

**JUSSARA VALENTE ROQUE**

**DESENVOLVIMENTO DE MODELOS DE REGRESSÃO MULTIVARIADA  
PARA DETERMINAÇÃO DE ÉSTERES DE FORBOL EM SEMENTES  
DE JATROPHA CURCAS L. USANDO ESPECTROSCOPIA E QUIMIOMETRIA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Agroquímica, para obtenção do título de Magister Scientiae.

**VIÇOSA  
MINAS GERAIS – BRASIL  
2015**

**Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal  
de Viçosa - Campus Viçosa**

T

R786d  
2015

Roque, Jussara Valente, 1988-

Desenvolvimento de modelos de regressão multivariada para  
determinação de ésteres de forbol em sementes de *Jatropha curcas* L.  
usando espectroscopia e quimiometria / Jussara Valente Roque. -  
Viçosa, MG, 2015.

xv, 67f. : il. (algumas color.) ; 29 cm.

Inclui anexo.

Orientador: Reinaldo Francisco Teófilo.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f.57-66.

1. Pinhão-mansô. 2. Éster de forbol. 3. Espectroscopia.  
4. Quimiometria. 5. Calibração - Análise multivariada. I. Universidade  
Federal de Viçosa. Departamento de Química. Programa de pós-  
graduação em Agroquímica. II. Título.

CDD 22. ed. 543

**JUSSARA VALENTE ROQUE**

**DESENVOLVIMENTO DE MODELOS DE REGRESSÃO MULTIVARIADA  
PARA DETERMINAÇÃO DE ÉSTERES DE FORBOL EM SEMENTES  
DE JATROPHA CURCAS L. USANDO ESPECTROSCOPIA E QUIMIOMETRIA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Agroquímica, para obtenção do título de Magister Scientiae.

APROVADA: 17 de julho de 2015.

---

Luiz Antônio dos Santos Dias  
(Coorientador)

---

Antônio Jacinto Demuner  
(Coorientador)

---

Gilmare Antônia da Silva

---

Reinaldo Francisco Teófilo  
(Orientador)

Dedico esta conquista  
À minha mãe Sônia, ao meu pai João e  
Aos meus irmãos Marco e Tatiany.

**A distância entre o sonho e a conquista chama-se atitude!**  
**(Autor desconhecido)**

## **AGRADECIMENTOS**

Agradeço aos meus pais e meus irmãos pelo amor, orações, apoio e paciência.

Ao Prof. Reinaldo F. Teófilo pela orientação, confiança, amizade, paciência e por estar sempre disposto a ensinar!

Aos Prof. Luiz Antônio S. Dias pela confiança e disponibilização de recursos para realização deste trabalho.

Ao Prof. Antônio J. Demuner pela valiosa contribuição na realização deste trabalho.

Ao Laboratório de Melhoramento de Oleaginosas e o Laboratório de Biotecnologia e Melhoramento Vegetal pela disponibilização dos equipamentos para realização deste trabalho.

Aos amigos do LINQ, em especial à Adriana, Gabriela, Larissa, Malu e Mariana, pela amizade, convívio e por estarem sempre dispostas a ajudar.

À Prof.<sup>a</sup> Gilmare A. da Silva, que gentilmente aceitou o convite para compor a banca examinadora.

À Universidade Federal de Viçosa, em especial o Departamento de Química, pela estrutura para poder realizar este trabalho.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela concessão da bolsa de estudo.

A todos aqueles que contribuíram direta ou indiretamente para a realização deste trabalho.

## **BIOGRAFIA**

JUSSARA VALENTE ROQUE, filha de João Batista Roque e Sônia Maris Valente Roque, nasceu na cidade de Viçosa, estado de Minas Gerais, em 26 de setembro de 1988.

Iniciou o curso de Farmácia em agosto de 2008 pela Universidade Federal de Ouro Preto (UFOP), em Ouro Preto, MG, diplomando-se como farmacêutica em julho de 2013. Em agosto do mesmo ano, iniciou o curso de pós-graduação em Agroquímica, com área de concentração em Química Analítica, em nível de Mestrado, na Universidade Federal de Viçosa, submetendo-se à defesa de dissertação em 17 de julho de 2015.

## SUMÁRIO

|                                                                  |      |
|------------------------------------------------------------------|------|
| LISTA DE ABREVIACÕES .....                                       | VIII |
| LISTA DE FIGURAS.....                                            | IX   |
| LISTA DE TABELAS .....                                           | XI   |
| LISTA DE EQUAÇÕES .....                                          | XII  |
| RESUMO.....                                                      | XIV  |
| ABSTRACT .....                                                   | XV   |
| 1 INTRODUÇÃO .....                                               | 1    |
| 2 REVISÃO DA LITERATURA .....                                    | 4    |
| 2.1 <i>Jatropha curcas</i> L. ....                               | 4    |
| 2.1.1 Ésteres de forbol .....                                    | 5    |
| 2.2 Espectroscopia .....                                         | 8    |
| 2.2.1 Infravermelho Próximo (NIR).....                           | 8    |
| 2.2.2 Ultravioleta Visível .....                                 | 10   |
| 2.3 Cromatografia Líquida de Alta Eficiência .....               | 11   |
| 2.4 Quimiometria.....                                            | 12   |
| 2.4.1 Planejamentos Experimentais .....                          | 12   |
| 2.4.2 Calibração Multivariada .....                              | 14   |
| 2.4.3 Quadrados Mínimos Parciais .....                           | 15   |
| 2.4.4 Seleção de Variáveis.....                                  | 18   |
| 2.4.5 Figuras de Mérito em Calibração Multivariada .....         | 20   |
| 3 MATERIAIS E MÉTODOS.....                                       | 23   |
| 3.1 Coleta e Identificação de Amostras .....                     | 23   |
| 3.2 Preparo das Amostras .....                                   | 25   |
| 3.2.1 Extração do Óleo das Sementes.....                         | 26   |
| 3.2.2 Extração dos Ésteres de Forbol do Óleo .....               | 26   |
| 3.3 Análise Espectral das Amostras.....                          | 28   |
| 3.4 Obtenção das Variáveis Dependentes (Ésteres de Forbol )..... | 29   |
| 3.4.1 Separação Cromatográfica .....                             | 29   |
| 3.4.2 Planejamento Experimental .....                            | 29   |
| 3.4.3 Curva Analítica .....                                      | 30   |
| 3.4.4 Quantificação e Identificação.....                         | 31   |
| 3.4.5 Validação .....                                            | 31   |
| 3.5 Construção dos Modelos de Calibração Multivariada .....      | 32   |
| 5 RESULTADOS E DISCUSSÃO .....                                   | 35   |
| 5.1 Planejamento Experimental.....                               | 35   |
| 5.2 Quantificação e Identificação dos Ésteres de Forbol .....    | 36   |
| 5.3 Modelos de Calibração .....                                  | 40   |

|       |                                      |    |
|-------|--------------------------------------|----|
| 5.3.1 | Espectros da Semente Sem Casca.....  | 41 |
| 5.3.2 | Espectros do Óleo .....              | 42 |
| 5.3.3 | Espectros da Semente Com Casca ..... | 44 |
| 5.3.4 | Espectros do Extrato do Óleo .....   | 48 |
| 6     | CONCLUSÃO .....                      | 55 |
| 7     | TRABALHOS FUTUROS.....               | 56 |
| 8     | REFERÊNCIAS .....                    | 57 |
| 9     | ANEXOS .....                         | 67 |

## LISTA DE ABREVIações

| Abreviaturas  | Termos em inglês                           | Termos em português                                |
|---------------|--------------------------------------------|----------------------------------------------------|
| <b>ACN</b>    | Acetonitrile                               | Acetonitrila                                       |
| <b>ANOVA</b>  | Variance Analysis                          | Análise de Variância                               |
| <b>GAB</b>    | Germplasm Active Bank                      | Banco Ativo de Germoplasma                         |
| <b>CCD</b>    | Composite Central Design                   | Planejamento Composto Central                      |
| <b>CM</b>     | Mean Center                                | Centrar na Média                                   |
| <b>CV</b>     | Cross Validation                           | Validação Cruzada                                  |
| <b>DAD</b>    | Diodo Array Detector                       | Detector de Arranjo de Diodos                      |
| <b>DOE</b>    | Design of Experiments                      | Planejamento de Experimentos                       |
| <b>GA</b>     | Genetic Algorithm                          | Algoritmo Genético                                 |
| <b>h</b>      | Number of Variable Latents                 | Número de Variáveis Latentes                       |
| <b>HCOOH</b>  | Formic Acid                                | Ácido Fórmico                                      |
| <b>HOMO</b>   | Highest Occupied Molecular Orbitals        | Orbital Molecular Ocupado Mais Alto                |
| <b>HPLC</b>   | High Performance Liquid Chromatography     | Cromatografia Líquida de Alta Eficiência           |
| <b>LOD</b>    | Limit of Detection                         | Limite de Detecção                                 |
| <b>LOQ</b>    | Limit of Quantification                    | Limite de Quantificação                            |
| <b>LUMO</b>   | Lowest Unoccupied Molecular Orbitals       | Orbital Molecular Não Ocupado Mais Baixo           |
| <b>MeOH</b>   | Methanol                                   | Metanol                                            |
| <b>MSC</b>    | Multiplicative Signal Correction           | Correção do Espalhamento Multiplicativo            |
| <b>NAS</b>    | Net Analyte Signal                         | Sinal Analítico Líquido                            |
| <b>NIR</b>    | Near Infrared                              | Infravermelho Próximo                              |
| <b>OPS</b>    | Ordered Predictors Selection               | Seleção dos Preditores Ordenados                   |
| <b>PCA</b>    | Principal Component Analysis               | Análise de Componentes Principais                  |
| <b>PE</b>     | Phorbol Ester                              | Éster de Forbol                                    |
| <b>PLS</b>    | Partial Least Squares                      | Quadrados Mínimos Parciais                         |
| <b>PMA</b>    | Phorbol-12-Myristate-13-Acetate            | Forbol-12-Miristato-13-Acetato                     |
| <b>R</b>      | Correlation Coefficient                    | Coefficiente de Correlação                         |
| <b>RMSE</b>   | Root Mean Square Error                     | Raiz Quadrática Média do Erro                      |
| <b>RMSECV</b> | Root Mean Square Error of Cross Validation | Raiz Quadrática Média do Erro de Validação Cruzada |
| <b>RMSEP</b>  | Root Mean Square Error of Prediction       | Raiz Quadrática Média do Erro de Predição          |
| <b>RPD</b>    | Ratio Performance Deviation                | Razão de Desempenho do Desvio                      |
| <b>RSD</b>    | Relative Standard Deviation                | Desvio Padrão Relativo                             |
| <b>RSM</b>    | Response Surface Methodology               | Metodologia de Superfície de Resposta              |
| <b>SDV</b>    | Standard Deviation of Validation Errors    | Desvio Padrão dos Erros de Validação               |
| <b>SEL</b>    | Selectivity                                | Seletividade                                       |
| <b>SEN</b>    | Sensibility                                | Sensibilidade                                      |
| <b>PT</b>     | Progeny Test                               | Teste de Progenie                                  |
| <b>UV/Vis</b> | Ultraviolet/Visible                        | Ultravioleta/Visível                               |

## LISTA DE FIGURAS

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Figura 1.</b> Estrutura base (12-desoxi-16-hidroxiforbol) e os seis ésteres de forbol presentes em <i>Jatropha curcas</i> L. ....                                                                                                                                                                                                                                                                                                                                   | <b>6</b>  |
| <b>Figura 2.</b> Espectro eletromagnético. ....                                                                                                                                                                                                                                                                                                                                                                                                                        | <b>8</b>  |
| <b>Figura 3.</b> Bandas e posições relativas de picos de absorção em infravermelho próximo. ....                                                                                                                                                                                                                                                                                                                                                                       | <b>9</b>  |
| <b>Figura 4.</b> Construção da matriz <b>X</b> para calibração multivariada.....                                                                                                                                                                                                                                                                                                                                                                                       | <b>14</b> |
| <b>Figura 5.</b> Método de seleção de variáveis usando a seleção dos preditores ordenados (OPS).....                                                                                                                                                                                                                                                                                                                                                                   | <b>19</b> |
| <b>Figura 6.</b> Semente de pinhão manso com a casca ( <b>A</b> ), semente sem a casca ( <b>B</b> ), semente cortada ao meio em seu eixo longitudinal ( <b>C</b> ) e semente com a superfície lisa e plana ( <b>D</b> ). ....                                                                                                                                                                                                                                          | <b>26</b> |
| <b>Figura 7.</b> Esquema do procedimento de extração dos PEs (♦) em óleo de <i>J. curcas</i> .....                                                                                                                                                                                                                                                                                                                                                                     | <b>27</b> |
| <b>Figura 8.</b> Esquema com a matriz <b>X</b> para cada conjunto de espectros e o vetor <b>y</b> com os teores de PEs. ....                                                                                                                                                                                                                                                                                                                                           | <b>33</b> |
| <b>Figura 9.</b> Gráfico de Pareto das influências dos fatores do HPLC sobre as respostas. Erro usado: Soma quadrática dos resíduos. Nível de significância, $\alpha$ : 0,05. ....                                                                                                                                                                                                                                                                                     | <b>35</b> |
| <b>Figura 10.</b> Superfície de resposta com a % ACN e % HCOOH em função da RD. ....                                                                                                                                                                                                                                                                                                                                                                                   | <b>36</b> |
| <b>Figura 11.</b> Curva analítica de PMA, para determinação de ésteres de forbol em <i>Jatropha curcas</i> . ....                                                                                                                                                                                                                                                                                                                                                      | <b>37</b> |
| <b>Figura 12.</b> Concentração dos ésteres de forbol em $\text{mg g}^{-1}$ na semente de <i>J. curcas</i> .....                                                                                                                                                                                                                                                                                                                                                        | <b>38</b> |
| <b>Figura 13.</b> Cromatograma obtido de uma amostra de pinhão-manso com os cinco picos referentes aos ésteres de forbol ( <b>A</b> ). Espectros dos picos cromatográficos dos fatores de <i>Jatropha curcas</i> ( <b>B</b> ): o primeiro pico corresponde ao fator C1; o segundo pico corresponde ao fator C2; o terceiro pico corresponde ao fator C3; o quarto pico corresponde ao fator C6 e o quinto pico corresponde aos fatores C4 e C5, que são epímeros. .... | <b>40</b> |
| <b>Figura 14.</b> ( <b>A</b> ) Espectros NIR das sementes de <i>Jatropha curcas</i> L. sem casca ( <b>X</b> <sub>NIRsc</sub> ) e ( <b>B</b> ) valores medidos e preditos dos teores de PEs para o conjunto de calibração do modelo PLS-OPS. ....                                                                                                                                                                                                                       | <b>41</b> |
| <b>Figura 15.</b> ( <b>A</b> ) Espectros UV-Vis-NIR das sementes de <i>Jatropha curcas</i> L. sem casca ( <b>X</b> <sub>UVsc</sub> ) e ( <b>B</b> ) valores medidos e preditos do teor de PEs para o conjunto de calibração do modelo PLS-OPS. ....                                                                                                                                                                                                                    | <b>42</b> |

|                                                                                                                                                                                                                                                      |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 16. (A)</b> Espectros NIR do óleo de <i>Jatropha curcas</i> L. ( $X_{NIRo}$ ) e <b>(B)</b> valores medidos e preditos do teor de PEs para o conjunto de calibração (●) e de predição (◆) para o modelo PLS-OPS. ....                       | 43 |
| <b>Figura 17. (A)</b> Espectros UV do óleo de <i>Jatropha curcas</i> L. ( $X_{UVo}$ ) e <b>(B)</b> valores medidos e preditos do teor de PEs para o conjunto de calibração (●) e de predição (◆) para o modelo PLS-OPS. ....                         | 43 |
| <b>Figura 18.</b> Espectros NIR das sementes de <i>Jatropha curcas</i> L. com casca ( $X_{NIRcc}$ ).....                                                                                                                                             | 44 |
| <b>Figura 19.</b> Variáveis selecionadas pelo OPS. ....                                                                                                                                                                                              | 46 |
| <b>Figura 20. (A)</b> Erros relativos (%) do conjunto de calibração e <b>(B)</b> conjunto de predição. ....                                                                                                                                          | 46 |
| <b>Figura 21.</b> Valores medidos versus preditos dos teores de PEs para o conjunto de calibração (●) e de predição (◆). ....                                                                                                                        | 47 |
| <b>Figura 22. (A)</b> Correlação por chance, <b>(B)</b> correlações e <b>(C)</b> valores de RMSECV obtidos para o teste de estabilidade do modelo.....                                                                                               | 48 |
| <b>Figura 23. (A)</b> Espectros UV dos extratos do óleo de <i>Jatropha curcas</i> L. ( $X_{UVe}$ ) e <b>(B)</b> espectro com a primeira derivada. ....                                                                                               | 49 |
| <b>Figura 24.</b> Variáveis selecionadas pelo algoritmo <b>(A)</b> OPS e <b>(B)</b> pelo GA.....                                                                                                                                                     | 50 |
| <b>Figura 25. (A)</b> Erros relativos para o conjunto de calibração, <b>(B)</b> erros relativos do conjunto de predição e <b>(C)</b> valores medidos versus preditos para o conjunto de calibração (●) e de predição (◆) para o modelo PLS-OPS. .... | 52 |
| <b>Figura 26. (A)</b> Correlação por chance, <b>(B)</b> correlações e <b>(C)</b> valores de RMSECV obtidos para o teste de estabilidade do modelo.....                                                                                               | 53 |

## LISTA DE TABELAS

|                                                                                                                                                                                                                     |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Tabela 1.</b> Propriedades físico-químicas do óleo de pinhão-manso. ....                                                                                                                                         | 5  |
| <b>Tabela 2.</b> Teores de ésteres de forbol encontrados na literatura. ....                                                                                                                                        | 7  |
| <b>Tabela 3.</b> Identificação dos acessos pertencentes ao Banco Ativo de Germoplasma (GAB). ....                                                                                                                   | 24 |
| <b>Tabela 4.</b> Identificação das progênies pertencentes ao teste de progênie (PT). ....                                                                                                                           | 25 |
| <b>Tabela 5.</b> Identificação das variáveis independentes (matriz <b>X</b> ) obtidas para cada tipo de material das amostras de <i>J. curcas</i> , faixa analisada e tipo de análise no NIR, UV e UV-VIS-NIR. .... | 28 |
| <b>Tabela 6.</b> Planejamento composto central para otimização da composição da fase móvel na separação cromatográfica: porcentagem de acetonitrila (% ACN) e porcentagem de ácido fórmico (% HCOOH). ....          | 30 |
| <b>Tabela 7.</b> Parâmetros analíticos de validação do método cromatográfico quantitativo. ....                                                                                                                     | 37 |
| <b>Tabela 8.</b> Parâmetros estatísticos dos modelos com todas as variáveis e com a seleção de variáveis pelos métodos OPS e GA. ....                                                                               | 45 |
| <b>Tabela 9.</b> Parâmetros estatísticos dos modelos com todas as variáveis e com a seleção de variáveis pelos métodos OPS e GA. ....                                                                               | 50 |
| <b>Tabela 10.</b> Figuras de mérito para o modelo PLS-OPS com o extrato do óleo. ....                                                                                                                               | 51 |
| <b>Tabela 11.</b> Concentração dos ésteres de forbol presentes nos acessos do Banco Ativo de Germoplasma (GAB). ....                                                                                                | 67 |
| <b>Tabela 12.</b> Concentrações dos ésteres de forbol presentes nas progênies do teste de progênie (PT). ....                                                                                                       | 67 |

## LISTA DE EQUAÇÕES

|                                                                                                   |    |
|---------------------------------------------------------------------------------------------------|----|
| $A = \varepsilon.c.b$ (1) .....                                                                   | 10 |
| $y \rightarrow X = URV^t$ (2) .....                                                               | 16 |
| $y = Xb \rightarrow \hat{X} = U_h R_h V_h^t \rightarrow \hat{b} = V_h R_h^{-1} U_h^t y$ (3) ..... | 16 |
| $RMSECV = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{I_{cv}}}$ (4) .....                               | 17 |
| $RPD = \frac{D_p}{RMSECV}$ (5) .....                                                              | 17 |
| $viés = \frac{\sum_{i=1}^{nv} (y_i^{ref} - \hat{y}_i)}{I_m}$ (6) .....                            | 17 |
| $SDV = \sqrt{\frac{\sum [(y_i^{ref} - \hat{y}_i) - viés]^2}{nv - 1}}$ (7) .....                   | 18 |
| $t_{calc} = \frac{ viés  \sqrt{nv}}{SDV}$ (8) .....                                               | 18 |
| $SEN = \frac{1}{\ b\ }$ (9) .....                                                                 | 21 |
| $\gamma = \frac{SEN}{\ \partial_x\ }$ (10) .....                                                  | 21 |
| $SEL = \frac{nas_{k,i}}{\ x_{k,i}\ }$ (11) .....                                                  | 21 |
| $LOD = 3 \cdot \ \delta r\  \cdot \ b\ $ (12) .....                                               | 22 |
| $LOQ = 10 \ \delta r\  \cdot \ b\ $ (13) .....                                                    | 22 |
| $LOD = \frac{\Delta(\alpha, \beta) w_{y_0} \sigma}{\hat{b}}$ (14) .....                           | 22 |
| $RD = \frac{trF}{np}$ (15) .....                                                                  | 30 |
| $C_f = \frac{C_i \cdot FD \cdot TO}{d}$ (16) .....                                                | 31 |
| $RSD = \left( \frac{s}{\bar{x}} \right) \times 100$ (17) .....                                    | 31 |

$$\text{LOD} = 3. \left( \frac{s_b}{b} \right) \quad (18) \dots\dots\dots 32$$

$$\text{LOQ} = 10. \left( \frac{s_b}{b} \right) \quad (19) \dots\dots\dots 32$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{I_m} (y_i - \hat{y}_i)^2}{I_m}} \quad (20) \dots\dots\dots 33$$

$$R = \frac{\sum_{i=1}^{I_m} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_m} (\hat{y}_i - \bar{\hat{y}})^2 (y_i - \bar{y})^2}} \quad (21) \dots\dots\dots 33$$

## RESUMO

ROQUE, Jussara Valente. Universidade Federal de Viçosa, julho de 2015. **DESENVOLVIMENTO DE MODELOS DE REGRESSÃO MULTIVARIADA PARA DETERMINAÇÃO DE ÉSTERES DE FORBOL EM SEMENTES DE JATROPHA CURCAS L. USANDO ESPECTROSCOPIA E QUIMIOMETRIA.** Orientador: Reinaldo Francisco Teófilo. Coorientadores: Luiz Antônio dos Santos Dias e Antônio Jacinto Demuner.

A construção de modelos de calibração multivariada usando espectroscopias na região do infravermelho próximo (NIR) e ultravioleta (UV), aliada à regressão por quadrados mínimos parciais (PLS), para estimar as concentrações de ésteres de forbol (PEs) em acessos/progênies de *Jatropha curcas* L. foi o objetivo deste trabalho. A composição da fase móvel para a separação dos compostos via cromatografia líquida de alta eficiência (HPLC) foi otimizada através de um planejamento composto central (CCD) com duas variáveis. A partir disso, análises cromatográficas foram realizadas para determinar a concentração dos PEs que foi expressa em  $\text{mg mL}^{-1}$  de equivalente de forbol-12-miristato-13-acetato (PMA). Os espectros foram obtidos a partir de diferentes materiais: semente com e sem a casca, óleo e extrato do óleo. Os espectros NIR foram obtidos na faixa de 4000 a  $10000 \text{ cm}^{-1}$ , os espectros UV, na faixa de 210 a 350 nm e os espectros UV-VIS-NIR, na faixa de 250 a 1000 nm. Os modelos foram construídos empregando a regressão PLS e dois algoritmos para seleção de variáveis foram testados: o algoritmo genético (GA) e o método de seleção dos preditores ordenados (OPS). Os modelos obtidos para a semente sem casca e para o óleo, tanto no NIR quanto no UV-VIS-NIR, não foram satisfatórios. A partir dos espectros NIR obtidos via refletância difusa foi possível construir um modelo para a semente com casca após seleção de 130 variáveis com o OPS. Os parâmetros do modelo foram: RMSEP de 0,48;  $R_p$  de 0,49; RPD de 0,66 e erro relativo médio na predição de 18,63%. A partir dos espectros UV de absorbância do extrato do óleo, foi possível construir um modelo após seleção de 85 variáveis com o OPS. Os parâmetros do modelo foram: RMSEP de 0,23;  $R_p$  de 0,96; RPD de 2,67 e erro relativo médio na predição de 9,26%. O modelo construído para predições a partir da casca pode ser usado para triagem e o modelo construído a partir do extrato do óleo pode ser usado para prever com exatidão os teores de PEs. O método OPS, em todos os casos, proporcionou a construção de modelos mais simples e preditivos quando comparados aqueles obtidos pela seleção de variáveis utilizando o GA. Assim, o método de referência para quantificação de ésteres de forbol pode ser substituído pelos modelos obtidos, uma vez que são mais simples, rápidos e ambientalmente corretos.

## ABSTRACT

ROQUE, Jussara Valente. Universidade Federal de Viçosa, July, 2015. **DEVELOPMENT OF REGRESSION MULTIVARIATE MODELS FOR PHORBOL ESTERS DETERMINATION IN JATROPHA CURCAS L. SEEDS USING SPECTROSCOPY AND CHEMOMETRICS.** Advisor: Reinaldo Francisco Teófilo. Co-advisors: Luiz Antônio dos Santos Dias e Antônio Jacinto Demuner.

The building of multivariate calibration models using near infrared (NIR) and ultraviolet (UV) spectroscopies, using partial least squares (PLS) to estimate the concentrations of phorbol esters (PEs) in genotype/progeny of *Jatropha curcas* L. was the goal of this work. The composition of the mobile phase for separating compounds via high performance liquid chromatography (HPLC) was optimized through a central composite design (CCD) with two variables. Based on this, chromatographic analysis were performed to determine the concentration of PEs that was expressed in  $\text{mg mL}^{-1}$  phorbol-12-myristate-13-acetate (PMA) equivalent. Spectra data were obtained from different materials: seed shell, kernel, oil and oil extract. The NIR spectra were obtained in the range  $4000\text{-}10000\text{ cm}^{-1}$ , UV spectra in the range  $210\text{-}350\text{ nm}$ , and UV-VIS-NIR spectra in the range  $250\text{-}1000\text{ nm}$ . The models were built using the PLS regression and two algorithms for variable selection were tested: Genetic Algorithm (GA) and the Ordered Predictors Selection (OPS). The models obtained for kernel and oil, in both regions, NIR and UV-VIS-NIR, were not satisfactory. Based on the NIR spectra obtained by diffuse reflectance, was possible to building a model for the seed shell after selection of 130 variables with OPS. The model parameters were: RMSEP 0.48;  $R_p$  0.49; RPD of 0.66 and medium relative error in the prediction of 18.63%. From oil extract UV absorbance spectra, was possible to build a model after selection of 85 variables with OPS. The model parameters were: RMSEP 0.23;  $R_p$  0.96; RPD of 2.67 and medium relative error in the prediction of 9.26%. The model built for predictions from seed shell can be used for screening and the model built from the oil extract can be used to accurately predict the PEs content. The OPS method, in all cases, provided the building simpler and predictive models when compared to those obtained by the selection of variables using the GA. Thus, the reference method for the quantification of phorbol esters may be replaced by the models obtained, because they are simple, fast and environmentally friendly.

