



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

**INPE-14669-TDI/1222**

**ESTIMATIVA DA PRODUTIVIDADE AGRÍCOLA DA CANA-DE-  
AÇÚCAR UTILIZANDO AGREGADOS DE REDES NEURAIAS  
ARTIFICIAIS: ESTUDO DE CASO USINA CATANDUVA**

Michelle Cristina Araújo Picoli

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto,  
orientada pelo Dr. Bernardo Friedrich Theodor Rudorff, aprovada em  
12 de setembro de 2006.

INPE  
São José dos Campos  
2007

Publicado por:

**esta página é responsabilidade do SID**

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) 3945-6911

Fax: (012) 3945-6919

E-mail: [pubtc@sid.inpe.br](mailto:pubtc@sid.inpe.br)

**Solicita-se intercâmbio  
We ask for exchange**

**Publicação Externa – É permitida sua reprodução para interessados.**



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

**INPE-14669-TDI/1222**

**ESTIMATIVA DA PRODUTIVIDADE AGRÍCOLA DA CANA-DE-  
AÇÚCAR UTILIZANDO AGREGADOS DE REDES NEURAIAS  
ARTIFICIAIS: ESTUDO DE CASO USINA CATANDUVA**

Michelle Cristina Araújo Picoli

Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto,  
orientada pelo Dr. Bernardo Friedrich Theodor Rudorff, aprovada em  
12 de setembro de 2006.

INPE  
São José dos Campos  
2007

528.711.7 : 681.3.019

Picoli, M. C. A.

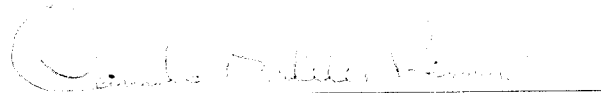
Estimativa da produtividade agrícola da cana-de-açúcar utilizando agregados de redes neurais artificiais: estudo de caso usina Catanduva / Michelle Cristina Araujo Picoli. – São José dos Campos: INPE, 2006.

90p. ; (INPE-14669-TDI/1222)

1.Modelos. 2.Estimativas. 3.Produtividade.  
4.Agrometeorologia. 5.Redes neurais. 6.cana-de-açúcar. 7.NDVI.  
8.MODIS. I.Título.

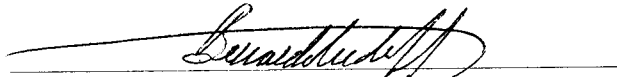
Aprovado (a) pela Banca Examinadora  
em cumprimento ao requisito exigido para  
obtenção do Título de **Mestre** em  
**Sensoriamento Remoto**

Dr. Camilo Daleles Rennó



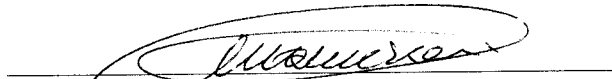
Presidente / INPE / SJC Campos - SP

Dr. Bernardo Friedrich Theodor Rudorff



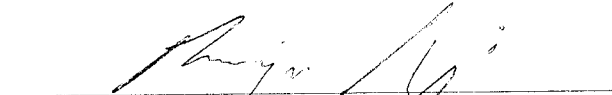
Orientador(a) / INPE / SJC Campos - SP

Dr. Maurício Alves Moreira



Membro da Banca / INPE / SJC Campos - SP

Dr. Rodrigo Rizzi



Convidado/Externo

Aluno (a): **Michelle Cristina Araujo Picoli**

São José dos Campos, 12 de setembro de 2006



*“Podam-se os galhos, colhem-se as frutas.  
E, outra vez se semeia”.*

CARTOLA





*Aos meus pais, à minha irmã e aos meus amigos.*



## **AGRADECIMENTOS**

À Fundação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela bolsa de mestrado.

Ao Instituto Nacional de Pesquisas Espaciais – INPE e em particular à Divisão de Sensoriamento Remoto, pela oportunidade de realizar este estudo e utilizar seus equipamentos e instalações.

Aos professores do INPE, pelo conhecimento compartilhado.

Ao meu orientador Dr. Bernardo F. T. Rudorff, pelo conhecimento passado, pela orientação e apoio na realização deste trabalho, e principalmente pela confiança em mim depositada quando aceitou ser o meu orientador.

Ao Dr. Rodrigo Rizzi, pelo apoio e incentivo na realização deste trabalho e principalmente pelo coleguismo.

À Usina Catanduva, por fornecer os mapas e dados utilizados neste trabalho e em especial aos senhores José Carlos B. de Oliveira, Julio Cezar Bressanin e Mário A. S. Pavani pela presteza em disponibilizar as informações da Usina.

Ao engenheiro agrônomo Luiz Antonio Dias Paes, do Centro de Tecnologia Canavieira – CTC, pelo apoio na realização deste trabalho e pelo fornecimento de informações referentes à produção da cana-de-açúcar.

Ao Prof. Dr. Fernando Von Zuben, ao Dr. Clodoaldo A. de Moraes Lima e ao Wilfredo J. Puma Villanueva, do Laboratório de Bioinformática e Computação Bio-Inspirada – LbiC, da Faculdade de Engenharia Elétrica e Computação da Universidade Estadual de Campinas, pelo grande apoio dado ao trabalhar com Agregados de Redes Neurais Artificiais, além da implementação dos algoritmos utilizados neste trabalho.

Aos meus amigos do Curso de Pós-Graduação em Sensoriamento Remoto, que tornaram o período que passamos juntos, no Laboratório de Tratamento de Imagens Digitais, mais agradável: Delano, Denis, Dani Máquina e Marcos.

Aos meus amigos de turma: Avelino, Camilinha, Conrado, Edu Mineiro, Edu Arraut, Elienê, Giovana, Ilka, José Wilson, Marco Aurélio, Murilo, Rodrigo, Samuca, Silvana, Vanessa, Wilson Lins, e principalmente ao Marcelo, pela enorme ajuda.

Ao pessoal do Laboratório de Monitoramento dos Recursos Agronômicos e Florestais: Daniel, Débora, Magog, Ramon, Vivi, e principalmente à Camila e à Tânia, pelo companheirismo e por me ajudarem sempre que precisei.

Aos colegas do INPE, pelos momentos de descontração: Liana, Felix, Annette, Mari, Luciana, Titico, Marcos e Priscila.

Às amadas amigas, Camila e Natália, para as quais não vêm palavras para expressar toda a minha gratidão, mas vou tentar: obrigada pela ajuda e compreensão nos momentos mais difíceis, pelas ricas discussões, pelo constante apoio e estímulo e por tantas outras coisas que, por vezes, somente bons amigos podem fazer por nós.

Aos grandes amigos de Campinas, pelos momentos felizes: Alan, Bruna, Aninha, João, Mayra, Lainas, Alininha, Joãozinho, Jojo, Taisinha, Pastel, e um agradecimento especial para as amigas Mari e Loli, pelo amor incondicional e compreensão desmedida.

À minha família, principalmente a família Matias Aires, por todo amor partilhado.

A meus pais, Cândida e Carlos, e à minha amiga, companheira e irmã Dedé por sempre acreditarem em mim e me apoiarem acima de tudo.

## RESUMO

A importância econômica da cultura da cana-de-açúcar para fornecer matéria-prima para a produção de álcool como fonte alternativa de energia tem crescido muito. Assim, informações precisas sobre a produção da cana são importantes para que as crescentes demandas de álcool possam ser adequadamente atendidas. A produção de cana em toneladas é dada pelo produto entre a área plantada (ha) e a produtividade agrícola ( $\text{ton ha}^{-1}$ ). O presente trabalho tem por objetivo estimar a produtividade da cana-de-açúcar na Usina Catanduva utilizando variáveis quantitativas e qualitativas em um modelo de agregados de Redes Neurais Artificiais (RNA) ou *ensembles*. Uma das variáveis quantitativas deste modelo foi estimada a partir de um modelo agrônomo-espectral. Este modelo utiliza uma série de variáveis meteorológicas e agrônomicas sendo que uma delas é o índice de área foliar (IAF) cujos valores foram estimados, para cada talhão de cana, a partir do NDVI das imagens MODIS obtidas ao longo do período de intenso crescimento da cultura canavieira. Outras variáveis utilizadas no modelo *ensemble* são: variedade plantada, tipo de solo, estágio de corte, aplicação de vinhaça, ano safra, produtividade de cana-de-açúcar estimada pelos técnicos da Usina, produtividade real do ano safra anterior e o NDVI de uma imagem Landsat-TM adquirida antes do início do período de colheita. O modelo *ensemble* foi criado, validado e a estimativa da produtividade do modelo foi avaliada. Este modelo explicou 66% da variabilidade da produtividade dos talhões com cana soca, cuja média foi de  $81,9 \text{ t ha}^{-1}$  e erro quadrático médio (EQM) de  $13,9 \text{ t ha}^{-1}$ . Outro modelo *ensemble* foi calculado sem utilizar a variável produtividade estimada pelos técnicos da Usina Catanduva, e este explicou 58% da variabilidade da produtividade real, com um EQM de  $15,6 \text{ t ha}^{-1}$ , e a média calculada foi igual a  $81,5 \text{ t ha}^{-1}$ . Já a produtividade estimada pelos técnicos da Usina explicou 55% da variabilidade da produtividade real com média de  $80,8 \text{ t ha}^{-1}$  e EQM de  $14,6 \text{ t ha}^{-1}$ . Sendo que o valor médio da produtividade real dos talhões foi de  $81,5 \text{ t ha}^{-1}$ , e, portanto o modelo superestimou essa média em apenas 0,5% ( $0,4 \text{ t ha}^{-1}$ ), o segundo modelo que não utilizou a variável produtividade estimada pelos técnicos da Usina Catanduva superestimou a produtividade média em 1,2% ( $1,5 \text{ t ha}^{-1}$ ), enquanto que a estimativa da Usina subestimou em 0,8% ( $0,7 \text{ t ha}^{-1}$ ). Por fim, os dois modelos propostos apresentaram resultados satisfatórios, permitindo assim a utilização destes no aprimoramento da estimativa da produtividade agrícola de talhões com cana-de-açúcar da Usina Catanduva.



## ESTIMATE SUGARCANE YIELD USING COMBINING ARTIFICIAL NEURAL NETWORKS: CASE STUDY CATANDUVA PLANT

### ABSTRACT

The economic relevance of sugarcane crop to provide raw material to produce alcohol as an energy alternative has grown much. Therefore, precise information on sugarcane production is important to adequately meet the growing demands of alcohol. Sugarcane production in tons is given by the product between planted area (ha) and crop yield (ton ha<sup>-1</sup>). The present work has the objective to estimate sugarcane yield in the Catanduva Plant using quantitative and qualitative variables in a combining Artificial Neural Networks (ANN) model or ensembles. One of the quantitative variables of this model was estimated through an agronomic-spectral model. This model uses several meteorological and agronomic variables being one of them the leaf area index (LAI) whose values were estimated for each sugarcane field, from the MODIS NDVI images acquired during the period of intense crop growth. Other variables that were used in the ensemble model are: planted variety, soil type, cutting stage, application of vinasse, crop year, estimated crop yield by Plant technicians, real crop yield of the year previous harvest and NDVI values from a Landsat-TM image acquired prior to the beginning of the harvest period. The ensemble model was created, validated and estimated crop yield was evaluated. The model explained 66% of the variability of sugarcane yield for ratoons, with mean value of 81.9 t ha<sup>-1</sup> and mean quadratic error (MQE) of 13.9 t ha<sup>-1</sup>. Another model ensemble was calculated without using the variable estimated crop yield by Plant technicians, and this explained 58% of the variability of sugarcane yield for ratoons, with a MQE of 15,6 t ha<sup>-1</sup>, with mean value of 81,5 t ha<sup>-1</sup>. The crop yield estimated by the Plant technicians explained 55% of the yield variability, with mean value of 80.8 t ha<sup>-1</sup> and MQE of 14.6 t ha<sup>-1</sup>. The mean observed sugarcane yield value was 81.5 t ha<sup>-1</sup>. The proposed model overestimated this mean by only 0.5% (0.4 t ha<sup>-1</sup>), as model that did not use the variable estimated crop yield by Plant technicians overestimated the mean in 1,2% (1,5 t ha<sup>-1</sup>), while the Plant underestimated it by 0.8% (0.7 t ha<sup>-1</sup>). The two proposed models presented a satisfactory resulted and can be used to improvement field sugarcane yield at field level in the Catanduva Plant.





## SUMÁRIO

### LISTA DE FIGURAS

### LISTA DE TABELAS

<b>CAPÍTULO 1 - INTRODUÇÃO</b> .....	<b>21</b>
<b>CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>23</b>
2.1 Cana-de-açúcar .....	23
2.1.1 Características da cultura e importância econômica.....	23
2.1.2 Fisiologia .....	25
2.1.3 Produtividade agrícola.....	27
2.1.4 Variedades .....	27
2.1.5 Estágio de corte .....	28
2.1.6 Época de plantio .....	29
2.1.7 Índice de Área Foliar .....	31
2.2 Sensoriamento Remoto aplicado à cultura da cana-de-açúcar .....	32
2.2.1 Comportamento espectral .....	32
2.2.2 Índices de Vegetação .....	33
2.2.3 Imagens do sensor MODIS .....	34
2.3 Modelos de estimativa de produtividade .....	35
2.3.1 Modelo agrônomico.....	37
2.3.2 Modelo espectral e agrônomico-espectral .....	37
2.4 Redes Neurais Artificiais.....	39
2.4.1 Arquiteturas .....	41
2.4.2 Treinamento (ou Aprendizado) .....	42
2.4.3 Função de ativação .....	42
2.4.4 Redes do Tipo <i>Perceptron</i> de Múltiplas Camadas (MLP) .....	43
2.4.5 Agregados de Redes Neurais Artificiais - <i>Ensembles</i> .....	47
<b>CAPÍTULO 3 - MATERIAIS E MÉTODOS</b> .....	<b>49</b>
3.1 Área de estudo .....	49
3.2 Materiais .....	51
3.2.1 Dados Meteorológicos.....	52
3.2.2 Dados Espectrais .....	53

3.2.3 Dados Qualitativos e Quantitativos .....	54
3.3 Método.....	54
3.3.1 Estimativa da Produtividade.....	54
3.3.1.1 Modelo agronômico-espectral .....	54
3.3.1.2 Organização e codificação dos dados de cana soca para a entrada nos <i>ensembles</i> .....	59
3.3.1.3 <i>Ensembles</i> para integração dos dados.....	60
3.3.1.4 Treinamento, Seleção, Validação e Teste.....	62
<b>CAPÍTULO 4 - RESULTADOS.....</b>	<b>65</b>
4.1 Análise preliminar e seleção dos dados .....	65
4.2 Modelo agronômico-espectral .....	66
4.3 Modelo de Agregados de Redes Neurais Artificiais - <i>Ensembles</i> .....	70
4.3.1 Criação do modelo.....	70
4.3.2 Validação do modelo.....	74
4.3.2.1 Conjunto de dados contendo a variável produtividade estimada pela Usina Catanduva .....	74
4.3.2.2 Conjunto de dados que não contém a variável produtividade estimada pela Usina Catanduva .....	77
<b>CAPÍTULO 5 - CONCLUSÕES E RECOMENDAÇÕES.....</b>	<b>85</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>87</b>
<b>APÊNDICE A .....</b>	<b>93</b>

## LISTA DE FIGURAS

2.1 – Maiores produtores mundiais de cana-de-açúcar na safra 2003/2004. A barra em azul representa a contribuição do Estado de São Paulo na produção de cana-de-açúcar no Brasil. ....	24
2.2 – Época das operações agrícolas para a cana-de-açúcar na região Centro-Sul. ....	30
2.3 – Curva de reflectância típica da vegetação. ....	33
2.4 – Célula neural biológica com a seqüência de propagação do sinal. ....	40
2.5 – Arquitetura MLP com duas camadas intermediárias. ....	44
2.6 – Ilustração das direções de propagação do sinal funcional e do erro. ....	45
2.7 – Exemplo de superfície do erro. ....	46
3.1 – (A) Estado de São Paulo; (B) Detalhamento da área em estudo e os talhões de cana-de-açúcar, na safra 2004/05. ....	50
3.2 – Disposição do tipo de cana-de-açúcar na Usina Catanduva na safra (A) 2004/05 e (B) 2005/06. ....	52
3.3 – Interface do modelo agrônômico-espectral. ....	58
3.4 – Interface do programa <i>Ensemble_Bagging</i> . ....	62
4.1 – Variação temporal do NDVI do MODIS para uma amostra de talhões das safras (A) 2004/05 e (B) 2005/06. ....	67
4.2 – Relação entre a produtividade real e a produtividade estimada pelo modelo agrônômico-espectral, safras (A) 2004/05 e (B) 2005/06. ....	69
4.3 – Produtividade estimada pelos ensembles 5, 9 e 11 (vermelho) e suas respectivas produtividades reais (azul), para as safras de 2004/05 e 2005/06, para o conjunto dos dados de entrada. ....	72
4.4 – Produtividade estimada pelos <i>ensembles</i> 1, 10 e 22 (vermelho) e suas respectivas produtividades reais (azul), para as safras de 2004/05 e 2005/06, para o conjunto dos dados de entrada sem a variável produtividade estimada pela Usina Catanduva. ....	74
4.5 – Produtividade estimada pelos <i>ensembles</i> 14, 21 e 30 e suas respectivas produtividades reais, para as safras de 2004/05 e 2005/06, para o conjunto dos dados de validação. ....	76
4.6 – Relação entre a produtividade real e a produtividade estimada pelo modelo ( <i>ensemble</i> 14), safras 2004/05 e 2005/06. ....	77
4.7 – Produtividade estimada pelos <i>ensembles</i> 9, 14, e 30 e suas respectivas	

produtividades reais, para as safras de 2004/05 e 2005/06, para o conjunto dos dados de validação sem a variável produtividade estimada pela Usina Catanduva. ....	79
4.8 – Relação entre a produtividade real e a produtividade estimada pelo modelo ( <i>ensemble</i> 14) sem a variável produtividade estimada pela Usina Catanduva, safras 2004/05 e 2005/06. ....	80

## LISTA DE TABELAS

2.1 – Produtividade média da cana-de-açúcar para diferentes estágios de corte, para os anos safras de 1998 até 2003. ....	29
3.1 – Valores dos parâmetros $k_c$ , $k_y$ e $D$ utilizados para o cálculo do modelo agrônômico-espectral, para a cana soca. ....	57
4.1– Resultado dos 30 melhores <i>ensembles</i> e seus respectivos EQM para o conjunto dos dados de entrada. ....	71
4.2 – Resultado dos 30 melhores <i>ensembles</i> e seus respectivos EQM para o conjunto dos dados de entrada sem a variável produtividade estimada pela Usina Catanduva...	73
4.3 – Coeficiente de determinação dos 30 melhores <i>ensembles</i> e seus respectivos EQM para o conjunto de dados validados. ....	75
4.4 – Coeficiente de determinação dos 30 melhores <i>ensembles</i> e seus respectivos EQM para o conjunto de dados validados sem a variável produtividade estimada pela Usina Catanduva. ....	78



## CAPÍTULO 1

### INTRODUÇÃO

A cana-de-açúcar (*Saccharum* spp) é uma gramínea que possui grande importância econômica para o Brasil por fornecer a matéria prima para a produção de açúcar e de álcool, além de exercer um papel relevante tanto no mercado interno quanto externo. São Paulo é o principal Estado produtor de cana, responsável por cerca de 60% de todo o açúcar e 62% de todo álcool produzido no País, e por 70% das exportações nacionais de açúcar (UNICA, 2004a).

A produção de açúcar e de álcool depende da quantidade de matéria prima disponível que por sua vez depende: da área plantada; da produtividade agrícola; e do açúcar total recuperável (ATR). Destes três fatores o mais complexo a ser estimado é a produtividade agrícola, pois depende de uma série de outros fatores relacionados, principalmente, a aspectos agrônômicos e climáticos cujos efeitos sobre a produtividade são difíceis de serem quantificados. Neste contexto, a estimativa de produtividade tem importância para o planejamento estratégico das empresas deste setor a fim de presumir o quanto da produção será processada e armazenada; a logística do transporte e também a tomada de decisão sobre a comercialização dos produtos finais.

Nas usinas de cana-de-açúcar, as estimativas de produtividade são feitas por métodos tradicionais, que demandam tempo e estão sujeitas a imprecisões que as comprometem. Estas estimativas são feitas geralmente por técnicos que percorrem os canaviais observando o desenvolvimento da cana-de-açúcar, agregando ao seu cálculo conhecimentos pessoais sobre a área cultivada e resultados registrados em safras anteriores. Essa maneira subjetiva de se estimar a produtividade é passível de manipulação e, portanto pode ser tendenciosa, além de não permitir uma análise do erro.

