



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21b/2015/02.02.19.00-TDI

**INFERÊNCIA EM REDES BAYESIANAS PARA  
MAPEAMENTO DE ÁREAS DE CULTIVO E ÁREAS  
APTAS PARA EXPANSÃO DE CANA-DE-AÇÚCAR NO  
ESTADO DE SÃO PAULO**

Alexsandro Cândido de Oliveira Silva

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelos Drs. Leila Maria Garcia Fonseca, e Marcio Pupin Mello, aprovada em 27 de maio de 2015.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3HSF79B>>

INPE  
São José dos Campos  
2015

## **PUBLICADO POR:**

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

## **COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):**

### **Presidente:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

### **Membros:**

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Amauri Silva Montes - Coordenação Engenharia e Tecnologia Espaciais (ETE)

Dr. André de Castro Milone - Coordenação Ciências Espaciais e Atmosféricas  
(CEA)

Dr. Joaquim José Barroso de Castro - Centro de Tecnologias Espaciais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos  
(CPT)

Dr<sup>a</sup> Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

### **BIBLIOTECA DIGITAL:**

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Clayton Martins Pereira - Serviço de Informação e Documentação (SID)

### **REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:**

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação  
(SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

### **EDITORAÇÃO ELETRÔNICA:**

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21b/2015/02.02.19.00-TDI

**INFERÊNCIA EM REDES BAYESIANAS PARA  
MAPEAMENTO DE ÁREAS DE CULTIVO E ÁREAS  
APTAS PARA EXPANSÃO DE CANA-DE-AÇÚCAR NO  
ESTADO DE SÃO PAULO**

Alexsandro Cândido de Oliveira Silva

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, orientada pelos Drs. Leila Maria Garcia Fonseca, e Marcio Pupin Mello, aprovada em 27 de maio de 2015.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3HSF79B>>

INPE  
São José dos Campos  
2015

Dados Internacionais de Catalogação na Publicação (CIP)

---

Silva, Alexsandro Cândido de Oliveira Silva.  
Si38i Inferência em redes bayesianas para mapeamento de áreas de cultivo e áreas aptas para expansão de cana-de-açúcar no estado de São Paulo / Alexsandro Cândido de Oliveira Silva . – São José dos Campos : INPE, 2015.  
xx + 68 p. ; (sid.inpe.br/mtc-m21b/2015/02.02.19.00-TDI)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2015.

Orientadores : Drs. Leila Maria Garcia Fonseca, e Marcio Pupin Mello.

1. Inteligência artificial. 2. Redes bayesianas. 3. Inferência. 4. Planejamento territorial. I.Título.

CDU 004.832.32:528.8

---



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](#).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#).

Aprovado (a) pela Banca Examinadora  
em cumprimento ao requisito exigido para  
obtenção do Título de **Mestre** em  
**Sensoriamento Remoto**

Dra. Leila Maria Garcia Fonseca



\_\_\_\_\_  
**Presidente / Orientador(a) / INPE / SJCampos - SP**

Dr. Márcio Pupin de Mello



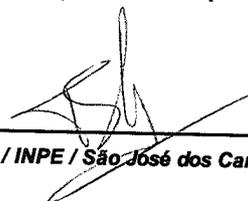
\_\_\_\_\_  
**Orientador(a) / Boeing Res & Techno / São José dos Campos - SP**

Dr. Antônio Roberto Formaggio



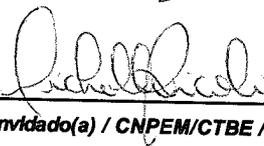
\_\_\_\_\_  
**Membro da Banca / INPE / SJCampos - SP**

Dr. Thales Sehn Korting



\_\_\_\_\_  
**Membro da Banca / INPE / São José dos Campos - SP**

Dra. Michelle Cristina Araujo Picoli



\_\_\_\_\_  
**Convidado(a) / CNPEM/CTBE / Campinas - SP**

**Este trabalho foi aprovado por:**

maioria simples

unanimidade

Aluno (a): **Alexsandro Cândido de Oliveira Silva**

**São José dos Campos, 23 de Fevereiro de 2015**



*“Failure is simply the opportunity to begin again.  
This time more intelligently.”  
Henry ford*



## **AGRADECIMENTOS**

A Deus, por dar-me paciência e coragem para enfrentar o caminho difícil de chegar até esta conquista.

Aos meus pais Ailton Cândido e Rosana de Oliveira e minha querida irmã Gabriella de Oliveira, muito obrigado pelo apoio e amor incondicional.

Aos mestres e doutores que passaram pela minha vida e sempre tinham algo a me ensinar. Em especial, meus orientadores Leila Fonseca e Marcio Mello, muito obrigado por me darem a oportunidade de trabalhar com vocês, muito obrigado pelas orientações, contribuições e pela amizade.

Estendo meus agradecimentos aos doutores Antônio Roberto Formaggio, Thales Sehn Korting e Michelle Cristina Araujo Picoli por terem aceito o convite. É uma honra tê-los como membros da banca de defesa

Aos amigos e amigas que fiz neste período no INPE. Aos amigos de longa data, em especial Fábio Henrique, muito obrigado por todo apoio e incentivo.

À unidade fomentadora desta pesquisa (134400/2013-5), o Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.



## RESUMO

O discurso sobre sustentabilidade tem ganhado projeção internacional com a necessidade de mitigação das emissões de gases do efeito estufa através da utilização de biocombustíveis. O Brasil se destaca por estar entre os maiores países produtores de etanol, com vasta experiência na produção do mesmo a partir da cana-de-açúcar. Devido ao aumento da demanda ao uso de biocombustíveis, até mesmo na área de transporte aéreo, há uma grande atenção voltada à ampliação sustentável do plantio da cana-de-açúcar. Portanto é fundamental que se faça um planejamento territorial que pode ser auxiliado por técnicas de Inteligência Artificial, amplamente utilizadas em Sensoriamento Remoto. Dentro deste contexto, este trabalho propõe mapear áreas de cultivo de cana-de-açúcar e áreas aptas ao seu plantio, no Estado de São Paulo, por meio de análises probabilísticas usando o modelo bayesiano *Enhanced Bayesian Network for Raster Data* – e-BayNeRD implementado no software R. O sistema e-BayNeRD é uma versão aprimorada do BayNeRD, cujas melhorias são descritas, com detalhes, neste trabalho. Os principais resultados do sistema e-BayNeRD são *layers* que mostram a distribuição espacial dos valores de probabilidade de ocorrência de cada classe do fenômeno sob estudo. O modelo de Redes Bayesianas (RB) proposto neste trabalho usou as seguintes variáveis: Zoneamento Agroambiental, Aptidão Edafoclimática, Declividade, *Crop Enhancement Index* (CEI), *Normalized Difference Vegetation Index* (NDVI) e ‘qualidade do solo’. Além disso, o mapa de áreas aptas à expansão da cana-de-açúcar, disponibilizados pela Embrapa Meio Ambiente, e o mapa áreas de cana-de-açúcar, produzido pelo projeto CANASAT, foram usados como referência para avaliar os resultados da classificação obtidos neste trabalho. Os resultados da avaliação mostraram que o modelo não foi capaz de inferir com precisão as áreas aptas à expansão da cana-de-açúcar. O principal problema foi quanto às limitações inerentes ao mapa de referência para as áreas aptas, que considerou áreas de preservação ambiental como área aptas à expansão, o que prejudicou a avaliação do mapa temático das áreas aptas geradas pelo modelo RB proposto. Por outro lado, o mapa de incertezas, calculado a partir dos valores de probabilidade das *layers*, indicou menores níveis de incerteza nas áreas de plantio de cana-de-açúcar. O sistema e-BayNeRD apresenta potencial para modelagem de sistemas complexos em diversas aplicações na área de observação da Terra



# **BAYESIAN NETWORK INFERENCE FOR MAPPING SUGARCANE AREAS AND SUITABLE AREAS FOR ITS EXPANSION IN SÃO PAULO STATE**

## **ABSTRACT**

Sustainability speech has gained international recognition with the need to mitigate greenhouse gas emissions using biofuels. Brazil stands out for being among the largest ethanol producing countries, with vast experience in the producing of it from sugarcane. With increasing demand for the use of biofuels, even in air transport area, there is an attention focused on sustainable expansion of sugarcane crops. Therefore, it is essential to make a territorial planning that can be aided by Artificial Intelligence techniques, widely used in Remote Sensing. Within this context, this work proposes to map sugarcane areas and suitable areas to cultivate it in the State of São Paulo, through probabilistic analysis using the e-BayNeRD (Enhanced Bayesian Network for Raster Data), a Bayesian model implemented in R software. The system e-BayNeRD is an enhanced version of BayNeRD, whose improvements are described in detail in this work. The main results of BayNeRD system are layers that show the spatial distribution of the probability values for each class of the phenomenon under study. The proposed Bayesian Network (BN) model in this work used the following variables: Agro-environmental Zoning, the edaphoclimatic aptitude, the Slope terrain, the Crop Enhancement Index (CEI), the Normalized Difference Vegetation Index (NDVI) and soil quality. In addition, the map of suitable areas suitable for the expansion of sugarcane, provided by Embrapa Environment, and the sugarcane map in 2006/07 harvest produced by CANASAT project, were used as reference maps to validate the classification results obtained in this work. The results showed that the model was not able to infer accurately the suitable areas for the expansion of sugarcane. The main problem was the inherent limitations in the reference map for suitable areas that considered areas of environmental preservation as areas suitable for the expansion, which hampered the assessment of suitable areas thematic map generated by the proposed RB model. On the other hand, the uncertainty map, computed from the probability values of the layers, indicated lower levels of uncertainty in the areas of sugarcane crops. The system e-BayNeRD shows potential for modeling complex systems in various applications in Earth Observation area.



## LISTA DE FIGURAS

Figura 2.1. – Matriz Energética Brasileira. ....	6
Figura 2.2 – Consumo final de energia de fontes de biomassa de 1970 a 2010.....	7
Figura 2.3 – Custos de produção de energia elétrica. ....	8
Figura 2.4 – Cadeia produtiva da cana-de-açúcar. ....	9
Figura 2.5 – Série histórica de produtividade.....	11
Figura 2.6 – Série histórica de área plantada de cana-de-açúcar.....	13
Figura 2.7 – Esquema conceitual do ZAE-Cana. ....	15
Figura 2.8 – Exemplo de modelo gráfico de uma Rede Bayesiana. ....	17
Figura 2.9 – Procedimentos usados na aplicação do BayNeRD. ....	19
Figura 3.1 – Diferentes possibilidades de intersecção entre as variáveis alvo e de contexto. ....	24
Figura 3.2 – Interface para definição do modelo gráfico da RB. ....	25
Figura 4.1 – Área de estudo: estado de São Paulo; dados de referência: áreas aptas à expansão, indicadas pelo Zoneamento Agroecológico da Cana-de-açúcar e áreas com cana-de-açúcar cultivada safra 2006/07, de acordo com o Projeto Canasat. ....	29
Figura 4.2 – Modelo gráfico da Rede Bayesiana.....	33
Figura 4.3 – Discretização da variável de contexto declividade (D). ....	34
Figura 4.4 – Dispersão das variáveis de contexto CEI (esquerda) e NDVI (direita) em função da variável alvo (ZC). ....	36
Figura 5.1 – BP1: probabilidade de ocorrência de áreas aptas à expansão. ....	38
Figura 5.2 – BP2: probabilidade de ocorrência de áreas de cana-de-açúcar. ....	38
Figura 5.3 – Região de alta probabilidade de ocorrência de áreas aptas à expansão da cana-de-açúcar. ....	40
Figura 5.4 – Região de alta probabilidade de ocorrência de áreas de cana-de-açúcar. ....	40
Figura 5.6 – Mapa temático criado a partir das Bandas de Probabilidade. ....	43
Figura 5.7 – Variação do índice Kappa e da acurácia geral. Simulação Monte Carlo para avaliação das incertezas da classificação (esquerda); Intervalo de Credibilidade (direita). ....	44
Figura 5.8 – Mapa de incertezas a partir da Entropia de Shannon. (A) Região com menores níveis de incerteza; (B) Região com maiores níveis de incerteza; ....	47
Figura 5.9 – Variação do índice Kappa e da acurácia geral da avaliação local. Simulação Monte Carlo para avaliação das incertezas da classificação (esquerda); Intervalo de Credibilidade (direita). ....	49

Figura 5.10 – Modelo gráfico da Rede Bayesiana para inferência de áreas de cana-de-açúcar e intervalos definidos no processo de discretização. *DA - Variável ‘distância do corpo d’água mais próximo’ (em Km). .....	50
Figura 5.11 – Dados de referência para o segundo modelo: áreas de cana-de-açúcar segundo Projeto Canasat. ....	50
Figura 5.12 – Banda de Probabilidade para a classe áreas de cana-de-açúcar, resultante do segundo modelo RB. ....	51
Figura 5.13 – Mapa Temático das áreas de cana-de-açúcar resultante do segundo modelo RB. ....	52
Figura 5.14 – Modelo gráfico da Rede Bayesiana para inferência de áreas aptas à expansão da cana-de-açúcar e intervalos definidos no processo de discretização. *DA - Variável ‘distância do corpo d’água mais próximo’ (em Km) .....	53
Figura 5.15 – Dados de referência para o terceiro modelo: áreas aptas ao plantio da cana-de-açúcar segundo o Zoneamento Agroecológico. ....	53
Figura 5.16 – Banda de Probabilidade para a classe áreas aptas, resultante do terceiro modelo RB.....	55
Figura 5.17 - Mapa Temático das áreas apta, resultante do terceiro modelo RB.....	55
Figura 5.18 – Recorte ampliado do Mapa Temático do primeiro modelo gerado e dos dados de Referência, onde a classe áreas de cana-de-açúcar apresenta os melhores resultados de classificação. ....	56
Figura 5.19 – Recorte ampliado do Mapa Temático do primeiro modelo gerado e dos dados de Referência, em uma região onde a classe áreas aptas apresentou altos níveis de incerteza.....	56

## LISTAS DE TABELAS

Tabela 4.1 – Variáveis de contexto .....	31
Tabela 4.2 – Probabilidade a priori $P(D = d)$ .....	35
Tabela 4.3 – Probabilidade condicional $P(ZC = zc D = d)$ .....	35
Tabela 4.4 – Discretização das variáveis de contexto .....	36
Tabela 5.1 – Divergência KL das variáveis de contexto. ....	41
Tabela 5.2 – Matriz de confusão com $\kappa_{\max}$ dentro do intervalo de credibilidade. ....	45
Tabela 5.3 – Matriz de confusão dentro do intervalo de credibilidade Kappa para avaliação local .....	49



## LISTA DE SIGLAS E ABREVIACOES

ANEEL	-	Agncia Nacional de Energia Eltrica
ANP	-	Agncia Nacional do Petrleo, Gs Natural e Bicom bustveis
APA	-	rea de Proteo Ambiental
BayNeRD	-	Bayesian Network for Raster Data
BEN	-	Balano Energtico Nacional
CONAB	-	Companhia Nacional de Abastecimento
DAG	-	Directed Acyclic Graph
e-BayNeRD	-	Enhanced Bayesian Network for Raster Data
EMBRAPA	-	Empresa Brasileira de Pesquisa Agropecuria
EPE	-	Empresa de Pesquisa Energtica
GHG	-	GreenHouse Gas
IA	-	Inteligncia Artificial
IBGE	-	Instituto Brasileiro de Geografia e Estatstica
MAPA	-	Ministrio da Agricultura, Pecuria e Abastecimento
MMA	-	Ministrio do Meio Ambiente
MME	-	Mnistrio de Minas e Energia
Prolcool	-	Programa Nacional do lcool
RB	-	Redes Bayesianas
UCPI	-	Unidade de Conservao e Proteo Integral
UDOP	-	Uno do Produtores de Biocombustveis
UNICA	-	Uno da Indstria de Cana-de-Acar
ZAE-Cana	-	Zoneamento Agroecolgico da Cana-de-acar



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
1.1.	Objetivos .....	3
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b> .....	<b>5</b>
2.1.	Matriz energética brasileira .....	5
2.2.	Biomassa sucroalcooleira .....	7
2.3.	Sustentabilidade do etanol brasileiro .....	10
2.4.	Zoneamento agroecológico da cana-de-açúcar .....	13
2.5.	Redes bayesianas .....	16
2.6.	Baynerd – algoritmo original .....	19
<b>3</b>	<b>ESTRUTURA DO MÉTODO DE INFERÊNCIA e-BayNeRD</b> .....	<b>23</b>
3.1.	Variável alvo .....	23
3.2.	Variáveis de contexto .....	24
3.3.	Modelo gráfico da RB .....	25
3.4.	Discretização e funções de probabilidade .....	26
3.5.	Imagem de probabilidade .....	27
<b>4</b>	<b>APLICAÇÃO DO e-BayNeRD PARA IDENTIFICAÇÃO DA CANA-DE-AÇUCAR E DE ÁREAS APTAS À SUA EXPANSÃO</b> .....	<b>29</b>
4.1.	Área de estudo .....	29
4.2.	Variáveis do modelo RB .....	30
4.3.	Modelo gráfico da RB para a aplicação proposta .....	33
4.4.	Discretização das variáveis .....	34
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>37</b>
5.1.	Bandas de probabilidade .....	37
5.2.	Mapa temático .....	42
5.3.	Avaliação das incertezas .....	43
<b>6</b>	<b>CONCLUSÃO E TRABALHOS FUTUROS</b> .....	<b>57</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>61</b>



## 1 INTRODUÇÃO

Nas últimas décadas tem-se verificado intenso e crescente desenvolvimento, industrialização e aumento demográfico, acarretando em uma ascendente demanda global por energia. Concomitantemente, parece haver uma vertente voltada à conservação dos recursos naturais que se preocupa em suprir este consumo por meio de fontes energéticas alternativas. Na busca da diversificação e modernização da matriz energética, o uso dos biocombustíveis é uma opção factível para aumentar a disponibilidade de energia e ajudar na mitigação das emissões de gases do efeito estufa (GHG, do inglês *GreenHouse Gas*) (RUDORFF et al., 2010). Em especial nos meios de transporte, os biocombustíveis têm provado sua viabilidade, haja visto o uso do etanol brasileiro fabricado a partir da cana-de-açúcar. Até mesmo a indústria aeroespacial já estabeleceu metas arrojadas para redução da emissões de GHG e, enxerga nos biocombustíveis, o principal caminho para o cumprimento dessas metas. (MORAES et al., 2014).

De acordo com Tilman et al. (2009), a procura por biocombustíveis deve-se concentrar em matérias-primas de biomassa sustentável que não competem com a produção de alimentos e nem causem desmatamentos, direta ou indiretamente. A cana-de-açúcar é uma potencial fonte de biomassa na produção de energia. Estudos atestam que a expansão da cultura tem-se dado em grande parte sobre pastagens e com uma mínima expressividade sobre áreas de florestas (RUDORFF et al., 2010; ADAMI et al., 2012). Além disso, no Estado de São Paulo (maior produtor de cana-de-açúcar) a expansão das áreas deste cultivo parece não ter provocado impacto nas áreas de produção de alimentos (GOLDEMBERG et al., 2008).

A expansão da cana-de-açúcar se deu com a implantação do Programa Nacional do Álcool – Proálcool, em 1975. Mais recentemente, a produção de etanol cresceu em larga escala com a implantação dos veículos *flexfuel* no mercado brasileiro em 2003. Hoje o Brasil tem destaque internacional na utilização de biomassa como recurso energético, sendo o maior produtor mundial de cana-de-açúcar, além de possuir uma vasta experiência na produção e uso do etanol. Por outro lado, o país ainda esbarra nas

barreiras comerciais e tarifárias impostas pelas certificações propostas por países importadores do etanol (ANEEL, 2008; RUDORFF et al., 2010).

A expansão sustentável das áreas plantadas com cana-de-açúcar deve observar alguns critérios como (i) não substituir áreas designadas para produção de alimentos e conservação da biodiversidade, conforme estabelece a Diretiva Europeia sobre energias renováveis (PARLAMENTO EUROPEU E DO CONSELHO, 2009) e (ii) eliminar a prática da queima da palha antes da colheita, no caso do estado de São Paulo, conforme dispõe a Lei Estadual N° 11241 de 19/09/2002 (SÃO PAULO, 2002). Com este propósito o Governo Federal determinou a elaboração do Zoneamento Agroecológico da Cana-de-açúcar (MANZATTO et al., 2009), o qual considerou a análise de variáveis ambientais, econômicas e sociais para estabelecer áreas apropriadas ou não para a expansão da cana-de-açúcar no país.

Descrever as relações entre estas variáveis para planejamento territorial requer o emprego de diferentes ferramentas como, por exemplo, técnicas de Inteligência Artificial (IA), amplamente utilizadas nas análises de produtos de Sensoriamento Remoto. No entanto, algumas aplicações em IA têm-se deparado com problemas ao operar com dados incompletos ou informações incertas. Algumas técnicas, como as Redes Bayesianas (RB), são capazes de lidar com essas e uma série de outras limitações em relação ao uso de dados espaciais. Matematicamente, uma RB é uma representação de uma tabela de probabilidades conjuntas do universo do problema. Mas, do ponto de vista do especialista, a RB constitui um modelo gráfico que representa as relações de (in)dependência entre as variáveis do sistema (NEAPOLITAN, 2004; PEARL; RUSSELL, 2011).

De fato, as RBs têm sido usadas em diferentes aplicações, tais como: diagnóstico médico (KAHN et al., 1997; ONÍŠKO; DRUZDZEL, 2013) e reconhecimento de fala e gestos (DAOUDI et al., 2003; SUK et al., 2010). Segundo Aguilera et al. (2011), as RBs são raramente usadas na Ciência de Observação da Terra e no processamento de dados de sensoriamento remoto. Neste contexto, Mello et al. (2013) desenvolveram um método bayesiano assistido por computador capaz de incorporar o conhecimento do

especialista para aplicações com uso de dados raster: *Bayesian Network for Raster Data* – BayNeRD. Mello et al. (2013) aplicaram o sistema BayNeRD para mapear soja no Estado do Mato Grosso e alcançaram bons resultados. Embora os principais conceitos de RB para aplicações em sensoriamento remoto tenham sido implementados por Mello et al. (2013), o algoritmo BayNeRD ainda apresenta possibilidade de aprimoramento.

De fato, a hipótese deste trabalho é de que o algoritmo BayNeRD (Mello et al., 2013) pode ser aperfeiçoado e apresenta potencial para aplicação na modelagem de sistemas complexos.

### **1.1. Objetivo**

O objetivo do presente trabalho consistiu em aprimorar o algoritmo BayNeRD (MELLO et al., 2013) e avaliar a versão aprimorada, nomeada como *Enhanced Bayesian Network for Raster Data* (e-BayNeRD), para identificar: (i) áreas de cultivo de cana-de-açúcar e (ii) áreas potenciais à expansão desta cultura no estado de São Paulo.

Para que o objetivo deste trabalho seja alcançado, foram estabelecidos objetivos de implementar um conjunto de melhorias, conforme expostas a seguir:

- Otimização da manipulação de dados *raster* no algoritmo, a fim de reduzir o pré-processamento dos dados;
- A possibilidade da definição de mais classes para os fenômenos sob estudo, visto que a complexidade de tais fenômenos não se restringe a presença e ausência;
- Um método automático e otimizado para definir as relações de (in)dependência entre as variáveis, de modo que o usuário possa fazê-la de maneira mais amigável;
- Um método para discretização e definição de funções de probabilidade, dada a possibilidade de a variável sob estudo poder ser do tipo contínua.
- Um método para calcular os principais resultados do e-BayNeRD, já que haverá a possibilidade de o fenômeno estudado ser descrito por mais de duas classes.



