

UNIVERSIDADE ESTADUAL DE CAMPINAS



PROGRAMA DE PÓS-GRADUAÇÃO DEPARTAMENTO DE QUÍMICA ANALÍTICA

TESE DE DOUTORADO

Planejamento quimiométrico para otimização do solvente extrator e análise exploratória da impressão digital cromatográfica de *Erythrina speciosa* Andrews

AUTORA: Patricia Kaori Soares

ORIENTADOR: Prof. Dr. Roy Edward Bruns

CO-ORIENTADORA: Profa. Dra. leda Spacino Scarminio

10 de Dezembro de 2010 CAMPINAS – SP

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE QUÍMICA DA UNICAMP

Soares, Patricia Kaori. Planejamento quimiométrico para otimização do solvente extrator e análise exploratória da impressão digital cromatográfica de *Erythrina speciosa* Andrews / Patricia Kaori Soares. -- Campinas, SP: [s.n], 2010.
Orientador: Prof. Dr. Roy Edward Bruns. Co-orientadora: Profa. Dra. Ieda Spacino Scarminio.
Doutorado - Universidade Estadual de Campinas, Instituto de Química.
1. Quimiometria. 2. Planejamento experimental de misturas.
3. Tucker3. 4. *Erythrina speciosa* Andrews. I. Bruns, Roy Edward. II. Scarminio, Ieda Spacino. III. Universidade Estadual de Campinas. Instituto de Química. IV. Título.

Título em inglês: Experimental design for solvent extractor optimization and chromatography fingerprint exploratory analysis of *Erythrina speciosa* Andrews

Palavras-chaves em inglês: Chemometrics, Experimental mixture design, Tucker3, *Erythrina speciosa* Andrews

Área de concentração: Química Analítica

Titulação: Doutor em Ciências

Banca examinadora: Prof. Dr. Roy Edward Bruns (orientador), Prof. Dr. Jez Willian Batista Braga (IQ-UNB), Profa. Dra. Maria Fernanda Pimentel Avelar (DEQ-UFPE), Profa. Dra. Isabel Cristina Sales Fontes Jardim (IQ-UNICAMP), Prof. Dr. Jarbas José Rodrigues Rohwedder (IQ-UNICAMP)

Data de defesa: 10/12/2010

Dedico este trabalho

Ao meu Pai, exemplo de homem, que com força, amor e coragem me ensinou a caminhar com os próprios pés. À minha Mãe (in memorian), que me ensinou a viver a vida e a enxergar sempre o lado bom das coisas. Você faz muita falta. Ao meu irmão, que até hoje me ensina o que é paciência. À minha irmã, guerreira, que enfrenta todas as batalhas da vida com um sorriso no rosto. Mesmo com toda distância e todas as dificuldades, sempre estivemos juntos. Ao Fê, que mudou a minha vida em todos os sentidos. É sempre por você. Amo muito todos vocês.

"Vocês são espertos, pena que eu também sou."

Roy Edward Bruns, na mesa do Bar para os alunos da engenharia mecânica

AGRADECIMENTOS

Ao Prof. Dr. Bruns. Não tenho palavras pra expressar minha gratidão por tudo que pude aprender ao seu lado. Um homem fantástico, apaixonado pela química e pela vida. Muito obrigada.

À Prof. Dra. leda, um exemplo de dedicação, paixão e acima de tudo, bom humor. Obrigada por tudo que fez por mim, seu apoio foi fundamental pra que eu chegasse até aqui.

Foi muito bom trabalhar ao lado de vocês, obrigada pela orientação, pelas conversas, pelas risadas e pelas broncas também. Espero que continuem fazendo parte da minha vida, como mestres e como amigos.

À Unicamp, ao Instituto de Química, e ao programa de pós-graduação.

À UEL por ceder a estrutura física para a realização da parte experimental.

Aos Professores das disciplinas cursadas, pelas contribuições pra minha formação.

Aos funcionários do IQ pela competência e seriedade com que realizam suas tarefas.

Aos Professores Doutores: Ronnei, Bel Jardim e Jarbas pelas contribuições no exame de qualificação de área de doutorado.

Ao CNPq pelo financiamento deste trabalho.

Ao meu Pai (Brother) e aos meus irmãos Cau e Nego, pelo amor incondicional e incentivo.

- À minha grande e querida família, tios, tias, primos e primas, em especial, Detão, Tia Elena, Chaveirinho, Tio Makio, Tia Fia,Tchanga, Rê, Chicão, Rô e Dinho, vocês são minha fonte de felicidade.
 - Às crianças, Lê, Naná, Gabi, Mila e o mais novo integrante Lucas, por alegrar e dar sentido à minha vida. Ao Durval, Augusta, Paty e Vini, pelo carinho e compreensão.
- Ao Fê pelo carinho, amor, amizade, bom humor, paciência e incentivo. Obrigada por estar sempre ao meu lado.

À Camilex, por tudo que passamos juntas...muitas histórias, muitas saídas e muitas gargalhadas que vão ficar pra sempre. Obrigada por tudo!

À Gleyce e ao Chicano, por terem me abraçado desde o dia em que cheguei.

Ao Mau, que em tão pouco tempo se tornou meu maior amigo.

À D. Adelaide e ao Sr. Alcides pelo carinho e pela paciência nas noites de música e churrasco.

- Aos amigos que fiz em Campinas, em especial, Marthola, Edson, Pedro V., Zeine, Nicola, Kátia e Zeique, obrigada pelos momentos maravilhosos e divertidos.
- Aos amigos do grupo, Joco Bosco, Kléber, Lívia, Dany, Helen, Vidal e os mais insanos, Serginho e Arnaldo pelas conversas e pelos momentos super descontraídos.

Súmula Curricular (2006-2010)

Dados Pessoais

Nome: Patricia Kaori Soares Naturalidade: Londrina/PR - Brasil Nome em citações bibliográficas : SOARES, Patricia Kaori; SOARES, P. K.

Formação Acadêmica/Titulação

- 2004 2006 Mestrado em Química dos Recursos Naturais. Universidade Estadual de Londrina, UEL, Londrina, Paraná, Brasil
 Título: Taxonomia por abordagem metabolômica para análise e rastreio de plantas do gênero Bauhinia, Ano de obtenção: 2006. Orientador: leda Spacino Scarminio Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- 1998 2003Graduação em Química Bacharelado. Universidade Estadual de Londrina, UEL, Londrina,
Paraná, Brasil
Título: Aplicação de métodos quimiométricos para discriminação de diferentes marcas
de chá de Boldo do Chile. Orientador: Dilson Norio Ishikawa

Formação complementar

- 2010 2010Curso de curta duração (15 horas) em Introdução ao SAS.
Centro Nacional de Processamento de Alto Desempenho em São Paulo, CENAPAD-SP,
Brasil
- 2009 2009Curso de curta duração em (15 horas) Introdução ao FORTRAN.
Centro Nacional de Processamento de Alto Desempenho em São Paulo, CENAPAD-SP,
Brasil

Atividades acadêmicas

- 1º semestre 2009 Monitoria na disciplina MR 676/A –Planejamento e Otimização de Experimentos, ministrada pelo Prof. Dr. Roy Edward Bruns no Programa de Mestrado Profissional em Engenharia Automobilística da FEM/UNICAMP. 45 h/aula
- 2º semestre 2008 Monitoria na disciplina MR 671/A Metodologia para Planejamento Experimental e Análise de Resultados, ministrada pelo Prof. Dr. Roy Edward Bruns no Programa de Mestrado Profissional em Engenharia Automobilística da FEM/UNICAMP. 45 h/aula

Lista de publicações

1. GARCIA, L. M. Z., SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S., **(ANEXO I)** Statistical mixture design - principal component determination of synergic solvent interactions for natural product extractions. Chemometrics and Intelligent laboratory Systems, v.103, p. 1-7, **2010**.

2. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S. (ANEXO II)

Statistical mixture design investigation of fractionated and total extracts from *Erythrina speciosa* Andrew leaves. Journal of Separation Science. , v.32, p.644 - 652, **2009**.

3. SOARES, P. K., SCARMINIO, I. S. (ANEXO III)

Multivariate chromatographic fingerprint preparation and authentication of plant material from the genus *Bauhinia*. Phytochemical Analysis. , v.19, p.78 - 85, **2008**.

4. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S. (ANEXO IV)

Statistical mixture design: Varimax factor optimization for selective compound extraction from plant material. Analytica Chimica Acta., v.613, p.48 - 55, **2008**.

5. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S. (ANEXO V)

Statistical mixture design - principal component optimization for selective compound extraction from plant material. Journal of Separation Science. , v.30, p.3302 - 3310, **2007**.

Manuscrito submetido

1. PASSARI, I. M. Z. G., SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S.**(ANEXO VI)** Estatística aplicada à química: dez dúvidas comuns. Química Nova.

Trabalhos em eventos internacionais

1. PASSARI, L. M. Z. G., OLIVEIRA, T. F., SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S. Binary synergic interaction effects in solvent mixtures for fractionated and total extracts of Mikania laevigata Sch. Bip. leaves, 12th Conference on Chemometrics in Analytical Chemistry, **2010**, Antuérpia, Bélgica.

2. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S.

Statistical mixture design – varimax factor optimization for selective compound extraction from plant material. 11th Conference on Chemometrics in Analytical Chemistry, **2008**, Montpellier, França.

3. SOARES, P. K, DELAROZA, F., SCARMINIO, I. S.

A comparative study of *Bauhinia* species by FTIR fingerprint and chemometric analysis. 10th International Conference on Chemometrics in Analytical Chemistry, **2006**, Águas de Lindóia, Brasil.

4. SOARES, P. K, DELAROZA, F., SCARMINIO, I. S.

Chromatographic fingerprint preparation and authentication of the plant material of *Bauhinia* genus by multivariate methodology, 10th International Conference on Chemometrics in Analytical Chemistry, **2006**, Águas de Lindóia, Brasil.

Apresentação oral em eventos internacionais

1. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S.

Statistical mixture design – varimax factor optimization for selective compound extraction from plant material. 11th Conference on Chemometrics in Analytical Chemistry, 2008, Montpellier. Experimental Design and Image Analysis. Montpellier: Cemagref, **2008**. v.4. p.29 - 33

Trabalhos em eventos nacionais

1. SOARES, P. K., DELAROZA, F., BRUNS, R. E., SCARMINIO, I. S.

Discriminação dos espectros UV-vis dos cromatogramas dos extratos de *B. variegata* L. preparados com base num planejamento centróide simplex modificado. 33° Reunião Anual da Sociedade Brasileira de Química, **2010**, Águas de Lindóia.

2. SOARES, P. K., GRACIA, L. M. Z., BRUNS, R. E., SCARMINIO, I. S.

Emprego de planejamento estatístico de misturas no processo de extração das folhas de *Mikania laevigata Sch.* Bip.. 33° Reunião Anual da Sociedade Brasileira de Química, **2010**, Águas de Lindóia.

3. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S., BORTOLOTI, J. A.

Planejamento mistura-processo e modelagem da área total cromatográfica dos extratos brutos de *Erythrina speciosa* Andrews. 15° Encontro Nacional de Química Analítica, **, 2009**, Salvador.

4. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S.

Planejamento estatístico de mistura para avaliação de rendimento de extrato bruto e das frações das folhas de *Erythrina speciosa* Andrews. 32° Reunião Anual da Sociedade Brasileira de Química: , **2009**, Fortaleza.

5. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S.

Uso de componentes principais com e sem rotação varimax na otimização de misturas extratoras de material vegetal. 31° Reunião Anual da Sociedade Brasileira de Química: , **2008**, Águas de Lindóia.

6. SOARES, P. K., BRUNS, R. E., SCARMINIO, I. S.

Planejamento experimental para otimização de misturas extratoras de material vegetal com uso de componentes principais. 14° Encontro Nacional de Química Analítica, **2007**, João Pessoa.

7. SOARES, P. K., BRUNS, Roy E, SCARMINIO, I. S.

Planejamento estatístico para estudo fitoquímico da *Erythrina speciosa* Andrews. 30° Reunião Anual da Sociedade Brasileira de Química, **2007**, Águas de Lindóia.

8. SOARES, P. K., SCARMINIO, I. S., SILVA, L. M. C.,

Comparação da eficiência de diferentes solventes extratores no estudo da *Baccharis trimera* por CLAE In: 30° Reunião Anual da Sociedade Brasileira de Química, **2007**, Águas de Lindóia.

9. SOARES, P. K., SCARMINIO, I. S., LEMOS, A. Y.

Discriminação de extratos de plantas do gênero *Bauhinia* por espectrofotometria UV-Vis e métodos quimiométricos. 29° Reunião Anual da Sociedade Brasileira de Química, **2006**, Águas de Lindóia.

10. SOARES, P. K., SCARMINIO, I. S.

Taxonomia por abordagem metabolômica para análise e rastreio de quatro variedades de *Bauhinia forficata Link*. 29° Reunião Anual da Sociedade Brasileira de Química, **2006**, Águas de Lindóia.

11.SOARES, P. K., SCARMINIO, I. S., DELAROZA, F.

Comparação de dois métodos de preparo de amostras de material vegetal para análise por CLAE. 29° Reunião Anual da Sociedade Brasileira de Química, **2006**, Águas de Lindóia.

RESUMO

Nos últimos anos muitos avanços foram feitos em termos da descoberta de novos princípios ativos de plantas medicinais. Mesmo com toda a tecnologia dos dias de hoje, uma das etapas mais importantes na busca pelo isolamento de novas substâncias continua sendo o processo de extração. Nesta etapa, além do método de extração, outra escolha fundamental é o solvente extrator a ser utilizado, pois a informação química a ser obtida, tanto qualitativa quanto quantitativa dependerá exclusivamente desses dois fatores. O planejamento de misturas neste caso permite identificar as interações entre os solventes utilizados, ou mesmo constatar que não há interação significativa entre eles, auxiliando na maximização ou minimização de uma resposta. Neste trabalho, a composição química das folhas da planta Erythrina speciosa Andrews foi investigada. Os extratos foram preparados com base em um planejamento experimental de misturas do tipo Centróide-Simplex com quatro componentes; etanol, diclorometano, hexano e acetona. Ao todo foram quinze misturas diferentes, sendo que o ponto central foi feito em quintuplicata, totalizando 19 extratos. Os extratos brutos passaram por um processo de fracionamento, que resultou nas fibras, nas frações neutra, básica e orgânica e também nas sobras, que foram calculadas subtraindo do extrato bruto, o peso das fibras e de todas as frações. Para avaliar os rendimentos através da metodologia de superfície de resposta, foram utilizados os pesos dos extratos brutos, das fibras, das frações neutra, básica e orgânica e também das sobras. Os resultados dos dados gravimétricos demonstraram a importância do planejamento experimental de misturas aplicado à obtenção dos extratos brutos de Erythrina speciosa e das frações em termos da quantidade extraída por cada um dos diferentes extratores, pois em todos os casos destacaramse as misturas binárias e ternárias. Os extratos brutos e as frações básica e orgânica foram analisados por cromatografia líquida. Os dados cromatográficos obtidos foram interpretados com a aplicação dos métodos de análise de componentes principais e análise hierárquica de agrupamentos e os dados cromatográficos de segunda ordem foram avaliados através da aplicação do método Tucker3. A análise exploratória dos dados cromatográficos permitiu discriminar as quinze misturas diferentes de acordo com os solventes utilizados e com base em alguns picos mais importantes.

XV

ABSTRACT

In recent years many advances have been made in terms of discovering of new substances from medicinal plants. Even with all the present day technology, one of the most important steps to achieve the isolation of new substances remains the extraction process. Besides the extraction method itself, another fundamental choice is the organic solvent to be used because the chemical information to be obtained, both qualitatively and quantitatively depend solely on these two factors. The experimental mixture design in this case, can identify the interactions between the solvents used, or even perceive no significant interaction between them, helping to maximize or minimize a response. In this work, the chemical composition of the leaves of *Erythring speciosa* Andrews was investigated. The extracts were prepared based on a Simplex Centroid mixture design with four components, ethanol, dichloromethane, hexane and acetone. A total of nineteen extractions were carried out with fifteen different mixtures. A five run replicate was performed at the center point. The extracts were submitted to a fractionation process, which resulted in the fibers, in the neutral, basic and organic fractions and also on the rest, which were calculated by subtracting from the crude extract, the weight of the fibers and fractions. To evaluate the yield by response surface methodology, the weights of the crude extract, the fibers, neutral, basic and organic fractions and also the rest were used. The results of the gravimetric data demonstrated the importance of mixture design applied to obtain the crude extracts of Erythrina speciosa and its fractions in terms of quantity extracted by each of the different extracts, because in almost all cases binary and ternary mixtures were more efficient. The extracts, basic and organic fractions were analyzed by liquid chromatography. The chromatographic data were interpreted by the application of principal component analysis and hierarchical cluster analysis and the second-order chromatographic data were evaluated by the application of the Tucker3 method. The exploratory analysis of the chromatographic data allowed discriminating the fifteen different mixtures according to the solvents used based on some major peaks.

SUMÁRIO

Índice de Tabelas xxii
Índice de Figurasxxiv
Lista de abreviaturasxxx
1 Introdução1
2 Cromatografia5
3 Métodos quimiométricos9
3.1 Planejamento de misturas do tipo Centróide-Simplex9
3.2 Pré-processamento13
3.3 Análise de componentes principais14
3.4 Análise hierárquica de agrupamentos16
3.5 Tucker3
4 Parte experimental
4.1 Coleta e secagem21
4.2 Planejamento experimental de misturas22
4.3 Procedimento experimental na fase de teste23
4.3.1 Preparo dos extratos23
4.3.2 Análise por cromatografia líquida de alta eficiência24
4.3.3 Reagentes e equipamentos
4.4 Procedimento experimental na fase exploratória25
4.4.1 Preparo dos extratos e fracionamento25
4.4.2 Análise por cromatografia líquida de alta eficiência27
4.4.3 Reagentes e equipamentos
4.5 Métodos estatísticos e programas computacionais29
5 Resultados

5.1 Resultados do teste inicial31
5.1.1 Análise de componentes principais dos cromatogramas dos extratos preparados com
base no planejamento Centróide-Simplex 32
5.1.2 Análise de componentes principais com rotação varimax dos cromatogramas dos
extratos preparados com base no planejamento Centróide-Simplex
5.2 Resultados da fase exploratória59
5.2.1 Análise dos dados gravimétricos do rendimento do extrato bruto61
5.2.2 Análise de dados gravimétricos do rendimento das fibras62
5.2.3 Análise dos dados gravimétricos dos rendimentos da fração neutra65
5.2.4 Análise dos dados gravimétricos do rendimento da fração orgânica67
5.2.5 Análise dos dados gravimétricos do rendimento da fração básica
5.2.6 Modelagem das sobras73
5.2.7 Análise de componentes principais dos dados gravimétricos dos rendimentos do extrato
bruto e das frações76
5.2.8 Análise de componentes principais dos dados cromatográficos do extrato bruto78
5.2.9 Análise de componentes principais dos dados cromatográficos da fração orgânica85
5.2.10 Análise de componentes principais dos dados cromatográficos da fração básica92
5.2.11 Análise de segunda ordem dos dados CLAE-DAD do extrato bruto
5.2.13 Análise de segunda ordem dos dados CLAE-DAD da fração orgânica
5.2.13 Análise de segunda ordem dos dados CLAE-DAD da fração básica
6 Conclusão
7 Referências137
8 Anexos141
I Statistical mixture design - Principal component determination of synergic solvent interactions
for natural product extractions141
II Statistical mixture design investigation of fractionated and total extracts from Erythrina
speciosa Andrews leaves149

III Multivariate Chromatographic Fingerprint Preparation and Authentication of Plant	: Material
from the Genus Bauhinia	163
IV Statistical mixture design - Varimax factor optimization for selective compound e	extraction
from plant material	172
V Statistical mixture design - principal component optimization for selective component	ompound
extraction from plant	
VI Estatística aplicada à Química: dez dúvidas comuns	197

ÍNDICE DE TABELAS

Tabela 1: Membros dos grupos determinados por Snyder6
Tabela 2. Análise da variância para um planejamento de misturas com replicatas11
Tabela 3: Composição do sistema extrator em termos de proporção. r1 a r5 corresponde àsreplicatas de 1 a 5 no ponto central
Tabela 4: Composição e força cromatográfica das duas fases móveis usadas na fase de testes 24
Tabela 5: Composição e força cromatográfica das três fases móveis testadas na fase exploratória 28
Tabela 6: Valores dos escores das três primeiras componentes principais e suas porcentagens devariância explicada para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvelMeOH:H2O 80:20 (v/v) e temperatura de 50°C)33
Tabela 7: Análise da variância para o ajuste do modelo quadrático aos valores de escores da CP1para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvel MeOH:H2O 80:20 (v/v)e temperatura de 50°C)
Tabela 8: Análise da variância para o ajuste do modelo quadrático aos valores de escores da CP2para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvel MeOH:H2O 80:20 (v/v)e temperatura de 50°C)41
Tabela 9: Análise da variância para o ajuste do modelo quadrático aos valores de escores da CP3para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvel MeOH:H2O 80:20 (v/v)e temperatura de 50°C)43
Tabela 10: Parâmetros utilizados na otimização simultânea dos escores da ACP para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvel MeOH:H2O 80:20 (v/v) e temperatura de 50oC) para os picos com tempo de retenção em 1,7, 3,1 e 6,6 min
Tabela 11: Composição do solvente extrator otimizado, respostas previstas pela análise dasuperfície de resposta e pela função de desejabilidade e valores máximos e mínimos dos escoresdas componentes principais 1, 2 e 347
Tabela 12: Valores dos escores rodados da CP1 ^r , CP2 ^r e CP3 ^r e suas porcentagens de variânciaexplicada para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvel MeOH:H2O80:20 (v/v) e temperatura de 50°C)49
Tabela 13 : Análise da variância para o ajuste do modelo quadrático aos valores de escores rodados das CP1 ^r , CP2 ^r e CP3 ^r para os dados cromatográficos dos extratos brutos de <i>Erythrina</i> (Fase móvel MeOH:H ₂ O 80:20 (v/v) e temperatura de 50°C)
Tabela 14: Composição do solvente extrator otimizado, respostas previstas pela análise dasuperfície de resposta e pela função de desejabilidade e valores máximos e mínimos dos escoresrodados das componentes principais 1, 2 e 357

Tabela 15: Rendimentos em gramas, média e desvio padrão dos rendimentos do extrato bruto, dasfibras, das frações neutra, orgânica e básica e dos resíduos60
Tabela 16: Correlações entre as seis variáveis no nível de 95% de confiança61
Tabela 17 : Análise da variância para o ajuste do modelo cúbico especial aos rendimentos das fibrasobtidas com o fracionamento dos extratos brutos de <i>Erythrina</i>
Tabela 18 : Análise da variância para o ajuste do modelo cúbico especial aos rendimentos da fraçãoneutra obtida com o fracionamento dos extratos brutos de <i>Erythrina</i> 65
Tabela 19 : Análise da variância para o ajuste do modelo quadrático aos rendimentos da fraçãoorgânica obtida com o fracionamento dos extratos brutos de <i>Erythrina</i> 69
Tabela 20 : Análise da variância para o ajuste do modelo cúbico especial aos rendimentos da fraçãobásica obtida com o fracionamento dos extratos brutos de <i>Erythrina</i>
Tabela 21 : Análise da variância para o ajuste do modelo cúbico especial aos resíduos obtidos pelocálculo da diferença entre o extrato bruto e a soma de todas as outras frações obtidas com ofracionamento dos extratos brutos de <i>Erythrina</i> 74
Tabela 22: Correlações entre os escores das três primeiras componentes principais da faseexploratória, os escores das três primeiras componentes principais sem rotação varimax da fase detestes e os rendimentos dos extratos brutos no nível de 95% de confiança
Tabela 23: Soma total de fatores, variância explicada e número de fatores nos modos A, B e C paraos dados cromatográficos dos extratos brutos de <i>Erythrina</i> (FM1)
Tabela 24. Importância dos elementos do core em ordem decrescente de acordo com a variânciado core para o modelo Tucker3 para os dados cromatográficos dos extratos brutos de Erythrina(FM1)
Tabela 25: Soma total de fatores, variância explicada e número de fatores nos modos A, B e C paraos dados cromatográficos da fração orgânica de <i>Erythrina</i> (FM1)112
Tabela 26. Importância dos elementos do core em ordem decrescente de acordo com a variânciado core para o modelo Tucker3 para os dados cromatográficos da fração orgânica de Erythrina(FM1)
Tabela 27: Soma total de fatores, variância explicada e número de fatores nos modos A, B e C paraos dados cromatográficos da fração básica de <i>Erythrina</i> (FM1)123
Tabela 28. Importância dos elementos do core em ordem decrescente de acordo com a variânciado core para o modelo Tucker3 para os dados cromatográficos da fração básica de Erythrina(FM1)

ÍNDICE DE FIGURAS

Figura 1: Etapas envolvidas no desenvolvimento de fitomedicamentos1
Figura 2: Triângulo de seletividade de Snyder para solventes usados em cromatografia7
Figura 3: Foto da planta Erythrina speciosa Andrews
Figura 4: Planejamento experimental do tipo Centróide-Simplex com 4 componentes
Figura 5: Procedimento geral do fracionamento prévio das diferentes classes de metabólitos 26
Figura 6: Cromatogramas originais das replicatas <i>r1</i> a <i>r5</i> do extrato bruto no ponto central obtidos no estudo da variação da composição da fase móvel e da temperatura
Figura 7: Gráfico dos escores das componentes principais (a) 1 e 2 e (b) 1 e 3 dos dados cromatográficos dos extratos brutos de <i>Erythrina</i> (Fase móvel MeOH:H ₂ O 80:20 (v/v) e temperatura de 50°C). A seta indica que as amostras <i>dh</i> e <i>dha</i> fazem parte do grupo II
Figura 8: Gráfico dos loadings da CP1, CP2 e CP3 dos dados cromatográficos dos extratos brutos de <i>Erythrina</i> (Fase móvel MeOH:H ₂ O 80:20 (v/v) e temperatura de 50° C)
Figura 9: Dendrograma baseado nos dados cromatográficos dos extratos brutos de <i>Erythrina</i> obtido com o método de agrupamento pareado igualmente ponderado
Figura 10: Cromatogramas dos extratos brutos de <i>Erythrina</i> separados com base nos grupos formados pela ACP (Fase móvel MeOH: H_2O 80:20 (v/v) e temperatura de 50°C)
Figura 11: Gráficos (a) dos resíduos deixados pelo ajuste do modelo quadrático e (b) das respostas previstas pelas respostas observadas para os dados cromatográficos dos extratos brutos de <i>Erythrina</i> (Fase móvel MeOH:H ₂ O 80:20 (v/v) e temperatura de 50°C)
Figura 12: Superfície de resposta para os componentes; diclorometano, hexano e acetona, aplicada aos escores da CP1
Figura 13: Gráficos (a) dos resíduos deixados pelo ajuste do modelo quadrático e (b) das respostas previstas pelas respostas observadas para os escores da CP2 para os dados cromatográficos dos extratos brutos de <i>Erythrina</i> (Fase móvel MeOH:H ₂ O 80:20 (v/v) e temperatura de 50°C)42
Figura 14: Superfície de resposta para os componentes; diclorometano, hexano e acetona, aplicada aos escores da CP242
Figura 15: Gráficos (a) dos resíduos deixados pelo ajuste do modelo quadrático e (b) das respostas previstas pelas respostas observadas para os escores da CP3 para os dados cromatográficos dos extratos brutos de <i>Erythrina</i> (Fase móvel MeOH:H ₂ O 80:20 (v/v) e temperatura de 50°C)44
Figura 16: Superfície de resposta para os componentes; etanol, hexano e acetona, aplicada aos escores da CP3

das amostras do planejamento para os picos em (a) 1,7 min, (b) 3,1 min e (c) 6,6 min (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C)48 Figura 18: Gráfico dos escores das componentes principais rodadas (a) 1 e 2 e (b) 1 e 3 para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvel MeOH:H₂O 80:20 (v/v) e Figura 19: Loadings da (a) CP1 e CP1^r, (b) CP2 e CP2^r e (c) CP3 e CP3^r para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C)...... 51 Figura 20: Superfície de resposta para os componentes; etanol, diclorometano e acetona, aplicada Figura 21: Superfície de resposta para os componentes; etanol, hexano e acetona, aplicada aos escores da CP2^r......55 Figura 22: Superfície de resposta para os componentes; etanol, hexano e acetona, aplicada aos escores da CP3^r......56 Figura 23: Cromatogramas dos experimentos confirmatórios comparados ao cromatograma da mistura otimizada dos escores da CP1 para o pico em 3,1 min (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C)......58 Figura 24: Cromatogramas dos pontos h, eh e ha do planejamento (Fase móvel MeOH:H₂O 80:20 Figura 25: Planejamento experimental Centróide-Simplex com os valores experimentais dos Figura 26: Gráfico (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das Figura 27: Superfície de resposta para os componentes; etanol, diclorometano e hexano, aplicada aos rendimentos das fibras......64 Figura 28: Gráficos (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das Figura 29: Superfície de resposta para os componentes; etanol, diclorometano e acetona, aplicada aos rendimentos da fração neutra67 Figura 30: Gráficos (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das respostas previstas pelas respostas observadas para os rendimentos da fração orgânica......70 Figura 31: Superfície de resposta para os componentes; etanol, diclorometano e hexano, aplicada aos rendimentos da fração orgânica70 Figura 32: Gráficos (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das respostas previstas pelas respostas observadas para os rendimentos da fração básica......72 XXV

Figura 17: Cromatogramas dos experimentos confirmatórios comparados com os cromatogramas

Figura 33: Superfície de resposta para os componentes; etanol, diclorometano e hexano, aplicada aos rendimentos da fração básica
Figura 34: Gráficos (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das respostas previstas pelas respostas observadas para os resíduos do fracionamento
Figura 35: Superfície de resposta para os componentes; etanol, diclorometano e acetona, aplicada aos resíduos do fracionamento
Figura 36: (a) Gráfico dos escores das duas primeiras componentes principais e (b) gráfico dos loadings das duas primeiras componentes principais
Figura 37: Cromatogramas das replicatas do extrato bruto no ponto central nas fases móveis 1, 2 e 3
Figura 38: Gráfico dos escores das componentes principais (a)1 e 2 e (b) 2 e 3 dos dados cromatográficos dos extratos brutos de <i>Erythrina</i> (FM1). A seta indica que as amostras <i>ed</i> e <i>edh</i> fazem parte do grupo IV
Figura 39: Gráfico dos loadings das componentes principais (a) 1, (b) 2 e (c) 3 para os dados cromatográficos dos extratos brutos de <i>Erythrina</i> (FM1)
Figura 40: Cromatogramas transformados dos extratos brutos de <i>Erythrina</i> separados com base nos grupos formados pela ACP
Figura 41: Dendrograma baseado nos dados cromatográficos dos extratos brutos de Erythrinaobtido com o método de Ward83
Figura 42: Gráfico dos valores de escores da CP3 e dos rendimentos do extrato bruto85
Figura 43: Cromatogramas da fração orgânica no ponto central (replicatas), nas fases móveis (a) 1, (b) 2 e (c) 3
Figura 44: Gráfico dos escores das componentes principais (a)1 e 2 e (b) 2 e 3 dos dados cromatográficos da fração orgânica de <i>Erythrina</i> (FM1). A seta indica que a amostra <i>a</i> faz parte do grupo I
Figura 45: Gráfico dos loadings das componentes principais (a) 1, (b) 2 e (c) 3 dos dados cromatográficos da fração orgânica de <i>Erythrina</i> (FM1)
Figura 46: Cromatogramas tratados das frações orgânicas de <i>Erythrina</i> separados com base nos grupos formados pela ACP
Figura 47: Dendrograma baseado nos dados cromatográficos da fração orgânica de <i>Erythrina</i> obtidos com o método de Ward
Figura 48: Gráfico dos rendimentos da fração orgânica com os valores de escores da (a) CP2 e (b) CP391
Figura 49: Cromatogramas da fração básica no ponto central (replicatas), nas fases móveis (a) 1, (b) 2 e (c) 392

Figura 50: Gráfico dos escores das componentes principais (a) 1 e 2 e (b) 2 e 3 dos dados cromatográficos da fração básica de Erythring (FM1). A seta indica que a amostra eda faz parte do Figura 51: Gráfico dos loadings das componentes principais (a) 1, (b) 2 e (c) 3 dos dados cromatográficos da fração básica de Erythrina (FM1)94 Figura 52: Cromatogramas das frações básicas de Erythrina separados com base nos grupos formados pela ACP......95 Figura 53: Dendrograma baseado nos dados cromatográficos da fração básica de Erythrina obtido pelo método de Ward......96 Figura 54: Gráfico dos valores de escore da CP1 e dos rendimentos da fração básica normalizados97 Figura 55: Gráfico tridimensional dos valores de absorvância em função do tempo de corrida e do comprimento de onda do extrato bruto em etanol puro98 Figura 56: Variância explicada para cada número total de fatores dos modelos Tucker3 para os dados cromatográficos dos extratos brutos de Erythrina (FM1)100 Figura 57: Loadings dos fatores 1, 2 e 3 do modo A do modelo Tucker3 para os dados tridimensionais dos extratos brutos de Erythrina (FM1).....102 Figura 58: Loadings dos fatores 1, 2 e 3 do modo B do modelo Tucker3 para os dados tridimensionais dos extratos brutos de Erythrina (FM1).....103 Figura 59: Escores dos fatores 1, 2 e 3 do modo C do modelo Tucker3 para os dados tridimensionais dos extratos brutos de Erythrina (FM1).....104 Figura 60: Cromatogramas dos extratos brutos de Erythrina que (a) possuem e que (b) não possuem acetona na composição do solvente extrator no intervalo de tempo de corrida de 1,5 a 7 Figura 61: Espectros referentes ao pico com tempo de retenção de 3,69 min dos extratos brutos de Erythrina que (a) possuem e que (b) não possuem acetona na composição do solvente extrator. 106 Figura 62: Cromatogramas dos extratos brutos de Erythrina que (a) possuem e que (b) não possuem diclorometano na composição do solvente extrator no intervalo de tempo de corrida de 7 Figura 63: Espectros referentes ao pico com tempo de retenção de 12,19 min dos extratos brutos de Erythrina que (a) possuem e que (b) não possuem diclorometano na composição do solvente Figura 64: Cromatogramas dos extratos brutos de Erythrina que possuem (a) alta e (b) baixa intensidade do pico com tempo de retenção de 4,05 min, no intervalo de tempo de corrida de 1,5 a

Figura 65: Espectros referentes ao pico com tempo de retenção de 4,07 min dos extratos brutos de *Erythrina* que (a) possuem e que (b) não possuem etanol na composição do solvente extrator ... 109

Figura 66: Espectros referentes ao pico com tempo de retenção de 2,73 min dos extratos brutos de <i>Erythrina</i> que apresentaram (a) menor, (b) média e (c) maior intensidade de absorção no comprimento de onda de 268 nm
Figura 67: Variância explicada para cada número total de fatores dos modelos Tucker3 para os dados cromatográficos da fração orgânica de <i>Erythrina</i> (FM1)113
Figura 68: Loadings dos fatores 1, 2, 3 e 4 do modo A do modelo Tucker3 para os dados tridimensionais da fração orgânica de <i>Erythrina</i> (FM1)114
Figura 69: Loadings do fator 1 do modo B do modelo Tucker3 para os dados tridimensionais da fração orgânica de <i>Erythrina</i> (FM1)115
Figura 70: Escores dos fatores 1, 2, 3 e 4 do modo C do modelo Tucker3 para os dados tridimensionais da fração orgânica de <i>Erythrina</i> (FM1)116
Figura 71: Cromatogramas da fração orgânica de <i>Erythrina</i> no intervalo de tempo de corrida de 4,5 a 6,5 min
Figura 72: Espectros referentes aos picos com tempo de retenção de (a) 4,80 e (b) 5,80 min das amostras da fração orgânica de <i>Erythrina</i> 118
Figura 73: Cromatogramas da fração orgânica de <i>Erythrina</i> no intervalo de tempo de corrida de (a) 2,5 a 3,0 min e (b) 9,2 a 10,6 min119
Figura 74: Espectros referentes aos picos com tempo de retenção de (a) 2,72 min, (b) em torno de 9,75 e 10,20 min das amostras da fração orgânica de <i>Erythrina</i>
Figura 75: Espectros normalizados referentes aos picos com tempo de retenção em torno de 9,75 e 10,20 min das amostras da fração orgânica de <i>Erythrina</i> , separados com base no perfil espectral121
Figura 76: Variância explicada para cada número total de fatores dos modelos Tucker3 para os dados cromatográficos da fração básica de <i>Erythrina</i> (FM1)123
Figura 77: Loadings dos fatores 1, 2 e 3 do modo A do modelo Tucker3 para os dados tridimensionais da fração básica de <i>Erythrina</i> (FM1)125
Figura 78: Loadings dos fatores 1 e 2 do modo B do modelo Tucker3 para os dados tridimensionais da fração básica de <i>Erythrina</i> (FM1)126
Figura 79: Escores dos fatores 1, 2 e 3 do modo C do modelo Tucker3 para os dados tridimensionais da fração básica de <i>Erythrina</i> (FM1)127
Figura 80: Cromatogramas da fração básica de <i>Erythrina</i> no intervalo de tempo de corrida de 1,5 a 10 min
Figura 81: Espectros referentes aos picos com tempo de retenção em torno de 2,55, 3,69, 4,77 e 5,81 min das amostras da fração básica de <i>Erythrina</i>

Figura 82: Cromatogramas da fração básica de <i>Erythrina</i> no intervalo de tempo de corrida de	e 4 a 7
min separados com base nos picos com tempo de retenção em (a) 4,77 e 5,76 min e (b) 5,12	e 6,26
min	130
Figura 83: Espectros referentes aos picos com tempo de retenção em torno de 5,12 e 6,26 m	in das
amostras da fração básica de <i>Erythrina</i>	131

LISTA DE ABREVIATURAS

ANOVA	Análise da variância
a	Acetona
AHA	Análise hierárquica de agrupamentos
CLAE	Cromatografia líquida de alta eficiência
СР	Componente principal
d	Diclorometano
da	Mistura diclorometano e acetona
DAD	Detector por arranjo de diodos
dh	Mistura diclorometano e hexano
dha	Mistura diclorometano, hexano e acetona
е	Etanol
еа	Mistura etanol e acetona
ed	Mistura etanol e diclorometano
eda	Mistura etanol, diclorometano e acetona
edh	Mistura etanol, diclorometano e hexano
eh	Mistura etanol e hexano
eha	Mistura etanol, hexano e acetona
FM	Fase móvel
h	Hexano
H ₂ O	Água
ha	Mistura hexano e acetona
MeOH	Metanol
min	Minuto
PCA	Principal component analysis (Análise de componentes principais)
r1	Replicata 1
r2	Replicata 2
r3	Replicata 3
r4	Replicata 4
r5	Replicata 5
u.a.	Unidade arbitrária
UV	Ultra-violeta

1. Introdução

A utilização das plantas ou partes de plantas na cura ou prevenção de doenças é praticada há tempos imemoriáveis. Estima-se que cerca de 25% das drogas prescritas no mundo todo sejam derivadas de plantas. Na lista de medicamentos essenciais da Organização Mundial da Saúde, 11% das 252 drogas são exclusivamente originárias de plantas¹. Dos 847 fármacos de baixa massa molar lançados no mercado entre 1981 e 2006, 43 eram produtos naturais, 232 eram produzidos por hemi-síntese a partir de produtos naturais e 572 obtidos por síntese total, dos quais 262 possuíam um grupo farmacofórico inspirado em produtos naturais ou poderiam ser considerados análogos de produtos naturais².

Nos últimos anos constatou-se que muitos avanços foram feitos em termos de descobertas de princípios ativos de plantas medicinais³⁻⁵. Porém, o desenvolvimento de medicamentos baseados nestes princípios ativos requer o conhecimento dos aspectos químicos, biológicos, genéticos e agronômicos da planta. O perfil químico e a produção dos metabólitos secundários são afetados por muitas variáveis, de caráter fisiológico e genético e principalmente pelas condições do ambiente ao redor (clima, tipo de solo, disponibilidade de nutrientes, etc). O conteúdo dos metabólitos secundários também é dependente de outros fatores como a época de colheita, forma de secagem, armazenagem e extração⁶⁻⁸. As etapas que levam da planta ao fitomedicamento, partindo da coleta até o isolamento do princípio ativo são mostradas na Figura 1¹.



Figura 1: Etapas envolvidas no desenvolvimento de fitomedicamentos.

O passo mais importante da sequência mostrada na Figura 1 é o processo de extração. No estudo de plantas medicinais, os objetivos do pesquisador podem ser os mais variados, além de buscar o isolamento do princípio ativo, o pesquisador pode querer diferenciar plantas do mesmo gênero⁹ ou espécie¹⁰, discriminar plantas com relação à origem geográfica¹¹, encontrar marcadores para uma dada espécie¹², sendo que neste caso, é desejável, mas não obrigatório que o próprio

princípio ativo seja o marcador¹³, enfim, o sucesso em qualquer um desses casos vai depender do processo de extração. A escolha do solvente extrator é que vai determinar a informação química obtida, tanto qualitativa quanto quantitativa.

Mas, em muitos casos, essa etapa não é explorada de forma adequada¹⁴. Geralmente, os pesquisadores desenvolvem todo este processo utilizando como extrator um único solvente puro^{15,16}, em seguida são feitas as etapas de fracionamento, purificação e isolamento¹⁷⁻¹⁹.

Uma alternativa interessante e extremamente segura é escolher o solvente extrator ótimo através de um planejamento estatístico de misturas²⁰. Neste tipo de planejamento, o produto final ou a resposta obtida será dependente das porcentagens ou proporções dos componentes da mistura. O uso desse tipo de planejamento possibilita o pesquisador a descobrir as interações entre os solventes utilizados, sejam elas sinérgicas ou antagônicas, e a usar essas informações para maximizar ou minimizar uma resposta, ou ainda descobrir que não existe uma interação significativa entre os solventes, de forma que a mistura não altera o resultado.

Uma das formas mais completas de se avaliar um extrato de planta medicinal é através da sua impressão digital cromatográfica²¹. Esta técnica foi introduzida pela Organização Mundial da Saúde e aceita por outras autoridades governamentais ou privadas, para auxiliar tanto na autenticação da qualidade de ervas medicinais²² como na identificação de ervas ou de seus produtos, pois permite obter o perfil químico não só dos marcadores, mas também dos compostos desconhecidos²³.

Outro desafio é conseguir separar as informações obtidas. Na análise da impressão digital, o perfil cromatográfico inteiro é utilizado^{24,25}, o que significa que em um único planejamento de misturas haveria um número muito alto de variáveis. Uma das alternativas para reduzir o número de respostas a serem analisadas e interpretar seus efeitos é utilizar métodos estatísticos de análise multivariada²⁶.

A aplicação de métodos quimiométricos nestes casos aumenta consideravelmente a qualidade da impressão digital obtida. Os métodos disponíveis permitem desde o alinhamento dos picos e correções de linha de base²⁷, até a discriminação de dezenas de amostras a partir de um número extremamente grande de variáveis^{21,28}. Hoje, a quimiometria não é mais uma opção na análise de matrizes complexas como o rastreamento da impressão digital de plantas medicinais, mas sim uma ferramenta essencial²⁹.

Uma das plantas medicinais, nunca antes explorada em estudo de impressão digital cromatográfica, mas muito utilizada na medicina alternativa é a *Erythrina speciosa* Andrews³⁰. O gênero *Erythrina* contém cerca de 110 espécies, que são vulgarmente conhecidas como "Mulungu", são árvores bastante ornamentais quando floridas, sendo muito utilizadas em paisagismo. O extrato das folhas, cascas e das raízes de várias espécies são utilizados na medicina popular para o tratamento de diversas doenças, como disenteria, asma, dor estomacal, infertilidade feminina e principalmente infecções microbianas³⁰⁻³², muitos desses efeitos são atribuídos aos chamados alcalóides de eritrina^{31,33}.

Dessa forma, este trabalho teve como objetivos:

- Realizar um planejamento experimental de misturas do tipo Centróide-Simplex com
 4 componentes na otimização do solvente extrator para a planta *Erythrina speciosa* Andrews.
- Avaliar os dados gravimétricos dos extratos brutos e também das frações através da metodologia de superfície de resposta para otimização dos rendimentos.
- Avaliar o desempenho de fases móveis com diferentes forças cromatográficas na separação dos compostos presentes nos extratos brutos e nas frações.
- Interpretar os resultados cromatográficos obtidos com a aplicação da análise de componentes principais e análise hierárquica de agrupamentos para discriminar as diferentes misturas com relação às variáveis.
- Interpretar os resultados CLAE-DAD de segunda ordem com a aplicação do método Tucker3 para discriminar as amostras com relação às variáveis cromatográficas e espectroscópicas.

2. Cromatografia

Nas análises por cromatografia líquida, na grande maioria das vezes, são empregadas misturas de solventes para facilitar ou melhorar a separação de solutos entre a fase estacionária e a fase móvel. Existem muitas razões para se utilizar misturas de solventes ao invés do solvente puro, especialmente em cromatografia líquida de alta eficiência (CLAE). Uma das mais importantes é a versatilidade, pois é possível variar a seletividade e aumentar a eficiência da separação de acordo com as necessidades.

A composição da mistura empregada é um grau de liberdade adicionado à escolha de diversos solventes disponíveis. Os solutos que serão separados são afetados pela composição da mistura e serão distribuídos de maneira diferente entre a fase móvel e a fase estacionária. Outra razão é que as misturas podem ser eluídas de forma isocrática ou por gradiente, fazendo com que o tempo necessário para uma separação eficiente seja reduzido e melhorando a resolução.

Outra razão implícita, extremamente importante na cromatografia líquida em fase reversa, é que um dos componentes, geralmente o menos polar (solvente orgânico) é seletivamente adsorvido pela fase estacionária, modificando sua estrutura de forma favorável à separação. Usualmente a fase móvel é composta de misturas aquosas, binárias ou ternárias, sendo que o modificador orgânico escolhido e a sua proporção na mistura representam a peça chave na eficiência da separação.

A modificação das características físicas da fase móvel, como a viscosidade, também pode requerer o uso de misturas ao invés de solventes puros³⁴.

Existem centenas de solventes disponíveis para definir os componentes da fase móvel. A seleção desses solventes, usualmente é afetada pelas características do solvente relacionadas à alta pureza, a não decomposição da fase estacionária, a compatibilidade com o detector empregado, a dissolução da amostra sem decompor seus componentes, baixa viscosidade e ponto de ebulição e miscibilidade completa em mistura. A disponibilidade comercial desses solventes na pureza adequada e a um preço razoável, também são fatores importantes³⁵.

Uma alternativa para a escolha da fase móvel é usar o esquema de polaridade proposto por Snyder. A classificação da propriedade do solvente por Snyder em termos de polaridade e interações químicas específicas é bem estabelecida e é descrito em livros texto de química

analítica³⁶. Além do índice de polaridade ele determinou outros três parâmetros experimentalmente, a basicidade (*xe*), a acidicidade (*xd*) e o momento dipolar (*xn*). A vantagem do esquema de Snyder é o agrupamento de solventes de acordo com o tipo de interação que eles possuem. Solventes quimicamente semelhantes que possuem parâmetros de seletividade similares fazem parte do mesmo grupo, embora possam apresentar diferentes valores de polaridade. A Tabela 1 mostra a discriminação desses grupos^{37,38}.

Tabela 1: Solventes dos grupos determinados por Snyder.

Grupos	Solventes
Ι	Éteres alifáticos (Éter dietílico)
11	Álcoois alifáticos (Metanol, Etanol, Octanol, 1-Propanol)
<i>III</i>	Derivados de piridina, THF, sulfóxidos
IV	Glicóis, ácido acético
V	Diclorometano
VI	Ésteres alifáticos, cetonas, nitrilas, dioxano (Acetonitrila)
VII	Hidrocarbonetos aromáticos, compostos aromáticos halogenados, éteres Aromáticos, compostos nitrogenados (Tolueno)
VIII	Água

A Figura 2 mostra os três parâmetros de seletividade representados em um diagrama triangular. O triângulo interno (amarelo) mostra os solventes utilizados na composição da fase móvel.



Figura 2: Triângulo de seletividade de Snyder para solventes usados em cromatografia.

Devido ao fato de que solventes do mesmo grupo possuem parâmetros de seletividade muito próximos, se uma separação cromatográfica não puder ser obtida com uma mistura em particular formada por dois solventes do mesmo grupo, será necessária a substituição de um dos solventes por um solvente de outro grupo.

Na cromatografia líquida em fase reversa, o solvente mais empregado é a água (grupo VIII), devido à sua abundância, disponibilidade, baixo custo, atoxicidade e segurança, bem como sua habilidade em dissolver uma grande variedade de substâncias, incluindo eletrólitos e substâncias orgânicas polares. O segundo solvente, ou o modificador orgânico é escolhido de um dos grupos restantes, sendo que os mais utilizados são o metanol (grupo II), a acetonitrila (grupo VI) e o tetraidrofurano (grupo III) todos com baixa viscosidade (η), 0,55, 0,38 e 0,55, respectivamente.

A composição da fase móvel pode afetar a separação devido à força do solvente e à seletividade. A força do solvente ou força cromatográfica de uma mistura eluente determina a variação dos fatores de retenção no qual todos os solutos de uma amostra são eluídos e dependem principalmente da fração de água na mistura.

A força do solvente de uma fase móvel composta por água e mais dois modificadores orgânicos A e B pode ser calculada pela seguinte equação:
$$S_T = \phi_A S_A + \phi_B S_B \tag{1}$$

onde ϕ_A e ϕ_B representam a fração de volume dos modificadores A e B na mistura e S_A e S_B referem-se à força dos solventes puros A e B, respectivamente.

Os parâmetros de força do solvente propostos para os quatro solventes mais utilizados na cromatografia líquida em fase reversa, isto é, para a água, metanol, acetonitrila e tetraidrofurano são 0, 2,6, 3,1 e 4,4 respectivamente³⁶.

3. Métodos quimiométricos

A aplicação de métodos quimiométricos a dados de retenção de sistemas cromatográficos, permite obter informações sobre o comportamento tanto das substâncias como do próprio sistema. Com o objetivo de descobrir como esses resultados de retenção de diferentes sistemas se relacionam, foram usados os métodos de análise de componentes principais³⁹, análise hierárquica de agrupamentos⁴⁰ e o método Tucker3⁴¹. Neste caso o sistema estudado é composto por diferentes extratos de plantas preparados com base num planejamento experimental de misturas do tipo Centróide-Simplex²⁰ com quatro componentes. Este capítulo apresenta de maneira simplificada, a descrição dos métodos estatísticos utilizados, sendo que para informações mais detalhadas o leitor pode consultar as referências.

3.1 Planejamento de misturas do tipo Centróide-Simplex^{20,42,43}

Uma mistura é um sistema multicomponente, no qual a soma de todos os componentes permanece constante para cada amostra, ou seja, a soma das proporções dos componentes da mistura deve ser 100%. A resposta medida para uma mistura é determinada pela proporção dos seus componentes e não por valores absolutos.

Num planejamento Centróide-Simplex com z componentes, o número de pontos distintos é $2^z -1$. Os pontos distintos do planejamento correspondem a z permutações do tipo (1,0,0,...,0) ou componentes puros, permutações do tipo (z/2) ou misturas binárias, do tipo (z/3) ou misturas ternárias e assim por diante até o ponto central (1/z, 1/z,..., 1/z) ou z-ésima mistura. Em outras palavras, este tipo de planejamento consiste apenas de misturas onde os componentes encontram- se em proporções iguais.

Em um planejamento Centróide-Simplex com 4 componentes, tem-se 15 misturas diferentes; 4 misturas de componentes puros, 6 misturas binárias, 4 misturas ternárias e 1 mistura quaternária.

Após a realização dos experimentos, os dados obtidos ou as respostas do planejamento são coletados e um polinômio com um número de termos ou parâmetros a estimar, que pode ser no

máximo igual ao número de pontos do planejamento é ajustado. Em geral o polinômio para um sistema com 4 componentes é definido pela expressão:

$$\hat{y} = \sum_{i=1}^{4} b_i x_i + \sum_{i < j=2}^{4} b_{ij} x_i x_j + \sum_{i < j < k=3}^{4} b_{ijk} x_i x_j x_k + \dots + b_{1234} x_1 x_2 x_3 x_4$$
(2)

onde, o parâmetro \mathbf{b}_i representa a resposta esperada para o componente puro *i* e é chamado de coeficiente linear do componente *i*, \mathbf{b}_{ij} é o coeficiente de interação entre os componentes *i* e *j*, o parâmetro \mathbf{b}_{ijk} é o coeficiente de interação entre os componentes *i*, *j* e *k* e \mathbf{b}_{1234} é o coeficiente de interação entre os componentes *i*, *j* e *k* e \mathbf{b}_{1234} é o coeficiente de interação entre os componentes 1, 2, 3 e 4.

Esses parâmetros (**b**) e seus respectivos erros padrão (V(**b**)) podem ser calculados pelas seguintes equações matriciais:

$$\boldsymbol{b} = \left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}\mathbf{X}^{T}\boldsymbol{y}$$
(3)

$$V(\boldsymbol{b}) = \left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}s^{2}$$
(4)

onde s^2 é a variância em y, **X** é a matriz contendo os pontos do planejamento, **y** é o vetor das respostas, o sobrescrito t indica a transposta da matriz e⁻¹ representa a inversão da matriz.

Após a obtenção de todos os parâmetros, é necessário testar a significância da regressão. O método mais usado para se avaliar numericamente a qualidade do ajuste de um modelo é a análise da variância, ou ANOVA. A análise da variância para um planejamento de misturas com replicatas é mostrado na Tabela 2.

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática
Modelo	$SQ_R = \sum_{i}^{m} \sum_{j}^{n_i} (\hat{y}_i - \overline{y})^2$	<i>p</i> -1	$MQ_R = SQ_R/p-1$
Resíduos	$SQ_r = \sum_{i}^{m} \sum_{j}^{n_i} (y_{ij} - \hat{y}_i)^2$	n-p	$MQ_r = SQ_r/n - p$
Falta de ajuste	$SQ_{faj} = \sum_{i}^{m} \sum_{j}^{n_i} (\hat{y}_i - \overline{y}_i)^2$	m-p	$MQ_{faj} = SQ_{faj} / m - p$
Erro puro	$SQ_{ep} = \sum_{i}^{m} \sum_{j}^{n_{i}} (y_{ij} - \overline{y}_{i})^{2}$	n-m	$MQ_{ep} = SQ_{ep} / n - m$
Total	$SQ_T = \sum_{i}^{m} \sum_{j}^{n_i} (y_{ij} - \overline{y})^2$	<i>n</i> -1	

Tabela 2. Análise da variância para um planejamento de misturas com replicatas.