Um modelo de estimativa de produtividade utilizando dados agrônômicos, meteorológicos e de sensoriamento remoto já foi desenvolvido anteriormente (Rudorff,

1985; Rudorff e Batista, 1990 e 1991; Rudorff et al., 1995; Berka et al., 2003; Rizzi, 2004), mas o seu potencial ainda não foi amplamente explorado. Em especial, no que diz respeito ao emprego de variáveis qualitativas (por exemplo: tipo de variedade plantada, estágio de corte, tipo de solo, entre outros) em uma estrutura de Agregados de Redes Neurais Artificiais (RNA) ou *ensembles*, conforme proposto neste trabalho.

Deste modo, o objetivo geral do trabalho foi estimar a produtividade agrícola da cana-de-açúcar por talhão utilizando variáveis quantitativas, qualitativas e espectrais em um modelo de agregados de RNA ou *ensembles* para as safras 2004/05 e 2005/06.

Os objetivos específicos foram:

- a) Ajustar um modelo agronômico-espectral para estimar a produtividade da cultura da cana-de-açúcar e utilizar esta estimativa como uma das variáveis no modelo de agregados de RNA;
- b) Criar e validar o modelo de agregados de RNA utilizando a estimativa da produtividade do modelo agronômico-espectral juntamente com outras variáveis quantitativas, qualitativas e espectrais relacionadas à produtividade e disponíveis na Usina;
- c) Avaliar a qualidade da estimativa do modelo de Agregados de RNA.



## **CAPÍTULO 2**

### **FUNDAMENTAÇÃO TEÓRICA**

#### **2.1 Cana-de-açúcar**

##### **2.1.1 Características da cultura e importância econômica**

O Brasil é o maior produtor de cana-de-açúcar (*Saccharum spp*) do mundo, produz cerca de 416 milhões de toneladas, seguido pela Índia e China. Em média, 55% da produção brasileira destina-se à produção de álcool e 45% à produção de açúcar. A cana-de-açúcar é cultivada nas regiões Centro-Sul e Nordeste, o que permite dois períodos de safra. Na região Centro-Sul, a safra ocorre de abril a novembro e na região Nordeste ela ocorre de novembro a abril (UNICA, 2004). Para o Brasil, e particularmente para o Estado de São Paulo, a cana-de-açúcar é uma cultura muito importante. Ela foi introduzida no Brasil em 1532 e já teve grande importância na economia do país no passado. Ao longo dos últimos anos, ela vem se destacando novamente sendo a principal cultura explorada no Estado de São Paulo, que produz aproximadamente 239 milhões de toneladas (safra 2003/04) e é o terceiro maior produtor mundial de cana-de-açúcar (Figura 2.1).

Países	Produção
Filipinas	28.000
Austrália	36.892
Colômbia	37.100
México	45.127
Paquistão	52.040
Tailândia	63.707
China	93.200
Índia	244.800
Brasil	239.528 - SP 416.256

\*valores em mil toneladas

FIGURA 2.1 – Maiores produtores mundiais de cana-de-açúcar na safra 2003/2004. A barra em azul representa a contribuição do Estado de São Paulo na produção de cana-de-açúcar no Brasil.

FONTE: FAO (2006).

Um dos principais indicadores de que o mercado da cana-de-açúcar só tende a aumentar é o protocolo de Quioto, assinado em 1997 por vários países membro das Nações Unidas, com uma proposta de estabilização das emissões de gases geradores do efeito estufa, e essa redução deve ocorrer entre 2008 e 2012.

A cana-de-açúcar fornece a matéria prima para a produção de álcool e uma das formas de reduzir a emissão de gases causadores do efeito estufa é misturar álcool a gasolina. Isto impulsiona programas de produção de etanol como fonte alternativa dos combustíveis fósseis. Além do fator ambiental, o fator econômico contribui para o uso do etanol como combustível devido à instabilidade na oferta do petróleo e suas elevações de preço (Figueira, 2005).

A atividade canavieira do Brasil é responsável por cerca de um milhão de empregos diretos, 511 mil apenas na produção de cana-de-açúcar. O restante está distribuído na agroindústria de açúcar e álcool. O Estado de São Paulo sozinho reúne 400 mil empregos diretos do setor, e o número de empregos indiretos pode chegar a 1,2 milhão.

No Brasil, o agronegócio é responsável por 20,6% do Produto Interno Bruto – PIB brasileiro e gera 14% dos empregos totais do País. Destaca-se que o agronegócio da cana-de-açúcar, reúne 6% dos empregos agroindustriais brasileiros e é responsável por 35% do PIB e do emprego rural do Estado de São Paulo (UNICA, 2006).

O plantio da cana-de-açúcar está distribuído entre variedades nacionais de três instituições que realizam melhoramento genético: o Instituto Agronômico de Campinas com a sigla inicial IAC, o Centro de Tecnologia Canavieira (CTC) com sigla inicial SP e o Plano Nacional de Melhoramento da Cana-de-Açúcar (Planalsucar), extinto em 1990, e com sigla inicial RB. Algumas variedades estrangeiras são representadas pelas siglas iniciais: B (Barbados), Co (Coimbratore), CP (Canal Point); F (Flórida), H (Havaí), M (Maurícios), Mex (México), NA (Norte da Argentina), POJ (Proofstation Oost Jawa), Q (Queensland), R (Reunion) e T (Tucuman). Todas as variedades são híbridos do gênero *Saccharum*, nos quais procura-se obter, além das características agrônomicas de produtividade agrícola, rusticidade, resistência a pragas e a doenças, algumas características industriais como alto teor de sacarose e teor médio de fibra, condições essenciais para uma boa exploração (Stupiello, 1987).

O cultivo da cana-de-açúcar apresenta nível tecnológico dos mais elevados dentro da agricultura brasileira, utilizando técnicas como gerenciamento das áreas produtoras por meio de banco de dados e setorização da produção agrícola em talhões (Joaquim, 1998).

Estimativas de produtividade acuradas de cana-de-açúcar são importantes por vários motivos incluindo políticas de preço e marketing para as indústrias, decisões de datas de início e fechamento das safras além de taxas de abastecimento e moagem das unidades industriais (Schmidt et al., 2001).

### **2.1.2 Fisiologia**

A cana-de-açúcar (*Saccharum* spp) é uma gramínea semi-perene e originária provavelmente do sudoeste da Ásia. Não é uma cultura exigente em solo, porém, se desenvolve melhor em solos com boa aeração, boa drenagem e com profundidade maior que um metro (Coopersucar, 1988).

O primeiro ciclo da cultura, chamada de cana planta, pode ser de 12 meses (cana de ano – CA) ou 18 meses (cana de ano e meio – CAM). Após o primeiro corte, o ciclo passa a ser de 12 meses para todas as variedades, conhecida como cana soca (CS). A mesma cana-de-açúcar pode ser colhida cinco ou mais vezes, porém, a cada ciclo devem ser feitos investimentos significativos em insumos e tratos culturais para manter a produtividade (UNICA, 2004).

Os fatores ambientais que afetam de maneira marcante a produção de cana-de-açúcar são: temperatura, luz (intensidade), e disponibilidade de água e nutrientes (Afonso et al., 1987). A cana-de-açúcar se desenvolve bem em regiões de clima quente, com temperatura oscilando entre 16°C e 33°C. No entanto, temperaturas inferiores a 21°C reduzem a taxa de alongamento dos colmos e promovem o acúmulo de sacarose (Magalhães, 1987).

A cana-de-açúcar apresenta uma grande capacidade fotossintética (Fortes, 2003) e responde bem a elevadas intensidades luminosas. A radiação solar afeta todos os estágios de desenvolvimento da cultura e quando se desenvolve sob condições de baixa luminosidade, apresenta colmos finos e longos, o que diminui o acúmulo de matéria seca (Lucchesi, 1995).

No decorrer do ciclo de crescimento e desenvolvimento, a cana-de-açúcar atravessa dois períodos distintos com relação ao teor de sacarose: o primeiro é assinalado por um intenso crescimento vegetativo acompanhado por uma gradual formação de sacarose, enquanto, no segundo, ocorre um predominante acúmulo de sacarose, motivado pela escassez dos principais fatores de desenvolvimento vegetativo, como temperatura e água disponível (Magalhães, 1987 e Lucchesi, 1995).

A maturação é um processo fisiológico afetado por vários fatores naturais, como a própria variedade, condições climáticas, tipo de solo e tratos culturais. As variedades de cana-de-açúcar têm comportamentos distintos, apresentando níveis diferentes de sacarose mesmo quando cultivada nas mesmas condições (Nunes Junior, 1987). O estágio de maturação é verificado principalmente pelos teores de sacarose, de açúcares redutores e umidade que os colmos apresentam no decorrer do período da safra.

### **2.1.3 Produtividade agrícola**

A cana-de-açúcar é insuperável em termos de produção de matéria seca e energia por hectare em um único corte. Nas condições de Brasil Central, a produção de cana-de-açúcar integral fresca por hectare por corte varia entre 60 e 120 toneladas, por um período de até cinco anos, obtendo maior produtividade no primeiro ano (Thiago e Vieira, 2002).

A produtividade da cana-de-açúcar (produtividade = produção de toneladas de colmos / área colhida em hectares) está diretamente relacionada com o seu perfilhamento. A adubação nitrogenada promove aumento de produção, porém pode aumentar o perfilhamento das plantas, o que nem sempre é bom, pois muitos perfilhos não indicam maior produtividade. O índice de perfilhamento da cana-de-açúcar é uma característica varietal. A capacidade que uma planta apresenta de mobilizar suas reservas, emitindo mais ou menos brotos em determinada condição ambiental é o que a caracteriza em alto, médio ou baixo índice de brotação. As recomendações de adubação nitrogenada para cana planta não são bem estabelecidas, por um provável desconhecimento das bases fisiológicas para as respostas ao nitrogênio aplicado como fertilizante, pois o excesso de dosagem poderia aumentar muito o número de perfilhos podendo prejudicar o peso individual dos colmos, ou seja, a produtividade da planta (Magalhães, 1987). O clima, a variedade, o tipo de solo, o estágio de corte, a data do plantio, entre outros, são fatores que também influenciam na produtividade agrícola da cana-de-açúcar (Maximiliano, 2002).

### **2.1.4 Variedades**

As necessidades da cana para uma boa produtividade são praticamente as mesmas de outras culturas: água, fertilizantes, solo descompactado, uso de defensivos agrícolas para controle de ervas daninhas, fungos, bactérias, insetos e nematóides. E o Brasil tem se preocupado bastante com a obtenção de novas variedades de cana-de-açúcar, que sejam mais produtivas e mais resistentes.

Segundo Mangelsdorf (1966), as variedades comerciais têm ciclos de cultivo bem

definidos, sendo que após este período começam a demonstrar sinais de degeneração. Esta é acompanhada de significativas perdas em produtividade agrícola e, por este motivo, precisam ser gradativamente substituídas por novos materiais mais estáveis e geneticamente superiores.

Por degenerescência ou por necessidade de aumento de produtividade, a busca de novas variedades conduz a uma verdadeira evolução no cultivo da cana-de-açúcar. O Estado de São Paulo é um dos melhores exemplos de ganho de produtividade devido substituição de variedades em declínio (Nunes Jr, 1987). Condições como o clima e o solo são fatores importantes para serem levados em conta na escolha da variedade. Programas de melhoramento têm sido criados e novas variedades cada vez mais produtivas estão aparecendo, variedades com maior potencial de sacarose, resistentes às principais doenças e adaptadas às atuais condições de manejo.

As variedades de cana-de-açúcar na região Centro-Sul são divididas em 4 grupos com relação à maturação, ou seja, o período dentro da safra em que atingem o máximo teor de sacarose nos colmos (Nunes Jr, 1987):

- G1 = Grupo de variedades de maturação precoce atingem o máximo teor de sacarose de abril a maio;
- G2 = Grupo variedades de maturação semiprecoce atingem o máximo teor de sacarose no final de maio ao início de julho;
- G3 = Grupo de variedades de maturação média atingem o máximo teor de sacarose no final de julho ao início de outubro;
- G4 = Grupo de variedades de maturação tardia atingem o máximo teor de sacarose de outubro a novembro.

### **2.1.5 Estágio de corte**

A cana-de-açúcar para produção de álcool e açúcar é cultivada, em média, por quatro a cinco cortes. Assim, a cana soca pode representar cerca de 85% da área cultivada.

O sistema radicular da cana-de-açúcar pode continuar ativo por um longo período de tempo depois do corte da planta, deixando de funcionar gradualmente, à medida que um sistema novo se forma, ao crescerem os perfilhos da soqueira (Humbert, 1974).

Num censo varietal realizado pelo Centro de Tecnologia Copersucar – CTC (2004), para uma área de aproximadamente 4 milhões de hectares cultivados com cana-de-açúcar no Brasil, pode ser observado que existe uma relação entre o estágio de corte e a produtividade da cana-de-açúcar (Tabela 2.1). Tal produtividade foi calculada utilizando as médias dos cortes das safras desde 1998 até 2003.

TABELA 2.1 – Produtividade média da cana-de-açúcar para diferentes estágios de corte, para os anos safras de 1998 até 2003.

Corte	Produtividade média (t ha <sup>-1</sup> )
1º - cana de ano	79
1º - cana de ano e meio	110
2º - 1ª soca	89
3º - 2ª soca	78
4º - 3ª soca	71
5º - 4ª soca	68
Outros cortes	68

FONTE: CTC (2004).

Na Tabela 2.1, observa-se que a produtividade da cana-de-açúcar tende a diminuir com o aumento do número de cortes.

### 2.1.6 Época de plantio

Ricaud e Cochran (1980) afirmam que a data de plantio exerce influência significativa na produtividade da cana-de-açúcar.

O clima do Estado de São Paulo permite o plantio da cana-de-açúcar num amplo período do ano. Normalmente, são recomendadas duas épocas de plantio. Uma menos ampla, no início da primavera, e outra, mais longa, no verão, quando as condições de temperatura e umidade são favoráveis à boa e rápida emergência dos colmos (Marchiori, 2004).

Barbiere e Villa Nova (1977) descrevem que o plantio da cana-de-açúcar na região Centro-Sul é feito em duas épocas, conhecidas como “plantio de cana de ano” e “plantio de cana de ano e meio”. O plantio feito em setembro-outubro permite a colheita da cana-de-açúcar com aproximadamente 11 a 14 meses, e é conhecida como “cana de ano”. O plantio feito em janeiro-março permite a colheita da cana-de-açúcar com aproximadamente 15 a 20 meses, e é conhecida como “cana de ano e meio” (Figura 2.2).

### EPOCA DAS OPERAÇÕES AGRÍCOLAS PARA A CANA-DE-ÂÇUCAR

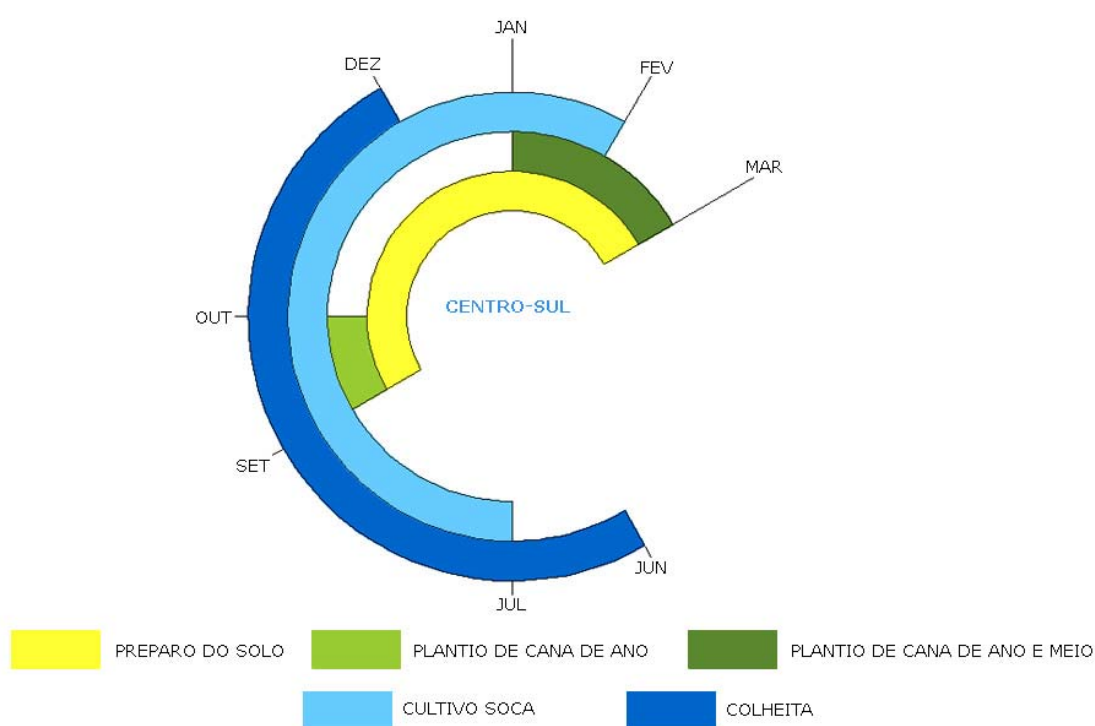


FIGURA 2.2 - Época das operações agrícolas para a cana-de-açúcar na região Centro-Sul.

FONTE: Marchiori (2004).

A cana de ano, plantada em setembro-outubro tem seu máximo de desenvolvimento de novembro a abril, diminuindo em seguida, devido às condições climáticas adversas ao crescimento, podendo ser colhida, dependendo da variedade, a partir de julho. A cana de ano e meio, plantada em janeiro-março, tem um crescimento inicial no primeiro período chuvoso (fevereiro a abril), mas desencadeia um acelerado crescimento no segundo



período chuvoso (outubro a abril) e embora não seja colhida no ano safra em que foi plantada ela tem quase o dobro da produtividade da cana de ano (Castro, 1999).

A influência da água para a cana-de-açúcar é grande, principalmente na fase de desenvolvimento vegetativo da cultura, que pode chegar a um período de nove meses no ano. Se nesta fase de desenvolvimento a escassez de água for grande pode ocorrer queda significativa na produtividade da cana-de-açúcar. Geralmente essa fase vai de novembro a abril, para a cana de ano e cana soca, e de setembro a abril para a cana de ano e meio (Alfonsi et al., 1987). Porém, devido à diferença entre os tipos de solos, a produtividade é menos afetada por déficit hídricos em solos argilosos do que em solos arenosos devido à diferença na capacidade de retenção de água destes solos.

### **2.1.7 Índice de Área Foliar**

O índice de área foliar (IAF) foi definido por Watson (1947) como a área foliar integrada do dossel por unidade de superfície projetada no solo ( $m^2/m^2$ ). O IAF é um dos principais parâmetros da vegetação, e é muito utilizado em modelos de crescimento vegetal e de evapotranspiração, pois está relacionado à produção de biomassa.

Formaggio e Epiphanyo (1988) afirmaram que o conhecimento das relações existentes entre os parâmetros físicos das culturas, como o IAF, e seu comportamento espectral é um passo importante para avaliar o desempenho e o desenvolvimento das culturas, utilizando modelos de dados de sensoriamento remoto com os índices de vegetação (IV).

Xavier (2002) estudou a relação entre o IAF de diferentes tipos de cobertura do solo (cana-de-açúcar, pastagem, milho, eucalipto e floresta ripária) com: o índice de vegetação da razão simples (RVI), o índice de vegetação por diferença normalizada (NDVI) e o índice de vegetação ajustado para o solo (SAVI), na microbacia Ribeirão dos Martins, Região de Piracicaba-SP. O autor observou que o padrão geral das curvas obtidas foi semelhante nas relações estudadas. Nas relações entre o IAF e o NDVI e também entre o IAF e o SAVI, notou que os valores desses dois índices de vegetação (NDVI e SAVI) apresentaram-se maiores para valores menores de IAF, mas tenderam a

se estabilizar para valores maiores de IAF. Os coeficientes de determinação encontrados para as relações entre IAF e SAVI foi 0,56, para IAF e RVI foi 0,70 e para IAF e NDVI foi 0,72, mostrando que o IAF apresentou melhor relação com o NDVI. Xavier (2002) ainda concluiu que o uso do NDVI, obtido de dados de sensoriamento remoto, foi eficaz para mapear o IAF em escala regional.

Xavier e Vettorazzi (2003) fizeram outro estudo, também na microbacia Ribeirão dos Martins, e obtiveram os valores de IAF para a cana-de-açúcar, que variaram de 0,47 a 3,34. Os autores observaram que além das medidas relativas fornecidas pelo equipamento LAI-2000, os valores mais elevados de IAF devem ser associados ao fato de que o tipo de material genético da cana é distinto e as técnicas de gerência utilizadas para a cultura são diferentes (por exemplo: espaçamento entre as plantas, adubação, irrigação, etc.), pois quando eles compararam os valores de IAF obtidos neste trabalho com os valores de IAF de outros estudos houve diferença.

Estes autores relacionaram os valores de IAF com os meses do ano. E a variação total do IAF explicada pelo mês do plantio e pelo mês da colheita foi 47% ( $P < 0.01$ ). Valores mais elevados de IAF se mostram, aproximadamente, 10 meses após o plantio. A conclusão que se chegou nesse trabalho é que o IAF da cana-de-açúcar tem muita variabilidade devido às características da própria cultura e ao seu manejo na área de estudo, pois cada variedade de cana tem sua própria estação e época de plantio e, conseqüentemente, valores distintos de IAF.