## **2 REVISÃO BIBLIOGRÁFICA**

Devido à rápida expansão da cana-de-açúcar, é importante discutir a questão da sustentabilidade da produção do etanol no Brasil, além das medidas governamentais adotadas para o desenvolvimento sustentável como o Zoneamento Agroecológico. Neste capítulo são discutidos, de forma resumida, as questões da matriz energética brasileira, a importância da biomassa sucroalcooleira como fonte de energia e a sustentabilidade do etanol. Finalmente, os conceitos de Redes Bayesianas, tema central da metodologia proposta neste trabalho, são também descritos aqui.

### **2.1. Matriz energética brasileira**

A preocupação com a segurança energética brasileira remete à década de 1970 quando houve a primeira Crise do Petróleo com um drástico aumento nos preços deste insumo (KOHLHEPP, 2010; SEGURA, 2012). Este fato desestabilizou a balança comercial brasileira visto que o petróleo era o principal produto nas importações do país naquele período (ANP, 2007; RAELE et al., 2014)

A reação do governo brasileiro foi desencadear ações e programas que reduzissem as importações de petróleo, extraindo-o em águas profundas como também a dependência do mesmo através da intensificação na construção de hidrelétricas. Outra ação importante foi a criação do Programa Nacional do Álcool – Proálcool, em 1975, para produção de etanol a partir da cana-de-açúcar (ANP, 2007; GOLDEMBERG et al., 2008; BORDONAL et al., 2013)

A partir de 2006, o Brasil alcançou a autossuficiência no fornecimento de petróleo pesado (aqueles de densidade leve ainda são importados) (SOARES; ANDREOZZI, 2011). A descoberta de grandes quantidades de petróleo e gás natural (pré-sal) entre o litoral do Espírito Santo e de São Paulo aumentou significativamente o volume das reservas do país (ANEEL, 2008; ARAUJO et al., 2013; KOHLHEPP, 2010). Tal descoberta teve importância estratégica para o Brasil no mercado internacional, já que o petróleo e seus derivados ainda são a principal fonte primária da matriz energética mundial e insumos para uma gama do setor industrial (ANEEL, 2008).

O consumo energético é um importante indicador do desenvolvimento econômico e da qualidade de vida de uma sociedade (ANEEL, 2008; OZTURK et al., 2010). A fim de aumentar a segurança do abastecimento, o país procura pela diversificação de sua matriz de energia, além de procurar torná-la mais “limpa” com o foco nas questões ambientais.

A Matriz Energética Brasileira é um exemplo em sustentabilidade (ARAUJO et al., 2013). As fontes primárias são altamente diversificadas com grande proporção de fontes renováveis, das quais um terço são hidroelétricas e dois terços são referentes à biomassa (HOLLANDA; ERBER, 2010; WALTER; DOLZAN, 2012). A Figura 2.1 apresenta a repartição da oferta interna de energia no país.



Figura 2.1. – Matriz Energética Brasileira.

Fonte: adaptada de EPE (2014)

De acordo com o último Balanço Energético Nacional (BEN) (EPE, 2014), a participação de fontes renováveis na Matriz Energética Brasileira se manteve entre as mais elevadas no mundo. No entanto, as condições hidrológicas desfavoráveis nos últimos anos causaram uma redução em 2013 comparado a 2012 e, adicionalmente, houve um aumento da geração de energia termoelétrica.

No Brasil houve um processo de transição do uso da biomassa na geração de energia. Iniciando-se com o uso doméstico conhecido como “biomassa convencional” e indo à “biomassa moderna”, que é utilizada na geração de energia e calor nas indústrias

(LORA; ANDRADE, 2009). A Figura 2.2 apresenta a contribuição da biomassa no consumo energético do Brasil. Pode-se observar a redução do consumo de madeira e a participação cada vez mais intensa do bagaço da cana-de-açúcar. Este aumento corresponde ao uso do bagaço como combustível para gerar energia térmica e elétrica nas usinas e para produção do etanol de segunda geração. (LORA; ANDRADE, 2009; WALTER; DOLZAN, 2012; VAKKILAINEN et al., 2013). Segundo Hollanda (2013), as formas de energia produzidas a partir da cana-de-açúcar são competitivas e têm perspectiva de crescimento, portanto, este setor possui amplas possibilidades para reduzir os custos de produção, assim como para intensificar sua importância na Matriz Energética Brasileira.

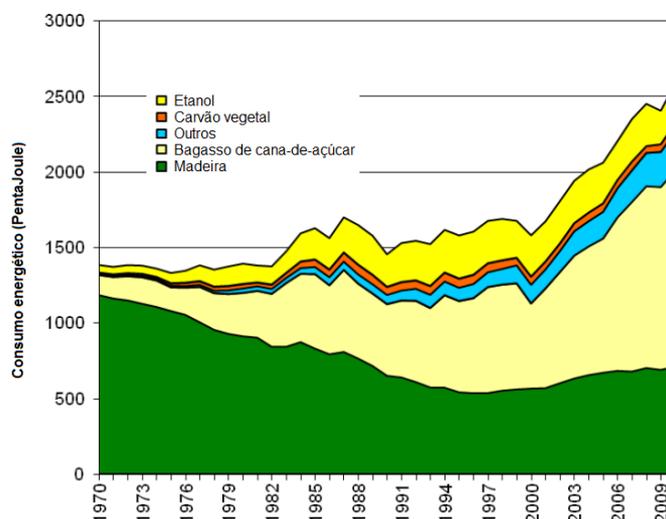
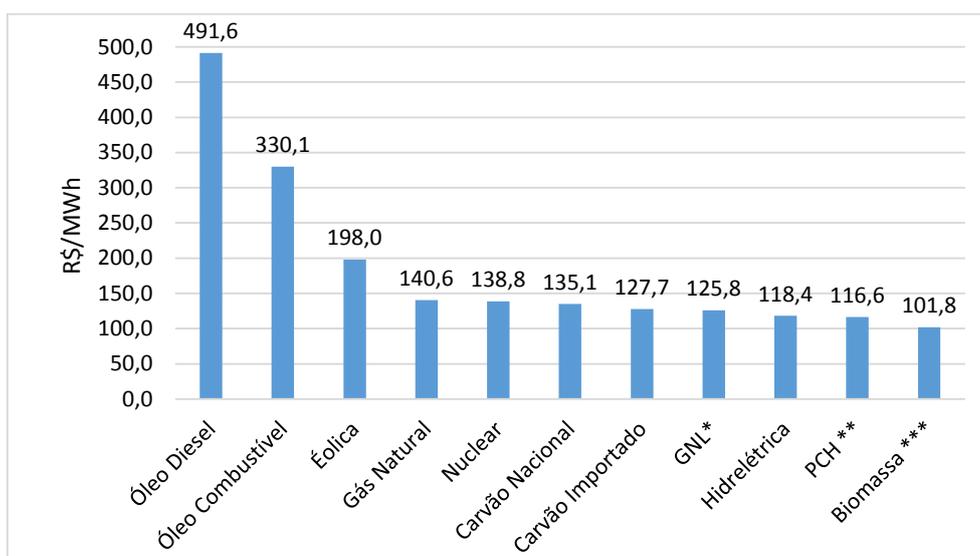


Figura 2.2 – Consumo final de energia de fontes de biomassa de 1970 a 2010.

Fonte: adaptada de Walter e Dolzan (2012)

## 2.2. Biomassa sucroalcooleira

O termo “biomassa” compreende todo recurso renovável provindo de matéria orgânica (animal ou vegetal, resíduos industriais ou urbanos) com a finalidade de produzir energia (MME, 2007; ANEEL, 2008; MMA, 2014). Conforme já mencionado, esta é uma das fontes com maior potencial de crescimento na matriz energética. Um dos motivos é o baixo custo de produção de energia, principalmente a partir do bagaço da cana-de-açúcar (Figura 2.3).



(\* Gás natural liquefeito; (\*\*) Pequenas centrais hidrelétricas; (\*\*\*) Bagaço de cana

Figura 2.3 – Custos de produção de energia elétrica.

Fonte: adaptada de ANEEL (2008).

Tais possibilidades de crescimento são ainda maiores no que se refere à cadeia energética de transformações e uso da cana-de-açúcar. A diferença entre a energia potencial da cana no campo e aquela efetivamente disponibilizada através do etanol, vapor e da energia elétrica deve-se ao aproveitamento limitado da energia da biomassa seca, o bagaço e a palha (ARAUJO et al., 2013; HOLLANDA, 2013).

A cadeia produtiva da cana-de-açúcar, apresentada na Figura 2.4, tem como principais produtos o açúcar e o etanol. A produção do açúcar é exclusivamente através da extração do caldo da cana-de-açúcar, enquanto que a produção de etanol pode ser diferenciada pela matéria-prima. Aquele produzido também pela extração do caldo é conhecido etanol de primeira geração e a produção através biomassa recuperada como etanol de segunda geração (DIAS et al., 2012; RAELE et al., 2014).

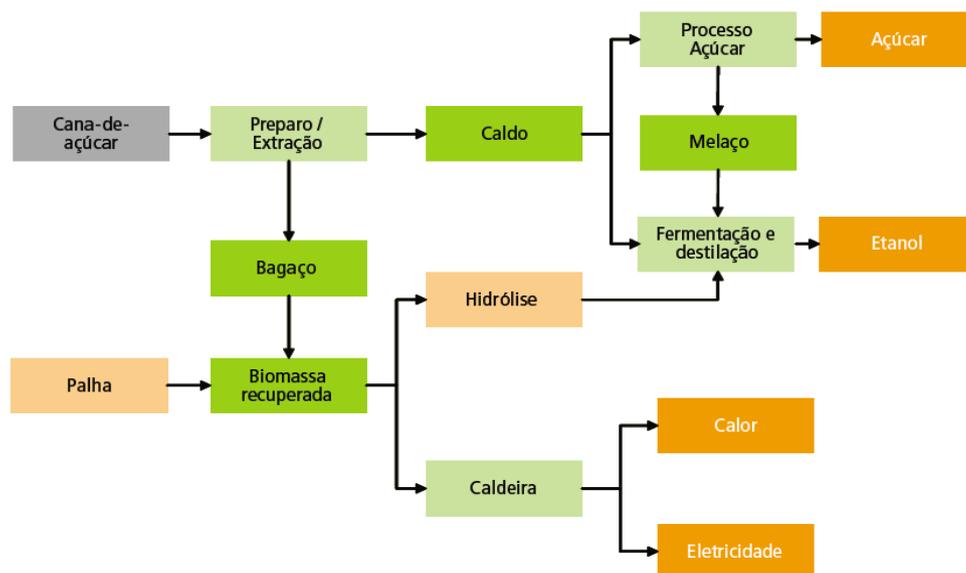


Figura 2.4 – Cadeia produtiva da cana-de-açúcar.

Fonte: MME (2007).

A produção do etanol através da hidrólise (etanol lignocelulósico) (RAELE et al., 2014) mostra-se como uma das alternativas com maior potencial para aumentar a produção total. O processo não depende da expansão da área de plantio da cana-de-açúcar e tão pouco compete com a produção de alimentos (DIAS et al., 2012; ARAUJO et al., 2013). Assim, o uso dos resíduos e subprodutos da cana-de-açúcar beneficia os aspectos ambientais e visa ao aumento de produtividade (ANEEL, 2008).

Segundo Manzatto et al. (2010), no Brasil, apenas um terço da cana-de-açúcar na produção do etanol é aproveitado. Os dois terços remanescentes referem-se ao bagaço produzido nas usinas e à palha colhida no campo, cuja real aplicação é servirem de combustível às caldeiras. Através desta matéria-prima é produzido todo o vapor e a eletricidade necessários à própria usina, que ainda pode vender a eletricidade excedente (HOLLANDA; ERBER, 2010; DIAS et al., 2012; WALTER; DOLZAN, 2012).

Vale ressaltar que uma oferta adicional de palha com a mecanização da colheita e a eliminação gradual das queimadas (DIAS et al., 2012) tem sido criada. Teoricamente, isto permitiria aumentar a oferta de biomassa seca. Porém, uma parte deste material deve permanecer na lavoura para cobertura do solo e atuando como adubo orgânico para

redução de custos de produção e danos ambientais (MME, 2007; MANZATTO et al., 2010).

Outra prerrogativa da cana-de-açúcar é que a safra desta cultura coincide com o período de estiagem na região Centro-Sul, principal produtora de cana-de-açúcar e onde se concentra a maior potência de hidrelétricas do país. Essa sazonalidade complementar (HOLLANDA, 2013) pode ajudar a preservar os níveis dos reservatórios das usinas hidrelétricas. No entanto, apesar do retorno econômico, do baixo custo de geração e da existência de linhas de crédito, o uso energético da biomassa seca continua inferior ao seu efetivo potencial (HOLLANDA, 2013).

### **2.3. Sustentabilidade do etanol brasileiro**

Em 2003, os veículos *flexfuel* entraram no mercado brasileiro, o que estimulou ainda mais o plantio da cana-de-açúcar para produção de etanol. A rápida expansão da cultura (RUDORFF et al., 2010) trouxe à tona questões sobre a sustentabilidade do etanol (GOLDEMBERG et al., 2008; BORDONAL et al., 2013).

O Brasil pode se tornar o maior exportador de etanol se quebrar as barreiras impostas pela certificação socioambiental sustentável proposta pelos países importadores. Para uma produção do etanol certificada, alguns critérios devem ser seguidos: (i) não substituir áreas designadas à produção alimentícia com cana-de-açúcar; (ii) conservar a biodiversidade; e (iii) eliminar a prática da queima da palha na pré-colheita. Segundo a avaliação de Smeets et al. (2008), os critérios (i) e (ii) são os maiores gargalos para produção sustentável e certificada.

Como consequência direta da rápida expansão, o cultivo da cana-de-açúcar substituiu áreas de pastagem e cultivos agrícolas (RUDORFF et al., 2010). Adami et al. (2012) avaliaram a expansão da área plantada de cana-de-açúcar entre 2005 e 2010 na região Centro-Sul do Brasil, responsável por quase 90% da produção nacional de cana-de-açúcar. Os autores concluíram que quase 70% das áreas de cana-de-açúcar substituíram áreas de pastagem, 25% áreas de culturas anuais, e 0,6% das áreas de expansão antes eram florestas.

Segundo Adami et al. (2012), apesar da redução das áreas de pastagem, houve aumento do número de cabeças de gado por hectare. De modo similar, a produção de grãos no Brasil vem continuamente crescendo. Kohlhepp (2010) também discute que a expansão da cana-de-açúcar no estado de São Paulo (principal estado produtor) se deu devido ao recuo das áreas de pastagem e a redução das áreas de gêneros alimentícios foi compensada pela alta produtividade, observada não apenas no estado de São Paulo, mas em todo o país. A série histórica, ilustrada na Figura 2.5, mostra que a produtividade brasileira de arroz, feijão, milho e trigo cresceu ao longo das safras 1976/77 a 2012/13.

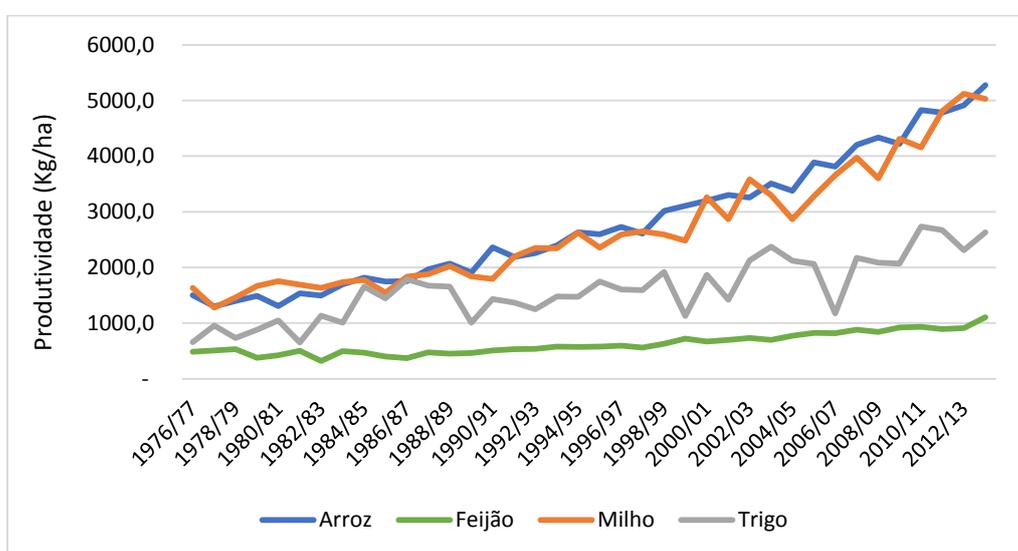


Figura 2.5 – Série histórica de produtividade.

Fonte: adaptada de CONAB (2014).

A produção de etanol a partir da cana-de-açúcar tem grande potencial na mitigação das emissões dos gases de efeito estufa (GHG). Comparado ao etanol produzido a partir de outras matérias primas, como o milho, beterraba e outros, o etanol da cana-de-açúcar possui uma alta eficiência (ADAMI et al., 2012; BORDONAL et al., 2013).

Muitos estudos têm quantificado o balanço das emissões de GHG do etanol brasileiro. Tais estudos indicam que a expansão da cana-de-açúcar e a produção do etanol levam a uma redução das emissões de GHG, mesmo quando se considera no balanço as perdas de carbono do solo pelas mudanças no uso da terra (FIGUEIREDO; SCALA JR, 2011; BORDONAL et al., 2013; EGESKOG et al., 2014).

Apesar dos progressos significativos em direção à prática da colheita mecanizada, a prévia queima da cana-de-açúcar ainda é praticada no Brasil e no Estado de São Paulo para tornar mais viável a colheita manual (AGUIAR et al., 2011; WALTER; DOLZAN, 2012). Em relação ao estado de São Paulo, a Lei Estadual Nº 11241 de 19/09/2002 (SÃO PAULO, 2002) dispõe pela eliminação da prática de queima em áreas aptas até 2021. Mas, o Programa Etanol Verde antecipou o fim desta prática, de 2021 para 2014 em áreas mecanizáveis e de 2031 para 2017 para áreas não mecanizáveis (AGUIAR et al., 2011; EGESKOG et al., 2014).

Outro ponto a ser discutido é a expansão da cana-de-açúcar nas áreas originalmente ocupados pelo cerrado. De acordo com a União da Indústria de Cana-de-Açúcar – UNICA (2007), aparentemente, ocorreu a substituição de coberturas que já haviam substituído o cerrado (pastagens). Ainda segundo a UNICA, a área total adicionada à cana-de-açúcar entre 1993 e 2003 nos estados com extensas regiões de cerrado (Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais) atingiu certa de 300 mil hectares. Enquanto que a expansão da cultura no Centro-Sul foi de 1,4 Milhões de hectares no mesmo período.

Mesmo com intensa atividade agrícola, o Brasil dispõe de terra agricultável (fronteira agrícola), sem prejudicar áreas de florestas e preservação. Segundo o Ministério de Minas e Energia (2007), a disponibilidade de terras adicionais ao cultivo é da ordem de 90 milhões de hectares, aproximadamente 11% do território brasileiro.

De acordo com Manzatto et al. (2010), o Brasil possui cerca de 64 milhões de hectares de áreas aptas à expansão do cultivo da cana-de-açúcar. Já o Ministério da Agricultura, Pecuária e Abastecimento – MAPA (2011) faz previsões de uma expansão da cultura de 2,1 milhões de hectares para os próximos anos, sendo os maiores aumentos de produção e área, 40,5% e 41,3%, respectivamente, no estado de Goiás.

Tais estimativas indicam que o país dispõe de áreas suficientes para atender às demandas futuras por açúcar e etanol sem afetar diretamente as terras utilizadas para produção de alimentos. Ainda há a possibilidade de garantir a segurança alimentar com a recuperação de pastagens degradadas, estimadas em mais de 80 milhões de hectares

somente na área tropical do Brasil. Com isso, não há necessidade de conflitos de interesse na produção de biocombustível com agricultura.

#### 2.4. Zoneamento agroecológico da cana-de-açúcar

O setor sucroalcooleiro do Brasil tem passado por um forte processo de expansão. A nova geopolítica energética internacional o impulsiona, já que objetiva reduzir a dependência dos combustíveis fósseis mediante à incorporação do etanol à matriz energética (SOUZA, 2010). Entre a safra 2005/06 e a previsão da safra 2013/14, a área de plantio da cana-de-açúcar aumentaria em torno de 50%, sendo a região Centro-Sul a maior responsável, como mostra o gráfico na Figura 2.6.

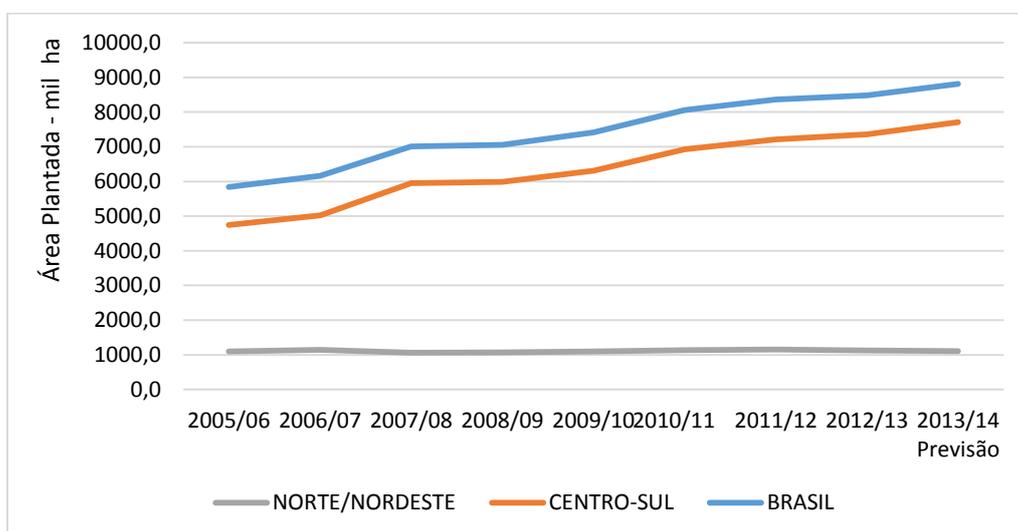


Figura 2.6 – Série histórica de área plantada de cana-de-açúcar.

Fonte: adaptado de CONAB (2014).

Com este cenário de crescimento, surge o desafio de planejar simultaneamente os impactos territoriais. A única interferência direta do governo no processo de autorização de novas usinas se dava pelo cumprimento da legislação ambiental brasileira, que, até então, não se mostrava suficiente para assegurar uma expansão harmônica do setor (MANZATTO et al., 2010).

Em 2009, o Governo Federal lançou o Zoneamento Agroecológico da Cana-de-Açúcar – ZAE-Cana. Tendo-o como instrumento de tomada de decisões, o governo visava

fornecer “subsídios técnicos para indicar as melhores áreas destinadas à expansão da cana-de-açúcar, com base nos fatores físicos, bióticos, socioeconômicos e jurídico-institucionais do território nacional” (MANZATTO et al., 2010, p. 194).