Nesta Tabela, \hat{y}_i é o valor previsto de *i*, \overline{y} é a média global, y_{il} é a *j*-ésima resposta obtida para o *i*-ésimo ensaio, \overline{y}_i é a média das respostas observadas no nível *i*, n_i é o número de repetições no nível *i*, *m* é o número de níveis distintos das variáveis independentes, *n* é o número total de observações e *p* é o número de parâmetros do modelo ajustado.

Agora, a significância da regressão pode ser testada através da comparação da razão MQ_R/MQ_r com o valor tabelado da distribuição F com o número apropriado de graus de liberdade, se $MQ_R/MQ_r > F_{tab}$, descarta-se a hipótese nula e acredita-se que as variáveis se relacionam de acordo com o modelo ajustado. A falta de ajuste é testada através da comparação da razão MQ_{faj}/MQ_{ep} com o valor tabelado da distribuição F com o número apropriado de graus de liberdade liberdade. Valores altos de MQ_{faj}/MQ_{ep} significam presença de falta de ajuste.

Todos estes cálculos são feitos de forma extremamente rápida com o uso de um computador, a maior dificuldade no entanto, é a representação gráfica dos resultados. Em alguns casos as equações podem ser simplificadas, sendo possível visualizar as curvas de nível fazendo cortes transversais em direções apropriadas do espaço tridimensional, mas isso se torna ainda mais complicado quando há várias respostas para o sistema.

Algumas respostas devem ser maximizadas, enquanto outras devem ser minimizadas. Em muitos casos, existe uma correlação negativa entre as variáveis, ou seja, aumentando uma resposta tem-se o efeito contrário na outra, o que torna a situação ainda mais complicada. Muitas aproximações foram feitas para tentar solucionar este problema. Uma delas é combinar todas as respostas em um único parâmetro através da função de desejabilidade, isto é, usar a metodologia de otimização simultânea proposta por G. C. Derringer e R. Suich^{20,44,45}.

A desejabilidade é fundamentada na idéia de "qualidade" de um produto ou processo que possua características múltiplas, sendo que o fato de uma delas estar fora dos limites desejados é inaceitável. Na transformação unilateral, a desejabilidade individual d_i aumenta com o aumento de y_i (da equação 2) quando o objetivo for a maximização ou com a diminuição de y_i quando o objetivo for a minimização. Nesse caso a função de desejabilidade da resposta é definida por:

$$d_{i} = \begin{cases} 0 & y_{i} \leq y_{i^{*}} \\ \left[\frac{y_{i} - y_{i^{*}}}{y_{i}^{*} - y_{i^{*}}}\right]^{r} & y_{i^{*}} \leq y_{i} \leq y_{i}^{*} \\ 1 & y_{i} \geq y_{i}^{*} \end{cases}$$
(5)

onde y_{i^*} é o menor valor aceitável de y_i e y_i^* o maior valor de y_i , na verdade, valores acima de y_i^* possuem pouco mérito adicional, ou seja, adiciona pouco à qualidade do processo ou ao controle de qualidade. O expoente r usado na transformação é definido pelo analista. No caso de maximização deve-se escolher um alto valor de r (r = 10) se for desejável que o valor de y_i cresça rapidamente acima do valor de y_{i^*} . Se escolher um r com valor pequeno (r = 0,1), significa que qualquer valor de y_i acima de y_{i^*} é tão desejável quanto qualquer outro valor de y_i acima de y_{i^*} , ou seja, não é de fundamental importância. Na equação, a função de desejabilidade d_i , para cada resposta é definida com valores restritos ao intervalo [0,1], onde zero significa um valor inaceitável e 1 o valor mais desejável. As desejabilidade global, normalmente dada pela média geométrica das u desejabilidades individuais:

$$D = \sqrt[n]{d_1 d_2 \dots d_n} \tag{6}$$

O valor de *D* aumenta quando o balanço das propriedades torna-se mais favorável, se qualquer $d_i = 0$ então D = 0, pois o produto global será inaceitável.

3.2 Pré-processamento

Muitas vezes, antes da análise dos dados através de métodos matemáticos, os dados originais são submetidos a algum tipo de pré-processamento^{48,51,52}. Alguns testes devem ser realizados para saber qual o melhor tratamento dos dados.

Existem dois tipos básicos de pré-processamento, dependendo de como são operados. A ferramenta de pré-processamento por linhas, opera sobre as amostras, enquanto que a ferramenta de pré-processamento por colunas opera sobre as variáveis. Dentre os tipos de pré-processamento por linhas está a normalização⁴⁸.

Neste trabalho, foram usados três tipos de pré-processamento por linhas, a normalização para 1, por comprimento do vetor e por área unitária e um tipo de pré-processamento por colunas, a centragem dos dados na média.

Na normalização para 1, cada variável de uma amostra é dividida pelo valor da variável que apresentou o valor mais alto, Equação 7. Após este procedimento para todas as amostras, a intensidade máxima em todas as amostras será igual a 1.

$$\mathbf{x'}_{ij} = \frac{\mathbf{x}_{ij}}{\mathbf{x}_{máx}} \tag{7}$$

Na equação, x'_{ii} refere-se aos dados normalizados da amostra *i* e variável *j*.

A normalização por área unitária é feita dividindo cada valor original pelo somatório de todos os valores absolutos das medidas, Equação 8. Este procedimento para os dados cromatográficos tem a finalidade de corrigir a variação do volume da amostra injetado no cromatógrafo, sem destruir a proporcionalidade das substâncias orgânicas,

13

$$x'_{ij} = \frac{x_{ij}}{\sum_{j=1}^{p} |x_j|}$$

Na normalização por comprimento do vetor, cada variável de uma amostra é dividida pelo comprimento do vetor da amostras, que é dado pela raiz quadrada do somatório do quadrado de todas as variáveis, Equação 9. A finalidade é remover a variação sistemática, normalmente associada com a quantidade total da amostra.

(8)

$$x'_{ij} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^{p} x_j^2}}$$
(9)

A centragem dos dados na média⁴⁸, é feita subtraindo de cada variável o valor médio de todas as variáveis, Equação 10. Geometricamente, equivale a fazer uma translação do sistema de eixos ao longo de um vetor para o centro do conjunto de dados.

$$x'_{ij} = x_{ij} - \overline{x}_j \tag{10}$$

3.3 Análise de componentes principais (PCA)

A PCA consiste essencialmente em transformar as coordenadas das amostras em outro sistema de eixos mais conveniente para análise de dados. É uma forma de identificar padrões nos dados e expressá-los de um modo que se possa enxergar as suas diferenças e similaridades, reduzindo a dimensionalidade sem perder muita informação.

Em outras palavras, reduz o número de variáveis originais a poucas variáveis chamadas de componentes principais (CP). A primeira CP é escolhida na direção de maior variância dos dados. A segunda CP é definida de tal maneira que seja ortogonal a primeira e representa a variância

máxima que não foi explicada pela primeira CP, ou seja, as componentes principais são obtidas em ordem decrescente de máxima variância⁴⁶⁻⁴⁸.

Matematicamente, cada CP pode ser descrita como uma combinação linear das variáveis originais onde a importância de cada uma dessas variáveis é dada pelos loadings. Esta combinação linear fornece ainda para cada objeto, valores chamados escores, que são as coordenadas no novo sistema de eixos.

A ACP, pode ser definida tanto no espaço das amostras como das variáveis, e a decomposição é calculada através da diagonalização da matriz de associação (AA^{t}) ou da matriz de covariância ($A^{t}A$), respectivamente³⁹. A matriz **A** (*n* x *k*), formada por *n* objetos e *k* variáveis é decomposta no produto de duas matrizes menores:

$$\mathbf{A}(n \times k) = \mathbf{T}(n \times q) \times \mathbf{P}^{\mathsf{t}}(q \times k) + \mathbf{E}(n \times k)$$
(11)

onde **P** é a matriz dos "loadings" (pesos) das componentes principais, **T** é a matriz dos escores das componentes principais, e **E** é a matriz dos resíduos. O sobrescrito *t* indica a transposta da matriz e q é um escalar que indica o número de componentes principais que descreve a maior parte da variância dos dados.

Após os cálculos, dois tipos de gráficos são obtidos, o gráfico dos escores, que fornece informações sobre os objetos, no caso deste trabalho as amostras do planejamento, e o gráfico dos loadings que representa a importância e a correlação das variáveis, neste caso as intensidades de absorbância dos picos cromatográficos.

Uma das etapas cruciais na análise de componentes principais é determinar o número de componentes principais que devem ser usadas. É equivalente a definir a quantidade de informação que é relevante para interpretar os dados. Porém, se a quantidade de informação ou variação explicada é suficiente ou não, vai depender do problema em si. A escolha de um número muito grande de componentes pode levar à interpretação de ruídos e o uso de poucas componentes pode ocasionar a exclusão de informações pertinentes⁴⁹.

Após a escolha do número de componentes principais, o resultado pode ser interpretado. No gráfico dos escores, é possível identificar as diferenças e as similaridades entre as amostras, através dele é definida a região no espaço que delimita a formação de alguns grupos. Geralmente,

15

os grupos são definidos pela sua posição em três regiões, de escores positivos, negativos e também a região intermediária que contem zero. Em alguns casos, pode-se ter somente valores positivos ou negativos de escores e a definição é feita da mesma maneira, região com valores de escore altamente positivo, menos positivo e intermediário. Esses grupos vão ser caracterizados pelo peso de algumas variáveis, que podem ser identificadas através do gráfico dos loadings. Amostras com valores de escore positivo sofrem influência das variáveis com loadings de sinal positivo. Loadings com mesmo sinal influenciam a componente principal na mesma direção, ou seja, se todos os pesos forem positivos, considera-se que a componente principal é um indicador de tamanho, ou concentração, dependendo do tipo de variável estudada.

Para auxiliar na interpretação dos pesos, os eixos das componentes principais de **P** ainda podem ser rotacionados, buscando uma matriz que tenha significado químico. A rotação varimax⁵⁰ gira rigidamente os eixos dos fatores, de modo que eles se aproximem o máximo dos vetores mais divergentes no espaço *q*-dimensional. O procedimento da rotação maximiza a variância dos pesos ou loadings dos fatores em cada vetor, sob a restrição de que os fatores permaneçam ortogonais. A matriz **P** é rodada, produzindo uma nova matriz, **F**:

$$\mathbf{F}(k \times q) = \mathbf{P}(k \times q) \times \mathbf{R}(q \times q)$$
(12)

onde **R** é a matriz de rotação e **F** é a matriz dos pesos dos fatores varimax. Cada linha de **F** corresponde ao peso de uma das variáveis originais e cada coluna de **F**, a um fator.

Geralmente este procedimento é adotado para uma melhor interpretação dos resultados pois os pesos se concentram em algumas poucas variáveis, resultando em um novo sistema de eixos menos complexos, facilitando possíveis interpretações físicas de cada fator⁵¹.

3.4 Análise hierárquica de agrupamentos (AHA)

A AHA⁴⁰ descreve a estrutura dos dados interligando as amostras por suas associações, revelando os agrupamentos naturais existentes no conjunto de dados. Com base na informação das variáveis medidas, a classificação é feita de tal maneira, que mesmo um pequeno grupo formado na separação do conjunto de dados possa ser inteiramente incluído num grupo maior formado na separação consecutiva.

Graficamente, a hierarquia pode ser representada na forma de um gráfico bidimensional chamado de dendrograma, onde as amostras semelhantes, segundo as variáveis escolhidas são agrupadas entre si.

Existem muitas maneiras de procurar agrupamentos no espaço *p*-dimensional^{26,53}. A maneira matematicamente mais simples consiste em agrupar os pares de pontos (amostras) que estão mais próximos, usando a distância Euclidiana, d_{ik} , definida na equação abaixo, e substituí-los por um novo ponto, localizado na metade da distância entre eles:

$$d_{ik} = \left[\sum_{j=1}^{p} \left(x_{ij} - x_{kj} \right)^2 \right]^{1/2}$$
(13)

O procedimento descrito é repetido até que todos os pontos sejam agrupados em um só ponto. O índice de similaridade é calculado com base na equação:

$$S_{ik} = \left[1 - \frac{d_{ik}}{\left(d_{ik}\right)_{máx}}\right]$$
(14)

onde S_{ik} é o índice de similaridade entre os pontos *i* e *k*, d_{ik} a distância Euclidiana e $(d_{ik})_{max}$ é a distância máxima entre dois pontos de um conjunto de dados no espaço *p*-dimensional.

Após o cálculo das distâncias entre todos os pares de amostras, as duas amostras mais similares são agrupadas. Uma vez que um grupo é ligado a outro grupo, eles formarão um novo grupo. Após essa verificação, a distância deste novo grupo com relação a todos os outros grupos é determinada e novamente, a menor distância entre os grupos é analisada e outra ligação é formada. Esse processo continua até que todas as amostras e grupos estejam ligados.

A distância entre dois grupos existentes e um novo grupo pode ser calculada de várias maneiras. Neste trabalho foram utilizados, o método de agrupamento pareado igualmente ponderado⁵³ e o método de Ward⁵⁴.

No método de agrupamento pareado igualmente ponderado, a distância entre dois grupos é calculada como a distância média entre todos os pares de objetos nos dois diferentes grupos. O método de Ward ou método de variância mínima usa uma aproximação da análise de variância para avaliar a distância entre os grupos. A estrutura hierárquica completa e a estimativa da quantidade perdida associada a cada estágio de agrupamento pode ser obtida de acordo com a Equação 15. Resumidamente, este método tenta minimizar a soma dos quadrados de quaisquer dois agrupamentos que possam ser formados em cada passo. Este método é muito eficiente, porém tende a formar agrupamentos bem pequenos.

$$ESS = \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2$$
(15)

onde *ESS* é a soma quadrática do erro, n é o número de amostras e x_i é o valor do *i*-ésimo indivíduo.

3.5 Tucker3^{55,56}

O modelo Tucker3 é uma generalização da ACP para dados multidimensionais⁵⁷⁻⁵⁹. Uma peculiaridade deste método é que o arranjo do tensor \underline{X} é decomposto num certo número de componentes, que pode ser diferente em cada uma das dimensões. Essa propriedade é muito útil em casos onde um dos modos possui um número muito menor de níveis com relação ao número de níveis dos outros modos⁶⁰.

O modelo Tucker3 para dados tridimensionais é aplicado a um tensor tridimensional <u>X</u> (I x J x K) que é decomposto em três matrizes de loadings, A (I x L), B (J x M) e C (K x N) cujos elementos são a_{il} , b_{jm} e c_{kn} , respectivamente. O número de colunas ortogonais ou número de fatores escolhidos em cada um dos modos A, B e C é dado por L, M e N. Esse modelo pode ser escrito da seguinte maneira⁶⁰:

$$x_{ijk} = \sum_{l=1}^{L} \sum_{m=1}^{M} \sum_{n=1}^{N} a_{il} b_{jm} c_{kn} g_{lmn} + e_{ijk}$$
(16)

na expressão, e_{ijk} é o elemento residual do tensor <u>E</u> (I x J x K) e g_{lmn} é um elemento do tensor ou core <u>G</u> (L x M x N). Especificamente, os elementos do core representam a importância das interações entre os fatores, ou seja, a sua magnitude é proporcional à importância da interação entre o *l-ésimo* vetor de loadings do primeiro modo, o *m-ésimo* vetor de loadings do segundo modo e o *n-ésimo* vetor de loadings do terceiro modo (um valor próximo de zero indica uma interação insignificante⁶¹).

A medida quantitativa da importância de cada elemento do core é dada pela chamada "variância do core", que é obtido dividindo g^2_{lmn} pela soma quadrática total dos elementos do core. O ajuste de um modelo é avaliado pela variância explicada, que é dada pela razão entre a soma quadrática do modelo e a soma quadrática total⁶⁰.

A determinação da complexidade do modelo, ou seja, do número de fatores em cada modo é uma das etapas mais importantes para a análise dos dados, não existindo para ela um critério absoluto⁶¹. Várias ferramentas podem ser usadas para esta finalidade, citando como exemplo, a variância explicada pelo modelo, o conhecimento químico do sistema e a validação cruzada⁶⁰⁻⁶¹. Neste trabalho optou-se por determinar o número de fatores através da variância explicada pelo modelo.

Uma vez determinada a complexidade do modelo, as interações mais importantes serão aquelas que apresentarem os valores mais altos de core⁶². O sinal do core deve ser levado em consideração, uma vez que ele vai determinar as possíveis combinações dos fatores. Por exemplo, se o core possui valor positivo e os loadings dos três modos apresentam valores positivos e negativos, existem 4 combinações de valores de loadings cuja multiplicação resulta no mesmo sinal do core (+), são elas (+,+,+), (-,-,+), (-,+,-) e (+,-,-), sendo que as outras 4 combinações resultam em valor negativo. Mas, mesmo com 4 combinações diferentes, é possível que apenas uma delas tenha significado relevante, caberá ao pesquisador identificar essa relevância.

4. Parte experimental

Para este estudo foram utilizadas as folhas secas da planta classificada como *Erythrina speciosa* Andrews, Figura 3. A exsicata desta planta está depositada no herbário da Universidade Estadual de Londrina e foi classificada por M.R.C. Paiva sob o número FUEL 35133.



Figura 3: Foto da planta Erythrina speciosa Andrews.

4.1 Coleta e secagem

A coleta para a fase de testes foi realizada em novembro de 2006 e a coleta após a definição do plano de trabalho foi realizada em janeiro de 2008.

As folhas de *Erythrina* foram selecionadas de forma a evitar amostras com danos causados por insetos, fungos ou danos mecânicos. Em seguida as folhas foram cortadas manualmente com o auxílio de uma tesoura para manter as substâncias orgânicas distribuídas de forma homogênea.

A secagem do material vegetal visa à durabilidade das substâncias da planta, sendo necessária para que não haja modificações físicas, químicas ou microbiológicas, evitando-se assim reações adversas, como oxidações, ataque de microorganismos e hidrólise das substâncias orgânicas⁶³.

Neste trabalho, a secagem das folhas foi realizada à temperatura ambiente por doze dias. As folhas foram cortadas com auxílio de uma tesoura, e em seguida dispostas em cima de papel comum e colocadas em local arejado e protegido de luz solar direta, já que a irradiação solar pode alterar a composição química do material. Durante o período de secagem, as folhas foram homogeneizadas duas vezes ao dia para evitar o ataque de fungos devido à umidade. Após os doze dias, as folhas foram trituradas com o uso de um liquidificador industrial e armazenadas em sacos de papel.

4.2 Planejamento experimental de misturas

Os extratos foram preparados com base em um planejamento experimental do tipo Centróide-Simplex²⁰ com 4 componentes, *e* etanol, *d* diclorometano, *h* hexano e *a* acetona, Figura 4.



Figura 4: Planejamento experimental do tipo Centróide-Simplex com 4 componentes.

Na Figura 4, os vértices do tetraedro (\circ) correspondem aos solventes puros, o centro das arestas (\bullet) correspondem às misturas binárias com os dois componentes em proporções iguais, os pontos no centro das faces dos triângulos (\bullet) correspondem às misturas ternárias com os três componentes em proporções iguais e o ponto no centro do tetraedro (\bullet) corresponde ao ponto central, isto é, à mistura contendo os quatro solventes em proporções iguais.

Os solventes puros foram selecionados com base no triângulo de seletividade de solventes de Snyder para cromatografia.

Este planejamento permitiu obter 15 misturas diferentes, cujas composições em termos de proporção podem ser vistas na Tabela 3. No total foram preparados 19 extratos (quintuplicata no ponto central), tanto na fase de testes quanto na fase exploratória.

Extrator		Solventes			
EXITATOS	Etanol	Diclorometano	Hexano	Acetona	
е	1	0	0	0	
d	0	1	0	0	
h	0	0	1	0	
а	0	0	0	1	
ed	1/2	1/2	0	0	
eh	1/2	0	1/2	0	
еа	1/2	0	0	1/2	
dh	0	1/2	1/2	0	
da	0	1/2	0	1/2	
ha	0	0	1/2	1/2	
edh	1/3	1/3	1/3	0	
eda	1/3	1/3	0	1/3	
eha	1/3	0	1/3	1/3	
dha	0	1/3	1/3	1/3	
r1 a r5	1/4	1/4	1/4	1/4	

Tabela 3: Composição das misturas extratoras em termos de proporção. *r*1 a *r*5 corresponde às replicatas de 1 a 5 no ponto central.

4.3 Procedimento experimental na fase de teste

4.3.1 Preparo dos extratos

Na fase de testes iniciais, foram preparados 19 extratos de *Erythrina* (cinco replicatas no ponto central) pesando-se 2,000 g de folhas secas e trituradas e adicionando-se 30 mL do solvente extrator de acordo com as composições do planejamento Centróide-Simplex. Essas misturas ficaram no banho ultrassônico por 30 minutos⁶⁴, sendo que a água do banho foi trocada a cada 10 min para evitar aquecimento. Em seguida os extratos foram filtrados em algodão e colocados em frascos limpos.

4.3.2 Análise por cromatografia líquida de alta eficiência

As amostras para análise por CLAE na fase de testes foram preparadas diluindo 50 µL do extrato bruto em 950 µL (1:20) da fase móvel. As amostras diluídas foram filtradas duas vezes em filtro Millipore Millex (0,22 µm de diâmetro de poro) para garantir que todas as partículas sólidas fossem retiradas e em seguida essa solução foi colocada em microtubo 3810 com capacidade para 1,5 mL da Eppendorf e armazenada em congelador até a análise para evitar a evaporação dos solventes.

Os 19 extratos foram analisados aleatoriamente. A análise por CLAE foi realizada nas seguintes condições: volume de injeção de 20 μL, monitoramento da eluição no comprimento de onda de 210 nm, coluna Metasil C18 ODS PN0380 Metachem (250 mm x 4,6 mm, 5 μm) e vazão da fase móvel de 1 mL/min. Foram testadas duas temperaturas de análise, 30°C e 50°C e duas fases móveis compostas por metanol e água em diferentes proporções, totalizando 4 blocos de experimentos. A Tabela 4 apresenta a composição das fases móveis e também a força cromatográfica calculada para cada uma delas.

Fase móvel	Solventes		Forca Cromatográfica	
	Água	Metanol		
1	20	80	2,08	
2	40	60	1,56	

Tabela 4. Composição e força cromatográfica das duas fases móveis usadas na fase de testes .

4.3.3 Reagentes e equipamentos

Os reagentes utilizados foram: acetona P.A (99,5%), diclorometano P.A (99,5%), hexano P.A (99%), metanol grau cromatográfico (99,9%), todos procedentes da VETEC, álcool etílico hidratado produzido pela Montenegro (92,8%) e água ultrapura.

Os equipamentos utilizados foram: Cromatógrafo a Líquido Shimadzu LC-10AD com detector por arranjo de diodo (DAD) SPD-M10AVP, liquidificador SKYMSEN TA-02, balança analítica Shimadzu AC200, ultrassom Unique modelo Ultracleaner 1400 e sistema Milli-Q.

4.4 Procedimento experimental na fase exploratória

4.4.1 Preparo dos extratos e fracionamento

Na fase exploratória, os 19 extratos de *Erythrina* foram preparados novamente, porém agora pesando-se 15,000 g de folhas secas e trituradas e adicionando-se 120 mL do solvente extrator aleatoriamente de acordo com as composições do planejamento Centróide-Simplex. Essas misturas ficaram no banho ultrassônico por 30 minutos, sendo que água do banho foi trocada a cada 10 minutos para evitar aquecimento, em seguida os extratos foram filtrados em papel de filtro comum, e às folhas foram adicionados mais 120 mL do solvente extrator, este procedimento foi repetido num total de 5 vezes, sendo que o volume de solvente extrator total adicionado às folhas foi de 600 mL.

Ao final do procedimento, os extratos foram concentrados a 1/10 do seu volume inicial em um evaporador rotativo a uma temperatura de 50°C (±2°C) e colocados em frascos limpos e pesados. Os frascos contendo os extratos foram colocados numa bancada de forma a evitar a incidência de luz solar e mantidos sob ventilação forçada durante alguns dias até atingir peso constante.

Após atingir peso constante, os extratos brutos de *Erythrina* foram fracionados⁶⁵ de acordo com a Figura 5.

Primeiramente cada extrato bruto de *Erythrina* foi homogeneizado em 60 mL de uma solução composta por metanol e água na proporção de 4:1 (v/v). Após a dissolução completa, cada mistura foi filtrada em papel de filtro comum previamente pesado.

O resíduo foi lavado com 20 mL de acetato de etila por três vezes. As fibras retidas no filtro eram compostas por polissacarídeos e a solução obtida, fração neutra, era composta por gorduras e graxas. Essa fração foi concentrada a 1/10 do seu volume inicial num evaporador rotativo e colocada em frascos limpos e pesados.

25



Figura 5: Procedimento geral do fracionamento prévio das diferentes classes de metabólitos.

O filtrado foi concentrado a 1/10 do seu volume inicial num evaporador rotativo e acidificado com ácido sulfúrico. Em seguida, essa solução foi extraída por três vezes com 20 mL de clorofórmio. Este procedimento resultou na fase orgânica e na fase aquosa.

A fase orgânica foi concentrada a 1/10 do seu volume inicial num evaporador rotativo e colocada em frascos limpos e pesados. Esta fase foi chamada de fração orgânica e era composta por terpenóides e compostos fenólicos.

A fase aquosa foi basificada com hidróxido de amônio a pH 10 e extraída por duas vezes com 20 mL de uma solução composta por clorofórmio e metanol na proporção de 3:1 (v/v) e uma vez em 20 mL de clorofórmio puro. Este procedimento resultou na fase orgânica e na fase aquosa.

A fase orgânica foi concentrada a 1/10 do seu volume inicial num evaporador rotativo e colocada em frascos limpos e pesados. Esta fase foi chamada de fração básica e era composta por alcalóides.

A fase aquosa foi concentrada a 1/10 do seu volume inicial num evaporador rotativo e colocada em frascos limpos e pesados. Esta fase foi chamada de fração polar e era composta por alcalóides quaternários e n-óxidos, porém devido à presença de água, houve o crescimento de fungos e esta fração foi descartada.

As fibras e os frascos contendo as frações, neutra, orgânica e básica foram colocados numa bancada de forma a evitar a incidência de luz solar e mantidos sob ventilação forçada durante alguns dias até atingir peso constante.

4.4.2 Análise por cromatografia líquida de alta eficiência

Os extratos brutos e as frações, básica e orgânica foram analisados por cromatografia líquida de alta eficiência. A fração neutra e as fibras não foram analisadas por cromatografia líquida de alta eficiência. A fração neutra é basicamente composta por gorduras e graxas, a caracterização desses lipídios pode ser feita através de cromatografia gasosa⁶⁶, determinando assim o perfil graxo. As fibras são compostas por polissacarídeos ou carboidratos e seu perfil pode ser determinado através de cromatografia de troca iônica⁶⁷.

Como os extratos brutos foram fracionados, para a análise por cromatografia líquida de alta eficiência, foi necessário preparar os extratos novamente usando as folhas secas e trituradas da mesma população das amostras do fracionamento. Os extratos foram preparados pesando-se 1g de folhas secas e trituradas e adicionando-se 15 mL do solvente extrator. Essas misturas ficaram no banho ultrassônico por 30 minutos, sendo que água do banho foi trocada a cada 10 minutos para evitar aquecimento, em seguida os extratos foram filtrados em papel de filtro comum, e 50 µL da solução foram diluídos em 950 µL da fase móvel.

As amostras diluídas foram filtradas duas vezes em filtro Millipore Millex (0,22 µm de diâmetro de poro para garantir que todas as partículas sólidas fossem retiradas e em seguida essa solução foi colocada em microtubo 3810 com capacidade para 1,5 mL da Eppendorf e armazenada em congelador até a análise para evitar a evaporação dos solventes.

Para a análise cromatográfica da fração orgânica, todo o conteúdo obtido foi dissolvido em 10 mL de metanol, em seguida, 50 μL da solução foram diluídos em 950 μL da fase móvel. As

27

amostras diluídas foram filtradas duas vezes em filtro Millipore Millex (0,22 μm de diâmetro de poro) e armazenadas em congelador.

Para a análise cromatográfica da fração básica, todo o conteúdo obtido foi dissolvido em 2 mL de metanol, em seguida, 50 μ L da solução foram diluídos em 950 μ L da fase móvel. As amostras diluídas foram filtradas duas vezes em filtro Millipore Millex (0,22 μ m de diâmetro de poro) e armazenadas em congelador.

A análise por CLAE dos extratos brutos e das frações básica e orgânica foi realizada nas seguintes condições: volume de injeção de 20 μL, monitoramento da eluição no comprimento de onda de 210 nm, coluna Gemini C18 ODS PN0380 Phenomenex (250 mm x 4,6 mm, 5 μm), coluna de guarda Bondclone AJO-4287 (40 mmx 3.0 mm)e vazão da fase móvel de 1 mL/min.

Nesta etapa do trabalho foram testadas três fases móveis. As fases móveis eram uma mistura de água, com os modificadores orgânicos metanol e acetonitrila, em proporções diferentes. A Tabela 5 apresenta a composição das três fases e também a força cromatográfica calculada para cada uma delas.

Fase móvel	Solventes			Força
	Água	Metanol	Acetonitrila	Cromatográfica
FM1	65	17,5	17,5	1,00
FM2	47	27	26	1,50
FM3	30	35	35	2,00
FM3	30	35	35	2,00

Tabela 5. Composição e força cromatográfica das três fases móveis testadas na fase exploratória.

4.4.3 Reagentes e equipamentos

Os reagentes utilizados foram: acetona P.A (99,5%), diclorometano P.A (99,5%), clorofórmio P.A (99,8%), hexano P.A (99%), etanol grau cromatográfico (95%), ácido sulfúrico P.A, todos procedentes da VETEC, metanol grau cromatográfico (99,8%) da Tedia, acetonitrila grau cromatográfico (99,8%) da J. T. Baker, hidróxido de amônio produzido pela CHEMCO e água ultrapura.

Os equipamentos utilizados foram: balança analítica Shimadzu AC200, liquidificador SKYMSEN TA-02, ultrassom Unique modelo Ultracleaner 1400, evaporador rotativo FISATOM 802,

pHmetro HANNA HI9321, Cromatógrafo a Líquido Finnigan Surveyor 61607 da Thermo Electron Coorporation com detector por arranjo de diodo Finnigan Surveyor PDA Plus e sistema Milli-Q.

4.5 Métodos estatísticos e programas computacionais

Neste trabalho os métodos estatísticos de análise de dados utilizados foram a análise de componentes principais, a análise hierárquica de agrupamentos e o método Tucker3, cuja rotina está disponível na referência 68. Para os cálculos, foram utilizados os programas computacionais STATISTICA[®] 6.0⁶⁹, Matlab[®] 6.5⁷⁰ e Pirouette[®] 3.11⁷¹ e também os programas computacionais desenvolvidos pela Profa. Dra. Ieda Spacino Scarminio do Laboratório de Quimiometria em Ciências Naturais (LQCN) do Depto. de Química da Universidade Estadual de Londrina.

5. Resultados

Os resultados deste trabalho estão divididos em duas partes, a primeira, referente aos testes iniciais e a segunda, que foi executada após a definição da parte experimental, chamada de fase exploratória.

5.1 Resultados do teste inicial

Na análise por cromatografia líquida de alta eficiência foram testadas duas fases móveis compostas por metanol e água nas proporções de 80:20 (v/v) e de 60:40 (v/v), e duas temperaturas, a 30° C e a 50° C. A Figura 6 mostra os cromatogramas originais das replicatas no ponto central obtidos em cada um dos quatro experimentos com tempo de corrida de 10 min.



Figura 6: Cromatogramas originais das replicatas *r1* a *r5* do extrato bruto no ponto central obtidos no estudo da variação da composição da fase móvel e da temperatura.

A Figura 6 mostra que os cromatogramas obtidos usando a fase móvel composta por metanol e água na proporção de 80:20 e na temperatura de 50°C forneceu maior informação química e também um menor tempo de corrida com boa separação dos picos. Dessa forma, serão mostrados somente os resultados das análises dos dados cromatográficos obtidos nessas condições.

5.1.1 Análise de componentes principais dos cromatogramas dos extratos preparados com base no planejamento Centróide-Simplex (Anexo V)

Os valores de absorbância dos picos cromatográficos dos 19 extratos foram transformados em uma matriz composta por 19 amostras e 1407 variáveis (valores de absorbância) correspondentes ao tempo de corrida de 15 minutos. Essa matriz foi normalizada por área unitária e depois submetida à análise de componentes principais. Os cromatogramas das replicatas no ponto central estão apresentados na Figura 6a.

As três primeiras componentes principais explicam 90,02% da variância contida nos dados. Os escores das três componentes principais para cada solvente extrator e as variâncias explicadas por cada componente principal estão apresentados na Tabela 6.

A Figura 7 apresenta o gráfico dos escores das componentes principais 1 e 2 e das componentes principais 1 e 3. Nos gráficos, é possível ver a formação de 5 grupos. Os grupos são definidos com base em sua localização no espaço e também na sua avaliação conjunta com os loadings e com os cromatogramas originais. O grupo I é formado pelas amostras *d* e *da*, o grupo III, pelas amostras *ed* e *eda*, o grupo IV pelas amostras *h*, *eh*, *ha* e *eha*, o grupo V é formado pelas amostras *e*, *a* e *ea*, e o grupo II é formado pelo restante das amostras, incluindo as replicatas no ponto central. As replicatas *r3*, *r4* e *r5* parecem formar um subgrupo dentro do grupo II, isso é devido a um efeito de bloco, pois essas replicatas foram preparadas um mês após as replicatas *r1* e *r2*. Além da variação das condições cromatográficas, o material vegetal pode ter perdido parte dos seus componentes voláteis, a diferença entre as replicatas pode ser vista na Figura 6a, principalmente para o pico em 6,6 min, que apresentou menor intensidade nas replicatas *r3*, *r4* e *r5*.

32

Tabela 6: Valores dos escores das três primeiras componentes principais e suas porcentagens de variância explicada para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Extrata	Componentes principais			
Extrato	CP1	CP2	CP3	
е	0,077	0,113	0,296	
d	0,434	-0,417	-0,190	
h	0,129	0,429	-0,262	
a	0,092	0,122	0,285	
ed	0,260	-0,132	0,160	
eh	0,113	0,297	-0,016	
еа	0,104	0,130	0,587	
dh	0,272	0,117	-0,251	
da	0,427	-0,337	-0,092	
ha	0,132	0,407	-0,198	
edh	0,201	0,135	-0,044	
eda	0,286	-0,138	0,219	
eha	0,122	0,314	0,015	
dha	0,271	0,090	-0,199	
r1	0,219	0,117	-0,052	
r2	0,214	0,117	-0,059	
r3	0,170	0,093	0,141	
r4	0,230	0,088	0,266	
r5	0,162	0,082	0,234	
% variância explicada	58,94	20,53	10,55	



Figura 7: Gráfico dos escores das componentes principais (a) 1 e 2 e (b) 1 e 3 dos dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C). A seta indica que as amostras *dh* e *dha* fazem parte do grupo II.

A componente principal 1 (CP1) explica sozinha 58,94% da variância dos dados, e discrimina o grupo I dos outros grupos. A componente principal 2 (CP2) consegue discriminar especialmente o grupo I do grupo IV com 20,53% de variância. A componente principal (CP3), com apenas 10,6% de

variância consegue discriminar o grupo V (especialmente a amostra *ea*) das amostras contendo, diclorometano, hexano e/ou acetona.

Os loadings das componentes principais 1, 2 e 3 estão apresentados na Figura 8. Esta Figura mostra que todos os valores dos loadings da CP1 são positivos, sendo que os valores mais altos correspondem às substâncias com tempo de retenção em 3,1, 1,7 e 6,6 min. Isto indica que as amostras do grupo I, alocadas na parte mais positiva da CP1, possuem os valores de escore mais elevados, e sofrem maior influência da variável com valor mais alto de loading, ou seja, são especialmente caracterizadas pela quantidade da substância ou mistura de substâncias com tempo de retenção em 3,1 min.



Figura 8: Gráfico dos loadings da CP1, CP2 e CP3 dos dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Os loadings da CP2 mostram picos em regiões positivas e negativas, as amostras com escores positivos permitem extrair uma maior quantidade das substâncias com tempo de retenção em 6,6 min, enquanto que as amostras localizadas na parte negativa são discriminadas principalmente devido à variável com tempo de retenção em 3,1 min.

Na parte positiva da CP3, a variável que discrimina as amostras é o pico com tempo de retenção em 1,7 min. Na parte negativa, as variáveis responsáveis pela separação são correspondentes aos tempos de retenção em 3,1 e 6,6 minutos.

Para confirmar os resultados da análise de componentes principais, a análise hierárquica de agrupamentos foi aplicada aos dados cromatográficos dos 19 extratos normalizados por área unitária utilizando o método de agrupamento pareado igualmente ponderado (Unweighted pair-group average).

O dendrograma obtido pelo conjunto de 19 amostras e 1407 valores de absorbância está apresentado na Figura 9, onde se observa no valor de distância de 0,07 os mesmos grupos formados na análise de componentes principais, exceto pelas amostras *e*, *a* e *ea*, que por pouco não foram incluídas no mesmo grupo, como indica a pequena diferença de distância entre elas. Também é possível observar que o grupo I é bem diferente dos demais.



Figura 9: Dendrograma baseado nos dados cromatográficos dos extratos brutos de *Erythrina* obtido com o método de agrupamento pareado igualmente ponderado.

Na Figura 10, estão apresentados os cromatogramas dos 19 extratos separados com base nos grupos formados pela análise de componentes principais, onde é possível constatar que os picos mais importantes na discriminação possuem tempo de retenção de 1,7, 3,1 e 6,6 min.

Os resultados da análise de componentes principais sugerem que o extrator pode ser otimizado para obter substâncias específicas do material vegetal. A metodologia de superfície de resposta que consiste na modelagem pela aplicação de modelos simples às respostas obtidas com o planejamento de misturas para fazer previsões foi aplicada aos escores de cada uma das três componentes principais. O erro padrão foi estimado pelo erro experimental determinado pelas cinco replicatas no ponto central.



Figura 10: Cromatogramas dos extratos brutos de *Erythrina* separados com base nos grupos formados pela PCA (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Utilizando os escores da CP1 como resposta, foram construídos os modelos linear, quadrático e cúbico especial. Nenhum modelo apresentou falta de ajuste no nível de 95% de confiança, porém, nenhum termo cúbico se mostrou significativo no modelo cúbico e o único termo de interação significativo foi observado entre os componentes diclorometano e acetona, logo o modelo preferido foi o quadrático. A equação completa para este modelo é dada por:

$$\begin{aligned} \mathbf{CP}_{1} &= \mathbf{0}, \mathbf{0827}e + \mathbf{0}, \mathbf{4423}d + \mathbf{0}, \mathbf{1375}h + \mathbf{0}, \mathbf{0987}a - 0, 0501ed - 0, 0270eh \\ & (\pm 0, 0268) \quad (\pm 0, 0268) \quad (\pm 0, 0268) \quad (\pm 0, 0268) \quad (\pm 0, 1132) \\ & + 0, 0332ea - 0, 1428dh + \mathbf{0}, \mathbf{5751}da + 0, 0006ha \\ & (\pm 0, 1132) \quad (\pm 0, 1132) \quad (\pm 0, 1132) \end{aligned}$$
(17)

onde e = etanol, d = diclorometano, h = hexano e a = acetona e os valores entre parênteses representam o erro padrão de cada coeficiente. Os termos em negrito representam significância no nível de 95% de confiança.

A análise de variância (ANOVA) para esta regressão é mostrada na Tabela 7.

Tabela 7: Análise da variância para o ajuste do modelo quadrático aos valores de escores da CP1 para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F*calculado
Modelo	0,1835	9	0,0203	29,00
Resíduos	0,0066	9	0,0007	
Falta de ajuste	0,0028	5	0,0005	0,55
Erro puro	0,0038	4	0,0009	
Total	0,1902	18	0,0105	

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{9,9}= 3,18 e F_{5,4} = 6,26.

O valor da razão da MQ_{faj}/MQ_{ep} é 0,55, o valor da distribuição F com 5 e 4 graus de liberdade é de 6,26 no nível de 95% de confiança. Como o valor de F tabelado é muito maior que a razão MQ_{faj}/MQ_{ep} , não há nenhuma indicação de falta de ajuste para o modelo quadrático. Além disso, o valor da razão MQ_R/MQ_r = 29, é muito maior se comparada ao valor da distribuição F com o mesmo número de graus de liberdade (9 e 9) que é 3,18 no nível de 95% de confiança, indicando que o modelo é significativo.

O coeficiente de determinação R^2 para este ajuste é calculado pela razão SQ_R/SQ_T que é igual a 0,9647 e o valor máximo que R^2 poderia assumir é dado por $(SQ_T - SQ_{ep})/SQ_T$ que neste caso corresponde a 0,9800.

Com a validação do modelo, pode-se interpretar a equação para determinar os fatores mais importantes na maximização dos escores da CP1, ou seja, determinar o solvente ou mistura de solventes capaz de extrair uma maior quantidade da substância com tempo de retenção em 3,1 min.

A equação mostra que os quatro termos lineares são significativos no nível de 95% de confiança. O solvente diclorometano apresentou o maior coeficiente linear. O coeficiente de

interação sinérgico altamente significativo entre o diclorometano e a acetona indica que a presença simultânea do diclorometano e da acetona na mistura é importante para extrair uma maior quantidade da substância com valor mais alto de loading na CP1, ou seja, da substância com tempo de retenção em 3,1 min, pertencente ao grupo I.

Na Figura 11 é possível ver os resíduos deixados pelo ajuste versus as respostas previstas pelo modelo, juntamente com o gráfico das respostas previstas pelo modelo pelas observadas experimentalmente. Na Figura 11a vê-se que a distribuição dos resíduos não revela tendências sistemáticas, também é possível visualizar o efeito de bloco na replicatas. Na Figura 11b vê-se claramente que as misturas contendo diclorometano na sua composição permitem extrair uma maior quantidade do pico em 3,1 min.



Figura 11: Gráficos (a) dos resíduos deixados pelo ajuste do modelo quadrático e (b) das respostas previstas pelas respostas observadas para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

As curvas de nível com a maior previsão para o modelo quadrático aplicado aos escores da CP1 podem ser vistas na Figura 12, o componente etanol foi igualado a zero com o intuito de representar a superfície em apenas duas dimensões.

O valor máximo de escore que a CP1 poderá alcançar é obtido com uma mistura de 75% de diclorometano com 25% de acetona. De todos os experimentos realizados, as amostras *d* e *da* foram as que apresentaram os maiores valores de escore.



Figura 12: Superfície de resposta para os componentes; diclorometano, hexano e acetona, aplicada aos escores da CP1.

Para a CP2, o modelo linear apresentou falta de ajuste, já para os modelos quadrático e cúbico especial, a equação foi bem ajustada. Como o termo cúbico não foi significativo no nível de 95% de confiança, foi escolhido o modelo quadrático, cuja equação completa está apresentada abaixo.

$$\begin{aligned} \textbf{CP}_{2} &= \textbf{0}, \textbf{1120}e - \textbf{0}, \textbf{4196}d + \textbf{0}, \textbf{4253}h + \textbf{0}, \textbf{1213}a + 0, 0920ed + \textbf{0}, \textbf{1552}eh \\ & (\pm 0, 0143) \ (\pm 0, 0143) \ (\pm 0, 0143) \ (\pm 0, 0143) \ (\pm 0, 0605) \ (\pm 0, 0605) \\ & + 0, 0549ea + \textbf{0}, \textbf{5050}dh - \textbf{0}, \textbf{7492}da + \textbf{0}, \textbf{5744}ha \\ & (\pm 0, 0605) \ (\pm 0, 0605) \ (\pm 0, 0605) \ (\pm 0, 0605) \end{aligned}$$

Os resultados da análise de variância (ANOVA), apresentados na Tabela 8, foram validados da mesma forma que o resultado anterior e mostram que o modelo quadrático descreve os pontos experimentais adequadamente. O coeficiente de determinação R² para este ajuste é igual a 0,9977 e o valor máximo que R² poderia assumir corresponde a 0,9987.

Tabela 8: Análise da variância para o ajuste do modelo quadrático aos valores de escores da CP2 para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F*calculado
Modelo	0,8578	9	0,0953	476,50
Resíduos	0,0019	9	0,0002	
Falta de ajuste	0,0007	5	0,0001	0,56
Erro puro	0,0011	4	0,0002	
Total	0,8597	18	0,0477	

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{9,9}= 3,18 e F_{5,4} = 6,26.

Na equação, todos os termos lineares são significativos, sendo que os mais importantes no nível de 95% de confiança são o hexano e o diclorometano. O coeficiente para o hexano possui valor positivo, indicando que este solvente puro atua de forma a aumentar os escores da CP2, enquanto que o coeficiente altamente negativo do diclorometano atua de forma a diminuir. Dos termos quadráticos, as misturas dos solventes diclorometano e hexano bem como hexano e acetona apresentam uma relação sinérgica altamente significativa, e nota-se uma interação antagônica ainda maior entre os solventes diclorometano e acetona.

A maximização dos escores da CP2 (valores positivos de escore) visa à obtenção de maiores quantidades de substâncias com tempo de retenção de 6,6 min, já a minimização (valores negativos de escore) corresponde à obtenção dos picos com tempo de retenção de 3,1 min.

Na Figura 13 é possível ver os resíduos deixados pelo ajuste versus as respostas previstas pelo modelo, juntamente com o gráfico das respostas previstas pelas observadas. A distribuição dos resíduos não revela tendências sistemáticas, novamente é possível observar o efeito da blocagem nas replicatas. Na Figura 13b observa-se que os extratos que possuem hexano na sua composição permitiram obter maiores quantidades do pico com tempo de retenção em 6,6 min e também que as misturas que foram ótimas para extração de maiores quantidades do pico com tempo de retenção de 3,1 min (Figura 11b) na otimização dos escores da CP1 foram menos eficientes.

41



Figura 13: Gráficos (a) dos resíduos deixados pelo ajuste do modelo quadrático e (b) das respostas previstas pelas respostas observadas para os escores da CP2 para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

As curvas de nível com o maior valor previsto para o modelo quadrático aplicado aos escores da CP2 podem ser vistas na Figura 14, e está representada em duas dimensões com os componentes diclorometano, hexano e acetona.



Figura 14: Superfície de resposta para os componentes; diclorometano, hexano e acetona, aplicada aos escores da CP2.

A superfície de resposta mostra que para maximizar os escores da CP2, ou seja, para extrair uma maior quantidade da substância com tempo de retenção de 6,6 min, o solvente extrator deve ser uma mistura composta por 75% de hexano e 25% de acetona. Porém, se for desejável minimizar os escores da CP2, que significa extrair uma maior quantidade da substância com tempo de retenção em 3,1 min, o solvente utilizado deverá ser diclorometano puro ou uma mistura de 80% de diclorometano e 20% de acetona. De acordo com a Tabela 6, o valor mais alto de escore na CP2 é obtido com a mistura binária hexano e acetona, correspondente ao extrato *ha* e os valores mais negativos de escore são obtidos utilizando-se como extrator o diclorometano puro e a mistura diclorometano e acetona.

Para os escores da CP3 também foram testados os modelos linear, quadrático e cúbico especial, sendo que nenhum dos três apresentou falta de ajuste. Porém, não foi verificado nenhum termo cúbico significativo. Mas devido à presença de um termo significativo de interação binária, o modelo quadrático foi escolhido para o ajuste dos escores da CP3, a equação ajustada a esse modelo está apresentada abaixo.

$$\begin{aligned} \textbf{CP}_{3} &= \textbf{0}, \textbf{2886}e - 0, 2094d - \textbf{0}, \textbf{2749}h + \textbf{0}, \textbf{2743}a + 0, 5674ed - 0, 1192eh \\ & (\pm 0, 1178) (\pm 0, 1178) (\pm 0, 1178) (\pm 0, 1178) (\pm 0, 4982) (\pm 0, 4982) \\ & + \textbf{1}, \textbf{1798}ea - 0, 1104dh - 0, 3710da - 0, 7726ha \\ & (\pm 0, 4982) (\pm 0, 4982) (\pm 0, 4982) (\pm 0, 4982) \end{aligned}$$

Os resultados da análise de variância (ANOVA), apresentados na Tabela 9, foram validados e mostram que o modelo quadrático descreve os pontos experimentais adequadamente. O coeficiente de determinação R² para este ajuste é igual a 0,8659 e o valor máximo que R² poderia assumir corresponde a 0,9003.

Tabela 9: Análise da variância para o ajuste do modelo quadrático aos valores de escores da CP3 para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F*calculado
Modelo	0,8337	9	0,0926	6,47
Resíduos	0,1291	9	0,0143	
Falta de ajuste	0,0331	5	0,0066	0,27
Erro puro	0,0959	4	0,0239	
Total	0,9628	18	0,0534	

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{9,9}= 3,18 e F_{5,4} = 6,26.
Na equação, observa-se que os coeficientes lineares etanol, hexano e acetona são significativos e muito parecidos em valor absoluto, porém, o etanol e a acetona são importantes na maximização dos escores da CP3, enquanto que o hexano é importante na minimização. Observa-se também a existência de uma interação sinérgica binária altamente significativa entre o etanol e a acetona.

Na Figura 15 é possível ver os resíduos deixados pelo ajuste versus as respostas previstas pelo modelo, juntamente com o gráfico das respostas previstas pelas observadas. A distribuição dos resíduos não revela tendências sistemáticas, novamente é possível observar o efeito da blocagem nas replicatas.



Figura 15: Gráficos (a) dos resíduos deixados pelo ajuste do modelo quadrático e (b) das respostas previstas pelas respostas observadas para os escores da CP3 para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

As curvas de nível para o modelo quadrático aplicado aos dados da CP3 podem ser vistas na Figura 16, onde o componente diclorometano foi igualado a zero para representar a superfície em apenas duas dimensões.



Figura 16: Superfície de resposta para os componentes; etanol, hexano e acetona, aplicada aos escores da CP3.

A superfície de resposta prevê que os maiores valores de escores para a CP3, que correspondem ao pico com tempo de retenção em 1,7 min, são obtidos utilizando como solvente extrator uma mistura composta por 50% de etanol e 50% de acetona. Já os valores mais negativos, que se refere aos picos em 3,1 min e 6,6 min podem ser obtidos utilizando o hexano puro como solvente extrator.

A Tabela 6 mostra que o maior valor de escore da CP3 dos dados experimentais foi obtido com a mistura etanol e acetona na proporção de 1:1, enquanto que o valor mais negativo é encontrado utilizando o hexano puro como solvente extrator, em seguida o menor valor é obtido com a mistura binária de hexano e diclorometano.

Até agora, a otimização dos escores das componentes principais foi feita individualmente, através da inspeção visual das superfícies de resposta, por exemplo, para maximizar os escores da CP1, que corresponde principalmente ao pico com tempo de retenção em 3,1 min, o solvente extrator deveria ser uma mistura composta por 80% de diclorometano e 20% de acetona. Porém, cada componente principal é responsável pela separação das amostras com base em duas ou três regiões mais relevantes no cromatograma. Dessa forma, deve-se pensar em descobrir os níveis dos fatores que produzirão o conjunto de respostas mais satisfatório, ou seja, os níveis dos fatores que permitem obter a maior quantidade das substâncias com tempo de retenção em 1,7, 3,1 e 6,6 minutos.

Usando a metodologia de otimização simultânea proposta por G. C. Derringer e R. Suich, foram avaliadas as funções de desejabilidade para os picos com tempo de retenção em 1,7, 3,1 e 6,6 min, que foram os mais importantes na discriminação dos extratos de *Erythrina*, por apresentarem maior peso nos loadings.

Supondo que o pico de interesse seja o pico com tempo de retenção em 3,1 min, deve-se otimizar as três respostas, que neste caso são os escores das componentes principais 1, 2 e 3. No gráfico dos loadings, observa-se que para obter maiores quantidades dessas substâncias seria desejável que as componentes principais 2 e 3 tivessem os mínimos valores de loadings possíveis, ou seja, que os escores das duas componentes fossem negativos, e que os escores da CP1 fossem maximizados.

A Tabela 10 mostra os parâmetros utilizados para o cálculo da desejabilidade para os três picos mais importantes.

Para a otimização da quantidade extraída da substância com tempo de retenção em 1,7 min, os escores da CP3 devem ser maximizados, enquanto que para os escores das componentes 1 e 2 é desejável que os escores estejam num valor médio ou baixo. Devido à CP3 caracterizar fortemente esta região, atribui-se a ela a importância máxima 5, e foi escolhido um peso superior igual a 10 para que a resposta atingisse rapidamente o valor desejável.

temperatura de 50oC) para os picos com tempo de retenção em 1,7, 3,1 e 6,6 min.							
Tempo de retenção	Resposta	Objetivo	Limite Inferior	Limite Superior	Peso Superior	Peso Inferior	Importância
	CP1	0,25588	0,07707	0,4347	1	1	1
1,7 min	CP2	0,00572	-0,41772	0,42916	1	1	1
	CP3	maximizar	-0,26281	0,58755	10		5
	CP1	maximizar	0,07707	0,4347	1		5
3,1 min	CP2	minimizar	-0,41772	0,42916		1	5
	CP3	minimizar	-0,26281	0,58755		1	4
	CP1	no intervalo	0,07707	0,4347			5
6,6 min	CP2	maximizar	-0,41772	0,42916	1		5
	CP3	minimizar	-0,26281	0,58755		1	4

Tabela 10: Parâmetros utilizados na otimização simultânea dos escores da PCA para os dados cromatográficos dos extratos brutos de Erythrina (Fase móvel MeOH:H2O 80:20 (v/v) e temperatura de 50oC) para os picos com tempo de retenção em 1,7, 3,1 e 6,6 min.

Para otimizar a extração da substância com tempo de retenção em 3,1 min é desejável maximizar os escores da CP1 e minimizar os escores da CP2 e CP3. Para a otimização da quantidade extraída da substância com tempo de retenção em 6,6 min é desejável maximizar os escores da CP2, minimizar os escores da CP3 enquanto que para a CP1 os valores dos escores podem estar dentro da faixa obtida experimentalmente.