## **2.2 Sensoriamento Remoto aplicado à cultura da cana-de-açúcar**

### **2.2.1 Comportamento espectral**

As propriedades ópticas de uma folha são função de sua composição, morfologia e estrutura interna as quais definem a variação da resposta espectral registrada pelo sensor ao longo da faixa do espectro eletromagnético analisada. Na região do visível (400 a 700 nm) se observa baixa reflectância devido à absorção da radiação incidente pelos pigmentos, como as clorofilas *a* e *b* e os carotenóides; na região do infravermelho próximo (700 a 1300 nm) a reflectância das folhas é alta devido às múltiplas reflexões

da radiação nas paredes das células; e na região do infravermelho médio o teor de água contido nas folhas é responsável pela absorção em 1400 e 1950 nm diminuindo a reflectância da folha (Gates et al., 1965), Figura 2.3.

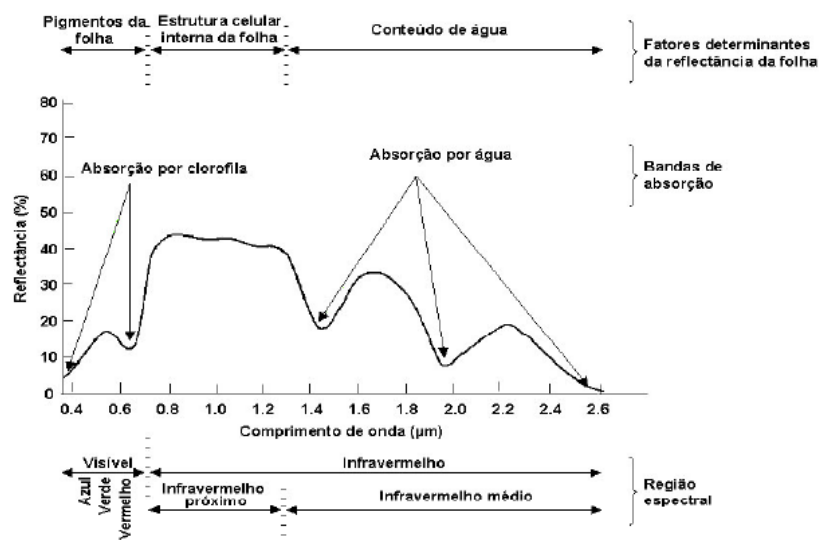


FIGURA 2.3 – Curva de reflectância típica da vegetação.

FONTE: Adaptada de Swain e Davis (1978, p. 232).

resposta espectral da vegetação também é influenciada pela resposta espectral do solo, principalmente na fase inicial do crescimento, quando a percentagem de cobertura do solo pela vegetação é pequena. Da mesma forma, a declividade e o relevo também influenciam a resposta espectral da vegetação (Ponzoni, 2002 e Valeriano, 2003).

Rudorff e Batista (1985) analisaram dados espectrais do sensor MSS a bordo do satélite Landsat 4 em cinco datas diferentes e concluíram que a melhor época para aquisição de imagens, com a finalidade de relacionar a resposta espectral com peso dos colmos, situa-se em torno do mês de fevereiro. Isto ocorre pelo fato da biomassa fotossinteticamente ativa encontrar-se bem desenvolvida nesta época e ter uma forte relação com a quantidade de colmos.

### 2.2.2 Índices de Vegetação

O conhecimento do comportamento espectral da cana-de-açúcar pode auxiliar na

estimativa da produtividade. Índices de vegetação, que podem ser calculados por meio das bandas espectrais das imagens de satélite, têm boa relação com a produtividade agrícola (Rudorff, 1985; Rudorff e Batista, 1990 e 1991; Rudorff et al., 1995; Berka et al., 2003; Rizzi, 2004).

Índices de vegetação (IV) são indicadores à presença e condição da vegetação e por isso têm sido frequentemente utilizados como variáveis auxiliares para estimar a produtividade de culturas agrícolas. Os IV se baseiam em combinações lineares, razões ou transformações ortogonais de várias bandas espectrais, de tal forma que resumem estes dados espectrais em um único valor. Os IV também têm a finalidade de destacar a contribuição espectral da vegetação minimizando a influência de fatores como solo, ângulo solar, vegetação senescente, atmosfera, e inclinação do terreno (Gutman et al., 1995 e Almeida et al. 2005). Porém, os efeitos da visada fora do nadir produzem diferenças significativas no cálculo do NDVI, como concluiu Galvão et al (2001).

O índice de vegetação por diferença normalizada (NDVI) é o IV mais utilizado. O NDVI é calculado pela combinação de valores de reflectância no comprimento de onda do vermelho (*verm*: 600 - 700 nm) e do infravermelho próximo (IVP: 700 - 1.300 nm) conforme a Equação 2.1. O NDVI realça os pixels com vegetação que possuem reflectância baixa no *verm* e alta no IVP.

$$NDVI = \frac{IVP - verm}{IVP + verm} \quad (2.1)$$

Pelo fato de ser um índice normalizado os valores de saída variam de -1 a 1. Áreas com vegetação apresentam valores positivos, pois o valor da reflectância no infravermelho próximo é maior do que na região do vermelho.

### **2.2.3 Imagens do sensor MODIS**

O primeiro sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*) a ganhar o espaço foi lançado a bordo do satélite TERRA em dezembro de 1999; o segundo sensor, presente na plataforma AQUA, foi lançado em maio de 2002. Tal sensor foi concebido pelo projeto *Earth Observing System* (EOS) da *National Aeronautical and*

*Spatial Administration* (NASA) para auxiliar os estudos das complexidades das mudanças globais. O MODIS foi projetado para obter informações sobre os três ambientes do sistema Terra: atmosfera, ecossistemas oceânico e terrestre; possui para isto bandas de resolução espectral e espacial específicas para o imageamento das feições de diversos alvos, como: propriedades das nuvens, dinâmicas e propriedades da vegetação e temperatura da superfície terrestre e oceânica. A partir dos dados deste sensor a NASA oferece 44 produtos dos quais 36 são disponibilizados ao público. Destes produtos, 3 são destinados a aplicações de calibração, 12 para estudos de ecossistemas terrestres, 6 a estudos atmosféricos e 15 para aplicações no estudo de ecossistemas oceânicos.

Esse sensor possui 36 bandas espectrais, estendendo-se da faixa do visível ao infravermelho termal. As sete bandas iniciais são voltadas principalmente para o sensoriamento remoto da superfície terrestre (denominadas de *land bands*). As resoluções espaciais variam da seguinte forma: 250 m para as bandas 1 e 2 (620-670 e 841-876 nm, respectivamente), 500 m para as bandas de 3 a 7 (459-479, 545-565, 1230-1250, 1628-1652 e 2105-2115 nm) e 1000 m para as demais bandas, sendo o período de revisita diário para as latitudes acima de 30° e de dois dias para as latitudes inferiores a 30° (MODIS, 2004). Isto permite que se faça a composição de várias imagens em um período de tempo (por exemplo: 16 dias) para resultar na geração de imagens livres de nuvens. As imagens diárias ou compostas são distribuídas gratuitamente no *web site* da EOS.

A EOS disponibiliza também outros produtos direcionados a aplicações específicas, como por exemplo, o produto MOD13Q1, que pode ser utilizado em aplicações na agricultura. Este produto contém informações sobre valores de NDVI obtidos pela composição de imagens no período de 16 dias, e com resolução espacial de 250 m.

### **2.3 Modelos de estimativa de produtividade**

As estimativas agrícolas são um importante meio para servir a agricultura, especialmente no que se refere ao planejamento. Instituições públicas, empresas agrícolas e outras entidades de planejamento necessitam saber, por exemplo, as datas de

ocorrência dos estágios fenológicos das culturas, para estimar o impacto de condições meteorológicas adversas ao desenvolvimento das mesmas, a fim de adotar medidas corretivas sempre que necessárias para fornecer condições adequadas de colheita, armazenagem, transporte, exportação, importação etc.

A estimativa da produtividade agrícola por meio de modelos matemáticos visa simular ou estabelecer relações entre as condições de crescimento das culturas e sua produtividade. Diversos trabalhos científicos relatam o desempenho destes modelos aplicando diferentes ênfases temáticas dependendo do interesse envolvido, tais como: agrometeorologia, fisiologia vegetal, estatística, economia, engenharia operacional, entre outras. Entretanto, o modelo se torna mais completo quando os vários aspectos relacionados com a formação da produtividade das culturas são considerados (Braga, 1995).

Modelos de estimativa de produtividade agrícola utilizando variáveis agrometeorológicas foram desenvolvidos, principalmente, ao longo dos últimos 30 anos e apresentam alguns resultados favoráveis para auxiliar na formação das estatísticas oficiais da produtividade de grandes culturas agrícolas (Braga, 1995). A maioria dos estudos sobre estimativa de produtividade estão relacionados aos cultivos anuais de cereais devido à sua função na produção mundial de alimentos e também por sua importância econômica no comércio internacional.

Os modelos de estimativa de produtividade ainda podem ser divididos em lineares e não lineares. Nos modelos lineares, para se estimar os parâmetros é necessário resolver um sistema de equações lineares com relação aos coeficientes de regressão desconhecidos. A solução do problema é única, ou seja, obtém-se uma forma analítica de estimação dos parâmetros, e esta forma é igual para qualquer modelo e qualquer conjunto de dados, um exemplo são os modelos de regressão lineares (Draper e Smith, 1966).

Já os modelos não-lineares não fazem uma descrição puramente empírica do fenômeno em estudo, trabalham na obtenção de uma relação teórica entre as variáveis observáveis de interesse a partir de suposições importantes sobre o problema. Uma vantagem desses modelos não-lineares é adquirir parâmetros que são facilmente interpretáveis. Além

disso, esses modelos geralmente necessitam de menos parâmetros do que os modelos lineares, o que facilita a interpretação e um bom ajuste do modelo (Draper e Smith, 1966).

### 2.3.1 Modelo agrônômico

Doorembos e Kassam (1979) propuseram um modelo agrônômico para estimar produtividade de diversas culturas. Tal modelo se baseia na penalização da produtividade máxima (PM) em função do estresse hídrico dado pela relação entre a evapotranspiração real (ET<sub>r</sub>) e a evapotranspiração máxima (ET<sub>m</sub>), limitado por um fator de resposta à produtividade (k<sub>y</sub>), cujo valor é dado em função da cultura e do seu estágio de desenvolvimento. A produtividade agrícola estimada (PR) pelo modelo desenvolvido pelos autores acima tem por base a Equação 2.2:

$$PR = PM \left[ 1 - k_y \left( 1 - \frac{ET_r}{ET_m} \right) \right] \quad (2.2)$$

Este modelo permite o cálculo da produtividade para diversas culturas. A produtividade final é o resultado do somatório das produtividades parciais obtidas em cada um dos intervalos de tempo desde o plantio até a colheita.

Delgado-Rojas e Barbieri (1999) desenvolveram um modelo com variáveis agrometeorológicas para estimar a produtividade agro-industrial da cana-de-açúcar baseado no modelo de penalização proposto por Jensen (1968). Os autores caracterizaram os efeitos da deficiência hídrica no solo sobre a produtividade agrícola (colmo) e a produtividade industrial (açúcar) em escala municipal. Os autores concluíram que, para a área experimental estudada, o modelo pode ser empregado para este tipo de estimativa, já que o erro médio absoluto representou 3,8% da produtividade média observada de colmo e 4,7% da produtividade média observada de açúcar. O erro máximo foi de 13% para colmo e 9,8% para o açúcar.

### 2.3.2 Modelo espectral e agrônômico-espectral

Nas últimas décadas, com o advento dos satélites de observação da Terra, que operam

de forma sistemática, considerável atenção foi dada ao desenvolvimento de modelos de produtividade que utilizam dados coletado por sensores abordo destes satélites. Muitos esforços foram concentrados no estabelecimento de relações empíricas entre parâmetros agronômicos (por exemplo: índice de área foliar e biomassa) e variáveis espectrais (por exemplo: índices vegetativos) para serem utilizados em modelos de estimativa de produtividade (Rudorff, 1985; Rudorff e Batista, 1990; Rudorff e Batista, 1991; Berka et al., 2003; Rizzi, 2004).

Rudorff e Batista (1990) propuseram um modelo de regressão com o uso de variáveis espectrais para melhorar a estimativa da produtividade de cana-de-açúcar obtida pelo modelo agronômico de Doorembos e Kassam (1979), numa área de 40.000 ha. A associação de dados agronômicos e espectrais gerou um modelo com maior poder explicativo.

Rudorff et al. (1995) utilizaram variáveis espectrais e também variáveis qualitativas (por exemplo: variedade e estágio de corte) e observaram melhora na estimativa produzida pelo modelo de regressão.

Berka et al. (2003) implementaram o modelo proposto por Doorembos e Kassam (1979) de forma espacial através do aplicativo SPRING utilizando a Linguagem Espacial para Geoprocessamento Algébrico (LEGAL) para a cultura da soja. O modelo foi utilizado para estimar a produtividade da cultura da soja em escala municipal e estadual, para o Estado do Paraná nas safras de 1996/97 a 2000/01 obtendo resultados satisfatórios em três dos cinco anos safra analisados.

Rizzi (2004), trabalhando com soja, adaptou o modelo de Berka et al. (2003) para o Estado do Rio Grande do Sul e inseriu o IAF gerado a partir de imagens adquiridas pelo sensor MODIS como variável espectral. O modelo foi calculado para os anos safra 2000/01, 2001/02 e 2002/03. Os resultados foram satisfatórios, e o modelo agronômico-espectral desenvolvido permitiu o monitoramento temporal espacial da evolução da produtividade durante todo o ciclo da cultura.

Modelos agronômicos implementados numa plataforma SIG são interessantes, pois



permitem examinar simulações de fenômenos espaciais e temporais, além de facilitar a manipulação, a análise e a visualização espacial dos dados (Hartkamp et al., 1999). Todavia, deve se tomar cuidado com a escala espacial e temporal do modelo, pois todos os dados devem estar na mesma escala. Também deve se tomar cuidado com o tipo de interpolação e o tipo de dado (raster ou vetorial) utilizado, entre outros.

## **2.4 Redes Neurais Artificiais**

Estudos envolvendo Redes Neurais Artificiais (RNA) têm aumentado nos últimos anos, porém a sua aplicabilidade no sensoriamento remoto ainda não é tão vasta. Uma Rede Neural Artificial é um sistema de processamento de informação não algorítmico, que se assemelha à estrutura do cérebro humano, inspirada nos neurônios biológicos.

O cérebro humano contém aproximadamente  $10^{11}$  neurônios. Cada um dos neurônios processa e se comunica com milhares de outros neurônios continuamente e em paralelo. Os neurônios são subdivididos em três partes, como mostra a Figura 2.4:

- Corpo da célula: processa a informação e gera impulsos nervosos;
- Dendrito: recebe as informações (impulsos nervosos), vindos de outros neurônios e conduz até o corpo celular;
- Axônio: transmite até os dendritos de outros neurônios os impulsos gerados pelo corpo da célula.

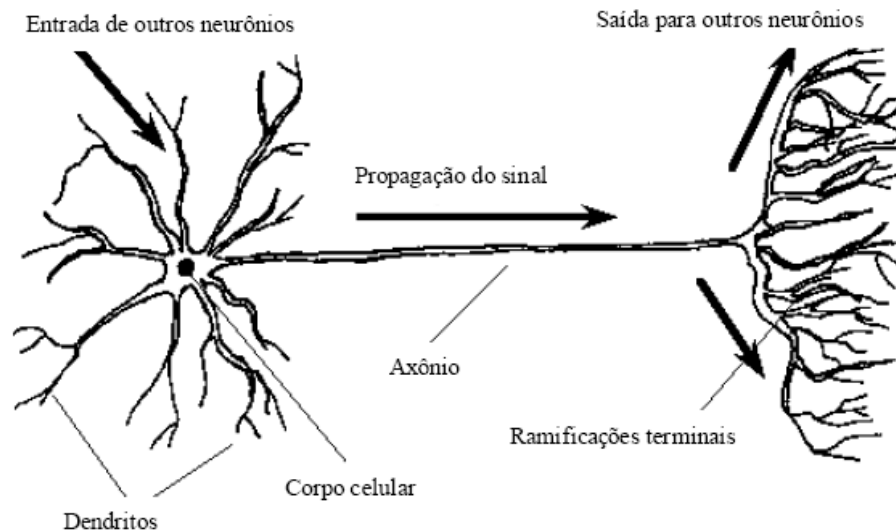


FIGURA 2.4 – Célula neural biológica com a seqüência de propagação do sinal.

FONTE: Silva (1998).

O ponto de contato entre a terminação axônica de um neurônio e o dendrito de outro é chamado de sinapse. São pelas sinapses que os neurônios se unem e formam as Redes Neurais, e também são elas que controlam as transmissões de impulsos (fluxo de informação) entre os neurônios da rede.

Os sinais originários dos neurônios pré-sinápticos passam para o corpo do neurônio, onde é feita uma comparação com outros sinais recebidos pelo neurônio. Se o percentual em um curto intervalo de tempo é suficientemente alto, a célula “dispara”, produzindo um impulso que é propagado para as células seguintes. As funções complexas surgem com a operação em paralelo de todos os  $10^{11}$  neurônio do cérebro humano (Carvalho et al., 1998).

As Redes Neurais Artificiais são compostas por sistemas paralelos distribuídos, constituídos por unidades de processamento simples (neurônios) que computam determinadas funções matemáticas (geralmente não-lineares). Esses neurônios são posicionados em uma ou mais camadas e interligados por conexões, geralmente unidirecionais. Na maioria das vezes, tais conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada

recebida por cada neurônio da rede (Carvalho et al., 1998).

Para se solucionar um problema utilizando RNA é necessário primeiramente passar pela etapa do aprendizado, na qual um grupo de exemplos é apresentado à rede, em que esta retira automaticamente as características indispensáveis para assim representar a informação necessária.

Uma das características mais relevantes de RNA é a sua capacidade em aprender por meio de exemplos e de generalizar a informação aprendida para se solucionar um problema, além da capacidade de auto-organização e de processamento temporal. Elas também são capazes de operar como mapeadores universais de funções multivariáveis, com um custo computacional que é linearmente proporcional ao número de variáveis (Haykin, 1999).

#### **2.4.1 Arquiteturas**

A arquitetura de uma RNA limita o tipo de problema que pode ser tratado pela mesma. Redes com apenas uma camada de neurônios resolvem apenas problemas linearmente separáveis, já as redes recorrentes resolvem problemas que envolvem processamento temporal (Haykin, 1999). Para definir a arquitetura de uma RNA é necessário determinar:

- Número de camadas:
- Número de neurônios em cada camada:
- Tipo de conexões entre os neurônios:
- Topologia da rede;
- Conectividade:

As quantidades de camadas e de neurônios em cada camada, a forma de conexão entre os neurônios na rede e os valores iniciais dos pesos, devem ser definidos antes do treinamento e dependem do problema que se deseja resolver.

A definição da rede é um processo de tentativa e erro. Pode acontecer de não se atingir uma significância desejável no treinamento, sendo necessária a repetição do processo usando uma arquitetura diferente e/ou com alterações nos valores dos pesos iniciais.

#### **2.4.2 Treinamento (ou Aprendizado)**

As redes aprendem por exemplos e fazem interpolações e extrapolações do que aprendem. Apresenta-se à rede um conjunto de procedimentos para que a mesma possa aprender uma determinada função. Esse conjunto é chamado de algoritmo de aprendizado. Tais algoritmos se diferem pela maneira que o ajuste dos pesos é feito.

É nesta fase que a rede extrai informações relevantes de padrões de informação apresentados para a mesma, criando uma representação própria para o problema. Nesta etapa também acontece o processo iterativo de ajuste de parâmetros da rede, quando os pesos das conexões entre as unidades de processamento, guardam, ao final do processo, o conhecimento que a rede adquiriu do ambiente que está operando.

Diversos métodos para o treinamento de redes foram desenvolvidos e podem ser agrupados em dois paradigmas principais (Carvalho et al., 1998):

- a) Aprendizado Supervisionado: quando é utilizado um agente externo que indica à rede a resposta desejada para o padrão de entrada;
- b) Aprendizado Não Supervisionado (auto-organização): quando não existe um agente externo indicando a resposta desejada para os padrões de entrada;

#### **2.4.3 Função de ativação**

Cada neurônio é capaz de processar um sinal de entrada e transformá-lo em um sinal de saída. O estado de ativação do neurônio é calculado a partir da aplicação de uma função de limiar ao valor de entrada fornecido ao neurônio, ou seja, a somatória dos valores de ativação dos neurônios precedentes, multiplicados pelos respectivos pesos.