No relatório produzido com os resultados do ZAE-Cana consta que o mesmo

“[...] surgiu da necessidade estratégica de se avaliar, indicar e espacializar o potencial das terras para a extensão da produção da cultura da cana-de-açúcar em regime sequeiro (sem irrigação plena) para a produção de etanol e açúcar com base para o planejamento do uso sustentável das terras, em harmonia com a biodiversidade” (MANZATTO et al., 2009, p. 8).

O estudo seguiu as seguintes diretrizes (MANZATTO et al., 2009, 2010):

- Indicação de áreas com potencial agrícola para o cultivo da cana-de-açúcar sem restrições ambientais;
- Exclusão de áreas com vegetação original e indicação de áreas sob uso antrópico;
- Exclusão de áreas para cultivos nos biomas Amazônia, Pantanal e Bacia do Alto Paraguai;
- Diminuição da competição direta com áreas de produção de alimentos;
- Indicação de áreas com potencial agrícola (solo e clima) para cultivo da cana em terras com declividade inferior a 12%, propiciando a produção ambientalmente adequada para colheita mecânica;
- Unidades industriais já instaladas, a produção de cana para seu suprimento e a expansão programada não foram objeto deste zoneamento.

A iniciativa governamental inédita no ordenamento da expansão de atividades agrícolas foi além de um tradicional zoneamento agrícola. A metodologia desenvolvida para o ZAE-Cana, apresentada na Figura 2.7, abrangeu um estudo mais completo do território, inserindo além das cartas de solos e dados climáticos, a análise integrada de uso e

cobertura vegetal, relevo, hidrografia e áreas de preservação (MANZATTO et al., 2009, 2010). Segundo Manzatto et al. (2010), a metodologia foi desenvolvida em quatro fases:

- Fase 1 – determinação da aptidão das terras para a cana, um dos principais componentes do Zoneamento, mostrando qual é a aptidão natural dos solos para a cultura;
- Fase 2 – determinação da aptidão climática, outro componente importante que relaciona as características meteorológicas das diferentes regiões para a cultura;
- Fase 3 – obtenção de uma máscara para excluir a Amazônia, o Pantanal e a Bacia do Alto do Paraguai, as áreas de proteção legal (unidades de conservação, terras indígenas e outras), os mapas de uso da terra dos quais desconsideraram-se as áreas agrícolas e os mapas de declividade da terra que tiveram um declive inferior a 12%;
- Fase 4 – integração temática que foi feita com os resultados das aptidões das terras e clima, sendo mapeadas as áreas resultantes da máscara de zoneamento obtida na fase anterior.

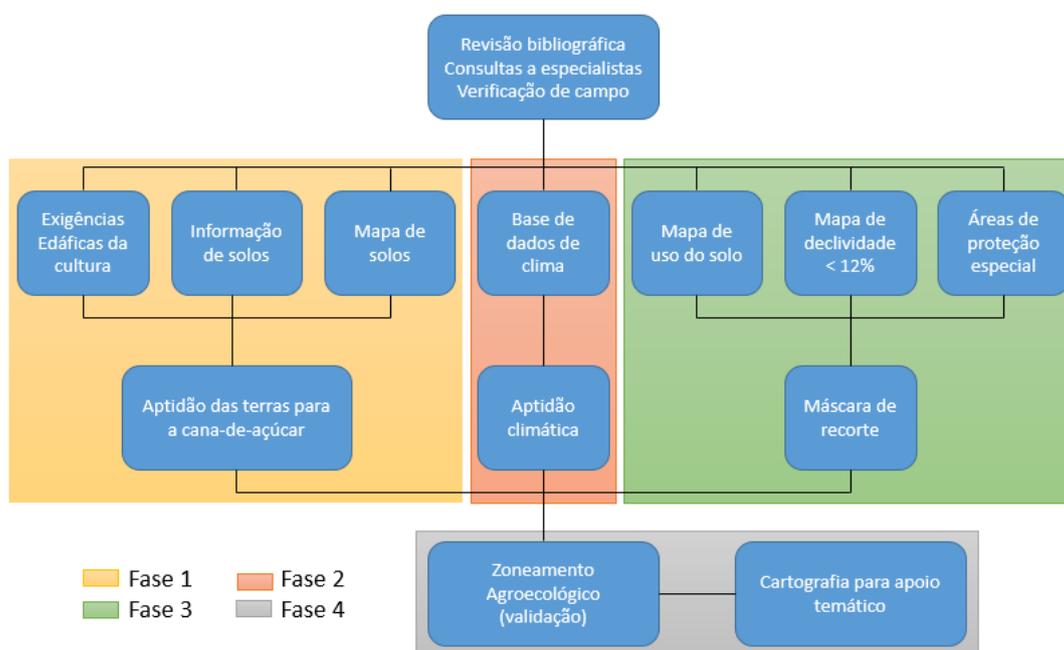


Figura 2.7 – Esquema conceitual do ZAE-Cana.

Fonte: adaptado de Manzatto et al. (2010).

## 2.5. Redes bayesianas

Culturas energéticas podem ocorrer em qualquer área agricultável. Mas torna-se indispensável um efetivo planejamento territorial a fim de se alcançar a produção e o desenvolvimento sustentável (MIURA, 2012). As tecnologias de sensoriamento remoto, combinadas com as técnicas de processamento de imagens e reconhecimento de padrões, fornecem informações importantes sobre alvos terrestres, além de dar uma ampla visão do espaço geográfico considerado. Seus produtos são base para estudos como agricultura de precisão (SEELAN et al., 2003), previsão de safra (BOLTON; FRIEDL, 2013), monitoramento das mudanças no uso e cobertura da terra, como a expansão da cana-de-açúcar (RUDORFF et al., 2010; ADAMI et al., 2012).

Técnicas de processamento de imagens e reconhecimento de padrões têm sido amplamente utilizadas nas fases de processamento e análise de produtos de sensoriamento remoto, sendo a classificação de imagens a aplicação mais comum. Entre os classificadores não-paramétricos mais comumente usados estão as Redes Neurais Artificiais, Árvores de decisão, Máquinas de Suporte a Vetor e Sistemas Especialistas (MELGANI; BRUZZONE, 2004; MAS; FLORES, 2008; WENG, 2012). Em algumas aplicações, o processo de análise de imagens pode se deparar com problemas ao operar com dados incompletos ou informações incertas. Para contornar tais problemas, os métodos de análise podem ser desenvolvidos a partir da teoria da probabilidade, como as Redes Bayesianas (RB) (AGUILERA et al., 2011; MELLO et al., 2013).

Redes Bayesianas são modelos matemáticos baseados em: (i) uma rede estruturada graficamente representada por um grafo acíclico direcionado (do inglês *Directed Acyclic Graph* – DAG) que denota as relações de (in)dependência entre as variáveis do modelo e (ii) um conjunto de funções de probabilidades associadas a cada variável que estabelece quão forte é a relação entre as variáveis (NEAPOLITAN, 2004; LANDUYT et al., 2013). Em um DAG, cada nó corresponde a uma variável no modelo, enquanto que um arco direcionado conectando duas variáveis indica a relação de dependência entre elas.

Uma importante vantagem das RBs sobre outros modelos preditivos como redes neurais, é que, ao contrário destas abordagens “caixa preta”, a RB é a representação direta do problema e não do processo de raciocínio (PEARL; RUSSELL, 2011; QIN et al., 2006). Com isto é possível representar a relevância das variáveis para alguma outra de interesse (AGUILERA et al., 2011). A Figura 2.8 apresenta um exemplo de uma RB, onde  $X_1, X_2$  e  $X_3$  representam as variáveis do modelo e os arcos uma suposta relação de dependência entre estas variáveis.

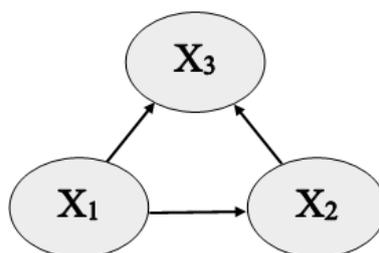


Figura 2.8 – Exemplo de modelo gráfico de uma Rede Bayesiana.

Uma vez que o modelo gráfico é definido, é necessário calcular as funções de probabilidades. Se um nó não possui arcos chegando a ele, (i.e. não possui pais), a função de probabilidade *a priori* é atribuída a este nó:  $P(X_i = x_i)$ , em que  $x_1, x_2, \dots, x_n$  são todos os valores possíveis que a variável  $X_i$  pode assumir. No entanto, se um nó possui um ou mais pais, designa-se a ele a função de probabilidade condicional. Para cada variável, a função expressa a probabilidade desta variável ser instanciada com um determinado valor dado as instanciações de seus pais. Se  $x_i$  denota os valores da variável  $X_i$  e  $pa_i$  denota o conjunto de valores dos pais de  $X_i$  então  $P(x_i|pa_i)$  corresponde à probabilidade condicional. Por exemplo,  $P(X_3 = x_3|X_1 = x_1, X_2 = x_2)$ , é a probabilidade condicional da variável  $X_3$  da Rede Bayesiana da Figura 2.8.

A regra fundamental da probabilidade condicional é:

$$P(a|b) = P(a, b)/P(b) \quad (2.1)$$

que equivale a:

$$P(a|b) * P(b) = P(a, b) \quad (2.2)$$

em que  $P(a, b)$  é a probabilidade conjunta do evento  $(a \wedge b)$ . Sabe-se que  $P(a, b) = P(b, a)$ , assim a partir da Equação 2.2 tem-se:

$$P(a|b) * P(b) = P(b|a) * P(a) \quad (2.3)$$

Os membros desta igualdade podem ser reorganizados de modo que:

$$P(a|b) = P(a) * \frac{P(b|a)}{P(b)}. \quad (2.4)$$

O grau de confiança em  $a$  tendo  $b$  ocorrido ( $P(a|b)$ ) é calculado considerando o grau de confiança inicial em  $a$  ( $P(a)$ ) e o suporte que  $b$  fornece para  $a$  ( $P(b|a)/P(b)$ ). A Equação 2.4 descreve o Teorema de Bayes (NEAPOLITAN, 2004), através do qual é possível atualizar o conhecimento a priori de um evento considerando uma nova/adicional evidência. A habilidade de calcular probabilidades à *posteriori* dada alguma evidência é chamada de inferência.

A eficácia de uma RB reside na possibilidade de calcular, através do Teorema de Bayes, a distribuição de probabilidade para os nós filhos dados os valores de seus pais como também a distribuição dos pais dado os valores de seus filhos. A RB permite conhecer os efeitos dado as causas e as causas dados os efeitos (UUSITALO, 2007; AGUILERA et al., 2011).

Considerando as (in)dependências representadas pelo modelo gráfico da RB, a distribuição conjunta sobre todas as variáveis é calculada pelo produto das probabilidades à *priori* e condicional para cada variável dado seus pais, conforme segue:

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i|pa_i), \quad (2.5)$$

em que  $x_i$  é o valor instanciado para a variável  $X_i$  e  $pa_i$  é o conjunto de valores instanciados para os pais de  $X_i$ . Tomando como exemplo a RB da Figura 2.8, a distribuição conjunta é dada por:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(x_1) * P(x_2|x_1) * P(x_3|x_1, x_2) \quad (2.6)$$

As vantagens das RBs incluem: a capacidade de lidar com grande número de dados mesmo que incompletos, de tratar dados numéricos e categóricos simultaneamente, além de incorporarem o conhecimento do especialista (MELLO et al., 2013).

## 2.6. Baynerd – algoritmo original

A primeira versão do BayNeRD – *Bayesian Network for Raster Data* – foi implementada por Mello et al. (2013). A Figura 2.9 mostra os procedimentos usados na aplicação do algoritmo BayNeRD para identificação das plantações de soja no estado do Mato Grosso.

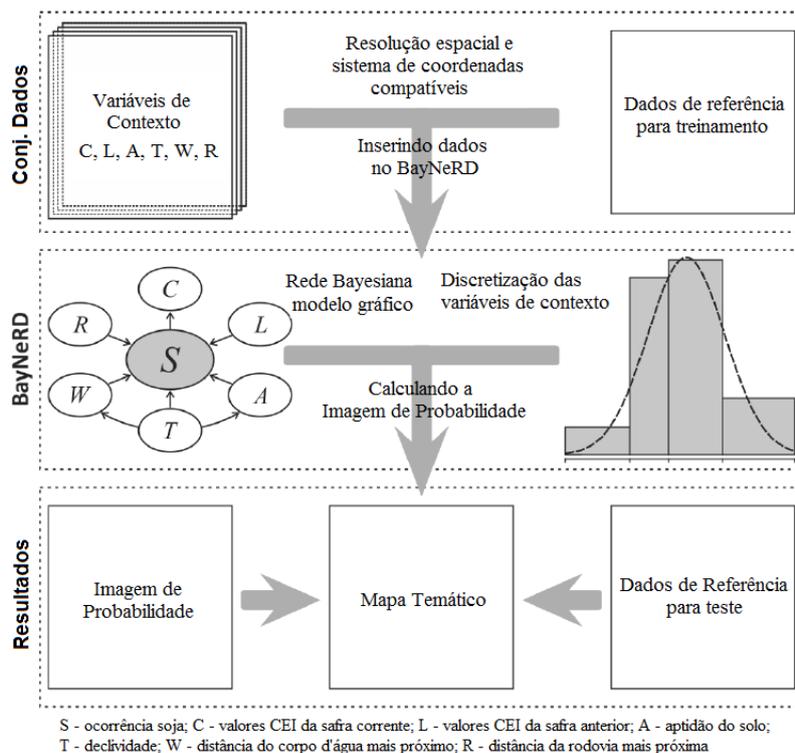


Figura 2.9 – Procedimentos usados em uma aplicação do BayNeRD para mapeamento de soja.

Fonte: adaptada de (MELLO et al., 2013).

O algoritmo foi implementado no software estatístico R (R CORE TEAM, 2014). O sistema manipula dados *raster* no formato GeoTiff. Cada GeoTiff corresponde a uma variável (nó) no modelo gráfico da RB. Mello et al. (2013) nomearam a variável que

representa o fenômeno sob estudo como *variável alvo* e as demais como *variáveis de contexto*.

- Variável alvo

O arquivo GeoTiff que representa a variável alvo deve ser fornecido como dado de referência para treinamento do algoritmo. Seus pixels podem ser divididos em quatro categorias: (i) presença do alvo; (ii) ausência do alvo; (iii) pixels sem informação; e (iv) pixels fora da área de estudo. Embora o arquivo GeoTiff possa conter mais categorias além destas, ao menos as dos itens (i) e (ii) são obrigatórias para que a variável alvo possa assumir os valores 1 para presença do alvo e 0 para ausência.

- Variáveis de contexto

As variáveis de contexto podem ter relações de dependência com a variável alvo e/ou entre si. Elas podem ser do tipo numérico (ex.: modelo digital de elevação) ou categórico (ex.: mapa temático) e também podem conter pixels sem informação. As estruturas de todos os arquivos GeoTiffs (variáveis alvo e de contexto) devem ser compatíveis. Isto é, elas devem ter o mesmo sistema de coordenadas, representar a mesma área geográfica e terem o mesmo tamanho de pixel.

- Modelo gráfico da RB

Após o algoritmo ler todas as variáveis, o BayNeRD interage com o usuário para definir o modelo gráfico da rede. Neste momento, o usuário é questionado sobre as relações de dependência entre todas as variáveis. Como estas relações são representadas por arcos no DAG, o algoritmo pergunta ao usuário se há um arco entre cada par de variáveis.

- Discretização e Funções de Probabilidade

De acordo com Mello et al. (2013), o BayNeRD foi desenvolvido para interagir com o usuário para que ele, utilizando seu conhecimento, defina as funções de probabilidade através do processo de discretização das variáveis de contexto. Nesta etapa, os valores observados de uma variável são representados por quantidades discretas. Em outras palavras, o *range* de valores é dividido em intervalos definidos pelo usuário e os valores são codificados de acordo com o intervalo ao qual pertence.

Estes intervalos podem ser calculados segundo três opções implementadas no BayNeRD: (i) intervalos equidistantes – os intervalos são divididos em tamanhos iguais; (ii) quantis – os intervalos têm o mesmo número de elementos (pixels); e (iii) manualmente – o usuário define os limites de cada intervalo. O processo de discretização influencia as funções de probabilidade, que são calculadas através da contagem de pixels levando em conta as relações de dependência entre as variáveis definidas no modelo gráfico e os intervalos definidos na discretização (MELLO et al., 2010). O usuário deve ser hábil na discretização de cada variável de contexto. Assim, todos os cenários, que envolve a combinação dos intervalos das variáveis com os intervalos de seus pais, terão dados representativos para calcular as funções de probabilidade.

- Imagem de probabilidade

Quando a probabilidade de ocorrência do alvo é calculada para cada pixel dentro da área de estudo, dado um cenário específico, i.e. os valores observados para o mesmo pixel nas variáveis de contexto, a Imagem de Probabilidade (IP) é formada. A IP consiste em um dado *raster* georreferenciado e é o principal resultado do BayNeRD. Supondo que, na Rede Bayesiana do exemplo na Figura 2.8, a variável  $X_3$  é a variável alvo, a IP seria formada por:

$$P(X_3 = 1 \mid X_1 = x_1, X_2 = x_2) \quad (2.7)$$

para cada pixel dentro da área de estudo; onde 1 é usado para representar a presença do alvo e  $x_i$  são os valores observados nas variáveis  $X_i$ .



### 3 ESTRUTURA DO MÉTODO DE INFERÊNCIA e-BayNeRD

Este capítulo apresenta as modificações que foram implementadas no sistema de inferência probabilística BayNeRD (MELLO et al., 2013). Serão descritas as mudanças em relação ao pré-processamento dos dados, à interação com o usuário e aos resultados do algoritmo aprimorado, aqui nomeado e-BayNeRD – *Enhanced Bayesian Network for Raster Data*.

Assim como o algoritmo original, o e-BayNeRD também foi desenvolvido no software R (R CORE TEAM, 2014). As melhorias serão descritas através de um processo comparativo entre as duas versões do algoritmo, a começar pelos pacotes de manipulação de dados *raster*. Existem alguns pacotes já implementados em R para este fim, como o pacote *rgdal* (BIVAND et al., 2014) utilizado no BayNeRD. Esta era uma questão que precisava ser aprimorada no algoritmo original. Neste caso, o pacote ‘*raster*’<sup>1</sup> (HIJMANS et al., 2014) mostrou-se mais adequado ao programa, porque fornece várias funções de alto nível para manipulação de dados *raster* além de permitir trabalhar melhor com volume grandes de dados *raster*.

As funções do pacote não carregam os valores das células dos dados *raster*, os pixels, na memória. O pacote cria um tipo de objeto, a partir destes dados, que contém informações como nome do arquivo, tamanho do pixel, sistema de coordenadas geográficas, extensão, etc. Em cálculos com os pixels, os dados são processados em blocos – *chunks* (HIJMANS, 2014). Considerando as vantagens do pacote ‘*raster*’ sobre o pacote *rgdal*, suas funções foram incorporadas ao e-BayNeRD a fim de otimizar o processamento das variáveis alvo e de contexto.

#### 3.1. Variável alvo

No algoritmo original, a variável alvo deve ser instanciada com dois valores representando a presença ou a ausência do alvo em estudo. No entanto, a complexidade dos fenômenos estudados pode dificultar a sua representação com apenas uma classe. Com o e-BayNeRD é possível contornar esta situação, já que a variável alvo pode

---

<sup>1</sup> Neste ponto, *raster* refere-se ao nome do pacote e não ao tipo do dado.

conter mais de duas instanciações, isto é, ela pode representar mais classes. Por exemplo, se um determinado estudo tiver interesse nas culturas de verão como a soja e o milho, a variável alvo pode ter o valor 1 para representar a soja, o valor 2 para representar o milho e o valor 0 para representar a ausência de ambos.

Esta mudança no algoritmo altera a definição das funções de probabilidade assim como os resultados (*outputs*) do algoritmo. Adicionalmente, o dado *raster* da variável alvo ainda pode conter as categorias que representem pixels sem informação e pixels fora da área de estudo. Esta última categoria é utilizada apenas para mascarar os pixels das variáveis de contexto que se encontram fora da área de estudo.

### 3.2. Variáveis de contexto

Um dos principais benefícios da incorporação do pacote '*raster*' ao sistema e-BayNeRD é a redução das operações na etapa de pré-processamento dos dados pelo usuário. No algoritmo aprimorado, os dados *raster* podem ter diferentes estruturas, ao contrário do algoritmo original, em que todos os dados devem ser compatíveis.

O e-BayNeRD toma como referência a variável alvo. Cada variável de contexto inserida é transformada para ser compatível com os parâmetros da variável alvo. Isto é, o algoritmo é capaz de transformar o sistema de coordenadas, reamostrar o tamanho do pixel e intersectar o dado *raster* de acordo com o retângulo envolvente da variável alvo, conforme mostra a Figura 3.1. Neste caso, cria-se um novo dado *raster* com os parâmetros da variável alvo e os pixels são interpolados usando o método do vizinho mais próximo para minimizar a distorção. Se uma região não possui intersecção entre as variáveis alvo e de contexto, os pixels desta região são preenchidos com valores NA (do inglês, *Not Available*), indicando pixels sem informação.



Figura 3.1 – Diferentes possibilidades de intersecção entre as variáveis alvo e de contexto.

### 3.3. Modelo gráfico da RB

O processo interativo para definição das relações de dependência entre as variáveis foi substancialmente aprimorado. Antes das melhorias, a interação com o usuário dava-se por linha de comando. Para cada variável de contexto, o usuário deveria responder a  $(n - 1)$  questões sobre a existência de um arco desta variável de contexto para as demais, onde  $n$  é o número de variáveis no modelo. Portanto, totalizam-se  $n * (n - 1)$  questionamentos, ou seja, o número de questões aumenta exponencialmente.

Com o propósito de deixar esta etapa mais amigável ao usuário, um modo mais eficiente para definir as relações entre as variáveis foi implementado. O software R também possui pacotes para manipular DAGs e executar análises Bayesianas. Deste modo, foi possível criar uma interface para o modelo gráfico da RB por meio da integração dos pacotes *'deal'* (BOTTCHER; DETHLEFSEN., 2013) e *'bnlearn'* (SCUTARI, 2015). O primeiro permite ao usuário especificar a RB utilizando o *mouse* e o segundo é usado nas análises e inferências Bayesianas. A Figura 3.2 apresenta a interface implementada no e-BayNeRD.

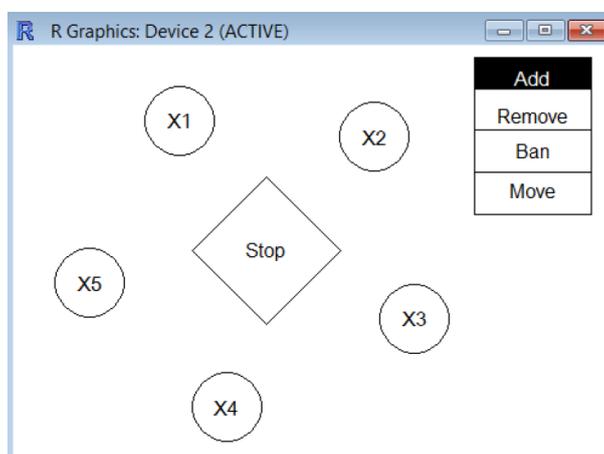


Figura 3.2 – Interface para definição do modelo gráfico da RB.