Para os cálculos foi utilizada uma grade de 40 pontos em cada um dos 4 fatores (etanol, diclorometano, hexano e acetona), o que significa que os valores das respostas e suas correspondentes desejabilidades foram calculados em 40 x 40 x 40 x 40 = $2,56x10^6$ combinações de níveis dos fatores. Todas as respostas foram ajustadas com modelos quadráticos completos.

Com os valores atribuídos à função de desejabilidade, foi possível obter uma desejabilidade global de 0,64 para o pico com tempo de retenção em 1,7 min, que não é muito alta, porém satisfatória, de 0,97 para o pico em 3,1 min e de 1 para o pico em 6,6 min. A Tabela 11 mostra a composição do solvente extrator otimizado pela função de desejabilidade, pela análise visual da superfície de resposta e também os valores máximos e mínimos encontrados experimentalmente para as três componentes.

Tabela 11: Composição do solvente extrator otimizado, respostas previstas pela análise da superfície de resposta e pela função de desejabilidade e valores máximos e mínimos dos escores das componentes principais 1, 2 e 3.

Método	Tempo de	Composição do	Escores			
	retenção (min)	solvente extrator	CP1	CP2	CP3	
Análise visual	3,1	80% d – 20% a	0,465	-0,431	-0,172	
da superfície	6,6	75% h – 25% a	0,128	0,456	-0,283	
de resposta	1,7	50% e – 50% a	0,099	0,130	0,576	
Algoritmo	3,1	83% d – 17% a	0,465	-0,433	-0,179	
Derringer e	6,6	78% h – 21% a	0,130	0,454	-0,286	
Suich	1,7	50% e – 50% a	0,099	0,130	0,576	
Valor máximo de escore obtido experimentalmente			0,434	0,429	0,587	
Valor mínimo de escore obtido experimentalmente			0,077	-0,417	-0,262	

Os resultados obtidos com a função de desejabilidade foram muito próximos daqueles previstos pela inspeção visual da superfície de resposta. Para o pico em 6,6 min as duas análises deram o mesmo resultado, ou seja, os valores ótimos de escore são alcançados utilizando-se como solvente extrator uma mistura composta por 50% de etanol e 50% de acetona. Para os picos em 3,1 e 6,6 min os resultados também demonstraram uma ótima concordância, sendo que as misturas otimizadas através do algoritmo tendem a diminuir a quantidade de acetona nas misturas.

Foram feitos três experimentos confirmatórios usando as misturas otimizadas através da aplicação do algoritmo Derringer & Suich. Os cromatogramas das misturas otimizadas foram comparados com os cromatogramas do planejamento que foram feitos usando os mesmos solventes na mistura. Os resultados podem ser vistos na Figura 17.



Figura 17: Cromatogramas dos experimentos confirmatórios comparados com os cromatogramas das amostras do planejamento para os picos em (a) 1,7 min, (b) 3,1 min e (c) 6,6 min (Fase móvel MeOH: H_2O 80:20 (v/v) e temperatura de 50°C).

A mistura otimizada para o pico em 1,7 min foi a mesma obtida com o planejamento, Figura 17a, o perfil cromatográfico é exatamente o mesmo, porém podemos observar que a intensidade dos picos da amostra otimizada é um pouco maior. Nas Figuras 17b e 17c verifica-se que as misturas otimizadas apresentaram maior intensidade dos picos em 3,1 e 6,6 min, respectivamente.

5.1.2 Análise de componentes principais com rotação varimax dos cromatogramas dos extratos preparados com base no planejamento Centróide-Simplex (Anexo IV)

A mesma matriz utilizada na análise de componentes principais, foi novamente submetida à análise de componentes principais porém agora com rotação varimax. Os resultados obtidos com a rotação varimax também foram analisados da mesma maneira. Os escores de cada extrator e as variâncias explicadas por cada uma das três primeiras componentes principais rodadas estão apresentados na Tabela 12.

Tabela 12: Valores dos escores rodados da $CP1^r$, $CP2^r$ e $CP3^r$ e suas porcentagens de variância explicada para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Extrata		Componentes principais	5
Extrato	CP1 ^r	CP2 ^r	CP3 ^r
е	0,052	0,137	0,265
d	0,944	-0,043	-0,059
h	0,016	0,988	-0,096
a	0,066	0,121	0,133
ed	0,876	0,005	0,311
eh	0,057	0,939	0,236
еа	0,057	0,079	0,874
dh	0,726	0,649	-0,107
da	0,967	-0,016	-0,044
ha	0,033	0,982	-0,060
edh	0,612	0,664	0,029
eda	0,823	-0,016	0,289
eha	0,036	0,903	0,226
dha	0,774	0,597	-0,069
r1	0,680	0,623	0,047
r2	0,669	0,625	0,013
r3	0,407	0,338	0,059
r4	0,390	0,225	0,126
r5	0,398	0,260	0,341
% variância explicada	57,25	18,55	11,87

O gráfico dos escores das componentes principais 1, 2, e 3 rodadas (CP1^r, CP2^r e CP3^r), podem ser vistos nas Figuras 18a e 18b.



Figura 18: Gráfico dos escores das componentes principais rodadas (a) 1 e 2 e (b) 1 e 3 para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

A projeção dos escores rodados mostra claramente a mesma discriminação dos grupos com relação aos grupos formados na projeção dos escores da análise de componentes principais sem rotação varimax, porém enfatizando ainda mais o efeito de bloco das replicatas *r3*, *r4* e *r5*. Como na análise anterior, os grupos foram definidos com base em sua localização no espaço e também na sua avaliação conjunta com os loadings e com os cromatogramas originais.

Na Figura 19 é possível observar a comparação dos loadings das três primeiras componentes principais obtidos com e sem rotação varimax. Os loadings rodados deram maior importância às mesmas variáveis, porém com maior peso.



Figura 19: Loadings da (a) CP1 e CP1^r, (b) CP2 e CP2^r e (c) CP3 e CP3^r para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Os loadings da CP1^r, Figura 19a, mostram um pico com alta intensidade e vários outros com intensidades mais baixas, já os loadings da CP1 mostram um pico com alta intensidade, dois com intensidade moderada e vários com baixa intensidade. Os loadings da CP2^r mostram um pico com

alta intensidade, enquanto que a CP2 mostra dois, um positivo e outro negativo. Os loadings da CP3^r também são mais simples e mostram um pico mais importante contra três nos loadings da CP3, sendo um com valor positivo e dois com valores negativos.

Na metodologia de superfície de resposta, aos escores das três primeiras componentes principais rodadas, foram aplicados os modelos linear, quadrático e cúbico especial, o melhor ajuste em todos os casos foi obtido com o modelo quadrático, a validação dos modelos foi feita da mesma maneira que no caso anterior. A análise de variância (ANOVA) para o ajuste do modelo quadrático aos valores de escores das CP1^r, CP2^r e CP3^r são mostrados na Tabela 13.

Tabela 13: Análise da variância para o ajuste do modelo quadrático aos valores de escores rodados das CP1^r, CP2^r e CP3^r para os dados cromatográficos dos extratos brutos de *Erythrina* (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F*calculado
CP1 ^r				
Modelo	2,1788	9	0,2420	18,5158
Resíduos	0,1176	9	0,0130	
Falta de ajuste	0,0259	5	0,0051	0,2262
Erro puro	0,0917	4	0,0229	
Total	2,2965	18	0,1275	
CP2 ^r				
Modelo	2,2800	9	0,2533	10,7085
Resíduos	0,2129	9	0,0236	
Falta de ajuste	0,0598	5	0,0119	0,3127
Erro puro	0,1530	4	0,0382	
Total	2,4929	18	0,1384	
CP3 ^r				
Modelo	0,8476	9	0,0941	6,3270
Resíduos	0,1339	9	0,0148	
Falta de ajuste	0,0645	5	0,0129	0,7430
Erro puro	0,0694	4	0,0173	
Total	0,9815	18	0,0545	

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{9,9}=3,18 e F_{5,4} = 6,26. O coeficiente de determinação R² para o ajuste do modelo quadrático aos escores da CP1^r é igual a 0,9487 e o valor máximo que R² poderia assumir corresponde a 0,9600. Para o ajuste do modelo quadrático aos escores da CP2^r, R² é igual a 0,9145 e o valor máximo que R² poderia assumir corresponde a 0,9386. E para o ajuste do modelo quadrático aos escores da CP3^r, R² é igual a 0,8635 e o valor máximo que R² poderia assumir corresponde a 0,9292.

A equação para o modelo quadrático ajustado aos escores rodados da CP1^r pode ser vista abaixo:

$$\begin{aligned} \textbf{CP}_{1}^{r} &= 0,0638e + \textbf{0}, \textbf{9445}d + 0,0253h + 0,0670a + \textbf{1}, \textbf{4786}ed - 0,1000eh \\ & (\pm 0,1125) \quad (\pm 0,1125) \quad (\pm 0,1125) \quad (\pm 0,4756) \quad (\pm 0,4756) \\ & -0,0458ea + 0,9858dh + \textbf{2}, \textbf{0045}da - 0,0329ha \\ & (\pm 0,4756) \quad (\pm 0,4756) \quad (\pm 0,4756) \end{aligned}$$

Este modelo se mostrou mais simples do que aquele ajustado aos escores não rodados, apresentando somente três termos significativos no nível de 95% de confiança enquanto que o outro apresentou cinco. Dos quatro solventes puros, o diclorometano seria o mais importante na otimização dos escores rodados, este solvente ainda participa das duas outras interações sinérgicas significativas, uma com a acetona e outra com o etanol. A interação sinérgica entre o diclorometano e a acetona é maior, e portanto a mistura ótima para maximizar os escores da CP1^r deve conter esses dois solventes. Os extratos preparados em essas misturas *d, ed* e *da* fazem parte dos grupos I e III, localizados na parte mais positiva da CP1^r.

As curvas de nível em termos da proporção dos solventes etanol, diclorometano e acetona podem ser vistas na Figura 20, o componente hexano foi igualado a zero com o intuito de representar a superfície em apenas duas dimensões.



Figura 20: Superfície de resposta para os componentes; etanol, diclorometano e acetona, aplicada aos escores da CP1^r.

Embora a mistura de 75% de diclorometano e 25% de acetona seja indicada como ótima na extração do pico em 3,1 min, ainda tem-se uma mistura alternativa composta por 80% de diclorometano e 20% de etanol. A mistura alternativa se mostrou quase tão eficiente quanto a mistura ótima. Comparando esta superfície de resposta com aquela obtida usando como resposta os escores não rodados com os mesmos solventes (não mostrado), foi constado que elas eram realmente diferentes. Apesar de indicarem a mesma região ótima, que corresponde a 75% de diclorometano e 25% de acetona, a superfície para os escores não rodados indica valores mais baixos nas redondezas da mistura 80% de diclorometano e 20% de etanol. Além do modelo para CP1^r ser mais simples, ainda prevê duas misturas diferentes com alto potencial na extração do pico em 3,1 min.

A equação para o modelo quadrático ajustado aos escores rodados da CP2^r pode ser vista abaixo:

$$\begin{aligned} \textbf{CP}_{2}^{r} &= 0,1485e - 0,0198d + \textbf{0}, \textbf{9979}h + 0,1394a - 0,3547ed + \textbf{1}, \textbf{5666}eh \\ & (\pm 0,1513) (\pm 0,1513) (\pm 0,1513) (\pm 0,1513) (\pm 0,6397) (\pm 0,6397) \\ & - 0,2813ea + 0,5482dh - 0,5238da + \textbf{1}, \textbf{6607}ha \\ & (\pm 0,6397) (\pm 0,6397) (\pm 0,6397) (\pm 0,6397) \end{aligned}$$

Neste modelo observa-se que o solvente hexano apresenta um alto coeficiente linear significativo e participa de duas interações sinérgicas binárias significativas com os solventes etanol e acetona. A substituição do solvente etanol pela acetona ou vice-versa possui pouco efeito no valor previsto pelo modelo. Esse comportamento é similar àquele observado para a CP1^r em misturas contendo grande parte de diclorometano. As misturas *h, eh* e *ha* fazem parte do grupo IV formado na projeção dos escores da CP2^r.

A curva de nível em termos da proporção dos solventes etanol, hexano e acetona pode ser vista na Figura 21, o componente diclorometano foi igualado a zero com o intuito de representar a superfície em apenas duas dimensões.



Figura 21: Superfície de resposta para os componentes; etanol, hexano e acetona, aplicada aos escores da CP2^r.

As misturas ótimas para extração do pico em 6,6 min são previstas na parte direita do triângulo com vértice representado pelo solvente hexano e os limites estão situados perto das misturas binárias entre hexano e acetona e entre hexano e etanol, ambas com proporções de 1:1.

A equação para o modelo quadrático ajustado aos escores rodados da CP3^r pode ser vista abaixo:

 $\begin{aligned} \mathbf{CP}_{3}^{r} &= \mathbf{0}, \mathbf{3088}e - 0, 0346d - 0, 0687h + 0, 1631a + 0, 3270ed + 0, 0612eh \\ & (\pm 0, 1200) \ (\pm 0, 1200) \ (\pm 0, 1200) \ (\pm 0, 1200) \ (\pm 0, 5074) \ (\pm 0, 5074) \\ & + \mathbf{2}, \mathbf{1173}ea - 0, 3363dh - 0, 5818da - 0, 6108ha \\ & (\pm 0, 5074) \ (\pm 0, 5074) \ (\pm 0, 5074) \ (\pm 0, 5074) \end{aligned}$

Esta equação mostra somente dois termos significativos no nível de 95% de confiança, já o modelo obtido para os escores da CP3 mostrou quatro termos significativos. Porém os dois mostraram uma forte interação sinérgica binária entre os solventes etanol e acetona, grupo V.

As curvas de nível em termos da proporção dos solventes etanol, hexano e acetona podem ser vistas na Figura 22, o componente diclorometano foi igualado a zero com o intuito de representar a superfície em apenas duas dimensões.



Figura 22: Superfície de resposta para os componentes; etanol, hexano e acetona, aplicada aos escores da CP3^r.

Comparando esta superfície com aquela obtida na otimização dos escores da CP3 (Figura 16), observa-se que elas são quase idênticas, e indicam a mesma mistura como ótima na extração do pico em 1,7 min, que seria composta por 50% de etanol e 50% de acetona.

A Tabela 14 mostra a composição do solvente extrator otimizado pela função de desejabilidade, pela análise da superfície de resposta e também os valores máximos e mínimos de escores rodados encontrados experimentalmente para as três componentes. A desejabilidade global para os picos em 3,1, 6,6 e 1,7 min foram 0,998, 1,00 e 0,988, respectivamente.

Novamente, os resultados obtidos com a análise da superfície de resposta foram muito próximos dos resultados obtidos com a metodologia de otimização simultânea. Os resultados obtidos com a otimização dos escores não rodados também são bastante similares a estes. Comparando os resultados obtidos, pode-se perceber que quando se usa os escores rodados na otimização existe uma leve diminuição na porcentagem de diclorometano na mistura otimizada para o pico que elui em 3,1 min.

Tabela 14: Composição do solvente extrator otimizado, respostas previstas pela análise da superfície de resposta e pela função de desejabilidade e valores máximos e mínimos dos escores rodados das componentes principais 1, 2 e 3.

Mátada	Tempo de	Composição do		Escores	
Metodo	retenção (min)	solvente extrator	CP1 ^r	CP2 ^r	CP3 ^r
Análise visual	3,1	75% d – 25% a	1,101	-0,048	-0,160
da superfície	6,6	75% h – 25% a	0,032	1,090	-0,117
de resposta	1,7	50% e – 50% a	0,063	0,083	0,765
Algoritmo	3,1	71% d – 29% a	1,101	-0,077	-0,096
Derringer e	6,6	83% h – 17% a	0,026	1,085	-0,106
Suich	1,7	50% e – 50% a	0,053	0,074	0,764
Valor máximo de escore rodado obtido experimentalmente			0,967	0,988	0,874
Valor mínimo de escore rodado obtido experimentalmente			0,033	-0,016	-0,044

Embora existam pequenas diferenças nas proporções estimadas por ambas as técnicas, na prática, os resultados da extração deverão ser essencialmente os mesmos levando em consideração as incertezas nos parâmetros do modelo.

Mais dois experimentos confirmatórios foram executados para o pico em 3,1 min, usando a mistura otimizada através do algoritmo Derringer & Suich para os escores rodados, que foi de 71% de diclorometano e 29% de acetona e também a mistura alternativa indicada pela análise visual da superfície de resposta apresentada na Figura 20, que foi de 80% de diclorometano e 20% de etanol. A Figura 23 mostra os cromatogramas das duas misturas bem como o cromatograma da mistura otimizada através do algoritmo Derringer & Suich para os escores não rodados para o pico em 3,1 min, que tinha sido de 83% de diclorometano e 17% de acetona.



Figura 23: Cromatogramas dos experimentos confirmatórios comparados ao cromatograma da mistura otimizada dos escores da CP1 para o pico em 3,1 min (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Pela comparação dos cromatogramas observa-se pouca diferença com relação à substituição da acetona pelo solvente etanol. A mistura otimizada através do algoritmo Derringer & Suich usando os escores não rodados mostrou-se um pouco mais eficiente do que as outras, permitindo obter uma área ligeiramente maior do pico em 3,1 min e consequentemente uma concentração mais elevada. Os três cromatogramas apresentaram praticamente o mesmo perfil cromatográfico.

Para os picos em 1,7 e 6,6 min não foi necessário executar experimentos confirmatórios. Para o pico em 1,7 min todas as análises indicaram a mesma mistura, 50% de etanol e 50% de acetona, que corresponde ao ponto *ea* do planejamento. Para o pico em 6,6 min, a análise visual da superfície de resposta indica que se pode usar como extratores toda a região compreendida entre hexano puro e as misturas contendo hexano e etanol e hexano e acetona em mesmas proporções, que são os pontos *h*, *eh* e *ha* do planejamento. Uma comparação dos cromatogramas desses três extratos pode ser vista na Figura 24.



Figura 24: Cromatogramas dos pontos *h*, *eh* e *ha* do planejamento (Fase móvel MeOH:H₂O 80:20 (v/v) e temperatura de 50°C).

Essa Figura mostra que o perfil cromatográfico desses três extratos é essencialmente o mesmo acima de 4 min e que a principal diferença entre eles é o fato de que a mistura contendo etanol permite extrair o pico em 1,7 min e a mistura contendo acetona permite extrair o pico em 2,6 min.

Embora os resultados otimizados usando os escores com e sem rotação varimax sejam quase equivalentes para este material vegetal, o uso dos escores rodados seria mais conveniente, pois além de permitir o uso de misturas alternativas, seus modelos foram mais simples de interpretar.

5.2 Resultados da fase exploratória

Nesta parte do trabalho, foram avaliados os rendimentos dos extratos brutos, das fibras, das frações neutra, orgânica e básica e também da sobra, que foi calculada subtraindo do peso do extrato bruto, o peso das fibras e de todas as frações. Os rendimentos em gramas podem ser vistos na Tabela 15, a média e o desvio padrão das 19 amostras para as seis variáveis são mostradas nas duas últimas linhas da Tabela.

Extrato	Rendimento (g)						
	Bruto	Fibras	Neutra	Orgânica	Básica	Sobra	
е	0,8268	0,0399	0,2422	0,2114	0,0104	0,3229	
d	0,5525	0,0775	0,2934	0,1747	0,0051	0,0018	
h	0,3645	0,0556	0,0971	0,2089	0,0037	-0,0008	
а	0,535	0,0428	0,3168	0,1454	0,0037	0,0263	
ed	1,2417	0,1015	0,3722	0,3078	0,0286	0,4316	
eh	1,0423	0,0947	0,3650	0,1749	0,0181	0,3896	
еа	0,9518	0,0952	0,2948	0,2427	0,022	0,2971	
dh	0,811	0,3402	0,3009	0,1301	0,0054	0,0344	
da	0,5222	0,0593	0,296	0,1504	0,0061	0,0104	
ha	0,4852	0,0267	0,2916	0,1176	0,0027	0,0466	
edh	1,1249	0,1571	0,3726	0,2439	0,0386	0,3127	
eda	0,9481	0,0697	0,4352	0,2036	0,0142	0,2254	
eha	0,8224	0,0753	0,4271	0,1531	0,0114	0,1555	
dha	0,7708	0,2329	0,2676	0,1301	0,0091	0,1311	
r1	0,8557	0,0611	0,3875	0,189	0,0172	0,2009	
r2	0,8600	0,0765	0,3602	0,2249	0,0118	0,1866	
r3	0,8713	0,1168	0,4312	0,1788	0,0199	0,1246	
r4	0,8435	0,081	0,3796	0,196	0,0228	0,1641	
r5	0,8657	0,0868	0,4332	0,1913	0,0159	0,1385	
Média Dosvio podrão	0,8050	0,0995	0,3349	0,1881	0,0140	0,1683	
Desvio paulao	0,2202	0,0745	0,0051	0,0409	0,0095	0,1222	

Tabela 15: Rendimentos em gramas, média e desvio padrão dos rendimentos do extrato bruto, das fibras, das frações neutra, orgânica e básica e dos resíduos.

Primeiramente, as correlações entre as seis variáveis foram calculadas, os resultados podem ser vistos na Tabela 16. Os valores em negrito indicam significância no nível de 95% de confiança.

Altas correlações foram encontradas entre os rendimentos do extrato bruto e ambos sobra e fração básica. Isso significa que um aumento no rendimento do extrato bruto conforme a mistura de solventes utilizada na extração causa um aumento no rendimento da fração básica e também um aumento das sobras. Foi ainda observada, uma correlação moderada entre as sobras e o rendimento da fração básica. As fibras não apresentaram correlação com nenhuma outra variável. Não foi constatada nenhuma correlação negativa significativa entre as variáveis.

Bruto	Neutra	Orgânica	Básica	Fibras	Sobra
1,00	0,62	0,63	0,84	0,26	0,88
0,62	1,00	0,11	0,50	-0,00	0,35
0,63	0,11	1,00	0,71	-0,19	0,70
0,84	0,50	0,71	1,00	0,09	0,75
0,26	-0,00	-0,19	0,09	1,00	-0,05
0,88	0,35	0,70	0,75	-0,05	1,00
	Bruto 1,00 0,62 0,63 0,84 0,26 0,88	Bruto Neutra 1,00 0,62 0,62 1,00 0,63 0,11 0,84 0,50 0,26 -0,00 0,88 0,35	BrutoNeutraOrgânica1,000,620,630,621,000,110,630,111,000,840,500,710,26-0,00-0,190,880,350,70	BrutoNeutraOrgânicaBásica1,000,620,630,840,621,000,110,500,630,111,000,710,840,500,711,000,26-0,00-0,190,090,880,350,700,75	BrutoNeutraOrgânicaBásicaFibras1,000,620,630,840,260,621,000,110,50-0,000,630,111,000,71-0,190,840,500,711,000,090,26-0,00-0,190,091,000,880,350,700,75-0,05

Tabela 16: Correlações entre as seis variáveis no nível de 95% de confiança.

Aos rendimentos do extrato bruto, das fibras, de todas as frações e também das sobras foram aplicados os modelos linear, quadrático e cúbico especial. A análise da variância foi utilizada para testar a significância do modelo e a falta de ajuste na validação dos modelos ajustados. O erro padrão foi estimado pelo erro experimental determinado pelas cinco replicatas no ponto central.

5.2.1 Análise dos dados gravimétricos do rendimento do extrato bruto (Anexo II)

Todos os modelos aplicados aos rendimentos do extrato bruto, apesar de significativos, apresentaram falta de ajuste no nível de 95% de confiança.

Uma forma de avaliar os resultados nesses casos, é montar a figura do planejamento substituindo os valores experimentais em seus respectivos lugares, como pode ser visto na Figura 25. Pela Figura vê-se que dos quatro solventes puros, o etanol é o extrator que permite o maior rendimento do extrato bruto, 0,826 g, seguido do diclorometano, acetona e por último o hexano. É interessante notar que os maiores rendimentos são obtidos com misturas binárias e ternárias. O maior rendimento foi obtido com a mistura binária entre etanol e diclorometano, 1,241 g, que é muito maior do que a média entre os rendimentos obtidos com os dois solventes puros, 0,689 g. A mistura binária entre etanol e hexano permitiu o segundo melhor rendimento, cerca de 1,042 g. Altos efeitos de interação sinérgica binária também são observados entre os solventes etanol e acetona e entre diclorometano e hexano. Por outro lado, as misturas entre hexano e acetona e entre diclorometano e acetona não apresentaram efeito de interação.

Há também um efeito de interação sinérgico nas misturas ternárias compostas por etanol, diclorometano e hexano e entre etanol, diclorometano e acetona. Dessa forma, tem-se evidência

para afirmar que a presença do etanol na mistura é importante para obtenção de maiores rendimentos, principalmente em misturas binárias e ternárias.



Figura 25: Planejamento experimental Centróide-Simplex com os valores experimentais dos rendimentos do extrato bruto de *Erythrina*.

5.2.2 Análise de dados gravimétricos do rendimento das fibras (Anexo II)

Os modelos linear e quadrático aplicados aos rendimentos das fibras apresentaram falta de ajuste no nível de 95% de confiança. Assim, o modelo escolhido foi o cúbico especial, e a equação para este modelo é dada por:

```
\begin{aligned} \textbf{Fibras} &= 0,0390e + 0,0766d + 0,0547h + 0,0419a + 0,2019ed + 0,2185eh + 0,2461ea + \textbf{1,1253}dh \\ & (\pm 0,0305) (\pm 0,0305) (\pm 0,0305) (\pm 0,0305) (\pm 0,1487) (\pm 0,1487) (\pm 0,1487) (\pm 0,1487) \\ & + 0,0273da - 0,0593ha - \textbf{2,5509}edh - \textbf{1,5843}eda - \textbf{1,0260}eha + 0,8267dha \\ & (\pm 0,1487) (\pm 0,1487) (\pm 0,9333) (\pm 0,9333) (\pm 0,9333) (\pm 0,9333) \end{aligned} \end{aligned}
```

Este modelo foi validado através da análise de variância (ANOVA), mostrada na Tabela 17. O coeficiente de determinação R² para este ajuste é igual a 0,9530 e o valor máximo que R² poderia assumir corresponde a 0,9832.

A equação mostra que os coeficientes lineares não são significativamente diferentes de zero. Isso indica que a extração das fibras é favorecida por efeitos interação sinérgica ou antagônica. A única interação binária significativa no nível de 95% de confiança ocorre entre os solventes diclorometano e hexano, 1,1253, e a única interação ternária significativa mostra que quando o solvente etanol é adicionado à mistura diclorometano e hexano, há um efeito de interação antagônico, -2,5509.

Tabela 17: Análise da variância para o ajuste do modelo cúbico especial aos rendimentos das fibrasobtidas com o fracionamento dos extratos brutos de *Erythrina*.

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F* calculado
Modelo	0,094799	13	0,007292	7,80
Resíduos	0,004670	5	0,000934	
Falta de ajuste	0,002998	1	0,002998	7,17
Erro puro	0,001672	4	0,000418	
Total	0,099470	18	0,005526	

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{13,5}= 4,65 e F_{1,4} = 7,71.

Para uma mistura contendo 50% de diclorometano e 50% de hexano a equação para o modelo cúbico especial prevê um rendimento de 0,347 g, que está em ótima concordância com o valor experimental de 0,3402 g (Tabela 15).

Na Figura 26 é possível ver os resíduos deixados pelo ajuste versus as respostas previstas pelo modelo, juntamente com o gráfico das respostas previstas pelas observadas. Em ambos os gráficos vê-se que o extrato *dh* apresenta a maior resposta prevista.



Figura 26: Gráfico (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das respostas previstas pelas respostas observadas para os rendimentos das fibras.

As curvas de nível para o modelo cúbico especial aplicado ao rendimento das fibras podem ser vistas na Figura 27, onde o componente acetona foi igualado a zero para representar a superfície em apenas duas dimensões.

Pela inspeção visual da superfície de resposta, observa-se que a mistura entre os solventes diclorometano e hexano na proporção de 1:1 permitiria obter o maior rendimento das fibras.



Figura 27: Superfície de resposta para os componentes; etanol, diclorometano e hexano, aplicada aos rendimentos das fibras.

5.2.3 Análise dos dados gravimétricos dos rendimentos da fração neutra (Anexo II)

O modelo linear aplicado aos rendimentos da fração neutra apresentou falta de ajuste. Os modelos quadrático e cúbico especial não apresentaram falta de ajuste no nível de 95% de confiança. Devido à presença de um termo de interação significativo entre os componentes etanol, diclorometano e acetona, o modelo escolhido foi o cúbico especial, e a equação para este modelo é dada por:

 $\begin{aligned} \mathsf{FN} &= \mathbf{0}, \mathbf{2420}e + \mathbf{0}, \mathbf{2932}d + \mathbf{0}, \mathbf{0969}h + \mathbf{0}, \mathbf{3166}a + \mathbf{0}, \mathbf{4233}ed + \mathbf{0}, \mathbf{7871}eh + 0, 0669ea + \mathbf{0}, \mathbf{4283}dh \\ & (\pm 0, 0294) \ (\pm 0, 0294) \ (\pm 0, 0294) \ (\pm 0, 0294) \ (\pm 0, 1431) \ (\pm 0,$

A análise de variância (ANOVA) para esta regressão é mostrada na Tabela 18. O coeficiente de determinação R² para este ajuste é igual a 0,9652 e o valor máximo que R² poderia assumir corresponde a 0,9661.

Tabela 18: Análise da variância para o ajuste do modelo cúbico especial aos rendimentos da fraçãoneutra obtida com o fracionamento dos extratos brutos de *Erythrina*.

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F* calculado
Modelo	0,120171	13	0,009244	10,69
Resíduos	0,004324	5	0,000865	
Falta de ajuste	0,000105	1	0,000105	0,099
Erro puro	0,004218	4	0,001055	
Total	0,124494	18	0,006916	

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{13,5}= 4,65 e F_{1,4} = 7,71.

A equação mostra que os quatro coeficientes lineares são significativos no nível de 95% de confiança, e que entre eles, o hexano é significativamente menor que os outros três, ou seja, ele é o menos eficiente em extrair maiores quantidades da fração neutra. São observadas ainda três interações binárias sinérgicas significativas, entre os solventes etanol e diclorometano, etanol e

hexano e entre diclorometano e hexano, entre elas a mais importante ocorre entre os solventes etanol e hexano. Experimentalmente, essa mistura na proporção de 1:1 extrai cerca de 0,3650 g, porém a mistura em etanol e diclorometano permite extrair uma quantidade um pouco maior, 0,3722 g, essa diferença é parcialmente explicada pelos valores dos coeficientes lineares do hexano e do diclorometano. Há ainda um coeficiente de interação cúbico significativo entre os solventes etanol, diclorometano e acetona, sendo que ambos interagem sinergicamente. O modelo cúbico prevê a extração de 0,4300 g de fração neutra com essa mistura, que está em ótima concordância com o valor experimental de 0,4352 g.

Na Figura 28 é possível ver os resíduos deixados pelo ajuste versus as respostas previstas pelo modelo, juntamente com o gráfico das respostas previstas pelas observadas. A distribuição dos resíduos não revela tendências sistemáticas. Na Figura 28b vê-se que o extrato *h* apresentou a menor resposta observada, ficando alocada bem distante das demais, enquanto que os extratos *eda* e *eha* foram os mais efetivos na extração da fração neutra.



Figura 28: Gráficos (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das respostas previstas pelas respostas observadas para os rendimentos da fração neutra.

As curvas de nível para o modelo cúbico especial aplicado ao rendimento da fração neutra podem ser vistas na Figura 29, onde o componente hexano foi igualado a zero para representar a superfície em apenas duas dimensões.



Figura 29: Superfície de resposta para os componentes; etanol, diclorometano e acetona, aplicada aos rendimentos da fração neutra.

Pela análise visual da superfície de resposta, vê-se que o máximo fica em torno do ponto central, e que para aumentar o rendimento da fração neutra o melhor solvente seria uma mistura entre etanol, diclorometano e acetona na proporção de 1:1:1.

5.2.4 Análise dos dados gravimétricos do rendimento da fração orgânica (Anexo II)

O modelo linear aplicado aos rendimentos da fração orgânica apresentou falta de ajuste no nível de 95% de confiança, ao contrário dos modelos quadrático e cúbico especial. Apesar de não apresentar nenhum termo cúbico significativo, quando se muda do modelo quadrático para o modelo cúbico especial, as interações binárias entre etanol e acetona e entre diclorometano e hexano passaram a ser significativas.

Os dois modelos reproduziram bem os dados experimentais. Supondo a exclusão das quintuplicatas no ponto central e aplicando os modelos quadrático e cúbico especial aos resultados restantes para prever a quantidade extraída para essa mistura no nível de 95% de confiança, o modelo quadrático indicou a extração de 0,1848 g, enquanto que o cúbico especial indicou 0,1927 g. Os dois valores estão em ótima concordância com o valor médio das quintuplicatas, que foi de 0,1960 g.

Para decidir qual dos dois modelos é mais adequado, pode-se usar um teste F, onde se calcula a redução nos resíduos causada pela ampliação do modelo. A relação de interesse é dada por:

$$F = \frac{\left(SQ_{r,quad} - SQ_{r,cub}\right)}{MQ_{r,cub}}$$
(25)

onde SQ_{r,quad} representa a soma quadrática do resíduos do modelo quadrático, SQ_{r,cub} a soma quadrática dos resíduos do modelo cúbico especial, MQ_{r,cub} a média quadrática dos resíduos do modelo cúbico especial e *p* representa o número de parâmetros que o modelo cúbico especial tem a mais do que o modelo quadrático. Substituindo os valores tem-se:

$$F = \frac{\begin{pmatrix} 0,004633 - 0,001476 \end{pmatrix}}{0,000295} = 2,675423$$
(26)

Como o valor de F calculado é menor que o $F_{4,5,95\%}$ = 5,19, o teste indica que não há diferença significativa nas variâncias explicadas pelos dois modelos, dessa forma pode-se concluir que não vale a pena introduzir os parâmetros do modelo cúbico na equação.

Sendo assim tem-se duas razões pelas quais se descarta o modelo cúbico especial; a insignificância do teste F no nível de 95% de confiança e a ausência de um termo cúbico significativo de acordo com o erro. Dessa forma optou-se pelo modelo quadrático:

$$FO = 0,2115e + 0,1721d + 0,2042h + 0,1475a + 0,4667ed - 0,0949eh (\pm 0,0223) (\pm 0,0223) (\pm 0,0223) (\pm 0,0223) (\pm 0,0943) (\pm 0,0943) + 0,1803ea - 0,1526dh - 0,0673da - 0,2287ha (\pm 0,0943) (\pm 0,0943) (\pm 0,0943) (\pm 0,0943)$$
(27)

A análise de variância (ANOVA) para esta regressão é mostrada na Tabela 19. O coeficiente de determinação R² para este ajuste é igual a 0,8832 e o valor máximo que R² poderia assumir corresponde a 0,9697.

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F* calculado
Modelo	0,035042	9	0,003894	7,56
Resíduos	0,004633	9	0,000515	
Falta de ajuste	0,003431	5	0,000686	2,28
Erro puro	0,001202	4	0,000301	
Total	0,039674	18	0,002204	

Tabela 19: Análise da variância para o ajuste do modelo quadrático aos rendimentos da fração orgânica obtida com o fracionamento dos extratos brutos de *Erythrina*.

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{9,9}= 3,18 e F_{5,4} = 6,26.

A equação mostra que a mistura dos componentes etanol e diclorometano apresenta uma forte sinergia no nível de 95% de confiança, já a interação entre os solventes hexano e acetona atua de forma a diminuir o rendimento da fração orgânica, ou seja, possuem um efeito antagônico. Essas relações podem ser explicadas pelo fato de que a mistura *ed* apresentou o maior valor experimental para o rendimento da fração orgânica, 0,3078 g, enquanto que a mistura *ha* se mostrou menos eficiente, extraindo apenas 0,1176 g. Para essas duas misturas, o modelo quadrático prevê que a quantidade extraída da fração orgânica seria de 0,3085 g e 0,1186 g, respectivamente. Os valores previstos estão em ótima concordância com os valores experimentais.

Na Figura 30 é possível ver os resíduos deixados pelo ajuste versus as respostas previstas pelo modelo, juntamente com o gráfico das respostas previstas pelas observadas. A distribuição dos resíduos não revela tendências sistemáticas. Esses gráficos mostram claramente que o maior rendimento é observado para o extrato *ed* e o menor rendimento foi obtido com a mistura *ha*.



Figura 30: Gráficos (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das respostas previstas pelas respostas observadas para os rendimentos da fração orgânica.

As curvas de nível para o modelo quadrático aplicado ao rendimento da fração orgânica podem ser vistas na Figura 31, onde o componente acetona foi igualado a zero para representar a superfície em apenas duas dimensões.

Pela análise visual da superfície de resposta pode-se confirmar que o melhor extrator em termos de rendimento da fração orgânica seria uma mistura composta por etanol e diclorometano na proporção de 1:1, extrato *ed* do planejamento.



Figura 31: Superfície de resposta para os componentes; etanol, diclorometano e hexano, aplicada aos rendimentos da fração orgânica.

5.2.5 Análise dos dados gravimétricos do rendimento da fração básica (Anexo II)

O modelo linear aplicado aos rendimentos da fração básica apresentou falta de ajuste. Os modelos quadrático e cúbico especial não apresentaram falta de ajuste no nível de 95% de confiança. Devido à presença de um termo de interação significativo entre os componentes etanol, diclorometano e hexano, o modelo escolhido foi o cúbico especial, e a equação para este modelo é dada por:

```
 \begin{aligned} \mathbf{FB} &= 0,0103e + 0,0050d + 0,0036h + 0,0036a + \mathbf{0},\mathbf{0852}ed + 0,0460eh + \mathbf{0},\mathbf{0616}ea + 0,0058dh \\ & (\pm 0,0040) \ (\pm 0,0040) \ (\pm 0,0040) \ (\pm 0,0040) \ (\pm 0,0195) \ (\pm 0,0195)
```

A análise de variância (ANOVA) para esta regressão é mostrada na Tabela 20. O coeficiente de determinação R² para este ajuste é igual a 0,9508 e o valor máximo que R² poderia assumir corresponde a 0,9576.

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F* calculado
Modelo	0,001549	13	0,000119	7,403284
Resíduos	0,000080	5	0,000016	
Falta de ajuste	0,000011	1	0,000011	0,665350
Erro puro	0,000069	4	0,000017	
Total	0,001629	18	0,000091	

Tabela 20: Análise da variância para o ajuste do modelo cúbico especial aos rendimentos da fração básica obtida com o fracionamento dos extratos brutos de *Erythrina*.

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{13,5}= 4,65 e F_{1,4} = 7,71.

A equação mostra que os coeficientes lineares não são significativamente diferentes de zero. Isso indica que a extração da fração básica é favorecida por efeitos de interação sinérgica. Os dois efeitos de interação binária significativos no nível de 95% de confiança ocorrem entre os solventes etanol e diclorometano e entre etanol e acetona, e o de interação ternária ocorre entre os solventes etanol, diclorometano e hexano. Pela Tabela 15, observa-se que o maior rendimento

da fração básica foi alcançado usando como extrator a mistura ternária *edh*, 0,0386 g, o modelo cúbico especial prevê a extração de 0,0428 g para essa mesma mistura, uma diferença de apenas 0,0042 g.

Na Figura 32 é possível ver os resíduos deixados pelo ajuste versus as respostas previstas pelo modelo, juntamente com o gráfico das respostas previstas pelas observadas. A distribuição dos resíduos não revela tendências sistemáticas. Pelo gráfico 32b pode-se observar que os extratos de número *edh* e *ed* são os que apresentaram os maiores rendimentos da fração básica.



Figura 32: Gráficos (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das respostas previstas pelas respostas observadas para os rendimentos da fração básica.

As curvas de nível para o modelo cúbico especial aplicado ao rendimento da fração básica podem ser vistas na Figura 33, onde o componente acetona foi igualado a zero para representar a superfície em apenas duas dimensões.

Pela análise visual da superfície de resposta pode-se confirmar os resultados. Um maior rendimento seria obtido com a mistura dos três solventes, porém, a proporção de etanol e diclorometano na mistura deve ser um pouco maior do que a proporção de hexano.



Figura 33: Superfície de resposta para os componentes; etanol, diclorometano e hexano, aplicada aos rendimentos da fração básica.

5.2.6 Modelagem das sobras (Anexo II)

A última coluna da Tabela 15 mostra os valores das sobras estimadas, ou seja, das quantidades que não foram extraídas com as quatro frações. O modelo cúbico especial forneceu o melhor ajuste dos dados.

```
 S = 0,3222e + 0,0174d - 0,0014h + 0,0257a + 1,0666ed + 0,9362eh + 0,5120ea + 0,1252dh 
(\pm 0,0335) (\pm 0,0335) (\pm 0,0335) (\pm 0,0335) (\pm 0,1631) (\pm 0,1631) (\pm 0,1631) (\pm 0,1631) 
- 0,0249da + 0,1574ha - 1,4329edh - 2,3107eda - 4,1848eha + 1,9444dha 
(\pm 0,1631) (\pm 0,1631) (\pm 1,0237) (\pm
```

A análise de variância (ANOVA) para esta regressão é mostrada na Tabela 21. O coeficiente de determinação R² para este ajuste é igual a 0,9823 e o valor máximo que R² poderia assumir corresponde a 0,9871.

Tabela 21: Análise da variância para o ajuste do modelo cúbico especial às sobras obtidas pelo cálculo da diferença entre o extrato bruto e a soma de todas as outras frações obtidas com o fracionamento dos extratos brutos de *Erythrina*.

Fonte de variação	Soma quadrática	Graus de liberdade	Média quadrática	F* calculado
Modelo	0,312050	13	0,024004	21,36
Resíduos	0,005619	5	0,001124	
Falta de ajuste	0,001549	1	0,001549	1,522
Erro puro	0,004069	4	0,001017	
Total	0,317668	18	0,017648	

* Valor de F para significância do modelo dado por MQ_R/MQ_r e para falta de ajuste MQ_{faj}/MQ_{ep} . Valores críticos de F correspondentes no nível de 95% de confiança são F_{13,5}= 4,65 F_{1,4} = 7,71.

Pela equação, observa-se que no nível de 95% de confiança, dos quatro coeficientes lineares, apenas o etanol é significativo, sendo dessa forma o mais importante na maximização das sobras. O etanol também apresentou efeito binário de interação sinérgica com os outros três solventes e um ternário, de interação antagônica com hexano e acetona. A Tabela 15 indica que o maior resíduo foi obtido com a mistura *ed*, 0,4316 g. O modelo cúbico especial prevê que a sobra deixada por esta mesma mistura será de 0,4364 g, um valor muito próximo do obtido experimentalmente.

Na Figura 34 é possível ver os resíduos deixados pelo ajuste versus as respostas previstas pelo modelo, juntamente com o gráfico das respostas previstas pelas observadas. A distribuição dos resíduos não revela tendências sistemáticas. Pelo gráfico 34b pode-se observar que as misturas *ed* e *eh* foram as que apresentaram os maiores valores de sobra do fracionamento.



Figura 34: Gráficos (a) dos resíduos deixados pelo ajuste do modelo cúbico especial e (b) das respostas previstas pelas respostas observadas para os resíduos do fracionamento.

As curvas de nível para o modelo cúbico especial aplicado aos resíduos podem ser vistas na Figura 35, onde o componente hexano foi igualado a zero para representar a superfície em apenas duas dimensões.



Figura 35: Superfície de resposta para os componentes; etanol, diclorometano e acetona, aplicada aos resíduos do fracionamento.

Pela análise visual da superfície de resposta pode-se confirmar os resultados. Um maior resíduo seria obtido com a mistura binária dos solventes etanol e diclorometano, com etanol em maior proporção.

5.2.7 Análise de componentes principais dos dados gravimétricos dos rendimentos do extrato bruto e das frações (Anexo II)

Os valores de rendimento ainda foram dispostos numa matriz contendo 19 amostras e 6 variáveis. Esta matriz foi submetida à análise de componentes principais. Os grupos são definidos com base em sua localização no espaço e também na sua avaliação conjunta com os loadings. A primeira componente principal contém 58,65% da variância e separa os grupos I e II dos grupos III e IV. A segunda componente principal discrimina o grupo IV dos outros grupos e contém 19,56% da variância. O gráfico dos escores e dos loadings para estas duas componentes pode ser visto na Figura 36.



Figura 36: (a) Gráfico dos escores das duas primeiras componentes principais e (b) gráfico dos loadings das duas primeiras componentes principais.

Analisando o gráfico dos loadings na Figura 36b, observa-se que as amostras do grupo I, extratos preparados em *ed* e *edh*, influenciam fortemente os rendimentos do extrato bruto, das frações básica e orgânica e também dos resíduos, de fato, as amostras do grupo I possuem os maiores valores de rendimento para estas variáveis, isto pode ser confirmado através dos dados experimentais mostrados na Tabela 15. Na análise das correlações, essas variáveis se relacionaram com coeficientes de correlação que variaram de 0,63 a 0,88, como mostrado na Tabela 15. Essas observações ainda são consistentes com a similaridade das superfícies de respostas para essas variáveis e com a alta sinergia entre os solventes etanol e diclorometano no rendimento do extrato bruto. Os extratos *dh* e *dha* do grupo IV aparentam ter um rendimento elevado de fibras, caracterizado pelo alto valor de loadings na CP2, por outro lado, permitem extrair pouca quantidade da fração orgânica, que apresentou o mais baixo valor de loading na CP2.

A localização do grupo III, altos valores de escore na CP1 e baixos na CP2, indica que estes extratos são os menos eficientes na obtenção de maiores rendimentos de todas as variáveis. O grupo II, que é composto pelo extrato *e*, pelas misturas binárias *eh* e *ea*, pelas misturas ternárias *eda* e *eha* e pelas replicatas *r1-r5* é influenciado pelos rendimentos da fração neutra, ou seja, permitem obter um alto rendimento desta fração. Aqui também foi possível constatar que a dispersão das replicatas dentro do grupo II dá uma boa idéia da dimensão do erro experimental comparado à dispersão global das amostras no gráfico dos escores, que mede a dispersão dos pontos com relação às diferentes misturas de solventes utilizados.

5.2.8 Análise de componentes principais dos dados cromatográficos do extrato bruto

Os cromatogramas das replicatas do extrato bruto no ponto central (amostras r1 - r5) obtidos nas três fases móveis descritas no item 4.4.2 podem ser vistos na Figura 37. Na Figura observa-se que a FM1 permitiu a melhor separação dentro de um tempo de corrida razoável, enquanto que a FM3 teve o menor tempo de retenção, porém com pouca separação dos picos. Sendo assim, serão apresentados apenas os resultados referentes aos extratos brutos analisados na FM1.

Os valores de absorbância dos picos cromatográficos dos 19 extratos analisados na FM1 foram transformados em uma matriz composta por 19 amostras e 1201 variáveis (valores de absorbância) correspondentes ao tempo de corrida de 20 minutos. Para algumas amostras foi necessário ajustar o tempo de retenção, isso foi feito utilizando a opção de alinhamento do software Pirouette^{®71}. A matriz original foi normalizada para 1 e centrada na média e em seguida submetida à análise de componentes principais com rotação varimax.



Figura 37: Cromatogramas das replicatas do extrato bruto no ponto central nas fases móveis 1, 2 e 3.

As três primeiras componentes principais explicam cerca de 56,81% da variância total. A Figura 38 apresenta os gráficos dos escores das componentes principais 1 e 2 e 2 e 3. Nestes gráficos, vê-se a formação de 7 grupos. Os grupos são definidos com base em sua localização no espaço e também na sua avaliação conjunta com os loadings e com os cromatogramas originais. O grupo I é formado pelas amostras *e* e *ea*, o grupo II pelas amostras *eh* e *eha*, o grupo III pelas replicatas *r1* a *r5*, o grupo IV é formado pelas amostras *ed* e *edh*, o grupo V é formado pelas amostras *dh*, *eda* e *dha*, o grupo VI pelas amostras *d* e *da* e o grupo VII é formado pelas amostras *h*, *a* e *ha*.


Figura 38: Gráfico dos escores das componentes principais (a)1 e 2 e (b) 2 e 3 dos dados cromatográficos dos extratos brutos de *Erythrina* (FM1). A seta indica que as amostras *ed* e *edh* fazem parte do grupo IV.

A CP1 explica sozinha 26,05% da variância dos dados e permite diferenciar três regiões; a primeira é representada pelo grupo I, localizado na parte mais positiva, a segunda pelos grupos V, VI e VII, na parte mais negativa, e a terceira pelos grupos II, III e IV, que estão localizados na parte intermediária próxima de zero. A CP2, que apresenta 21,46% de informação, separa os grupos I, II e

VII dos grupos III, IV, V e VI. Por fim, a CP3, que explica 9,29% da variância dos dados, permite a discriminação dos grupos I, II, III e IV dos grupos V, VI e VII.

Os loadings das componentes principais 1, 2 e 3 estão apresentados na Figura 39. Esta Figura mostra que os valores mais altos de loadings da CP1 correspondem às substâncias com tempo de retenção em 2,20 min. Isto indica que as amostras do grupo I, alocadas na parte mais positiva da CP1, possuem os valores de escore mais elevados, e sofrem maior influência da variável com valor mais alto de loading, ou seja, são especialmente caracterizadas pela quantidade da substância ou mistura de substâncias com tempo de retenção em 2,20 min. As regiões mais importantes com valor de loading negativo correspondem aos tempos de retenção de 3,62 e 4,09 min e caracterizam os grupos IV, V, VI e VII, que possuem valor de escores negativos.



Figura 39: Gráfico dos loadings das componentes principais (a) 1, (b) 2 e (c) 3 para os dados cromatográficos dos extratos brutos de *Erythrina* (FM1).

Os loadings da CP2 mostram que as amostras que possuem valores de escores positivos, sofrem maior influência basicamente da substância com tempo de retenção em 12,24 min, enquanto que as amostras localizadas na parte mais negativa, grupo VII, são discriminadas principalmente devido à variável com tempo de retenção em 2,22 min.

Na parte positiva da CP3 a variável com tempo de retenção de 4,04 min caracteriza principalmente o grupo VI. Na parte negativa, as variáveis responsáveis pela separação são correspondentes aos tempos de retenção de 4,12 min.

A Figura 40 mostra os cromatogramas transformados dos grupos formados pela projeção dos escores das componentes principais 1, 2 e 3 com tempo de corrida entre 1,5 e 15 min. Na Figura pode-se constatar a presença dos picos indicados pelos loadings de maior e menor valor. O pico em 2,20 min que caracteriza o grupo I, o pico em 4,09 min presente nos grupos I, II, III e IV e ausente nos grupos V, VI e VII e o pico em 12,24 min presente apenas nos grupos III, IV, V e VI.



Figura 40: Cromatogramas transformados dos extratos brutos de *Erythrina* separados com base nos grupos formados pela PCA.

Para confirmar os agrupamentos da PCA, a análise hierárquica de agrupamentos utilizando o método de Ward foi aplicada à mesma matriz utilizada na análise de componentes principais. O dendrograma obtido pelo conjunto de 19 amostras e 1201 valores de absorbância está apresentado na Figura 41.



Figura 41: Dendrograma baseado nos dados cromatográficos dos extratos brutos de *Erythrina* obtido com o método de Ward.

No valor de distância de 1,50 é possível observar os mesmos agrupamentos formados na análise de componentes principais, exceto pelos grupos V e VI, que formaram um único grupo, apresentado maior similaridade, realmente, esses dois grupos apresentaram localizações muito próximas na projeção dos escores das três componentes principais. Através da comparação dos cromatogramas das amostras desses dois grupos foi possível verificar que as amostras do grupo VI não apresentam o pico de baixa intensidade em 4,10 min e também apresentam o pico em 12,24 min com intensidade um pouco maior.

A análise deste resultado mostra que as duas amostras do grupo I contém o solvente etanol na mistura extratora, as amostras do grupo II possuem etanol e hexano, as do grupo III foram preparadas com a mistura contendo os quatro solventes, as amostras do grupo IV possuem etanol e diclorometano, do grupo V possuem o solvente diclorometano em alguma proporção, as do grupo VI foram preparadas com diclorometano puro e em mistura com acetona, e as amostras do grupo VII foram preparadas com hexano puro, acetona pura e também com a mistura desses dois solventes.

Para averiguar se há alguma relação entre o rendimento e os escores obtidos tanto na fase de testes quanto na fase exploratória, foram calculadas as correlações entre os escores das três primeiras componentes principais da fase exploratória (CP), os escores das três primeiras componentes principais sem rotação varimax da fase de testes (CPt) e também dos rendimentos dos extratos brutos (Rendim.), os resultados podem ser vistos na Tabela 22. O valor em negrito indica significância no nível de 95% de confiança.

Tabela 22: Correlações entre os escores das três primeiras componentes principais da fase exploratória, os escores das três primeiras componentes principais sem rotação varimax da fase de testes e os rendimentos dos extratos brutos no nível de 95% de confiança.

Variáveis	CP1	CP2	СРЗ	CP1t	CP2t	CP3t
Rendim.	0,42	0,39	-0,68	-0,11	-0,07	0,41

A variável rendimento apresentou correlação significativa de valor negativo, mas não muito pronunciada, apenas com a CP3. Isso significa que um aumento no rendimento do extrato bruto conforme a mistura de solventes utilizada na extração, causa uma diminuição nos escores da CP3. As amostras responsáveis por essa correlação negativa, ou seja, que aumentam no rendimento e diminuem na CP3, foram *d*, *h*, *ed*, *ea*, *dh*, *ha*, *edh*, *eda* e *dha*, as replicatas apresentaram pouca variação com relação ao rendimentos, enquanto que na CP3 a variação foi um pouco maior, Figura 42.



Figura 42: Gráfico dos valores de escores da CP3 e dos rendimentos do extrato bruto.

É interessante notar que das nove amostras, sete possuem o solvente diclorometano na sua composição. Os loadings da CP3 indicam que as variáveis mais importantes possuem pico em 4,04 e em 4,12 min, nas partes positiva e negativa, respectivamente. Porém, não se pode afirmar que a presença ou ausência desses picos está de fato relacionada com o rendimento dos extratos brutos.

5.2.9 Análise de componentes principais dos dados cromatográficos da fração orgânica

Os cromatogramas das replicatas fração orgânica no ponto central (amostras *r1-r5*) obtidos nas três fases móveis descritas no item 4.4.2, podem ser vistos na Figura 43. Novamente observase que a FM1 permitiu a melhor separação. Dessa forma, o estudo comparativo dos cromatogramas da fração orgânica será feito com base nos cromatogramas obtidos com a análise cromatográfica feita usando a FM1.