A função utilizada para o cálculo de ativação geralmente é algum tipo de função não-linear, a qual garante a plena funcionalidade das redes neurais com múltiplas camadas

de neurônios. Redes neurais que processam dados analógicos devem utilizar funções com formato sigmoïdal como tangente hiperbólica, seno, etc. Em redes utilizadas para processar valores discretos a função deve ser do tipo degrau (Von Zubem, 1996).

De acordo com uma ponderação dos sinais de entrada, realizada pela função de ativação, o neurônio pode ser ativado, enviando um sinal de saída. Este sinal de saída será propagado de acordo com a topologia de interconexão da rede de neurônios.

Na literatura estão descritos alguns tipos de função de ativação mais empregados nas RNAs, principalmente as que possuem arquiteturas do tipo Perceptron de Múltiplas Camadas (MLP), descritas em Kosko (1992) e Haykin (1999).

#### **2.4.4 Redes do Tipo *Perceptron* de Múltiplas Camadas (MLP)**

As arquiteturas do tipo *Perceptron* de Múltiplas Camadas (MLP - *Multi Layer Perceptron*) constituem os modelos neurais artificiais mais utilizados e conhecidos. Tipicamente, esta arquitetura consiste de um conjunto de unidades sensoriais que formam uma camada de entrada, uma ou mais camadas intermediárias (ou escondidas) de unidades computacionais e uma camada de saída. Os sinais de entrada são propagados camada a camada pela rede em direção a saída (Cybenko, 1988 citado por Carvalho et al., 1998).

A Figura 2.5 apresenta uma arquitetura do tipo MLP com duas camadas intermediárias, com conexão acíclica e está completamente conectada, o que significa que um neurônio em qualquer camada da rede está conectado a todas as outras unidades (neurônios) na camada anterior (Carvalho et al., 1998).

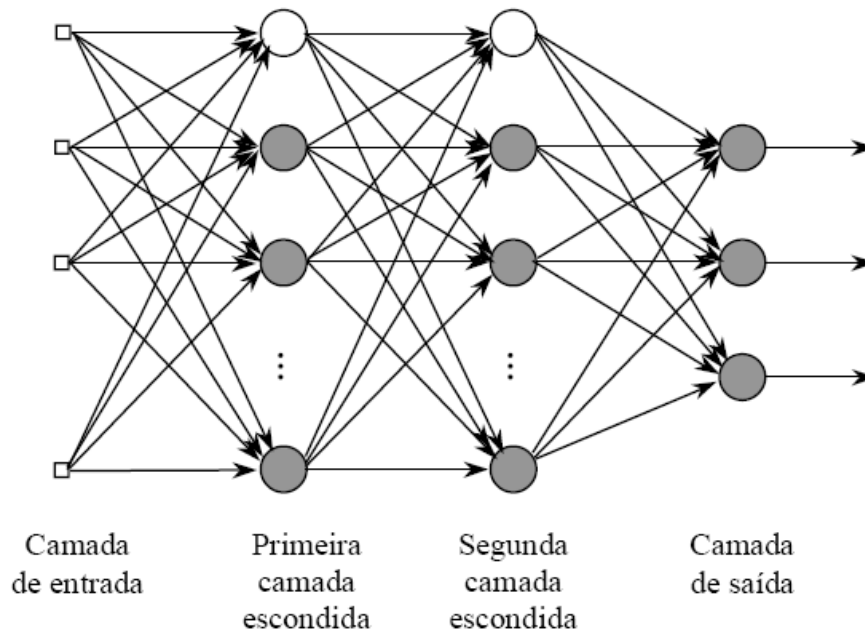


FIGURA 2.5 – Arquitetura MLP com duas camadas intermediárias.  
 FONTE: Adaptado de HAYKIN (1999).

A camada de entrada geralmente é composta por unidades que não modificam os sinais externos (neurônios sensoriais), apenas os distribuem para a primeira camada intermediária. As unidades de saída constituem a camada de saída da rede, e as demais unidades constituem as camadas intermediárias (Silva, 1998).

As redes do tipo MLP têm sido utilizadas com sucesso para a solução de vários problemas envolvendo altos graus de não-linearidade. O algoritmo de aprendizado mais utilizado para o treinamento destas redes é o algoritmo retro-propagação (*backpropagation*) (Rumelhart et al., 1986). Seu treinamento é do tipo supervisionado e é baseado numa regra de aprendizagem que “corrige” o erro durante o treinamento (Haykin, 1999).

O treinamento ocorre em duas fases, em que cada uma dessas fases percorre um sentido: uma fase de propagação do sinal funcional (*forward*) e uma de retro-propagação (*backward*). Na fase *forward*, os vetores de dados são aplicados às unidades de entrada, e seu efeito se propaga pela rede, camada a camada. Um conjunto de saídas é produzido

como resposta da rede. Durante esta fase os pesos das conexões são mantidos fixos. Na retro-propagação do erro, os pesos são ajustados de acordo com uma regra de correção do erro. Especificamente, a resposta da rede em um instante de tempo é subtraída da saída desejada (*target*) para produzir um sinal de erro, como mostra a Figura 2.6. Este sinal de erro é propagado da saída para a entrada, camada a camada, originando o nome “retro-propagação”. Os pesos são ajustados de forma que a distância entre a resposta da rede e a resposta desejada seja reduzida (Haykin, 1999).

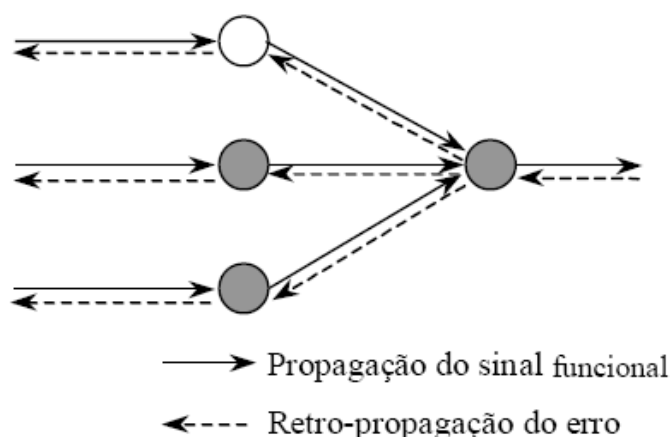


FIGURA 2.6 – Ilustração das direções de propagação do sinal funcional e do erro.

FONTE: Adaptado de Haykin (1999).

As redes que utilizam *backpropagation* trabalham com uma variação da regra delta (Windrow e Hoff, 1960 citado por Carvalho et al., 1998), apropriada para redes multicamadas: a regra delta generalizada. A regra delta padrão essencialmente implementa um gradiente descendente no quadrado da soma do erro para funções de ativação lineares. Redes sem camadas intermediárias, podem resolver problemas onde a superfície de erro tem a forma de um parabolóide com apenas um mínimo. Entretanto, a superfície do erro pode não ser tão simples, e suas derivadas mais difíceis de serem calculadas. Nestes casos, devem ser utilizadas redes com camadas intermediárias. Ainda assim, as redes ficam sujeitas ao problema de mínimos locais.

A regra delta generalizada funciona quando são utilizadas na rede unidades com uma função de ativação semi-linear, que é uma função diferenciável e não decrescente. A taxa de aprendizado é uma constante de proporcionalidade no intervalo  $[0, 1]$ , pois este

procedimento de aprendizado requer apenas que a mudança no peso seja proporcional à reta.

Entretanto, o verdadeiro gradiente descendente requer que sejam tomados passos infinitesimais. Assim quanto maior for essa constante, maior será a mudança nos pesos, aumentando a velocidade do aprendizado, o que pode levar a uma oscilação do modelo na superfície de erro. O ideal seria utilizar a maior taxa de aprendizado possível que não levasse a uma oscilação, resultando em um aprendizado mais rápido (Silva, 1998).

O treinamento das redes MLP com *backpropagation* pode exigir muitos passos no conjunto de treinamento, resultando num tempo de treinamento consideravelmente longo. Se for encontrado um mínimo local, o erro para o conjunto de treinamento pára de diminuir e estaciona em um valor maior que o aceitável. Uma maneira de aumentar a taxa de aprendizado sem levar à oscilação é modificar a regra delta generalizada para incluir o termo *momentum*, uma constante que determina o efeito das mudanças passadas dos pesos na direção atual do movimento no espaço de pesos (Haykin,1999).

O termo *momentum* torna-se útil em espaços de erro que contenham longas gargantas, com curvas acentuadas ou vales com descidas suaves, como o apresentado na Figura 2.7.

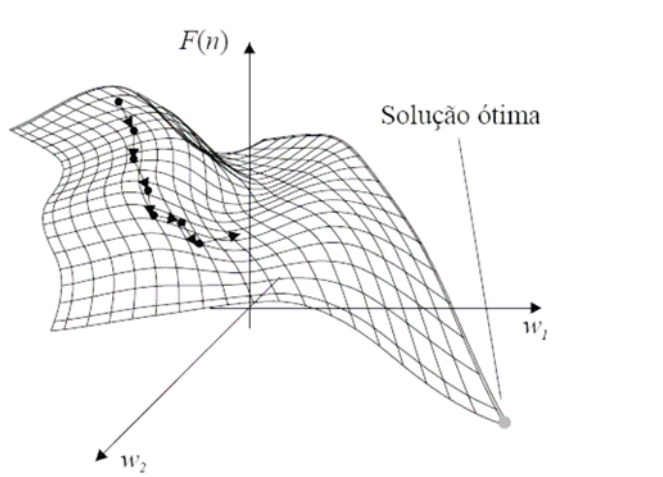


FIGURA 2.7 – Exemplo de superfície do erro.

FONTE: Carvalho (2005).



Outro problema enfrentado pelas RNA é o de “esquecimento”, em que a rede ao aprender novas informações, esquece as previamente aprendidas.

#### **2.4.5 Agregados de Redes Neurais Artificiais - *Ensembles***

*Ensemble* é um paradigma de aprendizado em que uma coleção finita de propostas alternativas para a solução de um problema, denominadas componentes do *ensemble*, é empregada em um conjunto na proposição de uma única solução para o problema (Sollich e Krogh, 1996). O trabalho de Hansen e Salamon (1990) deu origem aos *ensembles*, que comprovaram que a habilidade de generalização pode ser melhorada pela fusão de várias Redes Neurais Artificiais. Assim, as redes são treinadas independentemente e depois é feita uma composição das saídas individuais.

Geralmente *ensemble* é utilizado em pesquisas que envolvem problemas de classificação ou regressão, e pesquisas têm apresentado que um *ensemble* tende a ser mais exato que qualquer um dos classificadores ou regressores que o compõe.

Geralmente um *ensemble* é baseado em três passos: treinamento, seleção e combinação. Portanto, é necessário que o conjunto de dados seja dividido em três: a primeira parte para gerar os componentes, a segunda para a seleção e combinação e a terceira para testar o desempenho do *ensemble* (Lima, 2004).

Existem duas abordagens mais populares para a geração dos componentes: *boosting* e *bagging*. *Boosting* foi proposto por Schapire (1990) e aperfeiçoado por Freund (1995) e Freund e Schapire (1996). Neste método, os conjuntos de treinamento também são gerados a partir de amostragem com reposição, como no *bagging*. Porém a escolha de uma amostra depende da contribuição da mesma para o erro de treinamento dos componentes já treinados, ou seja, se uma amostra não foi classificada corretamente pelos componentes gerados, a probabilidade de que esta amostra seja novamente escolhida para compor novos componentes aumenta.

Breiman (1996) propôs o método *bagging* baseado na amostragem *bootstrap* (Efron e Tibshirani, 1993). Neste método, são gerados vários conjuntos de treinamento por meio

de amostragem com reposição do conjunto original dos dados, e a partir desses conjuntos de treinamentos se obtém uma proposta de solução. O número de amostras para os conjuntos é o mesmo, porém algumas amostras do conjunto de dados original podem aparecer mais de uma vez no mesmo conjunto de treinamento, e é essa distinção entre os conjuntos de treinamento que atribui a diversidade aos modelos de classificação ou regressão. Outras propostas de geração de componentes são descritas em Lima (2004).

Em relação aos métodos de combinação das redes neurais componentes, eles surgiram na Estatística e em Pesquisa Operacional. Há vários métodos diferentes de combinação. Esses métodos são específicos para problemas de classificação e regressão, e a abordagem mais utilizada é a média simples ou média ponderada, para problemas de regressão.

A grande motivação em se utilizar *ensemble* é que este apresente um incremento no desempenho, e isto é conquistado por meio de dois requisitos que os componentes do *ensemble* devem satisfazer: cada um dos componentes deve apresentar um bom desempenho quando analisado isoladamente e o componente deve ser o mais descorrelacionado possível quando comparado aos demais (Lima, 2004).

Um exemplo da aplicação de RNA foi o estudo de Braga (1995), que trabalhou com modelos de estimativa de produtividade da cultura de maçã nas principais fases fenológicas. O autor desenvolveu alguns modelos e comparou modelos de estimativa polinomial com Rede Neural Artificial, e também testou a combinação desses dois tipos de modelo na tentativa de minimizar os desvios e os erros relativos à estimativa. O autor concluiu que o desempenho da RNA foi semelhante aos dos modelos polinomiais, mas os modelos mistos (polinomiais + RNA) obtiveram um melhor desempenho na estimativa da produtividade.

## **CAPÍTULO 3**

### **MATERIAIS E MÉTODOS**

#### **3.1 Área de estudo**

A área de abrangência da Usina Catanduva abrange os seguintes municípios: Ariranha, Borborema, Catanduva, Catiguá, Cedral, Colina, Elisiário, Embaúba, Itajobi, Itápolis, Marapoama, Monte Azul Paulista, Novais, Olímpia, Palmares Paulista, Paraíso, Pindorama, Pirangi, Santa Adélia, Tabapuã, e Uchoa, totalizando uma área de aproximadamente 4.032 km<sup>2</sup> (Figura 3.1).

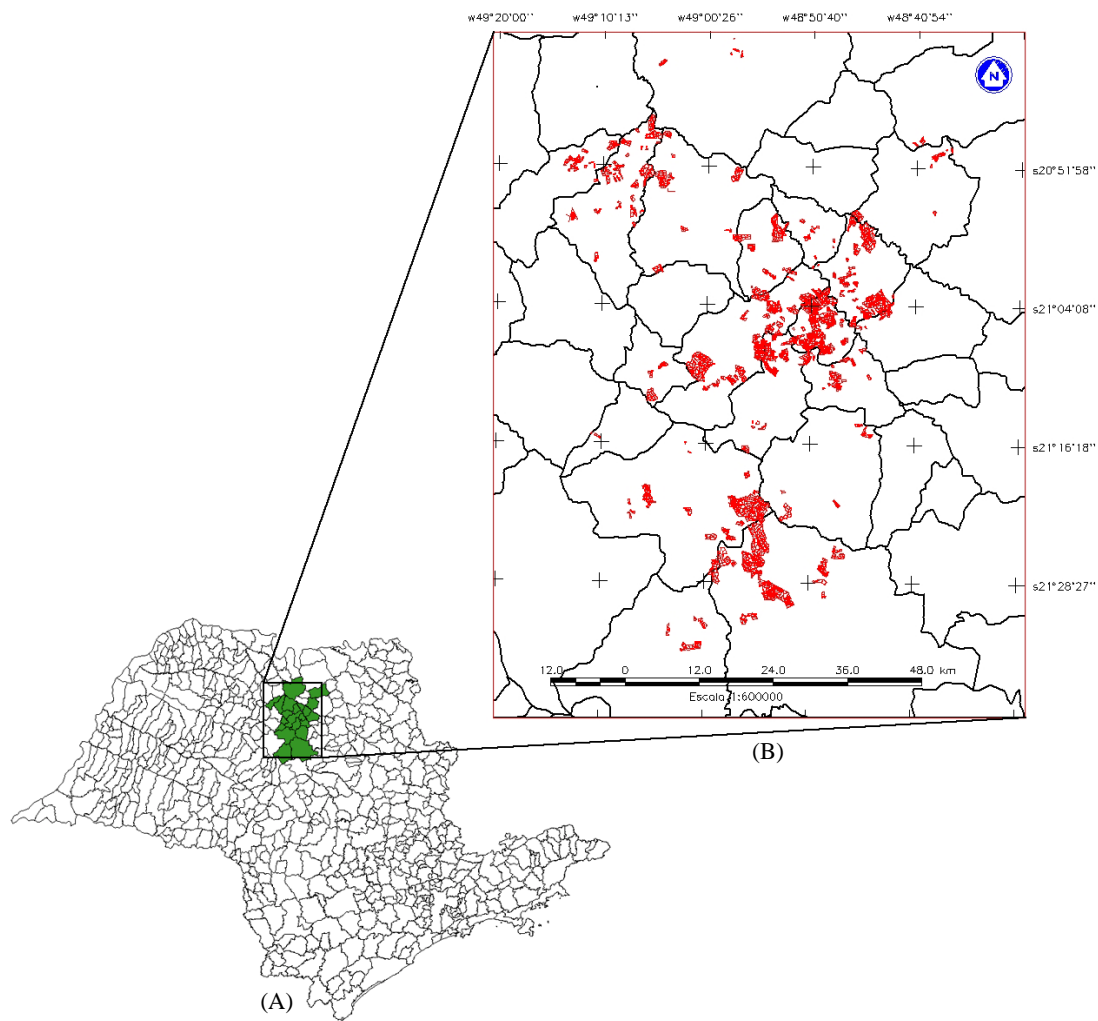


FIGURA 3.1 – (A) Estado de São Paulo; (B) Detalhamento da área em estudo e os talhões de cana-de-açúcar, na safra 2004/05.

Segundo o Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT (1981a), a área de estudo está localizada no Planalto Ocidental, o qual abrange cerca de 50% do Estado de São Paulo, ocupada por relevos monótonos de colinas e morrotes.

O relevo varia de 420 a 620 metros de altitude, a rede hidrográfica é organizada por rios conseqüentes, cursos d'água tributários do rio Tietê e do rio Grande. A região possui seis diferentes formas de relevo, sendo que três destas pertencem à classe de relevos de degradação, em planaltos dissecados (IPT, 1981a):

Segundo o IPT (1981b), a geologia da área em estudo, localiza-se na Bacia do Paraná,

pertencendo ao Grupo Bauru, Formação Adamantina (Ka). Descreve-se pela presença de depósitos fluviais com predominância de arenitos finos a muito finos, podendo apresentar cimentação e nódulos carbonáticos, com lentes de siltitos arenosos e argilitos, ocorrendo em bancos maciços; estratificação plano-paralela e cruzada de pequeno a médio porte.

A usina Catanduva possui um detalhado levantamento pedológico por talhão com os seguintes tipos de solos: Latossolo Vermelho Amarelo (LVA), Litólico (LI) e Podzólico Vermelho Amarelo (PVA).

A região apresenta o seguinte tipo climático segundo a classificação internacional de Köppen:

- Aw – clima tropical, com inverno seco, de maio a outubro, com temperatura média do mês mais frio superior a 18°C, e com estação chuvosa no verão, de novembro a abril;

A precipitação pluviométrica varia entre 1.100 e 1.300 mm anuais, e os meses de maior precipitação são dezembro, janeiro e fevereiro (CEPAGRI, 2006).

### **3.2 Materiais**

A Usina Catanduva possui três tipos de cana-de-açúcar quanto ao ciclo de desenvolvimento: cana de ano (CA), que são aquelas colhidas aproximadamente um ano após o plantio, cana de ano e meio (CAM), que são aquelas colhidas aproximadamente um ano e meio após o plantio, e por fim a cana soca (CS) que é a rebrota da cana de ano, da cana de ano e meio ou da própria soqueira. A Figura 3.2 apresenta uma análise preliminar feita nos tipos de cana-de-açúcar que a Usina Catanduva possui em porcentagem.

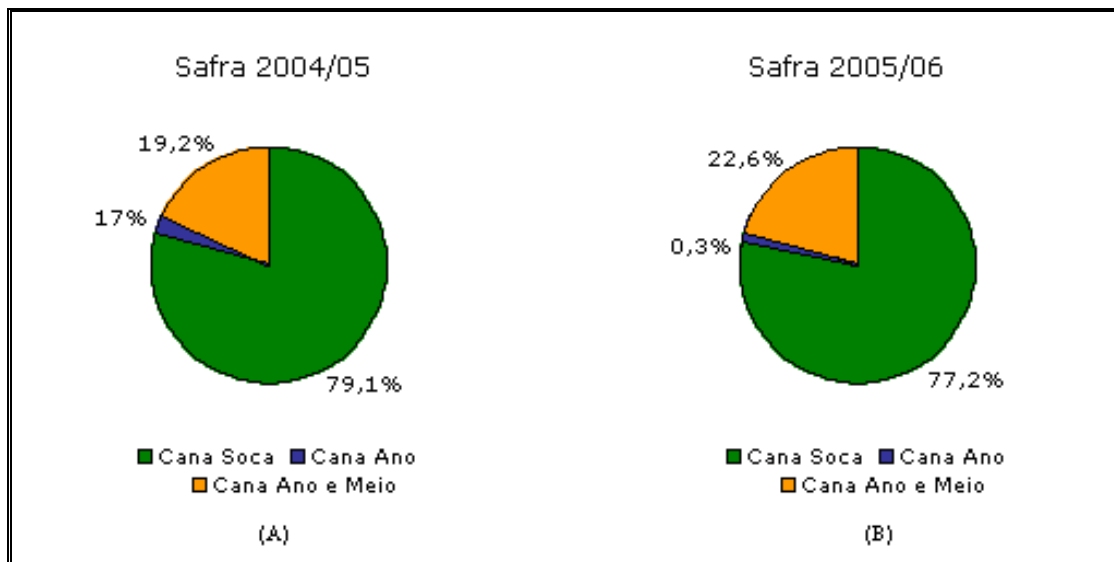


FIGURA 3.2 – Disposição do tipo de cana-de-açúcar na Usina Catanduva na safra (A) 2004/05 e (B) 2005/06.