Com esta interface, o usuário pode inserir ou remover arcos entre pares de variáveis. Para inserir/remover um arco de X1 para X2, primeiro ele deve clicar em *'Add'*/*'Remove'* e assim clicar no nó X1 e depois em X2. Após a construção do modelo

gráfico, o usuário deve certificar-se de que o mesmo está correto e finalizar o processo clicando em ‘*Stop*’.

Outra vantagem desta melhoria é que a interface ajuda o usuário a não cometer erros. Por exemplo, se o usuário inserir um arco que crie um ciclo no grafo, este arco não será desenhado uma vez que o modelo gráfico da RB deve representar um DAG – acíclico. Portanto, a tarefa de construção do modelo gráfico tornou-se mais fácil e mais rápida que a versão original.

### **3.4. Discretização e funções de probabilidade**

No processo de discretização das variáveis de contexto, os intervalos podem ser definidos manualmente pelo usuário ou segundo os critérios de intervalos equidistantes e de quantis. Estes critérios foram mantidos no e-BayNeRD e um novo foi adicionado. Agora o usuário também pode escolher a opção de *clustering* para converter variáveis contínuas em discretas. Esta opção é um método não-supervisionado, em que os valores observados em cada variável de contexto são particionados em  $k$  clusters definidos pelo usuário. Assim cada valor é codificado como pertencente ao cluster com a média mais próxima (HAHSLER et al., 2014).

Os intervalos são definidos por limites inferior e superior. Cada intervalo é codificado como uma categoria. Assim, a variável de contexto terá  $n$  categorias se for discretizada em  $n$  intervalos. As funções de probabilidade são computadas pela contagem de pixels categóricos para cada variável de contexto, levando em conta os pixels categóricos das variáveis que são pais desta. Com isto constrói-se a tabela de probabilidade condicional. Por exemplo, se uma variável não possui pais, a probabilidade de cada categoria é dada pelo número de pixels pertencentes a esta categoria dividido pelo número total de pixels. A soma das probabilidades de todas as categorias deve ser igual a 1. Mas, no caso de a variável ter um ou mais pais, a função de probabilidade condicional é calculada a partir de uma tabela de frequências.

Conforme mencionado anteriormente, a possibilidade de mais de uma classe para a variável alvo tem consequências nesta etapa de discretização. Agora o usuário também

deve executar a discretização da variável alvo, já que há a possibilidade de que ela seja contínua. Por exemplo, no caso do “índice de vegetação” ser a variável alvo, durante o processo de discretização o usuário deve definir os limites dos intervalos e definir qual(is) deles representa(m) a presença do alvo.

### 3.5. Imagem de probabilidade

O número de valores instanciados para a variável alvo também define o número de *layers* da Imagem de Probabilidade (IP). Vamos supor que na Rede Bayesiana, do exemplo na Figura 2.8, a variável  $X_3$  seja a variável alvo e a ocorrência do fenômeno seja representada por três classes:  $X_3 = 1$ ,  $X_3 = 2$ ,  $X_3 = 3$ . Neste caso, a IP será um *layer* com os valores de probabilidade de ocorrência (presença) para cada classe. O conceito de IP no e-BayNeRD é um pouco diferente. Cada *layer* é chamado de Banda de Probabilidade (BP) e a integração de todas as BPs produz a Imagem de Probabilidade.

Por meio do pacote *bnlearn*, consultas às tabelas de probabilidade podem ser executadas e o sistema retorna a probabilidade de um evento dada alguma evidência observada. Entende-se como evidência o conjunto de valores instanciados das variáveis de contexto. As combinações de todas as evidências possíveis formam os cenários. Ainda considerando a Rede Bayesiana da Figura 2.8, as Bandas de Probabilidade seriam calculadas, respectivamente, como:

$$P(X_3 = 1|X_1 = x_1, X_2 = x_2) \quad (3.1)$$

$$P(X_3 = 2|X_1 = x_1, X_2 = x_2) \quad (3.2)$$

$$P(X_3 = 3|X_1 = x_1, X_2 = x_2) \quad (3.3)$$

Para um dado cenário ( $X_1 = x_1, X_2 = x_2$ ), há três eventos para calcular as probabilidades associadas. Conseqüentemente, as Bandas de Probabilidade são computadas simultaneamente para otimizar o algoritmo. Se qualquer variável de contexto contiver pixels sem informação, a probabilidade de ocorrência do alvo ainda

assim é calculada, sendo possível até calcular  $P(X_3 = x_3)$  para pixels sem qualquer observação através da probabilidade a priori.

Assim como o algoritmo original, o e-BayNeRD calcula a influência de cada variável de contexto nas probabilidades para a variável alvo. O cálculo é feito usando a divergência de Kullback-Leibler (KL), que é uma medida não-simétrica da diferença entre duas distribuições de probabilidade (KULLBACK; LEIBLER, 1951). Assim, é possível mensurar a influência de cada variável de contexto calculando a divergência KL entre as probabilidades condicionais e marginais do modelo bayesiano (MELLO et al., 2013).

O e-BayNeRD gera uma Banda de Probabilidade para a ausência do alvo através do cálculo:  $BP_{\text{não-alvo}} = 1 - bp$ , onde  $bp$  corresponde à soma de todas as BPs anteriormente calculadas, sendo possível criar um mapa temático. A classificação é realizada em nível de pixel por meio do princípio básico da máxima verossimilhança. O pixel é classificado como pertencente à classe cuja Banda de Probabilidade possui o maior valor de probabilidade para este pixel.

## 4 APLICAÇÃO DO e-BayNeRD PARA IDENTIFICAÇÃO DA CANA-DE-AÇUCAR E DE ÁREAS APTAS À SUA EXPANSÃO

Neste Capítulo, descreve-se como o sistema de inferência e-BayNeRD foi usado para mapear áreas de cultivo da cana-de-açúcar e áreas potenciais à sua expansão sustentável no Estado de São Paulo.

### 4.1. Área de estudo

A escolha do estado de São Paulo como área de estudo justifica-se pelo fato deste estado ser o maior produtor de cana-de-açúcar no Brasil. Na safra de 2013, o estado de São Paulo foi responsável por, aproximadamente, 53% da área plantada (IBGE, 2015). Além disso, a possibilidade de poder identificar novas áreas para expansão da cana-de-açúcar neste estado atende a uma demanda do mercado energético, como foi tratado anteriormente no presente documento. A Figura 4.1 mostra a área de estudo e os dados de referência usados neste trabalho, os quais serão detalhados posteriormente.

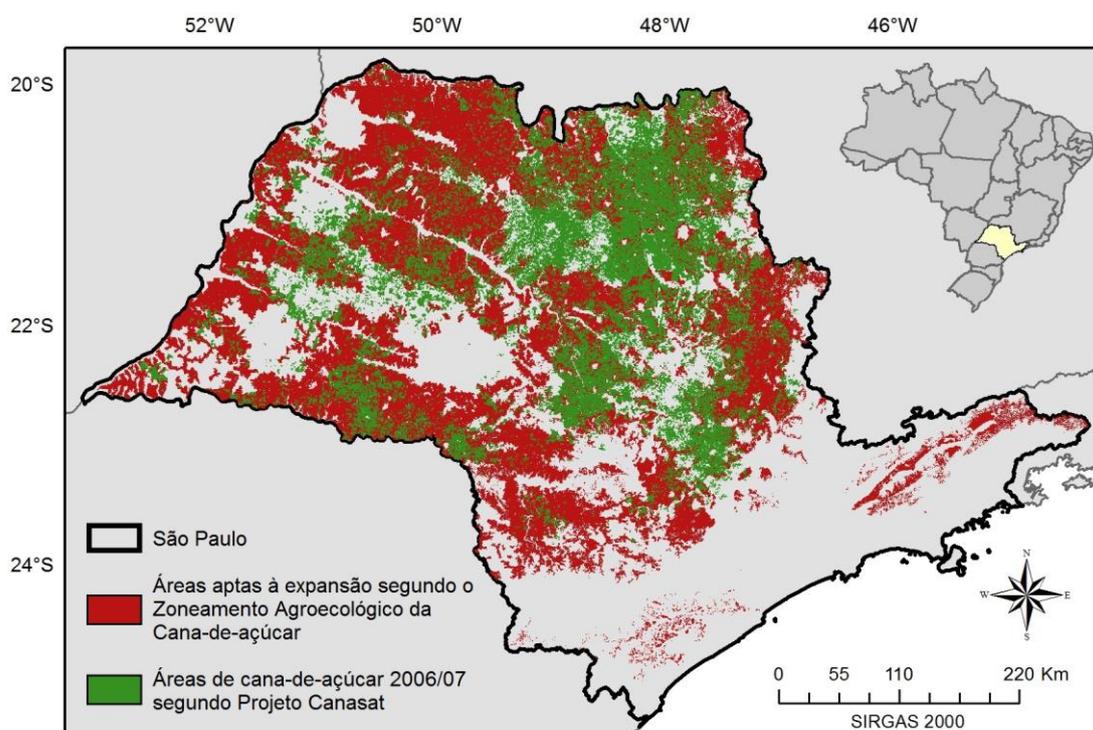


Figura 4.1 – Área de estudo: estado de São Paulo; dados de referência: áreas aptas à expansão, indicadas pelo Zoneamento Agroecológico da Cana-de-açúcar e áreas com cana-de-açúcar cultivada safra 2006/07, de acordo com o Projeto Canasat.

## 4.2. Variáveis do modelo RB

### a) Variável alvo (ZC)

Neste trabalho, dois alvos específicos são estudados e, portanto, a variável alvo (chamada como *ZC*) é composta por duas classes, como mostra a Figura 4.1: (i) áreas de alta aptidão à expansão da cana-de-açúcar, de acordo com o Zoneamento Agroecológico da Cana-de-açúcar – ZAE-Cana, fornecido pela Embrapa Meio Ambiente (<http://geo.cnpma.embrapa.br/>); e (ii) áreas de cultivo da cana-de-açúcar na safra 2006/07, fornecidas pelo Projeto Canasat ([www.dsr.inpe.br/laf/canasat/](http://www.dsr.inpe.br/laf/canasat/)) (RUDORFF et al., 2010).

A fim de evitar a correlação espacial, criou-se uma máscara sobre o dado *raster* contendo janelas em posições aleatórias de 5x5 pixels, de modo que dois terços dos pixels foram selecionados para compor o dado *raster* de treinamento e o restante usado para validação. Portanto, a variável alvo é representada pelas seguintes classes: (i) áreas aptas à expansão; (ii) áreas de cana-de-açúcar; (iii) ausência de ambos; (iv) pixels sem informação, que se refere aos pixels usados na validação; e (v) pixels fora da área de estudo. Assim, no e-BayNeRD, a variável alvo  $ZC = z_c$  pode assumir os seguintes valores:  $z_c = 1$  para áreas aptas à expansão,  $z_c = 2$  para áreas de cana-de-açúcar,  $z_c = 0$  para ausência de ambos,  $z_c = -2$  para pixels sem informação e  $z_c = -1$  para pixels fora da área de estudo. O *raster* GeoTiff representando a variável alvo foi formado conforme a grade do MODIS com resolução espacial nominal de 250x250 metros (JUSTICE et al., 2002).

### b) Variáveis de contexto

As variáveis selecionadas para compor o modelo são apresentadas na Tabela 4.1.

Tabela 4.1 – Variáveis de contexto usadas no modelo

<b>Variável</b>	<b>Descrição</b>
ZAA	<b>Z</b> oneamento <b>A</b> groambiental
ApEc	<b>A</b> ptidão <b>E</b> dafoclimática
CEI	<i>Crop Enhancement Index</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
D	<b>D</b> eclividade (em %)
S	<b>S</b> olos

O Zoneamento Agroambiental (ZAA) para o setor sucroalcooleiro foi elaborado em 2008 pela Secretaria do Meio Ambiente e pela Secretaria da Agricultura e Abastecimento do estado de São Paulo. Assim como o ZAE-Cana, o Zoneamento Agroambiental foi elaborado para orientar o planejamento sustentável do crescimento do setor sucroenergético no Estado. No ZAA foram considerados aspectos hidrográficos, físicos, topográficos e climáticos. O mapa temático contém diferentes classes com graus de aptidão agroambiental e adequação à instalação de unidades agroindustriais: (i) adequada; (ii) adequada com limitações; (iii) adequada com restrições; e (iv) inadequada. As classes (ii) e (iii) são adequadas do ponto de vista edafoclimático, entretanto possuem em seus territórios Áreas de Proteção Ambiental (APA) e Unidades de Conservação e Proteção Integral (UCPI), respectivamente.

O mapa de Aptidão Edafoclimática (*ApEc*) para a cana-de-açúcar no estado de São Paulo foi revisado durante a elaboração do Zoneamento Agroambiental (GOMES; MONTAÑO, 2012) e corresponde às restrições impostas pelo solo e pelo clima ao cultivo da cultura. Com isto, espera-se que ocorra uma relação de proximidade de ZAA e *ApEc* com ZC, de tal modo que ambas sejam pais de ZC no modelo gráfico da RB. Levando em conta que *ApEc* influencia ZAA, *ApEc* também deve ser pai de ZAA. Os mapas destas duas variáveis são disponibilizados pela Secretaria do Meio Ambiente ([www.ambiente.sp.gov.br/etanolverde/zoneamento-agroambiental/](http://www.ambiente.sp.gov.br/etanolverde/zoneamento-agroambiental/)).

O índice CEI – *Crop Enhancement Index* (RIZZI et al., 2009) foi desenvolvido para ter sensibilidade ao calendário agrícola. O índice CEI considera o comportamento temporal do índice EVI – *Enhanced Vegetation Index*, produto MOD13Q1 (HUETE et al., 2002),

ao longo do ciclo de desenvolvimento da cultura sob estudo, especificamente em dois momentos: quando há o preparo do solo para plantio ou a planta está em fase inicial – menores valores de EVI; e quando a cultura atinge seu maior vigor vegetativo – maiores valores de EVI. Os valores de CEI podem variar entre [-1, 1] e são calculados para cada pixel como:

$$CEI = G * \frac{(MaxEVI + R) - (MinEVI + R)}{(MaxEVI + R) + (MinEVI + R)} \quad (3.4)$$

em que  $G = 10^2$  é o fator de ganho;  $R = 10^2$  é o coeficiente de realce;  $MaxEVI$  corresponde ao máximo valor de EVI observado para o período de desenvolvimento da cana-de-açúcar, que atinge seu crescimento máximo em abril (RUDORFF et al., 2010);  $MinEVI$  é o mínimo valor de EVI no período em que acontece a colheita, que, no caso da cana-de-açúcar, estende-se até dezembro (RUDORFF et al., 2010). Assim, selecionou-se a série EVI, no Banco de Produtos MODIS da Embrapa ([www.modis.cnptia.embrapa.br](http://www.modis.cnptia.embrapa.br)), entre início de novembro de 2006 e fim de abril 2007 para compor a imagem de  $MaxEVI$ ; e entre início de abril de 2007 e fim de novembro de 2007 para compor a imagem de  $MinEVI$ . Espera-se que áreas com cana-de-açúcar apresentem maiores valores de  $CEI$ , assim  $ZC$  deve ser pai de  $CEI$  no modelo gráfico da RB.

Levando em conta que a refletância espectral pode variar conforme as mudanças fenológicas da cultura (RAHMAN et al., 2004), considerou-se a incorporação das imagens de mínimo e máximo  $NDVI$  no mesmo período utilizado para calcular o índice  $CEI$ . Mas, através da imagem de máximo  $NDVI$  não foi possível fazer qualquer discretização das regiões com presença e ausência dos alvos. Portanto, apenas a imagem de mínimo  $NDVI$  foi considerada, visto que, no período da colheita, as regiões de cultivo da cana-de-açúcar apresentam menores valores de  $NDVI$  se comparadas com as áreas de proteção, matas ciliares, Mata Atlântica. Neste caso,  $ZC$  deve ser pai de  $NDVI$ .

A declividade ( $D$ ) é uma variável crítica no cultivo da cana-de-açúcar, uma vez que define as áreas aptas à mecanização. A atual legislação estabelece a extinção da prática

de queima da cana-de-açúcar nas regiões com declividade  $> 12\%$  (AGUIAR et al., 2011). Considerando que  $D$  tem grande influência sobre as áreas aptas ao plantio sustentável da cana-de-açúcar,  $D$  deve ser pai de  $ZC$ , bem como de  $ApEc$  e  $ZAA$ . Os dados de declividade foram obtidos através do Banco de Dados Geomorfométricos do Brasil – Projeto TOPODATA ([www.dsr.inpe.br/topodata](http://www.dsr.inpe.br/topodata)).

Outra variável que influencia na aptidão do plantio da cana-de-açúcar é a qualidade do solo. Embora a cana-de-açúcar seja cultivada nos mais diversos solos, aqueles mais adequados ao seu cultivo devem ser profundos, bem arejados, com boa capacidade de retenção de água e nutrientes (UDOP, 2015). O mapa da variável solos ( $S$ ) foi criado levando em conta a qualidade do solo e a fertilidade. Considerando que  $S$  é uma variável de interesse na elaboração de um mapa de aptidão e zoneamento,  $S$  deve ser pai de  $ZC$ ,  $ZAA$  e  $ApEc$  no modelo gráfico da RB. Os dados referentes ao solo foram adquiridos no Ministério do Meio Ambiente ([mapas.mma.gov.br/i3geo/datadownload.htm#](http://mapas.mma.gov.br/i3geo/datadownload.htm#)).

#### 4.3. Modelo gráfico da RB para a aplicação proposta

Considerando as variáveis descritas na seção 4.2, o modelo gráfico da Rede Bayesiana para inferência de áreas de cultura de cana-de-açúcar e áreas aptas à sua expansão é representado na Figura 4.2. No modelo da RB, cada nó representa uma variável e os arcos representam suas relações de dependência descritas anteriormente.

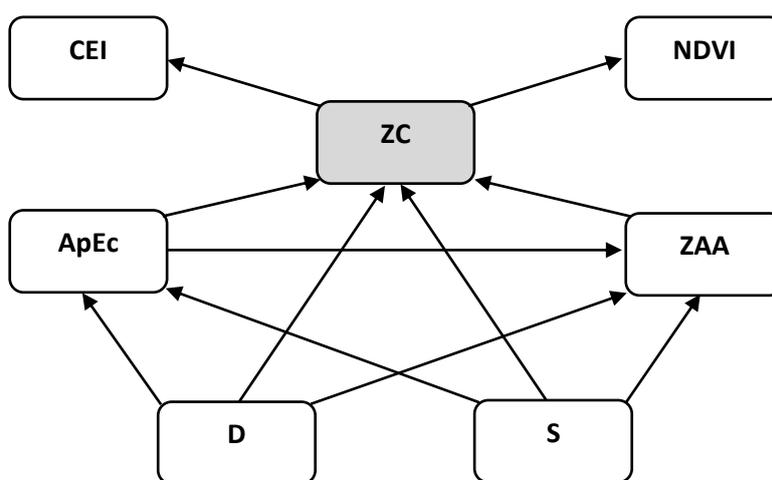


Figura 4.2 – Modelo gráfico da Rede Bayesiana.

#### 4.4. Discretização das variáveis

Dado o modelo gráfico da RB, o próximo passo a ser realizado é a discretização das variáveis contínuas. O número de intervalos deve ser escolhido baseado no conhecimento do usuário. A escolha de poucos intervalos pode não descrever corretamente o comportamento da variável em relação ao fenômeno estudado e um grande número de intervalos pode comprometer o cálculo das funções de probabilidade (MELLO et al., 2013). Com a definição manual dos limites dos intervalos, o conhecimento do especialista pode contribuir para gerar um modelo mais adequado à solução do problema.

Em relação à variável declividade ( $D$ ), assume-se um limite para mecanização da colheita de 12%, preferencialmente, o cultivo ocorre em regiões mais planas. As maiores declividades são observadas no extremo leste do Estado de São Paulo, coincidentes com as regiões sem cultivo da cana-de-açúcar. A partir deste conhecimento, os valores  $d$  da variável  $D$  foram divididos em três intervalos: (i)  $d < 6\%$ ; (ii)  $6\% \leq d < 12\%$ ; e (iii)  $d \geq 12\%$ . Pode-se observar na Rede Bayesiana da Figura 4.2, que a variável  $D$  não possui pais e, portanto, o e-BayNeRD calcula a função de probabilidade a priori das seguinte forma:  $P(d < 0,06)$ ,  $P(0,06 \leq d < 0,12)$  e  $P(d \geq 0,12)$ . A Figura 4.3 apresenta a discretização da variável  $D$  e as Tabelas 4.2 e 4.3 mostram as suas probabilidades a priori e a posteriori, respectivamente.

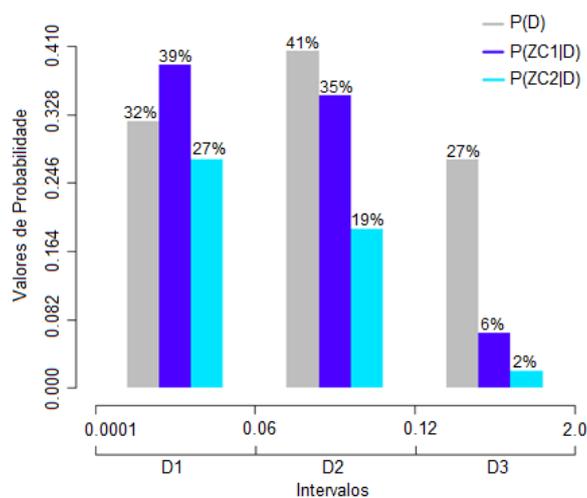


Figura 4.3 – Discretização da variável de contexto declividade ( $D$ ).

Tabela 4.2 – Probabilidade a priori  $P(D = d)$ 

<b>Intervalos</b>	<b>D1</b>	<b>D2</b>	<b>D3</b>
<b>Probabilidades</b>	0,3200	0,4051	0,2749

Tabela 4.3 – Probabilidade condicional  $P(ZC = zc|D = d)$ 

		<b>D</b>		
<b>Intervalos</b>		<b>D1</b>	<b>D2</b>	<b>D3</b>
<b>ZC</b>	<b>ZC0*</b>	0,3368	0,4579	0,9142
	<b>ZC1</b>	0,3884	0,3518	0,0656
	<b>ZC2</b>	0,2748	0,1903	0,0202

\*ZC0 refere-se à ausência de ambos os dados

Nota-se que, aproximadamente, 73% da área de estudo possui declividade inferior a 12%, sendo a maior parte no intervalo D2, ou seja, entre 6% e 12%. Espera-se que a probabilidade de ocorrência de áreas de cana-de-açúcar e áreas aptas à expansão seja maior em regiões com declividade inferior a 12%. Isto é, a probabilidade de  $ZC = zc$  condicionada a  $D = d$ , é calculada como:  $P(ZC = zc|D = d)$ .