Figura 43: Cromatogramas da fração orgânica no ponto central (replicatas), nas fases móveis (a) 1, (b) 2 e (c) 3.

Os valores de absorbância dos picos cromatográficos das 19 frações orgânicas analisadas na FM1 foram transformadas em uma matriz composta por 19 amostras e 1801 variáveis (valores de absorbância) correspondentes ao tempo de corrida de 30 minutos. Para algumas amostras, foi necessário ajustar o tempo de retenção. O conjunto de dados original foi normalizado por comprimento do vetor e centrado na média e em seguida submetido à análise de componentes principais.

Com esse tratamento, foi obtido um modelo que com três componentes explicou em torno de 74,36% da variância contida nos dados. A Figura 44 apresenta o gráfico dos escores das componentes principais 1 e 2 e 1 e 3. Neste gráfico, vê-se a formação de 5 grupos. Os grupos são definidos com base em sua localização no espaço e também na sua avaliação conjunta com os loadings e com os cromatogramas originais. O grupo I é formado pelas amostras *a*, *ha* e *dha*, o grupo II, pelas amostras *h*, *ea* e *r4*, o grupo III pelas amostras *e*, *eh* e *eda* e as replicatas *r1* a *r3* e *r5*, o grupo IV é formado pelas amostras *d*, *dh* e *da* e o grupo V é formado pelas amostras *ed*, *edh*, e *eha*.

A CP1 explica 40,08% da variância dos dados e permite a discriminação dos grupos I, II e III, dos grupos IV e V. A CP2, que apresenta 23,55% de informação, separa o grupo III dos grupos I e IV, sendo que os grupos II e V possuem amostras tanto na parte positiva como na negativa, porém próximas de zero. Por fim, na CP3, que explica 10,72% da variância dos dados, é possível discriminar os grupos II e V que estão na parte positiva, dos grupos I, III e IV.

86



Figura 44: Gráfico dos escores das componentes principais (a)1 e 2 e (b) 2 e 3 dos dados cromatográficos da fração orgânica de *Erythrina* (FM1). A seta indica que a amostra *a* faz parte do grupo I.

Os loadings das componentes principais 1, 2 e 3 podem ser visualizados na Figura 45. Os gráficos dos loadings analisados juntamente com o gráfico dos escores mostram que as amostras dos grupos I e II, alocadas na parte positiva da CP1, são especialmente caracterizadas pela quantidade da substância ou mistura de substâncias com tempo de retenção em 9,66 min. As duas variáveis mais importantes com valores de loading negativos na CP1 correspondem aos tempos de

retenção de 9,90 e 4,73 min, que caracterizam especialmente as amostras do grupo IV, que está localizado na parte mais negativa da CP1.



Figura 45: Gráfico dos loadings das componentes principais (a) 1, (b) 2 e (c) 3 dos dados cromatográficos da fração orgânica de *Erythrina* (FM1).

Os loadings da CP2 indicam que as amostras com escores positivos permitem extrair uma maior quantidade das substâncias com tempo de retenção em 9,88 min, enquanto que as amostras localizadas na parte negativa são discriminadas principalmente devido a variável com tempo de retenção em 4,80 min.

Na parte positiva da CP3 a variável que discrimina as amostras, em especial as amostras *ed* e *edh* do grupo V, é a substância com tempo de retenção em 4,73 min. Na parte negativa, as

variáveis responsáveis pela separação são correspondentes aos tempos de retenção em 4,63 e 5,78 min.

A Figura 46 mostra os cromatogramas transformados dos grupos formados pela projeção das componentes principais 1, 2 e 3 com tempo de corrida entre 1,5 e 20 min, onde é possível verificar a presença dos picos discriminados pelos loadings.



Figura 46: Cromatogramas tratados das frações orgânicas de *Erythrina* separados com base nos grupos formados pela PCA.

Para confirmar estes agrupamentos, a análise hierárquica de agrupamentos utilizando o método de Ward foi aplicada à mesma matriz utilizada na análise de componentes principais. O dendrograma obtido pelo conjunto de 19 amostras e 1801 valores de absorbância está apresentado na Figura 47.



Figura 47: Dendrograma baseado nos dados cromatográficos da fração orgânica de *Erythrina* obtidos com o método de Ward.

No valor de distância de 0,5 observa-se os mesmos agrupamentos formados na análise de componentes principais, sendo que o grupo III é subdividido em dois grupos. O primeiro formado pelas amostras *eh, eda, r2* e *r5* e o segundo pelas amostras *e, r1* e *r3*, os cromatogramas mostram que as amostras do segundo grupo possuem menor concentração dos picos em 5,78 min, essa subdivisão também pode ser vista no gráfico dos escores. A amostra *eha* do grupo V, também ficou separada, se alocando junto com as amostras do grupo IV, pelos cromatogramas desses dois grupos vê-se que o perfil cromatográfico é muito parecido e que a diferença entre eles é apenas com relação à intensidade do sinal. A projeção dos escores e o dendrograma mostraram que a replicata *r4* ficou separada do restante das replicatas. A visualização dos cromatogramas das replicatas mostrou que a amostra *r4* possui o mesmo perfil cromatográfico que as demais, porém com uma intensidade um pouco menor, e já em comparação com os perfis das amostras *h* e *ea* constatou-se uma ótima sobreposição.

A análise deste resultado mostra que a fração orgânica apresenta uma complexidade menor do que a do extrato bruto, pois, os cromatogramas não variam muito de um extrato para outro. Como a extração da fração orgânica pode ser vista como uma extração específica para compostos fenólicos e terpenóides, o que vai mudar é a capacidade que cada mistura extratora tem de extrair esses compostos em termos de concentração, ou seja, independente da mistura extratora o perfil cromatográfico da fração orgânica continuará praticamente o mesmo.

A correlação entre os escores das três primeiras componentes principais e o rendimento da fração orgânica foi calculada. As únicas correlações significativas foram entre o rendimento e ambos, CP2 e CP3, sendo que as correlações obtidas foram de -0,63 e 0,53, respectivamente. A Figura 48 mostra os gráficos dos valores de escore contra os rendimentos.

O valor de coeficiente negativo indica um comportamento oposto entre o rendimento e a CP2, as amostras que possuem esse comportamento contrário são *d*, *h*, *a*, *ed*, *dh*, *edh*, *eha*, *dha* e as replicatas *r*1 e *r*2, como pode ser visto na Figura 48a. É interessante notar que de todas essas misturas, apenas uma delas não possui o solvente diclorometano em sua composição.



Figura 48: Gráfico dos rendimentos da fração orgânica com os valores de escores da (CP2 e (b) CP3.

Os escores da CP3 e os valores de rendimento, apresentados na Figura 48b, se relacionam positivamente, com um coeficiente de 0,53. Analisando os loadings da CP3, observa-se que as amostras com valores positivos de escores possuem maior quantidade do pico com tempo de retenção em 4,73 min, enquanto que as amostras com valores de escore negativos possuem maior quantidade do pico em 5,78 min. A Figura 48b mostra que as amostra *ed, ea* e *edh* apresentaram os maiores valores tanto para os escores da CP3 quanto para o valores de rendimento, ou seja,

relaciona uma extração em maior quantidade da substância com pico no tempo de retenção de 4,73 min pelas amostras que apresentaram maiores rendimentos.

5.2.10 Análise de componentes principais dos dados cromatográficos da fração básica

Os cromatogramas das replicatas da fração básica no ponto central (amostras *r1-r5*) obtidos nas três fases móveis descritas no item 4.4.2, podem ser vistos na Figura 49. Novamente observase que a FM1 permitiu a melhor separação. Dessa forma, serão apresentados apenas os resultados referentes aos dados cromatográficos das frações básicas obtidos na FM1.



Os valores de absorbância dos picos cromatográficos das 19 frações básicas analisadas na FM1 foram colocadas na forma de uma matriz composta por 19 amostras e 1801 variáveis (valores de absorbância) correspondentes ao tempo de corrida de 30 minutos. Para estes dados foram testados alguns tipos de tratamento, porém, os melhores resultados foram obtidos sem nenhum pré-processamento. Dessa forma, a matriz original dos dados foi submetida à análise de componentes principais.

As três primeiras componentes principais explicam 96,60% da variância total contida nos dados. A Figura 50 apresenta o gráfico dos escores das componentes principais 1 e 2 e 2 e 3. Nestes gráficos, vê-se a formação de 8 grupos. Os grupos são definidos com base em sua

localização no espaço e também na sua avaliação conjunta com os loadings e com os cromatogramas originais.

O grupo I é formado pela amostra *edh*, o grupo II pelas replicatas *r1* a *r5*, o grupo III pelas amostras *eh*, *eda* e *eha*, o grupo IV é formado pelas amostras *a*, *h*, *ha* e *dha*, o grupo V é formado pelas amostras *d*, *dh* e *da* e os grupos VI, VII e VIII são formados pelas amostras *e*, *ea*, *ed*, respectivamente.



Figura 50: Gráfico dos escores das componentes principais (a) 1 e 2 e (b) 2 e 3 dos dados cromatográficos da fração básica de *Erythrina* (FM1). A seta indica que a amostra *eda* faz parte do grupo III.

A CP1 explica 87,86% da variância dos dados e apresenta três regiões importantes, que permitem discriminar o grupo I, localizado na parte mais positiva, dos grupos IV, V e VI localizados na parte menos positiva, e os grupos II, III, VII e VIII que se encontram na parte central. A CP2, que apresenta 5,50% de informação, separa principalmente o grupo I do grupo VIII. A CP3, que explica 3,23% da variância dos dados, permite a discriminação dos grupos I, III e IV dos grupos V, VI, VII e VIII, sendo que o grupo II, formado pelas replicatas, encontra-se em torno de zero.

Os loadings das componentes principais 1, 2 e 3 estão apresentados na Figura 51. Esta Figura mostra que todos os valores dos loadings da CP1 são positivos, sendo que os valores mais altos correspondem às substâncias com tempo de retenção em 4,80 e 3,61 min. Isto indica que a amostra *edh*, correspondente ao grupo I, alocada na parte mais positiva da CP1 é especialmente caracterizada pela presença dessas substâncias.



Figura 51: Gráfico dos loadings das componentes principais (a) 1, (b) 2 e (c) 3 dos dados cromatográficos da fração básica de *Erythrina* (FM1).

Os loadings da CP2 indicam que as amostras com escores positivos, caracterizando novamente a amostra *edh*, permitem extrair uma maior quantidade das substâncias com tempo de retenção em 5,83 min, enquanto que as amostras localizadas na parte negativa, especialmente a amostra *ed*, são discriminadas principalmente devido a variável com tempo de retenção em 3,70 min. Na parte positiva da CP3 a variável que melhor discrimina as amostras é a substância com tempo de retenção em 5,10 min. Na parte negativa, as variáveis responsáveis pela separação das amostras, principalmente amostra ea do grupo VIII, correspondem aos tempos de retenção de 4,76 e 5,73 min.

A Figura 52 mostra os cromatogramas originais dos grupos formados pela projeção das componentes principais 1, 2 e 3 com tempo de retenção entre 1,5 e 20,5 min, onde é possível averiguar a presença dos picos discriminados pela análise dos escores e dos loadings.



Figura 52: Cromatogramas das frações básicas de *Erythrina* separados com base nos grupos formados pela PCA.

Para confirmar estes resultados, a análise hierárquica de agrupamentos utilizando o método de Ward foi aplicada à matriz original utilizada na análise de componentes principais. O dendrograma obtido pelo conjunto de 19 amostras e 1801 valores de absorbância está apresentado na Figura 53. No valor de distância de 7,5_x10⁶ observa-se os mesmos agrupamentos formados na análise de componentes principais, exceto pela amostra *eda*, que ficou fora do grupo III.



Figura 53: Dendrograma baseado nos dados cromatográficos da fração básica de *Erythrina* obtido pelo método de Ward.

A análise deste resultado mostra que a fração básica apresenta perfis cromatográficos diferentes para as diferentes misturas extratoras usadas na obtenção do extrato bruto, ou seja, a diferença entre eles não é apenas quantitativa, como foi o caso dos cromatogramas obtidos para a fração orgânica e sim qualitativa. Dessa forma, pode-se afirmar que os compostos extraídos na fração básica vão depender da mistura extratora utilizada na preparação do extrato bruto.

A correlação entre os escores das três primeiras componentes principais e o rendimento da fração básica foi calculada. Foi constatada uma correlação altamente significativa entre o

rendimento e os escores da CP1, 0,91. Os loadings da CP1 indicam que a variável com maior peso na discriminação das amostras da fração básica possui tempo de retenção de 4,8 min. Esse resultado permite afirmar que as amostras com os maiores valores de rendimento apresentaram uma maior quantidade do pico com tempo de retenção em 4,8 min.

Os valores de escores e dos rendimentos da fração básica foram normalizados e colocados num gráfico, que pode ser visto na Figura 54. A Figura mostra que os valores de rendimento da fração básica variam essencialmente de acordo com a variação dos valores de escore da CP1.



Figura 54: Gráfico dos valores de escore da CP1 e dos rendimentos da fração básica normalizados.

5.2.11 Análise de segunda ordem dos dados CLAE-DAD do extrato bruto

Os dados cromatográficos dos extratos brutos e das frações foram obtidos através de um cromatógrafo equipado com detector por arranjo de diodos, dessa forma, foi possível obter os valores de absorbância em função do tempo de corrida em diferentes comprimentos de onda para cada um dos 19 extratos. Instrumentos como o cromatógrafo utilizado que fornecem como resposta uma matriz de dados para cada amostra geram dados de segunda ordem. A Figura 55 mostra os valores de absorbância em função do tempo de tempo de retenção e do comprimento de onda do extrato bruto usando como solvente extrator o etanol puro.



Figura 55: Gráfico tridimensional dos valores de absorbância em função do tempo de corrida e do comprimento de onda do extrato bruto em etanol puro.

Dentre vários métodos para análise de dados de segunda ordem, foram testados o método Tucker3 e a análise de fatores paralelos (PARAFAC). O PARAFAC em todos os casos se mostrou inferior. Avaliando o mesmo número de fatores, o PARAFAC indicava como importantes as mesmas variáveis em mais de uma combinação, ou seja, eram necessários 5 fatores para obter as mesmas informações de um modelo Tucker3 com 3 fatores em cada um dos 3 modos. Além disso, foram necessárias mais de 500 iterações até que o critério de convergência (10⁻⁶) fosse atingido, ou seja, até a diferença relativa entre duas iterações sucessivas fosse menor que 10⁻⁶. Dessa forma, o método Tucker3 foi o escolhido devido à sua capacidade de selecionar diferentes números de fatores para cada uma das dimensões e também por ser o mais indicado quando o objetivo é a análise exploratória dos dados.

As matrizes obtidas para cada amostra do extrato bruto foram colocadas lado a lado gerando um paralelepípedo. Sendo assim, o modelo Tucker3 foi aplicado a uma matriz tridimensional composta por 1200 valores de absorbância em função do tempo de corrida de 20 minutos (Modo A), 256 valores de absorbância em função do comprimento de onda correspondentes ao intervalo espectral de 190 a 450 nm (Modo B) e 19 amostras (Modo C).

Primeiramente, foi feito um teste para saber a combinação do número de fatores que permitiria uma maior porcentagem de variância explicada. O teste foi feito atribuindo 10 fatores

para cada uma das três dimensões. O teste permitiu 700 combinações diferentes entre os fatores, sendo que a soma máxima dos fatores das 3 dimensões foi 30 (10,10,10) e a mínima foi 3 (1,1,1). Dessas 700 combinações, foram selecionadas apenas aquelas que apresentaram a maior porcentagem de variância explicada para cada soma total dos fatores, Tabela 23. Por exemplo, para uma soma total de fatores igual a 5, tem-se as seguintes combinações (1,2,2), (2,1,2), (2,2,1), dessas três, é selecionada aquela com a maior porcentagem de variância explicada.

Os valores apresentados na Tabela 23 podem ser visualizados na Figura 56, onde está plotado o número total de fatores versus a porcentagem de variância explicada. Pela Figura, observa-se que a partir do número total 9 de fatores, a perda de variância explicada é de apenas 2,13% com relação à variância explicada usando um número total de 30 fatores, que é muito pouco se distribuído nos 3 modos, ou seja, não compensa adicionar mais fatores no modelo se a variância explicada é praticamente a mesma.

O número total 9 de fatores corresponde à combinação 3 x 3 x 3, ou seja, 3 fatores no modo A, 3 fatores no modo B e 3 no modo C. Esse modelo alcançou o critério de convergência após a quarta interação e explicou 97,72% da variância total.

Soma total	Variância	N ^o de fatores			Soma total	Variância explicada	N ^o de fatores		
de fatores	explicada (%)	A B C		C	de fatores	(%)	A	B	C
3	92,7260	1	1	1	18	99,6308	7	4	7
5	93,9420	2	2	1	19	99,6896	7	5	7
6	96,4559	2	2	2	20	99,7152	7	6	7
7	96,6384	3	2	2	21	99,7391	8	5	8
8	97,3606	3	2	3	22	99,7687	8	6	8
9	97,7174	3	3	3	23	99,7915	9	6	8
10	98,2122	4	2	4	24	99,8099	9	6	9
11	98,6226	4	3	4	25	99 <i>,</i> 8258	10	6	9
12	98,8954	5	3	4	26	99,8409	10	7	9
13	99,1024	5	3	5	27	99,8466	10	7	10
14	99,1773	6	3	5	28	99,8525	10	8	10
15	99,3330	6	3	6	29	99 <i>,</i> 8555	10	9	10
16	99,4954	6	4	6	30	99,8565	10	10	10
17	99,5507	6	5	6					

Tabela 23: Soma total de fatores, variância explicada e número de fatores nos modos A, B e C para os dados cromatográficos dos extratos brutos de *Erythrina* (FM1).



Figura 56: Variância explicada para cada número total de fatores dos modelos Tucker3 para os dados cromatográficos dos extratos brutos de *Erythrina* (FM1).

Para auxiliar numa interpretação mais detalhada acerca das relações entre os fatores nos diferentes modos, os valores de core das 27 combinações dos fatores foram considerados e podem ser vistos na Tabela 24. As dez primeiras combinações com os valores mais altos de core, ou seja, as dez combinações mais importantes dos fatores permitem explicar 99,94% da variação do core. A magnitude e o sinal do core devem ser levados em consideração na análise das diferentes combinações. A Tabela 24 mostra que algumas combinações apresentam valores positivos de core, enquanto outras apresentam valores de core com sinal negativo.

Ordom	Elemento do		Core	Variância	Ordom	Eler	nento	o do	Core	Variância	
Ordeni		core		(10 ⁷)	(%)	Ordeni		core		(10 ⁷)	(%)
1	1	1	1	-12,2089	94,85962	15	3	3	2	0,0385	0,00095
2	2	2	2	-2,0227	2,60395	16	1	1	3	-0,0336	0,00072
3	2	2	1	-1,3436	1,14887	17	2	3	2	0,0262	0,00044
4	3	1	3	-1,0731	0,73292	18	2	3	3	-0,0261	0,00044
5	1	3	3	0,6992	0,31121	19	3	2	1	-0,0234	0,00035
6	3	3	1	-0,4094	0,10670	20	1	2	1	0,0212	0,00029
7	3	1	2	-0,3555	0,08043	21	1	2	2	-0,0172	0,00019
8	2	2	3	0,2839	0,05132	22	2	1	3	-0,0146	0,00014
9	1	3	2	0,2827	0,05086	23	2	3	1	-0,0059	0,00002
10	1	1	2	0,2210	0,03110	24	1	3	1	-0,0039	0,00001
11	2	1	2	-0,1399	0,01246	25	3	2	3	-0,0023	0,00000
12	1	2	3	0,0661	0,00279	26	2	1	1	-0,0013	0,00000
13	3	3	3	0,0655	0,00273	27	3	1	1	0,0011	0,00000
14	3	2	2	0,0484	0,00150						

Tabela 24. Importância dos elementos do core em ordem decrescente de acordo com a variânciado core para o modelo Tucker3 para os dados cromatográficos dos extratos brutos de *Erythrina*(FM1).

A combinação de fatores mais importante é a (1,1,1) cujo valor do core é de $-12,2_x10^7$. O valor de core dos elementos da oitava e nona combinação são praticamente, isso quer dizer que as interações entre esses fatores possuem a mesma importância. As três primeiras combinações explicam 98,59% da variância total. Após a terceira combinação pouca variância é acrescentada à variância total, apenas 0,73%.

A Figura 57 apresenta os três fatores do modo A.

Analisando a Figura 57, observa-se que todas as variáveis possuem pesos com valores positivos nos loadings do fator 1 e que a variável com a maior importância possui tempo de retenção de 3,69 min. No fator 2 do modo A observa-se que as variáveis possuem pesos positivos e negativos, sendo que a variável com maior peso na parte positiva está no tempo de retenção de 12,19 min e as variáveis com maior peso na parte negativa, embora não sejam muito pronunciadas, possuem tempo de retenção de 3,65 e 3,74 min. No fator 3, verifica-se a existência de duas regiões mais importantes com valores de pesos negativos, que possuem tempos de

retenção em 3,64 e 3,75 min, e na parte positiva a variável mais importante possui tempo de retenção de 3,69 min.



Figura 57: Loadings dos fatores 1, 2 e 3 do modo A do modelo Tucker3 para os dados tridimensionais dos extratos brutos de *Erythrina* (FM1).

A Figura 58 apresenta os três fatores dos modos B, onde os fatores 2 e 3 apresentam variáveis com valores de pesos positivos e negativos, e o fator 1 apresenta variáveis importantes apenas com valores de pesos positivos. As variáveis mais importantes para o fator 1 possuem máximo nos comprimentos de onda de 198 e 260 nm. Para o fator 2, com valor de peso positivo o máximo é no comprimento de onda acerca de 201 nm e com valor negativo, embora não muito pronunciado, no comprimento de onda de 263 nm. A variável com valor de peso negativo mais importante para o fator 3 encontra-se em 268 nm, enquanto que as duas com valores de pesos positivos estão em 241 e 297 nm.



Figura 58: Loadings dos fatores 1, 2 e 3 do modo B do modelo Tucker3 para os dados tridimensionais dos extratos brutos de *Erythrina* (FM1).

Na análise de dados cromatográficos e espectroscópicos, deve-se lembrar que deslocamentos de picos devido a fatores como pequenas variações na composição de fase móvel, vazão irregular, etc., e desvios no comprimento de onda máximo de absorção devido ao efeito batocrômico, variações de temperatura, por exemplo, são suscetíveis de ocorrer nesse tipo de investigação e devem ser levados em conta na interpretação dos resultados, uma vez que não foi utilizado nenhum método para correção dos tempos de retenção.

A Figura 59 apresenta os três fatores do modo C.



Figura 59: Escores dos fatores 1, 2 e 3 do modo C do modelo Tucker3 para os dados tridimensionais dos extratos brutos de *Erythrina* (FM1).

No fator 1 do modo C, verifica-se que todos os objetos possuem valores negativos de escore. As magnitudes dos pesos ou escores num modo são proporcionais à sua contribuição para a construção daquele fator. Quanto maior o peso ou escore, maior a contribuição. Neste caso, as amostras ou objetos mais importantes são *a*, *ea*, *da*, *ha*, *eda*, *eha*, *dha* e as replicatas. Nos fatores 2 e 3 verifica-se valores de escores positivos e negativos. No fator 2, na parte positiva, o objeto ou amostra *a* apresenta o mais alto valor de escore e o valor de escore mais negativo é observado para a amostra *d*. No fator 3, os escores mais positivos foram obtidos com as amostras *a*, *d* e *da*, e os com valores negativos foram as replicatas acompanhadas das misturas ternárias *eha* e *dha*.

Para correlacionar essas informações, serão utilizadas as três combinações mais importantes em termos de valor do core, mostrados na Tabela 24.

A combinação dos fatores (1,1,1) possui um valor de core de $-12,2_{x}10^{7}$ e permite explicar cerca de 95% da variância do core. De acordo com os gráficos dos loadings dos fatores 1, 1 e 1 dos

modos A, B e C, respectivamente, esta combinação indica que a substância com pico no tempo de retenção de 3,69 min possui um perfil espectral muito parecido com o perfil representado pelos loadings do fator 1 do modo B com máximo em torno do comprimento de onda de 260 nm e é extraída em maior quantidade pelas amostras que apresentam valores mais negativos de escore no fator 1 do modo C. Dessa forma, tem-se que a extração do pico com tempo de retenção de 3,69 min está relacionada com a quantidade do solvente acetona na mistura. A amostra *a* possui o valor de escore mais negativo, seguido das misturas binárias, das misturas ternárias e por fim das misturas quaternárias contendo acetona.

As amostras com valores de escore pouco negativos apresentam em sua composição, todos os solventes, exceto a acetona, o solvente ou mistura de solventes utilizados no preparo dessas amostras permitem extrair uma quantidade muito baixa da substância ou mistura de substâncias representada pelo pico com tempo de retenção em 3,69 min. A avaliação dos dados originais confirma esse resultado, como pode ser visto na Figura 60, que mostra os cromatogramas das amostras que contém e que não contém acetona em sua composição no intervalo de 1,5 a 7 min.

A Figura 60b mostra que algumas amostras apresentam uma pequena absorbância no tempo de retenção de 3,69 min.



Figura 60: Cromatogramas dos extratos brutos de *Erythrina* que (a) possuem e que (b) não possuem acetona na composição do solvente extrator no intervalo de tempo de corrida de 1,5 a 7 min.

A Figura 61 mostra os espectros extraídos para os picos com tempo de retenção em aproximadamente 3,69 min no intervalo de comprimento de onda de 190 a 325 nm. Na Figura 61 é possível averiguar que os espectros dos extratos preparados com acetona em sua composição são diferentes daqueles obtidos das amostras que não contém acetona na mistura extratora, embora o perfil espectral seja parecido e a absorção máxima ocorra num comprimento de onda muito próximo. Isso é um indicativo de essas substâncias podem ser do mesmo grupo de compostos.



Figura 61: Espectros de absorção no UV referentes ao pico com tempo de retenção de 3,69 min dos extratos brutos de *Erythrina* que (a) possuem e que (b) não possuem acetona na composição do solvente extrator.

A segunda combinação de fatores mais importante é a (2,2,2), que explica 2,60% da variância total do core e apresenta um valor de core negativo. Combinando valores positivos, positivos e negativos dos loadings dos fatores 2, 2 e 2 dos modos A, B e C, respectivamente, podese afirmar que todas as amostras contendo diclorometano em sua composição, em especial a amostra *d*, permitem extrair uma maior quantidade da substância com pico no tempo de retenção em 12,19 min, e que o espectro dessa substância deve possuir um máximo de absorção em 201 nm. A Figura 62 confirma este resultado. As amostras *e*, *h* e *eh* apresentaram loadings negativos no fator 2 do modo C, mas pela Figura vê-se que estas amostras não apresentaram o pico com tempo de retenção em 12,19 min, porém a absorbância neste tempo de retenção não é exatamente zero, o fato delas se localizarem na região negativa também pode ser devido a outros picos, como os picos em 4,07 e em 2,73 min.



Figura 62: Cromatogramas dos extratos brutos de *Erythrina* que (a) possuem e que (b) não possuem diclorometano na composição do solvente extrator no intervalo de tempo de corrida de 7 a 15 min.

A Figura 63 mostra os espectros extraídos dos picos com tempo de retenção em 12,19 min para as amostras que contém diclorometano em sua composição e também os espectros de picos com tempo de retenção próximos de 12,19 min.



Figura 63: Espectros de absorção no UV referentes ao pico com tempo de retenção de 12,19 min dos extratos brutos de *Erythrina* que (a) possuem e que (b) não possuem diclorometano na composição do solvente extrator.

A Figura 63b mostra que os espectros dos picos em 12,15, 12,10 e 12,38 min das amostras *h*, *a* e *ha*, respectivamente, parecem ter o mesmo perfil espectral das amostras que possuem diclorometano em sua composição, porém em uma quantidade significativamente menor. Já as outras amostras apresentaram um perfil espectral diferente.

A combinação (2,2,1) apresenta valor de core negativo. Combinando somente valores de loadings negativos nos três modos, reforçam-se os resultados da combinação (1,1,1) que indicam que as amostras contendo o solvente acetona em sua composição permitem extrair uma maior quantidade da substância com pico no tempo de retenção em torno de 3,65 e 3,74 min, com espectro de absorção máxima variando em torno de 263 nm.

Das combinações que envolvem o fator 3 do modo B, as mais importantes são aquelas que envolvem o fator 2 do modo A, (2,3,1), (2,3,2) e (2,3,3), pois além do pico em 12,19 min, outros dois picos nos tempos de retenção de 2,73 e 4,07 min também podem ser considerados importantes, embora a intensidade destes seja muito menor.

A combinação (2,3,1) possui um valor de core de -5,9x10⁴, combinando valores positivos nos modos A e B e negativos no modo C, vê-se que todas as amostras são influenciadas pelos comprimentos de onda máximos em 198, 241 e 297 nm. Porém, é necessário averiguar à qual dos dois picos, em 2,73 ou 4,07 min, esta informação está relacionada. Para isso foram recuperados os espectros dos picos das amostras do extrato bruto nos tempo de retenção em torno de 2,73 e 4,07 min. Na Figura 64 é possível ver os cromatogramas das amostras do extrato bruto separados com base na intensidade do pico em 4,05 min, no intervalo de 1,5 a 7 min.

Ao comparar as Figuras, vê-se que as amostras que possuem etanol na composição do solvente extrator possuem maior quantidade do pico em 4,05 min e também que o pico em 2,73 min apresenta um ombro, indicando a possibilidade de sobreposição com outro pico. As outras amostras, que não possuem etanol, Figura 64b, apresentam ausência ou então uma baixíssima quantidade do pico em 4,05 min, e o pico em 2,73 min encontra-se sem o ombro e com boa resolução.

108



Figura 64: Cromatogramas dos extratos brutos de *Erythrina* que possuem (a) alta e (b) baixa intensidade do pico com tempo de retenção de 4,05 min, no intervalo de tempo de corrida de 1,5 a 7 min.

A Figura 65 mostra os espectros extraídos somente daquelas amostras que apresentaram picos em torno do tempo de retenção de 4,07 min.



Figura 65: Espectros de absorção no UV referentes ao pico com tempo de retenção de 4,07 min dos extratos brutos de *Erythrina* que (a) possuem e que (b) não possuem etanol na composição do solvente extrator.

Pela Figura 65 é possível observar que os espectros das amostras *d*, *h*, *dh* e *da* possuem um perfil espectral diferente das amostras da Figura 65a. A combinação (2,3,1) permite relacionar o pico em 4,07 min com os comprimentos de onda de 198, 241 e 297 nm, porém estes não são comprimentos de onda de absorção máxima e sim comprimentos de onda onde ocorrem os mínimos.

Ao avaliar os espectros das amostras que apresentaram picos em torno do tempo de retenção de 2,73 min, foi observado que estes possuíam máximo em 199 nm e mínimo em 241 nm.

A Figura 66 mostra os espectros para os picos com tempo de retenção em torno de 2,73 min separados com base na intensidade de absorção no comprimento de onda de 268 nm.



Figura 66: Espectros de absorção no UV referentes ao pico com tempo de retenção de 2,73 min dos extratos brutos de *Erythrina* que apresentaram (a) menor, (b) média e (c) maior intensidade de absorção no comprimento de onda de 268 nm.

De com acordo com os espectros, o resultado da combinação (2,3,1) também pode ser atribuído ao pico em 2,73 min. A Figura 66a mostra que as amostras que não possuem o solvente extrator etanol em sua composição tendem a apresentar uma menor absorção no comprimento de onda de 268 nm. A Figura 66c indica que os espectros das amostras *eha*, *r3*, *r4* e *r5* são diferentes das demais, pois além da absorção máxima ocorrer em 204 nm, o tempo de retenção também é diferente, estes espectros foram extraídos porque em muitos casos, a variação no tempo de retenção pode ser atribuída a um deslocamento no tempo de retenção.

Os resultados do método Tucker3 aplicado aos dados tridimensionais dos extratos brutos permitiram discriminar as amostras com relação aos solventes extratores etanol, diclorometano e acetona, isso demonstra a seletividade desses solventes na extração.

Comparando estes resultados com aqueles obtidos através da análise de componentes principais aplicada somente às absorbâncias em função do tempo de corrida, observa-se que em ambos os casos, foi possível discriminar as amostras com relação aos solventes extratores etanol, diclorometano e acetona. Os resultados da análise de componentes principais foram mais específicos, pois permitiu separar as 19 amostras em sete grupos diferentes com base em quatro picos mais importantes, sendo que os resultados do método Tucker3 permitiram discriminar três grupos. Claro que o método Tucker3 usa uma quantidade muito maior de informação, dessa forma os dois métodos tornam-se complementares.

5.2.12 Análise de segunda ordem dos dados CLAE-DAD da fração orgânica

O modelo Tucker3 também foi aplicado aos dados tridimensionais da fração orgânica. A matriz tridimensional era composta por 1500 valores de absorbância em função do tempo de corrida de 25 minutos (Modo A), 256 valores de absorbância em função do comprimento de onda correspondentes ao intervalo espectral de 190 a 450 nm (Modo B) e 19 amostras (Modo C).

O teste para saber a combinação do número de fatores que permitiria uma maior porcentagem de variância explicada foi feito atribuindo 10 fatores para cada uma das três dimensões. Das 700 combinações diferentes foram selecionadas apenas aquelas que apresentaram a maior porcentagem de variância explicada para cada soma total dos fatores, Tabela 25.

111

Soma total	Variância	N ^o de fatores			Soma total	Variância	N ^o de fatores		
de fatores	explicada (%)	nos modos A B C		dos C	de fatores	explicada (%)	n A	os mo B	dos C
3	59,270	1	1	1	18	98,820	8	3	7
5	93,111	2	1	2	19	98,929	9	3	7
6	93,216	2	2	2	20	99,016	9	4	7
7	95,862	3	1	3	21	99,096	10	4	7
8	96,024	3	2	3	22	99,164	10	4	8
9	96,943	4	1	4	23	99,225	10	5	8
10	97,172	4	2	4	24	99,262	10	6	8
11	97,531	5	2	4	25	99,294	10	6	9
12	97 <i>,</i> 855	5	2	5	26	99,314	10	6	10
13	97 <i>,</i> 963	6	2	5	27	99,33	10	7	10
14	98,284	6	2	6	28	99,335	10	8	10
15	98,413	7	2	6	29	99,34	10	9	10
16	98,549	7	3	6	30	99,343	10	10	10
17	98,672	8	3	6					

Tabela 25: Soma total de fatores, variância explicada e número de fatores nos modos A, B e C para os dados cromatográficos da fração orgânica de *Erythrina* (FM1).

Observando os 13 primeiros valores na Tabela 25, percebe-se que as diferenças mais representativas, embora pequenas, ocorrem quando se aumenta o número de fatores dos modos A e C simultaneamente, enquanto que quando os modos A e C são mantidos fixos e aumenta-se somente o número de fatores do modo B, pouca variância é acrescentada ao modelo. Os valores apresentados na Tabela 25, exceto a combinação (1, 1, 1), podem ser visualizados na Figura 67, onde está plotado o número total de fatores versus a porcentagem de variância explicada.



Figura 67: Variância explicada para cada número total de fatores dos modelos Tucker3 para os dados cromatográficos da fração orgânica de *Erythrina* (FM1).

Pela Figura 67, vê-se que a partir do número total 9 de fatores, a perda de variância explicada é de apenas 2,4% com relação à variância explicada usando um número total de 30 fatores. O número total 9 de fatores corresponde à combinação 4 x 1 x 4, ou seja, 4 fatores no modo A, 1 fator no modo B e 4 no modo C. Esse modelo alcançou o critério de convergência após a segunda interação e explicou 96,94% da variância total.

Considerando a variância explicada, das 16 combinações diferentes, as mais importantes foram as quatro primeiras, que podem ser vistas na Tabela 26. Juntas, as quatro combinações explicam 99,98% da variação do core.

Tabela 26. Importância dos elementos do core em ordem decrescente de acordo com a variância do core para o modelo Tucker3 para os dados cromatográficos da fração orgânica de *Erythrina* (FM1).

Ordem	Ele	mento do c	ore	Core (10 ⁷)	Variância (%)		
1	1	1	1	-4,06	61,10		
2	2	1	2	-3,07	34,94		
3	3	1	3	-0,87	2,83		
4	4	1	4	0,54	1,11		

A Tabela 26 mostra que as três primeiras interações possuem valores de core com sinal negativo. A combinação de fatores mais importante é a (1,1,1) cujo valor de core é de $-4,06_{x}10^{7}$. A Figura 68 apresenta os quatro fatores do modo A.



Figura 68: Loadings dos fatores 1, 2, 3 e 4 do modo A do modelo Tucker3 para os dados tridimensionais da fração orgânica de *Erythrina* (FM1).

Analisando a Figura 68a, observa-se que todas as variáveis possuem pesos com valores positivos nos loadings do fator 1 e que as variáveis mais importantes possuem tempo de retenção de 4,80, 4,97, 5,82 e 6,06 min. No fator 2 do modo A observa-se que as variáveis possuem pesos positivos e negativos, sendo que as mais importantes são as mesmas do fator 1, porém, as variáveis com tempo de retenção em 4,97 e 6,07 min apresentam maior peso na parte positiva e as variáveis em 4,80 e 5,81 min maior peso na parte negativa. No fator 3, a variávei mais importante com valor de peso positivo possui tempo de retenção de 5,89 min e as variáveis com valores de

pesos negativos possuem tempo de retenção em 2,33, 4,77 e 5,77 min. No fator 4, as variáveis com valores de pesos positivos mais importantes, possuem tempo de retenção de 2,72, 6,12 e 10,20 min e as variáveis com valores de pesos negativos em 4,94 e 6,01 min.

A Figura 69 apresenta o fator 1 do modo B, onde observa-se que todos os valores de loading são positivos. Os máximos ocorrem nos comprimentos de onda de 205, 234 e 282 nm.



Figura 69: Loadings do fator 1 do modo B do modelo Tucker3 para os dados tridimensionais da fração orgânica de *Erythrina* (FM1).

A Figura 70 apresenta os quatro fatores do modo C.


Figura 70: Escores dos fatores 1, 2, 3 e 4 do modo C do modelo Tucker3 para os dados tridimensionais da fração orgânica de *Erythrina* (FM1).

No fator 1 do modo C, vê-se que todos os objetos possuem valores negativos de escore, sendo que as amostras *ed* e *eh* possuem os valores mais negativos. Nos fatores 2, 3 e 4 verifica-se valores de escores positivos e negativos. No fator 2, na parte positiva, o objeto ou amostra *eh* apresenta o mais alto valor de escore e o valor de escore mais negativo é observado para a amostra *ed*. No fator 3, os escores mais positivos foram obtidos com as amostras *eh* e *r2*, e os com valores mais negativos foram obtidos com as replicatas *r1* e *r3*. No fator 4, as amostra *d*, *dh*, *da*, *edh* e *eha* foram as mais importantes com valores de escores positivos e a amostra *ed* a mais importante com valor de escore negativo.

A combinação dos fatores (1,1,1) possui um valor de core de $-4,06_{x}10^{7}$ e permite explicar 61,1% da variância do core. Esta combinação permite relacionar as substâncias com pico nos tempos de retenção de 4,80, 4,97, 5,82 e 6,06 min, às amostras *ed* e *eh*. Os picos com tempos de retenção em 4,80 e 4,97 min referem-se à mesma substância, pois a este conjunto de dados não

foi aplicado nenhum tipo de tratamento para alinhamento dos picos. O mesmo ocorre para os picos em 5,82 e 6,06 min. Na Figura 71 é possível ver o deslocamento desses picos nos cromatogramas das amostras da fração orgânica, no intervalo de tempo de corrida de 4,5 a 6,5 min.



Figura 71: Cromatogramas da fração orgânica de *Erythrina* no intervalo de tempo de corrida de 4,5 a 6,5 min.

Na Figura 71 observa-se que o deslocamento dos dois picos ocorre para as mesmas amostras. Os espectros referentes a esses dois picos foram extraídos e podem ser visualizados na Figura 72. Pela Figura constata-se que os espectros das amostras com deslocamento no tempo de retenção são similares àqueles das amostras que não apresentaram deslocamento. Também é possível notar a semelhança entre os espectros das substâncias com pico em 4,80 e 5,80 min, sendo que os máximos ocorrem no mesmo comprimento de onda, a diferença entre eles seria apenas com relação à intensidade de absorbância. Este fato é de se esperar, pois neste caso, a fração orgânica representa uma extração específica a partir do extrato bruto para terpenóides e compostos fenólicos. Dessa forma, a semelhança entre esses espectros indica que essas substâncias devem pertencer à mesma classe de compostos.



Figura 72: Espectros de absorção no UV referentes aos picos com tempo de retenção de (a) 4,80 e (b) 5,80 min das amostras da fração orgânica de *Erythrina*.

A combinação dos fatores (2,1,2) possui um valor de core de $-8,07_{x}10^{7}$ e permite explicar 34,94% da variância do core. Esta combinação separa as amostras da combinação anterior. A combinação de valores negativos, positivos e positivos, dos loadings do modo A, modo B e escores do modo C, respectivamente, permite relacionar as substâncias com pico nos tempos de retenção de 4,80, e 5,81 min à amostra *eh* e a combinação de valores positivos, positivos e negativos, relaciona os picos em 4,97 e 6,07 min à amostra *ed*.

A combinação dos fatores (3,1,3) possui um valor de core de $-0.87_{x}10^{7}$ e permite explicar 2,83% da variância do core. A combinação de valores negativos, positivos e positivos, indica que a amostra *eh* e a replicata *r2* permitem extrair uma maior quantidade da substância com pico no tempo de retenção de 5,77 min e a combinação de valores positivos, positivos e negativos, indica que as replicatas *r1* e *r3* permitem extrair uma maior quantidade da substância com pico no tempo de retenção de 5,89 min. Esses picos estão relacionados à mesma substância, porém com um pequeno deslocamento. Na Figura 71, essas afirmações podem ser verificadas.

Por fim, a combinação dos fatores (4,1,4) possui um valor de core de $0.54_{x}10^{7}$ e permite explicar somente 1,11% da variância do core. A combinação de valores negativos, positivos e negativos, relaciona os picos com tempo de retenção de 4,94 e 6,01 min com a amostra *ed* e a combinação de valores positivos nos três modos relaciona os picos em 2,72, 6,12 e 10,20 min com

o restante das amostras que apresentaram deslocamento no tempo de retenção, que foram, *d*, *dh*, *da*, *edh* e *eha*.

A Figura 73 mostra os cromatogramas das amostras da fração orgânica no intervalo de tempo de corrida de 2,5 a 3,0 min e 9,2 a 10,6 min. Na Figura 73a verifica-se que nos picos com tempo de retenção em 2,72 min não houve grandes deslocamentos, já na Figura 73b constata-se que as amostras com pico em 10,20 min são as mesmas que apresentaram deslocamento nos picos em 4,80 e 5,80 min, sendo que as outras amostras apresentaram este mesmo pico com um deslocamento generalizado em torno do tempo de retenção de 9,75 min.



Figura 73: Cromatogramas da fração orgânica de *Erythrina* no intervalo de tempo de corrida de (a) 2,5 a 3,0 min e (b) 9,2 a 10,6 min.

Os espectros dos picos com tempo de retenção em 2,72 e em torno de 9,75 e 10,20 min foram extraídos e estão apresentados na Figura 74.



Figura 74: Espectros de absorção no UV referentes aos picos com tempo de retenção de (a) 2,72 min, (b) em torno de 9,75 e 10,20 min das amostras da fração orgânica de *Erythrina*.

A Figura 74a mostra que todos os picos com tempo de retenção de 2,72 min das amostras da fração orgânica apresentaram o mesmo perfil espectral, sendo que a absorção máxima ocorre no comprimento de onda de 199 nm. Na Figura 74b são mostrados os espectros de absorção das amostras que apresentaram picos com tempo de retenção em torno 9,75 e 10,20 min, onde a amostra *ed* apresentou uma maior absorbância enquanto que a amostra *h* apresentou a menor. Apesar da similaridade entre esses espectros, é possível notar que existem pequenas diferenças entre alguns espectros. Para uma melhor visualização dessas diferenças, esses espectros foram normalizados para 1 e separados em três grupos com base nos perfis espectrais, como mostra a Figura 75.



Figura 75: Espectros de absorção no UV normalizados referentes aos picos com tempo de retenção em torno de 9,75 e 10,20 min das amostras da fração orgânica de *Erythrina*, separados com base no perfil espectral.

Os espectros das Figuras 75a e 75b são muito semelhantes, porém os máximos ocorrem em comprimento de onda diferentes. Este fato pode ser devido a algum tipo de deslocamento ou esses espectros são realmente de substâncias diferentes, o interessante é que o grupo da Figura 75b é formado pelas amostras *h*, *a*, *ha* e *dha*. Já o perfil espectral das amostras da Figura 75c é bem diferente das outras, sendo que as amostras desse grupo são aquelas que apresentaram deslocamento no tempo de retenção, ou seja, o pico dessa substância foi no tempo de retenção de 10,20 min. Isso indica que existe uma possibilidade de que haja três substâncias diferentes com

picos em tempos de retenção muito próximos. Uma alternativa neste caso é enfraquecer ainda mais a FM, para que estes compostos possam ser eluídos em diferentes tempos de retenção.

Os resultados do método Tucker3 aplicado aos dados tridimensionais da fração orgânica permitiram discriminar as amostras mesmo com um deslocamento significativo no tempo de retenção de alguns picos, porém as discriminações foram baseadas na capacidade extratora de alguns solventes em termos de absorbância, visto que a diferença entre as amostras foi apenas quantitativa, embora para o pico com tempo de retenção em torno de 10,20 min tenha sido constatado que os espectros eram diferentes daqueles com tempo de retenção em torno de 9,75 min. Ao comparar estes resultados com aqueles obtidos com a análise de componentes principais aplicada aos dados cromatográficos da fração orgânica, vê-se que neste caso foi possível discriminar as amostras com base em 4 picos mais importantes, contra 3 picos na PCA. Com o modelo Tucker3 também foi possível detectar que os picos em torno de 9,75 e 10,20 min, embora fossem muito próximos, eram provenientes de compostos diferentes, este fato não poderia ser detectado somente através da PCA.

5.2.13 Análise de segunda ordem dos dados CLAE-DAD da fração básica

O modelo Tucker3 ainda foi aplicado aos dados tridimensionais da fração básica, que consistiam numa matriz tridimensional composta por 1200 valores de absorbância em função do tempo de corrida de 20 minutos (Modo A), 256 valores de absorbância em função do comprimento de onda correspondentes ao intervalo espectral de 190 a 450 nm (Modo B) e 19 amostras (Modo C).

O teste para saber a combinação do número de fatores que permitiria uma maior porcentagem de variância explicada foi feito atribuindo 10 fatores para cada uma das três dimensões. Das 700 combinações diferentes foram selecionadas apenas aquelas que apresentaram a maior porcentagem de variância explicada para cada soma total dos fatores, Tabela 27.

os dados cromatográficos da fração básica de Erythrina (FM1). N^o de fatores Variância N^o de fatores Variância Soma total Soma total nos modos explicada nos modos de fatores explicada (%) de fatores С (%) Α В Α В С 85,63 99,11 99,19 91,33 92,33 99,27 94,41 99,32 95,49 99,37 95,92 99,43 97,04 99,47 97,44 99,49 98,00 99,52 98,28 99,52

Tabela 27: Soma total de fatores, variância explicada e número de fatores nos modos A, B e C para

Para uma melhor compreensão dos dados da Tabela 27, foram plotados o número total de fatores versus a porcentagem de variância explicada, Figura 76.

99,52

99,5

99,53

98,45

98,72

98,87

98,96



Figura 76: Variância explicada para cada número total de fatores dos modelos Tucker3 para os dados cromatográficos da fração básica de Erythrina (FM1).

Na Figura 76, observa-se que as maiores diferenças em termos de porcentagem de variância explicada ocorrem quando se passa do número total 3 de fatores para o número total 4 de fatores, isto também é observado indo do 5 para o 6 e do 9 para o 10. A partir do número total 10 de fatores a diferença é pequena, cerca de 2,49% com relação à variância explicada usando um número total de 30 fatores. Porém, ao comparar os resultados obtidos, foi constatado que o número total 8 de fatores era adequado para descrever a similaridade entre os grupos e que a adição de mais um fator, tanto no modo A como no modo C (modelo (4,2,4)) não melhorava o modelo em termos das relações entre os modos. O número total 8 de fatores corresponde à combinação 3 x 2 x 3, ou seja, 3 fatores no modo A, 2 fatores no modo B e 3 no modo C. Esse modelo alcançou o critério de convergência após a terceira interação e explicou 95,49% da variância total.

Considerando o valor do core, das 18 combinações diferentes, as oito mais importantes podem ser vistas na Tabela 28. Juntas, as oito combinações explicam 99,92% da variação do core.

Ordem	Elemento do core		Core (10 ⁷)	Variância (%)	Ordem	Elemento do core			Core (10 ⁷)	Variância (%)	
1	1	1	1	-22,37	89,66	5	1	2	2	-0,84	0,12
2	2	1	2	-5,73	5,89	6	3	2	2	0,59	0,06
3	3	1	3	-4,23	3,21	7	2	2	2	0,57	0,05
4	2	2	1	-2,27	0,92	8	3	1	2	0,28	0,01

Tabela 28. Importância dos elementos do core em ordem decrescente de acordo com a variância do core para o modelo Tucker3 para os dados cromatográficos da fração básica de *Erythrina* (FM1).

A Tabela 28 mostra que as cinco primeiras interações possuem valores de core com sinal negativo. A combinação de fatores mais importante é a (1,1,1) cujo valor de core é de $-22,37_{x}10^{7}$.

A Figura 77 apresenta os três fatores do modo A.



Figura 77: Loadings dos fatores 1, 2 e 3 do modo A do modelo Tucker3 para os dados tridimensionais da fração básica de *Erythrina* (FM1).

A Figura 77 mostra que todas as variáveis do fator 1 possuem pesos com valores positivos sendo que as mais importantes possuem tempo de retenção de 2,55, 3,60, 4,80 e 5,81 min. Nos loadings dos fatores 2 e 3, observa-se que as variáveis possuem pesos com valores positivos e negativos. As variáveis mais importantes com valores positivos no fator 2 possuem tempo de retenção em 4,82 e 5,84 min e com maior peso na parte negativa está a variável em 3,69 min. No fator 3, as variáveis mais importantes com valor de peso positivo possuem tempo de retenção de 5,12 e 6,26 min e as variáveis com valores de pesos negativos possuem tempo de retenção em 4,77 e 5,76 min.

A Figura 78 apresenta os fatores 1 e 2 do modo B.



Figura 78: Loadings dos fatores 1 e 2 do modo B do modelo Tucker3 para os dados tridimensionais da fração básica de *Erythrina* (FM1).

A Figura 78 mostra que todos os loadings do fator 1 do modo B possuem valores positivos, sendo que os máximos ocorrem nos comprimentos de onda de 205, 234 e 282 nm. Os loadings do fator 2 apresentam valores positivos e negativos. A variável com maior peso na parte positiva possui comprimento de onda de 243 nm, enquanto que a variável mais importante na parte negativa possui comprimento de onda de 201 nm.

A Figura 79 mostra os fatores 1, 2 e 3 do modo C.

No fator 1 do modo C, observa-se que todos os objetos possuem valores negativos de escore, sendo que a amostra *edh* possui o valor mais negativo, em seguida estão as amostras *ed* e *ea* juntamente com as replicatas. Nos fatores 2 e 3 tem-se valores de escores positivos e negativos. No fator 2, na parte positiva, o objeto ou amostra *ed* apresenta o mais alto valor de escore seguida das amostras *e, eh, eda* e *eha*, e o valor de escore mais negativo é observado para a amostra *edh*. No fator 3, o escore mais positivo foi obtido com a amostra *ea*, juntamente com a amostras *ed*, *dh*, *da* e *eda*, e os valores mais negativos foram obtidos com as amostras *eh*, *eha* e *dha*.



Figura 79: Escores dos fatores 1, 2 e 3 do modo C do modelo Tucker3 para os dados tridimensionais da fração básica de *Erythrina* (FM1).

Das oito combinações apresentadas na Tabela 28, apenas as 4 primeiras, que explicam 99,68% da variância total, serão interpretadas.

A combinação dos fatores (1,1,1) possui um valor de core de $-22,37_{x}10^{7}$ e permite explicar 89,66% da variância do core. Esta combinação permite relacionar as substâncias com pico nos tempos de retenção de 2,55, 3,60, 4,80 e 5,81 min, com um espectro de absorção no UV com máximos em 205, 234 e 282 nm, especialmente às amostras *edh*, *ed*, *ea* e as replicatas, pois estas apresentam os valores mais negativos de escore.

Uma análise do fator 1 do modo C junto com os cromatogramas permitiu identificar 4 grupos diferentes, que podem ser discriminados com base na magnitude do valor do escore. O primeiro grupo é formado pelas amostras *edh*, *ed*, *e*a e as replicatas, que apresentaram os quatro picos indicados pelo fator 1 do modo A e possuem os valores mais negativos de escore. O segundo grupo é formado pelas amostras *eh*, *eda* e *eha*, que apresentaram apenas os picos em 2,55, 3,69 e 4,77 min. O terceiro é formado pelas amostras *e*, *d*, *dh* e *da* que apresentaram os quatro picos,

porém com uma quantidade bem inferior à quantidade extraída pelas amostras do primeiro grupo. O quarto grupo é formado pelas amostras *h*, *a*, *ha* e *dha*, que apresentaram os picos em 2,55, 3,69 e 4,77 min em baixas quantidades e também os valores menos negativos de escore. A Figura 80 apresenta os cromatogramas dessas amostras agrupadas com base nos quatro grupos discriminados, todos em mesma escala e no intervalo de tempo de corrida de 1,5 a 10 min.



Figura 80: Cromatogramas da fração básica de *Erythrina* no intervalo de tempo de corrida de 1,5 a 10 min.

A Figura 80 mostra claramente a diferença entre os grupos. Para confirmar se esses picos são da mesma substância, foram extraídos os espectros de absorbância no UV correspondentes a cada um deles. A Figura 81 mostra os espectros agrupados com relação ao tempo de retenção.

