

**UNIVERSIDADE FEDERAL DE VIÇOSA**

**Gustavo Felipe Ferreira Vieira**

**Fenotipagem em cana-de-açúcar por espectroscopia no infravermelho próximo  
(NIR) – Modelagem estatística visando à predição**

**VIÇOSA – MINAS GERAIS**

**2016**

**Gustavo Felipe Ferreira Vieira**

**Fenotipagem em cana-de-açúcar por espectroscopia no infravermelho próximo  
(NIR) – Modelagem estatística visando á predição**

**Trabalho de conclusão de curso  
apresentado à Universidade Federal de  
Viçosa como parte das exigências para a  
obtenção do título de Engenheiro  
Agrônomo. Modalidade: trabalho  
científico.**

**Orientador: Luiz Alexandre Peternelli**

**Coorientadores:**

**Márcio Henrique Pereira Barbosa**

**Felipe Lopes da Silva**

**VIÇOSA – MINAS GERAIS**

**2016**

**Gustavo Felipe Ferreira Vieira**

**Fenotipagem em cana-de-açúcar por espectroscopia no infravermelho próximo  
(NIR) – Modelagem estatística visando à predição**

**Trabalho de conclusão de curso  
apresentado à Universidade Federal de  
Viçosa como parte das exigências para a  
obtenção do título de Engenheiro  
Agrônomo. Modalidade: trabalho  
científico.**

APROVADO: \_\_\_\_/\_\_\_\_/\_\_\_\_

---

Prof. Luiz Alexandre Peternelli

(orientador)

(UFV)

## RESUMO

O objetivo desse trabalho foi identificar alternativas para a seleção precoce de genótipos de cana-de-açúcar, para os caracteres teor de fibra (FIBRA) e polper cento cana (PC), utilizando-se dos espectros de reflectância no infravermelho próximo (NIR–NearInfraredReflectance) em diferentes tipos de amostras. Foram utilizados 160 clones de cana-de-açúcar contrastantes em relação aos caracteres FIBRA e PC. Os espectros NIR foram obtidos ao longo do tempo na folha+3, em três partes do colmo, sendo elas ápice, meio e base do colmo e no bagaço (cana moída) de todos os clones. O modelo de calibração foi gerado pelo método dos mínimos quadrados parciais (PLS) por meio do software R. A avaliação da capacidade preditiva do modelo foi realizada por meio da correlação entre os valores preditos e os valores reais. As amostras de colmo apresentaram resultados superiores a folha+3 para os dois caracteres e entre as partes avaliadas o meio do colmo foi a amostra com os maiores valores de coeficiente de determinação ( $R^2$ ) na predição de FIBRA, apresentando coeficiente de correlação ( $r$ ) de 0,31 na população teste. Para PC as análises da base do colmo apresentaram a menor raiz do erro quadrático médio (RMSE) e o maior  $R^2$ . As análises do bagaço proporcionaram os melhores resultados para FIBRA e PC, com correlação entre os valores reais e preditos na população teste de 0,37 e 0,80, respectivamente, demonstrando potencial para uso na predição simultânea de PC e FIBRA, pois apresentou os menores valores de RMSE, associado aos maiores valores de  $R^2$  na validação cruzada e aos maiores coeficientes de correlação entre os valores reais e preditos na população teste. Novas análises devem ser realizadas em um maior número de amostras de bagaço de cana-de-açúcar de forma a aumentar a representatividade das populações e melhorar a capacidade preditiva dos modelos de calibração.

Palavras-chave: NIR, Cana-de-açúcar, Fibra, PC.

## ABSTRACT

The aim of this study was to identify alternatives for the early selection of sugarcane genotypes for characters fiber (FIBER) and polpercent cane (PC), using near infrared reflectance (NIR) data in different types of samples. A total of 160 clones were used sugarcane contrasting relative to the fiber and character PC. NIR spectra were recorded over time in sheet +3, in three parts of the stalk (apex, mid and base), and bagasse (crushed sugar cane) of all clones. We considered partial least squares (PLS) from software R to fit the calibration models, and the correlation between the predicted and the observed (real) values as a measure to evaluate the predictive ability of each model. The stalk samples showed better results than +3 sheet for both characters, fiber and PC. Among the stalks segments, the middle part of the stalk was the kind of sample with the largest coefficient of determination values ( $R^2$ ) for fiber, showing a correlation coefficient ( $r$ ) of 0.31 in the test population. For PC, the samples taken at the base of the stalks showed the smallest values for the root mean squared error (RMSE) and the largest  $R^2$ . The bagasse analysis provided the best results for FIBER and PC, with correlation between real and predicted values in the test population of 0.37 and 0.80, respectively. Bagasse samples demonstrates potential for use in the simultaneous prediction of PC and FIBER, as it presented lower values of RMSE, associated to higher  $R^2$  values in the cross-validation analyzes, and the highest correlation coefficients between real and predicted values in the test population. One should conduct further analysis in a larger number of bagasse samples of sugarcane in order to increase the representativeness of the population and to improve the predictive ability of the calibration models.