Sendo assim decidiu-se por trabalhar apenas com os dados referentes à cana soca, pois representa a grande maioria da produção. Os dados referentes ao cultivo da cana de ano e o da cana de ano e meio são insuficientes para servir de entrada no modelo de estimativa de produtividade e, portanto, não foram considerados no presente estudo.

### 3.2.1 Dados Meteorológicos

As variáveis meteorológicas requeridas pelo modelo no cálculo da produtividade foram: temperatura média do ar ( $^{\circ}\text{C}$ ); velocidade do vento a 2 metros de altura ( $\text{km dia}^{-1}$ ); umidade relativa do ar (%); radiação ( $\text{W m}^{-2}$ ) e precipitação pluvial ( $\text{mm dia}^{-1}$ ). São necessários, ainda, dados adicionais referentes à altitude e à capacidade de armazenamento de água dos solos (CAD). Os dados referentes à CAD foram fornecidos pela Usina Catanduva. As variáveis meteorológicas foram obtidas junto ao Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) do INPE. Esses dados foram calculados pelo modelo ETA, que é um modelo de mesoescala, em ponto de grade de equações primitivas (Black, 1994). Porém, a versão do modelo ETA, que é executada

no CPTEC, é hidrostático e cobre a maior parte da América do Sul e oceanos adjacentes. Sua resolução horizontal é de 40 km e a vertical de 38 camadas. Os dados de altitude foram adquiridos do Eros Data Center do US Geological Survey (USGS).

Os dados meteorológicos foram organizados em médias quinzenais, com exceção da precipitação, que foram somados quinzenalmente.

Os dados fornecidos pelo modelo ETA representam amostras pontuais e foram inseridos ao modelo no formato ASCII (*American Standard Code for Information Interchange*), que contém as coordenadas geográficas e o dado meteorológico quinzenal. Com o uso dessas amostras pontuais foram geradas grades regulares com resolução de 250 m pelo método de interpolação do vizinho mais próximo.

### **3.2.2 Dados Espectrais**

Os dados espectrais referentes ao Índice de Vegetação por Diferença Normalizada (NDVI) foram adquiridos a partir das imagens do sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*), a bordo da plataforma TERRA. Foi utilizada a composição de 16 dias para os anos de 2003 a 2005, no formato HDF (*Hierarchical Data Format*) e reprojeta para o formato Geo-TIFF (*Georeferenced Tag Image File Format*) pelo *software* Modis Reprojection Tool 3.2, desenvolvido pela EOS.

Uma análise preliminar foi feita nos dados MODIS, e duas comparações foram feitas: 1) comparação entre o NDVI do MODIS da segunda quinzena de fevereiro/2004 com o NDVI do sensor *Thematic Mapper* (TM) do dia 27 de fevereiro de 2004; e 2) comparação entre os valores de NDVI do MODIS da primeira quinzena de fevereiro/2005 e do NDVI do TM do dia 13 de fevereiro de 2005. Segundo Rudorff (1985), as imagens adquiridas pouco antes do início da safra (fevereiro/março) fornecem a melhor correlação entre o índice vegetativo e a produtividade agrícola da cana-de-açúcar.

### **3.2.3 Dados Qualitativos e Quantitativos**

Os dados qualitativos utilizados neste trabalho e que estão relacionados com a produtividade foram: variedade plantada, tipo de solo, estágio de corte, aplicação de vinhaça e o ano safra. Os dados quantitativos se referem à produtividade de cana-de-açúcar estimada pela Usina além da produtividade real utilizada para avaliação dos resultados do modelo. Todos esses dados foram fornecidos pela Usina Catanduva. Por sua vez, estes foram importados no formato tabular relacionado com o mapa cadastral dos talhões de cana fornecidos pela Usina Catanduva.

## **3.3 Método**

### **3.3.1 Estimativa da Produtividade**

Primeiro foi calculado o modelo agronômico-espectral. Este modelo havia sido inicialmente proposto por Doorembos e Kassam (1979), em seguida ele foi adaptado por Rudorff (1985) e por fim, foi implementado por Berka et al. (2003) e por Rizzi (2004) no SPRING (Sistema de Processamento de Informações Georeferenciadas). Em seguida o resultado desse modelo agronômico-espectral foi integrado a dados qualitativos e quantitativos em um novo modelo de Agregados de RNA ou *ensembles*. Conforme mencionado anteriormente o modelo *ensembles* estimou a produtividade da cana-de-açúcar para as safras 2004/05 e 2005/06.

#### **3.3.1.1 Modelo agronômico-espectral**

O modelo agronômico-espectral foi executado com uma resolução espacial de 250 m, para ser compatível com a resolução dos dados MODIS. Para os dados meteorológicos foram geradas grades regulares (matrizes) onde cada pixel possuía uma área equivalente a 250 m, assim o modelo calculou a produtividade de cana-de-açúcar para cada pixel a cada 15 dias. Porém, a área dos pixels foram reamostradas para 30 m, por causa das limitações do *software* SPRING. A soma dos cálculos quinzenais é o valor da produtividade final da safra.

Além de georreferenciar os dados meteorológicos e os dados fornecidos pela Usina



Catanduva, foram geradas grades para dados de CAD e para dados de altitude, em que foi utilizado um modelo digital de elevação (DEM), com resolução de 1000 m, reamostrada para 250 m, disponibilizado pelo Eros Data Center do USGS.

Segundo o método proposto por Choudhury et al. (1994) para estimativa do índice de área foliar (IAF), as imagens de NDVI foram primeiramente transformadas em imagens de fração da cobertura do solo ( $F_c$ ) conforme a Equação 3.1:

$$F_c = 1 - \left( \frac{NDVI_{\max} - NDVI}{NDVI_{\max} - NDVI_{\min}} \right)^{0,9} \quad (3.1)$$

em que,

$F_c$  = fração do solo coberto pela cultura;

$NDVI_{\max}$  = valor máximo do NDVI da imagem;

$NDVI_{\min}$  = valor mínimo do NDVI da imagem;

$NDVI$  = valor do NDVI de cada pixel da imagem.

O IAF é estimado pela Equação 3.2, seguindo a metodologia sugerida por Norman et al. (2003).

$$IAF = -2 \ln(1 - F_c) \quad (3.2)$$

O fator de compensação do crescimento ( $F_{cc}$ ) independe da cultura. Esse fator é utilizado no cálculo da produtividade, e está relacionado ao índice de área foliar (IAF). Como indica a Equação 3.3, ajustada por Sugawara (2002), baseada nos dados de Doorenbos e Kassam (1979):

$$F_{cc} = 0,515 - e^{(-0,664 - (0,515 * IAF))} \quad (3.3)$$

cujos valores do IAF são obtidos a partir das imagens do sensor MODIS, conforme descrito anteriormente.

O programa em linguagem LEGAL utilizado neste trabalho que transformou os valores de NDVI em IAF pode ser visto no Apêndice A. Em seguida, foi confeccionado o mapa temático para cana soca.

Parâmetros específicos da cultura da cana-de-açúcar foram necessários para que o modelo agrônômico-espectral fosse executado. Tais parâmetros tiveram seus valores alterados de acordo com o estágio de desenvolvimento da planta. Um maior detalhamento sobre estes valores pode ser encontrado em Doorembos e Kassam (1979) e Rudorff (1985). Os parâmetros utilizados para o cálculo do modelo foram:

- Coeficiente de cultura ( $k_c$ ): apresenta a relação entre a evapotranspiração máxima da cultura em questão e a evapotranspiração de referência. Esta varia conforme o aumento de área foliar e o estágio de desenvolvimento da cultura.
- Fator de resposta à produtividade ( $k_y$ ): se refere à resposta da vegetação à disponibilidade hídrica e expressa a diminuição relativa da produtividade em função da diminuição da evapotranspiração.
- Profundidade do sistema radicular ( $D$ ): refere-se à profundidade que as raízes exploram a água no solo.

Os valores de  $k_c$ ,  $k_y$  e  $D$  utilizados neste trabalho estão na Tabela 3.1.

TABELA 3.1 – Valores dos parâmetros  $k_c$ ,  $k_y$  e  $D$  utilizados para o cálculo do modelo agronômico-espectral, para a cana soca.

Quinzenas	$k_c$	$k_y$	$D$
1	0,30	0,30	0,5
2	0,40	0,30	0,6
3	0,50	0,35	0,7
4	0,70	0,40	0,8
5	0,80	0,40	0,9
6	0,90	0,40	1,0
7	1,10	0,50	1,0
8	1,10	0,50	1,0
9	1,20	0,50	1,0
10	1,20	0,50	1,0
11	1,30	0,50	1,0
12	1,30	0,50	1,0

É importante salientar que pelo fato de ser um modelo de previsão da estimativa de produtividade este foi calculado até o mês de abril do ano de colheita, ou seja, o modelo estimou o valor da produtividade para o mês de abril, antes do começo da colheita. Por exemplo: um talhão com cana soca teve seu último corte na segunda quinzena do mês de agosto de 2003 e o modelo de estimativa de produtividade terá o valor de quanto esse talhão irá produzir no mês de abril de 2004, independente se este talhão será colhido em agosto, setembro ou outubro de 2004.

Além dos parâmetros citados acima ( $k_c$ ,  $k_y$  e  $D$ ) também foi utilizado o fator de produtividade agrícola (FPA), que tem um elevado peso no cálculo da estimativa da produtividade e tem o mesmo valor para todo o ciclo da cultura, e independe do ano safra. O modelo agronômico-espectral estima a produtividade com base na matéria seca da planta, e o FPA é multiplicado pelo valor de produtividade obtido em cada quinzena. Neste trabalho o FPA utilizado foi 2,9, este valor foi baseado no trabalho de Rudorff (1985).

Após o ajuste dos parâmetros, o modelo foi executado para os anos safra 2004/05 e 2005/06. A Figura 3.3 apresenta a interface do modelo agronômico-espectral presente

no software SPRING.

Modelo AgroMeteorologico

Leitura de Parâmetros

Cultura: Cana      Período: Quinzenal

Ano Inicial: 2004      Ano Final: 2005

Mês Inicial: OUT      Mês Final: MAR

Per. Inicial: 1º período      Per. Final: 2º período

Parâmetros de Entrada

C:\Michelle\dissertação\mapa-usina\MODELO\_CS\      Procurar...

Criar Plano de Informação Parcial

Média Zonal

PI Tematico: CS\_2004      Classe: cs\_2004

PI Cadastral: Talhao\_final

Objeto...: export\_poLO

Categoria de Saída

Categoria...: MNT\_Resultad

Executar      Fechar      Ajuda

**Parâmetros de entrada:**

- Kc;
- Ky;
- D;
- FPA;
- Precipitação;
- Temperatura;
- Umidade relativa do ar;
- Velocidade de vento;
- Radiação solar;
- CAD;
- Altitude do terreno;
- Valores de IAF (estimados NDVI – MODIS).

FIGURA 3.3 – Interface do modelo agrônômico-espectral.

A saída do programa é dada por meio de tabelas confeccionadas no aplicativo ACCESS e por meio de mapas em forma de plano de informação no programa SPRING.

Por fim, foi feito um estudo para verificar a relevância da utilização dos dados MODIS na estimativa da produtividade. Isso foi feito utilizando uma média dos valores de NDVI para cada quinzena, ou seja, calculou-se a média dos valores de NDVI de todos os talhões da primeira quinzena de outubro, da segunda quinzena de outubro, até a segunda quinzena de março. Por fim, obteve-se um único valor de NDVI para toda área de estudo por quinzena. Então se calculou o modelo agrônômico-espectral com esses valores médios.

Após a execução do modelo, fez-se uma análise comparando os resultados deste com os valores da produtividade real.

### **3.3.1.2 Organização e codificação dos dados de cana soca para a entrada nos *ensembles***

Um dos dados de entrada dos *ensembles*, para estimar a produtividade da cana, é o valor da produtividade estimada pelo modelo agrônômico-espectral para cada talhão. Outro dado disponível usado nos *ensembles* é a estimativa da produtividade da Usina obtida por meio de um levantamento de campo, por ocasião do início da safra. Uma variável também disponível na Usina e que poderia ajudar a explicar a variação da produtividade da cana é a produtividade real do ano safra anterior.

Rudorff e Batista (1990) obtiveram relações razoáveis do índice vegetativo de bandas do sensor MSS do Landsat-4 com a produtividade real para imagens adquiridas pouco antes do início da safra (fevereiro/março). Assim, no presente trabalho foi adquirido duas imagens do sensor TM a bordo do Landsat-5 de 27 de fevereiro de 2004 e de 13 de fevereiro de 2005. Estas imagens foram transformadas no NDVI e o valor médio de NDVI de cada talhão foi utilizado nos *ensembles*.

As variáveis qualitativas utilizadas e disponíveis na Usina são: tipo de solo; tipo de variedade da cana; se houve aplicação de vinhaça; estágio de corte; e o ano safra.

Para se chegar nessas nove variáveis (5 qualitativas, 3 quantitativas e 1 espectral) que iriam compor os Agregados de RNA (*ensembles*) foi preciso antes fazer várias associações entre elas e então executar os *ensembles* com cada uma das combinações. Depois foi observada com qual combinação se obtinha o menor Erro Quadrático Médio (EQM). E assim se chegou a essas nove variáveis, pois a princípio seriam dez, aumentaria a variável tipo de corte da cana.

Os dados qualitativos foram transformados em binários, para que pudessem ser introduzidos nas Redes Neurais Artificiais.

Cada um dos registros continha as informações por talhão, ou seja, cada linha

representava um talhão da Usina. Os dados das duas safras foram agregados e separou-se um conjunto com 80% dos dados para treinar, selecionar, validar e testar os *ensembles*. Os 20% restantes foram utilizados para testar e avaliar os melhores *ensembles*.

### 3.3.1.3 *Ensembles* para integração dos dados

Os dados utilizados no cálculo dos *ensembles* foram:

1. Produtividade estimada pelo modelo agronômico-espectral;
2. Estimativa da produtividade da Usina Catanduva;
3. Produtividade real do ano safra anterior;
4. NDVI do Landsat-TM (27 de fevereiro de 2004 e de 13 de fevereiro de 2005);
5. Tipo de solo (p. ex., solo do tipo PVA 25 + LI 7, Podzólico Vermelho Amarelo, textura arenosa/média, horizonte A:30 a 60cm, eutrófico (PV-4) + Solos Litólicos, textura média, substrato arenito de Bauru, eutrófico (LV-2));
6. Tipo de variedade da cana (p. ex., SP83-2847);
7. Aplicação de vinhaça;
8. Estágio de corte (2º corte ou 1º soca; 3º corte ou 2º soca; 4º corte ou 3º soca; 5º corte ou 4º soca);
9. Ano safra.

Organizou-se ainda o arquivo que continha os dados de saída com o valor da produtividade real e outro arquivo que armazenava os dados para validação.

Utilizou-se no *software* MATLAB 6.5 o pacote *Ensemble\_Bagging* desenvolvido por Von Zubem et al. (2005).

Assumiu-se que o número de neurônios de entrada fossem 9, com uma camada oculta

que continha 20 neurônios e um neurônio de saída, pelo fato da saída desejada ser o valor de produtividade da cana-de-açúcar.

Para a obtenção dos *ensembles* foi treinado um conjunto de 10 Redes Neurais Artificiais. Foram 354 dados de entrada, 80% dos dados do conjunto original, sendo que este grupo de dados foi subdividido da seguinte forma: 40% para o treinamento dos dados, 10% para validação, 30% para seleção e 20% para teste. O número máximo de épocas de treinamento foi 500 e o número de simulações foram 10. Dessas 10 simulações o programa apresentava apenas o melhor resultado dentre os 10. A Figura 3.4 apresenta a interface do programa *Ensemble\_Bagging*, onde foi possível entrar com os parâmetros acima citados, além de carregar os dados de entrada, saída desejada e o conjunto de dados para validação.

Esse mesmo processo descrito acima foi feito novamente só que utilizando oito das nove variáveis iniciais de entrada. A variável produtividade estimada pela Usina Catanduva foi retirada da análise com o intuito de observar se esta realmente influencia no cálculo da estimativa de produtividade agrícola da cana. Pois é uma variável que envolve custo, já que é necessária a contratação de técnicos especializados para estimá-la, além de ser uma variável subjetiva. Nessa nova estimativa a única diferença é que quando se calculou os *ensembles* a camada de entrada das RNA possuía 8 neurônios ao invés de 9.

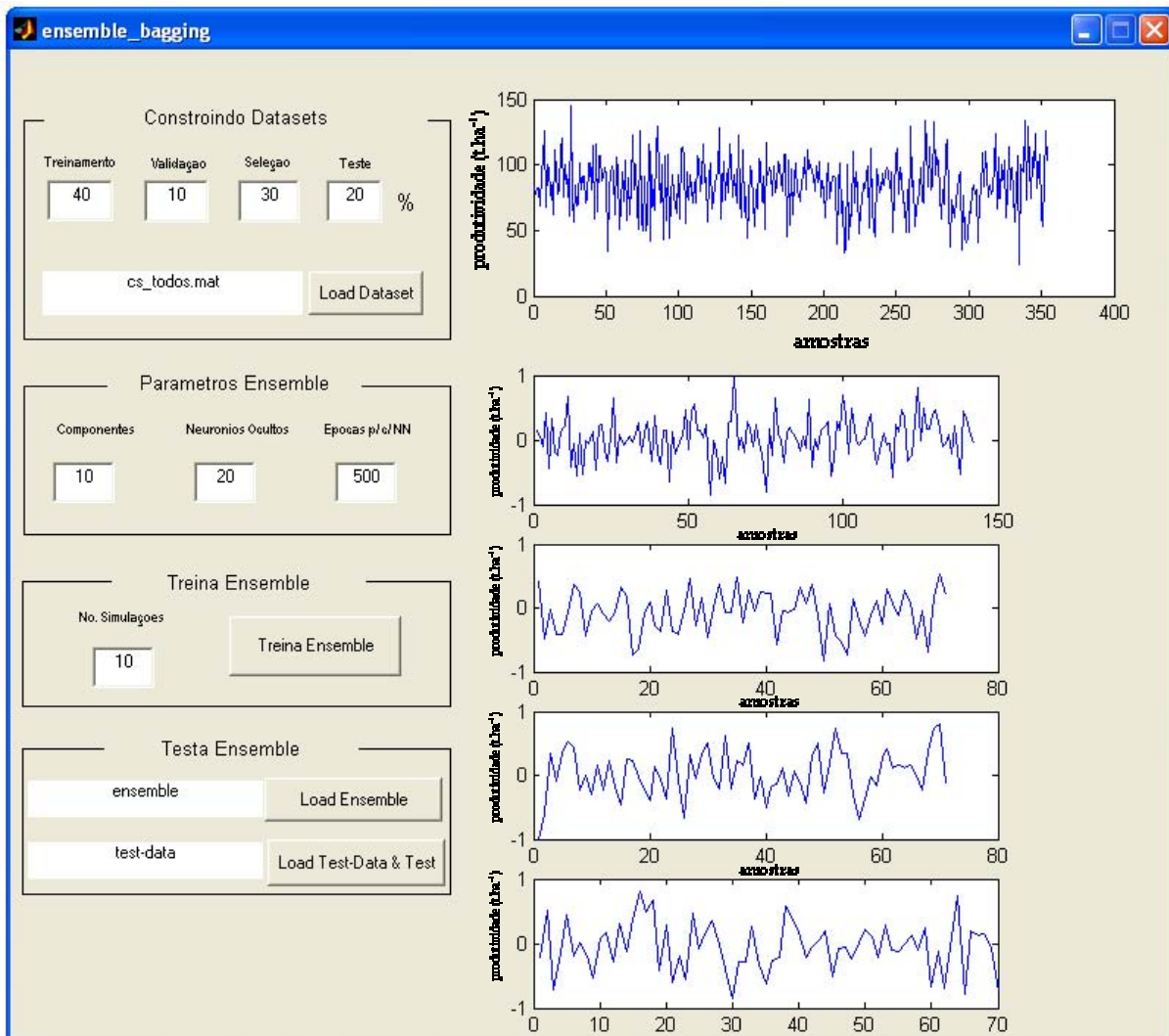


FIGURA 3.4 – Interface do programa *Ensemble\_Bagging*.