## 1 INTRODUÇÃO

Atualmente, aproximadamente 90% da energia consumida no planeta provêm de combustíveis fósseis. Isso indica um aumento dos níveis de emissão de gases do efeito estufa na atmosfera, com implicações que contribuem para o aumento da temperatura média e modificações complexas no clima (ANDERSON, 2015). Estes recursos fósseis, tais como gasolina e gás natural são limitados e insuficientes para a demanda de energia do mundo em um futuro próximo. Assim, os biocombustíveis, como o biodiesel, são relatados como uma estratégia que associa a preocupação com a segurança energética e a sustentabilidade (GUI; LEE; BHATIA, 2008; KUMAR; SHARMA, 2008).

O biodiesel derivado de óleos de sementes vegetais é uma alternativa ao diesel, como uma fonte renovável e ambientalmente limpa. Neste contexto, a *Jatropha curcas* L., conhecida também como pinhão-mansão, destaca-se como uma das mais promissoras oleaginosas não comestíveis para produção deste biocombustível (CONTRAN et al., 2013; FAIRLESS, 2007; RASHID et al., 2010).

Acredita-se que o pinhão-mansão seja originário da América Central, porém vegeta espontaneamente em diversas regiões do Brasil e de todo o mundo como a África, Índia, sudeste asiático e China (ACHTEN et al., 2008; DIVAKARA et al., 2010; HELLER, 1996). As sementes de *J. curcas* contêm cerca de 30 a 40% de óleo (KANDPAL; MADAN, 1995) sendo este de alta qualidade para a produção de biodiesel. Na extração do óleo da semente de pinhão-mansão é obtido de 60 a 80% de torta, sendo esta biomassa rica em proteínas. Dessa forma, este coproduto da extração do óleo apresenta potencial para aproveitamento na alimentação animal, agregando valor à cadeia produtiva do biodiesel (AHMED; SALIMON, 2009; OPENSHAW, 2000).

No entanto, a torta de *J. curcas* apresenta em sua composição, compostos tóxicos como os ésteres de forbol (MAKKAR; ADERIBIGBE; BECKER, 1998), limitando seu uso para a alimentação animal (MARTINEZ HERRERA et al., 2012). Felizmente, existem relatos na literatura da existência de germoplasmas livres desses compostos tóxicos (MAKKAR; BECKER, 1999).

Neste contexto, a determinação dos ésteres de forbol no óleo extraído das sementes de pinhão-mansão é importante a fim de identificar acessos (unidade vegetal de conservação do germoplasma, ou seja, o material genético de uma planta) atóxicos ou que contenham apenas traços destes compostos. Essa identificação é essencial para os programas de melhoramento genético visando a seleção de clones com menor teor dos ésteres de forbol, tornando possível o aproveitamento da torta como ração animal.

Dentre as técnicas de análise utilizadas para determinação de ésteres de forbol a que mais tem se destacado é a cromatografia líquida de alta eficiência com detecção na região ultravioleta com detector de arranjo de diodos (HPLC-DAD). A técnica HPLC-DAD tem sido usada por apresentar boa separação e bons limites de detecção (DEVAPPA; BINGHAM; KHANAL, 2013; PHASUKARRATCHAI; TONTAYAKOM; TONGCUMPOU, 2012; WANG et al., 2012). No entanto, essas análises são morosas, trabalhosas, onerosas e destrutivas, o que pode, em alguns casos, inviabilizar a avaliação e seleção de genótipos durante o desenvolvimento das cultivares. Por outro lado, métodos rápidos, não destrutivos e ambientalmente corretos são possíveis empregando métodos espectroscópicos e modelos de calibração multivariada (MONTES et al., 2013).

A espectroscopia na região do infravermelho próximo (NIR), tem sido amplamente utilizada para medir a qualidade de produtos agrícolas e alimentares. Esta técnica espectroscópica não invasiva pode fornecer rapidamente as informações físicas e químicas sobre amostras. Baye et al. (2006) usaram o NIR para predição da composição da semente de milho e Choung et al. (2001) empregaram o NIR para determinar o teor de proteína e do óleo da soja (BAYE; PEARSON; SETTLES, 2006; CHOUNG et al., 2001). Outra técnica espectroscópica, o ultravioleta - visível (UV/Vis), aliada à calibração multivariada tem se tornado bastante popular para quantificação de espécies de interesse. Na literatura existem diversos modelos de predição de componentes em alimentos, como a quantificação de taninos (ALEIXANDRE-TUDO et al., 2015) e a determinação da acidez em vinhos (MORAIS et al., 2015) e em produtos farmacêuticos, como a determinação simultânea de cafeína e paracetamol (MALUF et al., 2008).

No entanto, a construção de um modelo de calibração multivariada exige, na maioria das vezes, que análises da propriedade de interesse pelo método de referência sejam realizadas. Além disso, as amostras do conjunto de calibração precisam ter alta variabilidade nos valores da propriedade de interesse. O método de calibração multivariada usando informações de origem química pertence à uma área do conhecimento denominada quimiometria (BRERETON, 2003; FERREIRA, 2015).

Depois de construído e validado, o modelo pode ser usado para predições de amostras desconhecidas, sem a necessidade de realizar as análises de referência. Assim, por exemplo, a partir de um espectro, que pode ser obtido em segundos, obtêm-se os valores da propriedade de interesse calibrada para escolha do genótipo com a característica desejada.

Desta forma, o objetivo deste trabalho foi construir e validar modelos de calibração multivariada para análise de ésteres de forbol usando métodos espectroscópicos e quimiométricos. Os objetivos específicos foram: (1) separar e pré-tratar as amostras utilizadas para construir o modelo de calibração; (2) determinar com precisão o teor dos ésteres de forbol usando o método referenciado disponível na literatura; (3) obter os espectros NIR e UV das sementes de pinhão-mansinho (casca e núcleo), do óleo extraído das sementes e dos extratos obtidos do óleo; (4) importar os dados para softwares específicos para construir e validar os modelos de calibração multivariada empregando métodos quimiométricos; (5) construir e validar modelos de calibração multivariada para realizar as previsões do teor de ésteres de forbol em sementes de *Jatropha curcas* L.

## 2 REVISÃO DA LITERATURA

### 2.1 *Jatropha curcas* L.

O pinhão-mansão (*Jatropha curcas* L.) pertence à família Euphorbiaceae, a mesma da mamona (*Ricinus* sp.), da mandioca (*Manihot* sp.) e da seringueira (*Hevea* spp.), que compreende aproximadamente 8000 espécies e cerca de 320 gêneros. O gênero *Jatropha* possui cerca de 160 espécies de plantas herbáceas e arbustivas (HELLER, 1996).

Esta planta é nativa da América central e encontra-se difundida em diversas regiões do Brasil e do mundo (DIAS; MISSIO; DIAS, 2012; KUMAR; SHARMA, 2008). É uma oleaginosa perene e de crescimento rápido, e dentre as oleaginosas estudadas para a produção do biodiesel, o pinhão-mansão destaca-se como uma planta com alto potencial, uma vez que possui alto teor de óleo e existe de forma espontânea em áreas de solo pobre em nutrientes e de clima desfavorável (ARRUDA et al., 2004). Porém, para apresentar alta produtividade, necessita de aplicações tecnológicas como o controle de pragas, a adubação e as práticas de manejo (LAVIOLA; DIAS, 2008). Esta espécie ainda se encontra em processo de domesticação e somente nos últimos 30 anos começou a ter seus aspectos agrônômicos pesquisados (SATURNINO et al., 2005).

Segundo Carnielli (2003), o pinhão-mansão é uma planta oleaginosa viável para a obtenção de biocombustíveis pois produz, no mínimo, duas toneladas de óleo por hectare, levando de três a quatro anos para atingir a idade produtiva, que pode se estender por até 40 anos (CARNIELLI, 2003).

As sementes de pinhão-mansão possuem cerca de 30 a 40% de óleo (KANDPAL; MADAN, 1995) e este apresenta qualidades necessárias para ser transformado em matéria-prima para produção de biodiesel (ARRUDA et al., 2004; PURCINO; DRUMMOND, 1986). Esta qualidade deve-se à composição de ácidos graxos saturados, em particular o ácido oléico (45%), garantindo ótimas propriedades para obtenção de biodiesel e bioquerosene (BALAT, 2011; RAMOS et al., 2009). Na Tabela 1 são apresentadas algumas propriedades físico-químicas do óleo de *J. curcas* (CETEC, 2011).

A qualidade do óleo extraído a partir das sementes de pinhão-mansão é um aspecto que merece grande atenção, visto que é um parâmetro que afeta o seu valor, o valor da torta, o valor e também a eficiência do processo de produção do biodiesel (RAMOS et al., 2009). Do ponto de vista químico, um bom óleo para produção de biodiesel deve ser rico em cadeias longas e baixo índice de insaturação de ácidos graxos (PINTO et al., 2005).

**Tabela 1.** Propriedades físico-químicas do óleo de pinhão-mansão.

| Características Físico-Químicas                        | Valores |
|--------------------------------------------------------|---------|
| Teor de Ácidos Graxos Livres<br>(como ácido oléico, %) | 0,96    |
| Densidade à 25 °C (g cm <sup>-3</sup> )                | 0,9069  |
| Índice de refração à 25 °C                             | 1,468   |
| Índice de Saponificação                                | 189     |
| Índice de Iodo                                         | 97      |
| Insaponificáveis (%)                                   | 1,1     |
| Índice de Peróxido                                     | 9,98    |
| Ponto de Solidificação (25 °C )                        | < -10   |
| Cinzas (%)                                             | < 0,1   |

Fonte: CETEC (2011).

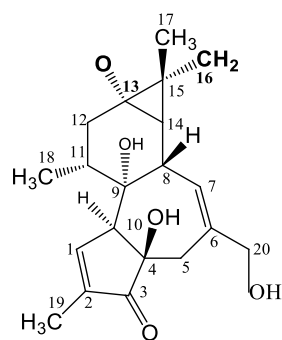
No entanto, a extração do óleo da semente de pinhão manso gera em torno de 62% de torta, rica em proteínas. Porém a torta contém compostos tóxicos como a lectina, inibidores de tripsina, saponinas, fitatos e os ésteres de forbol (MAKKAR; ADERIBIGBE; BECKER, 1998; MAKKAR et al., 1997), o que limita seu uso para a alimentação animal (MARTINEZ HERRERA et al., 2012).

#### 2.1.1 Ésteres de forbol

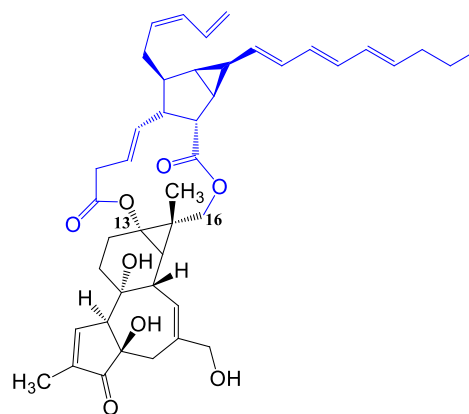
O termo “ésteres de forbol” (PEs) é usado para descrever um grupo de compostos derivados de diterpenos tetracíclicos com um esqueleto estrutural do tigliano os quais ocorrem naturalmente em muitas plantas da família Euphorbiaceae (AHMED; SALIMON, 2009; WAKANDIGARA; NHAMO; KUGARA, 2013).

Adolf et al. (1984) foram os primeiros a identificar ésteres de forbol em *J. curcas*. Seis ésteres de forbol foram caracterizados a partir de óleo de semente de pinhão-manso e designados como *Jatropha* fatores (C1, C2, epímeros C3 e C4; C5 e C6) com a fórmula molecular C<sub>44</sub>H<sub>54</sub>O<sub>8</sub> (ADOLF; OPFERKUCH; HECKER, 1984). Todos eles são diésteres intramoleculares da mesma estrutura base diterpenóide, o 12-desoxi-16-hidroxiforbol, como apresentado na Figura 1.

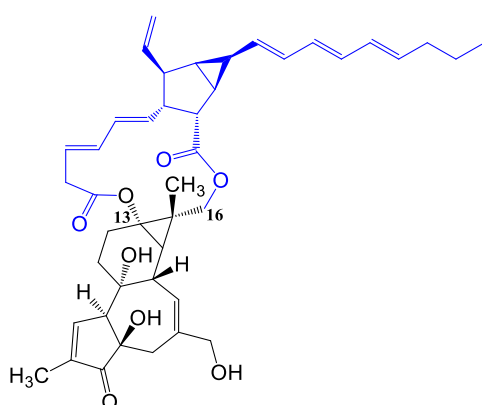
Dependendo da sua configuração ( $\alpha$  e  $\beta$ ), os ésteres de forbol são classificados como ativo ( $\beta$ ) e inativo ( $\alpha$ ) (GOEL et al., 2007), e no pinhão-manso, todos os seis fatores estão na forma ativa ( $\beta$ ) e são citados na literatura como tóxicos (DEVAPPA; MAKKAR; BECKER, 2010b; LI et al., 2010; MAKKAR; ADERIBIGBE; BECKER, 1998).



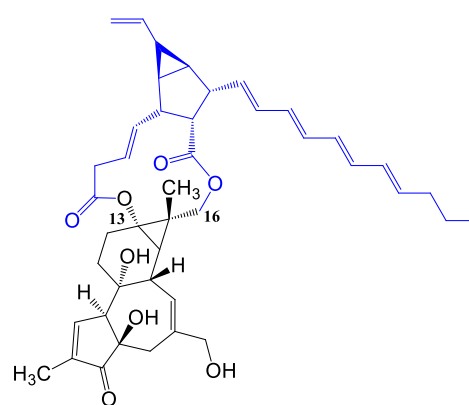
**Estrutura Base**



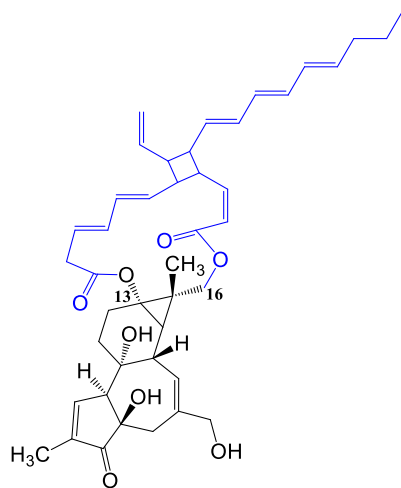
**Fator Jatropha C1**



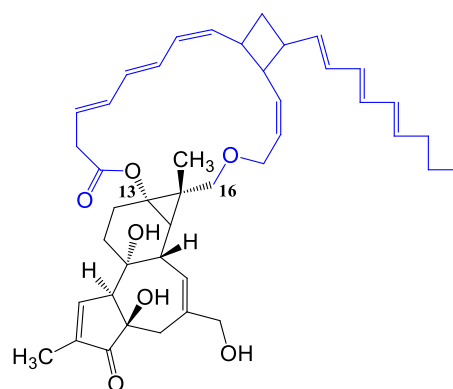
**Fator Jatropha C2**



**Fator Jatropha C3 e C4 (epímeros)**



**Fator Jatropha C5**



**Fator Jatropha C6**

**Figura 1.** Estrutura base (12-desoxi-16-hidroxiforbol) e os seis ésteres de forbol presentes em *Jatropha curcas* L.

Relatos de toxicidade humana por consumo acidental de sementes de pinhão-mansó são observados com sintomas tais como dor abdominal, vômitos e diarreia

(JOUBERT et al., 1984; KULKARNI et al., 2005; MAMPANE; JOUBERT; HAY, 1987).

A concentração dos PEs no pinhão-mansinho varia de acordo com a região em que a planta cresceu ou foi cultivada. Ahmed e Salimon (2009) mostraram a variação no teor dos ésteres de forbol no óleo da semente de *J. curcas* da Malásia (0,23%), Indonésia (1,58%) e Índia (0,58%) (AHMED; SALIMON, 2009). Saetae e Suntornsuk (2010) demonstraram uma variação de 0,21 a 0,48 mg g<sup>-1</sup> no teor dos PEs em diferentes províncias da Tailândia (SAETAET; SUNTORNUSUK, 2010). No Brasil, em um estudo de Ferrari et al. (2009), o teor de ésteres de forbol variou de 1,41 a 8,97 mg g<sup>-1</sup> nas sementes, demonstrando a presença no território brasileiro de plantas que produzem sementes com baixa e com elevada toxicidade (FERRARI et al., 2009).

Em diversas regiões por todo o mundo, são relatados teores de PEs com uma grande variabilidade de toxicidade. Na Tabela 2 estão listadas algumas referências com a quantidade dos ésteres relatada.

**Tabela 2.** Teores de ésteres de forbol encontrados na literatura.

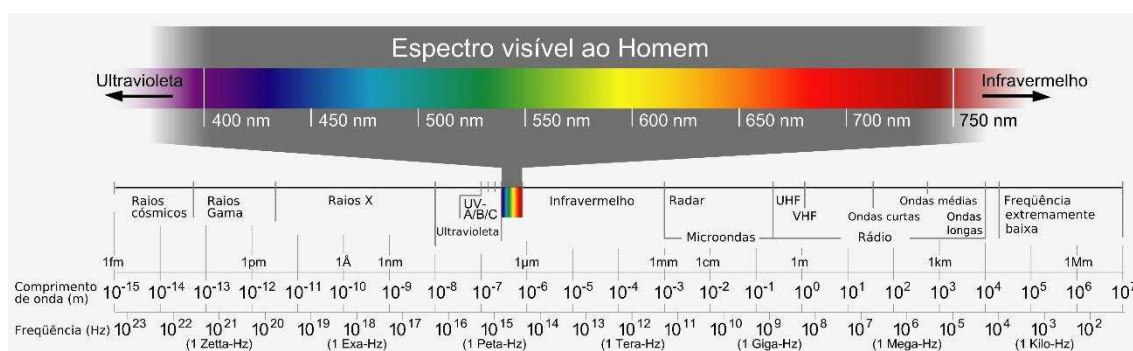
| Teor dos PEs                                     | Referências                         | Local de obtenção das amostras                             |
|--------------------------------------------------|-------------------------------------|------------------------------------------------------------|
| 2,84 – 3,16 mg g <sup>-1</sup> semente           | (OSKOEUEIAN et al., 2011))          | Malásia                                                    |
| 2 – 7 mg mL <sup>-1</sup> óleo                   | (DEVAPPA; MAKKAR; BECKER, 2010a)    | Índia                                                      |
| 2,17 – 2,30 mg g <sup>-1</sup> semente           | MAKKAR; ADERIBIGBE e BECKER, (1998) | Cabo Verde, Nicarágua, Nigéria e México                    |
| 0,11 mg g <sup>-1</sup> semente*<br>(não tóxica) | MAKKAR; ADERIBIGBE e BECKER, (1998) | Cabo Verde, Nicarágua, Nigéria e México                    |
| 0,87 – 3,32 mg g <sup>-1</sup> semente           | MAKKAR et al., (1997)               | Países da África, América do Norte, América Central e Ásia |
| 2,10 – 2,14 mg g <sup>-1</sup> óleo              | PRASAD et al., (2012)               | Índia                                                      |
| 3,10 – 3,38 mg g <sup>-1</sup> óleo              | DEVAPPA; BINGHAM e KHANAL, (2013)   | Estados Unidos                                             |
| 0 – 10,3 mg g <sup>-1</sup> semente              | MONTES et al., (2013)               | Alemanha                                                   |

Em destaque na Tabela 2, uma cultivar encontrada no México, contendo baixos teores dos ésteres de forbol, e dessa forma é considerada não tóxica.

## 2.2 Espectroscopia

A espectroscopia é a ciência que estuda as interações da radiação eletromagnética com a matéria. A informação obtida pela espectroscopia é conhecida como espectro. A partir da análise espectral é possível obter informações importantes sobre as ligações químicas, estrutura molecular, modo de interação entre as moléculas e também quantificar as espécies espectroativas.

A energia eletromagnética pode ser ordenada em função de seu comprimento de onda ou de sua frequência, sendo esta disposição denominada espectro eletromagnético, que apresenta subdivisões de acordo com as características de cada região.



**Figura 2.** Espectro eletromagnético.

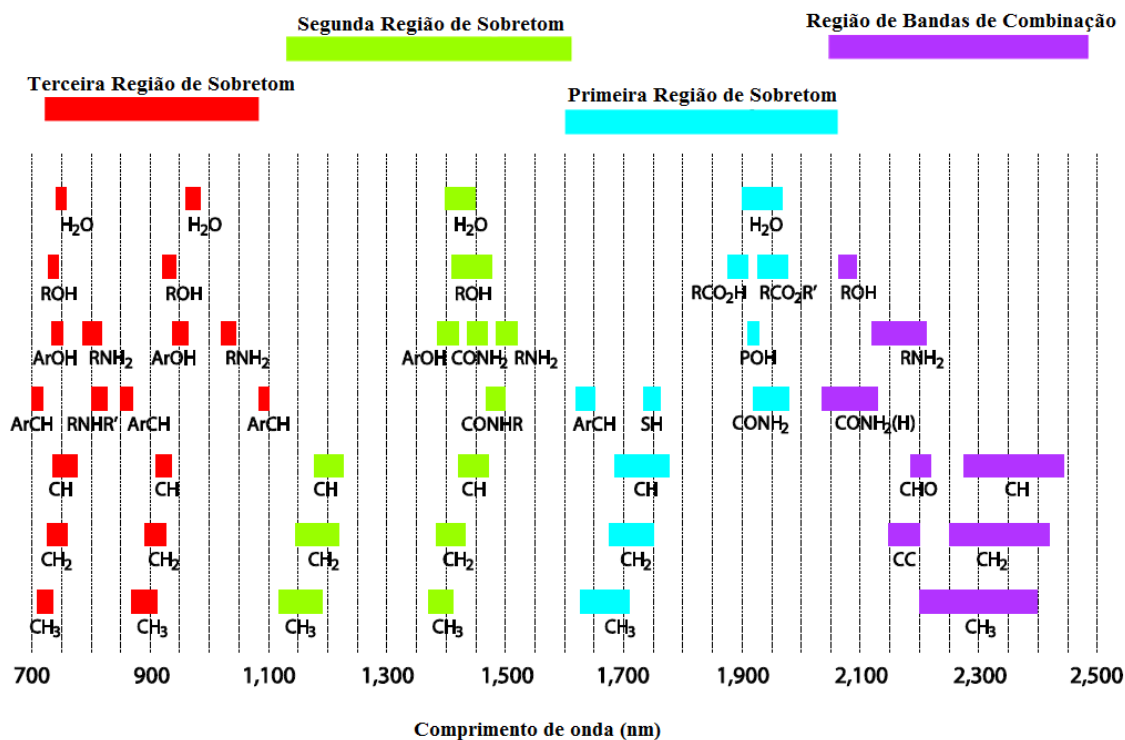
Neste trabalho, foram utilizadas as espectroscopias NIR e UV/Vis, que serão brevemente descritas a seguir.

### 2.2.1 Infravermelho Próximo

A espectroscopia na região do infravermelho próximo é uma técnica simples, rápida, não destrutiva, com alto poder de penetração do feixe de radiação, com exigências mínimas no preparo de amostras e que proporciona análise de vários tipos de matrizes com níveis de exatidão e precisão que são comparáveis aos métodos de referência (BLANCO; VILLARROYA, 2002; PASQUINI, 2003).

Esta técnica tem sido largamente aplicada para a resolução de diversos tipos de problemas analíticos encontrados nas mais diversas áreas como agricultura, indústria de alimentos, petróleo, têxtil, carvão, cosméticos, medicina, polímeros, química, tintas e farmacêuticas (ALVES; POPPI, 2013; AZIZIAN; KRAMER, 2005; COÛTEAUX et al., 2005; MCCLURE, 2003).

A região espectral do NIR compreende a faixa de radiação de  $12800$  a  $4000\text{ cm}^{-1}$  em número de onda ou de  $700$  a  $2500\text{ nm}$  em comprimento de onda, como pode ser visto na Figura 3 (BOKOBZA, 1998).



**Figura 3.** Bandas e posições relativas de picos de absorção em infravermelho próximo.

Como mostrado na Figura 3, as vibrações observadas na espectroscopia NIR são referentes a sobretons e combinações de ligações fundamentais de pequena intensidade. A ocorrência das bandas de absorção do NIR é devido às vibrações anarmônicas dos átomos, além da mudança no momento de dipolo. As regiões de transições de sobretons aparecem entre 700 e 2000 nm e as bandas de combinação entre 1900 e 2500 nm. Essas bandas de combinação aparecem em moléculas poliatômicas, onde vários modos de vibração podem interagir, sendo a banda resultante a soma de cada frequência de interação (AENUGU et al., 2011).

De forma geral, as ocorrências espectrais na região NIR provêm de ligações das moléculas em que o hidrogênio está envolvido, o que torna a técnica útil para a determinação de compostos orgânicos contendo ligações C-H, N-H, S-H e O-H (PASQUINI, 2003). Como essas ligações estão presentes na maior parte das moléculas, a utilização do NIR em análise qualitativa para identificação de compostos é bastante restrita. Além disso, para análise quantitativa, como os espectros de absorção NIR são complexos, é necessário o uso de procedimentos quimiométricos (DRENNEN; KRAEMER; LODDER, 1991; WETZEL, 1983).

### 2.2.2 Ultravioleta Visível

A espectroscopia de absorção molecular do ultravioleta e do visível compreende a região de 180 nm a 780 nm, sendo 180 a 380 nm a região do ultravioleta e entre 380 e 780 nm, correspondente a luz visível. Esta técnica utiliza a radiação eletromagnética e quando uma molécula é estimulada com esse tipo de radiação, ela pode sofrer transições eletrônicas por ocasião da absorção de energia quantizada. O espectro eletrônico de absorção é o registro gráfico da transição eletrônica após interação da radiação com a matéria.

Nestas regiões do espectro a absorção da radiação promove transições de níveis eletrônicos em seu estado de energia mais baixo ou estado fundamental para um estado de maior energia ou estado excitado. A transição ocorre do orbital molecular ocupado mais alto (HOMO) para o orbital molecular não ocupado mais baixo (LUMO) e a energia absorvida é exatamente igual à diferença de energia entre os dois níveis energéticos (PERKAMPUS, 1992).

Na espectroscopia de absorção molecular UV/Vis, obtemos informações sobre o analito medindo-se a quantidade de radiação eletromagnética absorvida pelas espécies espectroativas presentes na matéria analisada. O valor da absorção pode ser relacionada à concentração do analito (SKOOG et al., 2005).