A Figura 81 permite observar uma boa semelhança entre os perfis espectrais dos picos com tempo de retenção em 2,5 e 3,6 min e também dos picos com tempo de retenção em 4,7 e 5,8 min. Isto indica que essas substâncias podem pertencer ao mesmo grupo de compostos.



Figura 81: Espectros de absorção no UV referentes aos picos com tempo de retenção em torno de 2,55, 3,69, 4,77 e 5,81 min das amostras da fração básica de *Erythrina*.

Os quatro grupos de espectros apresentaram absorção máxima em torno de 205, 234 e 282 nm, que foi compatível com os máximos indicados pelos loadings do fator 1 do modo B.

A combinação (2,1,2) apresenta o segundo maior valor de core. Como o sinal do core é negativo, pode-se combinar valores negativos, positivos e positivos e também positivos, positivos e negativos dos fatores 2, 1 e 2 dos modos A, B e C, respectivamente. A primeira possibilidade indica que o pico em 3,69 min possui maior intensidade na amostra *ed* e a segunda indica que os picos em 4,82 e 5,84 min são extraídos em maior quantidade pela amostra *edh*. As duas indicações podem ser confirmadas através dos cromatogramas dessas amostras apresentados na Figura 81.

A combinação (3,1,3) apresenta um valor de core de $-4,2_{x}10^{7}$. A combinação de valores negativos, positivos e positivos permite associar os picos em 4,77 e 5,76 min a um espectro com

máximo em 205 nm às amostras *e*, *d*, *ed*, *ea*, *dh*, *da* e *eda* e às replicatas e a combinação de valores positivos, positivos e negativos permite dizer que os picos em 5,12 e 6,26 min estão presentes apenas nas amostras *h*, *a*, *eh*, *ha*, *eha* e *dha*, a amostra *edh* está alocada junto com essas amostras, porém seu perfil cromatográfico é mais similar às outras amostras. A presença desses picos nessas amostras pode ser visualizada na Figura 82, que mostra os cromatogramas das amostras separados com base nos dois grupos, no intervalo de tempo de corrida de 4 a 7 min.



Figura 82: Cromatogramas da fração básica de *Erythrina* no intervalo de tempo de corrida de 4 a 7 min separados com base nos picos com tempo de retenção em (a) 4,77 e 5,76 min e (b) 5,12 e 6,26 min.

Os espectros de absorção dos picos em 5,12 e 6,26 min também foram extraídos e são mostrados na Figura 83.



Figura 83: Espectros de absorção no UV referentes aos picos com tempo de retenção em torno de 5,12 e 6,26 min das amostras da fração básica de *Erythrina*.

A Figura 83 mostra que os espectros dos picos em 5,12 e 6,26 min são bastante similares. Comparando esses espectros com aqueles dos picos em 2,55 e 3,69 min mostrados na Figura 81, nota-se que também existe uma boa semelhança, apesar de que a absorção acerca do comprimento de onda de 234 nm é mais definida nos espectros dos picos com tempo de retenção de 5,12 e 6,26 min.

A combinação (2,2,1) possui um valor de core de $-2,27_{x}10^{7}$. A combinação de valores negativos nos três modos permite associar o pico em 3,69 min a um espectro com máximo em 201 nm à todas as amostras, em especial à amostra *edh* que apresenta o valor mais negativo de escore no fator 1 do modo C. Já a combinação de valores positivos, positivos e negativos reafirma a presença dos picos com tempo de retenção em 4,82 e 5,84 min que possuem um espectro com máximo em torno 243 nm, porém sabe-se que o pico em 5,84 não está presente em todas as amostras. Estes picos estão presentes nas amostras que em sua grande maioria contém os solventes etanol e diclorometano na mistura, grupos I e III da Figura 80.

Os resultados do método Tucker3 aplicado aos dados tridimensionais da fração básica permitiram discriminar as amostras com relação a dois "pares" de solventes extratores; etanol e diclorometano e hexano e acetona. Isso demonstra a seletividade desses solventes na extração da fração básica. Outro fato importante foi constatar que as maiores quantidades foram alcançadas a partir de misturas, sendo que os solventes puros foram os menos efetivos em extrair maiores quantidades. As misturas contendo o par de solventes etanol e diclorometano, permitiu extrair os

picos em torno dos tempos de retenção de 4,7 e 5,8 min, enquanto que as misturas contendo o par hexano e acetona foi seletivo para a extração simultânea dos compostos com tempo de retenção em torno de 5,1 e 6,2 min.

Comparando estes resultados com aqueles obtidos através da análise de componentes principais aplicada aos cromatogramas da fração básica, vê-se que em ambos os casos, foi possível discriminar as amostras com relação ao solvente extrator e que os resultados das duas técnicas foram bem similares. A análise de componentes principais permitiu separar as 19 amostras em oito grupos diferentes, enquanto que o método Tucker3 permitiu discriminar quatro grupos, em ambos os métodos os grupos foram discriminados com base nos mesmos picos.

6. Conclusão

Os resultados apresentados nesta tese de doutorado mostraram de forma geral a importância do planejamento de misturas do tipo Centróide –Simplex com quatro componentes aplicado à obtenção dos extratos brutos de *Erythrina speciosa* e das frações em termos da informação obtida por cada um dos diferentes extratores, pois em todos os casos destacaram-se as misturas binárias, ternárias e quaternárias e não somente a extração com solventes puros.

Embora os vários tipos de planejamento sejam populares entre graduandos de diferentes áreas, seu uso ainda é restrito a algumas aplicações. A maior parte dos usuários de planejamento está dentro das indústrias, pois estas, na grande maioria, usam o planejamento com o intuito de aumentar os lucros. No meio acadêmico poucos grupos de trabalho estudam e otimizam as suas condições experimentais através desses métodos. Dessa forma, esta tese também possui o objetivo de divulgar o planejamento de misturas dentro do meio acadêmico, mesmo que a aplicação tenha sido em sistemas de extração dos compostos químicos de *Erythrina*. Com certeza existe um tipo de planejamento adequado para cada sistema a ser estudado.

Neste trabalho, os resultados da otimização dos escores obtidos através da análise de componentes principais com e sem rotação varimax dos dados cromatográficos dos extratos brutos utilizando a análise visual da superfície de resposta e através da aplicação do algoritmo de Derringer & Suich foram comparados. Os resultados obtidos foram muito similares, demonstrando que a otimização individual dos escores através da análise visual das superfícies de resposta foi compatível com a otimização simultânea das respostas, os experimentos confirmatórios comprovaram a eficiência dos dois métodos. Vale lembrar que o uso deste algoritmo não se limita à otimização de apenas três respostas, que foi o caso mostrado aqui. Este algoritmo é uma ferramenta muito eficaz e pode ser utilizado na otimização simultânea de dezenas de respostas. Os modelos ajustados aos escores obtidos com a análise de componentes principais com rotação varimax se mostraram mais simples e fáceis de interpretar.

Com relação ao estudo dos dados gravimétricos dos rendimentos dos extratos brutos e das diferentes frações foi possível observar que, os solventes extratores ótimos foram diferentes, porém todos eram misturas binárias ou ternárias. O solvente extrator com menos destaque para obtenção de maiores rendimentos foi a acetona, porém ela se mostrou eficiente na análise

qualitativa dos dados cromatográficos dos extratos e das frações. Com a análise de componentes principais dos dados gravimétricos foi possível observar que maiores rendimentos do extrato bruto, das frações básica e orgânica e dos resíduos podem ser obtidos com os extratores *ed* e *edh*. Este fato é importante porque mostra que com esta metodologia foi possível selecionar uma única mistura de solventes capaz de extrair maiores quantidades de quatro das seis variáveis estudadas.

A análise de componentes principais aplicada aos dados cromatográficos do extrato bruto e das frações básica e orgânica permitiu a discriminação das amostras com relação à informação química extraída por cada mistura de solvente extrator. Através dos grupos formados foi possível perceber as características das misturas extratoras e reconhecer o poder de interação entre os solventes. Cada grupo se destaca por ser mais específico na extração de alguns picos, enquanto outros permitem a extração de todos os picos importantes, porém em baixas quantidades, etc. Aqui é importante ressaltar que essas informações somente podem ser obtidas com o préprocessamento correto dos dados. Algumas amostras do extrato bruto e da fração orgânica tiveram seus tempos de retenção ajustados, as matrizes de dados também foram submetidas a algum tipo de pré-processamento antes da análise de componentes principais, ao contrário, para avaliar fração básica, foram utilizados os dados originais. Hoje em dia existem inúmeras opções de tratamento dos dados, porém o pesquisador deve ter em mente que o uso dessas transformações deve ser limitado, caso contrário, o pré-processamento pode acabar destruindo toda a sua estrutura e característica.

No caso do extrato bruto foi possível discriminar as amostras em sete grupos diferentes com base em quatro picos mais importantes, com tempos de retenção de 2,20, 3,65, 4,10 e 12,24 min. Já os cromatogramas da fração orgânica foram discriminados em cinco grupos definidos basicamente pelos picos em 4,73, 5,78 e 9,90 min. A fração orgânica não apresentou muita variação qualitativa com relação ao solvente extrator. Ao contrário, os cromatogramas da fração básica apresentaram diferença qualitativa com relação ao solvente extrator, e os oito grupos formados foram classificados com base nos tempos de retenção em torno de 2,53, 3,61, 4,80, 5,83 e 6,26 min.

O método Tucker3 aplicado aos dados CLAE-DAD se mostrou superior à análise de componentes principais aplicada aos cromatogramas. Os dois métodos não são equivalentes, mas sim complementares. Por utilizar uma informação química muito maior, o método Tucker3 tornase menos específico, isso é constatado pelo número de grupos formados através dos dois métodos.

O método Tucker3 permitiu enxergar além dos picos dos cromatogramas, reconhecendo se os espectros de absorção no UV referentes aos picos com o mesmo tempo de retenção eram ou não da mesma substância. Neste caso, também não foi necessário ajustar nenhum tempo de retenção ou aplicar algum tipo de pré-processamento, pois o método permitiu saber se o deslocamento foi causado pelas variações das condições experimentais, ou se o deslocamento tratava-se na verdade, de uma substância diferente.

O modelo Tucker3 de complexidade (3,3,3) aplicado aos dados cromatográficos dos 19 extratos brutos mostrou que os cromatogramas puderam ser discriminados com relação aos solventes extratores etanol, diclorometano e acetona com base nos picos com tempo de retenção em 2,73, 3,69, 4,07 e 12, 19 min e seus respectivos espectros de absorção. Para a fração orgânica o melhor modelo Tucker3 foi o de complexidade (4,1,4), que discriminou as amostras quantitativamente com base nos picos com tempo de retenção em 2,72, 4,80, 5,80 e 10,20 min e seus espectros de absorção no UV. Para a fração básica, o modelo de complexidade (3,2,3) permitiu discriminar as amostras em quatro grupos com base no picos com tempo de retenção em 2,55, 3,60, 4,80, 5,81 e 6,26 min e seus espectros de absorção no UV.

Com essa parte dos resultados, pôde-se demonstrar a importância da análise exploratória dos dados cromatográficos da *Erythrina*, pois de outra forma não seria possível agrupar as amostras com relação à importância das variáveis medidas.

Finalmente, com este trabalho, esperamos que a comunidade científica como um todo, possa se beneficiar tanto das metodologias quimiométricas aplicadas na parte de análise de dados como das informações químicas e gravimétricas obtidas a respeito da *Erythrina*.

7. Referências

1- Sahoo, N., Manchikanti, P., Dey, S., Fitoterapia, v. 81, p. 462-471, 2010.

2- Costa, P. R. R., Rev. Virtual Quim., v. 1, p. 58-66, 2009.

3- Braz Filho, R., *Quim. Nova*, v.33, n. 1, p. 229-239, 2010.

4- Mukhtar, M., Arshad, M., Ahmad, M., Pomerantz, R. J., Wigdahl, B., Parveen, Z., *Virus Res.*, v. 131, p. 111–120, 2008.

5- Pan, L., Chai, H., Kinghom, D. A., *Phytochem. Letters*, v. 3, p. 1–8, 2010.

6- Drasar, P., Moravcova, J., J. Chromatogr. B, v.812, p.3-21, 2004.

7- Heyden, Y. V., Xu, C. J., Liang, Y. Z., Chau, F. T., J. Chromatogr. A, v.1134, p.253-259, 2006.

8- Heyden, Y. V., Alaerts, G., Matthijs, N., Smeyers-Verbeke, J., J. Chromatogr. A, v.1172, p.1-8, 2007.

9- Qin, X. M., Dai, Y.T., Zhang, L. Z., Guo, X. Q., Shao, H. X., Phytochem. Anal., v.20, p.307-313, 2009.

10- Ciesla, L., Bogucka-Kocka, A., Hajnos, M., Petruczynik, A., *J. Chromatogr. A*, v. 1207, p. 160-168, 2008.

11- Chen, C., Zhang, H., Xiao, W., Zheng-Ping, Y., Bai, N, J. Chromatogr. A, v. 1154, p. 250-259, 2007.

12- Heyden, Y. V., Ji, Y. B., Xu, Q. S., Hu, Y. Z., *J. Chromatogr. A*, v.1066, p.97–104; 2005.

13-Jiang, Y., David, B., Tu, P., Barbin, Y., Anal. Chim. Acta, v. 657, p. 9-18, 2010.

14- Tang, F., et al., *TrAC*, v. 28, n. 11, p. 1253-1262, 2009.

15- Lin, J. T., Liu, S. C., Tsay, G. J., Yang, D. J., Food Chem., v. 121, p. 659–665, 2010.

16- Lim, M. P., et al., *Phytochem.*, v. 66, p. 1560–1566, 2005.

17- Um, Y. R, et al., Food Chem., v. 120, p. 385–394, 2010.

18- Zhang, J., et al., J. Pharm. Biomed. Anal., v. 52, p. 446–451, 2010.

19- Zhang, H., et al., J. Sep. Sci., v. 32, p. 526 – 535, 2009.

20- Bruns, R. E., Scarminio, I. S. e Neto, B. B.. **Statistical design – Chemometrics**; Editora Elsevier, Amsterdam, 2006.

21- Dumarey, M., Smets, I., Heyden, Y. V., J. Chromatogr. B, article in press, 2010.

22-Wei, H., Sun, L., Tai, Z., Gao, S., Xu, W., Chen, W., Anal. Chim. Acta, v. 662, p. 97-104, 2010.

23- Wu, Y. J., Zheng, Y. L., Luan, L. J., Liu, X. S., Han, Z., Ren, Y. P., Gan, L. S., Zhou, C. X., *J. Sep. Sci.*, v. 33, p. 2734-2742, 2010.

24- Li, Y., et al., J. Pharm. Biomed. Anal., v. 52, p. 597–602, 2010.

25- Xie, P. S., et al., J. Pharm. Biomed. Anal., v. 52, p. 452–460, 2010.

26- Brereton, R. G.. **Chemometrics data analysis for the laboratory and chemical plant**, Editora John Wiley & Sons Ltd, New York, 2003.

27- Skov, T., **Mathematical resolution of complex chromatographic measuments**, Tese de doutorado, Universidade de Copenhagen, 2008.

28- Jin, W., Ge, R. L., Wei, Q. J., Bao, T. Y., Shi, H. M., Tu, P. F., *J. Chromatogr. A*, v.1132, p.320-324, 2006.

29- Mok, D. K. W., Chau, F.T., Chemom. Intell. Lab. Syst., v.82, p. 210-217, 2006.

30- Faria, T. J., et al., *Quim. Nova*, v.30, n. 3, p. 525-527, 2007.

31-Carvalho, A. C., C. S., Almeida, D. S., Melo, M. G. D., Cavalcanti, S. C. H., Marçal, R. M., *J. Ethnopharmacol.*, v. 122, p. 374-378, 2009.

32- Juma, B. J., Majinda, R. T., *Phytochem.*, v. 65, p. 1397-1404, 2004.

33- Chavez, R. S. M., Hernandez, M. S., Valdivia, A. C. R., Kite, G., *Phytochem. Rev.*, v. 6, p. 167-173, 2007.

34- Marcus, Y.. **Solvent mixtures. Properties and selective solvation**; Editora Marcel Dekker, Inc., New York, 2002.

35- Snyder, L. R., Kirkland, J. J., Glajch, J. L.; **Practical HPLC method development;** 2 ed.; Editora John Wiley & Sons Inc, New York, 1997.

36-Coenegeracht, P. M. J., Metting, H. J., Smilde, A. K., Coenegeracht-Lamers, P. J. M., *Chromatog.*, v.27, p.135-141, 1989.

37- Lough, W. J., Wainr, I. W.. **High performance liquid chromatography**; Editora Blackie Academic & Professional, London, 1995.

38- Snyder, L. R., Kirkland, J. J.. Introduction to modern liquid chromatography; 2 ed.; Editora John Wiley & Sons Inc, New York, 1979.

39- Reyment, R, Jöreskog, K. G.. **Applied factor analysis in the natural sciences**; Editora Cambridge University Press, Cambridge, UK, 1996.

40- Massart, D., L., et al.. Chemometrics: a textbook; Editora Elsevier, Amsterdam, 1988.

41- Kiers, H. A. L., Mechelen, I. I., Psychol. Methods, v. 6, n. 1, p. 84-110, 2001.

42- Cornell, J. A.: Experiments with mixtures: Designs, models, and the analysis of mixture data; 3

ed., Editora John Wiley & Sons, Inc. New York, 2002.

43- Myers, R. H., Montgomery, D. C.. Response surface methodology: Process and product optimization using designed experiments; 2 ed., Editora John Wiley & Sons, Inc. New York, 2002.
44- Derringer, S. e Suich, R., J. Quality Tech., v. 12, n. 4, p. 214-219.

45- Eren, I., Ertekin, F. K., *J. Food Eng.*, v. 79, p. 344-352, 2007.

46- Massart, D. L., Detroyer, A. e Heyden, Y. V., J. Chromatgr. A, v. 986, p. 227-238, 2003.

47- Soares, P. K.. Taxonomia por abordagem metabolômica e métodos quimiométricos para análise e rastreio de plantas do gênero *Bauhinia*. Dissertação de Mestrado, Universidade Estadual de Londrina, Londrina, 2006.

48- Beebe, K. R., Pell, R. J. e Seasholtz, M. B.. **Chemometrics: a practical guide**; Editora John Wiley & Sons Ltd, New York, 1998.

49- Pielou, E. C.. **The interpretation of ecological data: a primer on classification and ordination**; Editora John Wiley & Sons Ltd, New York, 1984.

50- Scarminio, I. S., Collins, C. H., et al., Quim. Nova, v.24, n. 3, p. 331-338, 2001.

51- Sharaf, M. A., Illman, D. L., Kowalski, B. R.. **Chemometrics;** Editora John Wiley & Sons Ltd, New York, 1986.

52- Brereton, R. G.. **Applied chemometrics for scientists**; Editora John Wiley & Sons Ltd, New York, 2007.

53- Hair, J. F., et al.. **Análise multivariada de dados**. Trad. Adonai S. Sant'Anna e Anselmo C. Neto. 5 ed.; Editora Bookman, Porto Alegre, 2005.

54- Ward Jr, J. H., J. Amer. Stat. Assoc., v. 58, p. 236-244, 1963.

55- Smilde, A., Bro, R., Geladi, P.. **Multi-way analysis. Applications in the chemical sciences;** Editora John Wiley & Sons Ltd, New York, 2004.

56- Kroonenberg, P. M.. **Applied multiway analysis**; Editora John Wiley & Sons Ltd, New York, 2008.

57- Oros, G., Cserháti, T., Chemom. Intell. Lab. Syst., v. 96, p. 1-5, 2009.

58- Bro, R., et al., Chemom. Intell. Lab. Syst., v. 76, p. 79-89, 2005.

59- Pravdova, V., Boucon, C., Jong, S., Walczak, B., Massart, D. L., *Anal. Chim. Acta*, v. 46, p. 133-148, 2002.

60- Walczak, B., et al., Chemom. Intell. Lab. Syst., v. 96, p. 203-209, 2009.

61- Sena, M. M., Trevisan, M. G., Poppi, R. J., Quim. Nova, v.28, n. 5, p. 910-920, 2005.

62- Kroonenberg, P. M., Harshman, R. A., Murakami, T., Applied Multivar. Res., v. 13, p. 5-41, 2009.

63- Lonni, A. A. S. G.. Caracterização e identificação química de carquejas do gênero Baccharis por

métodos quimiométricos, Dissertação de Mestrado, Universidade Estadual de Londrina, Londrina, 2002.

64- Camarão, E. B., et al., Ultrason. Sonochem., v. 13, p. 242-250, 2006.

65- Di Stasi, L.C.. **Plantas medicinais: Arte e Ciência. Um guia de estudo interdisciplinar**. São Paulo: UNESP, 1996.

- 66- D'oca, M. G. M., et al., Encontro de Química da Região Sul, v. 16, p. QO 116, 2008.
- 67- Buckeridge, M. S., et al., *Ciênc. Tecnol. Aliment.*, v. 25, p. 382-386, 2005.
- 68- http://www.models.kvl.dk/source, acessado em 15 de março de 2010.
- 69- Statistica for windows, Statsoft, Inc., Tulsa, OK, USA, 1999.
- 70- Matlab 6.5, The Mathworks Inc., Natick, MA, 2000.
- **71-** Pirouette 3.11, Infometrix, Inc., Woodinville, WA, 2003.

8. Anexos



Reproduzido com a permissão de Chemometrics and Intelligent Laboratory Systems 2010, 103, 1-7 © Copyright 2010 Elsevier Contents lists available at ScienceDirect



Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab



Statistical mixture design – Principal component determination of synergic solvent interactions for natural product extractions

Livia Maria Zambrozi Garcia^b, Talita Fogaça de Oliveira^b, Patricia Kaori Soares^a, Roy Edward Bruns^a, Ieda Spacino Scarminio^{b,*}

^a Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas, SP, Brazil

^b Laboratório de Quimiometria em Ciências Naturais, Departamento de Química, Universidade Estadual de Londrina, CP 6001, 86051-990 Londrina, PR, Brazil

ARTICLE INFO

Article history: Received 8 December 2009 Received in revised form 24 March 2010 Accepted 23 April 2010 Available online 10 May 2010

Keywords: Mixture design extractions Mikania laevigata Sch. Bip. Leaves Binary solvent interactions Principal components Simplex centroid design

ABSTRACT

Ethanol, ethyl acetate, dichloromethane, acetone and chloroform mixtures of a simplex centroid design have been used to extract total crude material as well as the neutral, organic, basic, fiber and residue fractions of *Mikania laevigata Sch. Bip.* leaves. The quantity of crude extracted material is found to depend on the solvent proportions according to a special cubic model. Synergic binary solvent effects between the ethanol–ethyl acetate, ethanol–dichloromethane, ethanol–chloroform, ethyl acetate–chloroform and ethyl acetate– dichloromethane pairs are found to explain the superior extraction results of the mixtures compared to results for the pure solvents. These significant synergic binary effects are seen to occur between solvents having disparate acidity, basicity and/or dipolarity parameters located at extremities of the domain defined by the mixtures within the Snyder-Rohrschneider solvent selectivity triangle.

So molecular interactions known to be important for chromatographic separations are also significant for separating crude material from its *Mikania laevigata Sch. Bip.* leave matrix. Principal component analysis shows these interaction effects to be important for obtaining higher yields of the fractionated materials. © 2010 Elsevier B.V. All rights reserved.

1. Introduction

Plants produce a great variety of organic compounds that are not directly involved in primary metabolic processes of growth and development. The roles played by these natural products or secondary metabolites in plants have only recently come to be appreciated in an analytical context. Natural products appear to function primarily in defense against predators and pathogens and in providing reproductive advantage as attractants of pollinators and seed dispersers [1,2]. Most natural products can be classified into three major groups: terpenoids, alkaloids and phenolic compounds (mostly phenylpropanoids). These chemical compounds have large economical importance. First of all, they are connected with important traits of the plant itself, e.g., color or fragrance of flowers, taste and color of food, resistance against pests and diseases, and also for the production of fine chemicals such as drugs, antioxidants, flavors, fragrances, dyes, insecticides and pheromones [3,4]. The effects of secondary metabolites on plants and their environments are currently being intensively researched by scientists. While valuable to humans, they are often very difficult to isolate in large quantities, and, as a result, some of the products of secondary metabolism can be very expensive to acquire, because they are available in such low amounts.

Mikania laevigata Schultz Bip. ex Baker belongs to the Asteraceae family and is a subarbustive plant native to South America [5]. In Brazil *Mikania laevigata* is popularly known as guaco and used to treat many complaints including asthma, bronchitis, chronic lung diseases, coughs and rheumatism [5,6]. Coumarin is the major chemical substance found in *Mikania laevigata*. There is a relationship between coumarin and the pharmacological action of this plant. For this reason it has been used as a chemical marker in pharmaceutical presentations [7–9]. Also anti-inflammatory and antioedematogenic pharmacological actions have been attributed to the *Mikania laevigata* [10].

In many extraction systems the solvents are employed in substantially pure form, as so called neat solvents. However, in certain extractions the neat solvents fail. In these cases it is expedient to use solvent mixtures, which may range from binary mixtures involving two solvents to ternary (three solvents) or even higher multicomponent mixtures [11]. An important aspect in the studies performed on mixed solvents is the role of solvent–solvent interactions in competition for the solvation of solutes [12–14].

In the literature, different solvent systems have been used for the extraction of secondary metabolites from plant materials, because their extraction efficacy depends on their chemical nature. For example, the most widely used solvents for extracting phenolic substances are methanol, acetone and their water mixtures [15]. Absolute methanol, absolute acetone, 90, 80 and 70% methanol or 90, 80 and 70% acetone were used to evaluate the effects of various solvent extraction systems on the recovery of beach pea tannins [16].

^{*} Corresponding author. Tel.: +55 43 33714366; fax: +55 43 33714286. *E-mail address:* ieda@qui.uel.br (I.S. Scarminio).

^{0169-7439/\$ –} see front matter 0 2010 Elsevier B.V. All rights reserved. doi:10.1016/j.chemolab.2010.04.016

Several extraction techniques have been reported for the extraction of phenolic compounds from different matrices using solvents with different polarities, such as methanol, water, ethyl acetate and petroleum ether [17,18]. Ethyl acetate, methanol, and water, were used to extract phenolic compounds from the pistachio hull [19].

Although the literature shows many studies on aqueous mixtures of pure organic solvents, relatively few investigations of these solvent systems have been carried out using statistical mixture designs. In previous work our group has shown that statistical mixture designs are useful tools for the study of solvent extraction behavior applied to *Erythrina speciosa* Andrews leaves. HPLC results for the extracted material of samples by solvent mixtures of an ethanol, dichloromethane, hexane and acetone simplex centroid design were investigated [20,21]. Later statistical mixture models and principal component analysis were carried out on quantities of fractionated and total extract of this plant for this same mixture design [22].

In this work a simplex centroid mixture design involving ethanol, ethyl acetate, dichloromethane, acetone and chloroform has been applied to the extraction of total and fractionated material from *Mikania laevigata Sch. Bip.* leaves. Interest is focused on the assessment of binary synergic interactions. For this five component system a total of ten binary interactions are possible. Some solvent pairs such as dichloromethane–chloroform and acetone–ethyl acetate have similar separation properties whereas others such as ethanol with any of the other four solvents have quite different separation capabilities. Statistical mixture models and principal component analysis are carried out to determine which of these interactions are significant for extracting total and individual fractionated materials. The extraction results are then interpreted using the acidity, basicity and dipolarity solvent selectivity parameters.

2. Experimental

For this study *Mikania laevigata Sch. Bip.* leaves were used. A voucher specimen of this plant was deposited in the herbarium of the Londrina State University and authenticated by A. O. Vieira and L. R. Bargoena and registered under the FUEL 44355 number. The collection was performed in October 2007. Leaves were selected avoiding material damaged by insects and/or fungus. Drying was carried out at room temperature for twelve days. During this period, the leaves were protected from light, humidity, fungus and insect attacks.

Extraction media were prepared using mixtures of five components, (1) ethanol, (2) ethyl acetate, (3) dichloromethane, (4) acetone and (5) chloroform whose proportions were varied according to a simplex centroid design. The proportions of each solvent used in the extraction mixtures are specified in Table 1. Thirty three extractions were carried out with 31 different mixtures. A three-run replicate was performed at the center point so experimental error could be estimated. Although a larger number of replicate determinations would increase the precision of the 95% confidence intervals of the model coefficient values this was not possible owing to operational difficulties and our constraint on limiting the total amount of solvent to be discarded.

Each extract was prepared by weighing 20 g exactly of dried and crushed leaves and adding 90 mL of the solvent mixtures in Table 1. These mixtures were placed in an ultrasonic bath (Unique, model USC 1400) for 30 min. The bathwater was changed every 10 min to maintain constant water temperature. The extracts were filtered through filter paper to separate the solution from small pieces of leaves and the solution was placed in an identified and weighted flask.

Table 1

Extractor mixture composition in mL and yield in g of the crude extracts, fibers, neutral, organic and basic fractions and residues.

$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	Extract	Solvents					Masses					
e A d a c e 90 0 0 0 0.0664 0.2620 0.5858 0.0068 0.0165 0.2382 A 0 90 0 0 0.6664 0.2611 0.2889 0.0090 0.02332 a 0 0 990 0 0.11128 0.8461 0.1644 0.0077 0.0168 0.0778 a 0 0 0 990 0.1615 0.8701 0.2345 0.01077 0.0168 0.0778 c 0 0 0 0 0 0.4306 0.6928 0.0045 0.0231 0.1169 ed 45 0 0 1.2679 0.4335 0.5666 0.0024 0.0231 0.1764 ec 45 0 0 1.2014 0.7868 0.3546 0.0038 0.0245 0.0317 Ac 0 45 45 0 0.2277 0.4455 0.4149 <td></td> <td>Ethanol</td> <td>E. A.</td> <td>Dichl.</td> <td>Acetone</td> <td>Chlorof.</td> <td>Crude</td> <td>Neutral</td> <td>Organic</td> <td>Basic</td> <td>Fiber</td> <td>Residue</td>		Ethanol	E. A.	Dichl.	Acetone	Chlorof.	Crude	Neutral	Organic	Basic	Fiber	Residue
e9000001.10330.26200.58580.00680.01650.2382A0900000.66460.29110.28890.00900.002440.0062a0090001.11280.84610.16440.00770.01680.0078a0090000.11280.84610.16440.00770.01840.0077eA45450001.26790.83150.56270.00640.00450.02310.11189ed45001.26790.63080.56660.00260.002310.1179ed45001.20140.76880.55870.00200.02940.2231dat045001.20140.76880.35460.00380.02440.2311dat045001.20140.76880.35460.00380.02440.2311dat00451.20140.76880.35460.00380.01220.0711dat00451.20140.76880.35460.00380.01220.0711dat00451.20280.32710.37660.56440.00710.02240.0749eac00451.14580.66230.35460.00320.04420.1059eac0030<		е	A	d	a	С						
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	е	90	0	0	0	0	1.1093	0.2620	0.5858	0.0068	0.0165	0.2382
	Α	0	90	0	0	0	0.6646	0.2911	0.2989	0.0090	0.0294	0.0362
a009000.913150.56270.00640.04560.0103c000901.16150.87010.23450.01070.03470.0115eA45450001.26790.43060.69280.00360.02310.1169ea45001.21120.43550.56660.00240.01230.1764ec45001.21120.45550.56660.00240.02430.0176da04545001.20140.78680.33660.00380.02450.0317Aa045009.2570.44550.41490.01020.03890.0162Ac04501.27750.66660.56460.00780.05430.0442ac0045451.20280.32730.37680.00910.05840.4421ac004501.14580.65270.26810.00710.0220.0713ed003001.14580.65230.31830.00610.02310.1137eac3000301.25560.62340.54970.00650.03210.1137eac303001.32580.62340.54970.00650.03210.1137eac303001.32590.55650.33860.0071	d	0	0	90	0	0	1.1128	0.8461	0.1644	0.0077	0.0168	0.0778
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	а	0	0	0	90	0	0.9193	0.3155	0.5627	0.0064	0.0456	-0.0109
eA450001.2690.43060.69280.00450.02310.1169ea45001.26570.63080.50630.00360.05210.4729ea45004501.21120.43350.56660.00240.022310.1764ea45000451.52760.71440.58870.00200.02940.2231Aa04545001.20140.78680.35460.00380.01220.03890.0162Ac0454500.92570.44550.41490.01020.03890.0162Ac0454501.27750.60660.56460.00780.05840.0431ac0045451.20280.32730.37680.00910.02890.1111dc0045451.20280.32730.37680.00910.02890.1137eAd3030001.14530.56520.31380.00610.02290.1137eac3000301.51930.94600.37860.00230.04920.1052eAc30300301.22040.73610.56500.00320.04920.1532eAc303030301.22040.73630.38660.00710.03650.03540.00120.1325<	С	0	0	0	0	90	1.1615	0.8701	0.2345	0.0107	0.0347	0.0115
ed45045001.6570.6380.5030.00360.00260.01210.4723ea4500451.52760.71440.55870.00200.02940.2231Ad04545002.01740.78680.35460.00380.02450.0017Aa04504500.92570.44450.01020.03890.0162Ac0450451.20750.60660.56460.00780.05840.4312ac0045451.20280.32750.28610.00710.0220.0741da0045451.20280.32750.28610.00710.0220.0743da0045451.20280.32750.28610.00710.0220.0743eAd303030301.32560.62340.54970.00650.03210.1137eac30300301.51930.94600.37860.00320.01420.1742eAa303030301.2520.89280.18320.00420.01520.1325dac03030301.2520.89280.18320.00110.0465-0.0111Ada03030301.2520.89280.13320.01460.01720.1065eda30	eA	45	45	0	0	0	1.2679	0.4306	0.6928	0.0045	0.0231	0.1169
ea45004501.2120.43550.56660.00240.01230.1764ec4500451.20160.71440.55870.00200.02940.2231Aa04545001.20160.78680.33460.00280.02250.02450.0131Aa0450001.20160.78680.33460.00280.01220.02950.1711Aa045000.56340.32450.01020.03890.01612Ac00454501.27750.60660.56460.00780.05430.0442ac00454501.27750.28610.00710.02280.0142ac004501.14530.68280.31380.00610.02890.1137eAc3030300301.45340.68280.31380.00610.02890.1137eAc30300301.45430.73610.38650.00320.04920.1059eAc30300301.45430.86280.31380.00610.02890.1137eAc30300301.45430.73610.33660.00710.3050.3060.00710.3050.3060.00710.3050.3060.00120.1132eAc3030	ed	45	0	45	0	0	1.6657	0.6308	0.5063	0.0036	0.0521	0.4729
ec 45 0 0 45 1,5276 0,7144 0,5587 0,0020 0,0294 0,2231 Ad 0 45 0 0 1,5276 0,7144 0,5587 0,0020 0,0294 0,0234 0,0317 Aa 0 45 0 0,9257 0,4455 0,4149 0,0102 0,0388 0,0162 Ac 0 45 0 1,2775 0,6066 0,5646 0,0078 0,0543 0,0442 ac 0 0 45 45 1,2028 0,3237 0,3768 0,0071 0,0228 0,0714 eac 30 30 30 0 45 1,1458 0,7557 0,2861 0,0071 0,022 0,0749 eac 30 30 30 30 1,313 0,6828 0,3138 0,0061 0,0289 0,1137 eac 30 30 0 30 1,3256 0,6234 0,5447 0,0065 0,0321 0,1139 eac 30 30 0 30 1,215	еа	45	0	0	45	0	1.2112	0.4535	0.5666	0.0024	0.0123	0.1764
Ad 0 45 45 0 0 1.2014 0.7868 0.3546 0.0038 0.0245 0.0317 Aa 0 45 0 45 0 0.9257 0.4455 0.149 0.0102 0.0389 0.0162 Ac 0 45 0 0 45 0.0666 0.5634 0.3208 0.0112 0.0295 0.0112 ac 0 0 45 45 1.2075 0.6066 0.5646 0.0071 0.0225 0.0121 eac 0 0 45 1.458 0.757 0.2861 0.0071 0.022 0.0741 eac 30 30 0 0 1.1453 0.6828 0.3138 0.0061 0.0229 0.1137 eac 30 0 0 30 1.255 0.826 0.3786 0.0032 0.0492 0.1559 edc 30 0 30 0 1.2204 0.7361 0.5650 0.0032 0.0492 0.1559 eda 30 0 30	ес	45	0	0	0	45	1.5276	0.7144	0.5587	0.0020	0.0294	0.2231
Aa 0 45 0 45 0 0.9257 0.4455 0.4149 0.0102 0.0389 0.0162 Ac 0 45 0 0.2775 0.6664 0.3208 0.0112 0.0295 0.1711 ac 0 0 45 45 1.2028 0.3273 0.3768 0.0091 0.0584 0.4422 ac 0 0 45 45 1.2028 0.3273 0.3768 0.0091 0.0584 0.4312 dc 0 0 45 0 45 1.2028 0.3273 0.3768 0.0091 0.0584 0.4312 dc 0 0 45 0 45 0.228 0.3757 0.2861 0.0071 0.022 0.0749 eAd 30 0 0 30 1.3256 0.6234 0.5497 0.0065 0.0321 0.1139 eAc 30 0 30 1.3256 0.6234 0.5497 0.0065 0.0321 0.1139 eAc 0 30 0 30 <t< td=""><td>Ad</td><td>0</td><td>45</td><td>45</td><td>0</td><td>0</td><td>1.2014</td><td>0.7868</td><td>0.3546</td><td>0.0038</td><td>0.0245</td><td>0.0317</td></t<>	Ad	0	45	45	0	0	1.2014	0.7868	0.3546	0.0038	0.0245	0.0317
Ac 0 45 0 0 45 1.0960 0.5634 0.3208 0.0112 0.0295 0.1711 da 0 0 45 45 0 1.2775 0.6066 0.5646 0.0078 0.0543 0.0442 ac 0 0 45 45 1.2028 0.3273 0.3768 0.0091 0.0548 0.4312 eAd 30 30 30 0 0 1.453 0.6828 0.3138 0.0061 0.0229 0.01139 eAc 30 0 0 30 1.4594 0.7361 0.5650 0.0032 0.0492 0.1593 eAc 30 0 0 30 1.4594 0.7361 0.5650 0.0032 0.0492 0.1593 eAc 30 0 30 0 1.252 0.8928 0.1763 0.0011 0.0425 0.1325 eAc 0 30 0 30 1.2524 0.8928 0.1763 0.0110 0.0485 -0.0111 Aac 0 30	Aa	0	45	0	45	0	0.9257	0.4455	0.4149	0.0102	0.0389	0.0162
da 0 45 45 0 1.2775 0.6066 0.5466 0.0078 0.0543 0.0443 ac 0 0 45 45 1.2028 0.3273 0.3768 0.0091 0.0584 0.4312 ac 0 0 45 0.1458 0.7557 0.2861 0.0071 0.022 0.0749 eAd 30 30 0 0 30 30 1.1453 0.6828 0.3138 0.0061 0.0289 0.1137 eac 30 0 0 30 1.4594 0.7361 0.5650 0.0032 0.0492 0.1059 edc 30 0 30 0.30 1.5193 0.9460 0.3786 0.0023 0.0182 0.1742 eAa 30 30 0 30 30 1.2252 0.8288 0.1832 0.0042 0.0125 0.1325 dac 0 30 30 0 1.2252 0.8288 0.1763 0.0110 0.0485 -0.0111 Aac 30 0 30 </td <td>Ac</td> <td>0</td> <td>45</td> <td>0</td> <td>0</td> <td>45</td> <td>1.0960</td> <td>0.5634</td> <td>0.3208</td> <td>0.0112</td> <td>0.0295</td> <td>0.1711</td>	Ac	0	45	0	0	45	1.0960	0.5634	0.3208	0.0112	0.0295	0.1711
ac 0 0 45 45 1.2028 0.3273 0.3768 0.0091 0.0584 0.4312 dc 0 0 45 0 45 1.1458 0.7557 0.2861 0.0071 0.022 0.0749 eAd 30 30 30 0 0 1.1453 0.6828 0.3138 0.0061 0.0229 0.1137 eac 30 0 0 30 30 1.3256 0.6234 0.5497 0.0065 0.0321 0.1139 eAc 30 0 30 0 30 1.5193 0.9460 0.5650 0.0323 0.0182 0.1742 eAa 30 30 0 30 0 30 1.2204 0.7396 0.3806 0.0071 0.0365 0.0556 dac 0 30 30 30 30 1.2204 0.7396 0.3806 0.0071 0.0365 0.0566 dac 0 30 30 30 1.2140 0.5405 0.5414 0.0048 0.0172 0.0165	da	0	0	45	45	0	1.2775	0.6066	0.5646	0.0078	0.0543	0.0442
dc 0 45 0 45 1.1458 0.7557 0.2861 0.0071 0.022 0.0749 eAd 30 30 30 0 0 1.1453 0.6828 0.3138 0.0061 0.0289 0.1137 eac 30 0 0 30 30 1.3256 0.6234 0.5497 0.0065 0.0321 0.1139 eac 30 0 0 30 1.4594 0.7361 0.5650 0.0032 0.0492 0.1059 edc 30 0 30 0 30 1.5193 0.9460 0.3786 0.0023 0.0182 0.1742 eda 30 30 0 30 30 1.2204 0.7396 0.3806 0.0071 0.0485 -0.0111 Aac 0 30 30 30 1.0769 0.5582 0.4703 0.0146 0.0176 -0.0085 Ada 0 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eAda 2	ас	0	0	0	45	45	1.2028	0.3273	0.3768	0.0091	0.0584	0.4312
eAd 30 30 30 0 0 1.1453 0.6828 0.3138 0.0061 0.0289 0.1137 eac 30 0 0 30 30 13256 0.6234 0.5497 0.0065 0.0321 0.1137 eac 30 0 30 0 30 1.4594 0.7361 0.5650 0.0032 0.0492 0.1059 edc 30 0 30 0 30 1.5193 0.9460 0.3786 0.0023 0.0125 0.1325 dac 0 0 30 0 30 1.2252 0.8928 0.1832 0.0042 0.0125 0.1325 dac 0 0 30 30 30 1.2244 0.7396 0.3806 0.0071 0.0365 0.0565 dac 0 30 30 30 30 1.0769 0.5582 0.4703 0.0110 0.0485 -0.0111 Adac 0 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1085 <td>dc</td> <td>0</td> <td>0</td> <td>45</td> <td>0</td> <td>45</td> <td>1.1458</td> <td>0.7557</td> <td>0.2861</td> <td>0.0071</td> <td>0.022</td> <td>0.0749</td>	dc	0	0	45	0	45	1.1458	0.7557	0.2861	0.0071	0.022	0.0749
eac 30 0 30 30 1.3256 0.6234 0.5497 0.0065 0.0321 0.1139 eAc 30 30 0 0 30 1.4594 0.7361 0.5650 0.0032 0.0492 0.1139 eAc 30 0 30 0 30 0 30 0.9460 0.3786 0.0021 0.01492 0.1742 eAa 30 30 0 30 0 1.5193 0.9460 0.3786 0.0021 0.0112 0.1325 dac 0 30 30 0 30 30 1.2204 0.7396 0.3806 0.0071 0.0365 0.0566 Aac 0 30 30 30 30 1.2204 0.7396 0.3806 0.0071 0.0365 0.0566 Aac 0 30 30 30 30 1.1898 0.9898 0.1763 0.0146 0.0176 -0.0185 Ada 0 30 0 1.2140 0.5405 0.5414 0.0084 0.0176 0.0274 <td>eAd</td> <td>30</td> <td>30</td> <td>30</td> <td>0</td> <td>0</td> <td>1.1453</td> <td>0.6828</td> <td>0.3138</td> <td>0.0061</td> <td>0.0289</td> <td>0.1137</td>	eAd	30	30	30	0	0	1.1453	0.6828	0.3138	0.0061	0.0289	0.1137
eAc 30 30 0 30 1.4594 0.7361 0.5650 0.0032 0.0492 0.1059 edc 30 0 30 0 30 1.5193 0.9460 0.3786 0.0023 0.0182 0.1742 eAa 30 30 0 30 0 1.2252 0.8928 0.1832 0.0042 0.0155 0.1325 dac 0 30 30 30 30 1.2204 0.7366 0.3806 0.0071 0.0365 0.0568 Aac 0 30 0 30 30 1.2204 0.7396 0.3806 0.0071 0.0365 0.0568 Adac 0 30 0 30 30 1.769 0.5582 0.4703 0.0116 0.0176 -0.0085 Ada 0 30 30 0 1.1898 0.9898 0.1763 0.0146 0.0172 0.1065 eda 30 0 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1087 ed	eac	30	0	0	30	30	1.3256	0.6234	0.5497	0.0065	0.0321	0.1139
edc 30 0 30 1.5193 0.9460 0.3786 0.0023 0.0182 0.1742 eAa 30 30 0 30 0 30 0 1.2252 0.8928 0.1832 0.0042 0.0125 0.1325 dac 0 0 30 30 30 1.2252 0.8928 0.1832 0.0042 0.0125 0.1325 dac 0 0 30 30 30 1.2254 0.7396 0.3806 0.0071 0.0365 0.0561 Aac 0 30 0 30 30 1.0769 0.5582 0.4703 0.0110 0.0485 -0.0111 Adc 0 30 30 0 1.1898 0.8989 0.1763 0.0116 0.0169 0.0828 Ada 0 30 30 0 1.3059 0.7556 0.3846 0.0071 0.0307 0.1279 eAda 22.5 22.5 22.5 0 1.2089 0.7604 0.3394 0.0094 0.0169 0.0828	eAc	30	30	0	0	30	1.4594	0.7361	0.5650	0.0032	0.0492	0.1059
eAa 30 30 0 30 0 1.2252 0.8928 0.1832 0.0042 0.0125 0.1325 dac 0 0 30 30 30 1.2204 0.7396 0.3806 0.0071 0.0365 0.0566 Aac 0 30 0 30 30 1.2204 0.7396 0.4703 0.0110 0.0485 -0.0111 Adc 0 30 30 0 30 1.898 0.9898 0.1763 0.0146 0.0172 0.1065 eda 0 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eda 30 0 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eda 22.5 22.5 22.5 0 1.2140 0.5405 0.3394 0.0094 0.0169 0.828 Adac 0 22.5 22.5 1.5251 0.9022 0.3566 0.0036 0.0161 0.1214 0.1229 0.4093 0.00	edc	30	0	30	0	30	1.5193	0.9460	0.3786	0.0023	0.0182	0.1742
dac 0 0 30 30 30 1.2204 0.7396 0.3806 0.0071 0.0365 0.0566 Aac 0 30 0 30 30 1.0769 0.5582 0.4703 0.0110 0.0485 0.0111 Adc 0 30 30 0 30 1.1898 0.9898 0.1763 0.0146 0.0176 0.0085 Ada 0 30 30 0 30 1.1898 0.9898 0.1763 0.0146 0.0176 0.0085 Ada 0 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eda 30 0 30 0 1.3059 0.7556 0.3846 0.0071 0.0307 0.1279 eAdac 0 22.5 22.5 0 1.2089 0.7604 0.3394 0.0094 0.0169 0.0828 eAdac 0 22.5 22.5 22.5 1.5251 0.9022 0.3566 0.0036 0.0182 0.1074 eA	eAa	30	30	0	30	0	1.2252	0.8928	0.1832	0.0042	0.0125	0.1325
Aac 0 30 0 30 10769 0.5582 0.4703 0.0110 0.0485 0.0111 Adc 0 30 30 0 30 1.1898 0.9898 0.1763 0.0146 0.0176 -0.0085 Ada 0 30 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eda 30 0 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eda 30 0 30 0 1.2140 0.5405 0.3846 0.0071 0.0307 0.1279 eAda 22.5 22.5 22.5 22.5 0 1.2089 0.7566 0.3846 0.0071 0.0307 0.1279 eAdac 0 22.5 22.5 22.5 22.5 1.511 0.5916 0.4093 0.0058 0.0214 0.1229 edac 22.5 0 22.5 22.5 1.5251 0.9022 0.3566 0.0035 0.0182 0.1074	dac	0	0	30	30	30	1.2204	0.7396	0.3806	0.0071	0.0365	0.0566
Adc 0 30 30 0 30 1.1898 0.9898 0.1763 0.0146 0.0176 0.0085 Ada 0 30 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eda 30 0 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eda 30 0 300 0 1.2140 0.5405 0.3846 0.0071 0.0307 0.1279 eAda 22.5 22.5 22.5 22.5 0 1.2089 0.7566 0.3846 0.0071 0.0307 0.1279 eAdac 0 22.5 22.5 22.5 22.5 1.510 0.5916 0.4093 0.0058 0.0214 0.1229 edac 22.5 0 22.5 22.5 1.5251 0.9022 0.3566 0.0036 0.0182 0.1074 eAac 22.5 22.5 0 22.5 1.3163 0.6304 0.5540 0.0063 0.0182 0.1074<	Aac	0	30	0	30	30	1.0769	0.5582	0.4703	0.0110	0.0485	-0.0111
Ada 0 30 30 30 0 1.2140 0.5405 0.5414 0.0084 0.0172 0.1065 eda 30 0 30 30 0 1.3059 0.7556 0.3846 0.0071 0.0307 0.1279 eAda 22.5 22.5 22.5 22.5 22.5 0 1.2089 0.7604 0.3394 0.0094 0.0169 0.828 Adac 0 22.5 22.5 22.5 22.5 1.510 0.5916 0.4093 0.0058 0.0214 0.1229 edac 22.5 0 22.5 22.5 1.5251 0.9022 0.3566 0.0063 0.0161 0.2111 eAac 22.5 0 22.5 22.5 1.3163 0.6304 0.5540 0.0063 0.0182 0.1074 eAdac 18 18 18 18 1.3587 0.8153 0.4345 0.0072 0.0229 0.0788 eAdac 18 18 <td>Adc</td> <td>0</td> <td>30</td> <td>30</td> <td>0</td> <td>30</td> <td>1.1898</td> <td>0.9898</td> <td>0.1763</td> <td>0.0146</td> <td>0.0176</td> <td>-0.0085</td>	Adc	0	30	30	0	30	1.1898	0.9898	0.1763	0.0146	0.0176	-0.0085
eda 30 0 30 30 0 1.3059 0.7556 0.3846 0.0071 0.0307 0.1279 eAda 22.5 22.5 22.5 22.5 22.5 0 1.2089 0.7604 0.3394 0.0094 0.0169 0.0828 Adac 0 22.5 22.5 22.5 22.5 11510 0.5916 0.4093 0.0058 0.0214 0.1229 edac 22.5 0 22.5 22.5 1.5251 0.9022 0.3566 0.0036 0.0182 0.1171 eAac 22.5 0 22.5 22.5 1.3163 0.6304 0.5540 0.0063 0.0182 0.1074 eAdac 22.5 22.5 0 22.5 1.4623 0.6751 0.5670 0.0035 0.0386 0.1781 eAdac 18 18 18 13587 0.8153 0.4345 0.0072 0.0229 0.0788 eAdac 18 18 18 <	Ada	0	30	30	30	0	1.2140	0.5405	0.5414	0.0084	0.0172	0.1065
eAda22.522.522.522.522.501.20890.76040.33940.00940.01690.0828Adac022.522.522.522.522.51.15100.59160.40930.00580.02140.1229edac22.5022.522.522.51.52510.90220.35660.00360.05160.2111eAac22.522.5022.522.51.31630.63040.55400.00630.01820.1074eAdc22.522.5022.51.46230.67510.56700.00350.03860.1781eAdac18181818181.35870.81530.43450.00720.02290.0788eAdac18181818181.33340.78200.38360.00610.02940.1134eAdac181818181381380.93170.30140.00480.02980.1134eAdac18181818181.33410.93170.30140.00480.02980.1129Standard deviation550.19660.19880.13590.01820.01310.1078	eda	30	0	30	30	0	1.3059	0.7556	0.3846	0.0071	0.0307	0.1279
Adac 0 22.5 22.5 22.5 22.5 1.1510 0.5916 0.4093 0.0058 0.0214 0.1229 edac 22.5 0 22.5 22.5 22.5 1.5251 0.9022 0.3566 0.0036 0.0016 0.2111 eAac 22.5 22.5 22.5 1.3163 0.6304 0.5540 0.0063 0.0182 0.1074 eAdc 22.5 22.5 0 22.5 1.3163 0.6304 0.5540 0.0063 0.0182 0.1074 eAdc 22.5 22.5 0 22.5 1.4623 0.6751 0.5670 0.0035 0.0386 0.1781 eAdac 18 18 18 18 13587 0.8153 0.4345 0.0072 0.0229 0.0788 eAdac 18 18 18 18 13344 0.9317 0.3014 0.0048 0.0298 0.11343 eAdac 18 18 18 18 18	eAda	22.5	22.5	22.5	22.5	0	1.2089	0.7604	0.3394	0.0094	0.0169	0.0828
edac 22.5 0 22.5 22.5 1.5251 0.9022 0.3566 0.0036 0.0516 0.2111 eAac 22.5 22.5 0 22.5 22.5 1.3163 0.6304 0.5540 0.0063 0.0182 0.1074 eAdc 22.5 22.5 0 22.5 1.4623 0.6751 0.5670 0.0035 0.0386 0.1781 eAdac 18 18 18 18 18 13587 0.8153 0.4345 0.0072 0.0229 0.0788 eAdac 18 18 18 18 1334 0.7820 0.3836 0.0061 0.0274 0.1343 eAdac 18 18 18 18 13841 0.9317 0.3014 0.0048 0.0298 0.1142 Average 1.2397 0.6622 0.4175 0.0097 0.0304 0.1229 Standard deviation U 0.1966 0.1988 0.1359 0.0182 0.0131 0.1078	Adac	0	22.5	22.5	22.5	22.5	1.1510	0.5916	0.4093	0.0058	0.0214	0.1229
eAac 22.5 22.5 0 22.5 22.5 1.3163 0.6304 0.5540 0.0063 0.0182 0.1074 eAdc 22.5 22.5 22.5 0 22.5 1.4623 0.6751 0.5670 0.0035 0.0386 0.1781 eAdac 18 18 18 18 18 1.3587 0.8153 0.4345 0.0072 0.0229 0.0788 eAdac 18 18 18 18 138 0.781 0.7820 0.3836 0.0061 0.0274 0.1313 eAdac 18 18 18 18 18 0.3344 0.7820 0.3836 0.0061 0.0274 0.1313 eAdac 18 18 18 18 13841 0.9317 0.3014 0.0048 0.0298 0.1164 Average 1.2397 0.6622 0.4175 0.0097 0.0304 0.1292 Standard deviation US 0.1966 0.1988 0.1359 0.0182 0.0131 0.1078	edac	22.5	0	22.5	22.5	22.5	1.5251	0.9022	0.3566	0.0036	0.0516	0.2111
eAdc 22.5 22.5 0 22.5 1.4623 0.6751 0.5670 0.0035 0.0386 0.1781 eAdac 18 18 18 18 18 1.3587 0.8153 0.4345 0.0072 0.0229 0.0788 eAdac 18 18 18 18 13334 0.7820 0.3836 0.0061 0.0274 0.1343 eAdac 18 18 18 18 13334 0.7820 0.3836 0.0061 0.0274 0.1343 eAdac 18 18 18 18 13841 0.9317 0.3014 0.0048 0.0298 0.1164 Average 1.2397 0.6622 0.4175 0.0097 0.0304 0.1229 Standard deviation U 0.1966 0.1988 0.1359 0.0182 0.0131 0.1078	eAac	22.5	22.5	0	22.5	22.5	1.3163	0.6304	0.5540	0.0063	0.0182	0.1074
eAdac 18 18 18 18 18 18 18 0.8153 0.4345 0.0072 0.0229 0.0788 eAdac 18 18 18 18 18 13334 0.7820 0.3836 0.0061 0.0274 0.1343 eAdac 18 18 18 18 13344 0.9317 0.3014 0.0048 0.0298 0.1164 Average 1.2397 0.6622 0.4175 0.0097 0.0304 0.1229 Standard deviation U U 0.1966 0.1988 0.1359 0.0182 0.0131 0.1078	eAdc	22.5	22.5	22.5	0	22.5	1.4623	0.6751	0.5670	0.0035	0.0386	0.1781
eAdac 18 18 18 18 18 18 13334 0.7820 0.3836 0.0061 0.0274 0.1343 eAdac 18 18 18 18 18 1.3334 0.7820 0.3836 0.0061 0.0274 0.1343 eAdac 18 18 18 1.3841 0.9317 0.3014 0.0048 0.0298 0.1164 Average 1.2397 0.6622 0.4175 0.0097 0.0304 0.1229 Standard deviation 0.1966 0.1988 0.1359 0.0182 0.0131 0.1078	eAdac	18	18	18	18	18	1.3587	0.8153	0.4345	0.0072	0.0229	0.0788
eAdac 18 18 18 18 1.3841 0.9317 0.3014 0.0048 0.0298 0.1164 Average 1.2397 0.6622 0.4175 0.0097 0.0304 0.1229 Standard deviation 0.1966 0.1988 0.1359 0.0182 0.0131 0.1078	eAdac	18	18	18	18	18	1.3334	0.7820	0.3836	0.0061	0.0274	0.1343
Average 1.2397 0.6622 0.4175 0.0097 0.0304 0.1229 Standard deviation 0.1966 0.1988 0.1359 0.0182 0.0131 0.1078	eAdac	18	18	18	18	18	1.3841	0.9317	0.3014	0.0048	0.0298	0.1164
Standard deviation 0.1966 0.1988 0.1359 0.0182 0.0131 0.1078	Average						1.2397	0.6622	0.4175	0.0097	0.0304	0.1229
	Standard d	eviation					0.1966	0.1988	0.1359	0.0182	0.0131	0.1078

This procedure was repeated two more times, so the total volume of solvent mixture added to the leaves was 270 mL.

