

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**ESTIMAÇÃO DOS PARÂMETROS DO MODELO F-SAC
UTILIZANDO UM ALGORITMO DE OTIMIZAÇÃO BASEADO EM
DERIVADAS**

TESE DE DOUTORADO

LUIZ FELIPE KUSLER POSSANI

**PORTO ALEGRE, RS
2019**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**ESTIMAÇÃO DOS PARÂMETROS DO MODELO F-SAC
UTILIZANDO UM ALGORITMO DE OTIMIZAÇÃO BASEADO EM
DERIVADAS**

LUIZ FELIPE KUSLER POSSANI

Tese de Doutorado apresentada como requisito parcial para obtenção do título de Doutor em Engenharia Química.

Área de Concentração: Pesquisa e Desenvolvimento de Processos

Orientador:
Prof. Rafael de Pelegrini Soares, D.Sc.

**PORTO ALEGRE, RS
2019**

Possani, Luiz Felipe K.

Estimação dos parâmetros do modelo F-SAC utilizando um algoritmo de otimização baseado em derivadas / Luiz Felipe Kusler Possani. -- 2019.

128 f.

Orientador: Rafael de Pelegrini Soares, D.Sc.

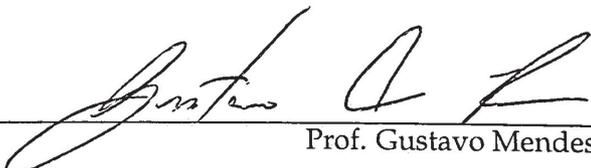
Tese (Doutorado) - Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Programa de Pós-Graduação em Engenharia Química, Porto Alegre, BR-RS, 2019.

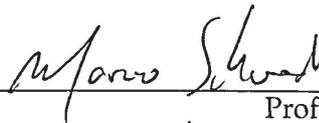
1. F-SAC. 2. Coeficiente de Atividade. 3. Estimação de Parâmetros. 4. Métodos Baseados em Derivadas. 5. Teorema da Função Implícita. I. Soares, Rafael de P., orient. II. Título.

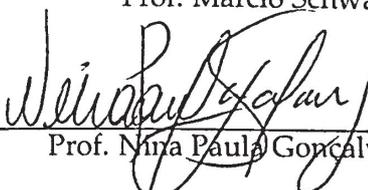
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

A Comissão Examinadora, abaixo assinada, aprova a Tese *Estimação dos parâmetros do modelo F-SAC utilizando um algoritmo de otimização baseado em derivadas*, elaborada por Luiz Felipe Kusler Possani como requisito parcial para obtenção do Grau de Doutor em Engenharia Química.

Comissão Examinadora:


Prof. Gustavo Mendes Platt, D.Sc.

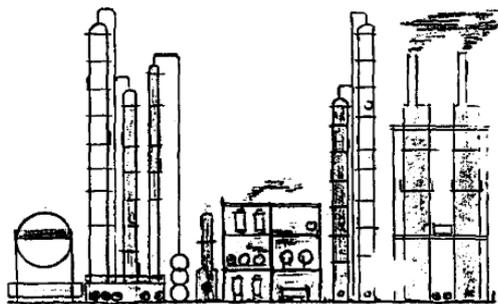

Prof. Marcio Schwaab, D.Sc.


Prof. Nina Paula Gonçalves Salau, D.Sc.

Green plants produce O₂, carbohydrates, fats, proteins from H₂O, CO₂ and minerals. Green plants are a demonstration of the success of evolution and we should look at them with humility and with joy.



A chemical plant is a technical monument invented and made by man. We should look at it with pride and with joy.



The chemical plant produces Plastics, Alcohol, Acid, Fertilizer synthetically from Air, Water and Fossil Fuel.

KNAPP, H. Physical properties in process design - past, present and future. Fluid Phase Equilib., v. 29, p. 1–21, 1986

Agradecimentos

Eu gostaria de agradecer a minha família, que tem me apoiado desde sempre; ao professor Rafael de Pelegrini Soares, pela dedicação e orientação; e a todos os colegas do LVPP.