Keywords: NIR, Sugarcane, Fiber, PC.

## CONTEÚDO

1. INTRODUÇÃO.....	7
2. OBJETIVOS.....	9
2.1. Objetivo geral .....	9
2.2. Objetivos específicos .....	9
3. MATERIAL E MÉTODOS.....	9
3.1. Material Vegetal .....	9
3.3. Obtenção dos espectros.....	10
3.4. Análise Estatística.....	10
4.RESULTADOS E DISCUSSÕES.....	10
5. CONCLUSÃO.....	14
6. REFERÊNCIAS BIBLIOGRÁFICAS .....	14

## 1. INTRODUÇÃO

A produção da cana-de-açúcar desempenha papel fundamental na economia do país, sendo grande aliada na busca pela balança comercial favorável. Sua importância tem sido cada vez maior, principalmente com o interesse na obtenção de fontes alternativas de energia renovável.

Segundo o relatório do quarto levantamento da safra 2015/2016 da Companhia Nacional de Abastecimento (CONAB), a lavoura de cana-de-açúcar está em expansão no Brasil, porém houve redução de plantio. A área plantada foi de aproximadamente 8.654,2 mil hectares na temporada 2015/2016, o que representa 3,9% de redução em relação à 2014/2015. Os Estados de São Paulo, Goiás, Minas Gerais e Mato Grosso do Sul foram os responsáveis pelo plantio em hectare de 4.498, 885, 866 e 596, respectivamente (CONAB, 2016).

Assim, é notável que existe uma ampla distribuição de lavouras pelo país e que novas áreas estão sendo incorporadas pela cultura. A produtividade média brasileira na safra 2015/16 foi de 75 t.ha<sup>-1</sup> (CONAB, 2016). Segundo Waclawovsky *et al.* (2010), a média de produtividade mundial é de 80 t.ha<sup>-1</sup>. Portanto, a produtividade média dos cultivares brasileiro está aquém do que se pode atingir com o potencial genético da cultura, o que enfatiza a necessidade do melhoramento genético para se alcançar tal objetivo, mesmo sendo o Brasil o maior produtor de cana (FAO, 2016). Para que a produção de interesse seja alcançada sem que ocorram grandes incorporações de novas áreas, clones cada vez mais produtivos e adaptados a essas novas regiões devem ser indicados pelos programas de melhoramento em menor espaço de tempo.

No entanto, o lançamento de novos cultivares tem ocorrido entre 11 e 13 anos (BARBOSA e SILVEIRA, 2010). Ademais, o procedimento de fenotipagem requerido no processo de melhoramento é extremamente oneroso devido ao grande número de experimentos e de mão-de-obra necessária para a avaliação de clones promissores. Além disso, a avaliação em laboratório dos chamados componentes tecnológica (PC e FIBRA descritos a seguir) envolve um grande número de mão-de-obra, tempo e recursos financeiros, o que muitas vezes limita a inclusão de potenciais genótipos para avaliação nas etapas do programa de melhoramento genético.

As características fenotípicas relacionadas aos teores de açúcares da cana-de-açúcar são as de maior importância para as indústrias canavieiras e tem sido foco dos programas de melhoramento genético. O principal parâmetro utilizado é a pureza do

caldo extraído, que pode ser obtido por meio do açúcar polarizável, denominado de polpercento cana (PC), que é uma medida indireta do teor de sacarose da cana-de-açúcar. O PC é obtido por métodos polarimétricos via polarímetros ou sacarímetros. A mensuração desse caracter demanda muita mão-de-obra, tempo e reagentes químicos de alto custo.

