; BULDA, A.; WU, Y.; LIN, Y.; YANG, W.; SONG, Y.; SHU, J.; QIN, W.; HANNAH, M. Using a conceptual Bayesian network to investigate environmental management of vegetable production in the Lake Taihu region of China. **Environmental Modelling & Software**, p.1-12, 2013.
- PINO, F. A. Estimativa subjetiva de safras agrícolas. **Informações Econômicas**, SP, v.31, n. 6, jun., 2001.
- PINO, F. A. Estatísticas agrícolas para o século XXI. **Agric. São Paulo**, SP, 46(2): 71-105, 1999.
- QIN, D.; JIANWEN, M.; YUN, O. Y. Remote sensing data change detection based on CI test of Bayesian Networks. **Computers & Geosciences**, n.32, p.195-202, 2006.
- QU, Y.; WANG, J.; WAN, H.; LI, X.; ZHOU, G. A Bayesian network algorithm for retrieving the characterization of land surface vegetation. **Remote Sensing of Environment**, n.112, p.613-622, 2008.
- R Development Core Team, **R: A Language and Environment for Statistical Computing**, R Foundation for Statistical Computing, Vienna, Austria, 2013.
- RISSO, J.; RIZZI, R.; RUDORFF, B. F. T.; ADAMI, M.; SHIMABUKURO, Y. E.; FORMAGGIO, A. R.; EPIPHANIO, R. D. V. Índices de vegetação MODIS aplicados na discriminação de áreas de soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.47, n.9, p.1317-1326, set. 2012.
- RITTER, N. D.; HEPNER, G. F. Application of an artificial neural network to land-cover classification of thematic mapper imagery. **Computers & Geosciences**, v.16, n.6, p.873-880, 1990.
- RIZZI, R.; RUDORFF, B. F. T. Estimativa da área de soja no Rio Grande do Sul por meio de imagens Landsat. **Revista Brasileira de Cartografia**, v.57, p.226-234, 2005.
- RIZZI, R.; RISSO, J.; EPIPHANIO, R. D. V.; RUDORFF, B. F. T.; FORMAGGIO, A. R.; SHIMABUKURO, Y. E.; FERNANDES, S. L. Estimativa da área de soja no Mato Grosso por meio de imagens MODIS. XIV SBSR. **Anais**. Natal. INPE p.387-394, 2009. Disponível em: <<http://marte.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2008/11.16.18.50.57/doc/387-394.pdf>>. Acesso em: 10 out. 2012.
- ROCCHINI, D.; FOODY, G. M.; NAGENDRA, H.; RICOTTA, C.; ANAND, M.; HE, K. S.; AMICI, V.; KLEINSCHMIT, B.; FOSTER, M.; SCHMIDTLEIN, S.; FEILHAUER, H.; GHISLA, A.; METZ, M.; NETELER, M. Uncertainty in ecosystem mapping by remote sensing. **Computers & Geosciences**, n.50, p.128-135, 2013.
- RODRIGUEZ, E.; MORRIS, C. S.; BELZ, J.E.; CHAPIN, E.C.; MARTIN, J.M.; DAFFER, W.; HENSLEY, S. An assessment of the SRTM topographic

products. **Technical Report JPL D-31639**, Jet Propulsion Laboratory, Pasadena, California, 143pg, 2005.

RUDORFF, B. F. T.; AGUIAR, D. A.; SILVA, W. F.; SUGAWARA, L. M.; ADAMI, M.; MOREIRA, M. A. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. **Remote Sensing**, v. 2, n. 4, p. 1057-1076, 2010.

SIGEL – Sistema de Informações Georreferenciadas do Setor Elétrico. Agência Nacional de Energia Elétrica. 2013. Disponível em: <<http://sigel.aneel.gov.br/>>. Acesso em: 10 mai. 2013.

UUSITALO, L. Advantages and challenges of Bayesian networks in environmental modeling. **Ecological Modelling**, n.203, p.312-318, 2007.

USDA – United States Department of Agriculture. Foreign Agricultural Service. Oils seeds: World markets and trade. Circular series, FOP 45-13, May, 2013. Disponível em: <<http://usda01.library.cornell.edu/usda/current/oilseed-trade/oilseed-trade-05-10-2013.pdf>>. Acesso em: 30 mai. 2013.

VALERIANO, M. M.; ALBUQUERQUE, P. C. G. TOPODATA: Processamento dos dados SRTM. **Instituto Nacional de Pesquisas Espaciais**, São José dos Campos, SP. 2010. INPE-16702-RPQ/854. Disponível em: <<http://mtc-m19.sid.inpe.br/col/sid.inpe.br/mtc-m19%4080/2010/05.10.18.42/doc/publicacao.pdf>>. Acesso em: 14 mai. 2013.

VALERIANO, M. M.; ROSSETTI, D. F. TOPODATA: Seleção de coeficientes geoestatísticos para o refinamento unificado de dados SRTM. **Instituto Nacional de Pesquisas Espaciais**, São José dos Campos, SP. 2010. INPE-16701-RPQ/853. Disponível em: <<http://mtc-m19.sid.inpe.br/col/sid.inpe.br/mtc-m19%4080/2010/05.10.18.35/doc/publicacao.pdf>>. Acesso em: 14 mai. 2013.

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.