Resumo

No desenvolvimento atual do modelo F-SAC, a estimação dos seus parâmetros é realizada com o uso de um clássico algoritmo de Nelder-Mead. Entretanto, à medida que o número de parâmetros a serem estimados aumenta, este método apresenta dificuldades de convergência. Assim, neste trabalho sugeriu-se a utilização de um algoritmo de otimização baseado nas derivadas da função objetivo para estimação dos parâmetros do modelo F-SAC. Isto demanda a obtenção de valores confiáveis das derivadas do modelo com relação aos seus parâmetros. Todavia, um método iterativo inerente ao F-SAC dificulta a sua derivação analítica. Logo, propõe-se a aplicação do Teorema da Função Implícita para a obtenção das derivadas implícitas do F-SAC, fácil e eficientemente, com relação a qualquer variável. O método proposto foi testado na derivação do modelo com relação à temperatura e ao número de mols de algum componente na mistura. As derivadas implícitas com respeito à temperatura foram comparadas com as respectivas derivadas numéricas, obtidas por diferenças finitas; os resultados exatos e numéricos apresentaram desvios na ordem de 10^{-5} . As derivadas implícitas com respeito ao número de mols foram inferidas por um teste de consistência de Gibbs-Duhem e comparadas com as derivadas numéricas; os valores de Gibbs-Duhem utilizando as derivadas numéricas foram inferiores a 10^{-2} e utilizando derivadas implícitas menores do que 10^{-5} . A partir disto, foi realizada a derivação do F-SAC com relação aos seus parâmetros. As derivadas implícitas foram comparadas com derivadas numéricas, para diversas misturas binárias, com diferentes composições, num intervalo de 273,15 K a 423,15 K; os resultados apresentaram um desvio médio de $1,083 \times 10^{-5}$. Os tempos computacionais para a obtenção das derivadas também foram comparados; as derivadas implícitas foram mais rápidas para todos os testes. Finalmente, as derivadas do modelo F-SAC com respeito aos seus parâmetros foram utilizadas em um método de otimização baseado em derivadas denominado MMA e disponível dentro do pacote computacional NLOpt. Para testar o MMA foram propostos 5 testes; em todos os testes realizados, o algoritmo convergiu para um mínimo local da função objetivo. Os resultados demonstraram que o MMA pode ser empregado na estimação de parâmetros do F-SAC. Além disto, foi realizado um estudo simplificado da influência do raio de contato padrão no comportamento do modelo. Espera-se simplificar a tarefa de estimação de parâmetros do modelo, e possibilitar a ampliação da tabela atual de parâmetros de maneira simples.

Palavras-chave: 1. F-SAC. 2. Coeficiente de Atividade. 3. Estimação de Parâmetros. 4. Métodos Baseados em Derivadas. 5. Teorema da Função Implícita.

Abstract

This requires, ultimately, reliable values of model's derivatives in relation to their parameters. However, an iterative method, inherent to F-SAC, difficults its analytical derivation. Therefore, the application of the Implicit Function Theorem is proposed in order to obtain FSAC's implicit derivatives, easily and efficiently, in relation to any variable. The proposed method was tested in the derivation of the model with respect to the temperature and the number of mols of some component in the mixture. The implicit derivatives with respect to temperature were compared with the respective numerical derivatives, obtained by finite differences; the implicit and numerical results showed deviations in the order of 10^{-5} . The implicit derivatives with respect to the number of moles were inferred by a Gibbs-Duhem consistency test and compared with numerical derivatives; the Gibbs-Duhem using the numerical derivatives were lower than 10^{-2} and the implicit derivatives were lower than 10^{-5} . From these results, the derivation of the F-SAC with respect to its parameters was performed. The implicit derivatives were compared with numerical derivatives for various binary mixtures with different compositions in a range of 273.15 K to 423.15 K; the implicit and numerical results showed an average deviation of 1.083×10^{-5} . The computational times to obtain the implicit and numerical derivatives were also compared; the implicit derivatives were faster for all tested systems. Finally, the derivatives of the F-SAC model with respect to its parameters were used in an derivative-based optimization method called MMA and available within the NLOpt computational package. To test the MMA, 5 tests were proposed; in all the tests performed, the algorithm converged to a local minimum of the objective function. The results demonstrated that the MMA can be used in the estimation of parameters of the F-SAC model. Besides, the influence of the standard contact radius on the model was studied. It is expected to simplify the model's parameters estimation task, as well as to allow the extension of the current table of parameters in a simple way.

Key-words: 1. F-SAC. 2. Activity Coefficient. 3. Parameters Estimation. 4. Derivative-based Methods. 5. The Implicit Function Theorem.