Nos últimos anos os programas de melhoramento genético da cana-de-açúcar, deram maior ênfase ao desenvolvimento de cultivares com maior rendimento de sacarose para a produção de açúcar e álcool (DAL-BIANCO et al., 2012). A demanda atual por maior quantidade de biomassa para produção de etanol de segunda geração e de energia elétrica faz com que novas linhas de pesquisa sejam iniciadas visando o desenvolvimento de cultivares de cana-de-açúcar com maiores teores de fibra (FIBRA), denominadas cana energia (LOUREIRO et al., 2011).

Nesse sentido métodos alternativos para medição dessas características tecnológicas vem sendo investigados. Dentre eles, a reflectância no infravermelho próximo (NIR–Near Infrared Reflectance) associado a métodos estatísticos multivariados (como o método dos mínimos quadrados parciais – PLS) apresenta grande possibilidade de aplicação devido a sua facilidade de utilização, rapidez, exatidão e não geração de resíduos químicos (VALDERRAMA et al., 2007; MORGANO et al., 2005). Além disso, após aquisição do aparelho que faz a leitura do NIR, não há nenhum gasto com químicos e outros materiais para se executar as leituras e posteriores mensurações.

Valderrama et al. (2007) construíram um modelo multivariado para predições por meio de NIR para POL e obtiveram correlação de 0,99 entre os valores reais e os preditos, demonstrando que o método pode ser muito útil na avaliação dessas características, sendo que 1 é o valor máximo na correlação, quanto mais próximo deste melhor é a correlação. Por outro lado, até o momento, nada relevante foi publicado relacionando a predição do teor de fibra via NIR, abrindo grande campo para pesquisa na área.

Portanto, visando facilitar a seleção de indivíduos com maior produtividade e qualidade da biomassa de maneira mais rápida e precoce possível, torna-se importante desenvolver procedimentos estatísticos envolvendo época de amostragem, número e tipo de amostras, e modelos preditivos associados à tecnologia NIR.



## **2. OBJETIVOS**

### **2.1. Objetivo geral**

O principal objetivo desse trabalho foi identificar alternativas para a seleção precoce de genótipos de cana-de-açúcar, para os caracteres teor de fibra (FIBRA) e polper cento cana (PC), utilizando-se dos espectros de reflectância no infravermelho próximo (NIR–NearInfraredReflectance) em diferentes tipos de amostras.

### **2.2. Objetivos específicos**

- Construir modelos estatísticos para determinação dos teores de fibra (FIBRA) e do polper cento cana (PC) em cana-de-açúcar;
- Avaliar qual parte do colmo (superior, médio, inferior) proporciona melhor leitura e consequente á mensuração dos componentes tecnológicos FIBRA e PC;
- Praticar o uso do NIR e sua modelagem com dados coletados ao longo do tempo na folha +3, avaliando, assim, a possibilidade de seleção precoce de genótipos para os caracteres FIBRA e PC;
- Avaliar o bagaço (cana moída) na predição de FIBRA e PC.

## **3. MATERIAL E MÉTODOS**

### **3.1. Material Vegetal**

Neste estudo foram utilizados 160 clones de cana-de-açúcar contrastantes em relação aos caracteres de FIBRA e PC. Os clones avaliados são oriundos de cruzamentos envolvendo genitores com alto teor de sacarose e genitores com elevado teor de fibra, visando á obtenção de cultivares de cana energia com maior teor de fibra e com os teores de sacarose das cultivares atuais, após a seleção na primeira fase de teste (T1), esses clones foram conduzidos para a segunda fase de teste (T2).

O T2 foi instalado em julho de 2014 no delineamento em blocos aumentados (DBA) com duas cultivares como testemunhas (RB867515 e C90-176) e com 18 repetições (blocos). Cada uma das 196 parcelas foi constituída por dois sulcos de 5m de comprimento, espaçados em 1,4m. A instalação do experimento foi realizado no Centro de Pesquisa e Melhoramento Genético da Cana-de-Açúcar (CECA) pertencente á

Universidade Federal de Viçosa, no município de Oratórios, MG, com latitude 20°25' S; longitude 42°48' W; altitude 494m; solo LVE.

### **3.2. Avaliação Fenotípica**

Os caracteres PC e FIBRA foram obtidos por meio da análise tecnológica realizada em uma amostra de 500g obtida com a moagem de 10 canas por parcela aos 10 meses após o plantio (FERNANDES, 2003).