A espectroscopia de absorção é fundamentada na lei de Lambert-Beer, que é a base matemática para medidas de absorção de radiação por amostras no estado sólido, líquido ou gasoso, nas regiões ultravioleta, visível e infravermelho do espectro eletromagnético. Para medidas de absorção de radiação em determinado comprimento de onda, tem-se:

$$A = \varepsilon \cdot c \cdot b \quad (1)$$

em que A é a absorbância,  $\varepsilon$  é a absortividade molar, que é uma grandeza característica da espécie absorvente, cuja magnitude depende do comprimento de onda da radiação incidente. O termo c é a concentração da espécie absorvente e b, a distância percorrida pelo feixe através da amostra, ou seja, o caminho ótico.

A espectroscopia de absorção na região UV/Vis é uma das técnicas analíticas mais empregadas em função da sua robustez, custo relativamente baixo e grande número de aplicações desenvolvidas para análise atômica e molecular (SALDANHA; ARAÚJO; NETO, 1999; SANTOS; DEMIATE; NAGATA, 2010).

No entanto, apesar da extensa utilização, várias inovações recentes relacionadas às medidas espectrofotométricas ainda são pouco conhecidas (ROCHA; TEIXEIRA, 2004). Em relação aos instrumentos, é notável a evolução desde os procedimentos baseados em

comparações visuais e dos fotômetros empregando filtros para a seleção de comprimentos de onda até os equipamentos mais recentes, que empregam arranjos de fotodetectores. Esses permitem a aquisição de espectros de absorção em intervalos de tempo da ordem de milésimos de segundos e são disponíveis, inclusive, em versões miniaturizadas acopladas (BROEKE; LANGERGRABER; WEINGARTNER, 2006).

### **2.3 Cromatografia Líquida de Alta Eficiência**

As técnicas cromatográficas de análise estão entre as principais técnicas de separação, especialmente na análise de substâncias presentes em matrizes complexas, tais como fluidos biológicos, produtos naturais, sedimentos de rio e outras. Isto se deve, principalmente, à sua capacidade de separação dos componentes presentes nas misturas em função da eficiência e do poder de resolução das colunas modernas (LANÇAS, 2009).

Essa técnica é utilizada pelos químicos para separar espécies em uma grande variedade de materiais orgânicos, inorgânicos e biológicos. Na cromatografia líquida, a fase móvel é um solvente líquido, ou mistura de solventes, os quais contêm a amostra na forma de uma mistura de solutos (SKOOG et al., 2005).

A separação usando HPLC é o tipo mais versátil e mais amplamente empregado de cromatografia por eluição. Ela está fundamentada na migração diferencial dos componentes de uma mistura, que ocorre devido a diferentes interações, entre duas fases imiscíveis, a fase móvel e a fase estacionária (DEGANI; CASS; VIEIRA, 1998).

Seu funcionamento inicia com o bombeamento da fase móvel sob alta pressão e vazão controladas, em seguida, a amostra é injetada em pequena quantidade e arrastada pela fase móvel através da coluna cromatográfica. A velocidade de eluição, na coluna, dos componentes da amostra é inversamente proporcional à interação destes com o revestimento da coluna. Por fim, a fase móvel, juntamente com os componentes da amostra, sai da coluna e vai para o detector, que gera um sinal proporcional à concentração do analito. Este sinal é enviado para um sistema de registro e tratamento de dados onde será gerado um cromatograma (COLLINS; BRAGA; BONATO, 2006).

O cromatograma é um gráfico gerado pelo detector e processamento de saída do instrumento, onde a resposta deste está em função do tempo de retenção, ou seja, o tempo gasto desde o momento da injeção até a saída do ponto máximo do pico, para uma amostra. A concentração dos analitos no cromatograma é proporcional à área ou à altura do pico (COLLINS; BRAGA; BONATO, 2006; LANÇAS, 2009).

Dentre os detectores mais empregados em HPLC está o de arranjo de fotodiodos (DAD). Este detector permite que a absorbância de uma amostra em todos os

comprimentos de onda seja determinada de modo simultâneo. O seu princípio é baseado na absorção de luz ultravioleta ou visível, por parte da amostra (LANÇAS, 2009). Para determinação de ésteres de forbol por HPLC são utilizados estes detectores uma vez que eles apresentam alta sensibilidade, não são destrutivos, permitem trabalhar com gradiente de fase móvel, respondem a uma larga faixa de concentração, dentre outras vantagens. Diversos trabalhos são encontrados na literatura com o uso do HPLC-DAD para quantificação dos PEs (BALDINI et al., 2014; DEVAPPA; BINGHAM; KHANAL, 2013; MONTES et al., 2013; WANG et al., 2012).

## **2.4 Quimiometria**

Com os avanços das técnicas instrumentais e a acentuada utilização de computadores que permitem o armazenamento de um grande volume de dados, torna-se necessário o uso de tratamentos matemáticos e estatísticos mais complexos, a fim de se obter informações mais completas e exatas. Dessa forma, a utilização de métodos multivariados em dados químicos está bem difundida e aceita pela comunidade científica, consistindo em uma área denominada quimiometria.

A quimiometria é a aplicação de métodos matemáticos e estatísticos a dados de origem química que tem por objetivo a obtenção de informação química de um conjunto complexo de informações (BRERETON, 2003; FERREIRA et al., 1999). Diversos métodos quimiométricos têm sido propostos para os mais diversos fins, como o planejamento e otimização de experimentos, a regressão multivariada e a seleção de variáveis (BARROS NETO; SCARMINIO; BRUNS, 2001; GEMPERLINE, 2006; MASSART; VANDEGINSTE, 1998) que serão apresentados brevemente a seguir.

### **2.4.1 Planejamentos Experimentais**

Os planejamentos experimentais (DOE) são definidos como uma série de experimentos nos quais determinadas mudanças são feitas nas variáveis de entrada do sistema (variáveis independentes) de tal forma que seja possível observar quais das variáveis independentes estudadas possuem efeito sobre a variável dependente, também denominada de resposta (MONTGOMERY, 1987).

Por definição, as variáveis independentes serão, neste texto, denominadas de variáveis ou fatores. A variável dependente será denominada de resposta.

Os DOE permitem identificar o efeito das variáveis individuais e de suas interações além de minimizar a quantidade de níveis em estudo de cada um dos fatores em um processo ou sistema, minimizando assim o tempo e o dinheiro gastos na execução do experimento. Estes métodos exigem a definição de níveis. Normalmente, usa-se em

planejamentos fatoriais apenas dois níveis os quais são nomeados pelos sinais - (menos) para o nível mais baixo e + (mais) para o nível mais alto (EBRAHIMI-NAJAFABADI; LEARDI; JALALI-HERAVI, 2014).

Os fatores significativos selecionados após execução de um planejamento adequado de triagem, como o planejamento fatorial completo ou fracionário, podem ser empregados em uma metodologia de superfície de resposta (RSM) para a otimização de um sistema ou processo. Dentre esses métodos, o planejamento composto central (CCD) e matriz Doehlert, são os mais utilizados (TEÓFILO; FERREIRA, 2006).

Um experimento para triagem é executado com o interesse em se determinar as variáveis de entrada e as interações entre variáveis que têm influência significativa sobre as diferentes respostas de interesse (LUNDSTEDT et al., 1998). Após selecionar as variáveis a serem estudadas e que provavelmente interferem no sistema, é preciso avaliar o método experimental e escolher o planejamento para estimar o efeito das diferentes variáveis no resultado (EBRAHIMI-NAJAFABADI; LEARDI; JALALI-HERAVI, 2014). Em estudos para triagem, modelos lineares e com interações de primeira ordem são os mais utilizados, tais como os planejamentos fatoriais completos e fracionários. Estes planejamentos são escolhidos por fornecerem condições de estudar o sistema com apenas dois níveis para cada variável. Além disso, pode-se incluir nestes tipos de planejamentos repetições no centro do planejamento (nível zero: 0), em que o valor médio dos níveis de todas as variáveis é empregado (LUNDSTEDT et al., 1998; TEÓFILO; FERREIRA, 2006).

Para encontrar o nível ótimo para cada variável é importante construir modelos empíricos lineares ou quadráticos. Estes modelos são obtidos com os planejamentos para metodologias de superfície de resposta. Tais planejamentos são mais abrangentes e, além de descrever todo o sistema estudado, o explora até a otimização (BEZERRA et al., 2008; TEÓFILO; FERREIRA, 2006). Em química analítica, as RSM mais utilizadas são simétricas e descrevem um domínio experimental esférico, tais como o CCD (FERREIRA et al., 2004).

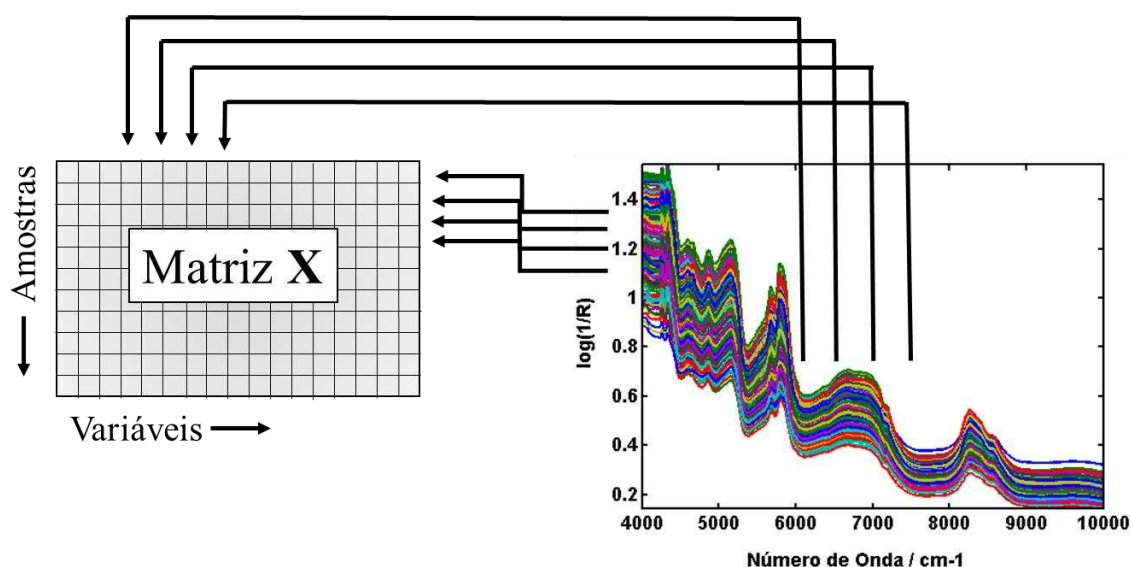
O CCD esférico e ortogonal é a combinação de um planejamento fatorial completo ou fracionário, com experimentos fora da região fatorial (pontos axiais) e pelo menos três experimentos internos e centrais à região fatorial (pontos centrais). Tais planejamentos apresentam propriedades como a rotabilidade ou ortogonalidade a fim de obter um ajuste quadrático (BEZERRA et al., 2008). Os pontos axiais para o CCD esférico estão situados nos eixos de sistema de coordenadas com distância  $\pm\alpha$  da origem, com  $\alpha = \sqrt[4]{2^k}$ , em que

$k$  é o número de variáveis a serem estudadas. Para construir o CCD é necessário definir o  $k$ , qual o planejamento fatorial será utilizado e quantos experimentos serão realizados no ponto central. O número de experimentos a ser realizado é dado por  $2^k + 2k + 1$  (TEÓFILO; FERREIRA, 2006).

#### 2.4.2 Calibração Multivariada

A calibração é o processo de construção de um modelo matemático para relacionar respostas instrumentais à(s) propriedade(s) conhecida(s) de padrões. O principal objetivo do modelo construído é realizar previsões das propriedades desconhecidas a partir de respostas obtidas de amostras (BEEBE; KOWALSKI, 1987).

A calibração multivariada inversa tem como objetivo resolver o sistema linear  $\mathbf{y} = \mathbf{X}\mathbf{b}$ . A matriz  $\mathbf{X}(n,m)$ , onde  $n$  é o número de linhas e  $m$  o número de colunas, é formada pelas respostas instrumentais dos padrões, conhecidas como variáveis independentes. O vetor  $\mathbf{y}(n,1)$  é conhecido como variável dependente e contém os valores da propriedade de interesse de cada padrão. O vetor  $\mathbf{b}(m,1)$  é a solução do sistema, o qual é denominado de vetor de regressão ou modelo da regressão. A Figura 4 ilustra como uma matriz  $\mathbf{X}$  pode ser construída a partir de sinais espectrais.



**Figura 4.** Construção da matriz  $\mathbf{X}$  para calibração multivariada.

A calibração consiste basicamente em duas etapas: (1) a modelagem, que estabelece uma regressão usando o conjunto de calibração, e (2) a validação, que avalia o desempenho do modelo de regressão construído. A validação é realizada com um conjunto de predição através da comparação dos valores de predição de padrões com seus valores conhecidos e assim os erros de previsões são avaliados (FERREIRA et al., 1999; GELADI; KOWALSKI, 1986; WOLD; SJÖSTRÖM; ERIKSSON, 2001).

A coleta dos dados para a obtenção do modelo é uma etapa muito importante e deve ser feita com atenção. Antes da construção do modelo, após serem coletados, os dados normalmente recebem algum tipo de pré-tratamento. Este procedimento é importante para eliminar variações indesejadas que não são eliminadas durante a análise e aquisição de dados e que podem interferir no resultado final (FERREIRA, 2015). Os pré-tratamentos que podem ser aplicados às linhas da matriz  $\mathbf{X}$  são chamados de transformações, como por exemplo a correção do espalhamento multiplicativo (MSC) e as derivadas primeira e segunda. Nas colunas de  $\mathbf{X}$ , são aplicados os pré-processamentos como o autoescalamento e a centragem na média (RINNAN; BERG; ENGELSEN, 2009).

Existem diversos métodos de regressão multivariada, dentre eles pode-se destacar a regressão por quadrados mínimos parciais (PLS), reconhecidamente o método mais utilizado para construção de modelos de calibração de primeira ordem a partir de dados de origem química (BRERETON, 2000).

### **2.4.3 Quadrados Mínimos Parciais**

A regressão por quadrados mínimos parciais está dentro da classe dos métodos de regressão inversa. Este método não requer um conhecimento exato de todos os componentes presentes nas amostras podendo realizar a predição de um analito de interesse mesmo na presença de interferentes, desde de que os interferentes também estejam presentes na construção do modelo (BRERETON, 2000). O PLS pode ser utilizado em dados altamente correlacionados e com ruídos experimentais, além de permitir a modelagem e predições simultânea de mais de um analito, i.e., uma matriz  $\mathbf{Y}$  (BEEBE; KOWALSKI, 1987; WOLD et al., 2001).

O PLS é calculado através de algoritmos computacionais. Diversos algoritmos PLS estão disponíveis na literatura e em softwares, tais como NIPALS, bidiagonal, SIMPLS e Kernel (DE JONG, 1993; LINDGREN; GELADI; WOLD, 1993; MANNE, 1987; WOLD et al., 1984). Martins et al. (2010) testaram alguns algoritmos PLS e concluíram que todos atingem o mesmo resultado, porém o algoritmo bidiagonal é o que possui menor tempo de pré-processamento computacional e também é o mais intuitivo (MARTINS; TEOFILLO; FERREIRA, 2010). Além disso, Golub et al. estudaram alguns algoritmos PLS e concluíram que o bidiagonal é o mais correto matematicamente (GOLUB; KAHAN, 1965; GOLUB; VAN LOAN, 1996). Dessa forma, a regressão através do algoritmo bidiagonal foi escolhida para a realização desse trabalho.

O algoritmo bidiagonal baseia-se na decomposição da matriz  $\mathbf{X}$  em três matrizes, conforme Equação 2 (BARLOW; BOSNER; DRMAC, 2005):

$$\mathbf{y} \rightarrow \mathbf{X} = \mathbf{URV}^t \quad (2)$$

em que  $\mathbf{UR}$  é a matriz de scores e  $\mathbf{V}$  a matriz de loadings.

Na Equação 2, a seta de  $\mathbf{y}$  para  $\mathbf{X}$  indica que durante a decomposição da matriz  $\mathbf{X}$  é incorporada a informação do vetor  $\mathbf{y}$  (WOLD; SJÖSTRÖM; ERIKSSON, 2001).

As colunas da matriz  $\mathbf{U}$  e as linhas da matriz  $\mathbf{V}^t$  criam um novo sistema de eixos da matriz original e são denominadas variáveis latentes ( $h$ ). As variáveis latentes são ortogonais e trazem informações decrescentes e complementares da variância da matriz  $\mathbf{X}$ . Normalmente, as primeiras variáveis latentes (entre 2 a 10) são capazes de reconstruir a matriz  $\mathbf{X}$  com praticamente 100% da informação original.

O algoritmo pode ser descrito na seguinte formulação resumida (MARTINS; TEOFILO; FERREIRA, 2010):

1. Inicialize o algoritmo para primeira componente,

$$\nu_1 = \mathbf{X}^t \mathbf{y} / \|\mathbf{X}^t \mathbf{y}\|; \alpha_1 \mu_1$$

2. Para  $i = 2, \dots, h$  componentes:

$$2.1 \quad y_{i-1} \nu_1 = \mathbf{X}^t \mu_{i-1} - \alpha_{i-1} \nu_{i-1}$$

$$2.2 \quad \alpha_i \mu_i = \mathbf{X} \nu_i - y_{i-1} \mu_{i-1}$$

$$\text{Com, } \mathbf{V}_h = (\nu_1, \dots, \nu_h), \quad \mathbf{U}_h = (\mu_1, \dots, \mu_h) \text{ e } \mathbf{R}_h = \begin{pmatrix} \alpha_1 & y_1 & & \\ & \ddots & \ddots & \\ & & \ddots & \\ & & & \alpha_{k-1} & y_{k-1} \\ & & & & \alpha_k \end{pmatrix} \text{ prova-se}$$

ainda que  $\mathbf{XV}_h = \mathbf{U}_h \mathbf{R}_h$  e, portanto,  $\mathbf{R}_h = \mathbf{U}_h^t \mathbf{XV}_h$ .

Com as matrizes  $\mathbf{U}$ ,  $\mathbf{V}$  e  $\mathbf{R}$  truncadas para  $h$  variáveis latentes, pode-se estimar a pseudo-inversa de Moore-Penrose de  $\mathbf{X}$  e resolver o problema dos quadrados mínimos, como mostrado a seguir, Equação 3.

$$\mathbf{y} = \mathbf{Xb} \rightarrow \hat{\mathbf{X}} = \mathbf{U}_h \mathbf{R}_h \mathbf{V}_h^t \rightarrow \hat{\mathbf{b}} = \mathbf{V}_h \mathbf{R}_h^{-1} \mathbf{U}_h^t \mathbf{y} \quad (3)$$

em que  $h$  é o número de variáveis latentes selecionado para a construção do modelo.

A suposição básica de qualquer modelo PLS é que o sistema estudado é conduzido por um pequeno número de variáveis latentes. Assim, a escolha do número de variáveis latentes  $h$  é de extrema importância para evitar sub ajuste ou sobre ajuste do modelo, pois as matrizes reconstruídas dependem do valor de  $h$ . O sub ajuste acontece quando a modelagem de dados é insuficiente para explicar toda informação e o sobre ajuste ocorre com a inclusão de excesso de informação no modelo, que pode ser aleatória ou estar relacionada à presença de erros sistemáticos (BRERETON, 2007).

Para escolha do número de variáveis latentes utiliza-se normalmente a técnica de validação cruzada, que se baseia no procedimento de reamostragem. Um gráfico relacionando os erros nesta reamostragem versus  $h$  é construído e o ponto com menor erro é então selecionado. Normalmente o cálculo do erro utilizado é a raiz quadrada do erro quadrático médio de validação cruzada (RMSECV), conforme apresentado na Equação 4:

$$\text{RMSECV} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{I_{cv}}} \quad (4)$$

em que  $\hat{y}_i$  é o valor estimado para a amostra  $i$ , e  $I_{cv}$  é o número de amostras da validação interna.

Os principais métodos de validação cruzada são: (1) leave-one-out, que remove uma amostra de cada vez; (2) blocos contíguos, que separa amostras em blocos de amostras sequenciais; (3) venezianas (venetian blinds), que separa amostras sistematicamente espaçadas; e (4) subconjuntos aleatórios, que separa, aleatoriamente, conjuntos de amostras (BRERETON, 2007).

Além disso, a qualidade dos modelos foi avaliada pelo RPD (razão de desempenho do desvio) (WILLIAMS; NORRIS, 1987). O RPD é calculado de acordo com a Equação 5:

$$\text{RPD} = \frac{D_p}{\text{RMSECV}} \quad (5)$$

em que  $D_p$  é o desvio padrão dos valores estimados pela validação cruzada.

Este parâmetro é utilizado para verificar a eficiência do modelo em relação à variação natural das amostras. De acordo com a literatura, valores de RPD acima de 2,0 já podem ser considerados como modelos excelentes, de 1,4 à 2,0 modelos aceitáveis e menor que 1,4 são modelos não confiáveis ou indicados para triagem (CHANG et al., 2001; DUNN et al., 2002).

O bias, ou viés, definido como a diferença entre o valor de uma medida e o valor de referência, é a medida de erros sistemáticos no modelo, sendo calculado apenas com as amostras de validação. O cálculo de viés para calibração multivariada é apresentado na Equação 6:

$$\text{viés} = \frac{\sum_{i=1}^{nv} (y_i^{\text{ref}} - \hat{y}_i)}{I_m} \quad (6)$$

em que  $y_i^{\text{ref}}$  é o valor de referência e  $\hat{y}_i$  é o valor previsto de cada amostra,  $I_m$  é número de amostras do conjunto de validação e  $nv$ , graus de liberdade.

$$SDV = \sqrt{\frac{\sum [(y_i^{\text{ref}} - \hat{y}_i) - \text{viés}]^2}{nv - 1}} \quad (7)$$

$$t_{\text{calc}} = \frac{|\text{viés}| \sqrt{nv}}{SDV} \quad (8)$$

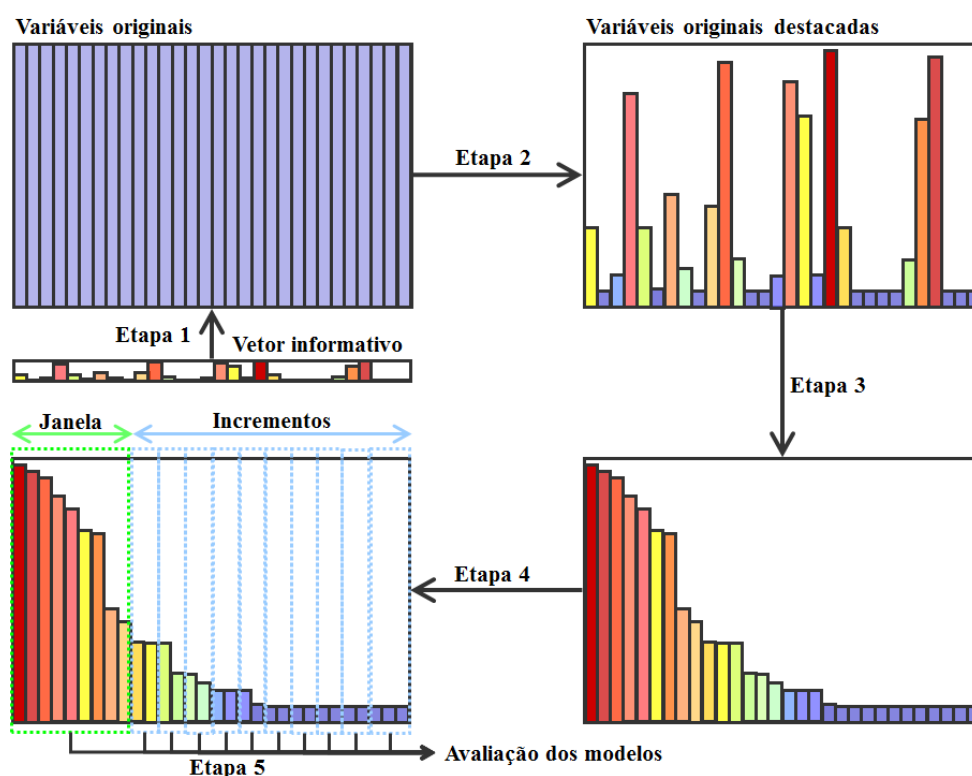
O SDV, desvio padrão de erros de validação (Equação 7), é utilizado para verificar se o viés é significativamente diferente de zero através de um teste t de Student (Equação 8).

#### 2.4.4 Seleção de Variáveis

Os métodos de regressão por compressão de dados como o PLS tornaram possível construir modelos de calibração multivariada com um grande número de variáveis altamente correlacionadas (colinearidade). Se considerarmos como exemplo uma matriz de dados contendo espectros, apesar da colinearidade, existem regiões espectrais onde variações espectrais não estão relacionadas com as concentrações dos componentes químicos de interesse presentes na amostra. Além disso, há também regiões em que não existe linearidade e ainda, há aquelas onde a relação sinal ruído é muito baixa. Estas evidências indicam que a escolha adequada de regiões espectrais pode melhorar significativamente a eficiência do modelo de calibração multivariada, além de torná-lo mais simples para interpretações. A seleção de variáveis envolve a escolha de determinadas regiões do espectro, que permitem ao modelo de calibração minimizar os erros de predição. Como consequência da seleção de variáveis, é possível produzir um modelo mais robusto, simples de interpretar e com melhor exatidão nas predições (OLIVEIRA et al., 2004; XIAOBO et al., 2010).

Na literatura, há diversos tipos de algoritmos para seleção de variáveis (ARAÚJO et al., 2001; LEARDI; LUPIÁÑEZ GONZÁLEZ, 1998; NORGAARD et al., 2000). Dentre eles, o algoritmo genético (GA) é um dos métodos mais utilizados. Este método resolve problemas de otimização com base em um processo de seleção natural e genética que imita a evolução biológica. Na primeira etapa, uma população inicial é constituída por uma série de diferentes indivíduos, e a quantidade de indivíduos é definido como o tamanho da população. Estes indivíduos são analisados e cruzados de acordo com os parâmetros introduzidos em algoritmo. No final do processo, um vetor contendo zeros e uns indica se variáveis devem ser incluídas (1) ou não (0).

No entanto, a maioria dos métodos de seleção de variáveis não são generalizados e eficientes para todos os tipos de variáveis. Por exemplo, em química, diversas técnicas analíticas fornecem diferentes tipos de preditores (colunas de  $X$ ) e cada um apresenta uma peculiaridade e características bem definidas (TEÓFILO; MARTINS; FERREIRA, 2009). Para solucionar este problema, Teófilo et al. (2009) desenvolveram o método de seleção dos preditores ordenados (OPS), um algoritmo para seleção de variáveis em regressão multivariada apresentado na Figura 5.