Assim como a variável  $D$ , as demais variáveis de contexto foram discretizadas com os intervalos descritos na Tabela 4.4. Os intervalos das variáveis  $CEI$  e  $NDVI$  foram definidos por meio de uma análise de dispersão destas variáveis em função da variável alvo, conforme é ilustrado na Figura 4.4. Pode-se observar que a discriminação entre as classes áreas aptas ( $ZC=1$ ) e áreas de cana-de-açúcar ( $ZC=2$ ) é confusa devido à similaridade entre a distribuição dos valores de  $CEI$  e  $NDVI$ . A classe não-alvo ( $ZC=0$ ) pode ser mais facilmente discriminada por meio da variável  $NDVI$ , diferentemente das classes  $ZC=1$  e  $ZC=2$ . Em relação às variáveis  $S$ ,  $ZAA$  e  $ApEc$ , estas já são do tipo categóricas.

Tabela 4.4 – Discretização das variáveis de contexto

Nº de Intervalos	CEI	NDVI	D (%)	S	ZAA	ApEc
1	[min; 0,175)	[min; 0,52)	[min; 0,06)	Bom	Adequado	Adequado
2	[0,175; 0,2375)	[0,52; max]	[0,06; 0,12)	Regular	Limitado	Limitado/Restrito
3	[0,2375; max]		[0,12; max]	Restrito	Restrito	Inadequado
4					Inadequado	

\*min e max referem-se aos valores mínimo e máximo da variável, respectivamente

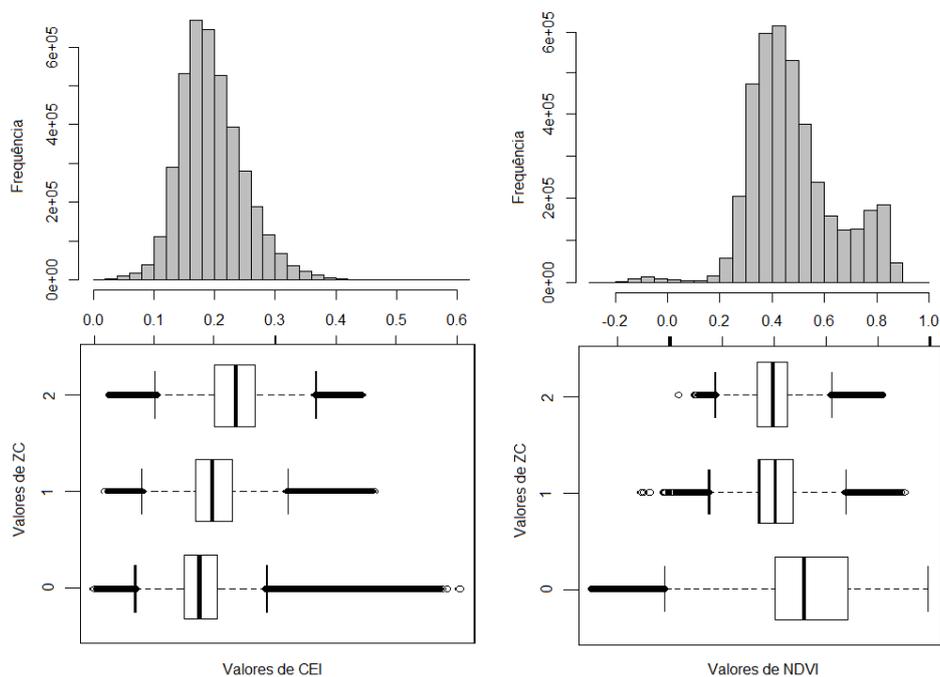


Figura 4.4 – Dispersão das variáveis de contexto CEI (esquerda) e NDVI (direita) em função da variável alvo (ZC).

## 5 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados da aplicação do sistema de inferência e-BayNeRD para o mapeamento das áreas de cana de açúcar e áreas potenciais ao seu plantio no estado de São Paulo.

### 5.1. Bandas de probabilidade

Baseando-se no modelo gráfico, ilustrado na Figura 4.2, e nas funções de probabilidade, o sistema e-BayNeRD calcula, para cada pixel na área de estudo, a probabilidade de ocorrência do alvo. Os valores de probabilidade de ocorrência de áreas aptas ( $ZC=1$ ) e áreas de cana-de-açúcar ( $ZC=2$ ) são calculados dados os valores observados das variáveis de contexto. Dadas as relações de dependência das variáveis no modelo gráfico da RB, pode-se definir que:

$$P(ZC = 1 | CEI = cei, NDVI = ndvi, D = d, S = s, ApEc = apec, ZAA = zaa) \quad (5.1)$$

$$P(ZC = 2 | CEI = cei, NDVI = ndvi, D = d, S = s, ApEc = apec, ZAA = zaa) \quad (5.2)$$

onde as letras minúsculas representam os valores instanciados das respectivas variáveis, que são representadas por letras maiúsculas. As Bandas de Probabilidade calculadas a partir das Equações 5.1 e 5.2 são apresentadas nas Figuras 5.1 e 5.2, respectivamente.

As Bandas de Probabilidade mostram a distribuição espacial da probabilidade de áreas aptas à expansão (BP1, Figura 5.1) e de áreas de cana-de-açúcar (BP2, Figura 5.2) no estado de São Paulo. Os pixels em cores mais fortes representam os maiores valores de probabilidade, baseados nas observações feitas sobre as variáveis de contexto

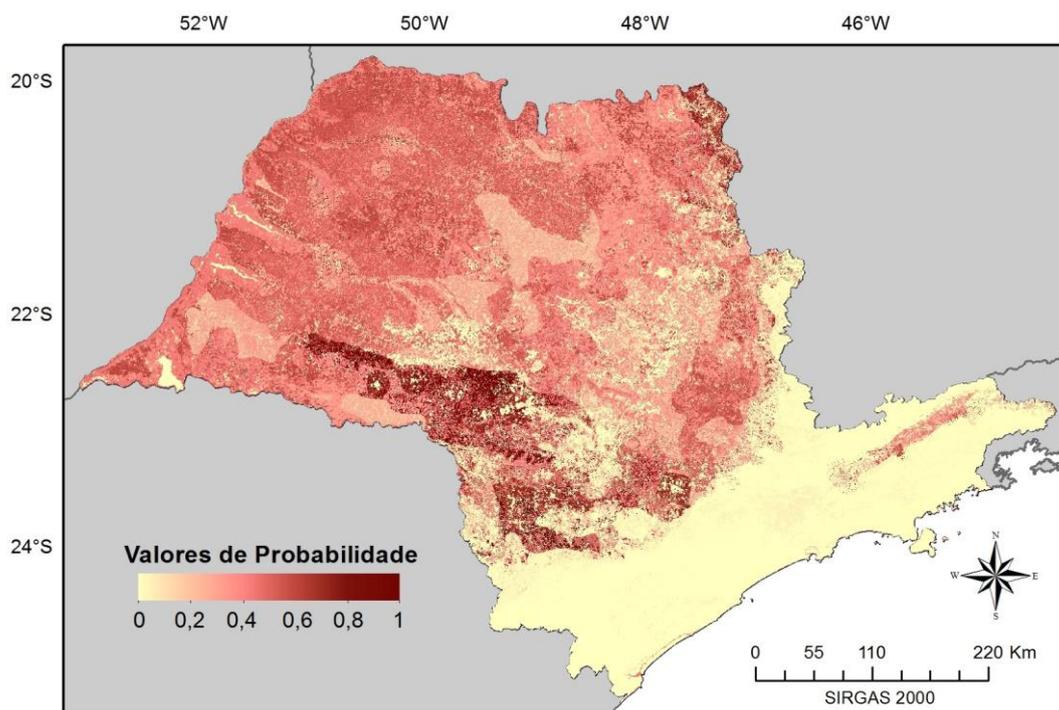


Figura 5.1 – BP1: probabilidade de ocorrência de áreas aptas à expansão.

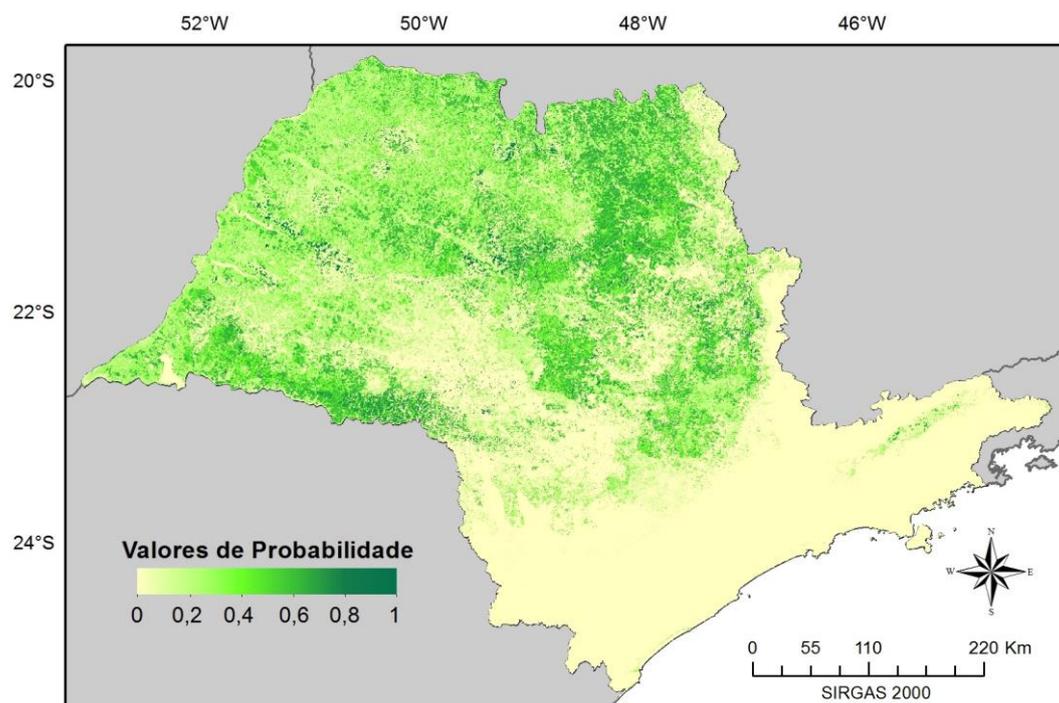


Figura 5.2 – BP2: probabilidade de ocorrência de áreas de cana-de-açúcar.

Pode-se observar que os resultados nas áreas de Mata Atlântica, na região Sul-Leste do estado e nas áreas de proteção localizadas no Centro-Sul e extremo Oeste do estado de

São Paulo receberam valores de probabilidade zero ou próximo de zero, que estão coerentes com a realidade. A maior concentração de áreas de cana-de-açúcar é verificada na região Centro-Norte (RUDORFF et al., 2010), região na qual os valores de probabilidade na BP2 são maiores, como mostra a Figura 5.2.

Por outro lado, em relação às áreas aptas à expansão da cana-de-açúcar, ilustradas na Figura 5.1, pode-se observar uma situação diferente. Os resultados apontam que o Oeste e Norte do Estado apresentam as maiores possibilidades de expansão do plantio da cana-de-açúcar, com probabilidades entre 50% e 60%. De fato, o Oeste do estado apresenta potencial de expansão para o plantio da cana-de-açúcar que já vem acontecendo nos últimos anos, como já evidenciado por Rudorff et al. (2010).

As Figuras 5.3 e 5.4 mostram alguns recortes ampliados das Bandas de Probabilidades para os dois alvos, juntamente com os valores observados das variáveis de contexto. Nestas áreas, observa-se que as regiões de alta declividade tiveram baixos valores de probabilidade para ocorrência das duas classes, mesmo quando outras variáveis de contexto apresentavam boas condições para ocorrência das classes. O mesmo ocorreu com as áreas de proteção que possuem valores de NDVI altos. Constatou-se, também, que os valores de probabilidade altos nas áreas de cana-de-açúcar (Figura 5.4) têm uma correlação com o comportamento do índice CEI.

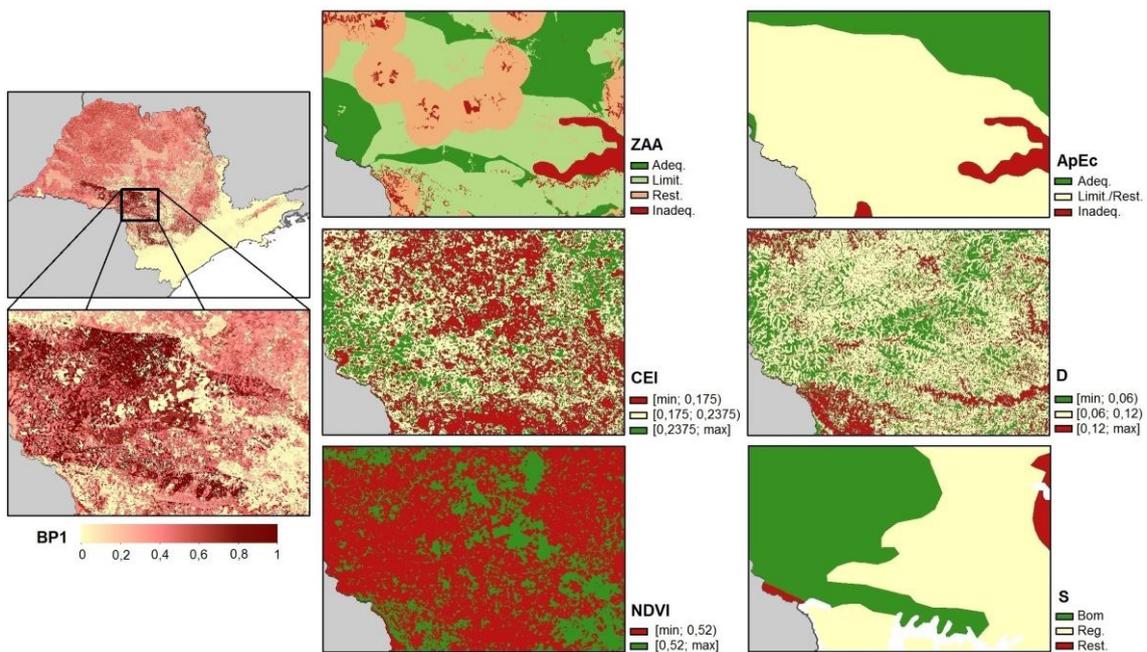


Figura 5.3 – Região de alta probabilidade de ocorrência de áreas aptas à expansão da cana-de-açúcar.

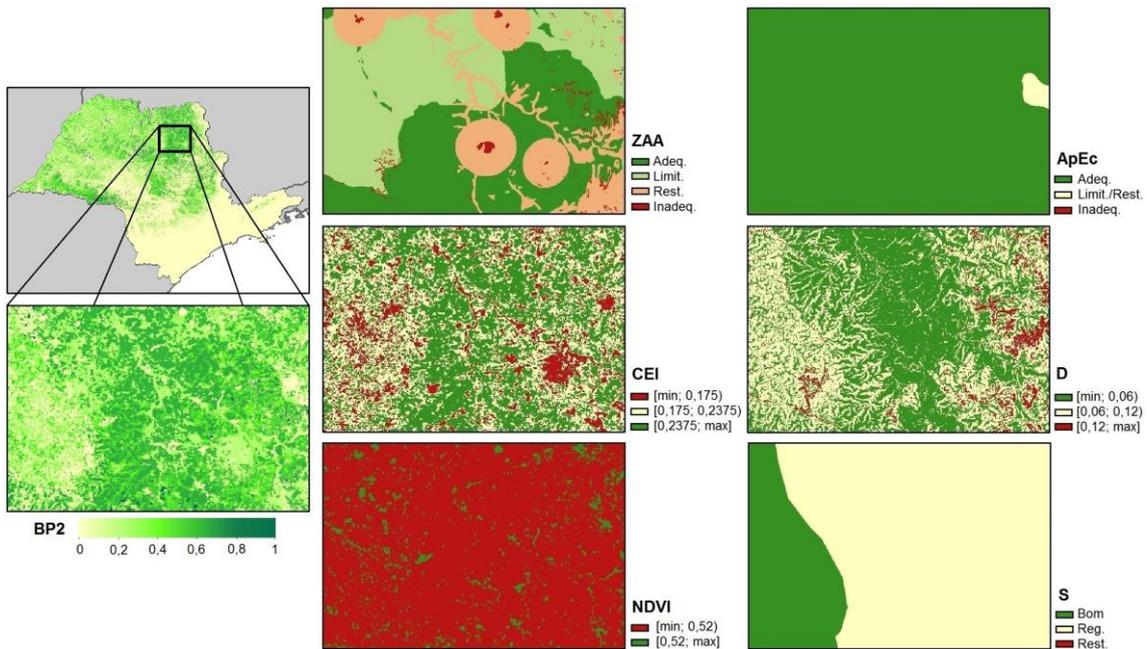


Figura 5.4 – Região de alta probabilidade de ocorrência de áreas de cana-de-açúcar.

A Tabela 5.1 mostra a influência das variáveis de contexto sobre a variável alvo segundo os valores de divergência KL. A variável ZAA teve maior importância (0,493), em relação às outras variáveis, para inferir a ocorrência das classes do alvo. Este

comportamento já era esperado devido ao fato da variável *ZAA* ser resultado de um planejamento sustentável – Zoneamento Agroambiental para a cultura da cana-de-açúcar.

Tabela 5.1 – Divergência KL das variáveis de contexto.

<b>Variável</b>	<b>Porcentagem</b>	<b>KL</b>
ZAA	27,04%	0,493
ApEc	24,33%	0,443
C	16,34%	0,298
D	14,83%	0,270
NDVI	12,29%	0,224
S	5,17%	0,094

Assim como *ZAA*, a variável *ApEc* também é produto da análise de outras variáveis, já que representa as restrições impostas pelo solo e pelo clima ao plantio da cana-de-açúcar. De forma similar, a variável *ApEc* (KL=0,433) também teve uma importante influência no modelo de inferência. Acredita-se que esta variável pode ter mascarado a influência da variável *S*, o que justificaria seu baixo valor de divergência KL (0,094). A influência da variável *D* (KL=0,27) pode ter sido minimizada devido ao fato que a maior parte das áreas de cana-de-açúcar e as áreas aptas à sua expansão se concentram em declividades inferiores a 12%.

As variáveis *CEI* e *NDVI* não tiveram influência tão expressiva sobre o alvo. Conforme já explicitado, a discriminação entre as classes áreas aptas (*ZC*=1) e áreas de cana-de-açúcar (*ZC*=2) foi confusa devido à similaridade entre a distribuição dos valores de *CEI* e *NDVI* em função da variável alvo.

O índice *CEI* foi elaborado para capturar a sazonalidade de culturas anuais como a soja (RIZZI et al., 2009). Este índice foi aplicado à cana-de-açúcar com o propósito de capturar o comportamento da cultura, numa tentativa de adicionar um componente temporal de observação à modelagem. Diferente das culturas anuais (às vezes chamadas de culturas de verão), a cana-de-açúcar é uma cultura que permanece no campo por todo o ano agrícola – seu ciclo pode ser de 12 a 18 meses. Além disso, sua colheita vai de

abril a dezembro no estado de São Paulo (RUDORFF et al., 2010). Como diferentes talhões apresentam manejo distinto ao longo do tempo, é possível que o uso do CEI tenha sido ineficiente na tentativa de inferir sobre a presença da cana-de-açúcar no pixel. Outro fator é que as áreas aptas ao plantio da cana-de-açúcar incluem áreas agrícolas. Portanto, o comportamento de outras culturas pode também ter sido incorporado na variável *CEI*.

## 5.2. Mapa temático

As Bandas de Probabilidade, apresentadas nas Figuras 5.1 e 5.2, são o principal resultado do e-BayNeRD e podem ser utilizadas para diversas aplicações como, por exemplo, dados de entrada para diversos classificadores; podem servir de máscara para regiões com baixos valores de probabilidades; podem ser gerados mapas temáticos definindo-se regras a partir dos valores de probabilidade.

Conforme mencionado anteriormente, a partir das BPs das classes, o algoritmo calcula a BP para a classe não-alvo:  $BP_{\text{não-alvo}} = 1 - bp$ , onde  $bp$  corresponde à soma de todas as BPs das classes do alvo. Portanto, para esta aplicação serão geradas três Bandas de Probabilidades: BP0 – para não alvo; BP1 – áreas aptas à expansão; e BP2 – áreas de cana-de-açúcar.

Após empilhar as BPs, o e-BayNeRD analisa um determinado pixel em profundidade. Neste caso, o pixel receberá o rótulo da classe cuja Banda de Probabilidade tiver o maior valor de probabilidade. O resultado é o mapa temático apresentado na Figura 5.6.

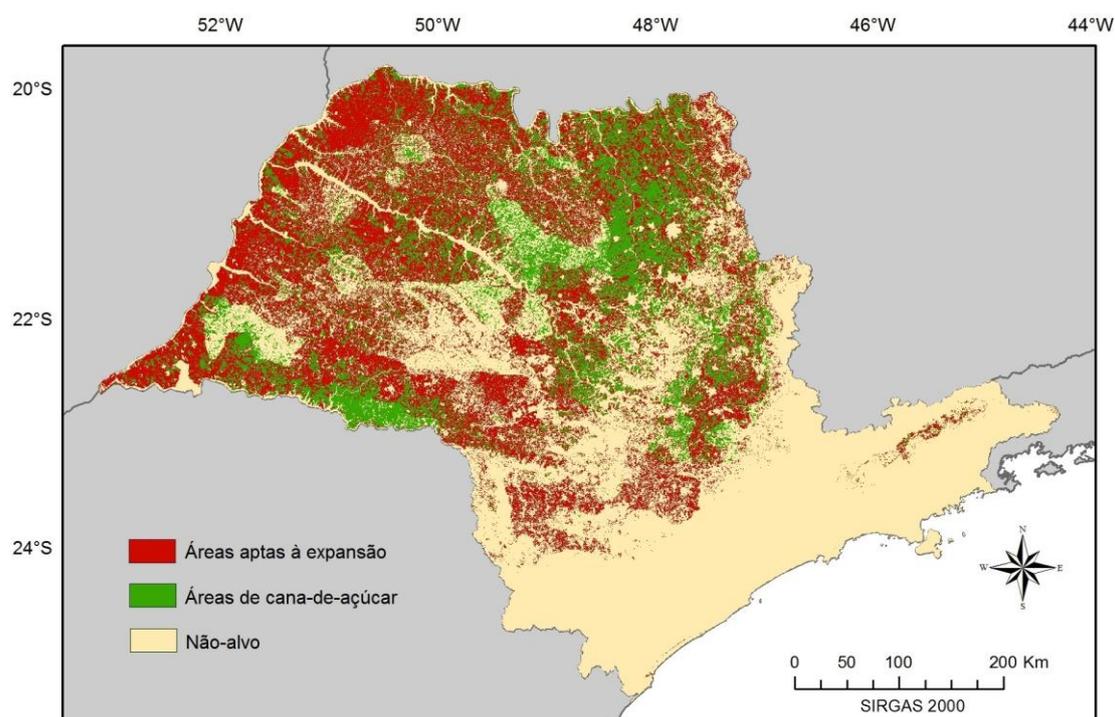


Figura 5.5 – Mapa temático criado a partir das Bandas de Probabilidade.