### **3.3. Obtenção dos espectros**

As leituras via NIR das amostras de folha +3 de cada clone foram realizadas mensalmente, de dezembro de 2014 a abril de 2015, coletando em campo duas amostras desta, uma no centro e outra na ponta da folha. As leituras das amostras de colmo foram realizadas em três partes (base, meio e ápice do colmo) de 1 colmo de cada um dos 160 clones, no mês de maio de 2015, cortando cada parte longitudinalmente para ser colocado no aparelho. Aproximadamente 3g de bagaço (cana moída) das amostras usadas nas análises tecnológicas foram usadas nas leituras via NIR, em maio de 2015, os colmos foram moídos através de um engenho presente na área experimental.

### **3.4. Análise Estatística**

O modelo de calibração foi gerado pelo método dos mínimos quadrados parciais (PLS), uma técnica que reduz os preditores para um conjunto menor de componentes não correlacionados e executa regressão de mínimos quadrados sobre estes componentes, em vez de fazer isso sobre os dados originais, por meio do *software* R. A avaliação da capacidade preditiva do modelo foi realizada por meio da correlação entre os valores preditos e os valores reais de FIBRA e PC. Para obtenção do modelo de calibração foram usados os dados padronizados de 140 clones na validação cruzada. Para avaliar a capacidade de predição foi usada uma população teste composta por 20 clones.

## **4. RESULTADOS E DISCUSSÃO**

Conforme pode ser observado nas tabelas 1, 2 e 3, o melhor tipo de amostra na predição do teor de fibra (FIBRA) e do polper cento cana (PC) foi o bagaço, pois proporcionou os maiores coeficientes de determinação ( $R^2$ ), as maiores correlações na

população teste (r) e os menores valores da raiz do erro quadrático médio (RMSE) na validação cruzada.

O número de variáveis latentes (VL) para cada modelo de predição variou entre 5 e 20 para os caracteres FIBRA e PC nas amostras de folha+3. Nas amostras de colmo o número de variáveis latentes variou de 8 a 18 e nas amostras de bagaço foi igual a 14 para os dois caracteres avaliados. Foi definido o número de variáveis latentes que proporcionou o menor valor de RMSE na validação cruzada para cada modelo de calibração (Tabelas 1, 2 e 3).

**TABELA 1** – Número de variáveis latentes (VL), raiz do erro quadrático médio na validação cruzada (RMSE), coeficiente de determinação na validação cruzada (R<sup>2</sup>) e coeficiente de correlação entre os valores reais e os valores preditos na população teste (r) para os caracteres teor de fibra (FIBRA) e polpercento cana (PC) em 5 meses de leitura via infravermelho próximo (NIR) na folha +3 de cana-de-açúcar.

CARACTER	MÊS	VL	RMSE	R <sup>2</sup>	r
FIBRA	dezembro	5	1,943	0,039	-0,075
	janeiro	18	1,813	0,158	-0,092
	fevereiro	16	1,683	0,270	0,146
	março	17	1,906	0,057	-0,149
	abril	20	1,917	0,056	0,036
PC	dezembro	7	2,516	0,090	0,269
	janeiro	18	2,391	0,150	0,548
	fevereiro	10	2,501	0,105	0,489
	março	16	2,427	0,084	0,442
	abril	11	2,453	0,127	0,052

**TABELA 2** – Número de variáveis latentes (VL), raiz do erro quadrático médio na validação cruzada (RMSE), coeficiente de determinação na validação cruzada (R<sup>2</sup>) e coeficiente de correlação (r) entre os valores reais e os valores preditos na população teste para os caracteres teor de fibra (FIBRA) e polpercento cana (PC) em 3 partes do colmo (ápice, meio e base) de cana-de-açúcar usadas na leitura do infra vermelho próximo (NIR).

CARACTER	AMOSTRA	VL	RMSE	R <sup>2</sup>	r
FIBRA	ápice	10	1,879	0,120	0,125
	meio	8	1,669	0,306	0,310

	base	18	1,673	0,245	0,142
	ápice	8	2,115	0,370	0,352
PC	meio	8	1,957	0,452	0,710
	base	11	1,673	0,587	0,563