**Figura 5.** Método de seleção de variáveis usando a seleção dos preditores ordenados (OPS).

O método se baseia na obtenção de um vetor informativo (Figura 5 – Etapa 1) que contenha a localização das melhores variáveis independentes para predição. Posteriormente, na segunda etapa (Figura 5 – Etapa 2), as variáveis originais são destacadas de acordo com os correspondentes valores absolutos dos elementos do vetor informativo obtidos anteriormente na etapa anterior. A partir deste vetor, as variáveis independentes originais (colunas da matriz  $X$ ) são decrescentemente ordenadas de acordo com a ordem de seus valores correspondentes no vetor informativo (Figura 5 – Etapa 3). Em seguida, as variáveis ordenadas são avaliadas usando uma estratégia de validação cruzada com um método de calibração multivariada. Assim, os modelos são construídos e avaliados sobre uma janela de variáveis e posteriormente sobre esta janela acrescida de

um incremento fixo de variáveis até que todas as variáveis ou uma porcentagem do total sejam adicionadas a janela (Figura 5 – Etapa 4). Por último (Figura 5 – Etapa 5), os conjuntos de variáveis avaliados são comparados utilizando os parâmetros de qualidade calculados e aquele conjunto que apresentar os melhores parâmetros é selecionado para construção do modelo (TEÓFILO; MARTINS; FERREIRA, 2009).

Na seleção de variáveis, ao se usar o vetor informativo atenção especial deve ser dada ao número de variáveis latentes,  $h$ , a ser utilizado na obtenção desse vetor. Primeiramente, deve-se determinar o número de variáveis,  $h = h_{\text{Mod}}$ , para construir e validar o modelo. Mas, geralmente,  $h_{\text{Mod}}$  não gera um vetor de regressão suficientemente informativo para seleção de variáveis. Para encontrar o melhor vetor de regressão para seleção de variáveis, um estudo deve ser realizado no conjunto completo de dados, aumentando o número de  $h$  do modelo. Então, a partir de  $h_{\text{Mod}}$  é realizada a seleção de variáveis usando o algoritmo OPS até um número de variáveis latentes ótimo para construir um vetor de regressão exclusivo para seleção de variáveis. Este número de variáveis latentes é definido como  $h_{\text{OPS}}$ , que é o número de variáveis latentes usado na construção do vetor de regressão exclusivo para seleção de variáveis no algoritmo OPS (MARTINS; FERREIRA, 2013). Assim, dois valores ótimos de  $h$  são empregados, um representando o número de  $h$  para construção do modelo ( $h_{\text{Mod}}$ ) e outro representando o número de  $h$  empregado para gerar o melhor vetor de regressão no método OPS ( $h_{\text{OPS}}$ ) (TEÓFILO, 2007).

#### **2.4.5 Figuras de Mérito em Calibração Multivariada**

A fim de assegurar a aplicabilidade e o alcance de um modelo de calibração multivariada durante as operações de rotina de um laboratório, os limites da qualidade das medidas instrumentais e a confiabilidade estatística dos cálculos envolvidos no seu processamento devem ser estabelecidos por meio da estimativa das figuras de mérito. Em definição, figuras de mérito são indicadores quantitativos da qualidade e desempenho de novos métodos analíticos e tem se tornado muito importantes a fim de caracterizar, comparar e desenvolver métodos analíticos multivariados (EURACHEM, 2002; RIBEIRO et al., 2008).

Dentre as principais figuras de mérito estão a sensibilidade, seletividade, limite de detecção (LOD) e limite de quantificação (LOQ) que estão relacionadas ao conceito de sinal analítico líquido (NAS). Este conceito exerce uma importante função na determinação de figuras de mérito para calibração multivariada e é definido como sendo

parte do sinal analítico que é ortogonal às contribuições de possíveis interferentes presentes na amostra (FABER, 1998; TEÓFILO, 2007).

A sensibilidade (SEN) é definida como a fração do sinal responsável pelo acréscimo de uma unidade de concentração à propriedade de interesse, ou seja, refere-se à medida de quão fortemente o analito de interesse responde para uma medida específica. Em calibração multivariada o modelo de regressão é inverso e, portanto, a sensibilidade (Equação 9) é o inverso dos coeficientes de regressão estimados pelo PLS (CURRIE, 1999b).

$$SEN = \frac{1}{\|\mathbf{b}\|} \quad (9)$$

em que  $\|\mathbf{b}\|$  é a norma Euclidiana do vetor dos coeficientes de regressão estimados pelo PLS.

A sensibilidade analítica (Equação 10),  $\gamma$ , é a sensibilidade do método em termos da unidade de concentração que é utilizada, sendo como a razão entre a sensibilidade e o desvio padrão do sinal de referência ( $\partial_x$ ) estimado através do desvio padrão do valor de NAS para espectros do sinal de referência (MUÑOZ DE LA PENA et al., 2002; RODRÍGUEZ et al., 1993).

$$\gamma = \frac{SEN}{\|\partial_x\|} \quad (10)$$

O inverso desse parâmetro, ou seja,  $\gamma^{-1}$ , permite estabelecer a menor diferença de concentração entre amostras, que pode ser distinguida pelo método.

Seletividade (SEL) é a capacidade de avaliar, de forma inequívoca, o analito de interesse na presença de componentes que podem interferir com a sua determinação em uma amostra complexa. Este parâmetro pode variar entre 0 e 1, em que 0 significa que o sinal analítico contém informações dos interferentes e um valor igual a 1, indica que não há interferentes. Em calibração multivariada a seletividade é definida pela razão entre o sinal analítico líquido e a resposta instrumental para determinada amostra. A seletividade (Equação 11) é dependente da amostra e como resultado final, é obtida a média e espera-se que esse valor represente bem o valor da seletividade do modelo (LORBER; FABER; KOWALSKI, 1997; RIDDER; BROWN; STEEG, 2005).

$$SEL = \frac{nas_{k,i}}{\|\mathbf{x}_{k,i}\|} \quad (11)$$

em que,  $nas_{k,i}$  é o valor escalar do sinal analítico líquido para a amostra  $i$  e  $\mathbf{x}_{k,i}$  representa o vetor de respostas instrumental para a amostra  $i$ .

O limite de detecção (LOD) e o limite de quantificação (LOQ) expressam as menores quantidades do analito de interesse que podem ser detectados e determinados quantitativamente, respectivamente (CURRIE, 1999a, 2004). Para um conjunto de dados que apresenta variância constante ao longo da faixa de trabalho e que seguem uma distribuição normal, os LOD e LOQ para calibração multivariada podem ser calculados pelas Equações 12 e 13, respectivamente.

$$\text{LOD} = 3 \cdot \|\delta r\| \cdot \|\mathbf{b}\| \quad (12)$$

$$\text{LOQ} = 10 \|\delta r\| \cdot \|\mathbf{b}\| \quad (13)$$

em que  $\|\delta r\|$  é a norma Euclidiana do vetor de desvios padrões das colunas da matriz de ruídos e  $\|\mathbf{b}\|$  é a norma Euclidiana do vetor de regressão (BOOKSH; KOWALSKI, 1994; CURRIE, 1999b).

O cálculo do LOD em calibração multivariada é o mais questionado e ainda não é muito bem definido. Assim, vários trabalhos têm avançado propondo cálculos mais precisos deste parâmetro (BOQUÉ; LARRECHI; RIUS, 1999; FABER; BRO, 2002; ORTIZ et al., 2003). Neste trabalho será utilizado o cálculo do limite de detecção usando as hipóteses de falso positivo e falso negativo, conhecido como método de Ortiz (ORTIZ et al., 2003).

Este método se baseia na extrapolação da equação de regressão obtida com os valores previstos e medidos da calibração multivariada para o valor zero de concentração. Assim, estimando a concentração a partir do sinal, obtemos a expressão do LOD (ORTIZ et al., 2003; SARABIA; ORTIZ, 1994).

$$\text{LOD} = \frac{\Delta(\alpha, \beta) w_{y_0} \sigma}{\hat{b}} \quad (14)$$

em que  $\sigma$  é o desvio padrão residual da regressão,  $\hat{b}$  é o coeficiente da regressão univariada,  $w_{y_0}$  é uma função das posições dos padrões. De maneira mais direta o produto  $w_{y_0} \sigma$  é uma estimativa do erro no intercepto. O fator  $\Delta(\alpha, \beta)$  depende da probabilidade  $\alpha$  e  $\beta$  e do número de graus de liberdade.

### **3 MATERIAIS E MÉTODOS**

#### **3.1 Coleta e Identificação de Amostras**

As amostras utilizadas neste trabalho são pertencentes ao Banco Ativo de Germoplasma (GAB) da Universidade Federal de Viçosa (UFV), constituído de 77 acessos de *J. curcas*, sendo 74 deles oriundos de diferentes regiões do Brasil e outros três do Camboja. Além destas, foram utilizadas amostras de um teste de progênies (PT) implantado em 2008 no Campo Experimental da UFV de Araponga, MG. A partir deste teste foram obtidas 121 progênies de polinização livre, assumidas meias-irmãs, de *J. curcas*. Essas amostras, obtidas em sementes, foram devidamente identificadas e armazenadas à temperatura ambiente até o momento das análises.

Para realização das análises, foram selecionadas 100 amostras do conjunto total de amostras (GAB + PT) e a identificação das amostras selecionadas está disponível nas Tabela 3 e 4 para GAB e PT, respectivamente.

**Tabela 3.** Identificação dos acessos pertencentes ao Banco Ativo de Germoplasma (GAB).

| <b>Amostra<br/>selecionada</b> | <b>Acesso</b> | <b>Observação</b> | <b>Amostra<br/>selecionada</b> | <b>Acesso</b> | <b>Observação</b> |
|--------------------------------|---------------|-------------------|--------------------------------|---------------|-------------------|
| <b>1</b>                       | T1            | GAB - I           | <b>35</b>                      | 32            | GAB - II          |
| <b>3</b>                       | 1             | GAB - I           | <b>36</b>                      | 33            | GAB - II          |
| <b>4</b>                       | 2             | GAB - I           | <b>39</b>                      | 36            | GAB - II          |
| <b>5</b>                       | 3             | GAB - I           | <b>41</b>                      | 38            | GAB - II          |
| <b>7</b>                       | 5             | GAB - I           | <b>42</b>                      | 39            | GAB - II          |
| <b>8</b>                       | 6             | GAB - I           | <b>43</b>                      | 40            | GAB - II          |
| <b>9</b>                       | 7             | GAB - I           | <b>44</b>                      | 41            | GAB - II          |
| <b>10</b>                      | 8             | GAB - I           | <b>46</b>                      | T2            | GAB - III         |
| <b>11</b>                      | 9             | GAB - I           | <b>48</b>                      | 43            | GAB - III         |
| <b>12</b>                      | 10            | GAB - I           | <b>50</b>                      | 45            | GAB - III         |
| <b>14</b>                      | 12            | GAB - I           | <b>51</b>                      | 46            | GAB - III         |
| <b>15</b>                      | 13            | GAB - I           | <b>52</b>                      | 47            | GAB - III         |
| <b>17</b>                      | 15            | GAB - I           | <b>53</b>                      | 48            | GAB - III         |
| <b>18</b>                      | 16            | GAB - I           | <b>56</b>                      | 51            | GAB - III         |
| <b>19</b>                      | 17            | GAB - I           | <b>57</b>                      | 52            | GAB - III         |
| <b>20</b>                      | 18            | GAB - I           | <b>60</b>                      | 55            | GAB - III         |
| <b>21</b>                      | 19            | GAB - I           | <b>61</b>                      | 56            | GAB - III         |
| <b>22</b>                      | 20            | GAB - I           | <b>63</b>                      | 58            | GAB - III         |
| <b>23</b>                      | T1            | GAB - II          | <b>69</b>                      | 62            | GAB - IV          |
| <b>26</b>                      | 22            | GAB - II          | <b>70</b>                      | 63            | GAB - IV          |
| <b>32</b>                      | 29            | GAB - II          | <b>85</b>                      | 83            | GAB - IV          |
| <b>33</b>                      | 30            | GAB - II          | <b>86</b>                      | 84            | GAB - IV          |
| <b>34</b>                      | 31            | GAB - II          |                                |               |                   |

**Tabela 4.** Identificação das progênie pertencentes ao teste de progênie (PT).

| <b>Amostra<br/>selecionada</b> | <b>Progênie</b> | <b>Amostra<br/>selecionada</b> | <b>Progênie</b> | <b>Amostra<br/>selecionada</b> | <b>Progênie</b> |
|--------------------------------|-----------------|--------------------------------|-----------------|--------------------------------|-----------------|
| <b>88</b>                      | 1               | <b>138</b>                     | 65              | <b>170</b>                     | 106             |
| <b>91</b>                      | 5               | <b>142</b>                     | 69              | <b>171</b>                     | 107             |
| <b>99</b>                      | 14              | <b>144</b>                     | 71              | <b>174</b>                     | 111             |
| <b>100</b>                     | 15              | <b>146</b>                     | 73              | <b>175</b>                     | 112             |
| <b>101</b>                     | 16              | <b>148</b>                     | 76              | <b>179</b>                     | 118             |
| <b>102</b>                     | 18              | <b>149</b>                     | 77              | <b>186</b>                     | 128             |
| <b>103</b>                     | 19              | <b>150</b>                     | 79              | <b>191</b>                     | 133             |
| <b>104</b>                     | 20              | <b>152</b>                     | 82              | <b>192</b>                     | 134             |
| <b>105</b>                     | 22              | <b>155</b>                     | 86              | <b>193</b>                     | 135             |
| <b>109</b>                     | 28              | <b>156</b>                     | 88              | <b>195</b>                     | 137             |
| <b>110</b>                     | 29              | <b>158</b>                     | 90              | <b>196</b>                     | 138             |
| <b>111</b>                     | 30              | <b>160</b>                     | 92              | <b>197</b>                     | 140             |
| <b>112</b>                     | 31              | <b>161</b>                     | 94              | <b>198</b>                     | 141             |
| <b>123</b>                     | 46              | <b>164</b>                     | 100             | <b>201</b>                     | 144             |
| <b>125</b>                     | 48              | <b>165</b>                     | 101             | <b>202</b>                     | 145             |
| <b>127</b>                     | 51              | <b>166</b>                     | 102             | <b>203</b>                     | 146             |
| <b>128</b>                     | 52              | <b>167</b>                     | 103             | <b>204</b>                     | 147             |
| <b>129</b>                     | 53              | <b>169</b>                     | 105             | <b>206</b>                     | 150             |
| <b>130</b>                     | 55              |                                |                 |                                |                 |

### **3.2 Preparo das Amostras**

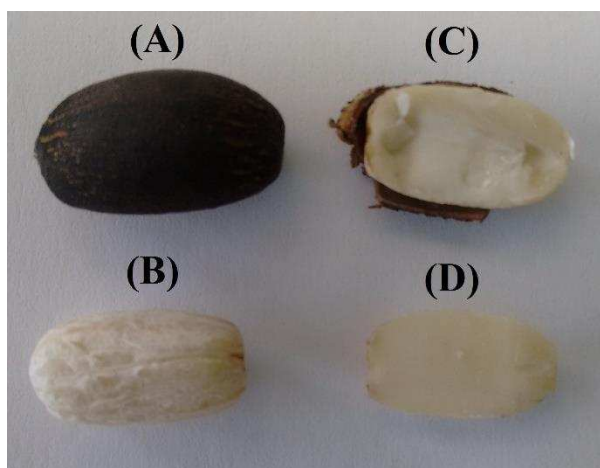
As análises espectrométricas foram realizadas nas partes externa e interna da semente, sem qualquer preparo; no óleo extraído de um conjunto de sementes e no extrato obtido do óleo.

Os espectros NIR da parte externa da semente, i.e., semente com a casca, foram obtidos em duas regiões, um espectro de cada lado da semente, sem qualquer preparo.

Os espectros NIR e UV da parte interna da semente, i.e., semente descascada, foram obtidos a partir de um corte da semente ao meio em seu eixo longitudinal resultando em uma superfície lisa e plana a qual foi colocada diretamente no local de obtenção dos espectros. As análises espectrométricas foram realizadas em três regiões diferentes da semente cortada: nas extremidades de seu eixo longitudinal e na região central.

Tanto a casca da semente quanto a parte interna da semente foram analisadas por reflectância difusa usando um acessório de esfera de integração dos instrumentos NIR e UV.

Na Figura 6 são apresentadas as semente com e sem a casca, com o seu corte longitudinal e sua a superfície lisa e plana.



**Figura 6.** Semente de pinhão manso com a casca **(A)**, semente sem a casca **(B)**, semente cortada ao meio em seu eixo longitudinal **(C)** e semente com a superfície lisa e plana **(D)**.

Para obtenção dos espectros do óleo e do extrato do óleo, extrações foram executadas e os procedimentos são descritos a seguir.

### **3.2.1 Extração do Óleo das Sementes**

O óleo de um conjunto de sementes com casca de um mesmo acesso/progênie foi extraído por prensagem manual com auxílio de uma prensa mecânica (Ribeiro®). Extraíu-se aproximadamente 20 mL de óleo, os quais foram colocados em frascos de vidro âmbar e armazenados sob refrigeração ( $-20\text{ }^{\circ}\text{C}$ ) até sua utilização.

Antes das análises espectrométricas foi necessário esperar o o óleo atingir a temperatura ambiente.

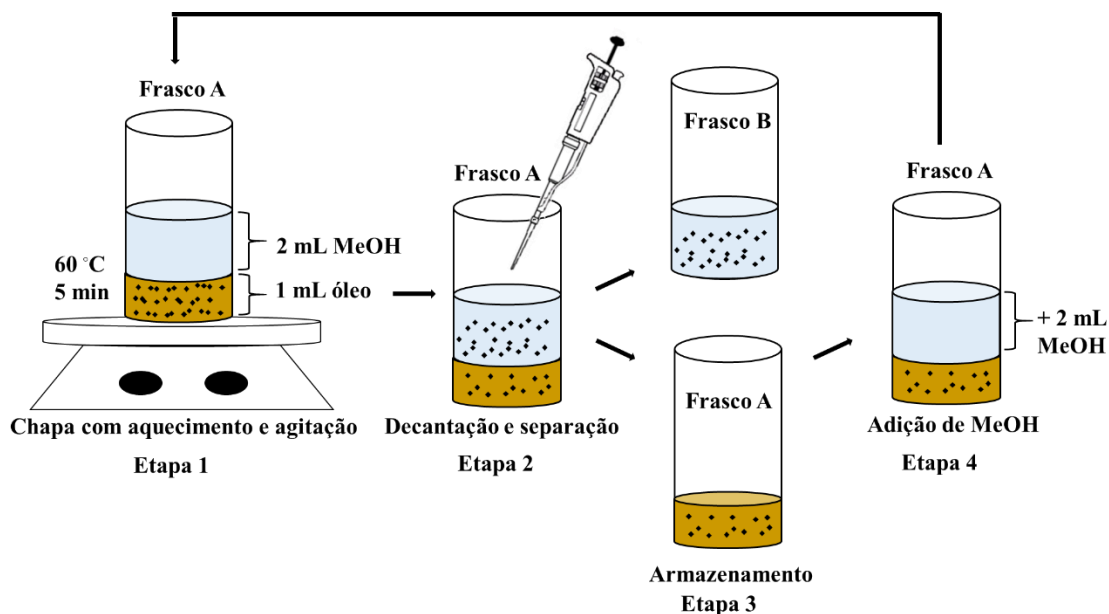
Os espectros NIR do óleo foram obtidos em duplicata em uma alíquota de 50  $\mu\text{L}$  colocada diretamente sobre o compartimento de amostra do instrumento. Os espectros foram obtidos por transflectância, empregando uma moeda de 5 centavos de reais com superfície refletora colocada sobre o óleo.

Para obtenção dos espectros UV/Vis do óleo, foi necessário diluí-lo em hexano (Cromato Produtos Químicos LTDA, PA) a fim de respeitar a lei de Beer. Uma alíquota de 300  $\mu\text{L}$  do óleo foi diluído para um volume final de 3 mL de hexano. Os espectros de absorbância na região de 250 a 330 nm foram obtidos dessa solução em duplicata usando uma cubeta de quartzo de 10 mm de caminho óptico.

### **3.2.2 Extração dos Ésteres de Forbol do Óleo**

A extração dos ésteres de forbol foi realizada por meio de adaptações da metodologia otimizada encontrada na literatura (DEVAPPA; MAKAR; BECKER,

2010a). A Figura 7 mostra o esquema da extração dos ésteres de forbol do óleo de J. curcas.



**Figura 7.** Esquema do procedimento de extração dos PEs (♦) em óleo de J. curcas.

Os PEs foram extraídos do óleo de J. curcas com metanol (J.T.Baker; 99,95%; grau HPLC). Em um frasco de vidro (Figura 7 – Etapa 1) foram adicionados 1,0 mL de óleo e 2,0 mL de metanol (MeOH). Este recipiente foi deixado em uma chapa de aquecimento com agitação magnética à 60 °C durante 5 min. Após este tempo, sob aquecimento e agitação, a mistura foi deixada em repouso até a separação das fases pelo processo de decantação (Figura 7 – Etapa 2). Em seguida, a fase superior metanólica, contendo os PEs de interesse, foi recolhida em outro frasco (Figura 7 – Etapa 3). Outra alíquota de 2,0 mL de metanol foi adicionada ao óleo já extraído e o processo descrito foi repetido (Figura 7 – Etapa 4) por mais três vezes a fim de extrair quantitativamente os PEs. Por fim, os extratos metanólicos coletados no frasco B (Figura 7) foram evaporados até resíduo e armazenado sob refrigeração à -20 °C.

Para quantificação dos ésteres de forbol, o resíduo do frasco B (Figura 7) foi redissolvido em 4,0 mL de metanol (solução A ou solução de trabalho) e em seguida a solução foi filtrada em vial, o qual foi levado para separação e análise dos PEs empregando HPLC com detecção UV no comprimento de onda de 280 nm.

Para obtenção dos espectros UV dos extratos, foi realizada uma diluição da solução A a fim de respeitar a lei de Lambert Beer. Uma alíquota de 600 µL dessa solução foi diluída em metanol para um volume final de 2 mL. Os espectros de absorbância na região

de 210 a 350 nm foram obtidos dessa solução em duplicata usando uma cubeta de quartzo de 10 mm de caminho óptico.

### 3.3 Análise Espectral das Amostras

Na Tabela 5 estão relacionados os tipos de matrizes analisadas por espectroscopia e a identificação da matriz **X** (variáveis independentes) obtida para as análises quimiométricas.

**Tabela 5.** Identificação das variáveis independentes (matriz **X**) obtidas para cada tipo de material das amostras de J. curcas, faixa analisada e tipo de análise no NIR, UV e UV-VIS-NIR.

| NIR               |               |                        |                 |
|-------------------|---------------|------------------------|-----------------|
| Amostras          | Identificação | Faixa analisada        | Tipo de análise |
| Semente Sem Casca | $X_{NIRsc}$   | 4000 – 10000 $cm^{-1}$ | Reflectância    |
| Semente Com Casca | $X_{NIRcc}$   | 4000 – 10000 $cm^{-1}$ | Reflectância    |
| Óleo              | $X_{NIRo}$    | 4000 – 10000 $cm^{-1}$ | Transflectância |
| UV                |               |                        |                 |
| Amostras          | Identificação | Faixa analisada        | Tipo de análise |
| Óleo              | $X_{UVo}$     | 250 – 330 nm           | Absorbância     |
| Extrato do óleo   | $X_{UVe}$     | 210 – 350 nm           | Absorbância     |
| UV-VIS-NIR        |               |                        |                 |
| Amostras          | Identificação | Faixa analisada        | Tipo de análise |
| Semente Sem Casca | $X_{UVsc}$    | 250 – 1000 nm          | Absorbância     |

As análises NIR da semente (com e sem casca) e a análise direta do óleo foram executadas em um espectrômetro Thermo Scientific Antaris II com transformada de Fourier e acessório para refletância difusa ou transflectância com esfera integradora. A faixa investigada foi de 10000 a 4000  $cm^{-1}$  com um incremento de 8  $cm^{-1}$ . Os espectros foram coletados usando o software TQ Analysis. Para cada amostra, um total de 32 varreduras foram realizadas e a média foi armazenada. Os espectros das sementes ( $X_{NIRsc}$  e  $X_{NIRcc}$ ) foram obtidos no modo refletância como  $\log(1/R)$ , em que R é a refletância coletada. Os espectros do óleo ( $X_{NIRo}$ ) foram obtidos no modo transflectância e a absorbância foi coletada ( $-\log(1/T)$  em que T é a transflectância).

Os espectros UV das matrizes líquidas i.e., óleo ( $X_{UVo}$ ) e extratos do óleo ( $X_{UVe}$ ), foram obtidos no modo de transmitância usando cubeta de quartzo de caminho óptico de 10 mm em um espectrógrafo Ocean Optics (USB4000-UV-VIS). A região espectral investigada foi de 210 a 350 nm, com incremento de 0,166 nm. Os espectros foram obtidos pelo software Spectra Suite Spectroscopy Operating, onde foram obtidos os valores de absorbância. Foram realizados 10 scans de cada amostra e a média foi armazenada.

Para obtenção dos espectros UV-VIS-NIR das sementes sem casca ( $X_{UVsc}$ ), foi utilizado um espectrofotômetro Evolution Array UV-Visible Spectrophotometer Thermo Scientific com acessório para refletância difusa com esfera de integração. A região espectral investigada foi de 186,9 a 1103,6 nm, com incremento de 0,895 nm. O software VISIONcollect foi utilizado para obter os espectros, em que a absorbância foi coletada. Para cada amostra, um total de 10 scans foram obtidos e a média foi armazenada.

### **3.4 Obtenção das Variáveis Dependentes (Ésteres de Forbol )**

#### **3.4.1 Separação Cromatográfica**

O sistema cromatográfico Shimadzu modelo 20AT prominence gerenciado pelo software LabSolutions foi usado nas análises de separação e detecção de PEs.

Para a separação cromatográfica foi utilizada uma coluna C18 (Kinetex®, Fenomenex), com 100 mm de comprimento; 4,6 mm de diâmetro e partícula de 2,6  $\mu\text{m}$ . O fluxo utilizado foi de 1  $\text{mL min}^{-1}$  e a temperatura da coluna de 30 °C. A composição da fase móvel empregada na separação foi aquela encontrada na otimização pelo CCD.

A solução de trabalho foi filtrada (filtro de nylon com poro de 0,45  $\mu\text{m}$  e diâmetro de 25 mm) para vials âmbar antes de ser levado para a análise cromatográfica.

A separação foi monitorada empregando o detector UV/Vis de arranjo de diodos (DAD) Shimadzu modelo SPD-M20A prominence. A faixa investigada foi de 190 a 380 nm, no entanto o comprimento de onda de 280 nm foi fixado para monitoramento dos PEs.

#### **3.4.2 Planejamento Experimental**

Para otimizar a separação cromatográfica, foi realizado um CCD com o objetivo de avaliar a composição da fase móvel, i.e., as porcentagens de acetonitrila (J.T.Baker; 99,95%; grau HPLC) e de ácido fórmico (Panreac;  $\geq 98,0\%$ ) a serem utilizadas na separação. Os níveis das variáveis codificados e decodificados (entre parênteses) estão descritos na Tabela 6.