Antes de avaliar a classificação, a rede hidrográfica e as áreas urbanas do Estado de São Paulo foram mascaradas. Apesar da declividade estar relacionada com a hidrografia, o modelo de RB gerado pelo sistema de inferência e-BayNeRD não foi capaz de atribuir baixos valores de probabilidade a essas regiões. Cabe ressaltar que vários modelos de Redes Bayesianas foram testados com diferentes variáveis, além das descritas nesta aplicação como, por exemplo, a variável ‘distância do corpo d’água mais próximo’. De fato, quando esta variável foi incluída na RB, as regiões próximas aos rios e reservatórios receberam baixos valores de probabilidade. Situação semelhante ocorreu com as áreas urbanas, quando a variável ‘distância da área urbana mais próxima’ foi incluída na RB. Apesar da melhor discriminação destas áreas nas Bandas de Probabilidade, o resultado da classificação geral não teve bons resultados. Assim o melhor resultado obtido é o apresentado na Figura 5.6.

### 5.3. Avaliação das incertezas

O mapa temático gerado a partir das Bandas de Probabilidade foi avaliado por meio do índice de concordância Kappa ( $\kappa$ ). Segundo Congalton e Green (2009), a análise do

índice Kappa é usada na avaliação da acurácia da classificação para determinar, estatisticamente, se uma matriz de confusão é significativamente diferente de outra. No entanto, este é um índice de exatidão global, ou seja, reflete a condição geral de acertos ou erros e pode esconder defeitos e qualidades locais. Para avaliar as incertezas oriundas da classificação, uma simulação Monte Carlo foi realizada, na qual o processo de avaliação é replicado diversas vezes. Assim, cria-se um intervalo de credibilidade Kappa.

A cada iteração, 100 amostras aleatórias de cada classe ( $ZC=0$ ,  $ZC=1$ ,  $ZC=2$ ) foram selecionadas e construídas as matrizes de confusão. O número de amostras foi baseado na distribuição multinomial (CONGALTON; GREEN, 2009). O intervalo de credibilidade Kappa foi construído após a exclusão de 5% dos valores extremos de  $\kappa$ , 2,5% de máximos e 2,5% de mínimos. Isto significa que, em 95% das vezes em que o Kappa foi calculado com base nos dados utilizados como referência para a validação, seu valor estava no intervalo  $\kappa = [0,3, 0,46]$ . Similarmente, o intervalo para a acurácia geral da classificação obtido foi de  $[0,53; 0,64]$ . A Figura 5.7 apresenta a variação do índice Kappa e da acurácia geral ao longo de 1500 iterações, juntamente com o intervalo de credibilidade.

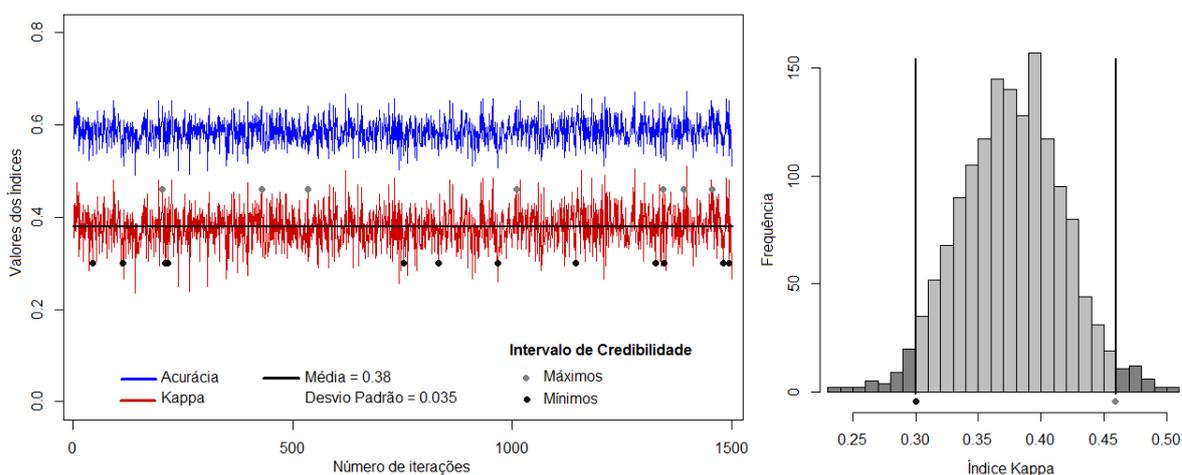


Figura 5.6 – Variação do índice Kappa e da acurácia geral. Simulação Monte Carlo para avaliação das incertezas da classificação (esquerda); Intervalo de Credibilidade (direita).

O intervalo de credibilidade Kappa apresentado é relativamente baixo, o que indica a ocorrência de confusões entre as classes. A Tabela 5.2 mostra a matriz de confusão que originou o  $\kappa_{max}$ . Nota-se que há uma confusão em relação às classes áreas aptas e áreas de cana, sendo que a primeira apresentou os maiores erros de omissão e inclusão. Ou seja, um número menor de amostras da classe áreas aptas foi classificado corretamente. No melhor caso,  $\kappa_{max}$ , 35% das amostras da classe áreas aptas foram classificadas como não-alvo e 13% se confundiu com áreas de cana-de-açúcar. Assim como a classe áreas aptas, a classe não-alvo é representada por regiões heterogêneas que incluem áreas urbanas, hidrografia, áreas de proteção ambiental, e outros.

Em relação a classe áreas de cana, para o  $\kappa_{max}$ , 25% das amostras de áreas de cana-de-açúcar foram classificadas como áreas aptas e 13% foram classificadas como não-alvo. A classe áreas de cana-de-açúcar se confundiu mais com a classe áreas aptas. Neste caso, a confusão maior entre as classes áreas de cana-de-açúcar e áreas aptas pode ser devido à inclusão de outras culturas agrícolas e pastagem nas áreas aptas que podem se confundir com a cana-de-açúcar. Neste ponto, pode-se dizer que o modelo multiclases para a aplicação proposta no presente trabalho não foi adequado para inferir os alvos com boa acurácia, considerando as observações feitas sobre as variáveis de contexto utilizadas.

Tabela 5.2 – Matriz de confusão com  $\kappa_{max}$  dentro do intervalo de credibilidade.

		Referência				Erro de Inclusão
		Não-alvo	Áreas aptas	Áreas cana	Total	
Classificação	Não-alvo	78	35	13	126	0,38
	Áreas aptas	17	52	25	94	0,45
	Áreas cana	5	13	62	80	0,18
	Total	100	100	100	300	
Erro de Omissão		0,22	0,48	0,38		

Segundo a análise da matriz de confusão e do intervalo de credibilidade Kappa, a acurácia da classificação foi baixa. Portanto, para analisar melhor este problema, criou-

se um mapa de incertezas para identificar e analisar regiões específicas da área de estudo.

As incertezas podem ser avaliadas comparando as regras de decisão que levaram a escolha de uma determinada classe. Segundo Wehenkel (1996), a entropia logarítmica ou Entropia de Shannon “é a base para várias interpretações no contexto de modelagem probabilística”. Suponha que  $\alpha$  seja uma variável aleatória e que possa assumir os valores  $a_1, a_2, \dots, a_n$ . Suponha também que  $p_i$  seja a probabilidade do evento  $\alpha = a_i$ . Assim, a Entropia de Shannon é definida como (HAMMER et al., 2000):

$$H(\alpha) = - \sum_i^n p_i * \log_2 p_i \quad (5.1)$$

O valor de entropia é máximo quando a probabilidade de um evento é a mesma dos demais. Assim, trazendo a Equação 5.1 para o contexto das Bandas de Probabilidade, incluindo a  $BP_{\text{não-alvo}}$ , aquelas regiões onde a classificação é mais duvidosa apresentarão máxima entropia, ou seja,  $P(ZC = 0) = P(ZC = 1) = P(ZC = 2) \cong 0, \bar{3}$ . O mapa de incertezas calculado a partir da entropia de Shannon é apresentado na Figura 5.8, onde as regiões mais claras indicam incerteza de classificação alta.

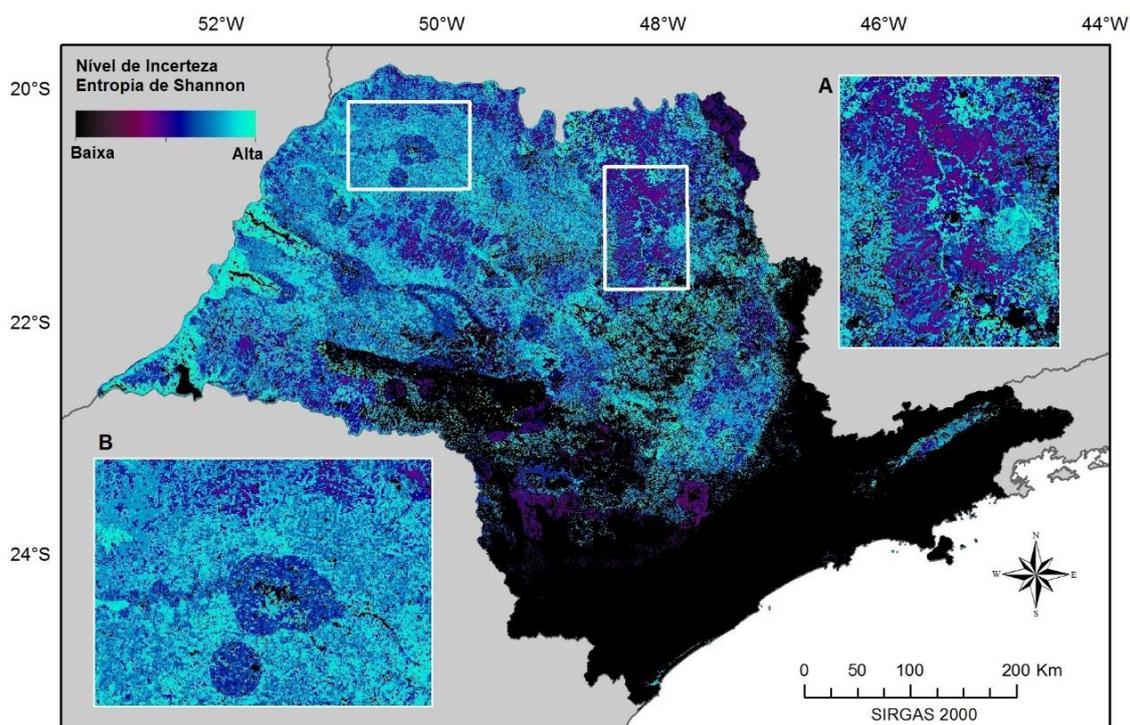


Figura 5.7 – Mapa de incertezas a partir da Entropia de Shannon. (A) Região com menores níveis de incerteza; (B) Região com maiores níveis de incerteza;

O mapa de incertezas corrobora o resultado obtido na análise das matrizes de confusão, onde as maiores incertezas na classificação estão na identificação de áreas aptas à expansão da cana-de-açúcar, conforme mostra a Figura 5.8-B. Apesar de os resultados sugerirem que o modelo e as variáveis utilizados na aplicação do e-BayNeRD não foram adequados para a correta identificação das áreas aptas à expansão da cana-de-açúcar no estado de São Paulo, existem argumentos a serem discutidos. O primeiro deles deve-se à limitação do algoritmo em gerar o mapa temático com base no maior valor de probabilidade encontrado nas BPs. A implementação de técnicas mais avançadas, com a definição de limiares mínimos de probabilidade para que determinado pixel fosse rotulado com determinada classe podem melhorar os valores dos índices de exatidão. Além disso, deve-se notar que os dados utilizados como referência possuem limitações inerentes. De fato, as áreas identificadas como cana-de-açúcar apresentaram melhores resultados e foram justamente as áreas em que a referência foi validada no Projeto Canasat (ADAMI et al., 2012). De qualquer forma o principal resultado do e-

BayNeRD é a IP, e esta apresentou-se coerente com os resultados encontrados na literatura, como discutido anteriormente (ver Figuras 5.1 e 5.2).

Uma parte considerável do território do estado de São Paulo apresenta valores altos de entropia, com grande incerteza na classificação. Na Figura 5.8, as áreas de não-alvo (regiões escuras no extremo leste) e algumas áreas aptas (regiões mais escuras no Centro-Sul) apresentaram baixos valores de entropia e, portanto, maior certeza na classificação. Pode-se também observar que as regiões, onde há predomínio de lavouras de cana-de-açúcar (regiões na cor roxo na Figura 5.8-A) também apresentaram baixos valores de entropia. Isto indica um resultado positivo da classificação ao definir essas regiões como áreas de cana-de-açúcar.

Dados os níveis de incerteza baixos das áreas de cana-de-açúcar, uma simulação Monte Carlo para avaliação local da região ilustrada na Figura 5.8-A foi realizada. Chegou-se ao intervalo de credibilidade Kappa  $\kappa = [0,54, 0,69]$  em 95% das vezes em que o índice foi calculado, como mostra a Figura 5.9. O intervalo de credibilidade Kappa obtido na avaliação local é melhor do que aquele obtido na avaliação de toda a área de estudo. Segundo a análise da matriz de confusão, apresentada na Tabela 5.3, pode-se observar que um número maior de amostras da classe área de cana-de-açúcar foi classificado corretamente e que houve também confusão entre as classes áreas aptas e áreas de cana-de-açúcar. Este comportamento é aceitável porque, nos dados de referência, áreas de agricultura e agropecuária estão representadas na classe áreas aptas, o que pode gerar confusão na classificação dos dois alvos.

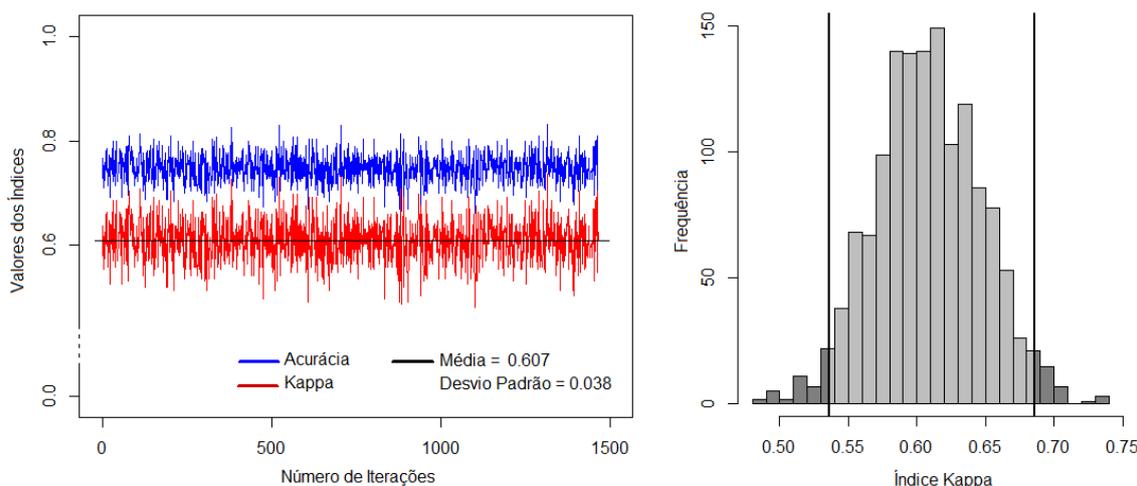


Figura 5.8 – Variação do índice Kappa e da acurácia geral da avaliação local. Simulação Monte Carlo para avaliação das incertezas da classificação (esquerda); Intervalo de Credibilidade (direita).

Tabela 5.3 – Matriz de confusão dentro do intervalo de credibilidade Kappa para avaliação local

		Referência				Erro de Inclusão
		Não-alvo	Áreas aptas	Áreas cana	Total	
Classificação	Não-alvo	68	7	2	77	0,12
	Áreas aptas	14	69	11	94	0,27
	Áreas cana	8	14	77	99	0,22
	Total	90	90	90	270	
Erro de Omissão		0,24	0,23	0,14		

Tendo em vista o melhor resultado da avaliação local da classificação de áreas de cana-de-açúcar, um novo modelo foi gerado, conforme apresentado na Figura 5.10, em que é ilustrado o modelo gráfico da RB juntamente com a discretização de cada variável de contexto. Neste modelo, considerou-se como referência apenas as áreas de cana-de-açúcar do Projeto Canasat (RUDORFF et al., 2010), conforme ilustrado na Figura 5.11, para avaliar o e-BayNeRD na inferência da classe cultivo da cana-de-açúcar.

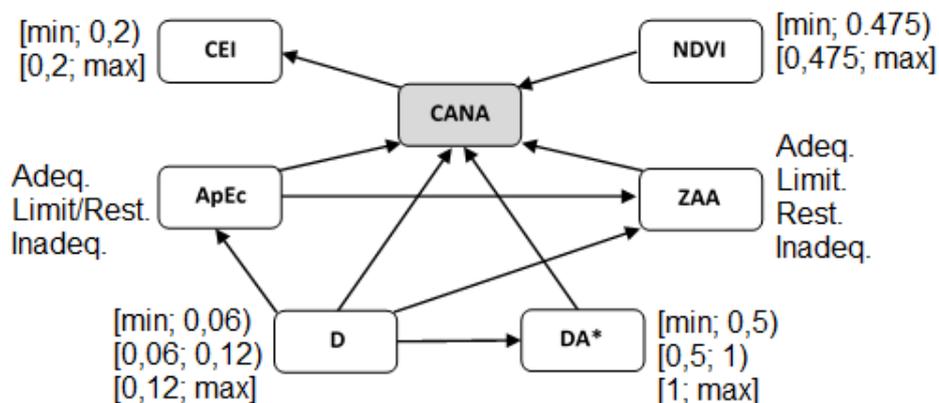


Figura 5.9 – Modelo gráfico da Rede Bayesiana para inferência de áreas de cana-de-açúcar e intervalos definidos no processo de discretização. \*DA - Variável ‘distância do corpo d’água mais próximo’ (em Km).

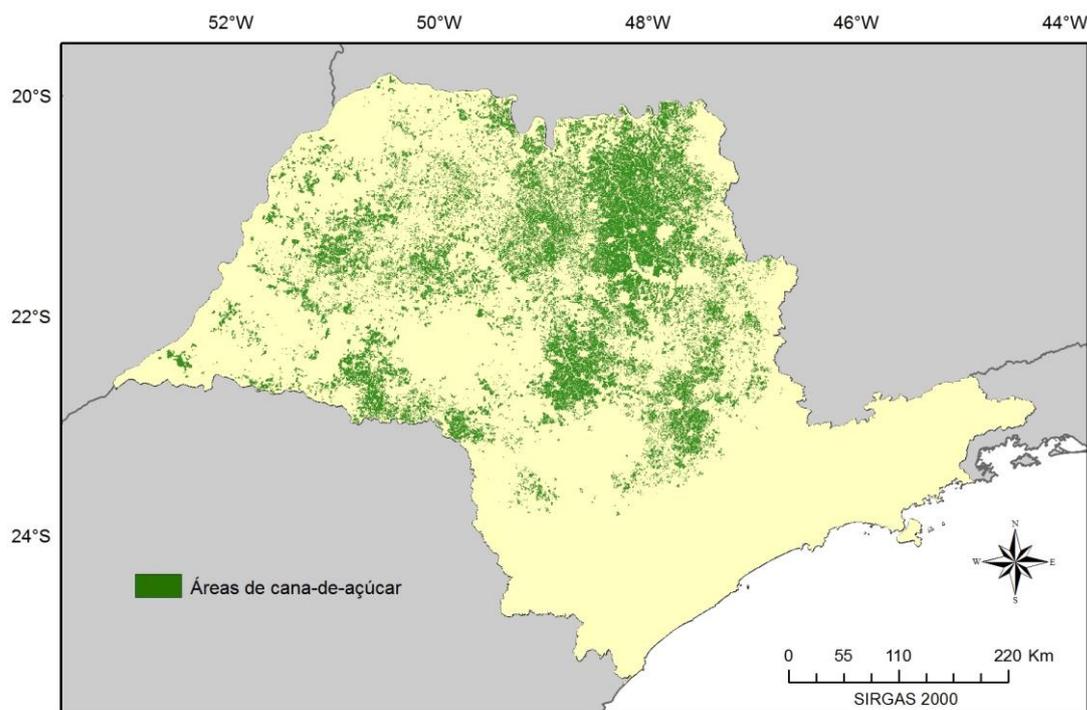


Figura 5.10 – Dados de referência para o segundo modelo: áreas de cana-de-açúcar segundo Projeto Canasat.

A Figura 5.12 mostra a Banda de Probabilidade resultante, onde as regiões com tonalidade verde escuro representam os maiores valores de probabilidade de ocorrência de áreas de cana-de-açúcar. A partir da BP, o mapa temático foi gerado definindo-se um limiar de probabilidade entre 0% e 100%, em que todos os pixels com valores de probabilidade acima de um determinado limiar foram classificados como áreas de cana-

de-açúcar. O objetivo é encontrar um limiar que gere um mapa temático mais adequado à classificação binária que indica presença e ausência de áreas de cana-de-açúcar. O mapa temático, ilustrado na Figura 5.13 foi gerado definindo-se um limiar de 0,34 e, selecionando-se aleatoriamente 100 amostras de cada classe, chegou-se à avaliação com um índice Kappa  $\kappa = 0,62$ , para toda a área de estudo (estado de São Paulo). Este resultado foi melhor do que o do modelo implementado para duas classes, cultivo da cana e aptidão para expansão.

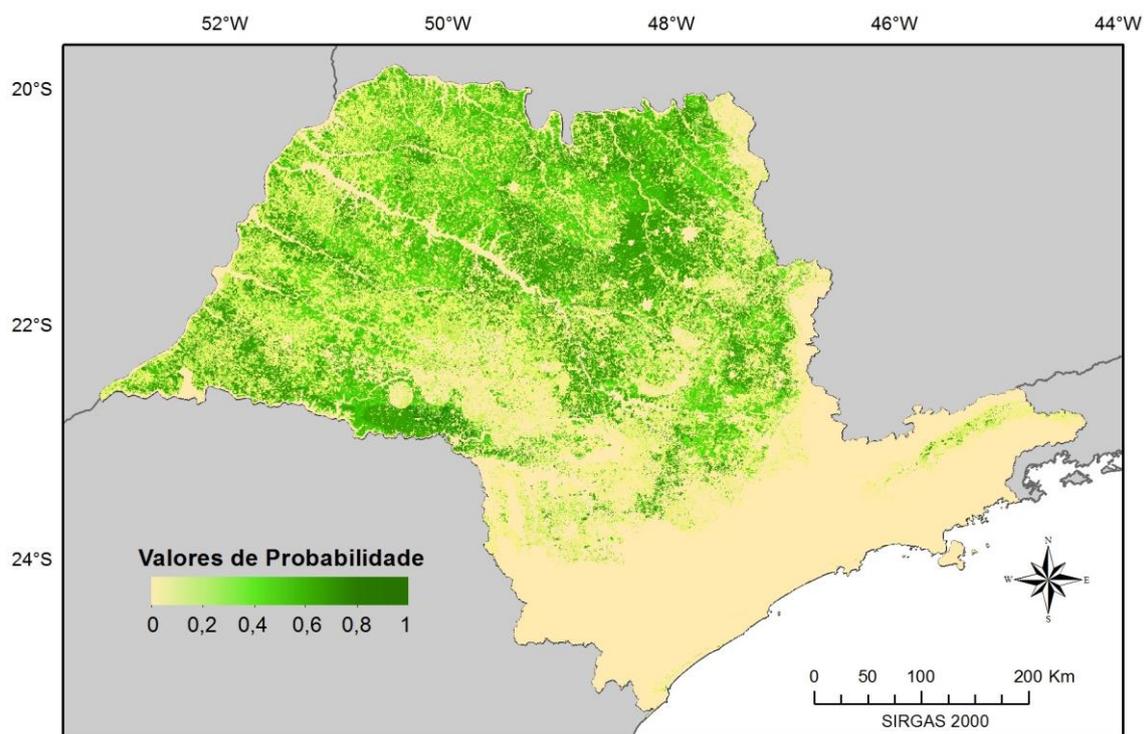


Figura 5.11 – Banda de Probabilidade para a classe áreas de cana-de-açúcar, resultante do segundo modelo RB.