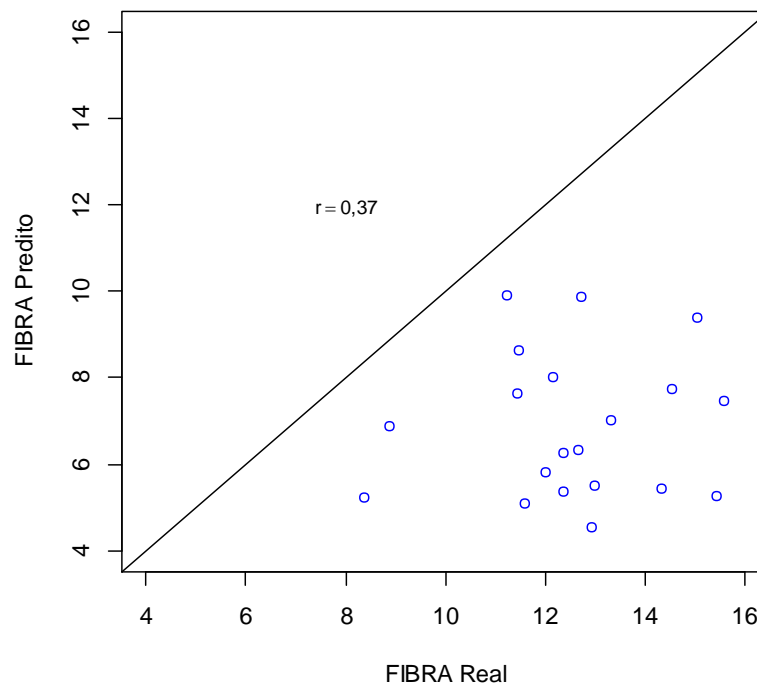
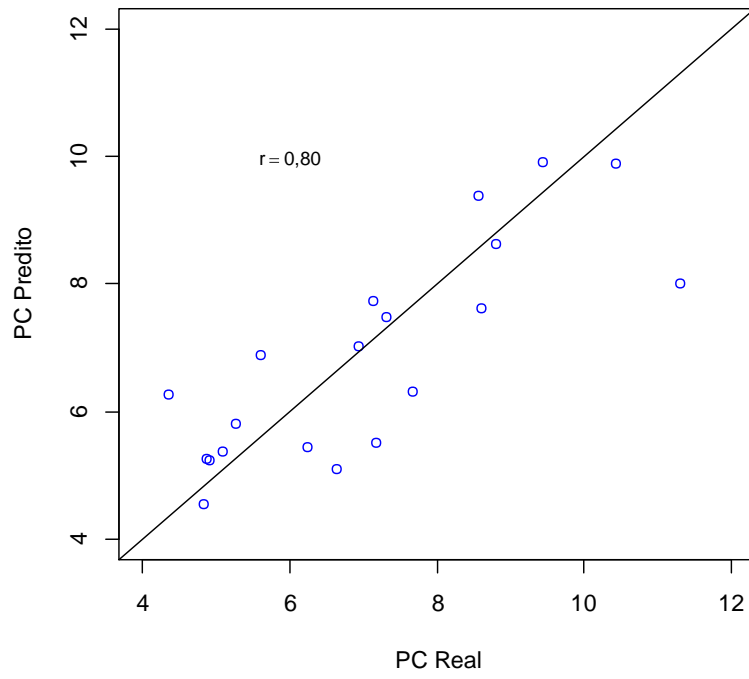
**TABELA 3** –Número de variáveis latentes (VL), raiz do erro quadrático médio na validação cruzada (RMSE), coeficiente de determinação na validação cruzada (R<sup>2</sup>) e coeficiente de correlação (r) entre os valores reais e os valores preditos na população teste para os caracteres teor de fibra (FIBRA) e polper cento cana (PC) no bagaço de cana-de-açúcar usado na leitura do infravermelho próximo (NIR).

CARACTER	VL	RMSE	R <sup>2</sup>	r
FIBRA	14	1,612	0,321	0,371
PC	14	1,649	0,599	0,805

Nas análises realizadas na folha +3, os coeficientes de determinação foram baixos para todos os meses de avaliação e para os dois caracteres, demonstrando que esse tipo de amostra apresenta baixo potencial na predição de FIBRA e PC, portanto a predição e seleção de genótipos de cana-de-açúcar para esses caracteres não pode ser realizada de forma eficiente.

As amostras de colmo apresentaram resultados superiores á folha+ 3 para os dois caracteres e entre as partes avaliadas o meio do colmo foi o tipo de amostra com os maiores valores de R<sup>2</sup> na predição do teor de fibra, apresentando correlação de 0,31 na população teste. Embora o meio do colmo tenha apresentado a maior correlação (0,71) na população teste para o caractere polper cento cana, nas análises da base do colmo foi obtido o menor erro quadrático médio e o maior coeficiente de determinação (Tabela 2).