These solutions were left at rest under forced ventilation until reaching constant weight. Then, the thirty three extracts were fractionated by liquid-liquid extraction, resulting in fibers and four more fractions denominated as neutral (fat and grease), organic (terpenoids and phenolic compounds), basic (alkaloids) and polar (quaternary alkaloids and n-oxide). Fractionation was performed according to the scheme showed in Fig. 1. Gravimetric measurements were performed on all the fractions except the polar one.

All reagents used were purchased from VETEC and were of analytical grade.

3. Mixture designs and models

Mixture models [23,24] have been shown to be useful in describing the behaviors of solvent mixtures to extract substances from plants [20–22]. The extracted crude or fractionated material has been found to depend on the proportions of the solvents in the extracting medium using quadratic or special cubic mixture models,

$$\hat{y} = \sum_{i} b_{i} x_{i} + \sum_{i} \sum_{\langle j} b_{ij} x_{i} x_{j} + \sum_{i} \sum_{\langle j} \sum_{\langle k} b_{ijk} x_{i} x_{j} x_{k}.$$
(1)

The first summation contains linear blending terms for all the mixture components. It describes the relative extraction effects of the pure components in the mixture. The second summation describes the binary interactions between solvent pairs in the extraction process. A significant positive term in this summation indicates that a solvent binary mixture extracts an excess of material relative to that extracted using each solvent separately. The last summation contains terms that will be important if ternary interactions among mixture solvents are important for the extraction process.

Modeling is based on gravimetric measurements of extracted material by solvent mixtures whose compositions are determined using statistical criteria. The simplex-centroid design used here is a basic one and permits the evaluation of models for mixtures having important binary and ternary interactions. The quality of the mixture models are judged by ANalysis Of VAriance (ANOVA) of the regression results.

4. Results

Table 1 contains the yields in grams of the crude extract and the neutral, organic, basic, fiber and residue fractions of the five component mixtures of the simplex centroid design. Linear, quadratic and special cubic models were used to obtain least square fits of the crude extract and each fraction as a function of the proportions of the mixture components. These models were subjected to an analysis of variance (ANOVA) for validation using statistical *F* tests for lack of fit and regression significance.

Quadratic and special cubic models could be fitted to the crude extract data with no significant lack of fit. Since the cubic interaction term involving ethanol, ethyl acetate and dichloromethane was significant at the 95% confidence level the quadratic model was eliminated from further consideration. All coefficients significant below the 90% confidence level were eliminated from the model (the ethyl acetate–acetone, dichloromethane–chloroform and acetone–



Fig. 1. Scheme showing general procedure for fractionating crude extracts.

chloroform binary terms and all ternary terms except the one for ethanol–ethyl acetate–dichloromethane) and the remaining terms were readjusted to the data. The second model had only one coefficient that was not significant at the 95% confidence level, the ethanol–acetone binary coefficient. Its term was eliminated from the model and after a second fitting the following special cubic model was chosen as the most appropriate,

$$Crude = \underbrace{1.15e}_{(\pm 0.07)} + \underbrace{0.69A}_{(\pm 0.06)} + \underbrace{1.11d}_{(\pm 0.06)} + \underbrace{1.20c}_{(\pm 0.06)} + \underbrace{1.62eA}_{(\pm 0.30)} \\ + \underbrace{2.03ed}_{(\pm 0.30)} + \underbrace{1.68ec}_{(\pm 0.28)} + \underbrace{1.23Ad}_{(\pm 0.30)} + \underbrace{0.75Ac}_{(\pm 0.28)} \\ + \underbrace{0.76da}_{(\pm 0.28)} - \underbrace{8.35eAd}_{(\pm 2.11)}$$
(2)

where *e*, *A*, *d*, *a* and *c* represent the ethanol, ethyl acetate, dichloromethane, acetone and chloroform proportions, respectively. It should be mentioned that both models discussed above have very similar coefficients with variations being less than the standard errors in Eq. (2). The ANOVA table for this model is given in Table 2. The ratio of the variance explained by the regression to the total variance, *R*-squared, is 0.93 and its adjusted value considering degrees of freedom is 0.89. The lack of fit mean square — pure error mean square ratio of 9.61 is less than the 95% confidence level $F_{19,2}$ value of 19.45 showing there is no significant lack of fit. Furthermore the regression mean square residual mean square ratio of 17.99 is larger than the 95% confidence value of 2.29 for 11 and 21 degrees of freedom. One can conclude that the varying crude extract yields do indeed have a significant dependence on the mixture component proportions.

Fig. 2 contains a graph of the observed vs. predicted values of the crude material yield. The mixture model predicts the crude extraction yields in Table 1 with a root mean square error of 0.055 g. The estimated experimental error from the triplicate crude results in Table 1 is \pm 0.025 g. Furthermore the residual values are randomly distributed about the line indicating exact agreement. The results are quite satisfactory considering that only some binary and one ternary interaction were used in modeling.

It is convenient to represent the results of Eq. (2) geometrically. The pentagon in Fig. 3 represents the five-component mixture. Each vertex stands for one of the pure solvents and its numerical value is the corresponding linear coefficient, b_i . The values at the midpoints of the lines are the binary coefficients involving the two solvents represented at its endpoints. Ternary coefficients can be given by numbers in triangles whose vertices correspond to the three solvents involved in the interaction. Fig. 3 only contains values for those coefficients that are significantly different from zero at the 95% confidence level. Note that six of the possible ten binary coefficients are statistically significant.

The largest crude yield in Table 1 is obtained using a 1:1 binary mixture of ethanol and dichloromethane. For the $e = d = \frac{1}{2}$ mixture

$$Crude = 1.09(1/2) + 1.12(1/2) + 2.06(1/2)(1/2) = 1.62$$
 (3)

compared with the experimental value of 1.67 g. As such this maximum crude yield in Table 1 is a result of the high extraction capabilities of the pure ethanol and dichloromethane solvents as well

as their large synergic interaction. The second largest crude extract yield of 1.53 occurs for the binary ethanol–chloroform mixture and can be attributed to the binary interaction between these two solvents as well as their individual extraction capabilities. These contributions in Eq. (2) predict an extraction yield of 1.58 g. These two binary interactions are extremely important for increasing the quantity of extracted material from *Mikania laevigata Sch. Bip.* leaves. The lowest yield from this plant occurs for the pure ethyl acetate solvent since it has the lowest extracting effect of the five solvents used here. These results illustrate the importance of using mixtures for extraction procedures since binary interactions can favor substantial increases in yields from plant material.

A response surface representation of the crude extract yields as a function of the five solvent proportions cannot be represented in three dimensional space. Analysis of the model in Eq. (1) indicates that the mixture response surface of crude extract yield as a function of the ethanol, dichloromethane and chloroform proportions, presented in Fig. 4 exhibits the highest yields. All these solvents have high linear extraction capabilities as well as large binary interactions between ethanol and dichloromethane and ethanol and chloroform. An alternative choice of ethanol, ethyl acetate and dichloromethane is also informative since significant binary synergic effects exist between all these solvents. But ethyl acetate has a much smaller individual extracting capability for the crude material of this plant than does chloroform and the ternary interaction among these three solvents is negative. The predicted maximum yield of crude extract material in Fig. 4 is indeed seen to occur for binary mixtures of ethanol and dichloromethane in almost equal proportions.

The linear, quadratic and special cubic models were examined for each of the fractions and for the residues. All three models for the basic fraction yields suffered from lack of fit. The ANOVA tables for the neutral and fiber fractions and the residues showed that quadratic models are to be preferred for all three. Their R^2 values ranged between 0.64 and 0.75. These models contain significant synergic ethanol-ethyl acetate, ethanol-acetone, ethanol-dichloromethane and acetone-chloroform interactions at the 95% confidence level as well as one atagonistic interaction between ethanol and acetone. The linear model was found to be most adequate for the organic fraction but had an R^2 value of only 0.37. Unlike the model for crude extract in Eq. (2) for which the ratio of the calculated F values for regression significance was more than five times the F 95% critical value, these models had calculated F values that were only slightly larger than their corresponding critical values. Rather than detail each model a principal component analysis is reported here that contains the relevant information about the fractionated yields as functions of the solvent proportions.

A biplot of the principal component loadings and scores is shown in Fig. 5. The first two principal components account for 67% of the total data variance. The loading graph of the first principal component which describes 38% of the variance has positive loadings for the crude, residue, fiber and organic fractions located on the right of the graph and a negative one for the basic fraction on the left. The score graph with points for the mixture compositions shows that the solutions containing ethanol are concentrated toward the right of the graph whereas the left side contains points representing mixtures

ANOVA for the regression results of the special cubic model for the total mass of crude extracts of Mikania laevigata Sch. Bip. leaves.

6	1		8	1	
Variation source	Sum of squares	Degrees of freedom	Mean square	Calculated ^a F-value	Probability ^b
Regression	1.118997	11	0.101727	17.99	0.000
Residuals	0.118746	21	0.005654		
Lack of fit	0.117461	19	0.006182	9.61	
Pure error	0.001285	2	0.000643		
Total	1.237743	32	0.038679		

^a Lack of fit mean square/pure error mean square ratio. Corresponding critical *F* values at the 95% confidence level for special cubic model fits are *F*_{11,21,0.05} = 2.29 *F*_{19,2,0.05} = 19.45. ^b Significance probability level.



Fig. 2. Observed values of the extract crude material plotted against those calculated from Eq. (1).

with predominating presences of dichloromethane, chloroform, ethyl acetate and acetone.

Solvent mixtures of ethanol with either chloroform or dichloromethane have score points close to the crude and residue loadings. As can be verified in Table 1 these mixtures have the largest quantities of both crude and the residue matter of all the solvents and mixtures tested. The *ec* and *ed* binary solutions extracted 1.53 g and 1.67 g of crude material compared to about 1.1 g extracted by pure ethanol, dicloromethane and chloroform. Extractions carried out for the 14 mixtures containing ethanol with chloroform and/or dichloromethane resulted in an average yield of 1.40 ± 0.14 g whereas the other 19 mixtures had an average yield of 1.12 ± 0.15 g. The *ec* and *ed* mixtures had 0.22 g and 0.47 g of residue, respectively. With the exception of the *ac* pair with 0.43 g all other pure solvents or mixtures in Table 1 resulted in residue masses with an average of 0.10 ± 0.07 g.

Although the organic fraction is efficiently extracted with mixtures containing ethanol, acetone-containing mixtures are also useful for this extraction. For mixtures containing either ethanol or acetone an average of 0.46 ± 0.12 g of organic fraction is extracted whereas this average is only 0.26 ± 0.07 g for the other mixtures. Since acetone and its binary mixtures all have negative second PC scores the organic fraction has a lower second PC loading than do the crude material and



Fig. 3. Summary of crude material model results for ethanol, ethyl acetate, dichloromethane, acetone and chloroform in the extraction mixtures. Points at vertices represent linear coefficients. Points at the center of a line represent binary coefficients for the solvents defining the line. Points at the center of a triangle represent ternary coefficients for the solvents at the vertices of the triangle. Only 95% significant coefficient values are presented.

the residue. Acetone-containing mixtures are also important for extracting the fiber fraction with the ac and da binary mixtures providing the largest amounts. Their score points are close to the fiber loading point.

The largest neutral fraction yields are extracted with ternary and quartenary mixtures, especially the *Adc*, *eAa* and *edc* ternary mixtures. These three mixtures have large positive score values on the second principal component. The *Adc* mixture also has a large negative first principal component score since this mixture extracts the largest amount of the basic fraction in Table 1. Pure chloroform and the *Ac* and *Aa* mixtures also yield relatively large quantities of this fraction. Their score points are close to the basic loading point in Fig. 5.

5. Discussion

Solvent composition extraction of plant material can be based on the solvent selectivity triangle, first described by Snyder [25] and Rohrschneider [26] in which organic solvents are classified according to their ability to interact with the solute as a proton donor (basicity, β), as a proton acceptor (acidity, α) or a dipole (dipolarity, Π).

This triangle is shown in Fig. 6 where the pure acidity, α , basicity, β , and dipolarity, π , selectivity parameters are represented at the vertices. The points in this figure correspond to the values of these parameters for the extraction mixtures in Table 1. The α , β and π values for the pure solvents were taken from [27] and component proportion weighted averages were calculated for the solvent mixtures. These points define an approximate inverted trapezoidal region with pure ethanol at one vertex, acetone and ethyl acetate at another and the chloroform–dichloromethane pair at vertices in close proximity at the bottom. Solvents that are located far away from each other in this triangle have the largest differences in these properties and, hence, greater differences in extraction selectivity will be found on mixing them. The model in Eq. (2) and the results of the principal component analysis can be interpreted in terms of the parameters of the selectivity triangle.

All the statistically significant binary interaction terms in Eq. (2) have solvent pairs from different extremities of the domain occupied by the mixture points. The 1.62 *eA* binary interaction term involves the ethanol solvent with strong α and β properties and low dipolarity and ethyl acetate with relatively high basicity and dipolarity values and low acidity. The 2.03 *ed* synergic interaction involves ethanol with dichloromethane that has high acidity and dipolarity parameters and low basicity values. Other binary coefficients that are significant at the 95% confidence level involve points from different extremities of the trapezoidal region in Fig. 6; ethanol-chloroform (1.68), ethyl acetate–dichloromethane (1.23), ethyl acetate–chloroform (0.75) and dichloromethane–acetone (0.76). It is interesting that solvents having similar α , β , and π parameters, the acetone–ethyl acetate and chloroform–dichloromethane pairs used in our study, do not have synergic interaction terms in Eq. (2).

The negative ethanol–ethyl acetate–dichloromethane ternary coefficient is also significant at the 95% confidence level. It shows that the *eA*, *ed* and *Ad* binary interactions are less effective when more than one of these interactions can occur in the solvent mixture. However the sum of these three binary contributions more than cancels the negative ternary effect enhancing solvent extraction.

In a more recent study [22] of a different sample of *Erythrina speciosa* Andrews leaves with a simplex centroid design involving four solvents validated mixture designs were found for the fiber, neutral, organic, basic and residue fractions. All these models had significant positive binary coefficients involving the dichloromethane–hexane, ethanol–dichloromethane, ethanol–hexane, dicloromethane–hexane and ethanol–acetone pairs. Only one significant negative binary coefficient was found in all these models involving the hexane–acetone pair.

L.M.Z. Garcia et al. / Chemometrics and Intelligent Laboratory Systems 103 (2010) 1-7



Fig. 4. Contour response surface for the extracted quantity of crude material from the plant as a function of the ethanol, dichloromethane and chloroform proportions.

The relative positions of the experimental design mixture points in the principal component biplot in Fig. 5 can also be interpreted in terms of the α , β , and π selectivity triangle parameters. The first principal component with negative scores for the [*A*, *a*, *Aa*, *Aac*] and the [*d*, *c*, *dc*, *Adc*] clusters correspond to the points on the lower right region of the selectivity triangle. These mixtures do not contain ethanol and have relatively high dipolarity values and either low basicity or acidity values. The ethanol containing mixtures have positive first principal component scores and correspond to mixtures in the selectivity triangle with low dipolarity values and higher acidity and/or basicity parameters.

The second principal component is characterized by low score values for the [*A*, *a*, *Aa*, *Aac*] points and high ones for the [*d*, *c*, *dc*, *Adc*] points. As such the direction of the second PC is characterized by changes from relatively low acidity–high basicity parameters to low basicity–high acidity parameters. This is consistent with the higher PC



Fig. 5. Principal component biplot graph of loadings and scores for the results of the simplex centroid design for the crude material, the organic, neutral, basic and fiber fractions and for the residue material. The *e*, *A*, *d*, *a* and *c* symbols stand for ethanol, ethyl acetate, dichloromethane, acetone and chloroform.



Fig. 6. The alpha, beta and gamma selectivity triangle with points corresponding to the mixtures of the simplex centroid design.

loadings for the neutral fraction and the lower one for the basic fraction.

Solvents with similar α , β and π parameters, *cd* and *Aa*, have binary solvent mixture score points very close to the points of their pure solvents and their mutual presences do not effectively enhance the amount of extracted material. On the other hand, solvent pairs having synergic interactions in Eq. (1), *eA*, *ed*, *ec*, *Ad*, *Ac* and *da*, are positioned farther from the points of their corresponding pure solvents. As such simplex centroid designs using ethanol and either acetone or ethyl acetate and either chloroform or dichloromethane would require only 15 different mixtures, about half of the number required by our five-component design.

6. Conclusions

Solvent mixtures have been shown to extract significantly larger amounts of crude material from *Mikania laevigata Sch. Bip.* leaves than the pure solvents. All solvent mixtures containing either ethanol, ethyl acetate, dichloromethane or acetone extracted more crude material than did the pure solvent. Of the 17 solvent mixtures containing chloroform only two extracted less crude than did pure chloroform. The molecular interactions existing in mixtures that most effectively enhance the extraction properties of mixtures are those with disparate acidity, basicity and dipolarity parameters. So the solvent selectivity triangle used in chromatographic work should be useful in planning efficient solvent extraction procedures for other biological samples.

Acknowledgements

The authors are grateful to the Conselho Nacional de Pesquisa (CNPq) and Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) for financial support that made the reported research possible. CNPq is acknowledged for providing a doctoral fellowship for P. K. S. and research fellowships to I. S. S. and R. E. B. L. M. Z. G. thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for a masters fellowship and T.F.O. thanks CNPq for a scientific initiation scholarship.

References

 R. Croteau, T.M. Kutchan, N.G. Lewis, Natural Products (Secondary Metabolites), Courier Companies Inc, Rockville, USA, 2000, pp. 1250–1318.

- [2] S.R. Rao, G.A. Ravishankar, Biotechnol. Adv. 20 (2002) 101-153.
- [3] J.B. Harborne, Biochemical Aspects of Plant and Animal Coevolution, Academic Press, London, 1978.
- [4] J.B. Harborne, F.A. Tomas-Barberan, Ecological Chemistry and Biochemistry of Plant Terpenoids, Oxford University Press, USA, 1991.
- [5] F. Oliveira, M.L. Saito, L.O. Garcia, Rev. LECTA 12 (1994) 23-65.
- [6] V. Lucas, Rev. Flora Med. 1 (1992) 101-132.
- [7] M.G.R. Leite, C.L. Souza, M.A.M. Silva, L.K.A. Moreira, F.J.A. Matos, G.S.B. Viana, Rev. Bras. Farm. 1 (1993) 12–15.
- [8] C.R. Silva, V.S. Gomes, I.C. Kulkamp, L.A. Kanis, Braz. J. Pharmacogn. 4 (2008) 594-599.
- [9] R.S. Moura, S.S. Costa, J.M. Jansen, C.A. Silva, C.S. Lopes, M. Bernardo-Filho, V. Nascimento da Silva, D.N. Criddli, N.B. Portela, L.M.S. Rubenich, R.G. Araújo, L.C.R.M. Carvalho, J. Pharm. Pharmacol. 54 (2002) 249–256.
- [10] E.S. Suyenaga, E. Reche, F.M. Farias, E.E. Schapoval, C.G.M. Chaves, A.T. Henriques, Phytother. Res. 16 (2002) 519-523.
- [11] Y. Markus, Solvent Mixtures: properties and selective solvatation, Marcel Dekker, ICN. New York. 2002.
- [12] Y. Marcus, J. Chem. Soc., Perkin Trans. 2 (1994) 1015-1021.

- [13] P. Chatterjee, A.K. Laha, S. Bagchi, J. Chem. Soc., Faraday Trans. 88 (1992) 1675-1678.
- Y. Marcus, Chem. Soc. Rev. 22 (1993) 409-416. [14]
- J.M. Čanadanović-Brunet, S.M. Djilas, G.S. Ćetković, V.T. Tumbas, A.I. Mandić, V.M. [15] Čanadanović, Int. J. Food Sci.Tech. 41 (2006) 667-675.
- U.D. Chavan, F. Shahidi, M. Naczk, Food Chem. 75 (2001) 509–512. [16]
- [17] L.M. Cheung, P.C.K. Cheung, V.E. Ooi, Food Chem. 81 (2003) 249–255.
 [18] R.P. Singh, K.N.C. Murthy, G.K. Jayaprakasha, J. Agric, Food Chem. 50 (2002) 81–86.
 [19] A.H. Goli, M. Barzegar, M.A. Sahari, Food Chem. 92 (2005) 521–525.
- [20] P.K. Soares, R.E. Bruns, I.S. Scarminio, J. Sep. Sci. 30 (2007) 3302–3310.
- [20] F.K. Soares, R.E. Bruns, I.S. Scarminio, Anal. Chim. Acta 613 (2008) 48–55.
 [22] P.K. Soares, R.E. Bruns, I.S. Scarminio, I.S. J. Sep. Sci. 32 (2009) 644–652.
- [23] J.A. Cornell, Experiments with mixtures, 3 rd ed.Wiley, NY, 2002.
- [24] R.E. Bruns, I.S. Scarminio, B.B. Neto, Statistical Design Chemometrics, Elsevier, Amsterdam, 2006.
- L.R. Snyder, J. Chromatog. 92 (1974) 223-230. [25]
- [26] L. Rohrschneider, Anal. Chem. 45 (1973) 1241–1247.
 [27] LR. Snyder, P.W. Carr, S.C. Rutan, J. Chromatog. 656 (1993) 537–547.

Versão pré revisada de Journal of Separation Science 2009, 32, 644-652 DOI: 10.1002/jssc.200800534 © Copyright 2009 WILEY-VCH

This is the pre-peer reviewed version of the following article:

Statistical mixture design investigation of fractionated and total extracts from *Erythrina speciosa* Andrews leaves.J. Sep. Sci., 32, 644-652, 2009.DOI: 10.1002/jssc.200800534

Statistical mixture design investigation of fractionated and total extracts from Erythrina speciosa Andrews leaves

Patricia Kaori Soares¹ Roy Edward. Bruns¹ Ieda Spacino Scarminio²

¹Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas, SP, Brazil ²Laboratório de Quimiometria em Ciências Naturais, Departamento de Química, Universidade Estadual de Londrina, CP 6001, 86051-990 Londrina, PR, Brazil

Running title: Mixture Design Extractions - Erythrina speciosa Andrew leaves

Correspondence: Dr. Roy Edward Bruns, Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970, Campinas, SP, Brazil.
E-mail: bruns@iqm.unicamp.br
Fax: +55-19-3521-3023

Keywords: extraction optimization / simplex centroid design / mixture modeling / principal components / *Erythrina speciosa* Andrews leaves

Abstract

The simplex centroid mixture design for the ethanol, dichloromethane, hexane and acetone solvents has been applied to the extraction of crude mass and the fiber, organic, neutral and basic fractions as well as the fractionation residues of *Erythrina speciosa* Andrews leaves. Binary and ternary synergic solvent interactions are seen to provide dominant contributions to the extraction of both crude mass and all the fractions. Quadratic and special cubic mixture models precisely predict the extracted quantities of each fraction and the residue as a function of the proportions of the four solvents. Different solvent mixtures are found to be the most efficient extractors for the different fractions: binary dichloromethane – hexane mixtures for the fiber fraction, ternary ethanol –

dichloromethane - acetone mixtures for the neutral fraction, binary ethanol – dichloromethane mixtures for the organic fraction, crude extract and residue values and ternary ethanol – dichloromethane – hexane mixtures for the basic fraction. Principal component analysis shows that the ethanol – dichloromethane mixtures are important for extracting large quantities of the basic and organic fractions as well as of the residue and crude masses.

1 Introduction

Secondary metabolites are economically important in phytomedical, nutraceutical and industrial applications. Most of the plant compounds that have been found to be medicinally useful and interesting tend to be secondary metabolites. It has been suggest that these compounds may have been synthesized by plants as part of their defense system [1-3]. Secondary metabolites represent features that can be expressed in terms of ecological, taxonomical and biochemical differentiation and diversity. The biosynthesis and accumulation of secondary metabolites provide a basis for biochemical systematics and chemosystematics of wide molecular diversity. In addition, the metabolites throughout the plant kingdom represent an extremely rich biogenic resource for the discovery of novel drugs and for the development of innovative drugs. Secondary metabolites can be classified into three groups on the basis of their chemical structure: alkaloids, terpenoids and phenolic compounds. [1]

Much plant analysis is devoted to the isolation and identification of secondary metabolites in a particular species or group of species with the expectation that some of the constituents may be novel or have unusual structures [4-7]. The classical chemical procedure for obtaining secondary metabolites from dried plant tissue is to continuously extract powdered material with a range of solvents, for example, stating in turn with ether, petroleum and chloroform (to separate lipids and terpenoids) and then using alcohol and ethyl acetate (for more polar compounds). However, one rarely achieves complete separation of constituents and the same compounds may be recovered in varying proportions in several fractions [8].

In spite of all the extraction applications reported in the literature few systematic studies have been made investigating the effect of changing solvent proportions in ternary and quaternary mixtures. Recently our group has shown that statistical mixture designs permit the development of rigorous but economical procedures for demonstrating the effects of solvent changes on the extracted constituents of plant material [9,10]. Experiments were carried out using mixtures of solvents that have been chosen using statistical criteria to minimize modeling error as well as reduce the number of experiments that must be carried out. Besides helping organize our effort, these designs permitted the identification and characterization of synergistic and antagonistic interaction effects between different solvents. Additive models obtained from experiments using only pure solvents could not have indicated the existence of these effects. Specifically, changes in the intensities of chromatographic peaks were observed for *Erythrina speciosa* Andrew leave extracts obtained by methodically changing ethanol, dichloromethane, hexane and acetone proportions according to a mixture design. Statistical models were then
found that permit precise determination of solvent extraction mixtures that were optimal for extracting constituents eluded for different chromatographic peaks. This procedure could also be applied to investigate the extraction of different plant fractions as well as the crude plant extract. Different solvent mixtures can be expected to be more efficient extractors for different fractions. In this investigation simplex centroid mixture designs for ethanol, dichloromethane, hexane and acetone are applied to the extraction of plant material from *Erythrina speciosa* Andrew leaves. The crude extract was then fractionated using standard procedures permitting detailed information about the solvent dependencies of the neutral, organic, basic and fibrous fractions as well as the residual material.

2 Experimental

For this study *Erythrina speciosa* Andrews leaves were used. A voucher specimen of this plant was deposited in the herbarium of the Londrina State University and authenticated by M. R. C. Paiva and registered under the FUEL 35133 number. The collection was performed in January 2008. Leaves were selected avoiding material damaged by insects and/or fungus. Drying was carried out at room temperature for twelve days. During this period, the leaves were protected from light, humidity, fungus and insect attacks.

Extraction mediums were prepared using mixtures of four components, (1) ethanol, (2) dichloromethane, (3) hexane and (4) acetone whose proportions were varied according to the simplex centroid design in Figure 1. The proportions of each solvent used in the extraction mixtures are specified in Table 1. Nineteen extractions were carried out with 15 different mixtures. A five-run replicate was performed at the center point so experimental error could be estimated.



Figure 1: Simplex centroid design used to optimize the extraction mixtures x_1 , x_2 , x_3 and x_4 represent the proportions of ethanol, dichloromethane, hexane and acetone, respectively.

		Solv	vents			Masses					
Extract	Ethanol	Dichlor.	hexane	Acetone	Crude	Fibers	Neutral	Organic	Basic	Residues	
1	120	0	0	0	0.8268	0.0399	0.2422	0.2114	0.0104	0.3229	
2	0	120	0	0	0.5525	0.0775	0.2934	0.1747	0.0051	0.0018	
3	0	0	120	0	0.3645	0.0556	0.0971	0.2089	0.0037	-0.0008	
4	0	0	0	120	0.535	0.0428	0.3168	0.1454	0.0037	0.0263	
5	60	60	0	0	1.2417	0.1015	0.3722	0.3078	0.0286	0.4316	
6	60	0	60	0	1.0423	0.0947	0.365	0.1749	0.0181	0.3896	
7	60	0	0	60	0.9518	0.0952	0.2948	0.2427	0.022	0.2971	
8	0	60	60	0	0.811	0.3402	0.3009	0.1301	0.0054	0.0344	
9	0	60	0	60	0.5222	0.0593	0.296	0.1504	0.0061	0.0104	
10	0	0	60	60	0.4852	0.0267	0.2916	0.1176	0.0027	0.0466	
11	40	40	40	0	1.1249	0.1571	0.3726	0.2439	0.0386	0.3127	
12	40	40	0	40	0.9481	0.0697	0.4352	0.2036	0.0142	0.2254	
13	40	0	40	40	0.8224	0.0753	0.4271	0.1531	0.0114	0.1555	
14	0	40	40	40	0.7708	0.2329	0.2676	0.1301	0.0091	0.1311	
15	30	30	30	30	0.8557	0.0611	0.3875	0.189	0.0172	0.2009	
16	30	30	30	30	0.8600	0.0765	0.3602	0.2249	0.0118	0.1866	
17	30	30	30	30	0.8713	0.1168	0.4312	0.1788	0.0199	0.1246	
18	30	30	30	30	0.8435	0.081	0.3796	0.196	0.0228	0.1641	
19	30	30	30	30	0.8657	0.0868	0.4332	0.1913	0.0159	0.1385	
		Average			0.8050	0.0995	0.3349	0.1881	0.0140	0.1683	
Standard deviation				0.2262	0.0743	0.0831	0.0469	0.0095	0.1339		

Table 1: Extractor mixture composition in mL and yield in g of the crude extracts, fibers, and neutral, organic and basic fractions.

Each extract was prepared by weighing 15.000 g of dried and crushed leaves and adding 120mL of the solvent mixtures listed in Table 1. These mixtures were placed in an ultrasonic bath (Unique, model USC 1400) for 30 min. The bathwater was changed every 10 min to maintain constant water temperature. The extracts were filtered through filter paper to separate the solution from small pieces of leaves and the solution was placed in an identified and weighted flask. Another 120mL of the solvent mixture were added to the leaves. This procedure was repeated five times, so the total volume of solvent mixture added to the leaves was 600mL.

These solutions were left at rest under forced ventilation until reaching constant weight. Then, the nineteen extracts were fractionated by liquid–liquid extraction, resulting in fibers and four more fractions denominated as neutral extract (fats and waxes), moderately polar extract (terpenoids and phenolic compounds), basic extract (most alkaloids) and polar extract (quaternary alkaloids and n–oxide), although the polar fraction was discarded since a fungus grew in the solution owing to the presence of water that was difficult to remove from the other constituents.

Fractionation of the crude extract was performed according to the scheme showed in Figure 2. All reagents used were purchased from VETEC and were of analytical grade.



Figure 2: Scheme showing general procedure for fractionating crude extracts.

3 Statistical mixture design

Solvent mixtures used in the extraction procedure followed the simplex centroid design [11,12] for four solvents shown in Figure 1. The four points at the vertices of the tetrahedron correspond to extractions carried out using each one of the pure solvents, ethanol, dichloromethane, hexane and acetone. The six midpoints of the sides of the tetrahedron correspond to 1:1 binary mixtures of these solvents. Ternary mixtures using equal proportions are also investigated and are represented in Figure 1 by points at the centroids of the four faces of the tetrahedron. The tetrahedron center point represents a four–component mixture in equal proportions that was investigated in pentuplicate to estimate experimental error. The proportions of the solvents are presented in the Table 1 for all the design mixtures. Statistical models were calculated using the Statistica 6.0 software [13].

4 Results and discussion

The masses of the crude extract and the fibers, neutral, moderately polar extract (organic) and basic fractions of the experiments for the simplex centroid design are given in Table 1.

The average and standard deviations for each of these extracts is given in the last two rows of Table 1. The final column lists the estimated residue values calculated from the mass of crude extract minus the sum of the masses of the fractions of all extracts.

Correlations between the six variables were calculated and the results can be seen in Table 2. Bold-faced correlation values indicate those that are significant at the 95% confidence level. High correlation values were found between crude extract masses and both the residue and basic fraction. This means that as the extracted crude mass increases on solvent change, so do the residue values and the basic fraction masses. As a consequence moderate correlation is seen for the residues and the basic fraction as well. The fiber fraction doesn't correlate with any other fraction. It is interesting that no significant negative correlations are observed

 Table 2: Correlations between crude extract masses, neutral, organic, fiber and basic fraction masses and residue values.

Variables	Crude	Organic	Basic	Residue	Neutral	Fibers
Crude	1.00	0.63	0.84	0.88	0.62	0.26
Organic	0.63	1.00	0.71	0.70	0.11	-0.19
Basic	0.84	0.71	1.00	0.75	0.50	0.09
Residue	0.88	0.70	0.75	1.00	0.35	-0.05
Neutral	0.62	0.11	0.50	0.35	1.00	0.00
Fibers	0.26	-0.19	0.09	-0.05	0.00	1.00

Mixture linear, quadratic and special cubic models were used to fit the crude extract and each fraction given in Table 1. Analysis of variance (ANOVA) was used to test for model lack of fit and model significance in the validation of all the preferred models discussed below. The ANOVA results for all the validated models are given in Table 3. All these models, except the one for crude extract, have calculated lack of fit mean square / pure error mean square ratios less than the F distribution critical value at the 95% significance level showing that they provide excellent fits to the experimental data. Residual plots showed random behavior consistent with these ANOVA results. Furthermore all have regression mean square / residual mean square ratios that are larger than their corresponding F distribution value at the 95% confidence level showing that all models describe significant dependencies of the extracted masses on the proportions of the solvents used in the extraction procedure.

Variation	Sum of	Degrees of	••	Calculated ^{a)}	b b b b b
source	squares	freedom	Mean square	F-value	Probability"
Crude					
Regression	0.912745	13	0.070211	40.24	0.0003
Residuals	0.008723	5	0.001745		
Lack of fit	0.008275	1	0.008275	73.88	0.001
Pure error	0.000448	4	0.000112		
Total	0.921468	18	0.051193		
Fiber					
Regression	0.094799	13	0.007292	7.80	0.016
Residuals	0.004670	5	0.000934		
Lack of fit	0.002998	1	0.002998	7.17	0.055
Pure error	0.001672	4	0.000418		
Total	0.099470	18	0.005526		
Neutral					
Regression	0.120171	13	0.009244	10.69	0.008
Residuals	0.004324	5	0.000865		
Lack of fit	0.000105	1	0.000105	0.099	0.767
Pure error	0.004218	4	0.001055		
Total	0.124494	18	0.006916		
Organic					
Regression	0.035042	9	0.003894	7.56	0.003
Residuals	0.004633	9	0.000515		
Lack of fit	0.003431	5	0.000686	2.28	0.222
Pure error	0.001202	4	0.000301		
Total	0.039674	18	0.002204		
Basic					
Regression	0.001549	13	0.000119	7.40	0.019
Residuals	0.000080	5	0.000016		
Lack of fit	0.000011	1	0.000011	0.665	0.460
Pure error	0.000069	4	0.000017		
Total	0.001629	18	0.000091		
Residue					

Table 3: ANOVA for the regression results of the preferred models for the total mass and the four fractions and residues from crude extracts of *E. speciosa* Andrews leaves.

a) Lack of fit mean square/pure error mean square ratio. Corresponding critical F values at the 95% confidence level for special cubic model fits are $F_{13,5,0.05}$ = 4.65 $F_{1,4,0.05}$ =7.71 and for quadratic model fits are $F_{9,9,0.05}$ = 3.18and $F_{5,4,0.05}$ = 6.26.

0.024004

0.001124

0.001549

0.001017

0.017648

21.36

1.522

0.002

0.285

b) Significance probability level.

0.312050

0.005619

0.001549

0.004069

0.317668

13

5

1

4

18

Regression

Residuals

Lack of fit

Pure error

Total

The linear, quadratic and special cubic models all presented significant lack of fit at the 95% confidence level for the crude extract masses. This can be seen in Table 3, for example, since the lack of fit mean square/pure error mean square ratio for the special cubic model of 73.88 is larger then the $F_{1,4}$ 95% critical value of 7.71.

The best way to analyze the crude extract masses is by examining Figure 3 where their masses are presented for each mixture design point. Of all the pure components ethanol extracts give the highest quantity of constituints, 0.826g followed by dichloromethane, 0.552g, acetone, 0.535g and hexane 0.364g. However the most striking feature of Figure 2 is that the highest masses are extracted by binary or ternary mixtures. The largest mass, 1.241g, extracted by a 1:1 binary mixture of ethanol and dichloromethane indicates a strong synergic effect between these solvents since this value is much larger than the average of the crude extract masses for pure ethanol and dichloromethane, 0.689g. The 1:1 ethanol–hexane mixture extracts 1.042g compared with the 0.595g average for extraction with these two pure components. Strong synergic effects are also indicated for the ethanol–acetone and hexane–dichloromethane pairs of solvents for crude extract masses. On the other hand the hexane–acetone and acetone–dichloromethane mixtures show very additive blending with the absence of interaction effects.





Synergic ternary effects are also indicated by the values in Figure 3. All the mass values for the center point of the four faces of the tetrahedron have mass values significantly larger than the means of the masses of the vertices of the face showing that more mass is extracted from the 1:1:1 ternary mixtures than that expected from additive blending of the three pure components. These binary and ternary synergic effects provide clear evidence that the mixtures are much more effective at extracting plant constituints than the pure solvents themselves.

The fiber fraction extraction masses are well reproduced by the special cubic model

F = 0.0390e + 0.0766d + 0.0547h + 0.0419a + 0.2019ed + 0.2185eh + 0.2461ea + 1.1253dh $(\pm 0.0305) (\pm 0.0305) (\pm 0.0305) (\pm 0.0305) (\pm 0.1487) (\pm 0.1487) (\pm 0.1487) (\pm 0.1487) (\pm 0.1487)$ + 0.0273da - 0.0593ha - 2.5509edh - 1.5843eda - 1.0260eha + 0.8267dha $(\pm 0.1487) (\pm 0.1487) (\pm 0.9333) (\pm 0.9333) (\pm 0.9333) (\pm 0.9333)$ (1) where *e*, *d*, *h* and *a* represent the ethanol, dichloromethane, hexane and acetone proportions respectively. Standard error estimates are presented in parenthesis directly below their corresponding model coefficients. Bold-faced coefficients indicate those that are significant at the 95% confidence level. All the linear blending terms are not significantly different from zero at the 95% confidence level. This means that fiber extraction is the result of synergic interaction effects. The only significant binary interaction is highly synergic, 1.1253g, and occurs between dichloromethane and hexane. For a 50%-50% dichloromethane–hexane mixture

Fibers =
$$0.0766 \left(\frac{1}{2}\right) + 0.0547 \left(\frac{1}{2}\right) + 1.1253 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = 0.347g$$
 (2)

compared with the experimental value of 0.340g. The latter value is the highest of all the fiber fraction extraction masses in Table 1. The response surface contour curves for the model in Eq.1 is presented in Figure 4a. As can be deduced there a 50%-50% dichloromethane-hexane mixture is predicted to extract the maximum amount of fiber mass.

The neutral fraction model is given by

$$\begin{split} N &= \mathbf{0.2420}e + \mathbf{0.2932}d + \mathbf{0.0969}h + \mathbf{0.3166}a + \mathbf{0.4233}ed + \mathbf{0.7871}eh + 0.0669ea + \mathbf{0.4283}dh \\ (\pm 0.0294) & (\pm 0.0294) & (\pm 0.0294) & (\pm 0.0294) & (\pm 0.1431) & (\pm 0.1431) & (\pm 0.1431) & (\pm 0.1431) \\ & - 0.0306da + 0.3443ha - 0.6629edh + \mathbf{2.5875}eda + 1.9191eha - 1.4789dha \\ & (\pm 0.1431) & (\pm 0.1431) & (\pm 0.8980) & (\pm 0.8980) & (\pm 0.8980) & (\pm 0.8980) \end{split}$$

Note that the ethanol, dichloromethane and acetone linear blending coefficients are significantly larger than the hexane linear coefficient. Of the pure solvents hexane is the least effective in extracting the neutral fraction, as can be seen in Table 1. Of the four 1:1 binary mixtures the ethanol–dichloromethane and ethanol–hexane mixtures extract about 0.37g of the neutral fraction, significantly more than the dichloromethane–hexane and dichloromethane–acetone mixture results of about 0.30g. This is partially explained by the significant ethanol–dichloromethane and ethanol–hexane binary coefficients in Eq. (3). Although the dichloromethane–hexane binary coefficient is also significant and positive the relatively low hexane linear coefficient explains why the 50%-50% dichloromethane–hexane mixture extracts less mass than the 50%-50% ethanol – dichloromethane–acetone mixture can be accounted for by the positive significant ternary interaction coefficient, 2.5875, of these three solvents, as well as by the lower order coefficients. The model in Eq. (3) predicts the extraction of a neutral fraction mass of 0.4300g in good agreement with the experimental value. The contour plot for the neutral fraction masses predicted by Eq. (3) is given in Figure 4b. As can be seen, there, a 1:1:1 ethanol–dichloromethane–acetone mixture is predicted to extract the largest quantity of neutral fraction components.

A quadratic model proved to be most adequate for describing the mass of the extracted organic fraction,

O = 0.2115e + 0.1721d + 0.2042h + 0.1475a + 0.4667ed - 0.0949eh $(\pm 0.0223) (\pm 0.0223) (\pm 0.0223) (\pm 0.0223) (\pm 0.0943) (\pm 0.0943)$ + 0.1803ea - 0.1526dh - 0.0673da - 0.2287ha $(\pm 0.0943) (\pm 0.0943) (\pm 0.0943) (\pm 0.0943)$ (4)

The most prominent features of this model are the significant synergic interaction effect between the ethanol and dichloromethane proportions and the antagonistic hexane–acetone coefficient. Indeed these coefficients are important in explaining why the largest mass for the organic fraction in Table 1 occurs for the 50%-50% ethanol–dichloromethane mixture, 0.3078g, and the smallest one occurs for the 1:1 hexane–acetone mixture, 0.1176g. The model in Eq. (4) predicts organic fraction masses of 0.3085g and 0.1186g respectively for these mixtures, in almost exact agreement with the experimental values. The response surface contour plot for this model is presented in Figure 4c where one finds that a 50%-50% ethanol–dichloromethane mixture results in a predicted maximum value for this four component mixture system.

A special cubic model gives the extracted mass of the basic fraction

$$\begin{split} \boldsymbol{B} &= 0.0103e + 0.0050d + 0.0036h + 0.0036a + \boldsymbol{0.0852}ed + 0.0460eh + \boldsymbol{0.0616}ea + 0.0058dh \\ & (\pm 0.0040) \ (\pm 0.0040) \ (\pm 0.0040) \ (\pm 0.0195) \ (\pm 0.01225) \ (\pm$$

All the linear terms are not significantly different from zero so the significant ethanol–dichloromethane and ethanol–acetone synergic terms and the ethanol–dichloromethane–hexane ternary synergic term account for the extracted basic fraction masses. The largest basic fraction mass in Table 1, 0.0386g occurs for the 1:1:1 ethanol–dichloromethane–hexane mixture. The above model predicts that 0.0428g will be extracted by this mixture, as can be seen in the response surface contour plot in Figure 4d as well.

The last column in Table 1 contains the estimated residue values not extracted with the four fractions. A special cubic model,

$$\mathbf{R} = \mathbf{0.3222}e + 0.0174d - 0.0014h + 0.0257a + \mathbf{1.0666}ed + \mathbf{0.9362}eh + \mathbf{0.5120}ea + 0.1252dh \\ (\pm 0.0335) (\pm 0.0335) (\pm 0.0335) (\pm 0.0335) (\pm 0.1631) (\pm 0.1631) (\pm 0.1631) (\pm 0.1631) \\ - 0.0249da + 0.1574ha - \mathbf{1.4329}edh - \mathbf{2.3107}eda - \mathbf{4.1848}eha + \mathbf{1.9444}dha \\ (\pm 0.1631) (\pm 0.1631) (\pm 1.0237) (\pm 1.0237) (\pm 1.0237) (\pm 1.0237) \\ \end{array}$$

provides the best fit to the data.

Of all the four pure solvents tested here, ethanol is the most important to maximize the residues. It is the only solvent with a significant linear blending term and also participates in three significant synergic terms, with the other three solvents and in a significant antagonistic term with hexane and acetone. Analyzing Table 1 one can see that the highest values for residues are found for solutions containing ethanol. The largest residue mass, 0.4316g occurs for ethanol–dichloromethane mixture. The above model predicts that 0.4364g will be left as residue by this mixture, as can be seen in the response surface contour plot as a function of ethanol, dichloromethane and acetone in Figure 4e.



Figure 4: Response surface contour plot curves for the (a) fiber, (b) neutral, (c) organic and (d) basic fraction masses and the (e) residue values. The response surfaces are functions of (a) dichloromethane-hexane-acetone, (b) and (d) ethanol-dichloromethane-hexane and (c) and (e) ethanol-dichloromethane-acetone.

Principal component analysis [14,15] was performed for the 19 samples and 6 variables. The score graph for these two components can be seen in Figure 5a and the corresponding loading plot is given in Fig. 5b. The first component contains 58.65% of the variance and separates groups I and II from groups III and IV. The second component discriminates group IV from the other groups and contains 19.56% of the variance.



Figure 5: (a) PC graph for the two first components. This graph explains 78.21% of the total data variance and (b) Loadings graph for the first and second principal components.

The mixtures of group I on the left of the score graph, a ternary ethanol–dichloromethane–hexane mixture and a binary ethanol–dichloromethane mixture are most effective in extracting larger masses of crude and the basic, organic and residual values. This is confirmed in Table 1 where these solutions are seen to extract the two largest values of crude mass and of the organic and basic fraction masses. They also provide two of the three highest residue values. As shown earlier in Table 2 all these fractions are related by correlation coefficients that vary between 0.63 to 0.88. These observations are consistent with the similarities found for the response surfaces in figures 4c - 4e and with the large synergic effect between ethanol and dichloromethane for the crude extract.

The binary dichloromethane-hexane and ternary dichloromethane hexane-acetone mixtures in the upper portion of the score plot extract the largest masses for the fiber fraction that is characterized by a high loading on the second PC. On the other hand they extract very small quantities of the organic fraction that has a very low loading on the second PC. The extraction mixtures with low scores on both the first and second PCs, notably all the pure solvents as well as the binary dichloromethane-acetone and hexane-acetone mixtures, extract small quantities of fiber, organic and basic fraction extracts, small quantities of crude masses and have low residue values.

The mixtures in Group II, ternary ethanol-dichloromethane-acetone and ethanol-hexane-acetone as well as the quartenary mixture, are most efficient for extracting neutral fraction masses. Furthermore the dispersion of the

replicate points in Group II gives us a measure of the size of the experimental error compared with the overall dispersion in the score plot that measures the size of the variations owing to changes in the solvent proportions.

These results show some similarity with the results obtained from our previous work [9,10]. There, we analyzed the crude extracts by high performance liquid chromatography and the data were subjected to principal component and varimax analysis. Comparing the results in Figure 4 of this manuscript with Figures 3 in ref. 9 and Figure 1 in ref. 10 one finds that some of the extraction mixtures remain in same group. For example (dh) and (dha), (e) and (ea) as well as (d), (h) and (ha) remain in the same groups for all these studies.

Acknowledgements

The authors are grateful to the Conselho Nacional de Pesquisa (CNPq) and Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) for financial support that made the reported research possible. CNPq is acknowledged for providing a doctoral fellowship for P. K. S. and research fellowships to I. S. S. and R. E. B.

5 References

[1] Gurib-Fankin A., *Molecular Aspects of Medicine* 2006, 27, 1-93.

[2] Endt D. V., Kijne J. W., Memelink J, Phytochemistry 2002, 61, 107-114.

[3] Barbosa, P. S., Abreu S.A., Batista, E. F., Guilhon, M. S. P., Muller, A. H., Arruda, M. S. P., Santos, L. S., Arruda, A. C., Seco, R. S., *Biochemical Systematics and Ecology* 2007, *35*, 887-890.

[4] Fernández-García E., Imhof, M., Schlichtherle-Cerny, H., Bosset, J. O., Nuñez, M., *International Dairy Journal* 2008, *18*, 147-157.

[5] Lendl A., Werner I., Glasl, S., Kletter, C., Mucaji, P., Presser, A., Reznicek. G., Jurenitsch, J., Taylor, D. W., *Phytochemistry* 2005, *66*, 2381-2387.

- [6] Simoneit, B. R. T., Otto, A., Wilde, V., Organic Geochemistry 2003, 34,121-129.
- [7] Otto, A., Simoneit, B. R. T., *Geochimica et Cosmochimica Acta* 2001, *6*, 3505-3527.
- [8] Harborne, J. B., *Modern Techniques of Plant Analysis*, Chapman and Hall, London 1984.
- [9] Soares, P. K., Bruns, R. E., Scarminio, I. S., J. Sep. Sci. 2007, 30, 3302-3310.
- [10] Soares, P. K., Bruns, R. E., Scarminio, *Anal. Chim. Acta* 2008, *613*, 48-55.

[11] Cornell J. A., *Experiments with Mixtures: Designs, Models, and Analysis of Mixture Data*, Second Ed., Wiley, New York 1990.

- [12] Bruns R. E., Scarminio I. S., Neto B. B., *Statistical Design Chemometrics*, Elsevier, Amsterdam 2006.
- [13] Statistica for Windows 6.0, **Statsoft**, Inc., Tulsa, OK, USA, 1999.
- [14] Wold S., Albano C., Dunn III W. J., Edlund U., ESbenson K., Geladi P., Hellberg S., Johansson E., Lindberg W.,

Sjöstrom M., Kowalski B. R. (Ed.) Chemometrics, Mathematics and Statistics in Chemistry, Reidel, Dordrecht 1984.

[15] Mardia K. V., Kent, J. T., Bibby, J. M., *Multivariate Analysis*, Academic Press, London 1979.

Reproduzido com a permissão de Phytochemical Analysis 2008, 19, 78-85 © Copyright 2007 John Wiley & Sons, Ltda



Multivariate Chromatographic Fingerprint Preparation and Authentication of Plant Material from the Genus *Bauhinia*

PATRICIA KAORI SOARES and IEDA SPACINO SCARMINIO*

Departamento de Química, Universidade Estadual de Londrina, Londrina, 86051-990 PR, Brazil

Received 17 August 2007; Revised 28 May 2007; Accepted 15 June 2007

Abstract: Multivariate analysis and statistical mixture designs were used for chromatographic fingerprint preparation and authentication of the plant material of three species of the genus *Bauhinia*. The extracts were analysed by reversed-phase high-performance liquid chromatography. Mixture design gave an optimum solvent composition for extracting components from the plants of 36% dichloromethane, 17% ethanol and 47% ethyl acetate (by volume), while an optimum mobile phase for chromatographic analyses was found to be 27% methanol, 27% acetonitrile and 46% of water (by volume). Results from principal component analysis, hierarchical analysis and soft independent modelling by class analogy showed that *Bauhinia candicans* cannot be synonymous with *B. forficata* Link. It was also possible to trace the metabolic profile without identifying its chemical constituents and to determine a chromatographic discriminating region. The characteristics responsible for discrimination between *B. candicans* and *B. forficata* were more polar substances that presented peaks with retention times around 1.65 and 1.81 min. Copyright © 2007 John Wiley & Sons, Ltd.

Keywords: Multivariate methodology; mixture design; fingerprint; authentication; Bauhinia forficata; Bauhinia candicans.