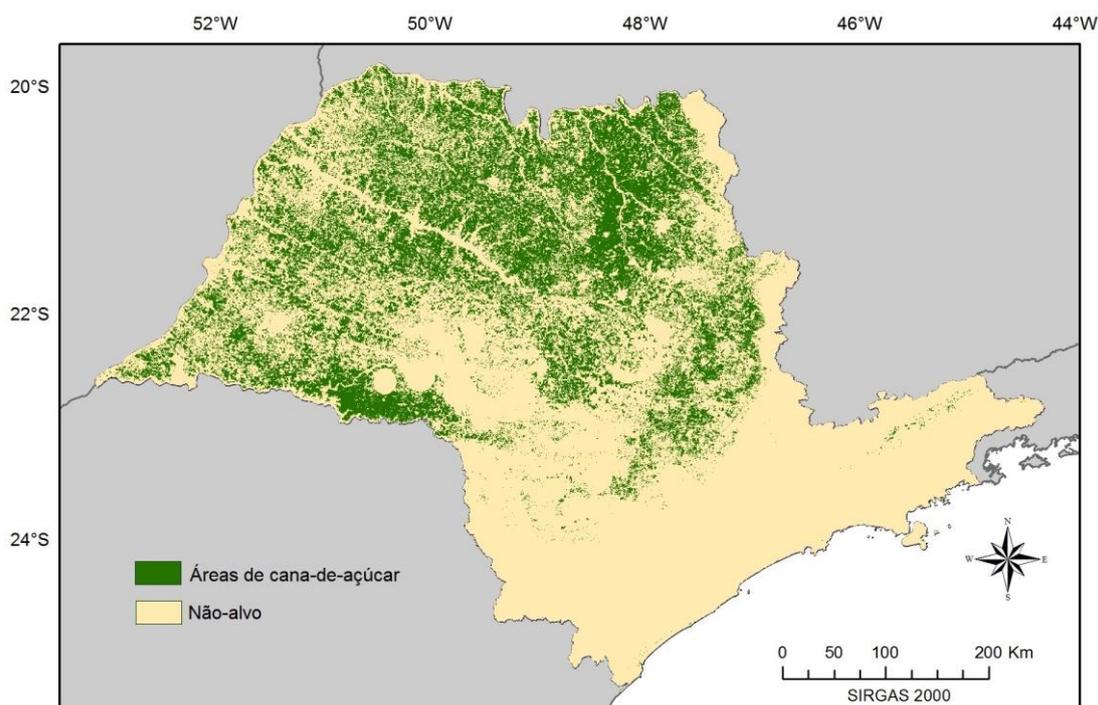


Figura 5.12 – Mapa Temático das áreas de cana-de-açúcar resultante do segundo modelo RB.

Pode-se observar que, no mapa temático, há maior inclusão de áreas de cana-de-açúcar, quando comparado aos dados de referência. Isto pode ser explicado pelo fato do modelo de inferência ter utilizado uma imagem NDVI de uma data específica e o índice CEI que pode ter incorporado o comportamento de outras culturas, conforme discutido anteriormente. Neste caso, áreas agrícolas podem ter sido classificadas como cana-de-açúcar.

Por outro lado, o mapeamento de áreas de cana-de-açúcar realizado pelo Projeto Canasat (RUDORFF et al., 2010) envolve processos de interpretação visual e análise de séries temporais, que produzem mapas de classificação com maior acurácia em torno de 98% (ADAMI et al., 2012). Uma forma de melhorar a acurácia da classificação da cana-de-açúcar poderia ser a inclusão de variáveis que modelam o comportamento do ciclo do cultivo da cana no modelo RB.

Similarmente ao modelo gerado para inferência apenas de áreas de cana-de-açúcar, também foi criado um modelo para inferência apenas das áreas aptas à expansão da cultura, conforme apresentado na Figura 5.14, em que é ilustrado o modelo gráfico da

RB juntamente com a discretização de cada variável de contexto. Neste modelo, considerou-se como referência apenas as áreas aptas, conforme mostra a Figura 5.15.

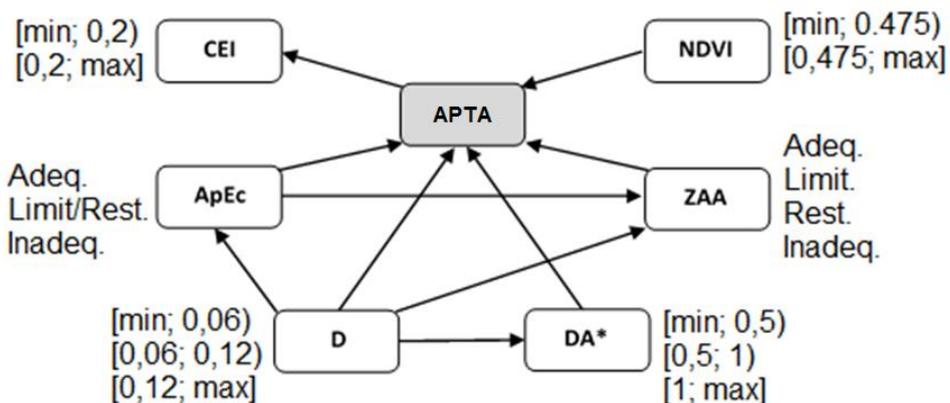


Figura 5.13 – Modelo gráfico da Rede Bayesiana para inferência de áreas aptas à expansão da cana-de-açúcar e intervalos definidos no processo de discretização. \*DA - Variável 'distância do corpo d'água mais próximo' (em Km)

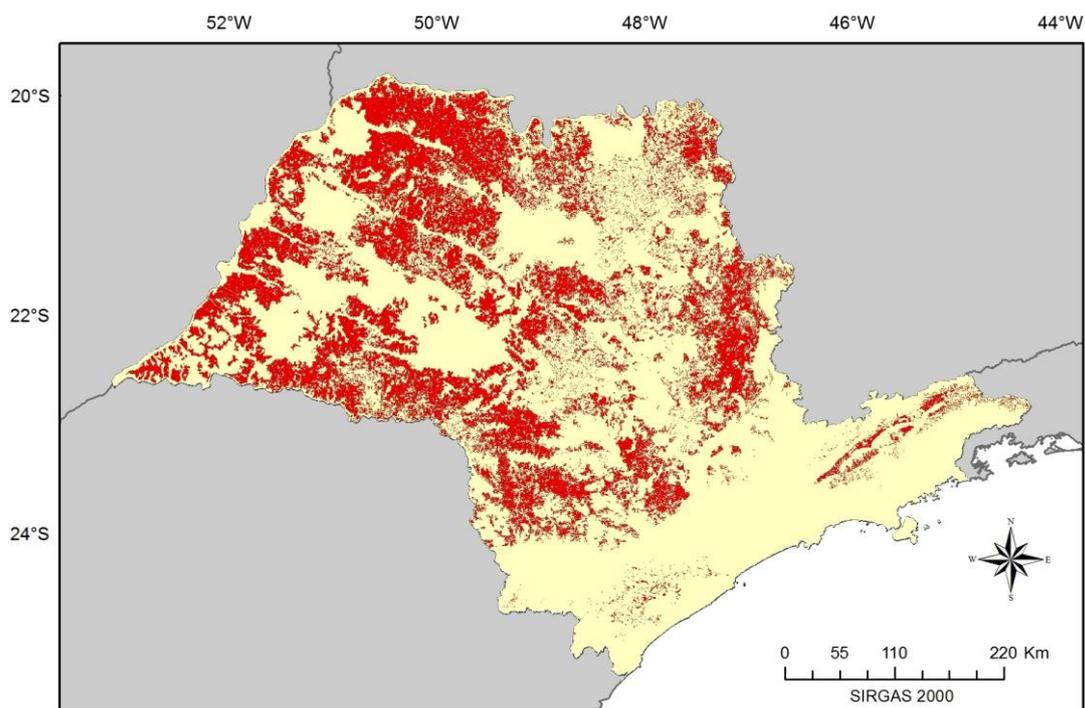


Figura 5.14 – Dados de referência para o terceiro modelo: áreas aptas ao plantio da cana-de-açúcar segundo o Zoneamento Agroecológico.

A Figura 5.16 mostra a Banda de Probabilidade resultante deste modelo, onde as regiões com tonalidade vermelho representam os maiores valores de probabilidade de

ocorrência de áreas aptas à expansão da cana-de-açúcar. De modo semelhante ao que foi discutido no modelo anterior, o mapa temático, ilustrado na Figura 5.17, foi gerado definindo-se um limiar de probabilidade. Nota-se que houve uma grande inclusão de áreas aptas. Isto pode ser explicado pelo baixo limiar selecionado de 0,29 no qual foi encontrado o melhor índice Kappa  $\kappa = 0,45$ . Após analisar os resultados, observou-se que os dados usados como referência apresentam alguns erros que podem ter levado ao baixo valor de Kappa nos mapas temáticos. A Figura 5.18 ilustra um recorte ampliado dos mapas temático e de referência do primeiro modelo RB (Figura 5.6) onde a classe áreas de cana-de-açúcar apresentam os melhores resultados de classificação. Os polígonos indicados nesta figura correspondem às áreas de preservação segundo o Ministério do Meio Ambiente (<http://mapas.mma.gov.br/i3geo/datadownload.htm>). Pode-se observar que, os dados do Zoneamento Agroecológico da Cana-de-açúcar, utilizados como mapa de referência, considera como áreas aptas (em vermelho) regiões que na verdade são áreas de preservação.

Da mesma forma, a Figura 5.19 mostra os polígonos de áreas de preservação sobrepostos ao mapa temático e de referência, para uma região onde houve maiores níveis de incerteza na classificação. O mesmo erro pode ser observado neste recorte, onde áreas de preservação ambiental podem ser observadas como áreas aptas à expansão da cana. Tendo isto em vista, pode-se concluir que os erros do mapa de referência prejudicaram a avaliação dos mapas temáticos produzidos pelo modelo RB, proposto no trabalho. Portanto, ele não pode ser usado como verdade na avaliação da classificação. É necessário usar dados de trabalho de campo para validar o método.

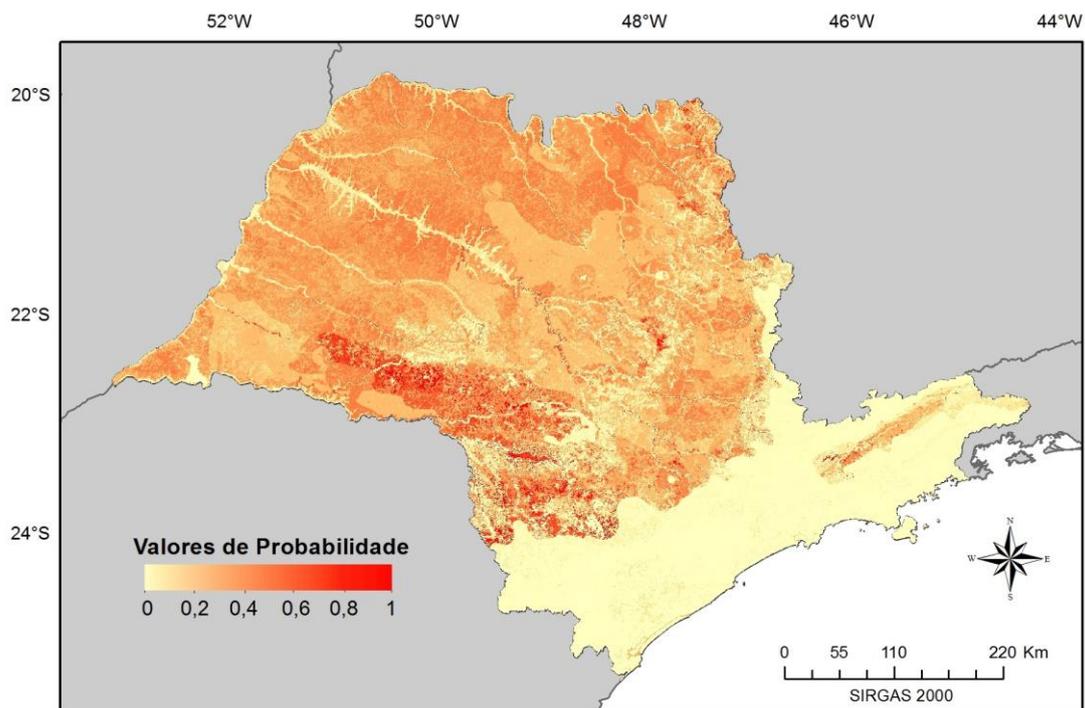


Figura 5.15 – Banda de Probabilidade para a classe áreas aptas, resultante do terceiro modelo RB

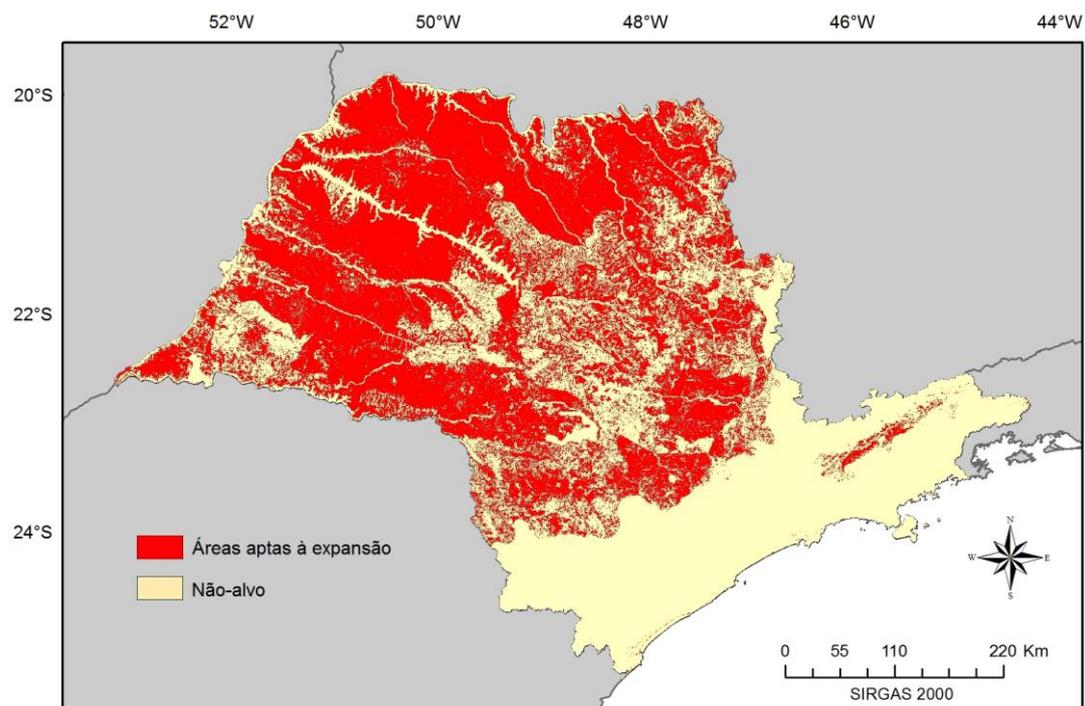


Figura 5.16 - Mapa Temático das áreas apta, resultante do terceiro modelo RB

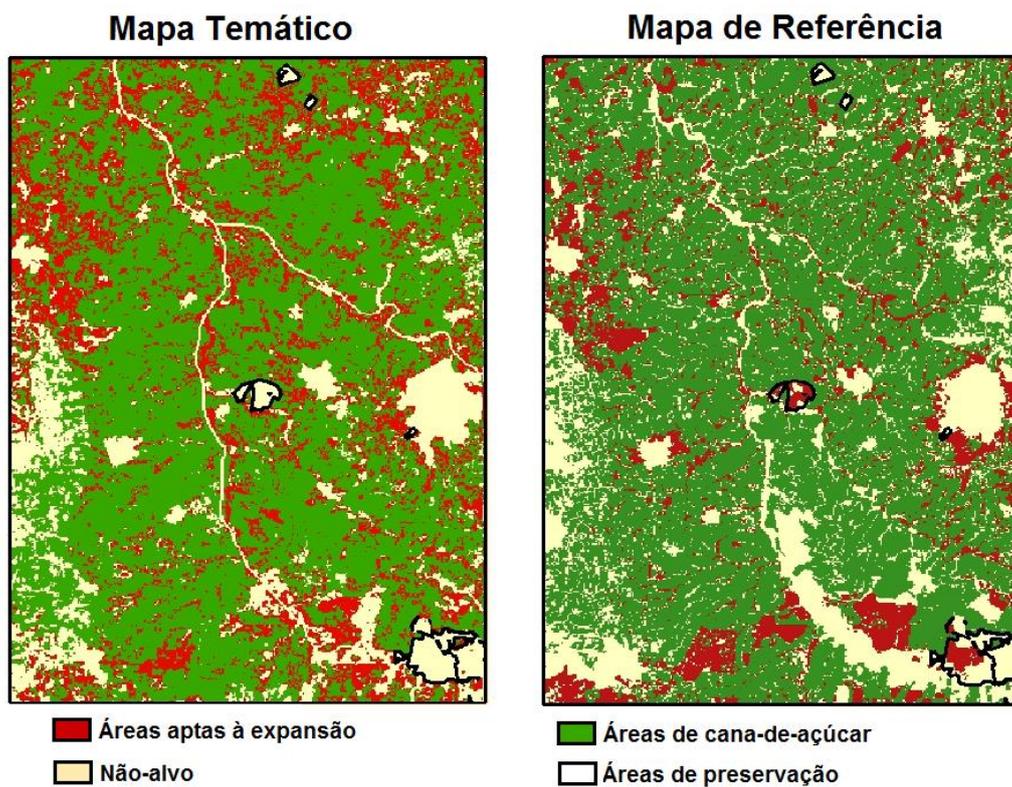


Figura 5.17 – Recorte ampliado do Mapa Temático do primeiro modelo gerado e dos dados de Referência, onde a classe áreas de cana-de-açúcar apresenta os melhores resultados de classificação.

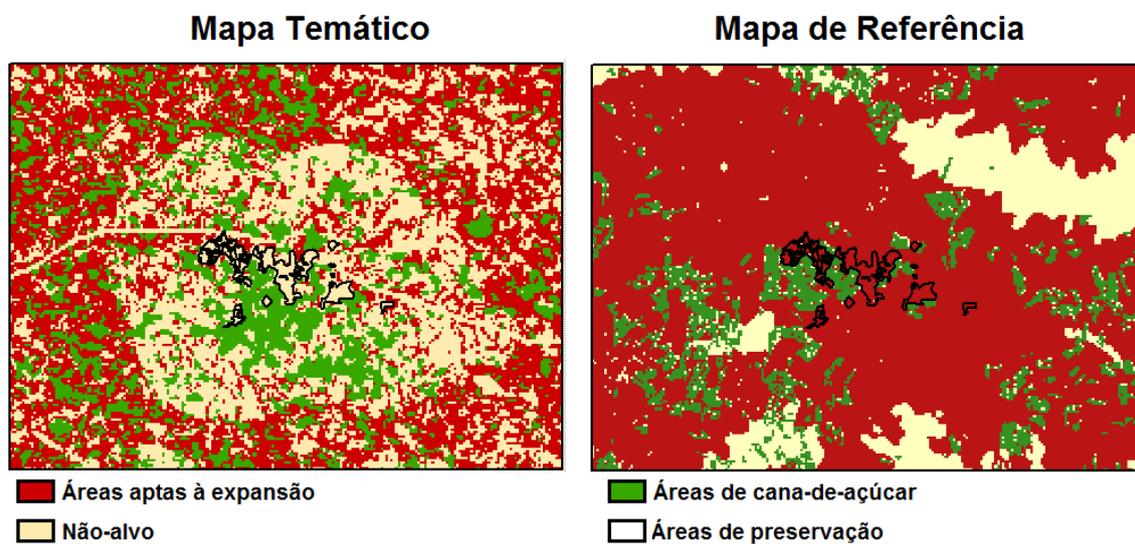


Figura 5.18 – Recorte ampliado do Mapa Temático do primeiro modelo gerado e dos dados de Referência, em uma região onde a classe áreas aptas apresentou altos níveis de incerteza.

## 6 CONCLUSÃO E TRABALHOS FUTUROS

O presente trabalho descreveu as melhorias que foram realizadas no algoritmo BayNeRD, desenvolvido por Mello et al. (2013). A partir destas implementações, criou-se uma versão aprimorada nomeada como *Enhanced Bayesian Network for Raster Data* – e-BayNeRD.

O algoritmo foi aprimorado para melhorar a manipulação de dados *raster*, bem como para minimizar a etapa de pré-processamento a ser realizada pelo usuário. Com estas modificações, o e-BayNeRD é capaz de manipular conjuntos de dados *raster* com diferentes sistemas de coordenadas, diferentes tamanhos de pixel e diferentes extensões. A implementação de uma interface para a definição do modelo gráfico da Rede Bayesiana tornou o sistema de inferência mais eficiente e amigável para o usuário. A capacidade de modelar mais de duas classes para a variável alvo também foi implementada no sistema. O algoritmo calcula a probabilidade de ocorrência de cada classe, considerando as observações feitas sobre as variáveis de contexto. Desta forma, no sistema e-BayNeRD, a probabilidade é calculada para todos os pixels, produzindo como resultado Bandas de Probabilidade que são integradas para gerar uma Imagem de Probabilidade de saída.

Neste trabalho, o sistema e-BayNeRD foi aplicado para mapear áreas de cana-de-açúcar e áreas aptas à sua expansão no estado de São Paulo. O resultado da classificação multiclassificada apresentou um intervalo de credibilidade Kappa baixo, principalmente, devido à confusão entre as classes áreas aptas à expansão e não-alvo, assim como à confusão entre as classes áreas de cana e áreas aptas. As áreas aptas e as áreas não-alvo são heterogêneas e incluem diferentes tipos de alvos, o que dificulta a modelagem destas classes. Além disso, as áreas de agricultura e pastagem inclusas nas áreas aptas, podem ter se confundido com as áreas de cana-de-açúcar. Tanto as áreas aptas quanto as áreas de cana localizam-se em regiões de baixa declividade, de características semelhantes de solo e clima, dificultando, assim, a classificação.

O mapa de incertezas indicou o maior nível de incerteza na classificação das áreas aptas do que as áreas de cana-de-açúcar. Devido à grande extensão da área de estudo, uma

avaliação local da classificação foi realizada para analisar acertos ou erros localmente. O resultado indicou maior acurácia do mapa temático, com intervalo de credibilidade igual a [0,54, 0,69], calculado para uma região que apresentou menor nível de incerteza da classe áreas de cana.