As análises do bagaço proporcionaram os melhores resultados para FIBRA e PC, com correlação entre os valores reais e preditos na população teste de 0,37 e 0,80, respectivamente (Figura 1), demonstrando que o bagaço apresenta o maior potencial entre os tipos de amostra avaliada nesse trabalho. Portanto, maiores esforços com esse tipo de amostra devem ser concentrados na obtenção de um modelo acurado de predição de FIBRA e PC em cana-de-açúcar. Para tanto, um maior número de genótipos deve ser avaliado visando aumentar a representatividade da variabilidade genética presente nas populações de melhoramento genético da cultura.



**FIGURA 1** – Correlação entre os valores reais e os valores preditos na população teste (20 clones) a partir dos modelos de predição para os caracteres polper cento cana (PC) e teor de fibra (FIBRA) em amostras de bagaço de cana-de-açúcar.

## 5. CONCLUSÃO

Entre as amostras avaliadas, o bagaço apresentou os melhores resultados, demonstrando elevado potencial para uso na predição simultânea de PC e FIBRA, pois apresentou os menores valores de RMSE, associado aos maiores valores de  $R^2$  e de correlação ( $r$ ) entre os valores reais e preditos na população teste.

Novas análises devem ser realizadas em um maior número de amostras de bagaço de cana-de-açúcar de forma a aumentar a representatividade das populações e melhorar a capacidade preditiva dos modelos de calibração.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- BARBOSA, M.H.P.; SILVEIRA, L.C.I. (2010) Melhoramento Genético e Recomendação de Cultivares. In: Santos, F.; Borém, A. e Caldas, C. Editores. **Cana-de-açúcar: Bioenergia, Açúcar e Álcool - Tecnologias e Perspectivas**. Viçosa, MG – Suprema, 578p.
- CONAB. (2016). Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de cana-de-açúcar. **Safra 2015/2016**. QuartolevantamentoAbr/2016Brasília: Conab, 2016. Disponível: em [http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16\\_04\\_14\\_09\\_06\\_31\\_boletim\\_cana\\_portugues\\_-\\_4o\\_lev\\_-\\_15-16.pdf](http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_04_14_09_06_31_boletim_cana_portugues_-_4o_lev_-_15-16.pdf) Acesso em: 15 de agosto 2016.
- DAL-BIANCO, M.; CARNEIRO, M.S.; HOTTA, C.T.; CHAPOLA, R.G.; HOFFMANN, H.P.; GARCIA, A.A.F.; SOUZA, G.M (2012) Sugarcane improvement: how far can we go?. **CurrentOpinion in Biotechnology** **23**: 265-270.
- FERNANDES, A. C. **Cálculos na agroindústria da cana-de-açúcar**. Piracicaba: STAB, 2003. 240p.
- GOLDSCHIMID, O. Ultravioletspectra. In: SARKANEN, K.V.; LUDWIG, C. H. Lignins: occurrence, formation, structureand reactions. New York: J. Wiley, 1971. p. 241-266.
- LOUREIRO, M.E.; BARBOSA, M.H.P.; LOPES, F.J.F.; SILVÉRIO, F.O. Sugarcane Breeding and Selection for more Efficient Biomass Conversion in Cellulosic Ethanol. In: BUCKERIDGE, M.S.; GOLDMAN, G.H. **Routes to Cellulosic Ethanol**. New York, Springer, 2011, cap. 13, p. 199-239.

- MORGANO, M. A.; FARIA, C. G.; FERRÃO, M. F.; BRAGAGNOLO, N.; FERREIRA, M. M. C. (2005). Determinação de proteína em café cru por espectroscopia NIR e regressão PLS. **Ciênc.Tecnol.Aliment.**, 25(1): 25-31.
- R Core Team (2014). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- TEW, T. L.; COBILL, R. M. Genetic Improvement of Sugarcane (*Saccharum*spp.) as an Energy Crop. *In: VERMERRIS, W. Genetic Improvement of Bioenergy Crops*. Springer, New York, 2008. cap. 9, p. 273-294.
- VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. (2007). Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy. **J. Braz. Chem. Soc.**, 18(2):259-266.
- WACLAWOVSKY, A. J.; SATO, P.M.; LEMBKE, C.G.; MOORE, P.H.; SOUZA, G.M. (2010). Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. **PlantBiotechnologyJournal**, 8(3):263-276.