A rede utilizada foi do tipo *Perceptron* de Múltiplas Camadas (MLP), com uma camada oculta.

### 3.3.1.4 Treinamento, Seleção, Validação e Teste

O treinamento escolhido para a rede MLP foi o supervisionado e o algoritmo de aprendizado utilizado para o treinamento desta rede foi o algoritmo retro-propagação (*backpropagation*).

O pré-processamento dos dados de treinamento foi feito com o objetivo de produzir conjunto de treinamentos distintos, para que os *ensembles* pudessem generalizar de diversas formas, mesmo que a arquitetura, os pesos iniciais e o algoritmo de



treinamento fossem os mesmos.

A técnica de reamostragem utilizada para criar subconjuntos de treinamento, utilizados para obtenção de componentes de um *ensemble*, a partir do único conjunto de treinamento, foi o método *bagging*, que consistiu em uma amostragem com reposição. Os componentes que foram treinados deviam ser instáveis, para produzirem comportamento distinto sempre que submetidos a conjuntos de treinamento distintos. Esse método buscou componentes que generalizassem de forma distinta, não necessariamente de forma ótima, ou seja, não precisou que a instabilidade fosse evitada. Mesmo que os componentes não apresentassem uma boa generalização, a agregação deles tendeu a uma generalização adequada. Porém a capacidade de generalização dos componentes não pode ser tão ruim, pois a melhora obtida com a agregação pode não ser suficiente (Lima, 2004).

Em seguida, as redes foram validadas no sentido de obter um Erro Quadrático Médio (EQM) que fosse utilizado como critério de parada para o conjunto de treinamento. Para isto foram executadas várias interações até encontrar o erro mínimo global do conjunto de validação. Posteriormente o treinamento foi novamente executado até o ponto onde o erro mínimo foi encontrado.

Na fase de seleção, que teve como objetivo maximizar o desempenho de generalização do *ensemble*, o método utilizado para “escolher” as redes que iriam compor o *ensemble* foi o método Poda, que se baseia na redução do número de componentes do Agregado de RNA (*ensemble*) visando um incremento na capacidade de generalização. Portanto, quando existem vários candidatos disponíveis, o uso de parte do conjunto de candidatos pode produzir melhores resultados que a combinação de todos eles. O método foi detalhado por Lima (2004).

Depois de selecionadas as redes que iriam compor o *ensemble* foi necessário optar por um método para se fazer a combinação das redes. O método escolhido foi o da média simples, que é um dos métodos mais utilizados, e consistiu na combinação linear sobre a saída de  $n$  componentes (Perrone e Cooper, 1993; Opitz e Shavlik, 1996).

A fase de teste foi feita para observar o Erro Médio Quadrático dos melhores *ensembles*. Observou-se o EMQ do melhor componente isolado; do *ensemble* que continha todos os componentes; e por fim o do *ensemble* que continha os melhores componentes juntos (melhor *ensemble*).

Todas essas fases citadas acima foram repetidas 30 vezes, ou seja, por fim havia 30 melhores *ensembles*. Para avaliar o desempenho desses melhores *ensembles* pegou-se 20% dos dados do conjunto original e os validou nesses 30 melhores *ensembles*. E finalmente comparou-se a estimativa fornecida por esses *ensembles* com os dados de produtividade real fornecidos pela Usina Catanduva.

Para a avaliação do desempenho do melhor *ensemble* na estimativa da produtividade da cana, utilizou-se análise de regressão e o índice de concordância  $d$ , proposto por Willmott et al. (1985), que quantifica numericamente a exatidão do modelo. Ele mostra como o modelo simula os valores observados, refletindo, numa escala de 0 a 1, o grau do desvio da linha 1:1, que representa a correlação perfeita entre os valores.

## **CAPÍTULO 4**

### **RESULTADOS**

#### **4.1 Análise preliminar e seleção dos dados**

Inicialmente foi feita uma análise dos dados fornecidos pela Usina para comparar os valores das áreas dos talhões de uma tabela com os valores das áreas dos respectivos talhões obtidos a partir de um mapa digital. Nesta comparação, foi observado que diversos talhões apresentaram diferença no valor da área de tal forma que foram eliminados do estudo, por apresentarem informação conflitante e, portanto, não confiável.

Talhões com produtividade real muito diferente da esperada, ou seja, muito acima ou abaixo do valor de produtividade inicialmente estimado pela Usina também foram eliminados da análise, pois no cálculo da produtividade real destes talhões pode ter havido influência da cana de algum talhão vizinho durante o processo de corte e transporte da cana para a Usina. Isto geralmente ocorre em talhões pequenos de tal forma que talhões com menos de 10 ha foram eliminados da análise. Além disso, a menor área vista pelo sensor MODIS é de 6,25 ha (250 x 250 m). Finalmente, talhões com soqueiras acima do quinto corte não foram considerados, pois a maioria destes talhões é renovada em função da significativa redução na produtividade, permanecendo apenas os talhões com elevada produtividade.

A fim de verificar a qualidade dos dados do NDVI do sensor MODIS em cada uma das quinzenas foi feita uma análise de correlação entre o NDVI de uma quinzena e o NDVI da quinzena seguinte uma vez que é de se esperar uma boa correlação entre datas subseqüentes. Esta análise permitiu observar que na primeira quinzena do mês de novembro de 2003 existia um problema, sem uma razão definida, com os valores de NDVI do MODIS. Para sanar este problema, os valores da primeira quinzena de novembro de 2003 foram substituídos pela média entre a segunda quinzena de outubro e

a segunda quinzena de novembro de 2003.

Por fim, foi feita análises entre os dados de sensor MODIS e do sensor TM. Foi possível observar nestas análises que existe uma correlação relativamente alta entre os dados de NDVI do MODIS e do TM. O coeficiente de correlação foi de 0,60 para a safra 2004/05 e de 0,79 para a safra 2005/06. Isto indica que as imagens MODIS, apesar da baixa resolução espacial, fornecem informações condizentes com sensores de resolução espacial média.

#### **4.2 Modelo agronômico-espectral**

A fim de definir o período no qual o modelo agronômico-espectral deveria estimar a produtividade foi feita uma análise da variação dos valores de NDVI ao longo do ano. Para tal, foi selecionado um conjunto de amostras aleatórias de aproximadamente 15% dos talhões de cada ano safra. A trajetória temporal dos valores de NDVI deste conjunto de amostras pode ser observada na Figura 4.1. Nota-se, nesta figura, que os valores de NDVI começam a aumentar com o início das chuvas em setembro/outubro e tendem a estabilizar em março/abril quando a colheita é iniciada. Ou seja, este é o período de maior acúmulo de biomassa pela cultura conforme também relatado por Castro (1999). Assim, decidiu-se calcular a produtividade pelo modelo agronômico-espectral para o período compreendido entre a primeira quinzena de outubro e a segunda quinzena de março, conforme indicado na Figura 4.1.

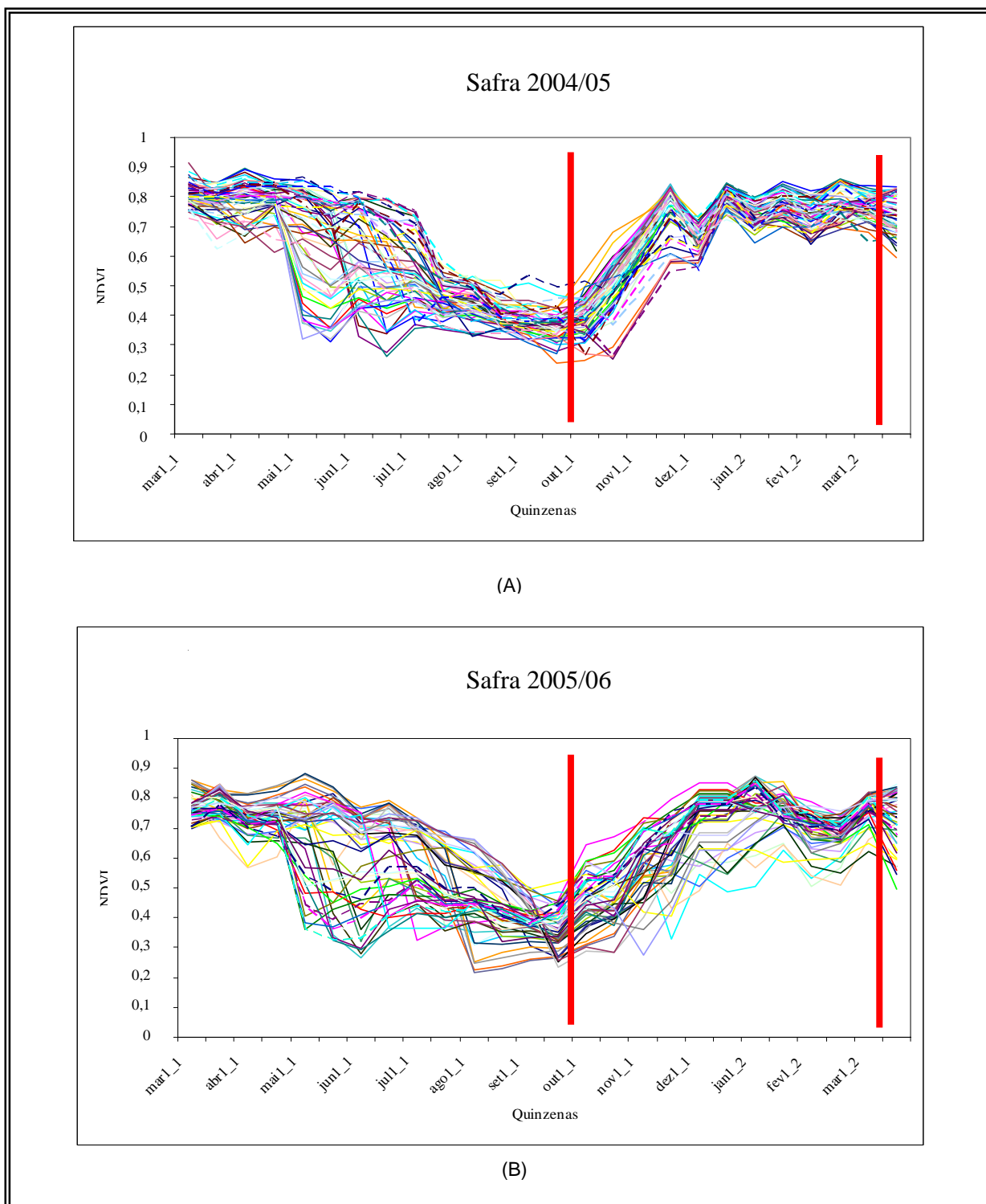


FIGURA 4.1 – Variação temporal do NDVI do MODIS para uma amostra de talhões das safras (A) 2004/05 e (B) 2005/06.

O cálculo da produtividade por talhão para os anos safra 2004/05 e 2005/06 foi feito através do modelo agronômico-espectral, implementado no *software* SPRING. O coeficiente de determinação para a regressão da produtividade do modelo agronômico-espectral com a produtividade real foi de 0,31 e 0,25 para os anos safra 2004/05 e 2005/06, respectivamente (Figura 4.2). A variação da produtividade não explicada por este modelo se deve em parte às suas próprias limitações, mas também à qualidade dos dados de entrada, por exemplo, a diferença existente entre os valores interpolados dos dados meteorológicos e os realmente observados. Outros fatores não considerados nesse modelo (por exemplo: estágio de corte, fertilidade, fitossanidade, etc.) também influenciam na variação da produtividade não explicada (Rudorff e Batista, 1990).

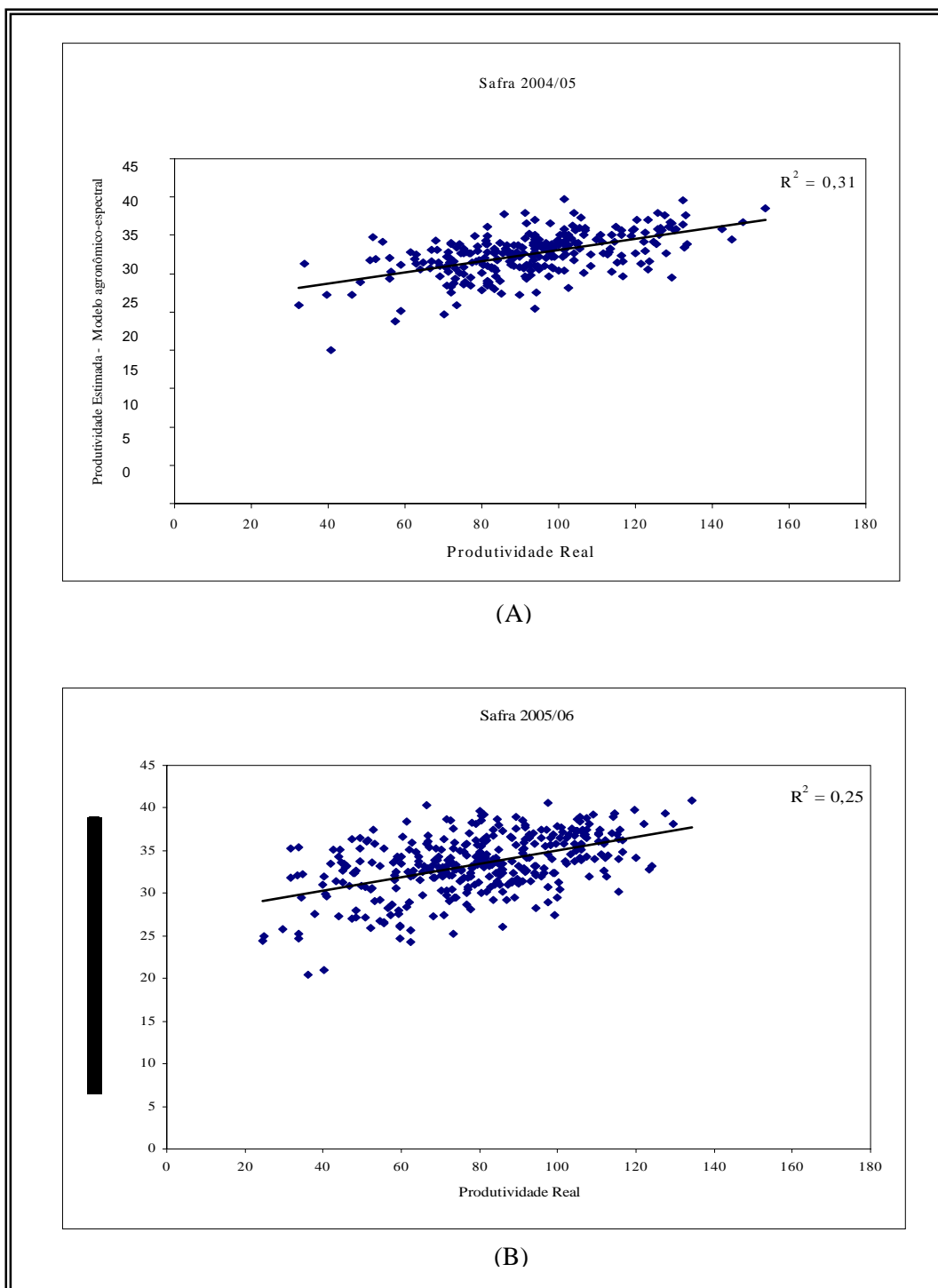


FIGURA 4.2 – Relação entre a produtividade real e a produtividade estimada pelo modelo agrônômico-espectral, safras (A) 2004/05 e (B) 2005/06.

A produtividade estimada pelo modelo agronômico-espectral para cana soca em cada ano safra foi utilizada como uma das variáveis quantitativas nos *ensembles*.

Em seguida, o modelo agronômico-espectral foi executado utilizando apenas uma média dos valores de NDVI para cada quinzena. Os valores dos coeficientes de determinação para a safra 2004/05 e 2005/06 não foram significativos quando calculou o modelo utilizando um único valor de NDVI por quinzena para toda a área de estudo. Portanto, a variação espacial dos dados quinzenais de NDVI originados do sensor MODIS é fundamental no cálculo do modelo de estimativa de produtividade.

### **4.3 Modelo de Agregados de Redes Neurais Artificiais - *Ensembles***

#### **4.3.1 Criação do modelo**

Os agregados de redes neurais artificiais ou *ensembles* foram utilizados para obter o modelo que melhor estime a produtividade da cana a partir das variáveis qualitativas e quantitativas relacionadas no item 3.3.1.3.

Conforme descrito na metodologia, os dados foram divididos da seguinte maneira: 80% para a entrada nos *ensembles* e 20% para a validação final. Após o cálculo dos *ensembles* para os dados de entrada (80% dos dados originais) foram obtidos os valores dos EQM dos 30 melhores *ensembles*, como apresenta a Tabela 4.1 e a Tabela 4.2 para o conjunto de dados sem a variável produtividade estimada pela Usina Catanduva.



TABELA 4.1 - Resultado dos 30 melhores *ensembles* e seus respectivos EQM para o conjunto dos dados de entrada.

<i>Ensembles</i>	EQM (ton ha <sup>-1</sup> )
1	14,22
2	16,58
3	14,79
4	16,30
5	13,69
6	14,51
7	14,59
8	15,96
9	13,71
10	16,19
11	13,79
12	15,88
13	14,80
14	13,88
15	15,41
16	16,23
17	14,39
18	14,77
19	15,87
20	14,47
21	13,90
22	16,14
23	16,39
24	15,26
25	16,01
26	15,66
27	15,21
28	16,06
29	14,66
30	14,64

Na Tabela 4.1 são destacados os *ensembles* com menor EQM (5, 9 e 11). A Figura 4.3 apresenta os gráficos dos valores da produtividade estimada pelos *ensembles* 5, 9 e 11

(traço vermelho) e os respectivos valores de produtividade real (traço em azul).

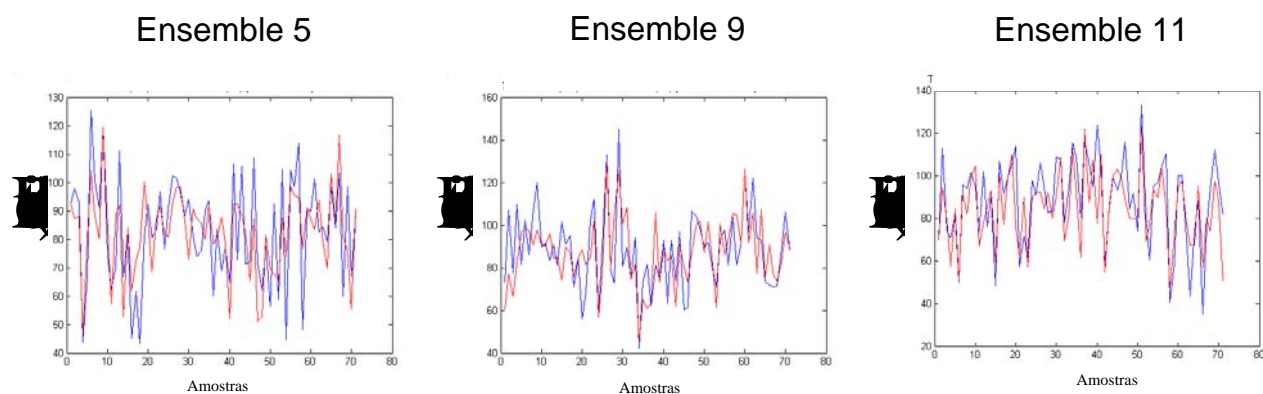


FIGURA 4.3 – Produtividade estimada pelos *ensembles* 5, 9 e 11 (vermelho) e suas respectivas produtividades reais (azul), para as safras de 2004/05 e 2005/06, para o conjunto dos dados de entrada.

TABELA 4.2 - Resultado dos 30 melhores *ensembles* e seus respectivos EQM para o conjunto dos dados de entrada sem a variável produtividade estimada pela Usina Catanduva.

<i>Ensembles</i>	EQM (ton ha <sup>-1</sup> )
1	13,83
2	15,23
3	16,22
4	14,42
5	15,68
6	17,32
7	14,83
8	21,87
9	14,53
10	14,14
11	14,86
12	15,76
13	19,08
14	17,21
15	15,44
16	15,50
17	15,48
18	14,60
19	16,54
20	14,85
21	15,37
22	13,83
23	15,23
24	16,22
25	14,42
26	15,68
27	17,32
28	14,83
29	21,87
30	14,53

São também destacados na Tabela 4.2 os *ensembles* com menor EQM (1, 10 e 22). A

Figura 4.4 apresenta os gráficos dos valores da produtividade estimada pelos *ensembles* 1, 10 e 22 (traço vermelho) e os respectivos valores de produtividade real (traço em azul).