**Tabela 6.** Planejamento composto central para otimização da composição da fase móvel na separação cromatográfica: porcentagem de acetonitrila (% ACN) e porcentagem de ácido fórmico (% HCOOH).

| Experimento | % ACN            | % HCOOH          |
|-------------|------------------|------------------|
| 1           | -1 (80,0)        | -1 (0,50)        |
| 2           | 1 (85,0)         | -1 (0,50)        |
| 3           | -1 (80,0)        | 1 (0,70)         |
| 4           | 1 (85,0)         | 1 (0,70)         |
| 5           | $-\alpha$ (79,0) | 0 (0,60)         |
| 6           | $\alpha$ (86,0)  | 0 (0,60)         |
| 7           | 0 (82,5)         | $-\alpha$ (0,45) |
| 8           | 0 (82,5)         | $\alpha$ (0,75)  |
| 9           | 0 (82,5)         | 0 (0,60)         |
| 10          | 0 (82,5)         | 0 (0,60)         |
| 11          | 0 (82,5)         | 0 (0,60)         |
| 12          | 0 (82,5)         | 0 (0,60)         |
| 13          | 0 (82,5)         | 0 (0,60)         |

Como resposta do planejamento, foi utilizada uma combinação de parâmetros cromatográficos a fim de obter a resposta desejada (RD) como apresentado na Equação 15:

$$RD = \frac{\text{trF}}{np} \quad (15)$$

em que trF é o tempo de retenção do último pico e np é o número de picos. Deseja-se encontrar o menor trF, pois isto corresponde a um menor tempo de corrida e o np igual a 5 é considerado ótimo, uma vez que indica a separação dos cinco ésteres de forbol presentes em J. curcas. Assim, o RD considerado ótimo, obtido pela equação, é o menor possível.

### 3.4.3 Curva Analítica

A fim de quantificar os ésteres de forbol foi construída uma curva analítica com o padrão de forbol-12-miristato-13-acetato (PMA). Sete níveis de concentração entre 0,05 e 2,80 mg mL<sup>-1</sup> foram utilizados para a construção da regressão linear univariada. As soluções foram levadas para separação e análise no cromatógrafo descrito no item 3.4.1. para obtenção da área cromatográfica de cada nível de concentração.

A curva analítica foi obtida por regressão inversa, em que o modelo é construído com os valores dos níveis das concentrações no eixo das ordenadas (eixo y) e os respectivos valores de área medidos no eixo das abscissas (eixo x).

### 3.4.4 Quantificação e Identificação

A composição da fase móvel empregada na separação foi aquela encontrada na otimização pelo método de superfície de respostas empregando o CCD.

O teor dos PEs foi encontrado através de predições da área encontrada via cromatografia no modelo de regressão construído (curva analítica). Os resultados foram expressos como equivalentes de PMA ( $\text{mg mL}^{-1}$ ).

O teor de óleo das sementes foi obtido utilizando a técnica de ressonância magnética nuclear (RMN MQC, Oxford Instruments) expresso em gramas de óleo por 100 gramas de semente ( $\text{g } 100 \text{ g}^{-1}$ ).

A partir do teor de óleo foi possível estabelecer uma relação matemática e obter a concentração dos ésteres de forbol em  $\text{mg g}^{-1}$  de semente, conforme Equação 16:

$$C_f = \frac{C_i \cdot \text{FD} \cdot \text{TO}}{d} \quad (16)$$

em que  $C_f$  é a concentração em  $\text{mg.g}^{-1}$  de semente;  $C_i$ , é a concentração de PEs em equivalente de PMA ( $\text{mg mL}^{-1}$ ), obtida empregando a curva analítica construída; FD é o fator de diluição dos extratos; TO é o teor de óleo ( $\text{g } 100 \text{ g}^{-1}$ ) para cada amostra e  $d$  é a densidade do óleo de *J. curcas*.

Para identificação dos picos cromatográficos, foram comparados os espectros UV dos picos cromatográficos com os correspondentes encontrados na literatura (BALDINI et al., 2014; DEVAPPA; BINGHAM; KHANAL, 2013; MAKKAR et al., 2009).

### 3.4.5 Validação

A validação do método quantitativo foi realizada segundo as recomendações da literatura (ANVISA, 2003; INMETRO, 2007; IUPAC, 1978). Foram avaliados a linearidade, a precisão, o limite de detecção (LOD) e o limite de quantificação (LOQ).

A linearidade foi avaliada levando em consideração o valor do coeficiente de correlação ( $R$ ) obtido pelo ajuste linear da curva analítica construída. A precisão foi avaliada através da repetitividade, definida como o grau de concordância entre os resultados de medições sucessivas, efetuadas sob as mesmas condições de análise. Este parâmetro foi expresso através do desvio padrão relativo (RSD), calculado pela Equação 17.

$$\text{RSD} = \left( \frac{s}{\bar{x}} \right) \times 100 \quad (17)$$

em que  $s$  é o desvio padrão das áreas dos picos cromatográficos de cinco repetições e  $\bar{x}$  é a média entre elas.

O limite de detecção (LOD) é a menor concentração que pode ser detectada do analito em uma amostra, mas não necessariamente quantificada, sob determinadas condições experimentais. O limite de quantificação (LOQ) do método analítico é a menor concentração do analito que pode ser determinada com exatidão e precisão aceitáveis. Neste trabalho, esses parâmetros, foram calculados pelas Equações 18 e 19:

$$\text{LOD} = 3 \cdot \left( \frac{s_b}{b} \right) \quad (18)$$

$$\text{LOQ} = 10 \cdot \left( \frac{s_b}{b} \right) \quad (19)$$

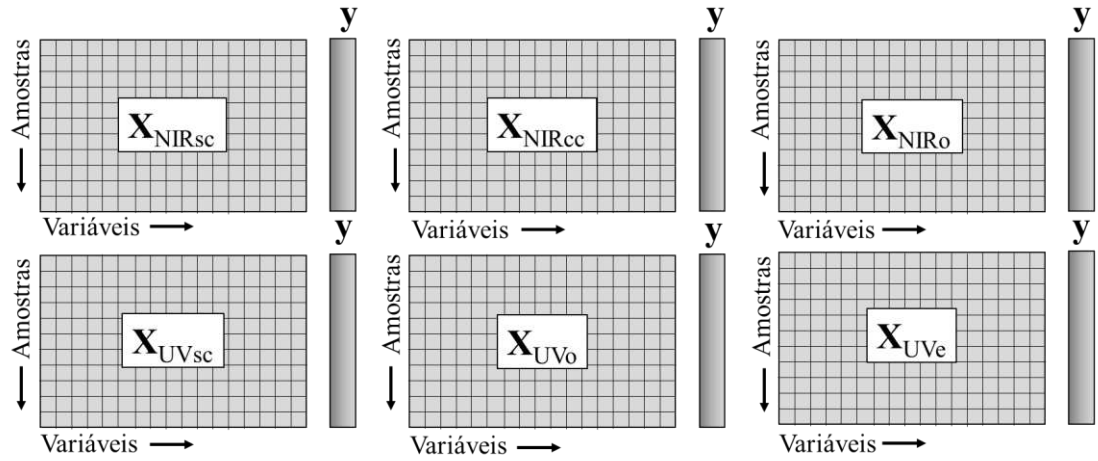
em que  $s_b$  é o valor do desvio padrão referente ao sinal cromatográfico de sete injeções do branco (MeOH) e  $b$  é o coeficiente angular da curva analítica.

### 3.5 Construção dos Modelos de Calibração Multivariada

Os espectros foram exportados dos softwares de cada instrumento e importados através de algoritmos escritos no Laboratório de Instrumentação e Quimiometria (LINQ) pelo software Matlab 7.9 (Math Works, Natick, USA). Uma matriz de dados para cada conjunto de espectros foi montada e denominada matriz **X**, que são as variáveis independentes. As linhas da matriz **X** correspondem às amostras e as colunas correspondem às variáveis (número de onda ou comprimento de onda). Um vetor contendo os respectivos valores determinados para os PEs foi construído e denominado **y**, que é a variável dependente. O vetor **y** possui um número de linhas igual ao número de amostras na matriz **X**. Todas as matrizes **X** e vetores **y** estão representados no esquema mostrado na Figura 8.

Para construção dos modelos foi utilizado a regressão PLS aliada a dois métodos de seleção de variáveis, o OPS e GA. Os algoritmos para a construção e validação dos modelos foram escritos no Laboratório de Instrumentação e Quimiometria (LINQ) em função .m para Matlab. Todos os cálculos foram realizados no Matlab 7.9. O pacote computacional OPS\_Toolbox foi usado para execução e cálculos do método OPS (TEÓFILO; MARTINS; FERREIRA, 2009).

As variáveis **X** e **y** foram pré-processadas usando a centragem na média em todos os cálculos. Transformações foram realizadas nas linhas da matriz **X** com o objetivo de encontrar o melhor modelo para predição. As transformações executadas foram: (1) primeira derivada; (2) segunda derivada e (3) (MSC).



**Figura 8.** Esquema com a matriz  $X$  para cada conjunto de espectros e o vetor  $y$  com os teores de PEs.

A qualidade dos modelos foi avaliada pela raiz quadrática média do erro (RMSE), a qual pode ser calculada de acordo com a Equação 20. Pela Equação 21 calcula-se o coeficiente de correlação ( $R$ ), em que  $\hat{y}$  e  $\hat{\mathbf{y}}$  são respectivamente os escalares e vetor de valores estimados,  $\bar{y}$  é um escalar dos valores médios de  $y$ , e  $I_m$  é o número das amostras.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{I_m} (y_i - \hat{y}_i)^2}{I_m}} \quad (20)$$

$$R = \frac{\sum_{i=1}^{I_m} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_m} (\hat{y}_i - \bar{\hat{y}})^2 (y_i - \bar{y})^2}} \quad (21)$$

Quando a validação cruzada (CV) é aplicada,  $I_m$  é o número de padrões no conjunto de calibração. Neste caso, o erro e o coeficiente de correlação são chamados RMSECV e  $R_{cv}$ , respectivamente. Para a validação externa (um novo conjunto de padrões),  $I_m$  é o número de padrões de predição ( $P$ ) e, neste caso, os coeficientes de correlação e o erro são nomeados  $R_p$  e RMSEP, respectivamente.

Para selecionar o número de variáveis latentes ( $h$ ) do modelo, foi utilizado o método de validação cruzada com remoção aleatória de três amostras. O gráfico número de componentes versus RMSECV foi utilizado para escolher o número de  $h$ .

A matriz de dados foi separada em dois conjuntos, um de calibração e outro de predição. A construção propriamente dita dos modelos foi realizada usando o conjunto de calibração. Os valores de RMSECV e o coeficiente de correlação dos valores medidos e preditos na validação cruzada ( $R_{cv}$ ) foram usados como parâmetros de verificação do

ajuste do modelo. Uma vez realizada a regressão PLS usou-se o conjunto de predição para verificar a capacidade preditiva do modelo e assim validar o modelo. Os valores de RMSEP e o coeficiente de correlação dos valores medidos e preditos na predição ( $R_p$ ), bem como os valores dos erros relativos das amostras foram usados como parâmetro para verificar a capacidade de predição do modelo construído.

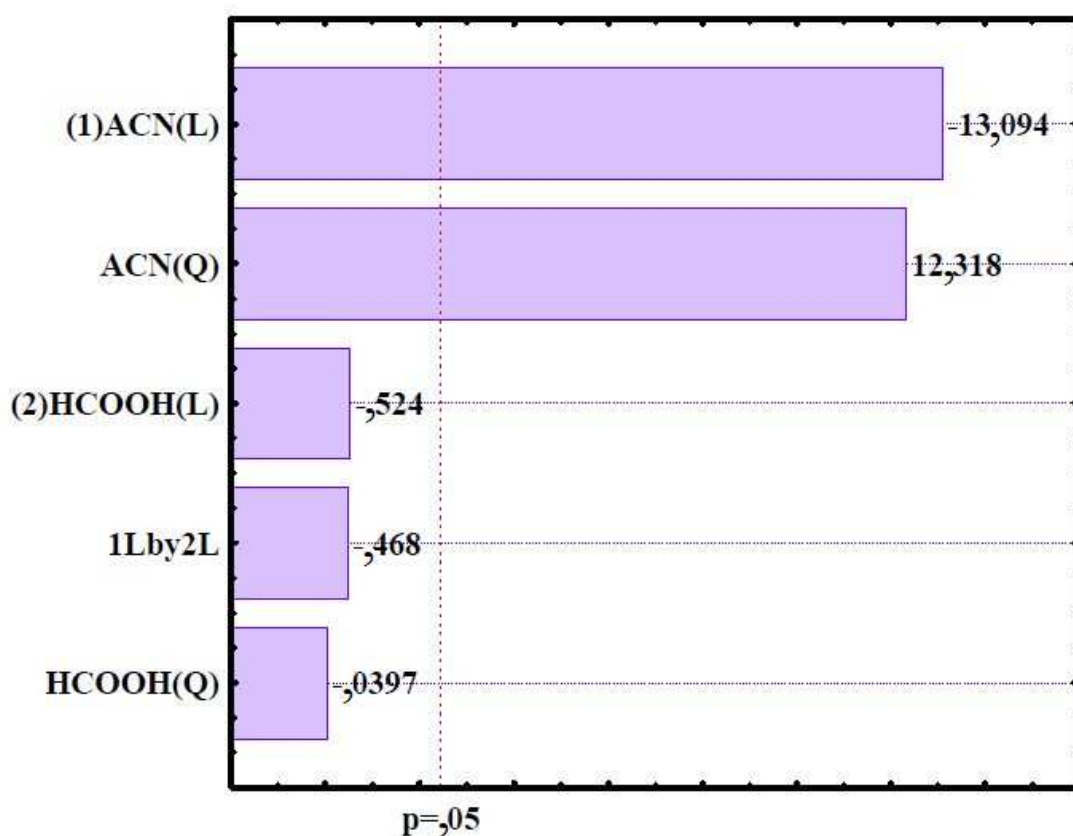
O número de amostras do conjunto de predição correspondeu a 15% do conjunto completo de amostras. A escolha das amostras para o conjunto de predição e calibração foi realizada utilizando o algoritmo Kennard-Stone (KENNARD; STONE, 1969), que seleciona as amostras com base em suas distâncias. A primeira amostra selecionada pelo algoritmo é a que apresenta a maior distância em relação à amostra média. A segunda amostra a ser selecionada será a que apresentar maior distância em relação à primeira amostra selecionada. A próxima amostra a ser selecionada apresentará maior distância em relação à última amostra selecionada, e assim sucessivamente até atingir o número de amostras desejadas.

O OPS foi executado com uma janela inicial de 10 variáveis e um incremento de 5. A seleção de variáveis pelo algoritmo genético (GA) (LEARDI; LUPIÁÑEZ GONZÁLEZ, 1998) foi realizada a fim de comparar o resultado obtido pela seleção realizada com o OPS. O GA foi empregado usando os seguintes parâmetros previamente otimizados: população = 54, gerações = 300, taxa de mutação = 0,008, largura da janela = 1, convergência = 80, inicialização = 50 e cross over = 2.

## 5 RESULTADOS E DISCUSSÃO

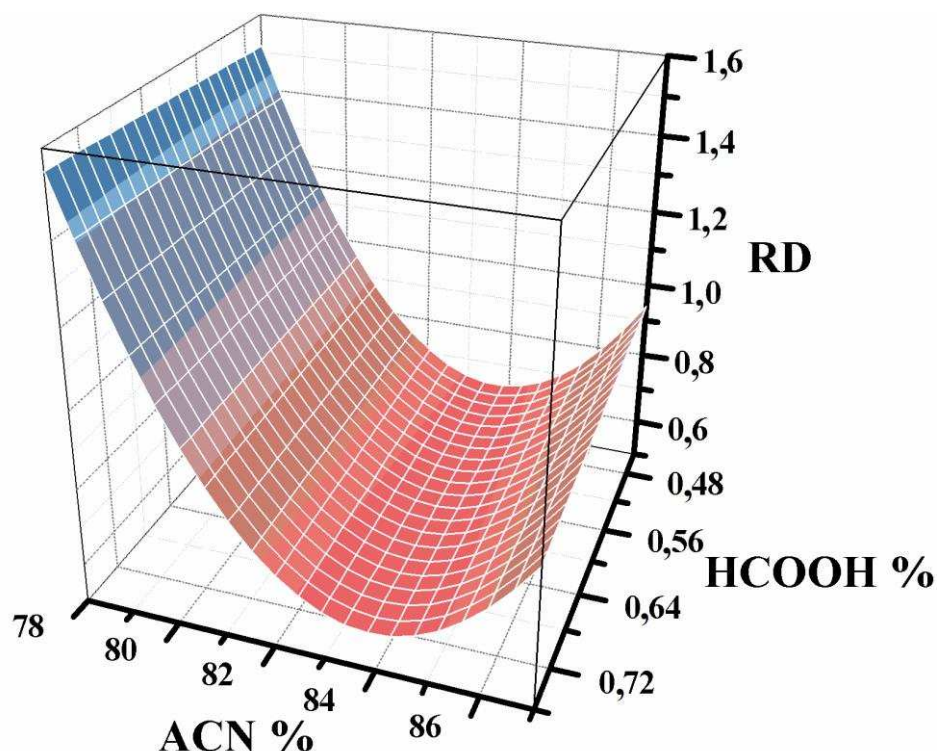
### 5.1 Planejamento Experimental

A análise estatística com o cálculo do erro pela soma quadrática dos resíduos e nível de significância ( $\alpha$ ) de 0,05 apontou apenas a %ACN como variável significativa, como pode ser visto na Figura 9. Este resultado indica que dentro dos níveis estudados o aumento da %ACN diminui o tempo de retenção do último pico e aumenta o número de picos, obtendo-se a resposta desejada, ou seja, menor RD.



**Figura 9.** Gráfico de Pareto das influências dos fatores do HPLC sobre as respostas. Erro usado: Soma quadrática dos resíduos. Nível de significância,  $\alpha$ : 0,05.

Pela análise de variância (ANOVA), a regressão foi significativa e ocorreu o ajuste do modelo com  $\alpha = 0,05$ . A partir da análise da superfície de resposta apresentada na Figura 10, pode-se perceber que a menor RD foi encontrado quando a %ACN variou de 82 a 86%. Como a variável %HCOOH não foi significativa, e a menor RD foi obtida em toda a faixa estudada, uma porcentagem intermediária desta variável (0,6%HCOOH) foi escolhida para as análises.

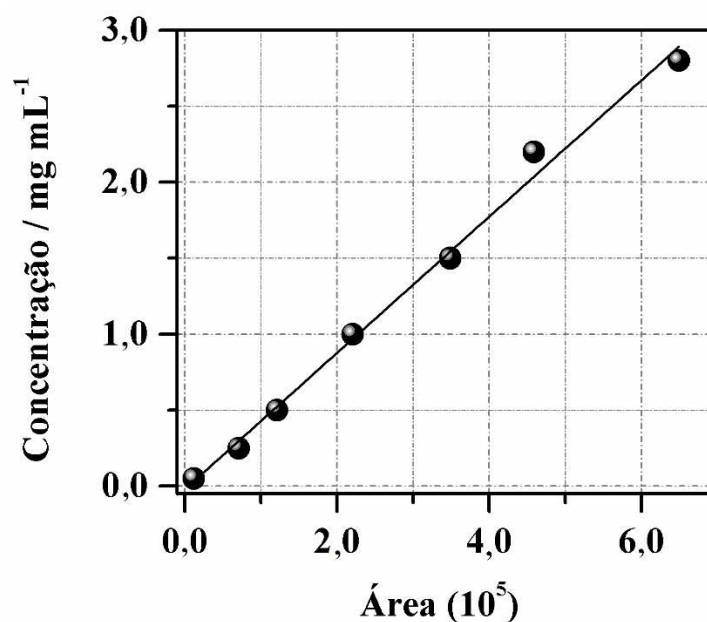


**Figura 10.** Superfície de resposta com a % ACN e % HCOOH em função da RD.

Dessa forma, as proporções ótimas de solventes para fase móvel utilizadas no experimento foram de 82,5% ACN e 0,6% HCOOH. Esta mistura forneceu uma boa separação cromatográfica das amostras de *J. curcas* em um curto tempo de corrida.

## 5.2 Quantificação e Identificação dos Ésteres de Forbol

A equação da regressão inversa obtida para quantificação dos ésteres de forbol foi  $[PE] = 4,482 \times 10^{-7} SA - 0,01897$ , em que SA é a soma das áreas dos picos cromatográficos e [PE] é a concentração de PEs. A Figura 1111 mostra a curva analítica, na qual nota-se uma relação linear satisfatória indicando excelente ajuste do modelo.



**Figura 11.** Curva analítica de PMA, para determinação de ésteres de forbol em *Jatropha curcas*.

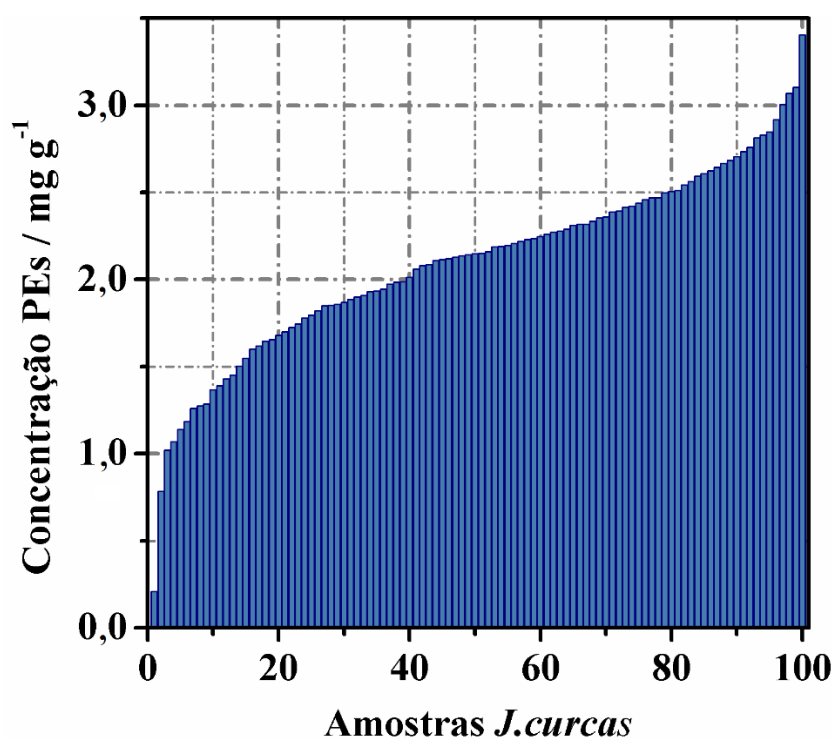
A linearidade, expressa pelo coeficiente de correlação (R), a precisão, obtida através do cálculo do desvio padrão relativo (RSD) e os limites de detecção e quantificação do método cromatográfico quantitativo estão apresentados na Tabela 7.

**Tabela 7.** Parâmetros analíticos de validação do método cromatográfico quantitativo.

| R      | RSD (%) | LOD (mg mL <sup>-1</sup> ) | LOQ (mg mL <sup>-1</sup> ) |
|--------|---------|----------------------------|----------------------------|
| 0,9968 | 11,16   | 0,0051                     | 0,0171                     |

O valor do coeficiente de correlação indica alta linearidade do modelo, dentro da faixa de concentração estudada, uma vez que valores de  $R \geq 0,99$  são aceitáveis (ANVISA, 2003). A precisão do método (RSD %), considerando a complexidade das amostras, está adequada e abaixo dos valores aceitáveis de até 20% descritos na literatura (HUBER, 1998).

O teor de óleo encontrado nas amostras de *J. curcas* variou em uma pequena faixa de 32,20 até 38,16 % (gramas de óleo por 100 gramas de semente). A partir da curva analítica foram obtidos teores de PEs na faixa de 0,133 a 2,123 mg mL<sup>-1</sup> em equivalentes de PMA. Através da Equação 16 foi obtida a concentração final ( $C_f$ ) dos ésteres de forbol em mg g<sup>-1</sup> de semente, conforme apresentado na Figura 12 e no Anexo 1 de forma mais detalhada.



**Figura 12.** Concentração dos ésteres de forbol em mg g<sup>-1</sup> na semente de *J. curcas*.

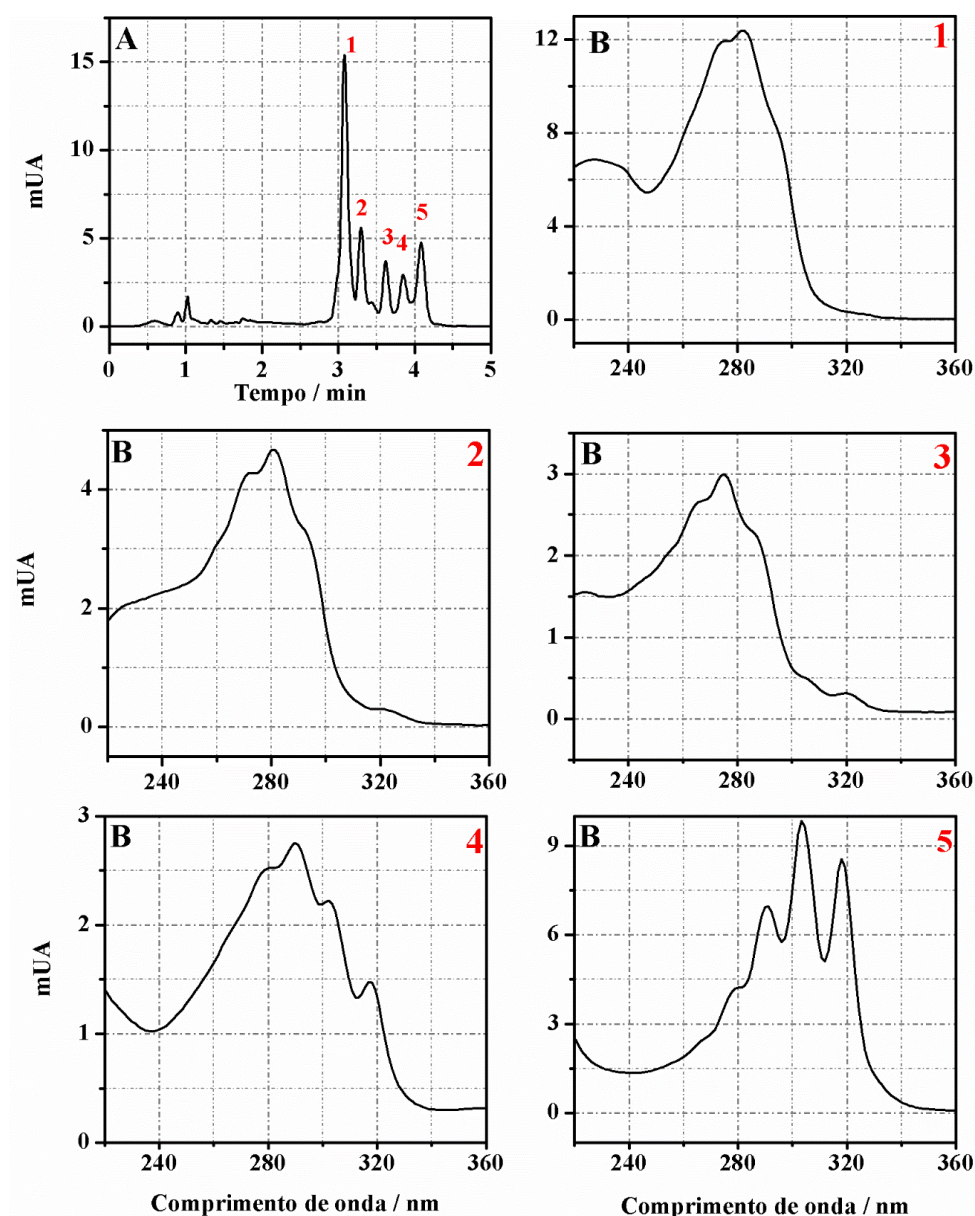
Neste trabalho, os acessos do GAB e as progênies do PT apresentaram concentrações de PEs variando de 0,206 a 3,406 mg g<sup>-1</sup>. Na literatura são relatadas diversas concentrações encontradas dos ésteres de forbol, variando da região onde as amostras foram obtidas. Na Malásia foi relatado em 2001 (OSKOUETIAN et al., 2011) a concentração de  $3 \pm 0,16$  mg g<sup>-1</sup>. Em 1997 Makkar et al. relataram concentrações entre 0,87 e 3,32 mg g<sup>-1</sup> em países da África Ocidental e Oriental, América do Norte e Central e Ásia (MAKKAR et al., 1997). Outro estudo realizado por Makkar et al. (MAKKAR; ADERIBIGBE; BECKER, 1998) avaliou quatro variedades de *Jatropha curcas* originárias de Cabo Verde, Nicarágua, Nigéria e México. Foram relatadas concentrações de 2,17; 2,30 e 2,70 mg g<sup>-1</sup> para as variedades da Nicarágua, Nigéria e Cabo Verde, respectivamente. A variedade mexicana foi descrita como não tóxica com apenas 0,11 mg g<sup>-1</sup>. Recentemente, em um estudo realizado na Alemanha, foram relatadas concentrações entre 0,00 até 10,3 mg g<sup>-1</sup> de amostras do Banco de Germoplasma (JatroSelect GmbH) contendo amostras a partir de uma ampla variedade de regiões geográficas de todo o mundo (MONTES et al., 2013). Ferrari et al. avaliaram sementes de *J. curcas* provenientes de diferentes regiões do Brasil, e a concentração de ésteres de forbol encontrada variou de 1,41 a 8,97 mg g<sup>-1</sup> (FERRARI et al., 2009).