INTRODUCTION

There is an increasing interest in the efficacy of many herbal medicines that have been used for centuries as natural remedies to treat a variety of ailments. Owing to the significant expansion in the use of herbal medicines, it is critical to develop a high standard of quality control to guarantee their identity, consistency and authenticity. A herbal medicine may consist of hundreds of complex phytochemicals, so it becomes difficult or impossible in most cases to identify the compounds that are most biologically active (Gong *et al.*, 2003), and to separate them from large amounts of proteins, sugars or tannins that do not contribute to the pharmaceutical effect.

In general, one or two markers or pharmacologically active components in herbs and or herbal mixtures are currently employed for evaluating the quality and authenticity of herbal medicines (Liang *et al.*, 2004). Even though this is an acceptable method for quality control, the presence of the chemical markers does not always guarantee that the product contains the actual herb stated on the label. Multiple constituents should be considered, and not only a few marker components, to evaluate the quality of herbal medicines.

Copyright © 2007 John Wiley & Sons, Ltd.

Recently the chromatographic fingerprint technique was introduced as a useful approach for the evaluation and quality control of multicomponent herbal materials and their finished products (Defernez and Colquhoun, 2003; Gong *et al.*, 2004; Ji *et al.*, 2005; Zhao *et al.*, 2005). As distinct from a conventional study, a chromatographic fingerprint for herbal medicine analysis is a chromatogram that represents the chemical characteristics of the herbal medicine (Li *et al.*, 2005).

Among all the quality control systems, fingerprints have gained the most attention due to their ability to identify a particular herb and, moreover, to distinguish it from closely related species (Drašar and Moravcov, 2004). Of central importance with respect to quality control is the correct identification of the species concerned, whether in the fresh, dried or powered state. Consequently, careful attention must be given to the establishment of identity standards and quality control profiles using a combination of techniques.

In this paper, a species of *Bauhinia* serves as an example of the approach to assessing quality control based on the chromatographic fingerprint concept. The genus *Bauhinia* belongs to the Caesalpiniaceae (formally Leguminosae) family and comprises about 500 species of shrubs, and small trees that grow 5–9 m tall. The leaves are 7–10 cm long and shaped like the hoof of a cow, a characteristic that is distinctive to the *Bauhinia* genus. Indeed, the Brazilian name, "pata de vaca", translates to "cow's foot". It can be found in rainforests and tropical parts of Peru and Brazil, as well as in tropical zones of Asia, eastern Paraguay and northeastern Argentina. It produces large, drooping



^{*} Correspondence to: I. S. Scarminio, Departamento de Química, Laboratório de Quimiometria em Ciências Naturais, Universidade Estadual de Londrina, Londrina, CP6001, CEP 86051-990, Paraná, Brazil.

E-mail: ieda@qui.uel.br

Contract/grant sponsor: Conselho Nacional de Pesquisa/Brasil; Contract/grant number: CNPq-472791/4 and CNPq-304008/2005-2.

flowers, with the narrow petals arranged to closely resemble an orchid and a brown seed pod resembling that of mimosa.

Bauhinia candicans is the Bauhinia species most used as an anti-diabetic herbal remedy in Brazil, although in the literature it is cited as *B. forficata* (Pepato *et al.*, 2002). Leaves and tea bags are readily purchased in pharmacies and supermarkets. Along with other species of *Bauhinia*, it is used in Chilean folk medicine for the treatment of diabetes mellitus (Montes and Wilkomirsky, 1985). *B. variegata* is used as an astringent and tonic and is useful for treating scrofula, skin diseases and ulcers (Peddy *et al.*, 2003). *B. guianensis* is employed for renal problems, respiratory syndrome and diarrhoea, and the tea of the root is used for amoeba treatment (Viana *et al.*, 1999).

Differentiation of the genus species is difficult to perform using classical methods. For example, in the literature certain authors consider B. forficata synonymous with B. candicans. This discrimination is important owing to the large number of people consuming "pata de vaca" and the similarity between species that could exert quite different effects on people's health. Therefore, the development of analytical strategies to obtain information about the authenticity of this plant material is crucial in order to ensure the quality, safety and efficacy of the raw material before it is converted into the final product. In many instances, consulting a taxonomic specialist for plant identification is highly desirable. However, taxonomic advice often does not guarantee unambiguity with regard to the investigated material. On the other hand, the use of chemical analysis has been restricted to the development of specialised methods to isolate specific compounds (Perry et al., 1997; Santos and Salatino, 2000; Nøbæk et al., 2002).

There are very few examples in the literature in which mixture design has been employed in order to develop a method providing the highest possible extraction efficiency for as many different classes of compounds. The objective of the present work was to develop a process for chromatographic fingerprint preparation and authentication of the plant material of three Bauhinia genus species based on the chromatographic fingerprint concept using multivariate analysis methods. The multivariate analysis was based on two fundamental principles: (1) the use of statistical mixture design for the establishment of optimal conditions of solvent extraction of chemical substances from Bauhinia and for optimal separation conditions for mobile phase chromatographic analysis; and (2) the use of pattern recognition and classification methods to search for hidden structures in the data set in order to describe them more completely. Phytochemical profiles obtained by chromatographic fingerprint concepts of individual species may help resolve some of the taxonomic confusion which exists for this genus.

EXPERIMENTAL

General experimental procedures

All reagents used were of analytical or HPLC grade. Mixture preparations were made using water prepared with the Millipore (São Paulo, Brazil) Milli-Q purification system.

Plant material. Leaves of six trees cultivated in the Universidade Estadual de Londrina (UEL), in Londrina, PR, Brazil were used in this work. The plants were identified by Dr. Manuel R. C. Paiva and voucher specimens of the species were deposited in the herbarium of the Universidade Estadual de Londrina. A summary of the features as well as the herbarium access number of each plant is given in Table 1. Samples of *Bauhinia* candicans were taken from commercial packages containing 15 g of tea leaves and having labels clearly identifying their contents.

Collection and drying. Collections were performed in 2005. Leaves of the plants were harvested, avoiding material damaged by insects or fungus. Drying was carried out at room temperature for 9 days, with the samples protected from light, humidity and insect and fungus attacks.

Extraction procedure. A simplex-centroid design with three axial mixture points was used to investigate the influence of the different solvent proportions on extract preparation, Fig. 1(A) (Bruns *et al.*, 2006). The solvent systems were divided into three groups: dichloromethane **1**, ethanol **2** and ethyl acetate **3**. The selection of each group was made considering Snyder's solvent selectivity triangle, since solvents from different groups in the triangle have different selectivity characteristics (Snyder and Kirkland, 1979). First, each extract was prepared by addition of 60.00 mL of solvent [according to the compositions specified in the experimental design, Fig. 1(A)], to 3.00 g of dried leaves and the mixture was left for 24 h. The extracts were filtered and concentrated in a rotatory evaporator at

Table 1 Description of the species of Bauhinia studied andtheir access number in the Herbarium of the UniversidadeEstadual de Londrina, Londrina, Paraná, Brazil

Access number	Flower colour		
36,475	White		
36,486	White		
37,339	White		
36,474	White		
36,473	Rose		
36,472	Purple		
	Access number 36,475 36,486 37,339 36,474 36,473 36,472		



Figure 1 (A) Simplex-centroid design with three axial mixture points used to investigate the influence of different solvent proportions on extract preparation and (B) simplex-centroid design with three axial mixture points used to optimise mobile phase composition.

a temperature of less than $65 \pm 2^{\circ}$ C. The extraction procedure was repeated four more times for each sample, and the extracts obtained from the five extractions were combined. The extracts were dried at room temperature until constant weight was attained. This procedure was also used for extract preparation under optimised conditions.

HPLC analysis. The experimental design chosen to optimise the mobile phase was the simplex-centroid design with three axial mixture points, illustrated in Fig. 1(B). Components 1 and 2 refer to pure methanol and acetonitrile, whilst 3 refers to a mixture of methanol:acetonitrile:water of composition 15:15:70 (v/v/v). Analyses were carried out using a Shimadzu (São Paulo, Brazil) model LC-10 AD liquid chromatograph equipped with an SPD-M10AVP diode array detector, and a Metachem (São Paulo, Brazil) C₁₈ ODSPN 0380 column ($250 \times 4.6 \text{ mm id.}; 5 \mu \text{m}$). Elution was at a flow-rate of 1 mL/min and the eluent was monitored at 210, 220, 230, 240, 250 and 254 nm. The temperature was fixed at 50°C. All of the prepared solutions were filtered twice using a $0.22\,\mu m$ Millipore filter. Each sample, prepared according to the experimental design shown in Fig. 1(B), was analysed with 10 mobile phases, Fig. 1(A). Each sample was prepared by weighing 30.0 mg of concentrated extract and re-dissolving it in 3.00 mL of the solvent used in extraction, storing for 1 h and filtering through Whatman no. 1 (São Paulo, Brazil) filter paper. A sample (50 µL) of this extract was diluted to 1.00 mL with the mobile phase. An aliquot (20 μ L) of this diluted solution was injected into the HPLC.

Chemometric methods

Pre-processing is a very important part of any chemometrics data analysis project. Available methods

Copyright © 2007 John Wiley & Sons, Ltd.

are divided into two basic types depending on whether they operate on samples or variables (height values of chromatographic peaks, absorbance, pH, concentrations, etc.; Beebe *et al.*, 1998). Mean centring is a common pre-processing tool that is applied to account for an intercept in the data. Mean centring a variable is accomplished by subtracting the mean of that variable vector from all of its elements. Variables can be scaled to variance, where each element in a variable vector is divided by the standard deviation of that variable. Autoscaling is the application of both variance scaling and mean centring. In this case, each data vector is scaled to a mean zero and unit variance.

Normalisation is a common pre-processing method that operates on each sample. Normalisation of a sample vector is accomplished by dividing each variable by a constant. Normalisation to unit area can be accomplished by dividing each element in the vector by the sum of all absolute values in the vector. Alternatively, normalisation to unit length can be accomplished by dividing each element in the vector by the square root of the sum of all the square values in the vector.

Four chemometric methods were used in this study to search for hidden structures in the data set. A large literature exists on each of the methods used (Jöreskog *et al.*, 1976; Davis, 1986; Beebe *et al.*, 1998; Brereton, 2004) and hence this section will provide only a summary of each technique. There are several advantages that result from application of chemometric methods for data interpretation. The methods can rapidly identify key variables that are important in making distinctions among samples. Another advantage of these techniques is that they are multivariate so they can be used to investigate many variables simultaneously.

Hierarchical analysis (HA). Unsupervised clustering techniques have been recognised as basic tools in data mining for knowledge discovery. Generally, the aim of clustering techniques is to group objects having

some natural relation to one another in a single group whereas objects in different groups are somewhat different from one another (Daszykowski et al., 2001). Hierarchical approaches decompose the database of a number of objects into several levels of nested clusters represented by a dendrogram. These results may be compared with the results obtained from the principal component plots. In general, the results of PCA and HA are complementary. There are many clustering techniques that may be employed in data interpretation. For brevity we shall illustrate only one here. The starting point is to calculate the distances between all pairs of points in the data set to define "similarity". Initially, each sample is considered as a single cluster. The next step is to find the shortest distance between any pair of points. These two points are then considered the most similar and are combined into a single group. The distance from this group to all other points is then calculated using some type of "average point" to represent the centre of the two point locations. The second shortest distance is then found (whether it is between this group and a point or two other points). This pair is considered to be the second most similar and is combined to form another group. This process is continued until the entire data set is included in a single group (Wolff and Parson, 1983).

Principal component analysis (PCA). PCA is probably the most widespread multivariate chemometric technique that is used and, owing to its importance for multivariate measurements in chemistry, it is regarded by many as the technique that most significantly changed the chemist's view of data analysis (Brereton, 2004). PCA is a multivariate statistical technique used for data reduction and for deciphering patterns within large sets of data. It reduces the dimensionality of the data by producing new linear combinations of the original variables while retaining the most significant information. It is performed without knowledge of sample classes. It is possible to plot the transformed data in order to indicate relationships between samples in multidimensional space (Bailey et al., 2003). PCA decomposes the matrix A into a product of two matrices,

$$\mathbf{X} = \mathbf{T}_{n \times q} \, \mathbf{P}_{q \times p}^{t} + \mathbf{E}_{n \times p} \tag{1}$$

where **T** is a score matrix, **P** a loading matrix and **E** a matrix of residuals. The number of factors describing the major part of the data variance is represented by q, and **P**^t is the transpose of **P** (Brereton, 2004).

Soft independent modelling by class analogy (SIMCA).

Classification is essentially determined by identification rules that are estimated by training on sets of objects for which the classes are known. SIMCA is a classification method for samples based on their similarities to principal component models (Lonni *et al.*, 2003). A principal component is fitted separately to each class in the training set and multivariate confidence envelopes are constructed around each class model containing the data points. The method is used to classify samples into predefined categories, to provide outlier detection, and to predict external properties. It is possible to use SIMCA to determine whether a sample fits into a predetermined class, or whether it does not fit into any of the classes or if indeed it fits into more than one class (Wold, 1976). The classification of test samples is performed by adjusting each of these samples to each model, and deciding, at the 95% confidence level, to which class each sample belongs.

Variance weights. These weights determine the individual importance of each variable or characteristic in the discrimination between each pair of classes (Lonni *et al.*, 2003). Feature weighting is of particular utility to the chemist in revealing whether or not a given list of measurements actually contains information necessary to solve a problem, and, if so, the minimum subset of the list that will do the job (Sharaf *et al.*, 1986).

Computer programs. The computer program, ARTHUR, modified for microcomputers (Scarminio and Bruns, 1989) was used for the SIMCA and weight calculations. For the PCA and pre-processing calculations, FORTRAN programs developed in our laboratory were used on PC computers.

RESULTS AND DISCUSSION

In this work, the mobile phase was chosen using TLC, since it is a simple and fast technique. Although pure water gave the best separation, it required an analysis time of about 2 h. The TLC results showed that a 15:15:70 (v/v/v) mixture of methanol:acetonitrile:water was adequate to analyse the extracts and the analysis time was drastically reduced. Therefore, this mixture composition was included in our investigation as the third corner of the triangle, Fig. 1(B). The choice of phase composition involves a compromise between efficiency and analysis time. In this way the mobile phase chosen was 27:27:46 (v/v/v) of methanol:acetonitrile: water [mobile phase number 10, Fig. 1(A)].

A larger number of peaks permits a better description of the system by furnishing more detailed fingerprints for botanical characterisation than any other response, for example, analysis time. Therefore, the number of peaks in the chromatogram observed for each extract of the mixture design was used as a response and was monitored at 210 nm. The most precise model for the number of peaks using 27:27:46 (v/v/v) of methanol:acetonitrile:water as mobile phase



Figure 2 Response surface of the quadratic model for the choice of the optimum extraction solvent employing, as criteria, the number of chromatographic peaks. The asterisk indicates the optimum mixture.

was the quadratic model. This model showed good fit at the 95% confidence level. The regression equation obtained was

$$\hat{y} = \mathbf{12.61}dc + \mathbf{9.89}et + \mathbf{12.79}ea + 18.70(dc)(et)$$

$$(\pm 1.50) \quad (\pm 1.50) \quad (\pm 1.50) \quad (\pm 6.92)$$

$$+ 16.52(dc)(ea) + 11.07(et)(ea) \quad (2)$$

$$(\pm 6.92) \quad (\pm 6.92)$$

where \hat{y} is the predicted number of peaks and dc, *et* and *ea* are the proportions of dichloromethane, ethanol and ethyl acetate, respectively. Standard error estimates are presented in parentheses directly below their corresponding model coefficient. Bold coefficients indicate those that are significant at the 95% confidence level.

The number of peaks increased with additions of the three solvents to the process extraction mixture. This may be confirmed by inspection of the response surface contour plot in Fig. 2, predicting the existence of 17 peaks. This model predicts an optimum solvent extraction composition of dichloromethane:ethanol:ethyl acetate of 36:17:47 (v/v/v) for *B. forficata* (36,475), Fig. 2.

Ten samples of each plant listed in Table 1 were prepared using 36:17:47 (v/v/v) of dichloromethane: ethanol:ethyl acetate as extracting solvent. The time for each chromatographic analysis was 30 min.

The data set contained 70 samples. Each sample was represented by 2813 height values of chromatographic peaks, resulting in a 70×2813 matrix. In this application four kinds of pre-processing techniques were

tested: normalisation to unit area, normalisation to the unitary vector, autoscaling and area normalisation followed by autoscaling. The best results were obtained using the unitary length vector. This matrix was subjected to a PCA. These results suggested that the characteristics responsible for the discrimination between the species were the more polar substances that presented peaks with retention times of up to 3.62 min. A new matrix was constructed with 70 × 326 dimensions and subjected to PCA, SIMCA and HA.

The three most important factors contained 97% of the total variance. Figure 3 shows the score graph of



Figure 3 PC1 vs PC2 score graph, which accounts for 94% of the total data variance of the seven plants obtained from 70 extracts prepared under otimum conditions (F, *B. forficata*; V, *B. variegata*; and C, *B. candicans*).



Figure 4 Graph of PC1 and PC2 loadings.

PC1 vs PC2, with 94% of the total variance. In this plot, a clear distinction exists between four groups, where F, V and C stand for B. forficata, B. variegata and B. candicans, respectively. Group I was composed of B. candicans, group II of B. forficata (access number 36,486), group III of B. variegata (36,473) and group IV of B. forficata (36,475), B. forficata (37,339), B. variegata (36,472) and B. variegata (36,474). PC2 discriminates groups I and II from III and IV. The scores suggest that B. candicans is distinctly different from the other species, and hence it cannot be synonymous with B. forficata. Figure 4 shows a graph of the loading values on PC1 and PC2 vs retention time. Positive values are situated between 1.55 and 1.96 min. In Fig. 5, it can be verified that, in the 1.55 and 1.96 intervals, the most intense analytical signals for B. candicans are larger than those of B. forficata (36,486) and *B. variegata* (36,473).

A similar situation also exists for *B. forficata* (36,486), with negative values situated between 2.19 and 2.64, whereas *B. forficata* (36,475) has a much lower score on PC2 The results confirm the difficulty of



Figure 5 Original chromatograms of *B. candicans*, *B. forficata* (36,486) and *B. variegata* (36,473) in the interval between 0.16 and 3.62 min. Composition of the extraction solvent was 36:17:47 (v/v/v) dichloromethane:ethanol:ethyl acetate whilst the mobile phase was 27:27:46 (v/v/v) mixture of methanol:acetonitrile:water. Group I is composed of *B. candicans*; group II of *B. forficata* (36,486); group III of *B. variegata* (36,473); and group IV of *B. forficata* (36,475), *B. forficata* (36,474).

classifying species based on conventional taxonomic techniques, since all three *B. forficata* were classified as different species.

These results suggest that the characteristics responsible for the discrimination between *B. candicans* and *B. forficata* (36,486) are the more polar substances that present peaks with retention times around 1.65 and 1.81 min.

Hierarchical analysis was performed in an attempt to define groups and compare them with those obtained from PCA. The Euclidean distance was chosen as the measure of similarity. Four clusters were obtained after hierarchical classification, Fig. 6. These results agree with those obtained with PCA. On the other hand, in group IV, the samples of *B. forficata* (37,339) do not

Plants	Number of samples	F (36,475)	F (36,486)	F (37,339)	V (36,474)	V (36,473)	V (36,472)	С
F (36475)	10	9	_	_	1	_	_	
F (36486)	10	_	10	_	_	_	_	_
F (37339)	10	_		10	_	_	_	_
V (36474)	10	_	_		5	_	5	_
V (36473)	10	_	_		_	10	_	_
V (36472)	10	_	_		_	_	10	_
С	10	_	_	—		_	_	10

Table 2 Results of the SIMCA classification method for the normalised chromatographic data set^a

^aDiagonal elements give the number of correct classifications whereas the off-diagonal elements give the number of incorrect classifications. F, V and C stand for *B. forficata*, *B. variegata* and *B. candicans*, respectively.



Figure 6 Dendrogram based on the chromatographic analysis of the 70 extracts prepared under optimum conditions.



Figure 7 PC2 vs PC3 score graph of the seven plants obtained from the 70 extracts prepared under optimum conditions.

overlap with samples of the others groups. This result can be confirmed in Fig. 7, where the samples of *B. forficata* (37,339) are seen to be clearly separated. From these data it is evident that there exist differences in secondary metabolite compositions of the three plants classified as *B. forficata*.

The SIMCA classification method was employed to confirm the qualitative discrimination of the species. Seven PCA models were constructed, one for each class. The number of PCs used in the modelling was determined by cross validation. The results are presented in Table 2. The SIMCA result confirmed that the region between 0.16 and 3.62 differentiates the five species, correctly classifying all the F (36,486), F (37,339), V (36,473), V (36,473) and V (36,472) species. For F (36,475) one sample was classified incorrectly, whereas for V (36,474), five samples were classified incorrectly. These results agree with those obtained with PCA and HA, where samples of V (36,474), overlap with samples of the group V (36,472).

Acknowledgements

The authors wish to express their gratitude to Dr. Roy E. Bruns for invaluable suggestions, to Jurandir P. Pinto for the assistance given in HPLC, to CNPq for the financial support of this work and to CAPES for a fellowship to Patricia K. Soares. We also thank Dr. Manuel R. C. Paiva for the identification of the botanical material.

REFERENCES

- Bailey NJC, Oven M, Holmes E, Nicholson J K, Zenk M H. 2003. Metabolomic analysis of the consequences of cadmium exposure in *Silene cucubalus* cell cultures via ¹H NMR spectroscopy and chemometrics. *Phytochemistry* 62: 851–858.
- Beebe KR, Pell RJ, Seasholtz MB. 1998. Chemometrics: a Practical Guide. Wiley: New York; 1–180.

- Brereton RG. 2004. Chemometrics: Data Analysis for the Laboratory and Chemical Plant. Wiley: Chichester; 184–198.
- Bruns RE, Scarminio IS, Barros BB. 2006. Statistical Design Chemometrics, Vol. 25. Elsevier: Amsterdam; 313–340.
- Daszykowski M, Walczak B, Massart DL. 2001. Looking for natural patterns in data: Part 1. Density-based approach. *Chemom Intell* Lab Syst 56: 83–92.
- Davis JC. 1986. Statistics: Data Analysis in Geology. Wiley: New York; 456–536.
- Defernez M, Colquhoun IJ. 2003. Factors affecting the robustness of metabolite fingerprinting using ¹H NMR spectra. *Phytochemistry* 62: 1009–1017.
- Drašar P, Moravcov J. 2004. Recent advances in analysis of Chinese medical plants and traditional medicines. J Chromatogr B 812: 3– 21.
- Gong F, Liang Y-Z, Xie P-S, Chau F-T. 2003. Information theory applied to chromatographic fingerprint of herbal medicine for quality control. J Chromatogr A 1002: 25–40.
- Gong F, Liang YZ, Fung YS, Chau FJ. 2004. Correction of retention time shifts for chromatographic fingerprints of herbal medicines. *J Chromatogr A* **1029**: 173–183.
- Ji YB, Xu QS, Hu YZ, Heyden YV. 2005. Development, optimisation and validation of a fingerprint of *Ginkgo biloba* extracts by highperformance liquid chromatography. J Chromatogr A 1066: 97– 104.
- Jöreskog KG, Klovan JE, Reyment RA. 1976. Geologycal Factor Analysis. Elsevier: New York; 178.
- Li W, Chen Z, Liao Y, Biu H. 2005. Separation methods for toxic components in traditional Chinese medicines. *Anal Sci* **21**: 1019–1030.
- Liang YZ, Xie P, Chan K. 2004. Quality control of herbal medicines. *J Chromatogr B* **812**: 53–70.
- Lonni AASG, Scarminio IS, Silva LMC, Trevisan DT. 2003. Differentiation of species of the *Baccharis* genus by HPLC and chemometric methods. *Anal Sci* 19: 1013–1018.

- Montes M, Wilkomirsky T. 1985. *Traditional Chilena Concepción*. Chile: University of Concepción Press.
- Nøbæk R, Brandt K, Nielsen JK, Ørgaard M, Jacobsen N. 2002. Flower pigment composition of *Crocus* species and cultivars used for a chemotaxonomic investigation. *Biochem Syst Ecol* 30: 763–791.
- Peddy MVB, Peddy MK, Gunasekar D, Caux C, Bodo B. 2003. A flavanone and a dihydrodibenzoxepin from *Bauhinia variegata*. *Phytochemistry* 64: 879–882.
- Pepato MT, Keller EH, Baviera AM, Kettelhut IC, Vendramini RC, Brunetti IL. 2002. Anti-diabetic activity of Bauhinia forficata decoction in streptozotocin–diabetic rats. J Ethnopharmac 81: 191–197.
- Perry NB, Vanklink JW, Brennan NJ, Harris W, Anderson RE, Douglas MH, Smallfield BM. 1997. Essential oils from new zealand manuka and kanuka: chemotaxonomy of *Kunzea. Phytochemistry* 45: 1605–1612.
- Santos DYAC, Salatino MLF. 2000. Foliar flavonoids of Annonaceae from Brazil: taxonomic significance. *Phytochemistry* 55: 567–573.
- Scarminio IS, Bruns RE. 1989. An adaptation of ARTHUR for microcomputers. Trends Anal Chem 8: 326–327.
- Sharaf MA, Illman DL, Kowalski BR. 1986. Chemometrics. Wiley: New York; 195–196.
- Snyder LR, Kirkland JJ. 1979. Introduction to Modern Liquid Chromatography. Wiley: New York; 261–265.
- Viana EP, Santa-Rosa RS, Almeida SSMS, Santos LS. 1999. Constituents of the stem bark of *Bauhinia guianensis*. Fitoterapia 70: 111–112.
- Wold S. 1976. Pattern recognition by means of disjoint principal components models. *Pattern Recogn* 8: 127–139.
- Wolff DD, Parsons ML. 1983. Pattern Recognition Approach to Data Interpretation. Plenum Press: New York; 71.
- Zhao L, Huang C, Shan Z, Xiang B, Mei L. 2005. Fingerprint analysis of *Psoralea corylifolia* L by HPLC and LC–MS. J Chromatogr B 821: 67–74.



Reproduzido com a permissão de Analytica Chimica Acta 2008, 613, 48-55 © Copyright 2008 Elsevier



Statistical mixture design—Varimax factor optimization for selective compound extraction from plant material

Patrícia Kaori Soares^a, Roy E. Bruns^{a,*}, Ieda S. Scarminio^b

^a Instituto de Química, Universidade Estadual de Campinas, CP 6154,
 13083-970 Campinas, SP, Brazil
 ^b Laboratório de Químiometria em Ciências Naturais, Departamento de Química, Universidade Estadual de Londrina,
 Caixa Postal 6001, 86051-990 Londrina, PR, Brazil

ARTICLE INFO

Article history: Received 21 November 2007 Received in revised form 30 January 2008 Accepted 27 February 2008 Published on line 4 March 2008

Keywords: Varimax rotation Principal component analysis Mixture designs Erythrina speciosa Andrews leaves Solvent extraction

ABSTRACT

Varimax-transformed chromatograms of compounds extracted from Erythrina speciosa Andrews leaves by simplex centroid design mixtures of dichloromethane, hexane, ethanol and acetone are reported and compared with principal component results. Six varimax factors were investigated focusing on the three main groups of extracted compounds with retention times of 1.7, 3.1 and 6.6 min. Varimax models provide chromatographic loading profiles that are simpler than their principal component counterparts. Furthermore varimax score models in terms of extraction medium compositions are simpler to interpret. The first varimax model results in an optimum extraction mixture of 71% dichloromethane–29% acetone although substitution of acetone by ethanol results in an almost identical extraction profile. The second varimax score model predicts optimum extraction binary mixtures with hexane proportions ranging from 50 to 100% for complementary proportions of either acetone or ethanol. The third varimax factor provides a response surface that is very similar to the one found for the third PC. Confirmatory experiments were performed to validate the model predictions.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

A great deal of effort is currently being devoted to systematically characterizing and quantifying metabolites in plants [1,2]. Chromatographic fingerprints have been considered as one of the most rational and powerful approaches for quality evaluation of herbal medicines. In crude extracts however, there are many peaks, leading to overlapped signals with complex profiles that are sometimes interpreted as 'fingerprints'. Fingerprinting generally requires chemometric interpretation of the complexity resulting from simultaneous acquisition of the analytical data on hundreds of metabolites. A possible approach is to look at the whole chromatographic fingerprint by using principal component analysis (PCA), in which an entire set of variables, is examined.

PCA achieves a reduction of dimensionality while finding out how samples are different from one another and which variables contribute the most to these differences [3]. Principal component methods are also used for selecting subsets of variables for modeling by regression methods. One such application is to perform a varimax rotation of the principal components, which are then used as predictors in the regression. Varimax rotation ensures that each variable is maximally correlated with only one principal component while having a near-zero association with the other components [4,5].

^{*} Corresponding author. Tel.: +55 19 37883106; fax: +55 19 37883023. E-mail address: bruns@iqm.unicamp.br (R.E. Bruns).

^{0003-2670/\$ –} see front matter © 2008 Elsevier B.V. All rights reserved. doi:10.1016/j.aca.2008.02.051

Principal component and varimax analyses have often been used to analyse infrared, Raman and UV-vis spectra in order to perform a variety of applications such as multivariate calibration and curve resolution [6–11]. However their combined use in liquid chromatography has been limited so far to multivariate curve resolution [9].

Recently our group has used PCA to transform chromatograms of extracts from Erythrina speciosa Andrews leaves obtained using solvent mixtures corresponding to a four component simplex centroid design [12]. The purpose of this investigation is to determine the characteristics of varimaxrotated transformed chromatograms in relation to these same chromatograms preprocessed using only principal components. This seems to be a necessary preliminary step in systematically characterizing and quantifying metabolites in plants using a chromatographic fingerprint. Varimax factors tend to concentrate their information in relatively fewer loadings than principal components. So one can expect varimax models that lend themselves to simpler interpretations in terms of the extraction component proportions than models obtained using principal component scores. This could permit more efficient optimization of extraction mixtures for selective compound extraction from plant material when chromatographic variations are planned using statistical mixture designs. This might be especially relevant for varimax factors corresponding to small chromatographic variations for which principal component results are more complex.

2. Experimental

Erythrina speciosa Andrews leaves were collected in November 2006. A voucher specimen of this plant was deposited in the herbarium of the Londrina State University and authenticated by M.R.C. Paiva and registered under the FUEL 35133 number.

Drying was carried out for 9 days at room temperature with the leaves protected from light, since solar irradiation can alter the chemical compositions of the plants.

Extraction mediums were prepared using mixtures of four components, (1) ethanol, (2) dichloromethane, (3) hexane and (4) acetone whose proportions were varied according to the simplex centroid design [13] in Fig. 1 of our previous work [12]. The proportions of each solvent used in the extraction mixtures are specified in Table 1. Nineteen extractions were carried out with 15 different mixtures. A five-run replicate was performed at the center point so experimental error could be estimated, although for operational reasons replicate numbers 17–19 were performed 1 month after the others. Three confirmatory experiments testing the models predictions were executed 6 months later.

Each extract was prepared by weighing 2.000 g of dried leaves and adding 30 mL of the solvent mixtures in Table 1. The mixture volumes were measured using a 0.1-mL graduated cylinder and were placed in an ultrasonic bath (Unique, model USC 1400) for 30 min [14]. The bathwater was changed every 10 min to maintain constant water temperature. The extracts were filtered in cotton, only to separate the solution from small pieces of leaves.

HPLC analysis was conducted on a Shimadzu Model LC-10 AD liquid chromatograph equipped with an SPD-



Fig. 1 – Score graph for the first three varimax factors. Group I contains extracts 2 and 9, group II contains extracts 8, 11, 14–19, group III contains 5 and 12, group IV contains 3, 6, 10 and 13 and group V contains 1, 4 and 7.

M10AVP diode array detector. A C18 ODS Metasil column (250 mm × 4.6 mm) 5 μ m particle size, was used for separation throughout this study, with a flow rate of 1 mL min⁻¹. All the prepared solutions were filtered twice through a 0.22- μ m Millipore Millex filter. The temperature was fixed at 50 °C. Elution was monitored with a 210 nm detector for retention times up to 20 min. This wavelength provides more information than the 208, 220, 230 and 254 wavelengths that were tested in an exploratory analysis. Samples (50 μ L) of extracts were diluted in 950 μ L of a 80:20 (v/v) methanol–water mobile phase. A 20- μ L aliquot of this diluted solution was injected into the HPLC. Elution was performed isocratically. All the reagents used were purchased from VETEC and were analytical or HPLC grade.

3. Statistical methods

Principal component-based methods are commonly used to reduce the number of variables in chromatography and spectroscopy problems so their full potentials can be exploited [15]. To obtain knowledge about the properties of a system, it is desirable to try to interpret the significance of the principal components by analysing their loadings. The task of interpreting the significance of principal components is not always straightforward, because they are complex linear combinations of the original variables. Varimax rotation perturbs the principal components so as to maximize the variance within each vector [16,17]. As a result, the number of variables with intermediate loadings is decreased for each vector, and the number of either very large (absolute magnitude) or very small loadings is increased. In other words, varimax rotation simplifies the principal components so they essentially depend on

Table 1 – Varimax score results for 19 chromatograms of 19 extraction mixtures for the simplex centroid design involving the ethanol, dichloromethane, hexane and acetone solvents

	Ethanol	Dichloromethane	Hexane	Acetone	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	1	0	0	0	0.052	0.137	0.265	0.055	0.195	0.929
2	0	1	0	0	0.944	-0.043	-0.059	-0.022	0.109	0.030
3	0	0	1	0	0.016	0.988	-0.096	-0.020	0.058	0.029
4	0	0	0	1	0.066	0.121	0.133	0.961	0.183	0.055
5	0.5	0.5	0	0	0.876	0.005	0.311	0.008	0.316	0.138
6	0.5	0	0.5	0	0.057	0.939	0.236	0.018	0.176	0.133
7	0.5	0	0	0.5	0.057	0.079	0.874	0.176	0.251	0.338
8	0	0.5	0.5	0	0.726	0.649	-0.107	-0.038	0.172	-0.010
9	0	0.5	0	0.5	0.967	-0.016	-0.044	0.079	0.214	-0.015
10	0	0	0.5	0.5	0.033	0.982	-0.060	0.119	0.073	0.032
11	0.333	0.334	0.333	0	0.612	0.664	0.029	0.024	0.395	0.086
12	0.333	0.333	0	0.334	0.823	-0.016	0.289	0.084	0.450	0.047
13	0.333	0	0.333	0.333	0.036	0.903	0.226	0.135	0.272	0.082
14	0	0.334	0.333	0.333	0.774	0.597	-0.069	0.058	0.159	-0.002
15	0.25	0.25	0.25	0.25	0.680	0.623	0.047	0.101	0.341	0.026
16	0.25	0.25	0.25	0.25	0.669	0.625	0.013	0.099	0.359	0.018
17	0.25	0.25	0.25	0.25	0.407	0.338	0.059	0.135	0.820	0.119
18	0.25	0.25	0.25	0.25	0.390	0.225	0.126	0.137	0.857	0.117
19	0.25	0.25	0.25	0.25	0.398	0.260	0.341	0.129	0.759	0.199
Expla	ained variand	ce (%)			57.25	18.55	11.87	4.70	3.65	2.23
Cumulative variance (%)				57.25	75.80	87.67	92.36	96.01	98.24	
Score	as larger that	n 0.7 are given in held fa	co turno							

a smaller number of the original variables and makes interpretation easier [18]. Varimax rotations are often employed in environmental applications so the transformed components can be identified with factors corresponding to different kinds of sources assumed to be contributing to pollution problems [19–22].

Most PCAs are transformed using the varimax rotation although others are available. In this approach, pairs of factor loadings are rotated in the two-dimensional space formed by their two axes such as to maximize the sum of the variances of the squared loadings. The factors are systematically rotated in pairs until changes are negligible. This procedure has the effect of producing factor loadings tending to be as extreme as possible. The final quantity to be maximized for producing simple structure via the varimax procedure is

$$s_{v}^{2} = p \sum_{j=1}^{k} \sum_{i=1}^{p} \left(\frac{b_{ij}}{h_{i}}\right)^{4} - \sum_{j=1}^{k} \left(\sum_{i=1}^{p} \frac{b_{ji}^{2}}{h_{i}^{2}}\right)^{2},$$
(1)

where s^2 is the variance, *b* is the rotated factor loading, *p* represents the number of variables, *k* the number of PC factors and h_i^2 is the communality ($h_i = \sqrt{b_{i1}^2 + b_{i2}^2 + \dots + b_{ik}^2}$). For any pair of factors, *j* and *l*, the quantity to be maximized is

$$s_{v_{jl}}^{2} = p \left[\sum_{i=1}^{p} \left(\frac{b_{ij}}{h_{i}} \right)^{4} + \sum_{i=1}^{p} \left(\frac{b_{il}}{h_{i}} \right)^{4} \right] - \left[\sum_{i=1}^{p} \left(\frac{b_{ij}^{2}}{h_{i}^{2}} \right)^{2} + \sum_{i=1}^{p} \left(\frac{b_{il}^{2}}{h_{i}^{2}} \right)^{2} \right]$$
(2)

4. Results and discussion

Chromatographic peak heights were registered at 1407 equally spaced times during a 15-min interval. The chemometrics data matrix, \mathbf{X} , consisted of 19 rows and 1407 columns. Each row corresponds to a chromatogram of an extract in a medium specified by the simplex centroid mixture design and was normalized to unit area to remove variations owing to slightly different injection volumes. The first six principal components were subjected to a varimax rotation.

Varimax scores for the first six factors accounting for 98.2% of the total data variance have been included in Table 1. Fig. 1 shows a score graph of the first three varimax factors. The same five groups identified from the PCA [12] are clearly discriminated in this figure. The largest group, group II, contains all the points corresponding to the center point as well as points for the 8, 11 and 14 extraction mixtures. Group II has subgroupings with the 17-19 center points on the left separated from 15 to 16 center points on the right that are close to points 8, 11 and 14. This separation really owes to a blocking effect since the 17–19 center point extractions were performed 1 month after those for experiments 15 and 16. Besides changes in chromatographic conditions the plant material likely lost volatile material since the peaks for samples 17–19 were less intense than those for samples 15 and 16. The other four groups, I, III-V can all be discriminated from each other by one or more varimax factors as can be seen in Fig. 2.

Principal component and varimax loadings for the chromatographic retention times are illustrated in Fig. 2 for the first six factors. Except for the fifth component the varimax



Fig. 2 - Varimax (factor) and principal component (CP) loading graphs for the first six factors.

loadings exhibit simpler behaviors than the PCA loadings. The first varimax loading graph has one strong peak and several weak ones whereas the PCA loadings shows one strong, two medium and several weak peaks. The second varimax factor has one strong-loading peak whereas the corresponding principal component has two strong peaks, one positive and the other negative. The third, fourth and sixth factors also show simpler loading graphs for the varimax factors compared with those for the principal components.

The varimax analysis suggests that the extraction mixtures can be optimized to separate specific substances from plant material. Optimization of the first factor would result in an almost exclusive extraction of compounds having a 3.1min retention time. Only small amounts of other compounds will be co-extracted. Optimization of the second varimax factor results in an extraction of the 6.6 min retention time compounds accompanied by smaller amounts of some other compounds, especially those with 4.0 and 6.3 min retention times.

The third varimax factor has a strong positive-loading peak at a 1.7-min retention time with very minor peaks, most of them negative, at longer retention times. Maximization of this varimax factor would result in an almost selective extraction of the 1.7 min retention time compounds whereas its minimization would be expected to yield mostly the 3.1 and 6.7 min retention time group of compounds.

The fourth and sixth varimax factors have simpler loading profiles than their corresponding PCs although they are not promising for selective compound extraction. Optimization of the fourth varimax factor results in the extraction of the 1.6, 2.7 and 4.5 min groups of compounds. The sixth varimax factor has positive large peaks at 1.8, 3.4, 4.0, 6.7, 7.9 and 14.3 min retention times. Both the fifth varimax factor and PC have complex profiles. The varimax factor explains the data variance describing the separation of the 15–16 and 17–19 subgroups of the replicated center point as can be seen in the score graphs.

Linear, quadratic and special cubic models were applied to the first six varimax factors. Lack of fit and regression significance F-tests and t-tests for model coefficient significance were used to determine the preferred models. A quadratic model was chosen to represent the concentration dependence of the first varimax factor scores on the extraction solvent proportions. The quadratic model is given by

$$\begin{aligned} CP_{1}^{r} &= \underbrace{0.0638e}_{(\pm 0.1125)} + \underbrace{0.9445d}_{(\pm 0.1125)} + \underbrace{0.0253h}_{(\pm 0.1125)} + \underbrace{0.0670a}_{(\pm 0.1125)} + \underbrace{1.4786ed}_{(\pm 0.4756)} \\ &- \underbrace{0.1000eh}_{(\pm 0.4756)} - \underbrace{0.0458ea}_{(\pm 0.4756)} + \underbrace{0.9858dh}_{(\pm 0.4756)} + \underbrace{2.0045da}_{(\pm 0.4756)} - \underbrace{0.0329ha}_{(\pm 0.4756)} \end{aligned}$$

where *e*, *d*, *h* and *a* are the proportions of ethanol, dichloromethane, hexane and acetone, respectively. Standard error estimates are presented in parenthesis directly below their corresponding model coefficients. Bold-faced coefficients indicate those that are significant at the 95% confidence level. The ANOVA of the regression results is presented in Table 2. The mean square lack of fit/pure error ratio of 0.23 is much smaller than the $v_{lof} = 5$ and $v_{pe} = 4.95\%$ F-distribution value of 6.26 indicating no significant lack of fit. Furthermore the mean square regression/residual ratio of 18.51 is much greater than the $v_R = 9$ and $v_r = 9.95\%$ confidence F-distribution value of 3.18 showing the regression is highly significant.

The model in Eq. (3) for the varimax scores as a function of the extraction solvent proportions is simpler than the one found for its corresponding PC because it only contains three 95% level statistically significant terms whereas the PC model contains five [12]. Of all the four pure solvents tested here dichloromethane is by far the most important to maximize this varimax factor. It is the only solvent with a significant linear blending term and also participates in two significant synergic terms, one with acetone and the other with ethanol. The dichloromethane–acetone synergic interaction term is larger and hence the optimum solvent mixture is binary containing these two solvents. The mixture response surface as a function of dichloromethane, acetone and ethanol proportions is shown in Fig. 3a. Although a 75% dichloromethane–25% acetone mixture is optimal for extraction of the 3.1 min retention time compounds an alternative mixture could be obtained substituting ethanol for acetone. As can be seen in Fig. 3a a 75% dichloromethane–25% ethanol mixture is expected to be almost as efficient in extracting these compounds.

The PC and varimax models have somewhat different response surfaces. Although both models show a maximum in the region of the 75% dichloromethane–25% acetone mixture point the PC model has relatively low response values in the vicinity of the 75% dichloromethane–25% ethanol point. The more complex principal component model does not predict the alternative binary mixture whereas the simpler varimax model predicts both potentially promising extraction mixtures.

Based on statistical criteria a quadratic model is also preferred for the second varimax factor,

$$\begin{aligned} CP_2^r &= \underbrace{0.1485e}_{(\pm 0.1513)} - \underbrace{0.0198d}_{(\pm 0.1513)} + \underbrace{0.9979h}_{(\pm 0.1513)} + \underbrace{0.1394a}_{(\pm 0.1513)} - \underbrace{0.3547ed}_{(\pm 0.6397)} \\ &+ \underbrace{1.5666eh}_{(\pm 0.6397)} - \underbrace{0.2813ea}_{(\pm 0.6397)} + \underbrace{0.5238da}_{(\pm 0.6397)} + \underbrace{1.6607ha}_{(\pm 0.6397)} + \underbrace{0.5238da}_{(\pm 0.6397)} + \underbrace{0.5238da}_{(\pm 0.6397)} + \underbrace{0.6397ha}_{(\pm 0.6397)} + \underbrace{0.5238da}_{(\pm 0.6397)} + \underbrace{0.6397ha}_{(\pm 0.6397)} + \underbrace{0.639ha}_{(\pm 0.639ha} + \underbrace{0.639ha}_{(\pm 0.639ha} + \underbrace{0.639ha}_{(\pm 0.639ha} + \underbrace{0.639ha}_{(\pm 0.639ha} + \underbrace{0.639$$

The mean square lack of fit/pure error ratio of 0.31 is much smaller than the 95% confidence level F-distribution value indicating no significant lack of fit. Furthermore the mean square regression/residual ratio of 10.71 is much larger than the 95% tabled value so the model is statistically significant at this level. The special cubic model is also adequate for describing the second varimax factor scores but its cubic terms are not significant at the 95% level showing that their inclusion in the model is unnecessary, leaving a simpler quadratic model. The model in Eq. (4) has a dominant linear blending term for hexane as well as two significant binary synergic terms involving this extractor with ethanol and acetone. Substitution of acetone by ethanol or vice versa in hexane rich solutions has little effect on the predicted varimax score. This is similar to the behavior observed for the first varimax factor scores in dichloromethane-rich solutions. Optimum extraction mixtures are predicted to be found for solvent concentrations in a lower right triangular region with vertices positioned at pure hexane and close to the 1:1 binary mixtures of hexane:acetone and hexane:ethanol as can be seen in Fig. 3b.

A quadratic model with only two significant coefficients at the 95% confidence level is the preferred model for the third varimax factor scores.

$$\begin{aligned} \text{CP}_{3}^{r} &= \underbrace{\textbf{0.3088e}}_{(\pm 0.1200)} - \underbrace{0.0346d}_{(\pm 0.1200)} - \underbrace{0.0687h}_{(\pm 0.1200)} + \underbrace{0.1631a}_{(\pm 0.1200)} + \underbrace{0.3270ed}_{(\pm 0.5074)} \\ &+ \underbrace{0.0612eh}_{(\pm 0.5074)} + \underbrace{\textbf{2.1173ea}}_{(\pm 0.5074)} - \underbrace{0.3363dh}_{(\pm 0.5074)} - \underbrace{0.5818da}_{(\pm 0.5074)} - \underbrace{0.6108ha}_{(\pm 0.5074)} \\ & (\pm 0.5074) \end{aligned}$$

Its mean square lack of fit/pure error and regression/residual ratios (See Table 2) indicate a highly significant

Table 2 – ANOVA tab	Table 2 – ANOVA tables for preferred models of all six factor loadings varimax									
Variation source	Sum of squares	Degrees of freedom	Mean square	F-value	Probability					
Factor loading 1 (quadra	tic model)									
Regression	2.1788	9	0.2420	18.5158	8.9×10^{-5}					
Residuals	0.1176	9	0.0130							
Lack of fit	0.0259	5	0.0051	0.2262	0.9326					
Pure error	0.0917	4	0.0229							
Total	2.2965	18	0.1275							
Factor loading 2 (quadra	tic model)									
Regression	2.2800	9	0.2533	10.7085	0.0008					
Residuals	0.2129	9	0.0236							
Lack of fit	0.0598	5	0.0119	0.3127	0.8828					
Pure error	0.1530	4	0.0382							
Total	2.4929	18	0.1384							
Factor loading 3 (quadra	tic model)									
Regression	0.8476	9	0.0941	6.3270	0.0056					
Residuals	0.1339	9	0.0148							
Lack of fit	0.0645	5	0.0129	0.7430	0.6305					
Pure error	0.0694	4	0.0173							
Total	0.9815	18	0.0545							

model with no lack of fit at the 95% confidence level. The corresponding model for the third PC has four statistically significant terms at this level. Both models have very dominant binary synergic terms involving the ethanol and acetone proportions so their response surfaces are almost identical with both predicting an optimum extraction medium using a 50:50% binary solution of ethanol and acetone for the 1.7 min retention time compounds [12].

Statistically significant models with no lack of fit at the 95% confidence level are observed for the fourth and sixth varimax factors. However their optimized scores are not useful for selective compound extraction since they contain several strong positive- and/or negative-loading peaks. The fifth varimax factor could not be modeled using conventional mixture models. Even the special cubic model was not statistically significant. However this could have been anticipated since the fifth varimax factor simply describes separation of the replicate subgroups and hence is not expected to depend on concentration changes. Table 3 contains the optimum solvent mixtures determined by visual inspection and the Derringer–Suich algorithm. Although small differences exist in the estimated proportions from both techniques, in practice, the extraction results are expected to be essentially the same on consideration of the uncertainties in the model's parameters.

Confirmatory extractions were carried out using the varimax-optimized mixture, 71% dichloromethane–29% acetone, the alternative varimax optimal mixture, 80% dichloromethane–20% ethanol and the optimized mixture obtained from the first principal component, 83% dichloromethane–17% acetone. The corresponding chromatograms are shown in Fig. 4.

There are only slight differences in the three chromatograms. The PCA-optimized mixture extracts slightly larger amounts of the low retention time compounds whereas the varimax-optimized mixtures are a bit more efficient in extracting compounds with retention times above 3.1 min. Substitution of 29% acetone by 20% ethanol in the

Table 3 – Optimized extractor solvent compositions, predicted varimax scores obtained by visual inspection of response surfaces and by Derringer–Suich desirability function and maximum and minimum scores of varimax-rotated principal components

•					
Method	Peaks (min)	Solvent extractor composition	Scores		
			PC_1^r	PC ₂ ^r	PC ₃
Visual analysis of response surfaces	3.1	75% d-25% a	1.101	-0.048	-0.160
	6.6	75% h–25% a	0.032	1.090	-0.117
	1.7	50% e–50% a	0.063	0.083	0.765
Derringer and Suich algorithm	3.1	71% d-29% a	1.101	-0.077	-0.096
	6.6	83% h–17% a	0.026	1.085	-0.106
	1.7	50% e–50% a	0.053	0.074	0.764
Maximum score value			0.967	0.988	0.874
Minimum score value			0.033	-0.043	-0.096



Fig. 3 – Response surface contour curves for varimax factors. The first and third varimax factor surfaces are shown as a function of dichloromethane, acetone and ethanol, whereas the second one is a function of hexane, acetone and ethanol.



Fig. 4 – Chromatographs for confirmatory experiments of the two alternative varimax-optimized and the PCA-optimized extraction mixtures.

dichloromethane-rich extraction solutions results in almost exactly the same chromatographic profiles as can be seen in the figure confirming predictions of the first varimax factor model. These mixtures also result in near minimum values for the second and third varimax factors. As such very small amounts of the 1.7 and 6.6 min retention time compounds are extracted. Even though the optimized results using PCs or varimax factors are almost equivalent for this plant material the varimax factor values are more convenient to optimize and their models are simpler to interpret.

Confirmatory experiments were really not necessary to test the model for the second varimax factor. The high plateau on the response surface in Fig. 3b covers the region including the 3, 6 and 10 experimental points of the group IV extraction mixtures. Their chromatographic profiles above 4 min retention time are essentially identical as can be seen in Fig. 5. A major difference in the profiles occurs at the 1.7 min retention time. The 50:50% hexane–ethanol mixture produces an extract with large amounts of the 1.7 min retention time com-



Fig. 5 – Chromatographs for the extracted compounds using pure hexane, 1:1 (v/v) hexane:ethanol and 1:1 (v/v) hexane:acetone mixtures.

pounds whereas pure hexane and the 50:50 hexane-acetone mixture do not. This occurs since the former mixture has not been minimized for varimax factor 3 (score of 0.268) whereas the other mixtures have minimum scores (-0.096 and -0.060).

Acknowledgements

The authors are grateful to the Fundação Araucária and the Conselho Nacional de Pesquisa (CNPq) for financial support that made this research possible. CNPq is also acknowledged for providing a doctoral fellowship for PKS and research fellowships to ISS and REB.

R E F E R E N C E S

- [1] M. Defernez, I.J. Colquhoun, Phytochemistry 62 (2003) 1009.
- [2] B.T. Bowie, D.B. Chase, P.R. Griffiths, Appl. Spectrosc. 54 (2000) 164A.
- [3] C.-C. Chuang, W.-C. Wen, S.-J. Sheu, J. Sep. Sci. 30 (2007) 1827.
- [4] R.L. Gorsuch, Factor Analysis, 2nd ed., Lawrence Erlbaum Associates, Hillsdale, 1983.
- [5] I.T. Jolliffe, Principal Component Analysis, Springer, NY, 1986.

- [6] J.J. Andrew, T.M. Hancewicz, Appl. Spectrosc. 52 (1998) 797.
- [7] B.K. Lavine, C.E. Davidson, J. Ritter, D.J. Westover, T. Hancewicz, Microchem. J. 76 (2004) 173.
- [8] S.R.C. Andrade, I.S. Scarminio, M.M. Nery, A.C. Oliveira, J. Pharm. Biomed. Anal. 33 (2003) 655.
- [9] B.K. Lavine, J.P. Ritter, E. Voigtman, Microchem. J. 72 (2002) 163.
- [10] J.R. Schoonover, R. Marx, W.R. Nichols, Vibr. Spectrosc. 35 (2004) 239.
- [11] P.H. Março, I.S. Scarminio, Anal. Chim. Acta 583 (2007) 138.
- [12] P.K. Soares, R.E. Bruns, I.S. Scarminio, J. Sep. Sci. 30 (2007) 3302.
- [13] J.A. Cornell, Experiments with Mixtures, 3rd ed., Wiley, NY, 2002.
- [14] T. Davies, Analusis Mag. 26 (1998) M17.
- [15] M.A. Sharif, D.L. Illman, B.R. Kowalski, Chemometrics, Wiley, NY, 1986.
- [16] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, 3rd ed., Academic Press Inc., NY, 1980.
- [17] R. Reyment, K.G. Jöreskog, Applied Factor Analysis in the Natural Sciences, 2nd ed., Cambridge University Press, 1993.
- [18] A. Báez, R. Belmont, R. García, H. Padilla, M.C. Torres, Atmos. Res. 6 (2007) 61.
- [19] T.-Y. Yu, L.-F.W. Chang, Atmos. Environ. 34 (2000) 4499.
- [20] M.D. Cheng, R.L. Tanner, Atmos. Environ. 36 (2002) 5795.
- [21] M.N. Kumru, M. Bakaç, J. Geochem. Explor. 77 (2003) 81.
- [22] I. Senaratne, D. Shooter, Atmos. Environ. 38 (2004) 3049.