Tendo em vista os resultados das análises, um novo modelo RB para inferir, tanto a presença de áreas aptas quanto à presença de áreas de cana-de-açúcar, foi criado. O modelo para cana-de-açúcar resultou em um mapa temático com índice Kappa de 0,62, melhor do que o resultado da classificação multiclases. Enquanto que o modelo para áreas aptas resultou um mapa temático com baixo índice Kappa de 0,45.

Verificou-se que os dados de referência utilizados para avaliar os resultados obtidos pela RB continham erros que prejudicaram a avaliação. Assim, recomenda-se o uso de dados de trabalho de campo para validar o modelo RB proposto nesta pesquisa. Contudo, o principal resultado do e-BayNeRD é a Imagem de Probabilidade (IP) e esta apresentou resultados coerentes com aqueles apresentados e discutidos na literatura (RUDORFF et al., 2010). O sistema e-BayNeRD apresenta potencial para modelagem de sistemas complexos em diversas aplicações na área de observação da Terra como: indicar áreas com maior probabilidade de deslizamento; indicar regiões mais propícias ao desmatamento e queimada; indicar pontos de derramamento de óleo nos mares; dentre outras aplicações relacionadas aos desastres ambientais.

Tendo em vista os resultados e os problemas identificados na aplicação proposta, algumas sugestões para melhorar o modelo RB e o sistema e-BayNeRD são:

- incluir características espaciais, como a influência de pixels vizinhos, no cálculo dos valores de probabilidade;
- incluir, como variáveis de contexto, atributos temporais que caracterizem melhor o comportamento da cana-de-açúcar ao longo de seu período de cultivo;

- incorporar variáveis políticas e socioeconômicas que, segundo Dr. Dalton<sup>2</sup>, foram utilizadas na geração do mapa de Zoneamento Agroecológico da Cana-de-açúcar.
- incorporar o cálculo do mapa de incertezas, tornando-o também um produto do e-BayNeRD
- implementar definição de limiares mínimos de probabilidade para a classificação dos pixels.
- Usar dados de campo para validar o modelo proposto neste trabalho.

---

<sup>2</sup> Consulta pessoal.



## REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMI, M.; MELLO, M. P.; AGUIAR, D. A.; RUDORFF, B. F. T.; DE SOUZA, A. F. A web platform development to perform thematic accuracy assessment of sugarcane mapping in South-Central Brazil. **Remote Sensing**, v. 4, p. 3201–3214, 2012.
- ADAMI, M.; RUDORFF, B. F. T.; FREITAS, R. M.; AGUIAR, D. A.; SUGAWARA, L. M.; MELLO, M. P. Remote Sensing Time Series to Evaluate Direct Land Use Change of Recent Expanded Sugarcane Crop in Brazil. **Sustainability**, v. 4, n. 12, p. 574–585, 2012.
- AGUIAR, D. A.; RUDORFF, B. F. T.; SILVA, W. F.; ADAMI, M.; MELLO, M. P. Remote Sensing Images in Support of Environmental Protocol: Monitoring the Sugarcane Harvest in São Paulo State, Brazil. **Remote Sensing**, v. 3, n. 12, p. 2682–2703, 2011.
- AGUILERA, P. A.; FERNÁNDEZ, A.; FERNÁNDEZ, R.; RUMÍ, R.; SALMERÓN, A. Bayesian networks in environmental modelling. **Environmental Modelling & Software**, v. 26, n. 12, p. 1376–1388, 2011.
- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL). **Atlas de Energia Elétrica do Brasil**. 3. ed. Brasília: ANEEL, 2008. 236 p. ISBN (978-85-87491-10-7). Disponível em: <[http://www.aneel.gov.br/visualizar\\_texto.cfm?idtxt=1687](http://www.aneel.gov.br/visualizar_texto.cfm?idtxt=1687)>. Acesso em: 07 ago. 2014.
- AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS (ANP). **Panorama atual da Indústria Brasileira de Álcool Combustível: avaliação do arcabouço regulatório da ANP referente à movimentação do produto**. Rio de Janeiro: ANP, [s.n.], 2007. 33 p.
- ARAÚJO, G. J. F.; NAVARRO, L. F. S.; SANTOS, B. A. S. O Etanol de segunda geração e sua importância estratégica ante o cenário energética internacional contemporâneo. In: FÓRUM AMBIENTAL DA ALTA PAULISTA, 9, 2013, [S.l.]. **Anais...** 2013. p. 1–11. Disponível em: <[http://amigosdanatureza.org.br/publicacoes/index.php/forum\\_ambiental/article/view/492](http://amigosdanatureza.org.br/publicacoes/index.php/forum_ambiental/article/view/492)>. Acesso em: 07 ago. 2014.
- BIVAND, R.; KEITT, T.; ROWLINGSON, B. **rgdal: bindings for the geospatial data abstraction library**. 2014. Disponível em: <<http://cran.r-project.org/package=rgdal>>. Acesso em: 19 abr. 2014.
- BOLTON, D. K.; FRIEDL, M. A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. **Agricultural and Forest Meteorology**, v. 173, p. 74–84, 2013.

BORDONAL, R. DE O.; BARRETTO DE FIGUEIREDO, E.; AGUIAR, D. A.; ADAMI, M.; THEODOR RUDORFF, B. F.; LA SCALA, N. Greenhouse gas mitigation potential from green harvested sugarcane scenarios in São Paulo State, Brazil. **Biomass and Bioenergy**, v. 59, p. 195–207, 2013.

BOTTCHER, S. G.; DETHLEFSEN, C. **deal**: learning bayesian networks with mixed variables. 2013. Disponível em: <<http://cran.r-project.org/package=deal>>. Acesso em: 22 mai. 2014.

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB). **Safras - séries históricas**. Disponível em: <[http://www.conab.gov.br/conteudos.php?a=1252&t=2&Pagina\\_objcmsconteudos=1#A\\_objcmsconteudos](http://www.conab.gov.br/conteudos.php?a=1252&t=2&Pagina_objcmsconteudos=1#A_objcmsconteudos)>. Acesso em: 31 out. 2014.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 2. ed. [S.l.]: CRC Press/Taylor & Francis, 2009. 183 p. ISBN (978-1-4200-5512-2).

DAOUDI, K.; FOHR, D.; ANTOINE, C. Dynamic Bayesian networks for multi-band automatic speech recognition. **Computer Speech & Language**, v. 17, n. 2-3, p. 263–285, 2003.

DIAS, M. O. S.; JUNQUEIRA, T. L.; JESUS, C. D. F.; ROSSELL, C. E. V.; MACIEL FILHO, R.; BONOMI, A. Improving second generation ethanol production through optimization of first generation production process from sugarcane. **Energy**, v. 43, n. 1, p. 246–252, 2012.

EGESKOG, A.; FREITAS, F.; BERNDDES, G.; SPAROVEK, G.; WIRSENIUS, S. Greenhouse gas balances and land use changes associated with the planned expansion (to 2020) of the sugarcane ethanol industry in Sao Paulo, Brazil. **Biomass and Bioenergy**, v. 63, p. 280–290, 2014.

EMPRESA DE PESQUISA ENERGÉTICA (EPE). **Balanco energético nacional 2014**: Ano base 2013. Rio de Janeiro: EPE, 2014. 288 p. Disponível em: <[https://ben.epe.gov.br/downloads/Relatorio\\_Final\\_BEN\\_2014.pdf](https://ben.epe.gov.br/downloads/Relatorio_Final_BEN_2014.pdf)>. Acesso em: 15 jan. 2015.

FIGUEIREDO, E. B. DE; SCALA JR, N. LA. Greenhouse gas balance due to the conversion of sugarcane areas from burned to green harvest in Brazil. **Agriculture, Ecosystems & Environment**, v. 141, n. 1-2, p. 77–85, 2011.

GOLDEMBERG, J.; COELHO, S. T.; GUARDABASSI, P. The sustainability of ethanol production from sugarcane. **Energy Policy**, v. 36, n. 6, p. 2086–2097, jun. 2008.

GOMES, P.; MONTAÑO, M. **A expansão da cana-de-açúcar eo zoneamento agroambiental do setor sucroalcooleiro do estado de São Paulo**. 2012, [S.l.: s.n.], 2012. Disponível em: <[http://avaliacaodeimpacto.org.br/wp-content/uploads/2012/10/068\\_zoneamento.pdf](http://avaliacaodeimpacto.org.br/wp-content/uploads/2012/10/068_zoneamento.pdf)>. Acesso em: 13 jan. 2015.

HAHSLER, M.; BUCHTA, C.; GRUEN, B.; HORNIK, K. **arules: mining association rules and frequent item sets**. 2014 Disponível em: <<http://cran.r-project.org/package=arules>>. Acesso em: 05 jun. 2014.

HAMMER, D.; ROMASHCHENKO, A.; SHEN, A.; VERESHCHAGIN, N. Inequalities for Shannon entropy and kolmogorov complexity. **Journal of Computer and System Sciences**, v. 60, p. 442–464, 2000.

HIJMANS, R. J. **Introduction to the “raster” package (version 2.2-31)**. 2014. Disponível em: <<http://cran.r-project.org/web/packages/raster/vignettes/Raster.pdf>>. Acesso em: 19 abr. 2014.

HIJMANS, R. J.; ETTEN, J. VAN; MATTIUZZI, M.; SUMNER, M.; GREENBERG, J. A.; LAMIGUEIRO, O. P.; BEVAN, A.; RACINE, E. B.; SHORTRIDGE, A. **raster: Geographic data analysis and modeling. R package version 2.2-12**. 2014. Disponível em: <<http://cran.r-project.org/web/packages/raster/>>. Acesso em: 19 abr. 2014.

HOLLANDA, J. B. DE. **Uma revolução energética para a cana de açúcar**. Rio de Janeiro: INEE - Instituto Nacional de Eficiência Energética. Disponível em: <[http://www.inee.org.br/biomassa\\_downloads.asp?Cat=biomassa](http://www.inee.org.br/biomassa_downloads.asp?Cat=biomassa)>. Acesso em: 20 set. 2014.

HOLLANDA, J. B. DE; ERBER, P. Energy Efficiency in Brazil. **Trade and Environment Review 2009/2010**, p. 68–77, 2010. Disponível em: <[http://www.inee.org.br/biomassa\\_downloads.asp?Cat=biomassa](http://www.inee.org.br/biomassa_downloads.asp?Cat=biomassa)>. Acesso em: 20 set. 2014.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, p. 195–213, 2002.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Área plantada/colhida - lavouras permanentes**. Disponível em: <[www.sidra.ibge.gov.br](http://www.sidra.ibge.gov.br)>. Acesso em 8 fev. 2015.

JUSTICE, C.; TOWNSHEND, J. R.; VERMOTE, E.; MASUOKA, E.; WOLFE, R.; SALEOUS, N.; ROY, D.; MORISETTE, J. An overview of MODIS Land data processing and product status. **Remote Sensing of Environment**, v. 83, n. 1-2, p. 3–15, 2002.

KAHN, C. E.; ROBERTS, L. M.; SHAFFER, K. A.; HADDAWY, P. Construction of a Bayesian network for mammographic diagnosis of breast cancer. **Computers in Biology and Medicine**, v. 27, n. 1, p. 19–29, 1997.

KOHLHEPP, G. Análise da situação da produção de etanol e biodiesel no Brasil. **Estudos Avançados**, v. 24, n. 68, p. 223–253, 2010. Disponível em: <<http://www.revistas.usp.br/eav/article/view/10476>>. Acesso em: 09 ago. 2014.

KULLBACK, S.; LEIBLER, R. On information and sufficiency. **The Annals of Mathematical Statistics**, v. 22, n. 1, p. 79–86, 1951. Disponível em: <<http://www.jstor.org/stable/2236703>>. Acesso em: 15 jan. 2015.

LANDUYT, D.; BROEKX, S.; D'HONDT, R.; ENGELEN, G.; AERTSENS, J.; GOETHALS, P. L. M. A review of Bayesian belief networks in ecosystem service modelling. **Environmental Modelling & Software**, v. 46, p. 1–11, 2013.

LORA, E. S.; ANDRADE, R. V. Biomass as energy source in Brazil. **Renewable and Sustainable Energy Reviews**, v. 13, n. 4, p. 777–788, 2009.

MANZATTO, C. V.; ASSAD, E. D.; BACCA, J. F. M.; ZARONI, M. J.; PEREIRA, S. E. M. (Orgs.). **Documentos 110** - Zoneamento agroecológico da cana-de açúcar - expandir a produção , preservar a vida , garantir o futuro. . Rio de Janeiro: EMBRAPA SOLOS, 2009. Disponível em: <[http://www.cnps.embrapa.br/zoneamento\\_cana\\_de\\_acucar/ZonCana.pdf](http://www.cnps.embrapa.br/zoneamento_cana_de_acucar/ZonCana.pdf)>.

MANZATTO, C. V.; BACA, J. F. M.; PEREIRA, S. E. M.; ASSAD, E. D.; MEIRELLES, M. S. P.; BACA, A. F. M.; NAIME, U. J.; MOTTA, P. E. F. DA. Zoneamento agroecológico da cana-de-açúcar: abordagem metodológica para integração temática de grandes áreas territoriais. In: PRADO, R. B.; TURETTA, A. P. D.; ANDRADE, A. G. (Orgs.). **Manejo e conservação do solo e da água no contexto das mudanças ambientais**. 1. ed. Rio de Janeiro: Embrapa Solos, 2010. p. 486..

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO (MAPA). **Brasil projeções do agronegócio 2010/2011 a 2020/2021**. Brasília: MAPA, 2011. 58 p. Disponível em: <[http://www.agricultura.gov.br/arq\\_editor/file/Ministerio/gestao/projecao/PROJECOES\\_DO\\_AGRONEGOCIO\\_2010-11\\_a\\_2020-21\\_-\\_2\\_0.pdf](http://www.agricultura.gov.br/arq_editor/file/Ministerio/gestao/projecao/PROJECOES_DO_AGRONEGOCIO_2010-11_a_2020-21_-_2_0.pdf)>. Acesso em: 09 ago. 2014.

MAS, J. F.; FLORES, J. J. The application of artificial neural networks to the analysis of remotely sensed data. **International Journal of Remote Sensing**, v. 29, n. 3, p. 617–663, 2008.

MELGANI, F.; BRUZZONE, L. Classification of hyperspectral remote sensing images with support vector machines. **IEEE Transactions on Geoscience and Remote Sensing**, v. 42, n. 8, p. 1778–1790, 2004.

MELLO, M. P.; RISSO, J.; ATZBERGER, C.; APLIN, P.; PEBESMA, E.; VIEIRA, C. A. O.; RUDORFF, B. F. T. Bayesian Networks for Raster Data (BayNeRD): Plausible Reasoning from Observations. **Remote Sensing**, v. 5, n. 11, p. 5999–6025, 15 nov. 2013. Disponível em: <<http://www.mdpi.com/2072-4292/5/11/5999/>>. Acesso em: 19 jan. 2014.

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. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2010, Honolulu. **Proceedings...** 2010. p. 351-354. DVD; On-line. ISBN 978-1-4244-9564-1, 978-1-4244-9565-8. ISSN 2153-6996. doi: <[10.1109/IGARSS.2010.5651814](https://doi.org/10.1109/IGARSS.2010.5651814)>.

MIURA, A. K. **Identificação de áreas potenciais para cultivos destinados à produção de bioenergia com apoio de geotecnologias**. 2012. 152 p. (sid.inpe.br/mtc-m19/2012/05.10.18.50-TDI). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012. Disponível em: <<http://urlib.net/8JMKD3MGP7W/3BS8BU5>>. Acesso em: 08 abr. de 2014.

MINISTÉRIO DO MEIO AMBIENTE (MMA). **Biomassa**. Brasília, [2015]. Disponível em: <<http://www.mma.gov.br/clima/energia/energias-renovaveis/item/118>>. Acesso em: 12 ago. 2014.

MINISTÉRIO DE MINAS E ENERGIA (MME). **Matriz energética nacional 2030**. Brasília: MME, 2007. 254 p. Disponível em: <<http://www.biblioteca.presidencia.gov.br/publicacoes-oficiais-1/catalogo/conselhos/conselho-nacional-de-politica-energetica/matriz-energetica-nacional-2030/view>>. Acesso em: 13 ago. 2014.

MORAES, M. A. F. D.; NASSAR, A. M.; MOURA, P.; LEAL, R. L. V.; CORTEZ, L. A. B. Jet biofuels in Brazil: Sustainability challenges. **Renewable and Sustainable Energy Reviews**, v. 40, p. 716–726, 2014.

NEAPOLITAN, R. E. **Learning Bayesian networks**. 2. ed. New Jersey: Prentice Hall, 2004. 674 p. ISBN (0-13-012534-2).

ONIŠKO, A.; DRUZDZEL, M. J. Impact of precision of Bayesian network parameters on accuracy of medical diagnostic systems. **Artificial intelligence in medicine**, v. 57, n. 3, p. 197–206, 2013.

OZTURK, I.; ASLAN, A.; KALYONCU, H. Energy consumption and economic growth relationship: Evidence from panel data for low and middle income countries. **Energy Policy**, v. 38, n. 8, p. 4422–4428, 2010.

PARLAMENTO EUROPEU E DO CONSELHO. Directiva 2009/28/Ce. **Jornal Oficial da União Europeia**, v. 2008, n. 2, p. 16–62, 2009.

- PEARL, J.; RUSSELL, S. **Bayesian networks**. Department of Statistics Papers. California: Department of Statistics, UCLA. Disponível em: <<https://escholarship.org/uc/item/53n4f34m>>. Acesso em: 07 jun. 2014.
- QIN, D.; JIANWEN, M.; YUN, O. Y. Remote sensing data change detection based on the CI test of Bayesian networks. **Computers & Geosciences**, v. 32, n. 2, p. 195–202, 2006.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. 2014. Vienna: R Foundation for Statistical Computing. Disponível em: <<http://www.r-project.org/>>. Acesso em: 15 nov. 2013.
- RAELE, R.; BOAVENTURA, J. M. G.; FISCHMANN, A. A.; SARTURI, G. Scenarios for the second generation ethanol in Brazil. **Technological Forecasting and Social Change**, v. 87, p. 205–223, 2014.
- RAHMAN, M. R.; ISLAM, A. H. M. H.; RAHMAN, M. A. NDVI Derived Sugarcane Area identification and Crop condition Assessment. **Plan Plus**, v. 1, n. 2, p. 1–12, 2004.
- 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. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14. (SBSR), 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. p. 387-394. DVD, On-line. ISBN 978-85-17-00044-7. (INPE-16002-PRE/10611). Disponível em:<<http://urlib.net/dpi.inpe.br/sbsr@80/2008/11.16.18.50.57>>. Acesso em: 08 nov. 2013..
- RUDORFF, B. F. T.; DE AGUIAR, D. A.; DA 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.
- SÃO PAULO. **Decreto-lei nº 11.241, de 19 de setembro de 2002**. Dispõe sobre a eliminação gradativa da queima da palha da cana-de-açúcar e dá providências correlatas. Poder Executivo, Seção I, São Paulo, 2002.
- SCUTARI, M. **bnlearn**: Bayesian network structure learning, parameter learning and inference. 2014. Disponível em: <<http://cran.r-project.org/web/packages/bnlearn/index.html>>. Acesso: 22 mai. 2014.
- SEELAN, S. K.; LAGUETTE, S.; CASADY, G. M.; SEIELSTAD, G. A. Remote sensing applications for precision agriculture: a learning community approach. **Remote Sensing of Environment**, v. 88, p. 157–169, 2003.

SEGURA, M. L. A evolução da matriz energética brasileira : O papel dos biocombustíveis e outras fontes alternativas. **Revista Âmbito Jurídico - Biodireito**, v. 15, n. 96, p. 1–7, 2012. Disponível em: <[http://www.ambitojuridico.com.br/site/?n\\_link=revista\\_artigos\\_leitura&artigo\\_id=11039&revista\\_caderno=6](http://www.ambitojuridico.com.br/site/?n_link=revista_artigos_leitura&artigo_id=11039&revista_caderno=6)>. Acesso em: 12 ago. 2014.

SMEETS, E.; JUNGINGER, M.; FAAIJ, A.; WALTER, A.; DOLZAN, P.; TURKENBURG, W. The sustainability of Brazilian ethanol—an assessment of the possibilities of certified production. **Biomass and Bioenergy**, v. 32, n. 8, p. 781–813, 2008.

SOARES, D. Z.; ANDREOZZI, S. L. Reflexões sobre o etanol e o biodiesel na matriz energética brasileira. **Revista Geográfica da América Central**, v. 2, p. 1–17, 2011.

SOUZA, M. A. A dinâmica territorial do agronegócio canavieiro e o Zoneamento Agroecológico da cana-de-açúcar: notas para um debate. **Revista de Geografia Agrária**, v. 5, n. 10, p. 148–167, 2010.

SUK, H.-I.; SIN, B.-K.; LEE, S.-W. Hand gesture recognition based on dynamic Bayesian network framework. **Pattern Recognition**, v. 43, n. 9, p. 3059–3072, 2010.

TILMAN, D.; SOCOLOW, R.; FOLEY, J. A.; HILL, J.; LARSON, E.; LYND, L.; PACALA, S.; REILLY, J.; SEARCHINGER, T.; SOMERVILLE, C.; WILLIAMS, R. Beneficial biofuels—the food, energy, and environment trilemma. **Science**, v. 325, p. 270–271, 2009.

UNIÃO DOS PRODUTORES DE BIOENERGIA (UDOP). **A cultura da cana-de-açúcar e os solos de Minas Gerais**. 2015. Disponível em: <<http://www.udop.com.br/index.php?item=noticias&cod=987>>. Acesso em: 05 jan. 2015.

UNIÃO DA INDÚSTRIA DE CANA-DE-AÇÚCAR (ÚNICA). **A energia da cana-de-açúcar - doze estudos sobre a agroindústria da cana-de-açúcar no Brasil e sua Sustentabilidade**. 2. ed. São Paulo: UNICA - União da Indústria de Cana-de-açúcar, 2007. 232. p.

UUSITALO, L. Advantages and challenges of Bayesian networks in environmental modelling. **Ecological Modelling**, v. 203, n. 3-4, p. 312–318, 2007.

VAKKILAINEN, E.; KUPARINEN, K.; HEINIMÖ, J. **Large industrial users of energy biomass**. Lappeenranta: IEA Bioenergy, 2013. 75 p. Disponível em: <<http://www.bioenergytrade.org/publications.html#Large>>. Acesso em: 21 set. 2014.

WALTER, A.; DOLZAN, P. **Brazil country report 2011**. Campinas: UNICAMP, 2012. Disponível em: <<http://www.bioenergytrade.org/publications.html#Large>>. Acesso em: 21 set. 2014.

WEHENKEL, L. On Uncertainty Measures Used for Decision Tree Induction. In: INFORMATION PROCESSING AND MANAGEMENT OF UNCERTAINTY IN KNOWLEDGE-BASED SYSTEMS, (IPMU). 1996, Granada, Espanha. **Proceedings...** 1996. p. 1–6. Disponível em: <<http://hdl.handle.net/2268/79723>>. Acesso em: 08 fev. 2015.

WENG, Q. Remote sensing of impervious surfaces in the urban areas: requirements, methods, and trends. **Remote Sensing of Environment**, v. 117, p. 34–49, 2012.