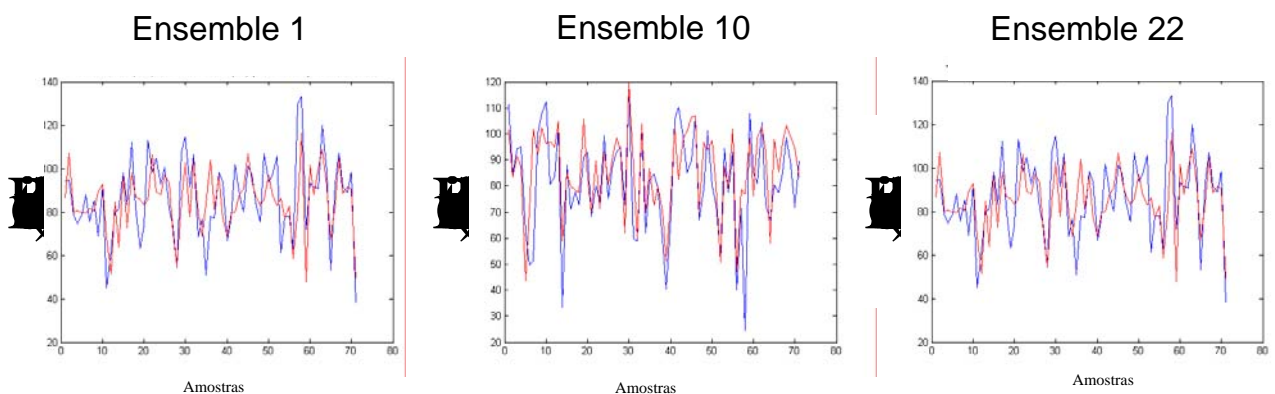


FIGURA 4.4 – Produtividade estimada pelos *ensembles* 1, 10 e 22 (vermelho) e suas respectivas produtividades reais (azul), para as safras de 2004/05 e 2005/06, para o conjunto dos dados de entrada sem a variável produtividade estimada pela Usina Catanduva.

#### 4.3.2 Validação do modelo

Os dados que haviam sido separados do conjunto inicial para validação final, cerca de 20% do conjunto de dados original (90 talhões), foram utilizados para validar os 30 melhores *ensembles*. Depois de validados os *ensembles*, os resultados foram comparados com os dados de produtividade real e obtido o coeficiente de determinação. Assim, pode ser feita uma análise para identificar os *ensembles* que melhor estimaram a produtividade.

##### 4.3.2.1 Conjunto de dados contendo a variável produtividade estimada pela Usina Catanduva

A Tabela 4.3 apresenta os 30 melhores *ensembles* destacando aqueles que apresentaram maior coeficiente de determinação.

TABELA 4.3 – Coeficiente de determinação dos 30 melhores *ensembles* e seus respectivos EQM para o conjunto de dados validados.

<i>Ensemble</i>	$R^2$	EQM
1	0,53	16,09
2	0,45	16,88
3	0,43	18,33
4	0,54	16,15
5	0,58	15,44
6	0,56	15,73
7	0,48	17,14
8	0,45	17,53
9	0,54	16,12
10	0,57	15,45
11	0,60	14,99
12	0,52	16,50
13	0,51	16,60
14	0,66	13,94
15	0,55	15,84
16	0,49	16,84
17	0,56	15,77
18	0,44	18,18
19	0,53	16,18
20	0,52	16,46
21	0,64	14,48
22	0,51	16,73
23	0,59	15,24
24	0,61	14,89
25	0,50	16,68
26	0,63	14,42
27	0,57	15,62
28	0,54	16,15
29	0,48	17,40
30	0,65	14,18

Observou-se que os *ensembles* que possuíam os menores EQM foram os que melhor estimaram a produtividade agrícola da cana-de-açúcar. Estes foram: *ensemble* 14, 21 e 30, Figura 4.5.

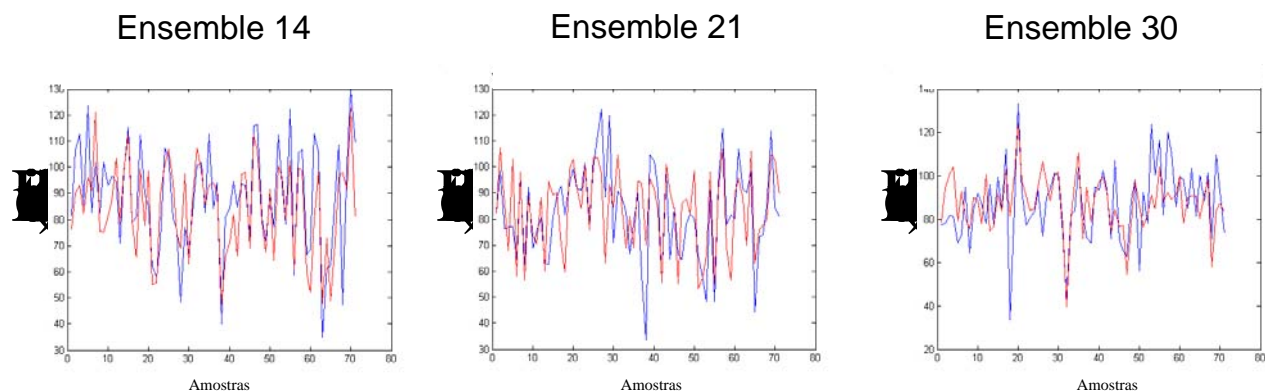


FIGURA 4.5 - Produtividade estimada pelos *ensembles* 14, 21 e 30 e suas respectivas produtividades reais, para as safras de 2004/05 e 2005/06, para o conjunto dos dados de validação.

Utilizando a análise de regressão e o índice de concordância ( $d$ ), foi verificado que o *ensemble* 14 foi o mais satisfatório, pois conseguiu explicar 66% da variação da produtividade real com um EQM 13,94 t/ha. O valor do índice  $d$  foi igual a 0,87, com 95% de confiança, o que indica uma boa concordância entre os dados de produtividade real e a estimativa da produtividade obtida pelo modelo (*ensemble* 14) proposto neste trabalho (Figura 4.6).

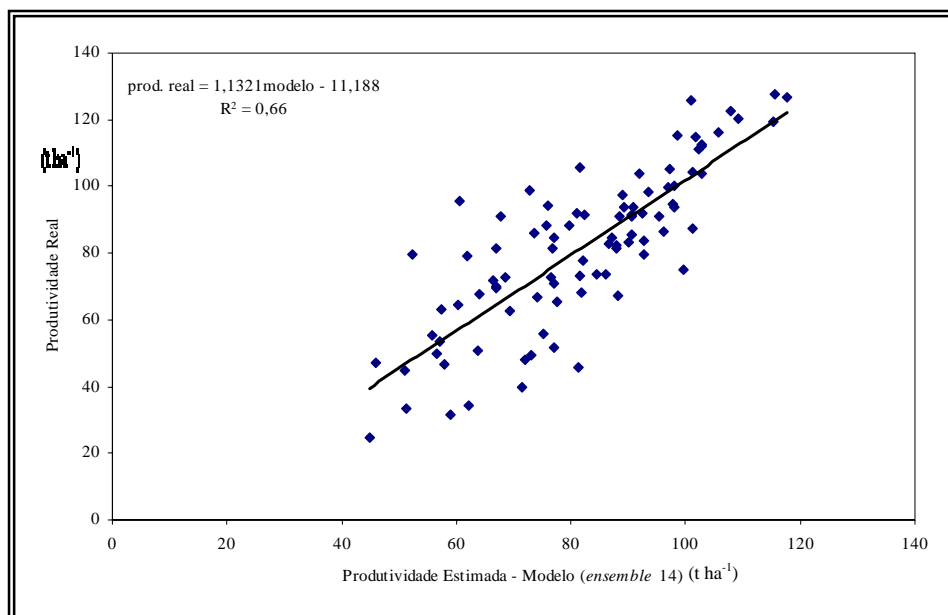


FIGURA 4.6 – Relação entre a produtividade real e a produtividade estimada pelo modelo (*ensemble 14*), safras 2004/05 e 2005/06.

Cabe ressaltar que a estimativa da produtividade feita pelos técnicos da Usina nos dois anos safra explicou 55% da variabilidade da produtividade real, com um EMQ de 14,6 t ha<sup>-1</sup>, enquanto que o modelo aqui proposto explicou 66% desta mesma variabilidade e com um EQM de 13,9 t ha<sup>-1</sup>, ou seja, com um erro de 0,7 toneladas por hectare a menos. Já o valor médio da produtividade por talhão estimado pela Usina foi 80,8 t ha<sup>-1</sup>, o estimado pelo modelo foi 81,9 t ha<sup>-1</sup>, sendo que, a produtividade real média foi 81,5 t ha<sup>-1</sup>, ou seja, a produtividade estimada pelos técnicos da Usina subestimou a média em 0,8% (0,7 t ha<sup>-1</sup>) enquanto que o modelo proposto superestimou em apenas 0,5% (0,4 t ha<sup>-1</sup>).

#### 4.3.2.2 Conjunto de dados que não contém a variável produtividade estimada pela Usina Catanduva

A Tabela 4.4 apresenta os 30 melhores *ensembles* para o conjunto de dados que não possuía a variável produtividade estimada pela Usina Catanduva, destacando aqueles que apresentaram maior coeficiente de determinação.

TABELA 4.4 - Coeficiente de determinação dos 30 melhores *ensembles* e seus respectivos EQM para o conjunto de dados validados sem a variável produtividade estimada pela Usina Catanduva.

<i>Ensemble</i>	R <sup>2</sup>	EQM
1	0,48	17,12
2	0,45	17,86
3	0,46	17,40
4	0,47	17,36
5	0,56	15,98
6	0,48	17,96
7	0,56	15,80
8	0,51	16,62
9	0,57	15,58
10	0,46	17,61
11	0,56	15,88
12	0,44	18,38
13	0,47	17,36
14	0,58	15,57
15	0,53	16,24
16	0,56	15,82
17	0,40	18,59
18	0,49	16,98
19	0,56	15,91
20	0,43	17,93
21	0,54	16,07
22	0,48	17,12
23	0,44	17,86
24	0,46	17,40
25	0,47	17,36
26	0,56	15,98
27	0,48	17,96
28	0,56	15,80
29	0,51	16,62
30	0,57	15,58



Os *ensembles* que melhor estimaram a produtividade agrícola da cana-de-açúcar foram: *ensemble* 9, 14 e 30, Figura 4.7.

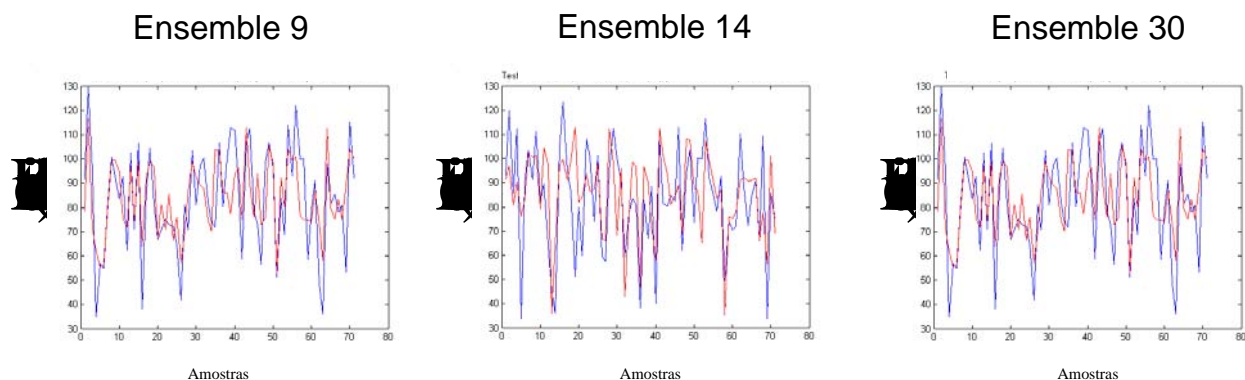


FIGURA 4.7 - Produtividade estimada pelos *ensembles* 9, 14, e 30 e suas respectivas produtividades reais, para as safras de 2004/05 e 2005/06, para o conjunto dos dados de validação sem a variável produtividade estimada pela Usina Catanduva.

Desses três melhores *ensembles* o escolhido foi o 14. O *ensemble* 14 explicou 58% da variação da produtividade real com um EQM  $15,57 \text{ t ha}^{-1}$ , mesmo não utilizando a variável produtividade estimada pela Usina Catanduva (Figura 4.8). O valor do índice  $d$  foi igual a 0,83, com 95% de confiança, o que indica uma boa concordância entre os dados de produtividade real e a estimativa da produtividade obtida pelo modelo (*ensemble* 14) proposto neste trabalho.

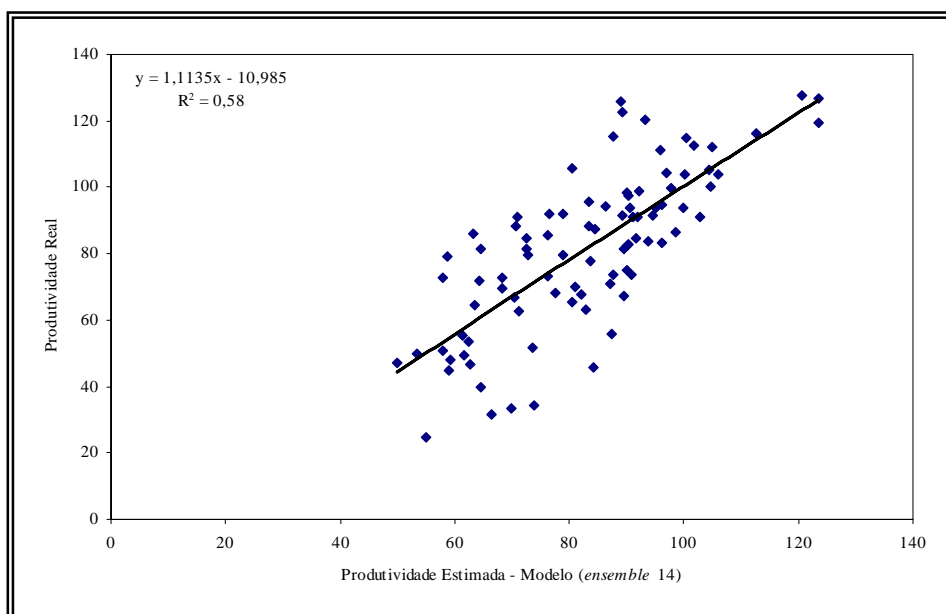


FIGURA 4.8 - Relação entre a produtividade real e a produtividade estimada pelo modelo (*ensemble 14*) sem a variável produtividade estimada pela Usina Catanduva, safras 2004/05 e 2005/06.

Sabe-se que o coeficiente de correlação dos dados de produtividade estimados pela Usina com os dados de produtividade real foi de 0,55, com um EMQ de  $14,6 \text{ t ha}^{-1}$ , enquanto que o coeficiente de variação do modelo estimado, sem utilizar como dado de entrada essa estimativa da Usina, foi de 0,58. Este resultado mostra que a variável produtividade estimada pela Usina Catanduva quando retirada do modelo reduz os custos da estimativa da produtividade e, ainda assim, a estimativa do modelo consegue explicar melhor a variabilidade da produtividade real. Além do modelo agrônômico-espectral não ser um modelo subjetivo, e que por isso proporciona uma análise dos erros.

O valor médio da produtividade estimado pela Usina foi  $80,8 \text{ t ha}^{-1}$  e o estimado pelo modelo foi  $83 \text{ t ha}^{-1}$ , sendo que a produtividade real média foi  $81,5 \text{ t ha}^{-1}$ . Portanto, a produtividade estimada pelos técnicos da Usina Catanduva subestimou a média em 0,8% ( $0,7 \text{ t ha}^{-1}$ ) enquanto que o modelo proposto superestimou 1,2% ( $1,5 \text{ t ha}^{-1}$ ).

## CAPÍTULO 5

### CONCLUSÕES E RECOMENDAÇÕES

O trabalho conseguiu atingir o seu objetivo principal que foi estimar a produtividade agrícola da cana-de-açúcar utilizando variáveis quantitativas e qualitativas em um modelo de agregados de RNA ou *ensembles*. Outras conclusões do trabalho são:

- O modelo agrônômico-espectral foi adaptado com sucesso para a cultura da cana-de-açúcar, e este conseguiu explicar 31% da variabilidade da produtividade real para a safra 2004/05 e 25% para a safra 2005/06;
- O modelo *ensemble* obteve uma estimativa satisfatória dos valores de produtividade agrícola da cana-de-açúcar explicando 66% da variabilidade da produtividade real para as duas safras com EQM de 13,9 t ha<sup>-1</sup>;
- Sem a variável produtividade estimada pela Usina Catanduva o modelo *ensemble* conseguiu explicar 58% da variabilidade da produtividade real para as duas safras com EQM de 15,6 t ha<sup>-1</sup>. O que indica que é possível fazer uma estimativa melhor do que a da Usina Catanduva com modelo que não seja subjetivo.
- A estimativa da produtividade da cana feita pelos técnicos da Usina explica 55% da variabilidade da produtividade real para as duas safras com EQM de 14,6 t ha<sup>-1</sup> indicando que os modelos *ensembles* apresentaram uma melhora na estimativa da produtividade;
- Ao comparar as estimativas de produtividade obtida pelo modelo *ensemble* e pela Usina Catanduva com a produtividade real, concluiu-se que o modelo *ensemble* superestimou a produtividade em 0,4 t ha<sup>-1</sup> (0,5%), já o modelo *ensemble* sem a variável produtividade estimada pela Usina Catanduva superestimou a produtividade em 1,5 t ha<sup>-1</sup> (1,2%), enquanto que a estimativa da

Usina subestimou a produtividade em  $0,7 \text{ t ha}^{-1}$  (0,8%).

Um modelo mais robusto poderá ser obtido por meio da repetição deste trabalho em anos safra futuros para melhor auxiliar os *ensembles* no reconhecimento dos padrões e assim fornecer estimativas mais ajustadas.

Melhorar a resolução espacial da grade de dados pluviométricos, pois a grade utilizada não permite detalhar a variabilidade da precipitação.

Aprimorar a implementação do *software* SPRING em relação ao cálculo da média zonal por grades.

## REFERÊNCIAS BIBLIOGRÁFICAS

Alfonsi, R. R.; Pedro, M. J.; Brunini, O.; Barbieri, V. Condições climática para a cana-de-açúcar. In: Paranhos, S. B (Ed.). **Cana-de-açúcar: cultivo e utilização**. São Paulo: Fundação Cargill. cap.1 v.1, p.42-87. 1987. 856p.

Almeida, T. S.; Fontana, D. C.; Martorano, L.G.; Bergamasch, H. Índices de vegetação para a cultura da soja em diferentes condições hídricas e de sistema de manejo do solo. In: Simpósio Brasileiro de Sensoriamento Remoto, 12., 2005, Goiânia. **Anais...** São José dos Campos: INPE, 2005. p.17-24.

Berka, L.M.S.; Rudorff, B.F.T.; Shimabukuro, Y.E. Soybean yield estimation by an agrometeorological model in a GIS. **Scientia Agricola**, v.60, n.3, p.433-440. 2003.

Black, T. L. The new NMC mesoscale Eta model: Description and forecast examples. **Weather and Forecasting**, v. 9, n. 2, p. 265-278. 1994.

Braga, H. J. Previsão agrícola: uma nova abordagem - **Uso de scanner aerotransportável e redes neurais**. 1995. 243 p. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 1995. Disponível na biblioteca digital de teses e dissertações da UFSC:  
<<http://www.eps.ufsc.br/teses/braga/index> >. Acesso em jan/2005.

Breiman, L. Bagging predictors. **Machine Learning**, v. 24, n. 2, p. 123-140, 1996.

Carvalho, A. P. L. F.; Braga, A. P.; Ludermir, T. B. **Fundamentos de redes neurais artificiais**. In: Escola de Computação, 11., 1998. Rio de Janeiro: UFRJ. 1998. 246 p.

Carvalho, A. P. L. F. **Redes neurais artificiais**. Disponível em:  
<<http://www.icmc.usp.br/~andre/research/neural/MLP.htm#back>>. Acesso em: out/2005.

Castro, P. R.C. Maturadores químicos em cana-de-açúcar. In: Semana da cana-de-açúcar de Piracicaba, 4., 1999. Piracicaba. **Anais...** Piracicaba: ESALQ, 1999. p. 12-16.

Centro de Tecnologia Copersucar – CTC. **Censo varietal Copersucar 2003**. Piracicaba: Centro de Tecnologia Copersucar, 2004. 16 p.

Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (CEPAGRI). **Clima dos municípios paulistas**. Disponível em:  
<http://www.cpa.unicamp.br/portal/index.php>. Acesso em: ago/2006.

Choudhury, B.J.; Ahmed N.U.; Idso S.B.; Reginato R.J.; Daughtry C.S.T. Relations between evaporation coefficients and vegetation indices studied by model simulations. **Remote Sensing of Environment**, v.50, p.1-17, 1994.

Cooperativa dos produtores de cana, açúcar e álcool do Estado de São Paulo (Copersucar). **Recomendação de adubação para a cultura da cana-de-açúcar**. Piracicaba – SP: Centro de Tecnologia Copersucar, 1988. 7 p.