Dessa forma, as amostras estudadas neste trabalho apresentaram concentrações de PEs dentro da faixa reportada pela literatura sobre *J. curcas* encontrado no Brasil, mas

também foram encontradas amostras com concentrações inferiores de PEs, bem próximas a cultivar encontrada no México (MAKKAR; ADERIBIGBE; BECKER, 1998).

A concentração de PEs é comumente reportada na literatura como equivalente de PMA, um éster de forbol encontrado na planta cróton (*Croton tiglium*) e que não está presente em *J. curcas*. O fator C1 compreende a fração principal (51%) dos ésteres de forbol encontrados no pinhão-mansão e dessa forma Roach et al. (2012) sugeriram a substituição do PMA pelo uso do fator C1 como padrão externo na quantificação dos PEs. Neste mesmo trabalho e em outro publicado por Devappa et al. (2013), os valores quantificados pelo novo método para diferentes amostras de *J. curcas* estavam dentro da gama de valores anteriormente relatados. Além disso, foi obtido um fator de conversão entre valores de concentração obtidos pelo método de equivalente de PMA e pelo uso do fator C1. Dessa forma, pôde-se concluir que os PEs do pinhão manso e o PMA possuem coeficiente de absorvidade similares permitindo a quantificação dos ésteres de forbol com exatidão.

A fim de identificar os ésteres de forbol, os espectros UV dos picos cromatográficos foram comparados com os correspondentes encontrados na literatura (BALDINI et al., 2014; DEVAPPA; BINGHAM; KHANAL, 2013; MAKKAR et al., 2009). O cromatograma com os cinco picos de ésteres de forbol de uma amostra de *J. curcas* e os respectivos espectros de cada um dos *Jatropha* fatores (C1 a C6) são apresentados na Figura 13.



**Figura 13.** Cromatograma obtido de uma amostra de pinhão-mansó com os cinco picos referentes aos ésteres de forbol (A). Espectros dos picos cromatográficos dos fatores de *Jatropha curcas* (B): o primeiro pico corresponde ao fator C1; o segundo pico corresponde ao fator C2; o terceiro pico corresponde ao fator C3; o quarto pico corresponde ao fator C6 e o quinto pico corresponde aos fatores C4 e C5, que são epímeros.

Os espectros apresentados na Figura 13 estão de acordo com os respectivos encontrados na literatura, confirmando a identidade dos ésteres de forbol em amostras de *J. curcas* utilizadas neste trabalho.

### 5.3 Modelos de Calibração

Nesta seção, serão descritos primeiramente os modelos obtidos para a semente sem casca e para o óleo, tanto no NIR quanto no UV-Vis-NIR. Para estes modelos, os resultados não foram satisfatórios e portanto serão apresentados somente os valores

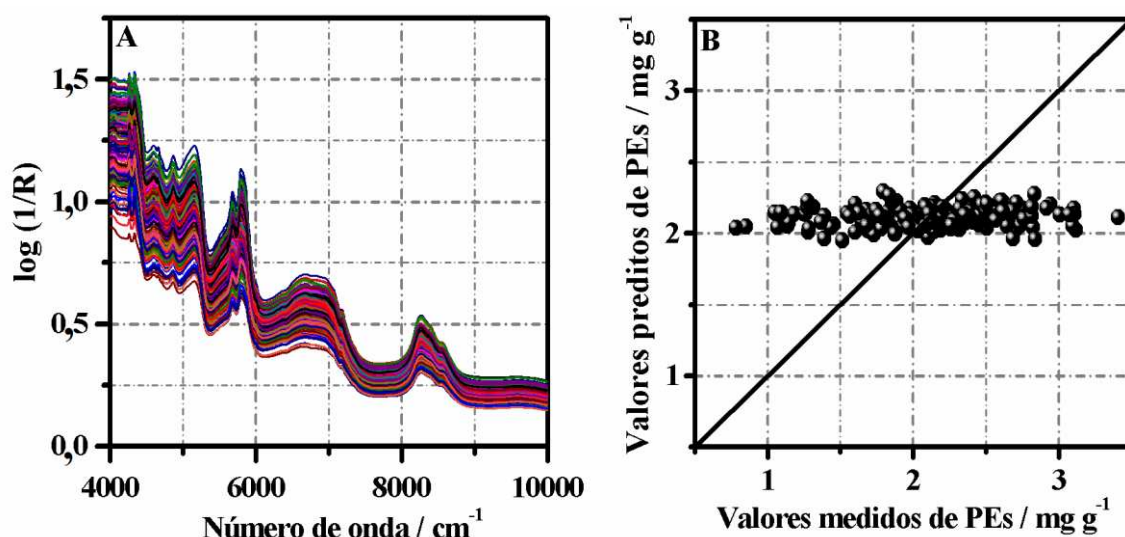
medidos e preditos obtidos pelo método dos quadrados mínimos parciais e seleção dos preditores ordenados (PLS-OPS).

Em seguida será apresentado o modelo obtido para a semente com casca no NIR, que apresentou resultados para triagem das concentrações de ésteres de forbol com capacidade de decisão rápida para ser empregado em programas de melhoramento genético.

Por último será apresentado o modelo obtido com o extrato do óleo no UV, que apresentou resultados satisfatórios e condições de prever com exatidão a concentração dos ésteres de forbol.

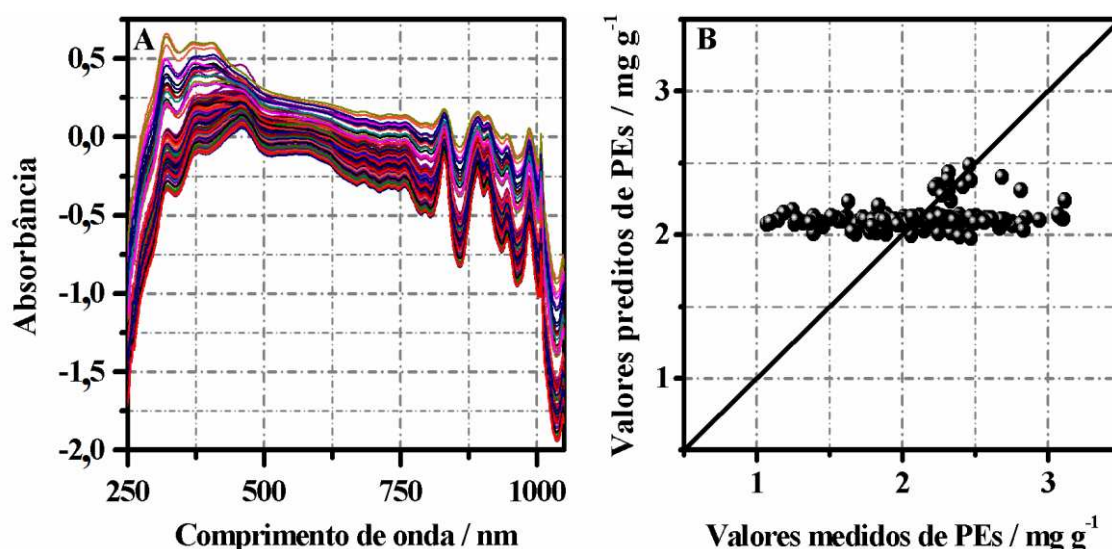
### 5.3.1 Espectros da Semente Sem Casca

Os espectros NIR das amostras de semente de pinhão-manso sem casca, na faixa de 4000 a 10000  $\text{cm}^{-1}$ , são apresentados na Figura 14 (A). Na Figura 14 (B) são apresentados os valores medidos e preditos para o conjunto de calibração.



**Figura 14. (A)** Espectros NIR das sementes de *Jatropha curcas* L. sem casca ( $X_{\text{NIRsc}}$ ) e **(B)** valores medidos e preditos dos teores de PEs para o conjunto de calibração do modelo PLS-OPS.

Os espectros UV-Vis-NIR das amostras de semente de pinhão-manso sem casca, na faixa de 250 a 1100 nm, são apresentados na Figura 15 (A). Na Figura 15 (B) são apresentados os valores medidos e preditos para o conjunto de calibração.

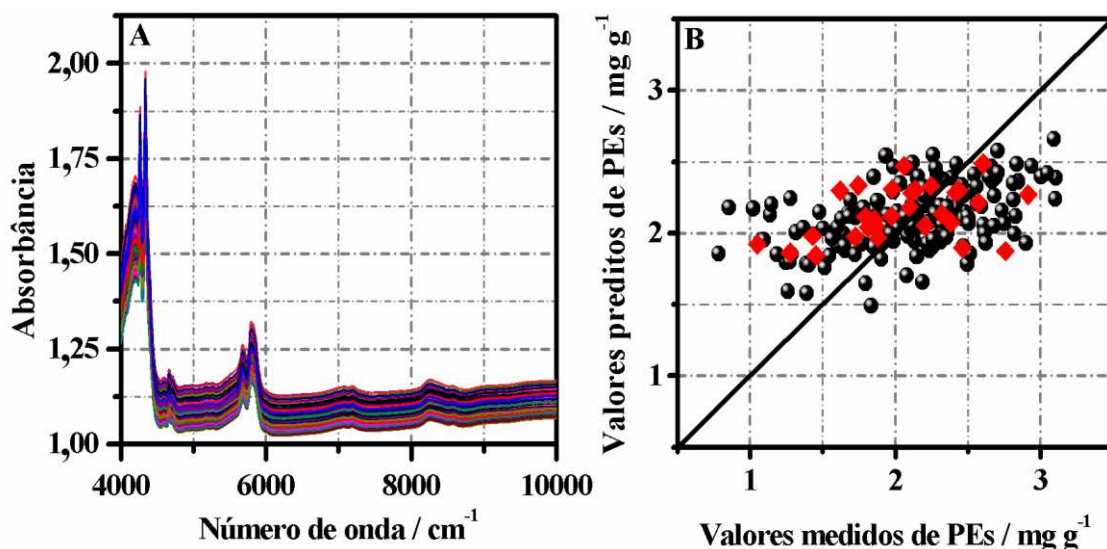


**Figura 15. (A)** Espectros UV-Vis-NIR das sementes de *Jatropha curcas* L. sem casca ( $X_{UVsc}$ ) e **(B)** valores medidos e preditos do teor de PEs para o conjunto de calibração do modelo PLS-OPS.

Pelo gráfico de valores medidos versus preditos, tanto para o NIR quanto para o UV-Vis-NIR, pode-se perceber que os modelos obtidos apresentam altos erros relativos indicando que o modelo não está adequado para ser utilizado em predições dos ésteres de forbol. Assim, não foi possível obter um modelo utilizando as espectroscopias NIR e UV-Vis-NIR para prever a partir da semente sem casca os teores de PEs. Provavelmente, isso ocorreu pela grande variedade e quantidade de interferentes presentes nas amostras de semente de pinhão-manso sem casca, impossibilitando a modelagem normalmente realizada na presença de alguns interferentes. Além disso, a concentração dos ésteres de forbol está na faixa de 0,02 a 0,3 % (m/m) e segundo Pasquini (2003), o NIR apresenta baixa sensibilidade, apresentando limite de detecção para análises quantitativas de aproximadamente 0,1% (PASQUINI, 2003). Portanto, a baixa quantidade dos PEs presentes na semente também pode ter sido responsável pelo modelo com baixa capacidade de predição obtida para o NIR.

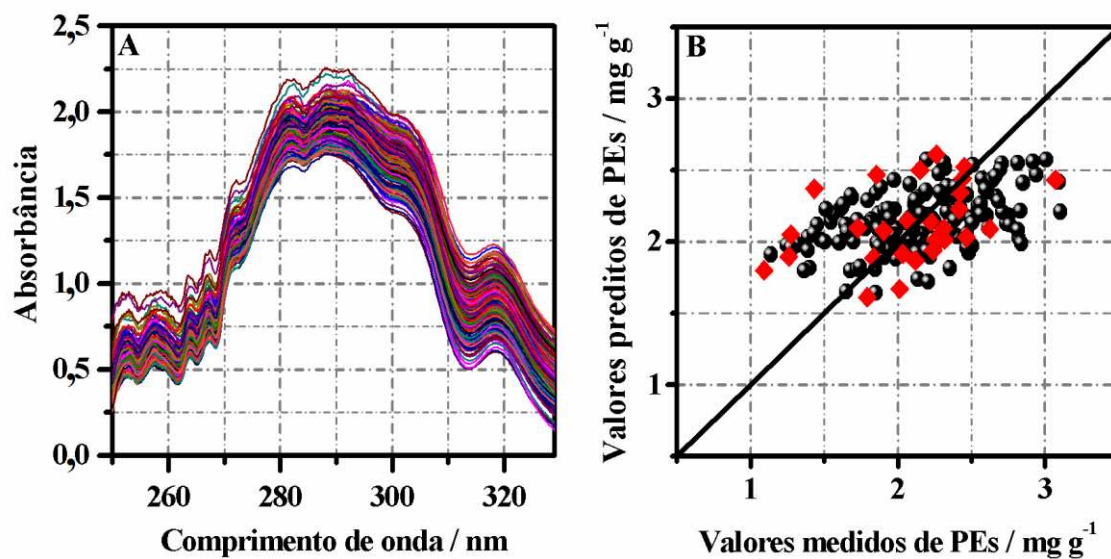
### 5.3.2 Espectros do Óleo

Os espectros NIR das amostras de óleo de pinhão-manso, na faixa de 4000 a 10000  $\text{cm}^{-1}$ , são apresentados na Figura 16 (A). Na Figura 16 (B) são apresentados os valores medidos e preditos para o conjunto de calibração e predição.



**Figura 16.** (A) Espectros NIR do óleo de *Jatropha curcas* L. ( $X_{\text{NIR}_0}$ ) e (B) valores medidos e preditos do teor de PEs para o conjunto de calibração (●) e de predição (♦) para o modelo PLS-OPS.

Os espectros UV das amostras de óleo de pinhão-manso diluído em hexano, na faixa de 240 a 340 nm, são apresentados na Figura 17 (A). Na Figura 17 (B) são apresentados os valores medidos e preditos para o conjunto de calibração e predição.



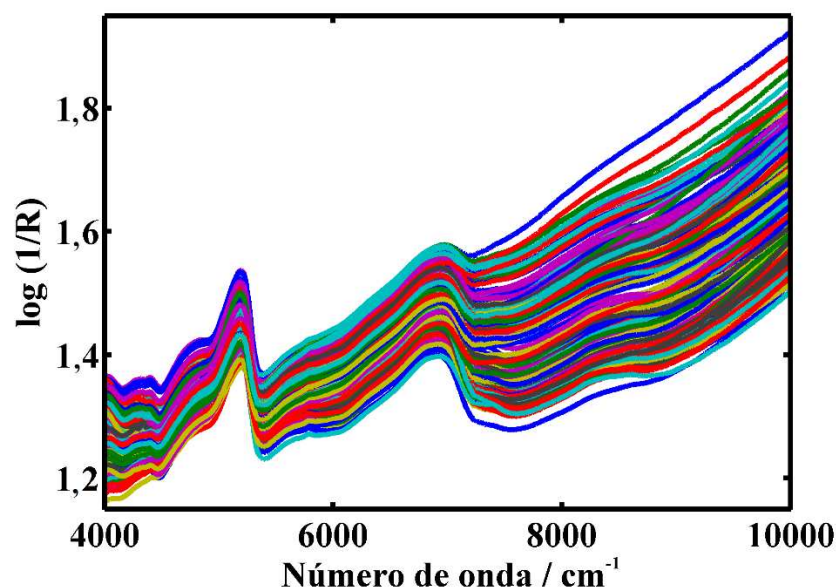
**Figura 17.** (A) Espectros UV do óleo de *Jatropha curcas* L. ( $X_{\text{UV}_0}$ ) e (B) valores medidos e preditos do teor de PEs para o conjunto de calibração (●) e de predição (♦) para o modelo PLS-OPS.

Analisando as Figuras 16 (B) e 17 (B), os valores medidos e preditos não estão concordantes, indicando que o modelo não está apto a realizar predições dos teores de PEs com exatidão. Dessa forma, pode-se concluir que também não é possível prever os teores de PEs no óleo extraído das sementes de pinhão-manso utilizando as espectroscopias NIR e UV. Os modelos obtidos para o óleo possivelmente também não foram satisfatórios

pela presença de muitos interferentes não modelados e pela baixa concentração de ésteres de forbol, que provavelmente não está sendo detectada.

### 5.3.3 Espectros da Semente Com Casca

Os espectros NIR das amostras de semente de pinhão-mansinho com casca ( $n = 186$ ), na faixa de 4000 a 10000  $\text{cm}^{-1}$ , são apresentados na Figura 18.



**Figura 18.** Espectros NIR das sementes de *Jatropha curcas* L. com casca ( $X_{\text{NIRcc}}$ ).

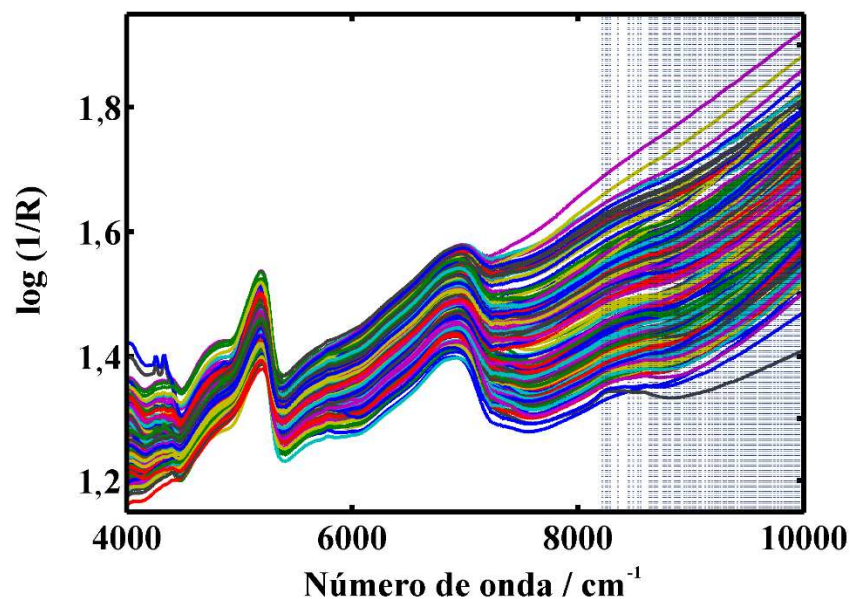
Para a construção do modelo, com o objetivo de obter o menor valor de RMSECV, foram testadas diferentes transformações como, por exemplo, MSC e primeira e segunda derivadas. Em todos os testes, as colunas de  $X$  foram centradas na média (CM). A partir da análise dos valores de RMSECV, verificou-se que as transformações aumentavam o erro e dessa forma os espectros foram apenas centrados na média.

Na construção dos modelos para determinação de PEs na semente com casca foram utilizados os algoritmos OPS e GA para seleção de variáveis. Dessa forma, foi possível a seleção de regiões do espectro que apresentam informações relevantes e que melhor estão correlacionadas com a concentração dos ésteres de forbol em amostras de pinhão-mansinho. Os parâmetros estatísticos calculados para todos os modelos estão apresentados na Tabela 8.

**Tabela 8.** Parâmetros estatísticos dos modelos com todas as variáveis e com a seleção de variáveis pelos métodos OPS e GA.

| <b>Modelos</b> |                 |              |            |
|----------------|-----------------|--------------|------------|
|                | <b>Completo</b> | <b>OPS</b>   | <b>GA</b>  |
| <b>h</b>       | 4               | 5 (h OPS=19) | 4          |
| <b>nVars</b>   | 1556            | 130          | 85         |
| <b>RPD</b>     | 0,3109 (C)      | 0,6590 (C)   | 0,3693 (C) |
| <b>RMSECV</b>  | 0,4408          | 0,4335       | 0,4121     |
| <b>RMSEC</b>   | 0,4242          | 0,3131       | 0,4016     |
| <b>Rc</b>      | 0,2957          | 0,7528       | 0,3518     |
| <b>Rcv</b>     | 0,1776          | 0,4409       | 0,2861     |
| <b>RMSEP</b>   | 0,4441          | 0,4858       | 0,5271     |
| <b>Rp</b>      | 0,0076          | 0,4976       | 0,1276     |
| <b>Viés</b>    | 0,0026          | 0,0001       | 0,0003     |
| <b>%ER</b>     | 20,36           | 18,63        | 23,12      |

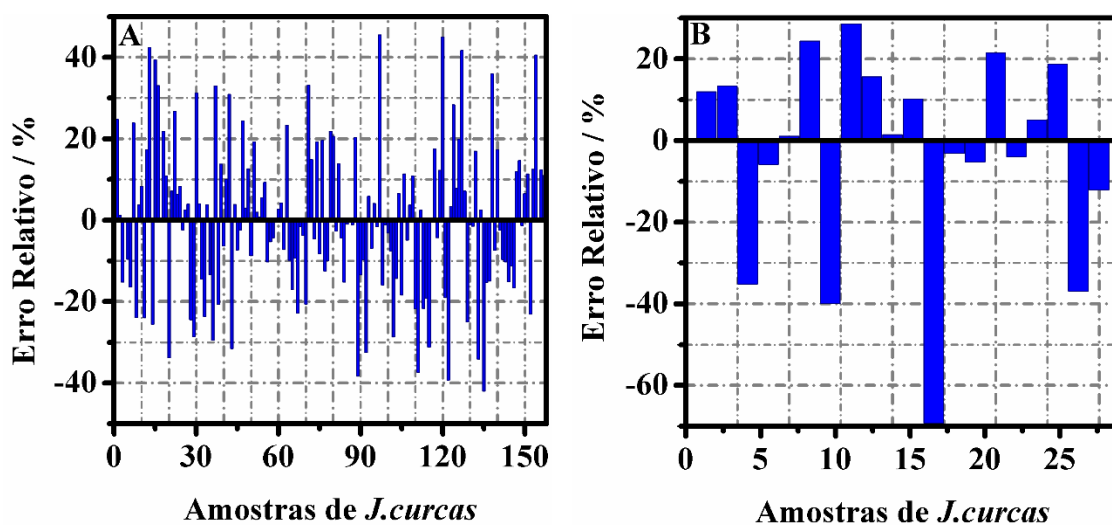
Pela Tabela 8, pode-se notar que todos os modelos obtidos apresentaram  $RPD < 1,4$  e são considerados modelos não confiáveis (CHANG et al., 2001; DUNN et al., 2002), podendo ser utilizados apenas para triagem. Os coeficientes de correlação da calibração, da validação cruzada e da predição, Rc, Rcv e Rp, respectivamente, são consideravelmente maiores para o OPS. Além disso, a média dos erros relativos (% ER) está um pouco elevada em todos os casos, porém os erros para o modelo OPS estão menores. Assim, o modelo obtido pelo OPS pode ser considerado o mais eficiente e robusto. As variáveis selecionadas pelo OPS estão representadas na Figura 19.



**Figura 19.** Variáveis selecionadas pelo OPS.

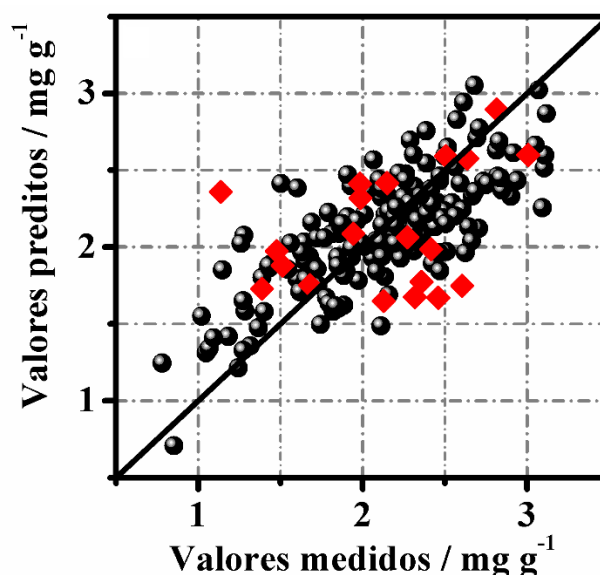
Na região selecionada pelo OPS no espectro NIR apresentado acima, ocorrem vibrações da segunda região de sobretons com estiramentos C-H.

A Figura 20 apresenta os valores de erro relativo para o conjunto de calibração e predição.