Versão pré revisada de Journal of Separation Science 2007, 30, 3302-3310 DOI: 10.1002/jssc.200700236 © Copyright 2007 WILEY-VCH

This is the pre-peer reviewed version of the following article:

Statistical mixture design – principal component optimization for selective compound extraction from plantmaterial. J. Sep. Sci., 30, 3302-3310, 2007.DOI: 10.1002/jssc.200700236

Statistical mixture design – principal component optimization for selective compound extraction from plant material

Patricia Kaori Soares¹ Roy Edward. Bruns¹ Ieda Spacino Scarminio²

¹Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas, SP, Brazil ²Laboratório de Quimiometria em Ciências Naturais, Departamento de Química, Universidade Estadual de Londrina, CP 6001, 86051-990 Londrina, PR, Brazil

Correspondence: Dr. Roy Edward Bruns, , Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970, Campinas, SP, Brazil.
E-mail: bruns@iqm.unicamp.br
Fax: +55-19-3521-3023

Keywords: extraction optimization / simplex centroid design / mixture modeling / principal components / *Erythrina speciosa* Andrews leaves

Abstract

A simplex centroid design is used to optimize solvent mixtures for selective extraction of compounds from *Erythrina speciosa* Andrews leaves. Three main groups of compounds characterized by chromatographic retention times of 1.7, 3.1 and 6.6 min. were extracted. The chromatographic peak heights registered at 1407 equally-spaced times for each design extract were converted into principal components (PC). Three PC account for 89.1% of the data variance and have important loading values at the above retention times. Quadratic mixture models were found to adequately describe the response surfaces of their PC score values. Maximizing the first PC scores while minimizing the second and third ones results in an optimum 83% dichloromethane-17% acetone mixture that selectively extracts the group of compounds with retention times around 3.1 min. Maximizing the second PC scores and minimizing the third PC scores leads to a 78% hexane-22% acetone mixture adequate for extracting the 6.6 min. retention time compounds as well as other compounds with high retention times. The 1.7 min. group of compounds

is most efficiently extracted on maximizing the third PC that results in a 50% ethanol – 50% acetone mixture. The Derringer and Suich algorithm confirmed the determination of these optimized mixtures.

Introduction

Optimization of extraction techniques is important for a variety of applications. The extraction of selected compounds of commercial and pharmaceutical interest with a minimum of impurities or co-extracted materials provides an outstanding example. Fingerprint identification of new biomarkers that can discriminate between botanical species is also receiving increased interest. In recent years statistically designed experiments [1,2] are being used to minimize the number of experiments as well as identify interaction effects among experimental factors that can be important to optimize extraction procedures. Li et al. [3] applied response surface methodology increasing both the yield and purity of ultra-sound extraction of polysaccharides by about 20% compared to classical extraction methods. Silveira et al. optimized phycocyanin extraction obtaining higher yield and purity using both factorial design and response surface methodology [4] Factorial design was applied to the ultrasound extraction of phenolic compounds from coconut shell powder [5]. Response surface analysis was applied to solvent free microwave extraction of essential oil [6]. Response surface methodology has been used to optimize extraction procedures for secondary metabolites from several plants [7-11]. All these statistical studies have concentrated on optimizing process factors that can be manipulated independently of one another. The optimization of proportions of mixtures making up the extraction medium remains a novel almost unexplored route to optimize extraction procedures. Statistical mixture analysis [12] is a response surface methodology applied to constrained variables that has been shown to be useful for chromatographic fingerprint development for the discrimination of botanical species [13].

It is relatively simple to find the optimum conditions for a single response using experimental designs. However the researcher confronts more complex problems trying to simultaneously optimize several responses that are normally of interest in extraction applications. The simplest strategy is visual inspection of surfaces for each response of interest. If the number of significant factors allows the graphical visualization of the adjusted models and if the number of responses is not too large, the surfaces can be overlapped and the optimum conditions found by inspection [2]. However if the optimum conditions of each response differ a compromise solution must be found since changes in the proportions of one mixture component that improves some response values will have negative and/or neutral effects on the others. Derringer and Suich [14] propose the use of a desirability function for simultaneous response optimization that has been already applied to the solid-phase extraction of 18 organochlorine and nine organophosphorus pesticides[15].

183

Principal component analysis (PCA) is a chemometrics tool that has been extensively used for classification [16], pattern recognition [17] and multivariate calibration [18] applications. Here this chemometrics technique and statistical design are used in a pioneering attempt to optimize solvent mixtures for the selective extraction of compounds from plant material. Extracts are often analysed by chromatographic techniques for which peak displacements are common and there is a problem of defining the analytical signals that provide the most adequate responses to guide the optimization procedure. Principal component loadings identify those retention times that have variations in analytical signal highly correlated with specific changes in component proportions of the more polar components increase whereas others will diminish. The mixture component proportions that maximize or minimize the analytical signals for these retention times are the optimized ones for selective extraction. The response surfaces that contain the information necessary to determine these proportions are determined from the principal component scores.

Experimental

Erythrina speciosa Andrews leaves were collected in November of 2006. Care was taken to avoid leaves with insect, fungus and mechanical damage. A voucher specimen of this plant was deposited in the herbarium of the Londrina State University and classified by M. R. C. Paiva and registered under the FUEL 35133 number.

Drying was carried out avoiding physical, chemical and microbiological modifications so that adverse reactions, such as oxidation or hydrolysis by organic substances, and attacks by microorganisms did not take place.

Drying was carried out for nine days at room temperature with the leaves protected from light, humidity and attacks by fungus, insects and rodents. Since solar irradiation can alter the chemical composition of the plant the drying was carried out in the shade.

Extraction mediums were prepared using mixtures of four components, 1) ethanol, 2) dichloromethane, 3) hexane and 4) acetone whose proportions were varied according to the simplex centroid design in Figure 1. The proportions of each solvent used in the extraction mixtures are specified in Table 1. Nineteen extractions were carried out with fifteen different mixtures. A five-run replicate was performed at the center point so experimental error could be estimated.

2.000 g of dried leaves chopped into small pieces by a food processor were added to 30 mL of the solvent mixtures in Table 1. The mixtures were placed in an ultrasonic bath for 30 minutes [19]. The bath's water was changed every ten minutes to maintain constant room temperature. The extracts were filtered in cotton. 50 μ L of these extracts were diluted in 950 μ L of a 80:20 v/v methanol-water mobile phase. This mixture was passed twice through a disposable Millipore Millex 0.22 μ m filter. The samples were analyzed by HPLC using a Shimadzu Model LC-10 AD liquid chromatograph equipped with a SPD-M10AVP diode array detector. The chromatographic conditions

184

were: 250 mm x 4.6mm Metasil C18 ODS column, 5 μm particle size, 50° C furnace temperature, 20 μL injection volume, 210 nm detector, 20 min. retention time. Elution was performed isocratically with a flow rate of 1 mL min⁻¹.



Figure 1. Simplex centroid mixture design used to optimize the extraction mixtures. x_1 , x_2 , x_3 and x_4 represent the proportions of ethanol, dichloromethane, hexane and acetone, respectively.

Table 1.	Extraction	solution	compositions,	scores	of the	first	three	principal	components	and	their	percentages	of
explaine	d variance.												

		Solvente		Comp	onentes prir	ncipais		
Extrato	Ethanol	Dichloromethane	Hexane	Acetone	DC	DC	DC	Ordem
	X ₁	X ₂	X 3	X 4	PC_1	PC ₂	PC3	
1	1	0	0	0	0.077	0.113	0.296	1
2	0	1	0	0	0.434	-0.417	-0.190	2
3	0	0	1	0	0.129	0.429	-0.262	12
4	0	0	0	1	0.092	0.122	0.285	8
5	1/2	1/2	0	0	0.260	-0.132	0.160	7
6	1/2	0	1/2	0	0.113	0.297	-0.016	4
7	1/2	0	0	1/2	0.104	0.130	0.587	11
8	0	1/2	1/2	0	0.272	0.117	-0.251	6
9	0	1/2	0	1/2	0.427	-0.337	-0.092	14
10	0	0	1/2	1/2	0.132	0.407	-0.198	10
11	1/3	1/3	1/3	0	0.201	0.135	-0.044	13
12	1/3	1/3	0	1/3	0.286	-0.138	0.219	3
13	1/3	0	1/3	1/3	0.122	0.314	0.015	16
14	0	1/3	1/3	1/3	0.271	0.090	-0.199	5
15	1/4	1/4	1/4	1/4	0.219	0.117	-0.052	9
16	1/4	1/4	1/4	1/4	0.214	0.117	-0.059	15
17	1/4	1/4	1/4	1/4	0.170	0.093	0.141	18
18	1/4	1/4	1/4	1/4	0.230	0.088	0.266	17
19	1/4	1/4	1/4	1/4	0.162	0.082	0.234	19
Percentage variance 58,94 20,53 10,55								

Statistical methods

Principal component based methods can be used to reduce the number of variables in chromatography and spectroscopy problems so their full analytical potentials can be exploited. Near infrared spectroscopy provides one outstanding example of this [20]. These methods have been less commonly applied to chromatographic data.

Principal components are obtained by diagonalizing the correlation matrix [21], **Y**^t**Y**, of an experimental data matrix, **Y**. **Y** consists of n rows, one for each chromatogram, and p columns, one for each retention time of the digitalized chromatogram. The elements of the **Y** matrix are the values of the chromatographic signal for the different chromatograms and retention times.

On diagonalizing **Y**^t**Y** a small number of principal components are selected that are capable of describing a large portion of the data variance. Ideally the variance of the selected components describes signal whereas that of the discarded components pertain to noise. For the principal component model

$$\mathbf{Y} = \mathbf{T}\mathbf{P}^t + \boldsymbol{\varepsilon} \tag{1}$$

where the product of scores and loadings, **T** and **P**^t, are expected to contain useful information and ε , the noise. The loading matrix contains elements that are correlation coefficients between the principal components and the original variables. The score matrix contains information about the chromatograms that can be applied to optimization problems.

Statistical mixture designs are often used to optimize the proportions of the components that make up a mixture [12]. The same mixture design can be used to optimize different responses of a mixture problem. One of the goals of statistical mixture designs is to obtain a desired result on formulating a minimum number of mixtures.

Mixture models differ from most regression models since their variables are component proportions that sum to unity or 100%. The least square matrix equations for calculating model coefficients, **b**, and their standard errors, **V(b)**, are

$$\mathbf{b} = \left(\mathbf{X}^{\mathsf{t}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{t}}\mathbf{t} \quad \text{and} \quad \mathbf{V}\left(\mathbf{b}\right) = \left(\mathbf{X}^{\mathsf{t}}\mathbf{X}\right)^{-1}s^{2}$$
(2)

where **X** is the design matrix and **t** is the response vector, in our application one of the first three columns of the principal component score matrix. For mixture designs the $\mathbf{X}^t \mathbf{X}$ matrix is singular and Eqs. 1 cannot be solved. For this reason the restraint condition is incorporated into the matrix changing the form of **X** so that $\mathbf{X}^t \mathbf{X}$ is no longer singular. One consequence of this procedure is that mixture models do not have constant terms like most regression models.

The most sophisticated model investigated in our work is the special cubic model

$$y = \sum_{i=1}^{m} b_i x_i + \sum_{i=1}^{m} \sum_{j \neq i}^{m} b_{ij} x_i x_j + \sum_{i=1}^{m} \sum_{j \neq i}^{m} \sum_{k \neq j}^{m} b_{ijk} x_i x_j x_k + \mathcal{E}$$
(3)

Linear models using only the first summation and quadratic models that contain the first two summations are also investigated in this work.

Calculation of the mixture response surfaces and principal components were carried out using the Statistica 6.0 program [22]

Results and Discussion

Chromatographic peak heights were registered at 1407 equally-spaced times during a 15 min. interval. The chemometrics data matrix, **X**, consisted of 19 rows and 1407 columns. Each row corresponds to a chromatogram of an extract obtained using a simplex centroid mixture and was normalized to unit area to remove variations owing to slightly different injection volumes. The chromatograms are shown in Figure 2. Three large peaks contain most of the eluted compounds and five characteristic chromatograms are obtained. The group I mixtures are efficient at extracting only those compounds being eluted in about 3.1 min., group II mixtures those eluted in 3.1 and 6.6 min., group III those in 1.7 and 3.1 min., Group IV those in 1.7 and 6.6 min. and Group V eluted predominantly the 1.7 min. eluted compounds.



Figure 2. Chromatograms of the Erythrina speciosa Andrews extracts for the mixtures of the simplex centroid design.

The first three principal components account for 89.1% of the total variance in the data. A graph of the scores of these three principal components is shown in Figure 3 and their values have been included in Table 1. Five groups can be observed. The largest one, Group II, contains all the points corresponding to the center point as well
as points for extraction mixtures 8, 11 and 14. The dispersion of points 15-19 of this group gives us a measure of variation owing to experimental error. The first component contains 58.9% of the variance and clearly separates groups I and III from the other groups. The second component best discriminates groups I and IV and contains 20.5% of the variance. The third component, even though it describes only 10.6% of the variance, is important in discriminating group V from the others although this is not so evident in the projection used in Figure 3.



Figure 3. Principal component graph for the first three components. This graph explains 89.1% of the total data variance. The numbers represent the mixtures in Table 1.

The principal component loadings are shown in Figure 4 for the 1407 retention time variables. The first principal component scores are all positive. The values are especially high at retention times of 6.6, 3.1 and 1.7 min. On the other hand, the second component has negative loadings in the region of 3.1 min. while the third one has negative loading around 3.1 and 6.6 min. Mixture 2 and 9 of group I with high PC₁ scores are expected to be especially efficient in extracting the chemical substances with a 3.1 min. retention time since they have high loadings on this component. The substances with retention time around 6.6 min. are extracted by the Group IV mixtures that have high PC₂ scores. Finally Group V, containing the highest PC₃ scores, contains mixtures that are capable of extracting the 1.7 min. retention time substances.



Figure 4. Loading graph of the first three principal components with retention times indicated for the largest peaks.

A hierarchical cluster analysis was performed to complement the PCA results. The dendrogram in Figure 5, obtained using the unweighted pair-group average algorithm, confirms the existence of groups I, II, III and IV that form individual clusters. Although mixtures 1 and 7 are placed in a different group than mixture 4 they are very similar as indicated by the small difference in their similarity values, given by the horizontal lines joining mixture 4 and mixtures 1 and 7 with the main cluster containing groups II, III and IV. The dendrogram clearly shows that Group I is much different than the other groups.



Figure 5. Unweighted pair-group average dendrogram of the chromatograms of the simplex centroid design.

The PCA suggests that the extraction mixtures can be optimized to separate specific substances from the plant material. Since Group I falls in a region with high PC₁ and low PC₂ and PC₃ scores its mixtures should be useful in extracting the compounds with retention times around 3.1 min. from the plant material. On the other hand, the mixtures of group IV with high PC₂ and low PC₁ and PC₃ scores can be expected to preferentially separate the compounds with a 6.6 min. retention time. It was decided to apply response surface analysis to obtain models for

the PC scores with the objective of optimizing the extraction mixtures for isolating compounds with retention times of 1.7, 3.1 and 6.6 min.

Linear, quadratic and special cubic models were determined for the scores of the first three principal components. This involves performing a linear regression, via Eq. (2), of the scores in Table 1 for each principal component on the mixture proportions for the linear model. The quadratic and special cubic model regressions involve including second and third-ordered cross terms of Eq. (3) in addition to the component proportions in the regression. Standard errors were propagated from the experimental error estimate made at the replicated center point. The model coefficients were tested for statistical significance by comparing the model coefficient/standard error ratio with the t-distribution of the appropriate number of degrees of freedom. All models were validated using analysis of variance (ANOVA) of the regression results. Lack of fit/pure error mean square ratios were compared with F-distribution values with the appropriate numbers of degrees of freedom. Finally residual plots were examined for evidence of possible lack of fit for all tentative models.

The linear, quadratic and special cubic models applied to the first principal component scores all resulted in models that presented no significant lack of fit at the 95% confidence level. The quadratic model is preferred since none of the cubic term coefficients were significant whereas the dichloromethane – acetone quadratic level term was significant at the 95% confidence level. The quadratic model is given by

$$PC_{1} = \underbrace{\mathbf{0.0827}}_{(\pm 0.0268)} e + \underbrace{\mathbf{0.4423}}_{(0.0268)} d + \underbrace{\mathbf{0.1375}}_{(\pm 0.0268)} h + \underbrace{\mathbf{0.0987}}_{(\pm 0.0268)} a - \underbrace{\mathbf{0.0501}}_{(\pm 0.1132)} e - \underbrace{\mathbf{0.0207}}_{(\pm 0.1132)} h + \underbrace{\mathbf{0.0332}}_{(\pm 0.1132)} e - \underbrace{\mathbf{0.1428}}_{(\pm 0.1132)} d + \underbrace{\mathbf{0.0006}}_{(\pm 0.1132)} h + \underbrace{\mathbf{0.0332}}_{(\pm 0.1132)} e - \underbrace{\mathbf{0.1428}}_{(\pm 0.1132)} d + \underbrace{\mathbf{0.0006}}_{(\pm 0.1132)} h + \underbrace{\mathbf{0.0006}}_{(\pm 0.1132)} h$$

where *e*, *d*, *h* and *a* are the proportions of ethanol, dichloromethane, hexane and acetone, respectively. Standard error estimates are presented in parenthesis directly below their corresponding model coefficient. Bold face coefficients indicate those that are significant at the 95% confidence level. The ANOVA of the regression results is presented in Table 2. The mean square lack of fit/ pure error ratio of 0.55 is much smaller than the $v_1 = 5$, $v_2 = 4$ 95% confidence F-distribution value of 6.26 indicating no significant lack of fit. Furthermore the mean square regression/residual ratio of 29.00 is much larger than the $v_1 = v_2 = 9$ 95% confidence F-distribution value of 3.18 showing that the regression is highly significant.

Since the model in Eq. (4) is validated it can be interpreted to determine those effects that are important for increasing the PC₁ scores and hence the amount of extracted 3.1 min. retention time substances. Of the linear blending coefficients the one for dichloromethane is much larger than those for the other solvents that have very similar values. Of all the four pure solvents tested here dichloromethane is by far the most important to maximize the PC₁ score and extract the 3.1 min. retention time substances. The presence of acetone in the extraction mixture can be important because of the strong synergic interaction it has with dichloromethane. The mixture response

surface for PC₁ is given in Figure 6a. For clarity, the ethanol proportion has been set to zero in order to have a twodimensional representation of the mixture response surface. The maximum score value is predicted to occur for a 80-85% dichloromethane: 20-15% acetone mixture. This result is consistent with the results of the PCA. Of all the mixtures tested the 2^{nd} (pure dichloromethane) and 9^{th} (50% dichloromethane:50% acetone) mixtures provided the chromatograms with the largest PC₁ scores.

Variation source	Sum of squares	Degrees of freedom	Mean square		
PC ₁					
Regression	0.1835	9	0.0203		
Residuals	0.0066	9	0.0007		
Lack of fit	0.0028	5	0.0005		
Pure error	0.0038	4	0.0009		
Total	0.1902	18	0.0105		
PC ₂					
Regression	0.8578	9	0.0953		
Residuals	0.0019	9	0.0002		
Lack of fit	0.0007	5	0.0001		
Pure error	0.0011	4	0.0002		
Total	0.8597	18	0.04776		
PC ₃					
Regression	0.8337	9	0.0926		
Residuals	0.1291	9	0.0143		
Lack of fit	0.0331	5	0.0066		
Pure error	0.0959	4	0.0239		
Total	0.9628	18	0.0534		

Table 2. Analysis of variance (ANOVA) for the quadratic model regression data of the first three principal components of the chromatographic data of the *Erythrina speciosa* Andrews extracts.

The linear model for PC_2 showed significant lack of fit. However both the quadratic and special cubic models were well-adjusted to the second component scores. The quadratic model was preferred since none of the cubic interaction terms were significant at the 95% confidence level. The quadratic model is given in Eq. (5) and its ANOVA results have been included in Table 2.

$$PC_{2} = \mathbf{0.1120}e - \mathbf{0.4196}d + \mathbf{0.4253}h + \mathbf{0.1213}a + \mathbf{0.0920}ed + \mathbf{0.1552}eh$$

$$+ 0.0549 ea + 0.5050 dh - 0.7492 da + 0.5744 ha .$$
(5)

The mean square lack of fit/pure error ratio of 0.5 is smaller than the $F_{5,4} = 6.26$ tabled value at the 95% confidence level. Besides representing the scores accurately the quadratic model is highly significant with a mean square regression/residual ratio of 477 compared to a 95% confidence level $F_{9,9}$ reference value of 3.18. Graphs showed that the residuals have random behavior further supporting the validity of the model.

The model in Eq. (5) shows that controlling the proportions of three solvents, hexane, dichloromethane and acetone, is important to optimize the PC_2 scores. Of the linear blending terms two are most important, the positive hexane term and the negative dichloromethane term. Strong synergistic interactions can be found for the dichloromethane – hexane and hexane – acetone binary mixtures. An even more important antagonistic term is found for dichloromethane – acetone mixtures. Maximum PC_2 values can be expected for mixtures rich in hexane with moderate amounts of acetone.

Figure 6b illustrates the mixture response surface for the PC_2 scores as a function of the hexane, acetone and dichloromethane proportions. Maximum PC_2 scores are predicted for binary mixtures close to the 75% hexane -25% acetone composition. Of the mixtures in Table 1 the pure hexane and 1:1 hexane:acetone mixtures of Group IV have the largest PC_2 scores. Figure 6b also shows that minimum PC_2 scores can be expected to occur for binary mixtures of dichloromethane and acetone close to a 80%:20% ratio. The pure dichloromethane and 1:1 dichloromethane: acetone binary mixtures give the most negative PC_2 scores in Table 1.

The linear, quadratic and special cubic models were all capable of adequately representing the third principal component scores, PC_3 with no evidence of lack of fit. No significant ternary effects and only one binary effect were significant at the 95% confidence level. So the quadratic model, given in Eq. (6) was preferred to represent the PC_3 scores.

$$PC_{3} = \underbrace{\mathbf{0.2886}}_{(\pm 0.1178)} e^{-0.2094} d - \underbrace{\mathbf{0.2749}}_{(\pm 0.1178)} h + \underbrace{\mathbf{0.2743}}_{(\pm 0.1178)} a^{+0.5674} e^{-0.1192} e^{-0$$

The ANOVA of the regression results for this model have been included in Table 2. The mean square lack of fit/ pure error ratio of 0.28 is much smaller than the $F_{5,4}$ critical value of 6.26 at the 95% confidence level indicating no significant lack of fit. Residual plots did not present any evidence of lack of fit since they displayed a random distribution. The regression is significant with a mean square regression to residual ratio of 6.5 that is larger than the 95% confidence level $F_{9,9}$ value of 3.18.

The absolute values of all the linear blending coefficients in Eq. (6) are about the same. However the coefficients for ethanol and acetone are positive whereas the ones for dichloromethane and hexane are negative. The only significant interaction involves ethanol and acetone. So it is very simple to find a maximum value for the

 PC_3 score. It can be expected for a 1:1 binary mixture of ethanol and acetone. This is confirmed by the response surface plot of PC_3 as a function of the ethanol, acetone and hexane proportions. The predicted PC_3 score from Eq. (6) for this mixture is 0.577 close to the 0.587 result for the 1:1 ethanol-acetone mixture in Table 1. Minimum PC_3 scores in Table 1 are found for pure dichloromethane or pure hexane or their binary mixture as can be confirmed in Table 1.



Figure 6. Mixture response surfaces for the first three principal component score values, a) PC₁, b) PC₂ and c) PC₃.

Table 3 contains a summary of the results obtained from the response surface analyses for the three mixtures that are predicted to maximize the three principal component scores. Predicted values of these scores as calculated from Eq. (4) – (6) for each mixture are presented in the first three rows of the table. Maximum and minimum score values from Table 1 have been included in the last two rows. The loading graph in Figure 4 shows that the $T_R = 3.1$ retention compounds are highly correlated with PC₁ whereas they are negatively correlated with PC₂ and PC₃. As such maximum PC₁ values and minimum PC₂ and PC₃ values are required for an efficient extraction of the $T_R = 3.1$ compounds. The 80% dichloromethane – 20% acetone mixture is seen to furnish these results since the predicted model score for the first principal component is indeed close to the maximum value in Table 1 while the predicted PC₂ and PC₃ scores are close to the minimum values. So this extraction mixture is optimum (or close to

it) for extracting the $T_R = 3.1$ min. compounds. The 75% hexane – 25% acetone mixture has a predicted PC₂ score very close to its maximum and a PC₃ score close to its minimum value in Table 1. This mixture is almost optimum for extracting the $T_R = 6.6$ min. compounds since the loading graph in Figure 4 has a very positive peak for the second PC and a negative one for the third PC at this retention time. The loading graph exhibits three positive peaks for the $T_R = 1.7$ retention time compounds. Of these the one for PC₃ is by far the strongest. Efficient extraction of these compounds would require a mixture with response surface predicted values that are maximum for the third PC and positive for the first and second components. The 50% ethanol – 50% acetone mixture has predicted score values consistent with these criteria.

Table 3	 Predicted 	principal	component	scores fo	r three	optimized	extraction	mixtures	and	maximum	and	minimum
score va	alues from ⁻	Table 1.										

Mixture	PC1	PC ₂	PC ₃
80% d - 20% a ^a	0.465	-0.431	-0.172
75% h – 25% aª	0.128	0.456	-0.283
50% e – 50% aA	0.099	0.130	0.576
Maximum (Table 1)	0.434	0.429	0.587
Minimum (Table 1)	0.077	-0.417	-0.262
83% d – 17% a ^b	0.465	-0.433	-0,179
78% h – 21% a ^b	0.130	0.454	-0.286
50% e – 50% a ^b	0.099	0.130	0.576

^a Visual inspection, ^b Derringer-Suich algorithm.

Included in Table 3 are optimum mixtures obtained using the algorithm of Derringer and Suich and their predicted principal component scores calculated using Eqs. (4)-(6). A 83% dichloromethane – 17% acetone solution is indicated as optimal on maximizing the first principal component score and minimizing the second and third ones. The 78% hexane – 22% acetone mixture was obtained on maximizing the PC₂ scores and minimizing the PC₃ scores while requiring the first component scores to remain positive. Maximizing the third principal component scores while requiring the scores of PC₁ to be positive and allowing the PC₂ scores to vary freely resulted in the 50%-50% ethanol – acetone mixture. As can be seen these indications confirm the conclusions made by visual inspection of the mixture response surfaces.

Confirmatory experiments were run using those mixtures that were predicted to be optimum for selective extraction. Chromatograms obtained for extracts of erythrina speciosa leaves have been included in Figure 2. The 50%:50% ethanol-acetone mixture extracts the 1.7 min. retention time compounds along with other compounds with low retention times. The 83%:17% dichloromethane acetone mixture almost exclusively extracts the 3.1 min.

retention time compounds. Finally the 78%:22% hexane-acetone mixture extracts those compounds with high retention times, expecially the 6.6 min. retention time compounds.

Acknowledgements

The authors thank Márcia C. Breitkreitz for helping us with the Derringer-Suich calculations. The authors are grateful to the Fundação Araucaria, Conselho Nacional de Pesquisa (CNPq) and Fundação de Amparo á Pesquisa do Estado de São Paulo (FAPESP) for financial support that made the reported research possible. CNPq is acknowledged for providing a doctoral fellowship for PKS and research fellowships to ISS and REB.

References

- 1 Box G. E. P., Hunter J. S., Hunter W. G., *Statistics for experimenters*, Wiley, 2005.
- 2 Bruns R. E., Scarminio I. S., Neto B. B., Statistical Design Chemometrics, Elsevier, Amsterdam, 2006.
- 3 Li J-W, Ding S-D, Ding X-L, J. Food. Eng, 2007, 80, 176-183.
- 4 Silveira S. T., Burkert J. F. M., Costa J. A. V., Burkert C. A. V., Kalil S. J., *Bioresource Technol.*, 2007, 98, 1629-1634.
- 5 Rodrigues S., Pinto G. A. S., J. Food. Eng., 2007, 80, 869-872.
- 6 Lucchesi M. E., Smadja J., Bradshaw S., Louw W., Chemat F., J. Food. Eng, 2007, 79, 1079-1086.
- 7 Bandeira K. F., Tininis A. G., Bolzani V. S., Cavalheiro A. J., *Phytochem. Anal.* 2006, 17, 168-175.
- 8 Jones N. M., Bernardo-Gil M. G., Lourenço M. G., JAOAC Int. 2000, 84, 309-316.
- 9 Liu F. F., Ang C. Y. W., Springer D., J Agric Food Chem. 2000, 48, 3364–3371.
- 10 Moldão-Martins M., Palavra A., da Costa M. L. B., Bernardo-Gil M. G., J. Supercrit. Fluids, 2000, 18, 25-34.
- 11 Del Valle J. M., Bello S., Thiel J., Allen A., Chordia L., Braz. J. Chem. Eng. 2000, 17, 335-348
- 12 Cornell J. A., *Experiments with Mixtures: Designs, Models, and Analysis of Mixture Data*, Second ed., Wiley, New York, 1990.
- 13 Borges C.N., Bruns R. E., Almeida A. A., Scarminio I. S., Anal. Chim. Acta ,2007, doi:10.1016/j.aca.2007.02.067
- 14 Derringer G., Suich R., J. Qual. Technol. 1980, 12, 214-219.
- 15 Carro A. M., R. A. Lorenzo, Analyst, 2001, 7, 1005-1010.
- 16 Scarminio I. S., Lonni A. A. S. G., e Silva L. M. C., Ferreira D. T., Anal. Sci. 21 (2005) 235-239.
- 17 Wold S., Albano C., Dunn III W. J., Edlund U., ESbenson K., Geladi P., Hellberg S., Johansson E., Lindberg W., Sjöstrom M., Kowalski B. R. (Ed.) *Chemometrics, Mathematics and Statistics in Chemistry*, Reidel, Dordrecht, 1984.
- 18 Martins N., *Multivariate Calibration*, John Wiley & Sons, Chichester, 1992.

- 19 Melecchi M. I. S., Péres V. F., Dariva C., Zini C. A., Abad F. C., Martinez M. M., Camarão E. B., *Utrason.* Sonochem., 2006, 13, 242-250.
- 20 Davies T., Analusis Magazine, 1998, 26, M17-M19.
- 21 Mardia K. V., Ken J. T. t, Bibby J. M., *Multivariate Analysis*, Academic Press, London, 1979.
- 22 Statistica for Windows, **Statsoft**, Inc., Tulsa, OK, USA, 1999.



Versão submetida *Química Nova*

Estatística aplicada à química: dez dúvidas comuns

Livia Maria Zambrozi Garcia Passari¹, Patricia Kaori Soares¹, Roy Edward Bruns^{1*}, Ieda Spacino Scarminio² ¹Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas, SP, Brasil ²Laboratório de Quimiometria em Ciências Naturais, Departamento de Química, Universidade Estadual de Londrina, CP 6001, 86051-990 Londrina, PR, Brasil

Correspondência autor. R. E. Bruns Tel.: (19) 35213106; Fax: (19) 35213023 E-mail : bruns@iqm.unicamp.br

ABSTRACT

Ten common doubts of chemistry students and professionals about their statistical applications are discussed. The use of the N-1 denominator for the standard deviation instead of N is described. The reason why scientists and engineers use the average instead of the median is explained. Several problematic aspects about regression and correlation are treated. The popular use of triplicate experiments in teaching and research laboratories seen to have its origin in statistical confidence intervals. Nonparametric statistics and bootstrapping methods round out the discussion.

Key words: regressão linear, mediana, correlação, triplicatas, estatística não-paramétrica

INTRODUÇÃO

As aplicações da estatística se desenvolveram de tal forma que praticamente todas as áreas de pesquisa e produção se beneficiam da utilização de seus métodos. Freqüentemente estudantes e pesquisadores que aplicam métodos estatísticos como ferramenta de análise dos dados, encontram dificuldades para compreender e interpretar alguns conceitos estatísticos importantes.

O objetivo deste trabalho é reunir em um só texto, dez dúvidas e respostas que são frequentes entre químicos quando métodos estatísticos são aplicados aos seus dados. Essas informações, embora possam ser encontradas, estão dispersas em várias publicações fora da área de química e muitas das quais num vocabulário pouco familiar aos químicos.

1 Porque o denominador do desvio padrão amostral é N-1?

Cursos de estatística normalmente explicam o denominador $(N-1)^{1/2}$ na equação do desvio padrão em termos de graus de liberdade,

$$s_{x} = \left\{ \sum \left(x_{i} - \overline{x} \right)^{2} / N - 1 \right\}^{\frac{1}{2}}$$
(1)

onde, s_x é o desvio padrão de x, x_i é a *i*-ésima observação, N é o número de observações e \overline{x} é a média das observações, definida como o somatório de todas as observações dividido pelo número total de observações.

Para entender a razão do denominador não ser o número total de observações, *N*, e sim *N*-1 imagine cinco amostras com os seguintes teores de ferro: 70,2%, 71,0%, 70,8%, 73,5% e 70,6%. Normalmente, a estimativa da quantidade de ferro nessas amostras é a média, 71,22%. As amostras apresentam cinco resultados de porcentagem de ferro que não podem ser preditos antes da realização das análises, isto é, elas possuem cinco valores não conhecidos ou graus de liberdade para serem especificados. Mesmo conhecendo a porcentagem de ferro em cinco amostras analisadas, não é possível prever a porcentagem da próxima amostra que será analisada.

Com a soma dos valores de todos os desvios a situação é diferente, porque este somatório é igual a zero

$$\sum_{i} d_{i} = \sum_{i} (x_{i} - \overline{x}) = \sum_{i} x_{i} - \sum_{i} \overline{x} = \sum_{i} x_{i} - N\overline{x} = \sum_{i} x_{i} - \sum_{i} x_{i} = 0$$
 (2)

O termo $N \overline{x}$ foi substituído pelo somatório $\sum_{i} x_i$ e $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$.

Sendo assim, caso se conheça o valor da média das cinco amostras será necessário executar somente quatro análises para saber a porcentagem de ferro da quinta amostra, ou seja, temos agora quatro graus de liberdade. A restrição imposta pela Equação 2, que vem do cálculo da média retira um grau de liberdade do conjunto de desvios. Considerando que dos *N* desvios só *N*-1 podem flutuar aleatoriamente, é natural que o denominador na definição da variância amostral seja *N*-1 e não *N*.

É importante notar que as considerações feitas acima não constituem uma prova que a Equação (1) seja uma estimativa sem tendências no desvio padrão. A prova matemática disto pode ser encontrada no livro de Montgomery¹. Pela mesma razão, a raiz quadrada do erro médio quadrático de calibração tem um denominador (*N*p)^{1/2} e não *N*,

$$RMSEC = \left[\frac{\sum (y_i - \hat{y}_i)^2}{N - p}\right]^{1/2}$$
(3)

onde, y_i é a *i*-ésima observação, \hat{y}_i é o *i*-ésimo valor previsto pelo modelo, N é o número de observações e p é o número de parâmetros do modelo de calibração.

É necessário relembrar que essa fórmula é a mesma para a raiz quadrada do erro médio quadrático de validação, Equação 4, salvo que neste caso o denominador é $N^{1/2}$ porque os valores de y_i não foram utilizados para determinar o modelo de calibração.

$$RMSEV = \left[\frac{\sum (y_i - \hat{y}_i)^2}{N}\right]^{1/2}$$
(4)

2 Porque utilizar a média e não a mediana?

Considerando as análises dos teores de ferro descritas no item 1, observa-se que o valor de 73,5% é consideravelmente maior do que os resultados das outras determinações. Se forem feitos testes usando os métodos de Dixon ou de Grubbs², entre muitos outros, o valor de 73,5% seria considerado um "outlier" em relação aos outros resultados determinados para o teor de ferro.

A mediana é obtida listando-se os *N* valores em ordem de magnitude e selecionando o valor do centro se *N* for ímpar ou a média dos dois valores centrais se *N* for par. Sendo assim, a utilização do valor mediano é menos sensível à inclusão ou não de valores extremos, como por exemplo, o valor da mediana incluindo o teor de ferro 73,5% no cálculo é 70,6%, excluindo este valor a mediana é 70,4%. Para o valor médio utilizando os cinco valores dos

teores de ferro a média é 71,22%, enquanto que rejeitando o valor suspeito a média é 70,65%, um valor bem diferente.

Podemos concluir que o valor mediano é bem mais robusto do que o valor da média para a presença de valores suspeitos no conjunto de dados. Porém, a média é preferencialmente utilizada no lugar da mediana porque existe uma equação simples para calcular o erro no valor médio, s_{τ} , Equação 5,

$$s_{\overline{x}} = \left(\frac{s_x^2}{N}\right)^{\frac{1}{2}}$$
(5)

onde, N é o número de observações e s_x é o desvio padrão de x, apresentado na Equação 1. Não existe uma equação, mesmo complexa, que calcule o erro no valor mediano, lembrando que o erro da média é sempre menor do que de uma medida individual.

3 Por que em métodos de regressão, a melhor reta é aquela que minimiza as soma dos quadrados das distâncias verticais entre os pontos e a reta?

A Figura 1 mostra um gráfico que ilustra o princípio do ajuste por mínimos quadrados normalmente empregado em análises de regressão.



Figura 1: Reta ajustada por mínimos quadrados.

Nesse método a melhor localização da reta especificada pelos coeficientes b_0 e b_1 da equação, $\hat{y}_i = b_0 + b_1 X_i$, é aquela que minimiza a soma dos quadrados dos comprimentos dos segmentos verticais,

indicado pelas linhas sólidas, que ligam os pontos experimentais à reta. Este critério implica que os valores de X_i sejam valores fixos definidos pelo pesquisador no planejamento estatístico e os valores de y_i as variáveis aleatórias afetadas por erros experimentais.

Se ambas as variáveis X e y forem afetadas por erros experimentais da mesma grandeza, a minimização da soma quadrática das distâncias ortogonais, representadas pelas linhas tracejadas, seria a mais apropriada². No entanto, na prática isto não ocorre, pois os químicos possuem outras preocupações mais relevantes.

O aluno normalmente se pergunta por que não minimizar a soma das distâncias verticais em lugar da soma dos quadrados. A resposta é que minimizando as distâncias verticais, teríamos um número infinito de retas que satisfariam a condição $\sum (y_i - \hat{y}_i) = 0$, pois desvios positivos da reta cancelariam os desvios negativos. Este cancelamento poderia ser evitado, minimizando-se a soma dos valores absolutos das distâncias verticais. Na prática isto não é feito porque não existem derivadas para as Equações 6 e 7, necessárias para gerar as equações lineares que determinam os valores de $b_0 e b_1$.

$$\frac{\partial \sum \left| y_i - \hat{y}_i \right|}{\partial b_0} = 0 \tag{6}$$

$$\frac{\partial \sum |y_i - \hat{y}_i|}{\partial b_1} = 0 \tag{7}$$

4 É possível determinar um modelo quadrático usando regressão linear?

O modelo quadrático é freqüentemente representado pela seguinte equação genérica:

$$\hat{y} = b_0 + b_1 x + b_2 x^2 \tag{8}$$

Os primeiros dois termos do lado direito representam o modelo linear muito usado pelos químicos para calibração de métodos analíticos, determinação de quantidades físicos-químicas bem como para relacionar medidas empíricas. Uma vez que a aproximação linear tem sua validade limitada, o terceiro termo ($b_2 x^2$), pode ser adicionado ao modelo permitindo melhor ajuste dos dados.

O termo "regressão linear" usado pelos estatísticos corresponde aos parâmetros *b*, ou seja, as incógnitas no modelo. As variáveis independentes *x* são valores fixados pelo experimentador em diferentes níveis, como por exemplo, as concentrações para a curva de calibração ou temperaturas para determinar as mudanças na pressão de vapor, enquanto que as respostas, valores de *y*, são medidas experimentais.

Um exemplo de modelo não linear nos parâmetros estatísticos é, $y = b_0 (1 - e^{b_1 x})$, embora esse modelo possa ser linearizado resolvendo o logaritmo.

5 Como um cientista ou engenheiro pode testar se uma curva de calibração é realmente uma reta?

Muitos modelos de calibração são baseados na suposição de que a relação entre a propriedade medida (sinal analítico, logaritmo da pressão de vapor, etc) e o nível do fator controlado pelo experimentador (concentração, inverso da temperatura, etc) é linear. Mas será que isto é sempre verdade? Como obter evidência objetiva de que a relação entre a propriedade medida e o fator controlado pelo pesquisador é realmente linear?

A resposta para essas perguntas poderá ser obtida apenas se os experimentos forem executados em replicatas, pois só assim os resultados fornecerão uma estimativa do erro experimental da propriedade que está sendo investigada. Se este erro for da mesma grandeza das diferenças entre os valores experimentais e aqueles previstos pelo modelo, podemos afirmar que a suposição sobre a linearidade está correta. Este procedimento é equivalente a fazer experimentos confirmatórios para testar um modelo. Mesmo assim é preciso fazer réplicas para determinar se os resultados confirmatórios estão dentro dos limites do erro experimental.

Na Tabela 1 encontram-se os valores da pressão de vapor de tetracloreto de carbono (CCl₄) para diferentes valores de temperatura (T)³.

Ensaio	Т (К)	p _{vap} (torr)
1	273	0,044
2	283	0,075
3	293	0,122
4	303	0,190
5	313	0,288
6	323	0,422
7	333	0,601
8	343	0,829
9	353	1,124

Tabela 1: Variação da pressão de vapor do CCl₄ com a temperatura

Se a entalpia de vaporização for constante e não depender da temperatura e a equação de Clausius – Clapeyron for validada nessas condições, o gráfico de ln p_{vap} vs. (1/T) será uma reta.

Uma regressão linear usando ln p_{vap} como variável dependente e (1/T) como a variável independente resultará na equação: ln $p_{vap} = 11,20-3901,98 \left(\frac{1}{T}\right)$, com R² = 0,9997.

Este resultado corresponde a um calor de vaporização de $32,44 \pm 0,22$ KJ.mol⁻¹.

A Figura 2 contém o gráfico dos valores previstos pelos observados, juntamente com o gráfico dos resíduos deixados pelo ajuste versus os valores previstos pelo modelo linear. Apesar da excelente concordância entre os valores observados e previstos e o alto valor de R², Figura 2a, há necessidade de incluir um termo quadrático no modelo apresentado, pois o gráfico dos resíduos, Figura 2b, deixa claro que eles não estão distribuídos aleatoriamente. Supondo que não existam erros sistemáticos nos resultados e que a execução dos experimentos foi feita em ordem aleatória, os resíduos não podem ser explicados como sendo devido ao erro experimental. Concluímos que o modelo linear é falho para representar os dados da Tabela 1.



Figura 2: Gráfico (a) dos valores previstos pelos valores observados e (b) dos resíduos deixados pelo ajuste do modelo linear.

Fazendo um novo ajuste aos dados e adicionando um termo quadrático temos uma nova equação:

$$\ln p_{vap} = 8,00 - 1954 \left(\frac{1}{T}\right) - 300311 \left(\frac{1}{T}\right)^2, \text{ com } \mathbb{R}^2 = 1,0000$$

A Figura 3 mostra o gráfico dos resíduos deixados pelo ajuste do modelo quadrático. Nessa figura os resíduos estão distribuídos aleatoriamente ao redor da linha no valor zero e não existe razão para suspeitar que existe falta de ajuste do modelo quadrático dos dados da Tabela 1.



Figura 3: Resíduos do modelo quadrático ajustado aos dados da Tabela 1.

Usando portanto o modelo quadrático e admitindo que a derivada $\frac{\partial (\ln p_{\text{vap}})}{\partial (\frac{1}{T})} = -1,945 - 600,622 (\frac{1}{T}) \acute{e}$

uma estimativa mais realista de $\frac{-\Delta H_{vap}}{R}$, podemos concluir que o calor de vaporização do CCl₄ na verdade varia entre 30,39 e 34,54 KJmol⁻¹, no intervalo de temperaturas consideradas.

Atualmente a verificação de linearidade da reta de calibração em química analítica é frequentemente analisada usando o coeficiente de correlação fornecido pelo programa computacional usado para fazer a regressão. Infelizmente não existe um teste estatístico que possa ser aplicado a este coeficiente para comprovar a linearidade da reta num dado nível de confiança. Cada vez mais esta sendo exigido a utilização do critério da falta de ajuste recomendado por Pimentel e Barros Neto⁴ e Danzar e Currie⁵.

6 Quando o coeficiente de correlação é zero, significa que não existe relação entre as duas variáveis?

Não. O coeficiente de correlação é limitado para investigar relações lineares entre as variáveis. Imagine, por exemplo, a relação entre a energia potencial de uma ligação química e sua distorção na geometria do equilíbrio. Os dados estão representados graficamente pelos pontos de uma curva anharmônica na Figura 4:



Figura 4: Relação entre a energia potencial de uma ligação química e sua distorção na geometria do equilíbrio.

O coeficiente de correlação linear é dado pela equação:

$$r(x, y) = \frac{1}{N-1} \sum \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i - \overline{y}}{s_y} \right)$$
(9)

onde, N é o número de pontos na Figura 4 e s_x e s_y são os desvios padrão das variáveis x e y.

Aplicando a equação da correlação para os pontos deste gráfico, observa-se que as contribuições dos dois pontos, (x_a , y_a) e (- x_a , y_a), na equação irão praticamente se cancelar. Este argumento é válido para qualquer par de pontos que tem o mesmo valor de y_i . De fato, se os pontos seguissem exatamente uma parábola, a aplicação da equação acima resultaria em um coeficiente de correlação igual a zero. Porém, como as ligações químicas não obedecem à risca a lei de Hooke, e por isso não vibram igual a um oscilador harmônico, o coeficiente de correlação dos dados representados na Figura 4 não será exatamente zero e sim, um coeficiente de correlação pequeno que não reflete o comportamento sistemático dos dados apresentados no gráfico.

7 Porque é importante examinar os gráficos ao invés de apenas calcular os parâmetros estatísticos?

Parâmetros estatísticos como a média, o desvio padrão e o coeficiente de correlação são representações numéricas de grande quantidade de dados. O coeficiente de correlação *r*, por exemplo, é um número que expressa a relação entre duas variáveis, obtido pela Equação 9. Na realidade, as relações entre variáveis são normalmente muito complexas para serem representadas por um único número.

A Figura 5 contém três gráficos de uma variável y plotada em função de uma variável x. A Figura 5a mostra uma relação evidentemente linear entre as variáveis x e y, porém, com grandes dispersões do modelo de regressão representado pela reta. A Figura 5b mostra um gráfico com pontos que apresentam uma relação não linear.

Observa-se que os quatro pontos na esquerda têm um arranjo linear, diferente dos pontos na direita que também apresentam uma distribuição quase linear, porém com uma inclinação diferente dos pontos da esquerda. A Figura 5c tem um arranjo de pontos completamente diferente, o ponto na direita tem um valor bem distinto dos pontos na esquerda.

Estes conjuntos de dados possuem algo em comum. Todos apresentam um coeficiente de correlação de Pearson igual a 0,87, porém, as situações físicas apresentadas nos gráficos são completamente diferentes. A Figura 5a mostra uma simples relação linear, a 5b uma interseção de dois modelos lineares, enquanto que a Figura 5c mostra a possível existência de um "outlier", ou seja, um ponto fora do padrão. Se este ponto for retirado do conjunto de dados o coeficiente de correlação dos pontos na esquerda cai para 0,27.



Figura 5: Três conjunto de dados com o mesmo coeficiente de correlação, r = 0,87, mas

8 Por que nas aulas de laboratório de química as determinações são feitas em triplicata?

A realização de experimentos em triplicatas é recomendada nos laboratórios de ensino porque é um compromisso aceitável entre a precisão e o trabalho. O valor médio da triplicata é a melhor estimativa do teor do analito na amostra, enquanto que o desvio padrão é a estimativa do erro experimental em uma determinação, sendo que o erro padrão no valor médio da triplicata é menor pelo fator de $1/\sqrt{3}$. Isto pode ser visto na equação que representa o intervalo de confiança do valor médio:

$$\mu = \overline{x} \pm t_{N-1} \frac{s}{\sqrt{N}} \tag{10}$$

onde, \overline{x} representa a média, *s* corresponde ao desvio padrão, *N* é o número de réplicas (três no caso da triplicata) e *t* é o valor crítico da distribuição *t* de Student com *N*-1 graus de liberdade.

Aumentando o valor de *N*, o intervalo de confiança irá diminuir por causa da diminuição do valor t_{N-1} e do fator $1/\sqrt{N}$. No nível de 95% de confiança, os valores de t_{N-1} são 12,71; 4,30; 3,18; 2,78 e 2,57 quando *N* vai de 2 para 6. Para o mesmo intervalo de *N*, a expressão $1/\sqrt{N}$ vai de 1,000; 0,707; 0,577; 0,500 até 0,408. Os produtos de t_{N-1} $1/\sqrt{N}$ ficam progressivamente menor, indo de 12,71; 3,04; 1,83; 1,39 até 1,13. Sendo assim, quando se realiza uma triplicata ao invés de uma duplicata, ocorre um melhoramento da precisão de 12,71 para 3,04 (fator de 4). Entretanto, realizando uma quadruplicata ao invés de uma triplicata, ocorre um melhoramento da precisão de 12,71 para 3,04 (fator de 4).

A Figura 6 apresenta um gráfico de t_{N-1}/\sqrt{N} vs. N ilustrando como a precisão relativa diminui quando se aumenta o número de replicas. Nota-se que a precisão é muito pouco melhorada com a execução de cinco ou mais replicatas.



Figura 6: Gráfico de t/\sqrt{N} contra o número de replicas *N* ilustrando como a precisão relativa diminui quando se aumenta o número de replicas.

9 Qual é a diferença entre a estatística paramétrica e a não paramétrica?

A grande maioria dos químicos utiliza métodos de estatística paramétrica para resolver problemas no laboratório como, por exemplo, determinar se um valor experimental é igual a um valor padrão, se os resultados provenientes de diferentes laboratórios são iguais, se o erro médio quadrático de validação de um método analítico é maior do que o erro de um outro método, etc. Nestes casos, considera-se que os dados seguem uma distribuição normal uma vez que a comparação de valores médios não é tão problemática porque estes valores tendem a seguir distribuições normais devido ao teorema do limite central, mesmo se os dados das determinações individuais não seguirem nenhuma distribuição conhecida. Por isto, planejamentos estatísticos de experimentos produzem resultados confiáveis. Como os valores dos parâmetros dos modelos associados a estes planejamentos são combinações lineares de resultados experimentais eles tendem a seguir uma distribuição normal.

Os métodos de estatística não paramétrica são usados quando o pesquisador precisa testar dados que não seguem uma distribuição bem caracterizada (normal ou não). Muitos métodos básicos de estatística paramétrica têm seu análogo não-paramétrico, como por exemplo, o "Wilcoxon rank sum test" é o equivalente não paramétrico do teste *t* para duas médias⁶. O teste *t* pareado paramétrico corresponde ao "Wilcoxon signed rank test". O "Kruskal-Wallis h test" é o equivalente não paramétrico da análise de variância (ANOVA) para dados obtidos em ordem completamente aleatória. Para dados obtidos com aleatorização por blocos pode ser usado o método não-paramétrico "Friedman-R". Também existe um coeficiente de correlação não-paramétrico chamado "Spearman Rank Coefficient of Correlation" análogo ao coeficiente de correlação de Pearson.

Todos estes métodos não paramétricos utilizam os postos dos dados em lugar dos dados em si. A Tabela 2 apresenta os valores e postos dos dados *x* e *y* correspondendo aos pontos no gráfico da Figura 5a. Os postos foram determinados colocando os dados em ordem crescente e atribuindo posto 1 para o menor valor, posto 2 o segundo menor, etc. Depois de determinar os postos para ambos os dados, *x* e *y*, calcula-se a diferença entre os postos que estão apresentados na última coluna *d*. O coeficiente de correlação é obtido por meio da equação 11,

$$r_{s} = 1 - \frac{6\sum d^{2}}{N(N^{2} - 1)}$$
(11)

onde *N* é o número de observações, neste caso 6 e o somatório de d^2 é igual a 4. Dessa forma temos um coeficiente de Spearman de 0,89. Nota-se que este resultado está em boa concordância com o valor do coeficiente de correlação de Pearson, 0,87, dado no item 6.

Número	X	Posto	у	Posto	d
1	4,3	1	4,0	1	0
2	5,0	2	5,5	3	-1
3	6,0	3	5,0	2	1
4	7,0	4	7,4	5	-1
5	7,9	5	6,3	4	1
6	8,5	6	8,4	6	0

Tabela 2: Dados para calcular o coeficiente de correlação de Spearman.

10 Métodos que utilizam o poder do computador para fazer cálculos podem resolver problemas que não são resolvidos usando estatística clássica?

Esta questão refere-se a aplicações para as quais não existem equações analíticas para resolver o problema de interesse. Por exemplo, seria possível saber se o valor mediano de um conjunto de resultados é estatisticamente igual ao valor mediano de um grupo de controle? Usando a estatística clássica a resposta para esta questão é não, pois não existem equações analíticas que possam ser utilizadas para determinar se os dois valores medianos são realmente diferentes ou se a diferença é uma mera flutuação estatística causada pelo erro experimental. Entretanto, é possível resolver este problema empregando a metodologia do "bootstrap"⁷. Para cada grupo, usa-se um computador para criar centenas ou até milhares de amostras do mesmo tamanho por amostragem aleatória com substituição. Por exemplo, para um valor mediano de seis resultados do laboratório (x₁, x₂, x₃, x₄, x₅, x₆) uma amostra "bootstrap" pode ser (x₃, x₅, x₁, x₆, x₅, x₂). Este procedimento é feito várias vezes para cada grupo e depois é construído um histograma das diferenças dos cálculos medianos, onde podem ser determinados intervalos no nível de 95% de confiança. O mesmo histograma serve para intervalos com outros níveis de confiança.

Referências Bibliográficas

- 1) Montgomery, D. C.; Design and Analysis of Experiments; 3^{*a*} ed, p.22, Wiley, New York, 1991.
- 2) Irvin, J. A.; Quickenden, T. I. J. Chem. Educ. *60*, (1983) 711.
- 3) Neto, B. B., Bruns, R. E., Scarminio, I. E.; Como fazer experimentos; 4^a ed, p. 260-263, Artmed, Porto Alegre, RS, 2010.
- 4) Pimentel, M. F., Barros Neto, B. de, Quim. Nova 19, (1996) 268.
- 5) Danzar, K., Currie, L. A., Pure & Appl. Chem. 70, (1998) 993-1014.

- 6) Wagner, S. F.; Introduction to Statistics; *Capítulo 14, Harper Collins, New York, 1992*.
- 7) Efron, B., Tibshirani, R. J.; An Introduction to the Bootstrap; *Chapman & Hall/CRC, Boca Raton,* 1994.