- Draper, N. R.; Smith, H. **Applied regression analysis**. New York: John Wiley & Sons. 1966. 407 p.
- Delgado-Rojas, J. S.; Barbieri, V. Modelo agrometeorológico de estimativa da produtividade da cana-de-açúcar. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 7, n. 1, p. 67-73, 1999.
- Doorenbos, J.; Kassam, A. H. **Yield response to water**. Rome: Food and Agriculture Organization of the United Nations, 1979. 193 p. (FAO-Irrigation and Drainage Paper n. 33).
- Efron, E.; Tibshirani, R. **An introduction to the Bootstrap**. New York: 1993. Chapman & Hall.
- Figueira, S. R. **Os programas de álcool como combustíveis nos EUA, no Japão e na União Européia e as possibilidades de exportação do Brasil**. 2005. 245 p. Tese (Doutorado em Ciências) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2005. Disponível na biblioteca digital de teses e dissertações da USP: <<http://www.teses.usp.br/>>. Acesso em mar/2006.
- Food and Agricultural Organizations of the United Nations – FAO. **Food and Agricultural commodities production**. Disponível em: <http://www.fao.org/es/ess/top/commodity.html;jsessionid=051B3E79C2FD549C004A92353A80DEEC?lang=en&item=156&year=2004>. Acesso em: mar/2006.
- Formaggio, A. R.; Epiphanio, J. C. N. Estudos Radiométricos das culturas de trigo e de feijão em três tipos de solos. In: Simpósio Brasileiro de Sensoriamento Remoto, 5., 1988. **Anais...** Natal: INPE, 1988.
- Fortes, C. **Discriminação varietal e estimativa de produtividade agroindustrial de cana-de-açúcar pelo sensor orbital ETM+/LANDSAT 7**. 2003. 131 p. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2003. Disponível na biblioteca digital de teses e dissertações da USP: <<http://www.teses.usp.br/>>. Acesso em: 19 dez/2004.
- Freund, Y. Boosting a weak algorithm by majority. **Information and Computation**, v. 121, n. 2, p. 256-286, 1995.
- Freund, Y.; Shapire, R. Experiments with a new boosting algorithm. In: International Conference on Machine Learning, 13., 1996. **Proceedings...** Morgan Kaufmann, 1996. p. 149-156.
- Gates, D. M.; Keegan, H. J.; Schleter, J. C.; Weidner, V. R. Spectral properties of plants. **Applied Optics**, v. 4, p. 11-20, 1965.
- Gutman, G.; Tarpley, D.; Ignatov, A. The enhanced NOAA global land dataset from the Advanced Very High Resolution Radiometer. **Bulletin of the American Meteorological Society**, v. 76, n. 7, p. 1141-1156, July 1995. Disponível em: [http://ams.allenpress.com/pdfserv/10.1175%2F1520-0477\(1995\)076%3C1141:TENGLD%3E2.0.CO%3B2](http://ams.allenpress.com/pdfserv/10.1175%2F1520-0477(1995)076%3C1141:TENGLD%3E2.0.CO%3B2). Acesso em ago/2005.

Hartkamp, A.D.; White, J.W.; Hoogenboom, G. Interfacing geographic information systems with agronomic modellings: A review. **Agronomy Journal**, v. 91, n. 5, p. 761-722, 1999.

Hansen, L. K.; Salamon, P. Neural network ensembles. **IEEE Transactions Pattern Analysis and Machine Intelligence**, v. 12, n. 10, p. 993-1001, 1990.

Haykin, S. **Neural networks** – a comprehensive foundation. Ed. Tom Robbins, 2. New Jersey: Prentice Hall, 1999. p. 842.

Hoogenboom, G. Contribution of agrometeorology to the simulation of crop production and its applications. **Agricultural and Forest Meteorology**, v. 103, n. 1-2, p. 137-157, Jun. 2000.

Humbert, R. P. **El cultivo de la caña de azucar**. México, D. F.: Companhia Editorial Continental, 1974. 719p.

Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT). **Mapa geomorfológico do Estado de São Paulo**. Volumes I e II. São Paulo. 1981.

Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT). **Mapa geológico do Estado de São Paulo**. Volumes I e II. São Paulo. 1981a.

Jensen, M.E. Water consumption by agricultural plants. In: Kozlowski, T. T.(Ed.) **Water deficits and plants growth**. New York: Academic Press, 1968. v. 2, p. 1-22.

Joaquim, A. C. **Identificação de variedades de cana-de-açúcar em três classes texturais de solos, na região de Araraquara-SP, através de análise de nível de cinza em imagens LANDSAT/TM**. 1998. 118 p. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual de Campinas, Campinas, 1998.

Kosko, B. **Neural Networks and Fuzzy Systems: a dynamical Systems approach to machine intelligence**. California: Prentice-Hall, 1992.

Lima, C. A. M. **Comitê de Máquinas: uma abordagem unificada empregando máquina de vetores-suporte**. 2004. 314 p. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, Campinas, 2004.

Lucchesi, A. A. Processos fisiológicos da cultura da cana-de-açúcar (*Saccharum spp*). **Boletim Técnico ESALQ/CENA**, n.7, p.1-50, 1995.

Magalhães, A. C. N. Ecofisiologia da cana-de-açúcar: aspectos do metabolismo do carbono da planta. In: Castro, P. R. C.; Ferreira, S. O.; Yamada, T. (Coord.) **Ecofisiologia da Produção**. Piracicaba: Associação Brasileira para pesquisa do Potássio e do Fósforo, 1987. p. 113-118.

Mangelsdorf, A. J. **Um programa de melhoramento da cana-de-açúcar para a agroindústria canavieira do Brasil**. Rio de Janeiro: Instituto do Açúcar e do álcool. 1966. 63 p.

- Marchiori, L. F. S. **Influência da época de plantio e corte na produtividade da cana-de-açúcar.** 2004. 275 p. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2004. Disponível na biblioteca digital de teses e dissertações da USP: <<http://www.teses.usp.br/>>. Acesso em dez/2004.
- Matteo, K. C. **Sistemas de informação geográfica para monitoramento da cultura da cana-de-açúcar.** 1998. 107 p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1998.
- Maximiliano, S. S. **Modelos para a previsão da produtividade da cana-de-açúcar (*Saccharum spp.*) através de parâmetros climáticos.** 2002. 79 p. Dissertação (Mestrado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002. Disponível na biblioteca digital de teses e dissertações da USP: <<http://www.teses.usp.br/>>. Acesso em dez/2004.
- MODIS – Moderate Resolution Imaging Spectroradiometer. **About MODIS.** Disponível em: <http://modis.gsfc.nasa.gov/about>. Acesso em: jun/ 2004.
- Norman, J. M.; Anderson, M. C.; Kustas, W. P.; French, A. N.; Mecikalski, J.; Torn, R.; Diak, G. R.; Achmugge, T. J. **Remote Sensing of Evapotranspiration for Precision-Farming Applications.** In: International Geoscience & Remote Sensing Symposium. Toulouse, 2003. 308p.
- Novo, E.M. de M. **Sensoriamento remoto: princípios e aplicações.** São Paulo: Edgard Blucher. 1989. 308p.
- Nunes Jr, D. M. S. Variedades de cana-de-açúcar. In: Paranhos, S. B. **Cana-de-açúcar: cultivo e utilização.** São Paulo: Fundação Cargill, 1987. 856p. cap.2 v.1, p.187-259.
- Opitz D. W.; Shavlik J. W. Connectionist theory refinement: genetically searching the space of network topologies. **Journal of Artificial Intelligence Research**, v. 6, p. 177-209, 1997. Disponível em: <http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume6/opitz97a.pdf>. Acesso em out/2005.
- Ponzoni, F. J. Comportamento Espectral da Vegetação. In: Paulo Roberto Meneses; José da Silva Madeira Netto. (Org.). **Sensoriamento remoto: reflectância de alvos naturais.** 1 ed. Brasília: Editora UNB, 2002, v. 1, p. 157-199.
- Ricaud, R. B.; Cochran, M. Methods of planting sugarcane for sugar and biomass production in Louisiana. In: Congresso of International Society of sugar cane technologists, 1980, Louisiana. **Proceedings...** Louisiana: ISSCT, 1980. p. 118-120.
- Rizzi, R. **Geotecnologias em um sistema de estimativa da produção de soja: estudo de caso no Rio Grande do Sul.** 2004. 204 p. Tese (Doutorado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais, São José dos campos. 2004.
- Rudorff, B. F. T. **Dados Landsat na estimativa da produtividade agrícola da cana-de-açúcar.** 1985. 114p. (INPE-3744-TDL/202). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1985.



Rudorff, B.F.T.; Batista, G.T. Resposta Espectral da Cana-de-açúcar Através de Dados Multiespectrais do MSS do Landsat-4. In: Reunião da Sociedade de Especialistas Latinoamericanos em Sensoriamento Remoto – Selper, 4, 1984. Chile. **Anais...** São José dos Campos: INPE, 1985. 12p. Publicado como: (INPE-3474-PRE/717).

Rudorff, B. F. T.; Batista, G. T. Yield estimation of sugar-cane based on agrometeorological - Spectral models. **Remote Sensing of Environment**, v. 33, n. 3, p. 183-192, Set/1990.

Rudorff, B. F. T.; Batista, G. T. Wheat yield estimation at the farm level using Landsat-TM and agrometeorological data. **International Journal Of Remote Sensing**, v. 12, n. 12, p. 2477-2484, 1991.

Rudorff, B. F. T., Shimabukuro, Y. E., Batista, G. T., Lee, D. The contribution of qualitative variables to a sugarcane yield model based on spectral vegetation index. In: Simposio Latinoamericano de Percepción Remota, 7., 1995. Puerto Vallarta. Memorias del Simposio Latinoamericano de Percepción Remota, 7., SELPER e SIE, 1995. **Anais...** v.1. p.705 – 708.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagation errors. **Nature**, v. 323, p. 533–536, 1986.

Schapire, R. E. The strength of weak learn ability. **Machine Learning**, v. 5, n. 2, p. 197-227, 1990.

Schmidt, E.J.; Gers, C.; Narciso, G.; Frost, P. Remote sensing in the south African Sugar Industry. In International Society of Sugarcane Technologists Congress, 2001. Brisbane. **Proceedings...** Brisbane, v. 2, p. 241-246. Disponível em: <http://www.sasa.org.za/sasri/ageng/pub/CG/ISSCT2001.pdf>. Acesso em junho/2005.

Silva, L. N. C. **Análise e síntese de estratégias de aprendizado para redes neurais artificiais**. 1998. 210 p. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e Computação. Universidade Estadual de Campinas, Campinas, 1998.

Sollich, P.; Krogh, A. Learning with ensembles: how overfitting can be useful. In: Touretzky, D. S.; Mozer, M. C.; Hasselmo, M. E. (Eds.) **Advances in Neural Information Processing Systems 8**. Cambridge: MIT Press, p. 190-196, 1996.

Stupiello, J. P. A cana-de-açúcar como matéria prima. In: Paranhos, S. B. **Cana-de-açúcar: cultivo e utilização**. São Paulo: Fundação Cargill, 1987. 856p. cap.7 v.2, p.761-804.

Sugawara, L. M. **Avaliação de Modelo Agrometeorológico e Imagens NOAA/AVHRR no Acompanhamento e Estimativa de Produtividade da Soja no Estado do Paraná**. 2002. 181 p. (INPE-8702-TDI/794). Dissertação (Mestrado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2002.

Swain, P.H.; Davis, S. M. **Remote sensing: the quantitative approach**. New York: McGraw-Hill, 1978. 396 p.

Thiago, L. R. L. S.; Vieira, J. M. **Cana-de-açúcar: uma alternativa de alimento para a seca.** EMBRAPA Gado de Corte - Comunicado Técnico nº 73, 2002. Disponível em: <http://www.cnpqg.embrapa.br/publicacoes/cot/COT73.html>. Acesso em jan/2005.

União da Agroindústria Canavieira de São Paulo – UNICA. **A alta competitividade canavieira do Estado de São Paulo.** Disponível em: [http://www.unica.com.br/pages/agroindustria\\_alta.asp](http://www.unica.com.br/pages/agroindustria_alta.asp). Acesso em out/2004.

União da Agroindústria Canavieira de São Paulo – UNICA. **Cana-de-açúcar: produtos.** Disponível em: [http://www.unica.com.br/pages/cana\\_produtos.asp](http://www.unica.com.br/pages/cana_produtos.asp). Acesso em nov/2004a.

União da Agroindústria Canavieira de São Paulo – UNICA. **Geração descentralizada de emprego e renda.** Disponível em: [http://www.unica.com.br/pages/sociedade\\_mercado1.asp](http://www.unica.com.br/pages/sociedade_mercado1.asp). Acesso em fev/2006.

Valeriano, M.M. Reflectância espectral de culturas agrícolas anuais: espectrorradiometria. **Espaço e Geografia**, v.6, n.1, p.7-27, 2003.

Von Zuben, F. J. **Modelos paramétricos e não-paramétricos de redes neurais artificiais e aplicações.** 1996. 244 p. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, Campinas, 1996.

Von Zuben, F. J.; Lima, C. A. M.; Villanueva, W. J. P. **Programa computacional – Ensemble\_Bagging.** Laboratório de Bioinformática e Computação Bio-Inspirada – LbiC. Campinas: Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, 2005. Programa não publicado.

Watson, D. J. Comparative physiological studies on growth of field crops: I. Variation in net assimilation rate and leaf area between species and varieties, and within and between years. **Annals of Botany**. London, v. 11, n. 44, p. 41-76, 1947.

Weber, L. S. **Integração de dados espectrais e indicadores meteorológicos por meio de redes neurais para a estimativa de produtividade de cana-de-açúcar.** 2005. 106 p. Tese (Doutorado em Engenharia Agrícola) – Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, 2005.

Willmott, C. J.; Ackelson, S. G.; Davis, R. E. Statistics for the evaluation and comparison of models. **Journal of Geophysical Research**. v.90, n.C5, p. 8995-9005, 1985.

Xavier, A. C. **Estimativa da dinâmica do índice de área foliar em uma microbacia hidrográfica por meio de técnicas de sensoriamento remoto.** 2002. 111 p. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2002.

Xavier, A. C.; Vettorazzi, C. A. Leaf area index of ground covers in a subtropical watershed. **Scientia Agricola**, v. 60, n. 3, p. 425-431, 2003.

## APÊNDICE A

### Transformação dos valores de NDVI em valores de IAF (dados de 2004).

{

Numerico ndvi1, ndvi2, ndvi3, ndvi4, ndvi5, ndvi6, ndvi7, ndvi8, ndvi9, ndvi10, ndvi11, ndvi12, ndvi13, ndvi14, ndvi15, ndvi16, ndvi17, ndvi18, ndvi19, ndvi20, ndvi21, ndvi22, ndvi23 ("MNT\_NDVI");  
Numerico IAF1, IAF2, IAF3, IAF4, IAF5, IAF6, IAF7, IAF8, IAF9, IAF10, IAF11, IAF12, IAF13, IAF14, IAF15, IAF16, IAF17, IAF18, IAF19, IAF20, IAF21, IAF22, IAF23 ("MNT\_IAF");

NDVIMAX = 9950;

NDVIMIN = 0;

N = 0.9;

Ndvi1 = Recuperar (Nome="2004\_001");

IAF1 = Novo (Nome= "IAF\_2004\_001", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF1 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi1) / (NDVIMAX - NDVIMIN))^N)))$ ;

Ndvi2 = Recuperar (Nome="2004\_017");

IAF2 = Novo (Nome= "IAF\_2004\_017", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF2 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi2) / (NDVIMAX - NDVIMIN))^N)))$ ;

Ndvi3 = Recuperar (Nome="2004\_033");

IAF3 = Novo (Nome= "IAF\_2004\_033", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF3 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi3) / (NDVIMAX - NDVIMIN))^N)))$ ;

Ndvi4 = Recuperar (Nome="2004\_049");

IAF4 = Novo (Nome= "IAF\_2004\_049", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF4 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi4) / (NDVIMAX - NDVIMIN))^N)))$ ;

Ndvi5 = Recuperar (Nome="2004\_065");

IAF5 = Novo (Nome= "IAF\_2004\_065", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF5 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi5) / (NDVIMAX - NDVIMIN))^N)))$ ;

Ndvi6 = Recuperar (Nome="2004\_081");

IAF6 = Novo (Nome= "IAF\_2004\_081", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF6 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi6) / (NDVIMAX - NDVIMIN))^N)))$ ;

Ndvi7 = Recuperar (Nome="2004\_097");

IAF7 = Novo (Nome= "IAF\_2004\_097", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF7 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi7) / (NDVIMAX - NDVIMIN))^N)))$ ;

Ndvi8 = Recuperar (Nome="2004\_113");

IAF8 = Novo (Nome= "IAF\_2004\_113", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF8 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi8) / (NDVIMAX - NDVIMIN))^N)))$ ;

ndvi9 = Recuperar (Nome="2004\_129");

IAF9 = Novo (Nome= "IAF\_2004\_129", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

IAF9 =  $-2 * \log(1 - (1 - (((NDVIMAX - ndvi9) / (NDVIMAX - NDVIMIN))^N)))$ ;

Ndvi10 = Recuperar (Nome="2004\_145");

IAF10 = Novo (Nome= "IAF\_2004\_145", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);

```

IAF10 = -2*log(1-(1-((( NDVIMAX - ndvi10 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi11 = Recuperare (Nome="2004_161");
IAF11 = Novo (Nome= "IAF_2004_161", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF11 = -2*log(1-(1-((( NDVIMAX - ndvi11 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi12 = Recuperare (Nome="2004_177");
IAF12 = Novo (Nome= "IAF_2004_177", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF12 = -2*log(1-(1-((( NDVIMAX - ndvi12 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi13 = Recuperare (Nome="2004_193");
IAF13 = Novo (Nome= "IAF_2004_193", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF13 = -2*log(1-(1-((( NDVIMAX - ndvi13 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi14 = Recuperare (Nome="2004_209");
IAF14 = Novo (Nome= "IAF_2004_209", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF14 = -2*log(1-(1-((( NDVIMAX - ndvi14 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi15 = Recuperare (Nome="2004_225");
IAF15 = Novo (Nome= "IAF_2004_225", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF15 = -2*log(1-(1-((( NDVIMAX - ndvi15 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi16 = Recuperare (Nome="2004_241");
IAF16 = Novo (Nome= "IAF_2004_241", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF16 = -2*log(1-(1-((( NDVIMAX - ndvi16 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi17 = Recuperare (Nome="2004_257");
IAF17 = Novo (Nome= "IAF_2004_257", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF17 = -2*log(1-(1-((( NDVIMAX - ndvi17 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi18 = Recuperare (Nome="2004_273");
IAF18 = Novo (Nome= "IAF_2004_273", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF18 = -2*log(1-(1-((( NDVIMAX - ndvi18 )/( NDVIMAX - NDVIMIN ))^N)));

ndvi19 = Recuperare (Nome="2004_289");
IAF19 = Novo (Nome= "IAF_2004_289", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF19 = -2*log(1-(1-((( NDVIMAX - ndvi19 )/( NDVIMAX - NDVIMIN ))^N)));

Ndvi20 = Recuperare (Nome="2004_305");
IAF20 = Novo (Nome= "IAF_2004_305", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF20 = -2*log(1-(1-((( NDVIMAX - ndvi20 )/( NDVIMAX - NDVIMIN ))^N)));

Ndvi21 = Recuperare (Nome="2004_321");
IAF21 = Novo (Nome= "IAF_2004_321", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF21 = -2*log(1-(1-((( NDVIMAX - ndvi21 )/( NDVIMAX - NDVIMIN ))^N)));

Ndvi22 = Recuperare (Nome="2004_337");
IAF22 = Novo (Nome= "IAF_2004_337", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF22 = -2*log(1-(1-((( NDVIMAX - ndvi22 )/( NDVIMAX - NDVIMIN ))^N)));

Ndvi23 = Recuperare (Nome="2004_353");
IAF23 = Novo (Nome= "IAF_2004_353", ResX= 30, ResY= 30, Escala=60000, Min=0, Max=255);
IAF23 = -2*log(1-(1-((( NDVIMAX - ndvi23 )/( NDVIMAX - NDVIMIN ))^N)));

}

```

## **PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE**

### **Teses e Dissertações (TDI)**

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

### **Manuais Técnicos (MAN)**

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

### **Notas Técnico-Científicas (NTC)**

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programa de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

### **Relatórios de Pesquisa (RPQ)**

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

### **Propostas e Relatórios de Projetos (PRP)**

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

### **Publicações Didáticas (PUD)**

Incluem apostilas, notas de aula e manuais didáticos.

### **Publicações Seriadas**

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

### **Programas de Computador (PDC)**

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. São aceitos tanto programas fonte quanto executáveis.

### **Pré-publicações (PRE)**

Todos os artigos publicados em periódicos, anais e como capítulos de livros.