**Figura 20.** (A) Erros relativos (%) do conjunto de calibração e (B) conjunto de predição.

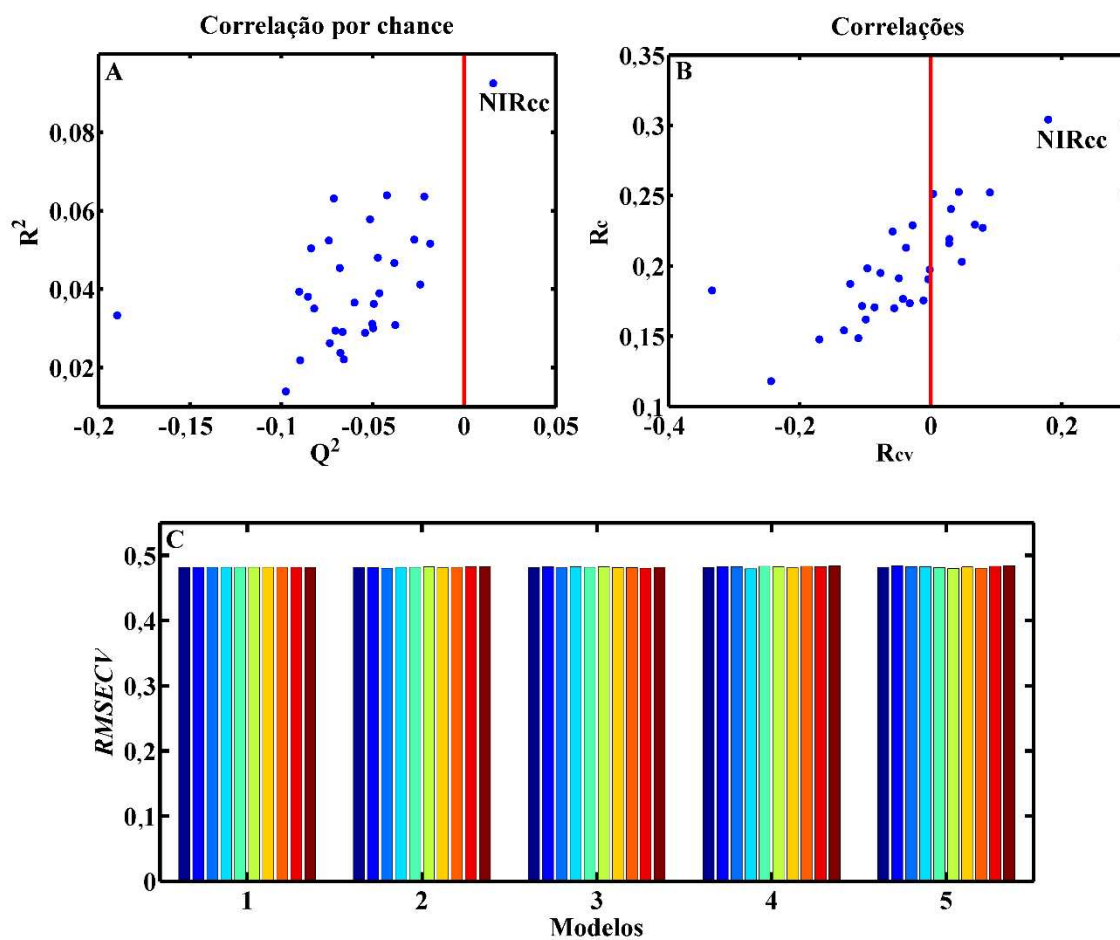
Observando os gráficos da Figura 20, nota-se que os erros relativos estão altos tanto para o conjunto de calibração quanto de predição, indicando que o modelo não está apto a realizar predições exatas dos teores de PEs.



**Figura 21.** Valores medidos versus preditos dos teores de PEs para o conjunto de calibração (●) e de predição (◆).

Pelo gráfico de valores medidos versus preditos apresentado na Figura 21, pode-se perceber que embora o modelo apresente altos erros de predição, há um ajuste linear entre esses valores. Dessa forma, apesar destes erros e um valor de RPD baixo, este modelo pode ser utilizado na discriminação entre amostras com altos e baixos teores de ésteres de forbol, auxiliando na tomada de decisão rápida para o programa de melhoramento genético de plantas.

O modelo foi avaliado para verificar se houve uma correlação por chance. Deste modo aplicou-se o método da aleatorização do vetor  $y$  (variáveis dependentes) e construção de modelos com estas variáveis aleatórias. Dessa forma, avaliam-se os parâmetros de correlação dos modelos construídos com valores de  $y$  aleatórios com o valor de  $y$  medido. Se os parâmetros de correlação de  $y$  medido se distanciarem dos valores de correlação dos  $y$  aleatórios, é um indicativo que o modelo não ocorreu por chance. Além disso, avaliou-se também a estabilidade do modelo através do cálculo do RMSECV para  $n$  modelos construídos quando amostras são retiradas para validação cruzada.



**Figura 22. (A)** Correlação por chance, **(B)** correlações e **(C)** valores de RMSECV obtidos para o teste de estabilidade do modelo.

Pela Figura 22 (A e B) pode-se observar que o modelo obtido por espectroscopia NIR da semente com casca (NIRcc), está separado dos outros modelos construídos. Portanto, o modelo NIRcc não foi obtido ao acaso. Analisando a Figura 22 (C), nota-se que o valor de RMSECV não variou significativamente, portanto o modelo obtido é estável e robusto.

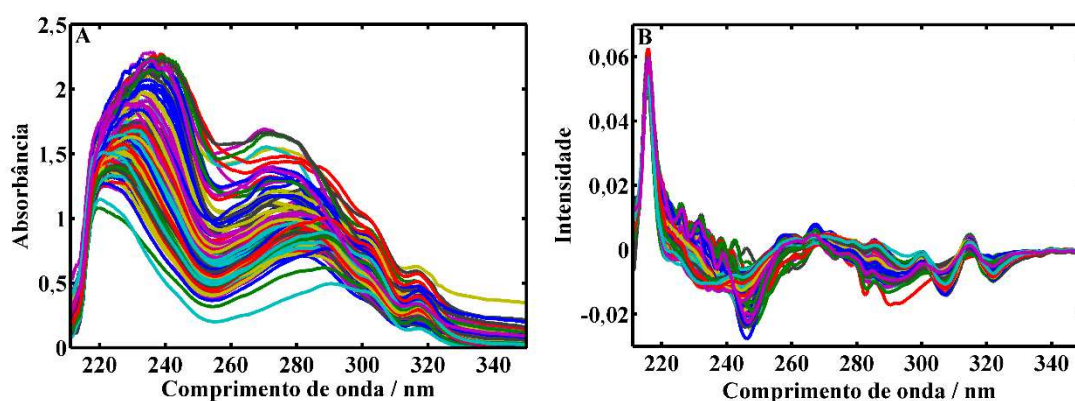
Além disso, acredita-se que a obtenção deste modelo está relacionada a presença de algum componente da casca da semente correlacionado com a quantidade dos ésteres de forbol no interior da mesma.

Dessa forma, a partir deste modelo é possível obter uma faixa do valor da concentração dos ésteres de forbol com a simples obtenção do espectro NIR da semente de pinhão-manso com casca.

### 5.3.4 Espectros do Extrato do Óleo

Os espectros UV do extrato do óleo das amostras de pinhão-manso ( $n = 138$ ), na faixa de 210 a 350 nm, estão apresentados na Figura 23 (A). Para construção do modelo

para este conjunto de dados, foi necessário excluir as amostras que apresentavam concentração inferior a  $0,72 \text{ mg g}^{-1}$ , pois estas afetavam o modelo aumentando consideravelmente os erros.



**Figura 23. (A)** Espectros UV dos extratos do óleo de *Jatropha curcas* L. ( $X_{UVe}$ ) e **(B)** espectro com a primeira derivada.

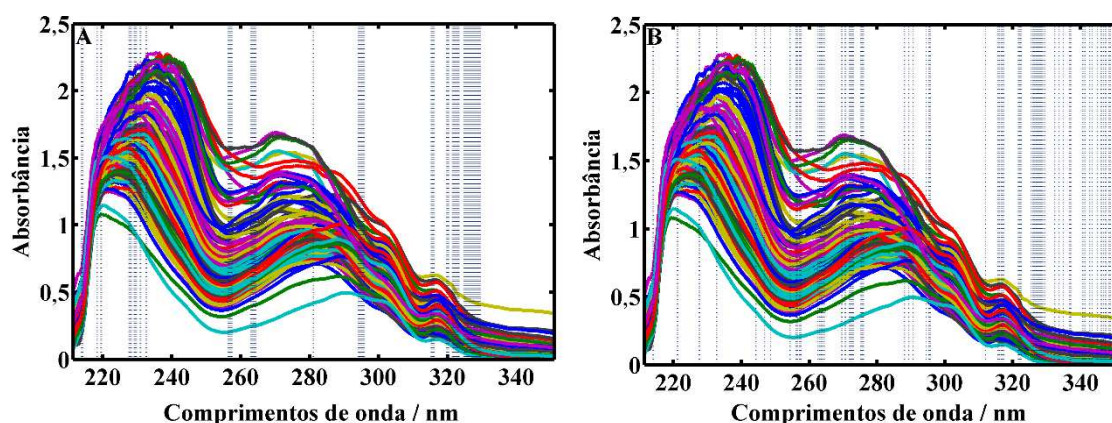
Para a construção do modelo, com o objetivo de obter melhores predições, foram testadas diferentes transformações como, por exemplo, MSC e primeira e segunda derivadas. Em todos os testes as colunas de  $X$  foram centradas na média (CM). A partir da análise dos valores de RMSECV, verificou-se que a melhor transformação foi a segunda derivada. Logo, o modelo foi construído com base nesta transformação, apresentada na Figura 23 (B).

Na construção dos modelos para determinação de PEs em extratos do óleo, foram utilizados os algoritmos OPS e GA para seleção de variáveis. Dessa forma, foi possível a seleção de regiões do espectro que apresentam informações relevantes e que estão melhor correlacionadas com a concentração dos ésteres de forbol em amostras de pinhão-manso. Os parâmetros estatísticos calculados para todos os modelos estão apresentados na Tabela 9.

**Tabela 9.** Parâmetros estatísticos dos modelos com todas as variáveis e com a seleção de variáveis pelos métodos OPS e GA.

|               | Modelos  |              |         |
|---------------|----------|--------------|---------|
|               | Completo | OPS          | GA      |
| <b>h</b>      | 8        | 8 (h OPS=11) | 8       |
| <b>nVars</b>  | 656      | 85           | 85      |
| <b>RPD</b>    | 2,12 (A) | 2,67(A)      | 2,46(A) |
| <b>RMSECV</b> | 0,2798   | 0,2222       | 0,2289  |
| <b>RMSEC</b>  | 0,1923   | 0,1893       | 0,1862  |
| <b>Rc</b>     | 0,9510   | 0,9525       | 0,9501  |
| <b>Rcv</b>    | 0,8950   | 0,9341       | 0,9236  |
| <b>RMSEP</b>  | 0,3004   | 0,2337       | 0,2535  |
| <b>Rp</b>     | 0,9305   | 0,9626       | 0,9637  |
| <b>Viés</b>   | 0,0091   | 0,0002       | 0,0027  |
| <b>%ER</b>    | 12,72    | 9,26         | 12,48   |

Analisando a Tabela 9, o valor de RPD foi  $> 2$  para todos os modelos apresentados. Os valores de RMSECV, RMSEC e RMSEP estão muito próximos para todos os modelos, assim como os valores dos coeficientes de correlação ( $R_c$ ,  $R_{cv}$  e  $R_p$ ). A média dos erros relativos (% ER) está um pouco elevada em todos os casos, porém os erros para o modelo OPS estão consideravelmente menores. O número de variáveis selecionadas pelos dois algoritmos foi o mesmo, porém não foram exatamente as mesmas, como pode ser visto na Figura 24.



**Figura 24.** Variáveis selecionadas pelo algoritmo (A) OPS e (B) pelo GA.

Dessa forma, é importante ressaltar que o OPS conseguiu realizar os cálculos para este conjunto de dados em minutos, enquanto o GA demorou cerca de uma hora, desconsiderando o tempo para otimizar as entradas exigida pelo algoritmo. Assim, o

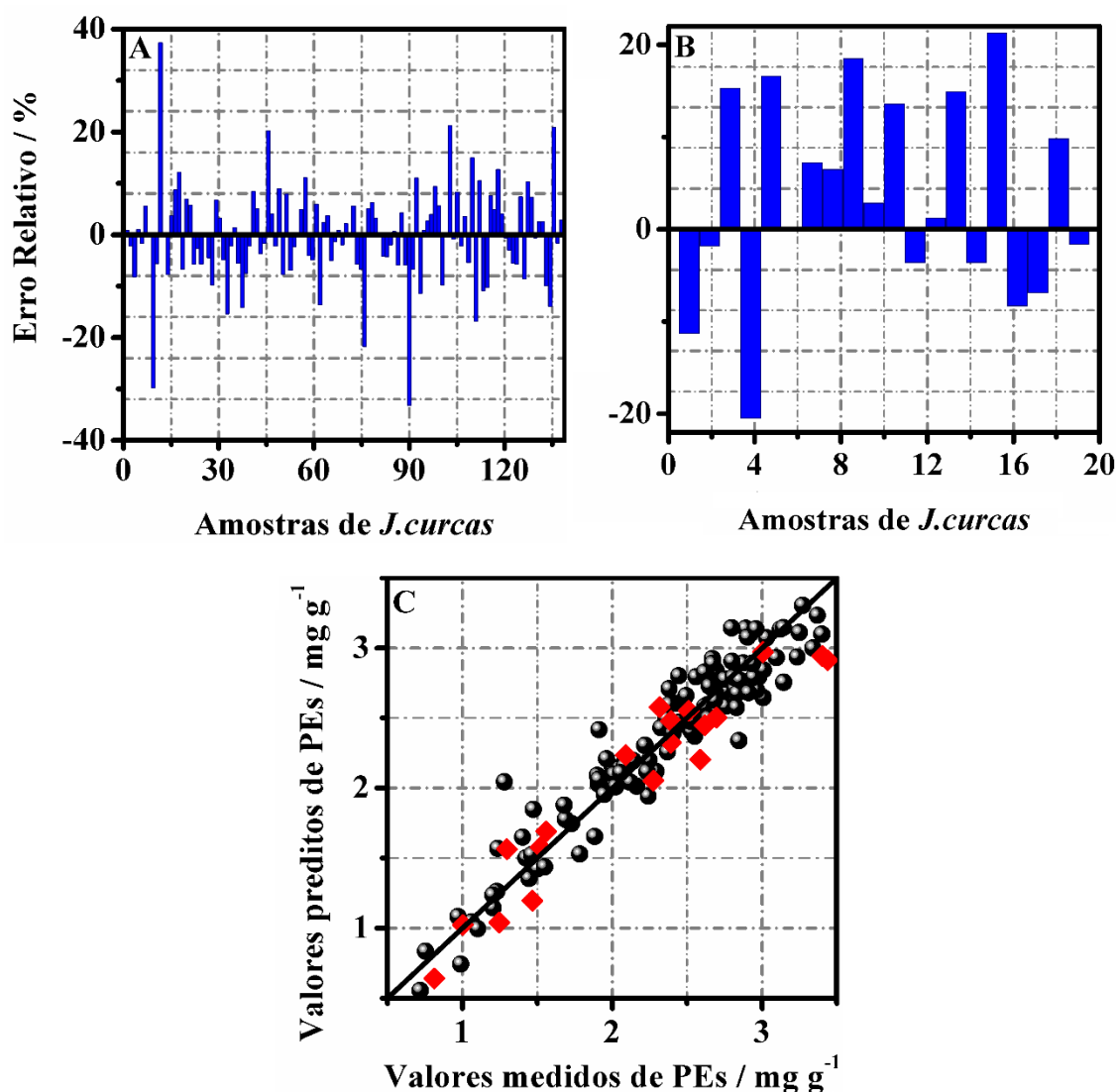
modelo utilizando o algoritmo OPS foi considerado mais eficiente e robusto e os cálculos de validação/figuras de mérito (Tabela 10) foram realizados para o modelo OPS.

**Tabela 10.** Figuras de mérito para o modelo PLS-OPS com o extrato do óleo.

| <b>Extrato do óleo - OPS</b>    |         |
|---------------------------------|---------|
| <b>RPD</b>                      | 2,67    |
| <b>RMSEP</b>                    | 0,2337  |
| <b>Rp</b>                       | 0,9626  |
| <b>SEL</b>                      | 0,1664  |
| <b>SEN.</b>                     | 0,0065  |
| <b><math>\gamma</math></b>      | 12,1775 |
| <b><math>\gamma^{-1}</math></b> | 0,0821  |
| <b>LOD</b>                      | 0,2666  |

Os resultados apresentados na Tabela 10 indicam que o modelo apresenta alta capacidade para prever ésteres de forbol em amostras de J. curcas com exatidão em relação ao método de referência. O inverso da sensibilidade analítica  $\gamma^{-1}$  mostrou que o modelo é sensível. Como era esperado, o modelo apresentou baixa seletividade, uma vez que ele é construído na presença de interferentes.

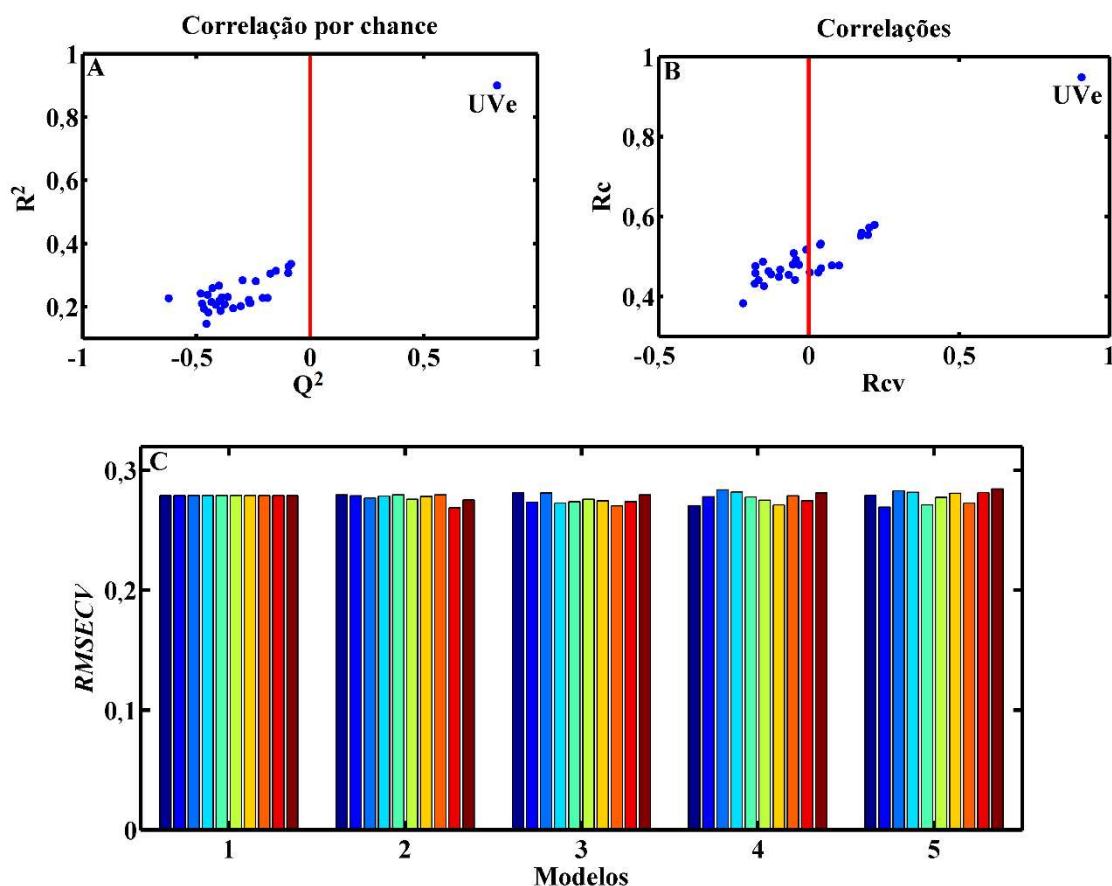
A Figura 25 apresenta os valores dos erros relativos para o conjunto calibração e predição e os valores medidos e preditos.



**Figura 25.** (A) Erros relativos para o conjunto de calibração, (B) erros relativos do conjunto de predição e (C) valores medidos versus preditos para o conjunto de calibração (●) e de predição (♦) para o modelo PLS-OPS.

Pela análise dos gráficos da Figura 25 percebe-se que o maior erro encontrado (em módulo) foi de 37,35 % e o menor 0,0009 % (conjunto calibração). Já para o conjunto predição, os valores variavam na faixa de (em módulo) 21,29% e 0,0115%. Estes erros podem ser considerados altos em um primeiro momento, no entanto, levando em consideração a complexidade da amostra analisada e os baixíssimos teores de ésteres de forbol encontrados, o modelo pode ser considerado apto a realizar predições com exatidão.

Assim como para o modelo anterior, foi realizada a avaliação deste modelo através da investigação da possibilidade de ocorrer uma correlação por chance e a estabilidade do mesmo. Os resultados são apresentado na Figura 26.



**Figura 26.** (A) Correlação por chance, (B) correlações e (C) valores de RMSECV obtidos para o teste de estabilidade do modelo.

Pela Figura 26 (A e B) pode-se observar que o modelo obtido por espectroscopia UV do extrato do óleo (Uve) está separado dos outros modelos construídos e assim o modelo não foi obtido ao acaso. Analisando a Figura 26 (C), nota-se que o valor de RMSECV não variou significativamente, portanto o modelo obtido é estável e robusto.

Na literatura foi encontrado apenas um trabalho de calibração multivariada para quantificação dos ésteres de forbol (MONTES et al., 2013). O modelo obtido, com espectroscopia NIR, apresentou capacidade de classificar as amostras qualitativamente, semelhante ao modelo obtido no presente trabalho com a semente com casca. Além disso, no modelo obtido por Montes et al. foi necessário moer e peneirar as sementes para adquirir os espectros NIR. Para o modelo da semente com casca apresentado neste trabalho, não foi necessário qualquer preparo da amostra.

A utilização da espectroscopia UV, aliada a modelos de calibração multivariada para predição da concentração de PEs, pelo melhor que conhecemos, não foi publicado na literatura. Além disso, não foram encontrados trabalhos para predição de PEs em qualquer parte das sementes, do óleo ou extrato do óleo.

Pelo nosso conhecimento, a construção de modelos de calibração multivariada para quantificar teores muito baixos do analito de interesse diretamente na matriz, sem qualquer preparo de amostras, não foi encontrado na literatura. Para quantificações de níveis baixos e até mesmo traços, é necessário algum tipo de pré-concentração do analito, de forma semelhante ao realizado neste trabalho. Assim, o modelo para prever o teor de ésteres de forbol em amostras de *J. curcas* obtido pela matriz de espectros com o extrato do óleo ( $X_{UVe}$ ) pode perfeitamente substituir o método tradicional, que exige um instrumento de relativo alto custo, alto gasto com solventes, colunas e consumíveis.

Então, a fim de conseguir uma resposta rápida e confiável para incluir ou excluir amostras no programa de melhoramento genético, o espectro NIR da semente com casca é suficiente para tomar esta decisão. Para obter informações mais exatas da concentração dos ésteres de forbol, é necessária a extração dos mesmos a partir do óleo e a obtenção do espectro UV, que fornecerá, a partir do modelo construído, um valor de concentração ao invés de apenas classificar a amostra.

## 6 CONCLUSÃO

Neste trabalho a quantificação dos ésteres de forbol em amostras de pinhão-manso de diversas regiões do Brasil e do Camboja empregando HPLC-DAD proporcionou a obtenção de resultados precisos que foram utilizados na construção de modelos de calibração multivariada.

A espectroscopia NIR não foi capaz de realizar boas predições de PEs sem uma etapa de pré-concentração deste analito, devido a sua sensibilidade limitada. O modelo obtido pelo uso do NIR a partir da semente com casca apresentou capacidade de prever concentrações de ésteres de forbol apenas para comparações relativas para tomadas de decisões em programas de melhoramento genético.

Por outro lado, a espectroscopia UV foi explorada pela primeira vez, juntamente com modelos de calibração multivariada a fim de prever as concentrações de PEs em J. curcas. O modelo obtido a partir do extrato do óleo, utilizando o UV, apresentou boa capacidade de predição quando utilizados para prever o teor de ésteres de forbol em pinhão-manso, podendo ser aplicados, substituindo a técnica HPLC-DAD normalmente utilizada para determinar estes compostos.

Em relação aos modelos de calibração multivariada construídos, o método OPS aliado à regressão PLS proporcionou a construção de modelos mais simples e preditivos quando comparados aqueles obtidos pela seleção de variáveis utilizando o GA.

## **7 TRABALHOS FUTUROS**

Outras estratégias de concentração dos ésteres de forbol como a extração dos ésteres de forbol diretamente da semente, serão avaliadas a fim de obter modelos com maior capacidade preditiva e menor preparo de amostra.

Além disso, será estudada a faixa de concentração de 0,0 a 0,72 mg g<sup>-1</sup> no extrato do óleo, uma vez que esta faixa foi excluída do modelo do extrato do óleo obtido neste trabalho.

## 8 REFERÊNCIAS

- ACHTEN, W. M. J. et al. Jatropha bio-diesel production and use. **Biomass and Bioenergy**, v. 32, n. 12, p. 1063–1084, dez. 2008.
- ADOLF, W.; OPFERKUCH, H. J.; HECKER, E. Irritant phorbol derivatives from four Jatropha species. **Phytochemistry**, v. 23, n. 1, p. 129–132, jan. 1984.
- AENUGU, H. P. R. et al. Near InfraRed Spectroscopy-An Overview. **International Journal of ChemTech Research**, v. 3, n. 2, p. 825–836, 2011.
- AHMED, W. A.; SALIMON, J. Phorbol ester as toxic constituents of tropical Jatropha curcas seed oil. **European Journal of Scientific Research**, v. 31, n. 3, p. 429–436, 2009.
- ALEIXANDRE-TUDO, J. L. et al. Robust ultraviolet-visible (UV-Vis) partial least-squares (PLS) models for tannin quantification in red wine. **Journal of Agricultural and Food Chemistry**, v. 63, n. 4, p. 1088–1098, 2015.
- ALVES, J. C. L.; POPPI, R. J. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). **Talanta**, v. 104, p. 155–161, 2013.
- ANDERSON, L. G. Effects of using renewable fuels on vehicle emissions. **Renewable and Sustainable Energy Reviews**, v. 47, p. 162–172, jul. 2015.
- ANVISA. Agência Nacional de Vigilância Sanitária. **Resolução RE n 899, 29 de maio de 2003**, p. 1–15, 2003.
- ARAÚJO, M. C. U. et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 57, n. 2, p. 65–73, jul. 2001.
- ARRUDA, F. P. et al. Cultivo de pinhão manso (Jatropha curcas L.) como alternativa para o semi-árido nordestino. **Revista Brasileira de Olericultura Fibrose**, v. 8, n. 1, p. 789–799, 2004.
- AZIZIAN, H.; KRAMER, J. K. A rapid method for the quantification of fatty acids in fats and oils with emphasis on trans fatty acids using fourier transform near infrared spectroscopy (FT-NIR). **Lipids**, v. 40, n. 8, p. 855–867, 2005.
- BALAT, M. Potential alternatives to edible oils for biodiesel production – A review of current work. **Energy Conversion and Management**, v. 52, n. 2, p. 1479–1492, fev. 2011.
- BALDINI, M. et al. Determination of phorbol esters in seeds and leaves of Jatropha curcas and in animal tissue by high-performance liquid chromatography tandem mass spectrometry. **Industrial Crops and Products**, v. 59, p. 268–276, ago. 2014.
- BARLOW, J. L.; BOSNER, N.; DRMAC, Z. A new stable bidiagonal reduction algorithm. **Linear Algebra and its Applications**, v. 397, p. 35–84, 2005.

BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. **Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria**. Campinas: Editora Unicamp, 2001.

BAYE, T. M.; PEARSON, T. C.; SETTLES, A. M. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. **Journal of Cereal Science**, v. 43, n. 2, p. 236–243, 2006.

BEEBE, K. R.; KOWALSKI, B. R. An introduction to multivariate calibration and analysis. **Analytical Chemistry**, v. 59, n. 17, p. 1007 A–1017 A, 1987.

BEZERRA, M. A. et al. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. **Talanta**, v. 76, n. 5, p. 965–977, 2008.

BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. **TrAC Trends in Analytical Chemistry**, v. 21, n. 4, p. 240–250, 2002.

BOKOBZA, L. Near infrared spectroscopy. **Journal of Near Infrared Spectroscopy**, v. 6, n. 1, p. 3–17, 1998.

BOOKSH, K. S.; KOWALSKI, B. R. Theory of analytical chemistry. **Analytical Chemistry**, v. 66, n. 15, p. 782A–791A, 1994.

BOQUÉ, R.; LARRECHI, M. S.; RIUS, F. X. Multivariate detection limits with fixed probabilities of error. **Chemometrics and Intelligent Laboratory Systems**, v. 45, n. 1-2, p. 397–408, jan. 1999.

BRERETON, R. G. Introduction to multivariate calibration in analytical chemistry. **The Analyst**, v. 125, n. 11, p. 2125–2154, 1 jan. 2000.

BRERETON, R. G. **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. England: John Wiley & Sons, Ltd., 2003. v. 8

BRERETON, R. G. **Applied chemometrics for scientists**. England: John Wiley & Sons, Ltd., 2007.

BROEKE, J. VAN DEN; LANGERGRABER, G.; WEINGARTNER, A. On-line and in-situ UV / vis spectroscopy for measurements : a brief review. **Spectroscopy Europe 1**, v. 18, n. 4, p. 4–7, 2006.

CARNIELLI, F. **O combustível do futuro**. Belo Horizonte: Boletim UFMG, 2003.

CETEC. **Propriedades Físico-Químicas de Jatropha curcas/Pinhão-manso**. Curitiba. Disponível em: <http://www.pinhaomanso.com.br/propiedades.html>, 2011.

CHANG, C.-W. et al. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. **Soil Science Society of America Journal**, v. 65, n. 2, p. 480, 2001.

CHOUNG, M.-G. et al. Determination of Protein and Oil Contents in Soybean Seed by Near Infrared Reflectance Spectroscopy. **Korean J. Crop Sci.**, v. 46, n. 2, p. 106–111, 2001.

COLLINS, C. H.; BRAGA, G. L.; BONATO, P. S. **Fundamentos de Cromatografia**. 2. ed. Campinas, SP: Editora da Unicamp, 2006.

CONTRAN, N. et al. State-of-the-art of the *Jatropha curcas* productive chain: from sowing to biodiesel and by-products. **Industrial Crops and Products**, v. 42, n. 1, p. 202–215, 2013.

COÛTEAUX, M. M. et al. Determination of water-soluble and total extractable polyphenolics in biomass, necromass and decomposing plant material using near-infrared reflectance spectroscopy (NIRS). **Soil Biology and Biochemistry**, v. 37, n. 4, p. 795–799, 2005.

CURRIE, L. International recommendations offered on analytical detection and quantification concepts and nomenclature. **Analytica Chimica Acta**, v. 391, n. 2, p. 103, maio 1999a.

CURRIE, L. A. Nomenclature in evaluation of analytical methods including detection and quantification capabilities. **Analytica Chimica Acta**, v. 391, n. 2, p. 105–126, maio 1999b.

CURRIE, L. A. Detection and quantification limits: basic concepts, international harmonization, and outstanding (“low-level”) issues. **Applied Radiation and Isotopes**, v. 61, n. 2-3, p. 145–9, jan. 2004.

DE JONG, S. SIMPLS: An alternative approach to partial least squares regression. **Chemometrics and Intelligent Laboratory Systems**, v. 18, n. 3, p. 251–263, mar. 1993.

DEGANI, A. L. G.; CASS, Q. B.; VIEIRA, P. C. Cromatografia - um breve ensaio. **Química Nova na Escola**, n. 7, p. 21–25, 1998.

DEVAPPA, R. K.; BINGHAM, J.-P.; KHANAL, S. K. High performance liquid chromatography method for rapid quantification of phorbol esters in *Jatropha curcas* seed. **Industrial Crops and Products**, v. 49, p. 211–219, ago. 2013.

DEVAPPA, R. K.; MAKKAR, H. P. S.; BECKER, K. Optimization of conditions for the extraction of phorbol esters from *Jatropha* oil. **Biomass and Bioenergy**, v. 34, n. 8, p. 1125–1133, 2010a.

DEVAPPA, R. K.; MAKKAR, H. P. S.; BECKER, K. *Jatropha* toxicity - a review. **Journal of toxicology and environmental health. Part B, Critical reviews**, v. 13, n. 6, p. 476–507, ago. 2010b.

DIAS, L. A. S.; MISSIO, R. F.; DIAS, D. C. Antiquity, botany, origin and domestication of *Jatropha curcas* (Euphorbiaceae), a plant species with potential for biodiesel production. **Genetics and molecular research : GMR**, v. 11, n. 3, p. 2719–2728, 2012.

DIVAKARA, B. N. et al. Biology and genetic improvement of *Jatropha curcas* L.: A review. **Applied Energy**, v. 87, n. 3, p. 732–742, mar. 2010.

DRENNEN, J. K.; KRAEMER, E. G.; LODDER, R. A. Advances and perspectives in near-infrared spectrophotometry. **Critical Reviews in Analytical Chemistry**, v. 22, n. 6, p. 443–475, 1991.

DUNN, B. W. et al. The potential of near-infrared reflectance spectroscopy for soil analysis: a case study from the Riverine Plain of south-eastern Australia. **Australian Journal of Experimental Agriculture**, v. 42, n. 5, p. 607–614, 1 jan. 2002.

EBRAHIMI-NAJAFABADI, H.; LEARDI, R.; JALALI-HERAVI, M. Experimental design in analytical chemistry - part I: theory. **Journal of AOAC International**, v. 97, n. 1, p. 3–11, 2014.

EURACHEM, C.-W. G. **Guide to Quality in Analytical Chemistry - An Aid to Accreditation**. 2. ed. [s.l: s.n.].

FABER, N. (KLAAS) M.; BRO, R. Standard error of prediction for multiway PLS. **Chemometrics and Intelligent Laboratory Systems**, v. 61, n. 1-2, p. 133–149, fev. 2002.

FABER, N. M. Efficient computation of net analyte signal vector in inverse multivariate calibration models. **Analytical Chemistry**, v. 70, n. 23, p. 5108–5110, 1998.

FAIRLESS, D. Biofuel: The little shrub that could - maybe. **NATURE**, v. 449, n. 7163, p. 652–655, 2007.

FERRARI, R. A. et al. Evaluation of the chemical composition and toxic constituent in physic nut plants from different localities. **Brazilian Journal of Food Technology**, v. 12, n. 1/4, p. 309–314, 2009.

FERREIRA, M. M. C. et al. Quimiometria I: Calibração multivariada, um tutorial. **Química Nova**, v. 22, n. 5, p. 724–731, 1999.

FERREIRA, M. M. C. **Quimiometria – Conceitos, Métodos e Aplicações**. Campinas, SP: Editora Unicamp, 2015.

FERREIRA, S. L. C. et al. Doehlert matrix: A chemometric tool for analytical chemistry - Review. **Talanta**, v. 63, n. 4, p. 1061–1067, 2004.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 186, p. 1–17, 1986.

GEMPERLINE, P. **Practical Guide to Chemometrics**. 2. ed. New York: Taylor e Francis, 2006.

GOEL, G. et al. Phorbol esters: structure, biological activity, and toxicity in animals. **International Journal of Toxicology**, v. 26, n. 4, p. 279–288, 2007.

GOLUB, G. H.; KAHAN, W. Calculating the singular values and pseudo-inverse of a matrix. **SIAM Journal on Numerical Analysis**, v. 2, n. 2, p. 205–224, 1965.

GOLUB, G. H.; VAN LOAN, C. F. **Matrix Computations**. London, England: The Johns Hopkins University Press, 3 ed, 1996.

GUI, M. M.; LEE, K. T.; BHATIA, S. Feasibility of edible oil vs. non-edible oil vs. waste edible oil as biodiesel feedstock. **Energy**, v. 33, n. 11, p. 1646–1653, nov. 2008.

HELLER, J. **Physic nut, Jatropha curcas L.** Volume 1 ed. Bioversity international, 1996.

HUBER, L. Validation of analytical methods: review and strategy. **LC/GC International**, v. 11, p. 96–105, 1998.

INMETRO. **Orientação sobre validação de métodos de ensaios químicos**. DOQ-CGCRE-008, 2007.

IUPAC. Nomenclature, symbols, units and their usage in spectrochemical analysis - II. **Spectrochimica Acta, part B**, v. 33, n. 6, p. 241–245, 1978.

JOUBERT, P. H. et al. Acute poisoning with Jatropha curcas (purging nut tree) in children. **South African Medical Journal**, v. 65, n. 18, p. 729–730, 1984.

KANDPAL, J. B.; MADAN, M. Jatropha curcas: a renewable source of energy for meeting future energy needs. **Renewable Energy**, v. 6, n. 2, p. 159–160, mar. 1995.

KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, v. 11, n. 1, p. 137–148, 1969.

KULKARNI, M. L. et al. Jatropha curcas - poisoning. **The Indian Journal of Pediatrics**, v. 72, n. 1, p. 75–76, 2005.

KUMAR, A.; SHARMA, S. An evaluation of multipurpose oil seed crop for industrial uses (Jatropha curcas L.): A review. **Industrial Crops and Products**, v. 28, n. 1, p. 1–10, jul. 2008.

LANÇAS, F. M. **Cromatografia líquida moderna: HPLC/CLAE**. Campinas, SP: Editora Átomo, 2009.

LAVIOLA, B. G.; DIAS, L. A. S. Teor e acúmulo de nutrientes em folhas e frutos de pinhão-mansão. **Revista Brasileira de Ciencia do Solo**, v. 32, n. 5, p. 1969–1975, 2008.

LEARDI, R.; LUPIÁÑEZ GONZÁLEZ, A. Genetic algorithms applied to feature selection in PLS regression: How and when to use them. **Chemometrics and Intelligent Laboratory Systems**, v. 41, n. 2, p. 195–207, 1998.

LI, C. Y. et al. Toxicity of Jatropha curcas phorbol esters in mice. **Food and Chemical Toxicology**, v. 48, n. 2, p. 620–625, 2010.

LINDGREN, F.; GELADI, P.; WOLD, S. The kernel algorithm for PLS. **Journal of Chemometrics**, v. 7, n. 1, p. 45–59, 1993.

LIU, X. et al. Quantitative determination of phorbol ester derivatives in Chinese *Jatropha curcas* seeds by high-performance liquid chromatography/mass spectrometry. **Industrial Crops and Products**, v. 47, p. 29–32, maio 2013.

LORBER, A; FABER, K.; KOWALSKI, B. R. Net Analyte Signal Calculation in Multivariate Calibration. **Analytical Chemistry**, v. 69, p. 1620–1626, 1997.

LUNDSTEDT, T. et al. Experimental design and optimization. **Chemometrics and Intelligent Laboratory Systems**, v. 42, n. 1-2, p. 3–40, 1998.

MAKKAR, H. P. .; BECKER, K. Nutritional studies on rats and fish fed diets containing unheated and heated *Jatropha curcas*. **Plant Foods for Human Nutrition**, v. 53, p. 183–192, 1999.

MAKKAR, H. P. S. et al. Studies on nutritive potential and toxic constituents of different provenances of *Jatropha curcas*. **Journal of Agricultural and Food Chemistry**, v. 45, n. 8, p. 3152–3157, 1997.

MAKKAR, H. P. S. et al. Removal and degradation of phorbol esters during pre-treatment and transesterification of *Jatropha curcas* oil. **Journal of the American Oil Chemical Society**, v. 86, p. 173–181, 2009.

MAKKAR, H. P. S.; ADERIBIGBE, A. O.; BECKER, K. Comparative evaluation of non-toxic and toxic varieties of *Jatropha curcas* for chemical composition, digestibility, protein degradability and toxic factors. **Food Chemistry**, v. 62, n. 2, p. 207–215, jun. 1998.

MALUF, D. F. et al. Determinação simultânea de paracetamol e cafeína por espectrometria UV-Vis associada a ferramentas matemáticas. **Revista Brasileira de Farmácia**, v. 89, n. 1, p. 39–43, 2008.

MAMPANE, K. J.; JOUBERT, P. H.; HAY, I. T. *Jatropha curcas*: use as a traditional Tswana medicine and its role as a cause of acute poisoning. **Phytotherapy Research**, v. 1, n. 1, p. 50–51, 1987.

MANNE, R. Analysis of two partial-least-squares algorithms for multivariate calibration. **Chemometrics and Intelligent Laboratory Systems**, v. 2, n. 1, p. 187–197, 1987.

MARTINEZ HERRERA, J. et al. Evaluation of the nutritional quality of nontoxic kernel flour from *jatropha curcas* L. in rats. **Journal of Food Quality**, v. 35, n. 2, p. 152–158, 2012.

MARTINS, J. P. A; TEOFILO, R. F.; FERREIRA, M. M. C. Computational performance and cross-validation error precision of five PLS algorithms using designed and real data sets. **Journal of Chemometrics**, v. 24, n. 6, p. 320–332, 2010.

MARTINS, J. P. A.; FERREIRA, M. M. C. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. **Química Nova**, v. 36, n. 4, p. 554–560, 2013.

MASSART, L.; VANDEGINSTE, B. **Handbook of Chemometrics and Qualimetrics Part A**. Amsterdam: Elsevier, 1998.

MCCLURE, W. F. 204 years of near infrared technology: 1800-2003. **Journal of Near Infrared Spectroscopy**, v. 11, n. 6, p. 487–518, 2003.

MONTES, J. M. et al. Grain quality determination by means of near infrared spectroscopy in *Jatropha curcas* L. **Industrial Crops and Products**, v. 43, p. 301–305, maio 2013.

MONTGOMERY, D. C. **Design and analysis of experiments**. 5. ed. New York: John Wiley e Sons, Inc., 1987.

MORAIS, C. D. S. et al. Multivariate calibration applied in UV-Vis spectroscopy data to determine total acidity in wines. **Revista Brasileira de Pesquisa em Alimentos**, v. 6, n. 1, 2015.

MUÑOZ DE LA PENA, A. et al. Comparative study of net analyte signal-based methods and partial least squares for the simultaneous determination of amoxycillin and clavulanic acid by stopped-flow kinetic analysis. **Analytica Chimica Acta**, v. 463, n. 1, p. 75–88, 2002.

NORGAARD, L. et al. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. **Applied Spectroscopy**, v. 54, n. 3, p. 413–419, 2000.

OLIVEIRA, F. C. C. et al. A escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos. **Química Nova**, v. 27, n. 2, p. 218–225, abr. 2004.

OPENSHAW, K. A review of *Jatropha curcas*: an oil plant of unfulfilled promise. **Biomass and Bioenergy**, v. 19, n. 1, p. 1–15, jul. 2000.

ORTIZ, M. et al. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. **Chemometrics and Intelligent Laboratory Systems**, v. 69, n. 1-2, p. 21–33, nov. 2003.

OSKOEIAN, E. et al. Bioactive compounds and biological activities of *Jatropha curcas* L. kernel meal extract. **International journal of molecular sciences**, v. 12, n. 9, p. 5955–70, jan. 2011.

PASQUINI, C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, abr. 2003.

PERKAMPUS, H.-H. **UV-VIS spectroscopy and its application**. Berlin, Germany: Springer Berlin, 1992.

PHASUKARRATCHAI, N.; TONTAYAKOM, V.; TONGCUMPOU, C. Reduction of phorbol esters in *Jatropha curcas* L. pressed meal by surfactant solutions extraction. **Biomass and Bioenergy**, v. 45, p. 48–56, 2012.

PINTO, A. C. et al. Biodiesel: an overview. **Journal of the Brazilian Chemical Society**, v. 16, n. 6b, p. 1313–1330, 2005.

PRASAD, L. et al. Experimental assessment of toxic phorbol ester in oil, biodiesel and seed cake of *Jatropha curcas* and use of biodiesel in diesel engine. **Applied Energy**, v. 93, p. 245–250, 2012.

PURCINO, A. A. C.; DRUMMOND, O. A. **Pinhão Manso**. Belo Horizonte: EPAMIG, 1986.

RAMOS, M. J. et al. Influence of fatty acid composition of raw materials on biodiesel properties. **Bioresource technology**, v. 100, n. 1, p. 261–268, jan. 2009.

RASHID, U. et al. *Jatropha Curcas* seed oil as a viable source for biodiesel. **Pakistan Journal of Botany**, v. 42, n. 1, p. 575–582, 2010.

RIBEIRO, F. A. DE L. et al. Planilha de validação: uma nova ferramenta para estimar figuras de mérito na validação de métodos analíticos univariados. **Química Nova**, v. 31, n. 1, p. 164–171, 2008.

RIDDER, T. D.; BROWN, C. D.; STEEG, B. J. V. E. R. Framework for Multivariate Selectivity Analysis, Part I: Theoretical and Practical Merits. **Applied Spectroscopy**, v. 59, n. 6, p. 787–803, 2005.

RINNAN, A.; BERG, F. VAN DEN; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC - Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201–1222, 2009.

ROACH, J. S. et al. Isolation, stability and bioactivity of *Jatropha curcas* phorbol esters. **Fitoterapia**, v. 83, n. 3, p. 586–592, 2012.

ROCHA, F. R. P.; TEIXEIRA, L. S. G. Estratégias para aumento de sensibilidade em espectrofotometria UV-VIS. **Química Nova**, v. 27, n. 5, p. 807–812, 2004.

RODRÍGUEZ, L. C. et al. Estimation of performance characteristics of an analytical method using the data set of the calibration experiment. **Analytical letters**, v. 26, n. 6, p. 1243–1258, 1993.

SAETAE, D.; SUNTORNSUK, W. Variation of PEs contents in *J. curcas* from Thailand.pdf. **American-Eurasian Journal Agriculture & Environmental Science**, v. 8, n. 5, p. 497–501, 2010.

SALDANHA, T. C. B.; ARAÚJO, M. C. U. DE; NETO, B. B. Análise multicomponente simultânea por espectrofotometria de absorção molecular UV-VIS. **Química Nova**, v. 22, n. 6, p. 847–853, 1999.

SANTOS, M. E. DOS; DEMIATE, I. M.; NAGATA, N. Determinação simultânea de amarelo tartrazina e amarelo crepúsculo em alimentos via espectrofotometria UV-VIS e métodos de calibração multivariada. **Ciência e Tecnologia de Alimentos**, v. 30, n. 4, p. 903–909, dez. 2010.

SARABIA, L.; ORTIZ, M. C. DETARCHI: A program for detection limits with specified assurance probabilities and characteristic curves of detection. **TrAC Trends in Analytical Chemistry**, v. 13, n. 1, p. 1–6, jan. 1994.

SATURNINO, H. M. et al. Cultivation of *Jatropha curcas* L. **Informe agropecuário**, v. 26, n. 229, p. 44–78, 2005.

SKOOG, D. A. et al. **Fundamentos de Química Analítica**. 8. ed. Thomson, 2005.

TEÓFILO, R. F. **Métodos quimiométricos em estudos eletroquímicos de fenóis sobre filmes de diamante dopado com boro**. Campinas, SP: Tese (Doutorado em Química) - Universidade Estadual de Campinas, 2007.

TEÓFILO, R. F.; FERREIRA, M. M. C. Quimiometria II: Planilhas eletrônicas para cálculos de planejamentos experimentais, um tutorial. **Química Nova**, v. 29, n. 2, p. 338–350, 2006.

TEÓFILO, R. F.; MARTINS, J. P. A; FERREIRA, M. M. C. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. **Journal of Chemometrics**, v. 23, n. 1, p. 32–48, 2009.

WAKANDIGARA, A.; NHAMO, L. R. M.; KUGARA, J. Chemistry of phorbol ester toxicity in *Jatropha curcas* seed – a Review. **International Journal of Biochemistry Research & Review**, v. 3, n. 3, p. 146–161, 2013.

WANG, Z. et al. Metabolic profiling assisted quality control of phorbol esters in *Jatropha curcas* seed by high-performance liquid chromatography using a fused-core column. **Journal of Agricultural and Food Chemistry**, 2012.

WETZEL, D. L. Near-infrared reflectance analysis. **Analytical Chemistry**, v. 55, n. 12, p. 1165A–1176A, 1983.

WILLIAMS, P.; NORRIS, K. **Near-infrared technology in the agricultural and food industries**. American Association of Cereal Chemists, Inc., 1987.

WOLD, S. et al. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. **SIAM Journal on Scientific and Statistical Computing**, v. 5, n. 3, p. 735–743, 1984.

WOLD, S. et al. Some recent developments in PLS modeling. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 131–150, 2001.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: A basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.

XIAOBO, Z. et al. Variables selection methods in near-infrared spectroscopy.  
**Analytica Chimica Acta**, v. 667, p. 14–32, 2010.

## 9 ANEXOS

**Tabela 11.** Concentração dos ésteres de forbol presentes nos acessos do Banco Ativo de Germoplasma (GAB).

| Amostra selecionada | Acesso | Observação | PEs (mg g <sup>-1</sup> ) | Amostra selecionada | Acesso | Observação | PEs (mg g <sup>-1</sup> ) |
|---------------------|--------|------------|---------------------------|---------------------|--------|------------|---------------------------|
| 1                   | T1     | GAB - I    | 1,553                     | 35                  | 32     | GAB - II   | 2,048                     |
| 3                   | 1      | GAB - I    | 2,393                     | 36                  | 33     | GAB - II   | 2,167                     |
| 4                   | 2      | GAB - I    | 2,275                     | 39                  | 36     | GAB - II   | 2,485                     |
| 5                   | 3      | GAB - I    | 2,110                     | 41                  | 38     | GAB - II   | 2,006                     |
| 7                   | 5      | GAB - I    | 2,528                     | 42                  | 39     | GAB - II   | 2,470                     |
| 8                   | 6      | GAB - I    | 2,392                     | 43                  | 40     | GAB - II   | 2,227                     |
| 9                   | 7      | GAB - I    | 1,582                     | 44                  | 41     | GAB - II   | 2,400                     |
| 10                  | 8      | GAB - I    | 3,250                     | 46                  | T2     | GAB - III  | 2,271                     |
| 11                  | 9      | GAB - I    | 2,976                     | 48                  | 43     | GAB - III  | 2,253                     |
| 12                  | 10     | GAB - I    | 2,443                     | 50                  | 45     | GAB - III  | 2,459                     |
| 14                  | 12     | GAB - I    | 2,544                     | 51                  | 46     | GAB - III  | 1,523                     |
| 15                  | 13     | GAB - I    | 2,188                     | 52                  | 47     | GAB - III  | 2,240                     |
| 17                  | 15     | GAB - I    | 2,349                     | 53                  | 48     | GAB - III  | 2,774                     |
| 18                  | 16     | GAB - I    | 0,220                     | 56                  | 51     | GAB - III  | 2,545                     |
| 19                  | 17     | GAB - I    | 1,866                     | 57                  | 52     | GAB - III  | 2,565                     |
| 20                  | 18     | GAB - I    | 1,080                     | 60                  | 55     | GAB - III  | 2,655                     |
| 21                  | 19     | GAB - I    | 2,691                     | 61                  | 56     | GAB - III  | 2,490                     |
| 22                  | 20     | GAB - I    | 1,213                     | 63                  | 58     | GAB - III  | 1,563                     |
| 23                  | T1     | GAB - II   | 1,869                     | 69                  | 62     | GAB - IV   | 2,629                     |
| 26                  | 22     | GAB - II   | 1,941                     | 70                  | 63     | GAB - IV   | 2,776                     |
| 32                  | 29     | GAB - II   | 2,058                     | 85                  | 83     | GAB - IV   | 2,446                     |
| 33                  | 30     | GAB - II   | 2,117                     | 86                  | 84     | GAB - IV   | 1,812                     |
| 34                  | 31     | GAB - II   | 2,121                     |                     |        |            |                           |

**Tabela 12.** Concentrações dos ésteres de forbol presentes nas progênes do teste de progênie (PT).

| Amostra selecionada | Progênie | PEs (mg g <sup>-1</sup> ) | Amostra selecionada | Progênie | PEs (mg g <sup>-1</sup> ) | Amostra selecionada | Progênie | PEs (mg g <sup>-1</sup> ) |
|---------------------|----------|---------------------------|---------------------|----------|---------------------------|---------------------|----------|---------------------------|
| 88                  | 1        | 1,789                     | 138                 | 65       | 1,966                     | 170                 | 106      | 2,256                     |
| 91                  | 5        | 2,160                     | 142                 | 69       | 1,349                     | 171                 | 107      | 2,294                     |
| 99                  | 14       | 1,772                     | 144                 | 71       | 2,043                     | 174                 | 111      | 2,763                     |
| 100                 | 15       | 1,325                     | 146                 | 73       | 1,938                     | 175                 | 112      | 1,357                     |
| 101                 | 16       | 2,283                     | 148                 | 76       | 2,487                     | 179                 | 118      | 1,665                     |
| 102                 | 18       | 2,556                     | 149                 | 77       | 2,234                     | 186                 | 128      | 1,993                     |
| 103                 | 19       | 2,038                     | 150                 | 79       | 2,470                     | 191                 | 133      | 2,050                     |
| 104                 | 20       | 2,974                     | 152                 | 82       | 0,938                     | 192                 | 134      | 1,919                     |
| 105                 | 22       | 2,098                     | 155                 | 86       | 2,311                     | 193                 | 135      | 1,840                     |
| 109                 | 28       | 1,162                     | 156                 | 88       | 2,066                     | 195                 | 137      | 2,633                     |
| 110                 | 29       | 1,717                     | 158                 | 90       | 2,244                     | 196                 | 138      | 2,345                     |
| 111                 | 30       | 1,468                     | 160                 | 92       | 1,689                     | 197                 | 140      | 2,172                     |
| 112                 | 31       | 0,961                     | 161                 | 94       | 2,401                     | 198                 | 141      | 1,846                     |
| 123                 | 46       | 2,123                     | 164                 | 100      | 1,623                     | 201                 | 144      | 1,438                     |
| 125                 | 48       | 2,093                     | 165                 | 101      | 2,344                     | 202                 | 145      | 2,027                     |
| 127                 | 51       | 2,892                     | 166                 | 102      | 2,624                     | 203                 | 146      | 1,892                     |
| 128                 | 52       | 2,367                     | 167                 | 103      | 2,243                     | 204                 | 147      | 2,106                     |
| 129                 | 53       | 2,294                     | 169                 | 105      | 2,604                     | 206                 | 150      | 2,197                     |
| 130                 | 55       | 1,324                     |                     |          |                           |                     |          |                           |