Sumário

Lista de Figuras	xv
Lista de Tabelas	xvi
Lista de Símbolos	xviii
1 Introdução	1
1.1 Motivação e Objetivos	1
1.2 Estrutura do Trabalho	5
2 Revisão Bibliográfica	7
2.1 Termodinâmica das Soluções	7
2.1.1 Propriedades Parciais Molares e Propriedades de Mistura	7
2.1.2 Potencial Químico e Coeficiente de Fugacidade	10
2.1.3 Propriedades em Excesso e Coeficiente de Atividade	12
2.1.4 Modelos de Coeficiente de Atividade	14
2.2 Modelos de Coeficiente de Atividade Baseados em COSMO	15
2.2.1 Introdução	15
2.2.2 Equacionamento do Modelo F-SAC	20
2.2.3 Resolução da Equação Auto-Consistente	26
2.3 Estimação de Parâmetros	30
2.3.1 Problema de Otimização	30
2.3.2 Dados Experimentais	33
2.4 Métodos de Otimização	34
2.4.1 Introdução	34
2.4.2 Nelder-Mead	37
2.4.3 Interior Point OPTimizer (Ipopt)	38
2.4.4 NonLinear optimization (NLOpt)	41
2.5 Obtenção das Derivadas e o Teorema da Função Implícita	43
3 Metodologia	47
3.1 Descrição do Problema de Otimização	47
3.2 Derivada da Função Objetivo	50
3.3 Derivada do Modelo F-SAC	50
3.3.1 Introdução	50
3.3.2 Equacionamento Geral	52
3.3.3 Limites e Restrições	54

3.3.4	Derivada com Relação à Temperatura	55
3.3.5	Derivada com Relação ao Número de Mols	56
3.3.6	Derivada com Relação à Área Total de um Subgrupo	58
3.3.7	Derivada com Relação às Áreas Positivas e Negativas de um Grupo	59
3.3.8	Derivada com Relação à Carga Positiva de um Grupo	61
3.3.9	Derivada com Relação à Ligação de Hidrogênio	62
3.3.10	Obtenção dos Intervalos de Confiança dos Parâmetros Estimados	62
4	Resultados e Discussão	65
4.1	Resultados da Derivação	65
4.1.1	Introdução	65
4.1.2	Derivadas com Relação à Temperatura	66
4.1.3	Derivadas com Relação ao Número de Mols – Aplicação da Equação de Gibbs-Duhem	67
4.1.4	Derivadas com Relação aos Parâmetros do Modelo F-SAC	69
4.1.5	Tempos Computacionais	70
4.2	Resultados das Otimizações	72
4.2.1	Introdução	72
4.2.2	Descrição dos Testes	72
4.2.3	Resultados dos Testes	73
4.3	Efeito do Raio de Contato Padrão no Modelo F-SAC	77
5	Conclusões e Trabalhos Futuros	81
5.1	Conclusões	81
5.2	Trabalhos Futuros	83
5.3	Principais Publicações	84
	Referências Bibliográficas	85
A	Tabelas utilizadas nas Estimações de Parâmetros	89
B	Tabelas Resultantes das Estimações de Parâmetros	97

Lista de Figuras

Figura 2.1	Cargas induzidas nas superfícies das moléculas de benzeno e acetona, de acordo com a teoria COSMO.	15
Figura 2.2	Possível contato entre as moléculas de benzeno e acetona, de acordo com a teoria COSMO.	17
Figura 2.3	Centros dos segmentos de carga superficial (esferas) para algumas moléculas, de acordo com a metodologia COSMO.	18
Figura 2.4	Perfis- σ contínuos das moléculas de n-hexano e acetona.	19
Figura 2.5	Perfis- σ das moléculas de n-hexano e acetona com $\Delta\sigma^{51}=0,001 e/\text{\AA}^2$ e $\Delta\sigma^{11}=0,005 e/\text{\AA}^2$.	20
Figura 2.6	Perfil- σ para o benzeno calculado com o COSMO e ajustado para o F-SAC.	23
Figura 2.7	Representação COSMO para acetona, clorofórmio e etanol.	26
Figura 2.8	Fluxograma da substituição sucessiva.	28
Figura 2.9	Tela inicial do F-SAC <i>Optimizer</i> . Extraído de Possani (2014).	36
Figura 3.1	Esquema do programa implementado.	49
Figura 4.1	Logaritmo natural do IDAC experimental <i>versus</i> o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.	75
Figura 4.2	Logaritmo natural do IDAC experimental <i>versus</i> o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.	75
Figura 4.3	Logaritmo natural do IDAC experimental <i>versus</i> o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.	76
Figura 4.4	Logaritmo natural do IDAC experimental <i>versus</i> o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.	76
Figura 4.5	Logaritmo natural do IDAC experimental <i>versus</i> o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.	77
Figura 4.6	Valores da função objetivo depois da estimação de parâmetros <i>versus</i> r_{eff} .	79

Lista de Tabelas

Tabela 3.1	Valores da derivada da área do segmento m com relação à $Q_k^{+/-}$	60
Tabela 3.2	Valores da derivada da carga do segmento m caso este seja positivo, negativo ou neutro.	61
Tabela 4.1	Desvios Médios Absolutos (AAD_{hE}) e Maiores Desvios ($ h_{i,implcito}^E - h_{i,numérico}^E $).	66
Tabela 4.2	Resultados de Gibbs-Duhem, calculado pela Equação 4.4, para o sistema acetato de metila + água, a 330,5 K.	68
Tabela 4.3	Sistemas escolhidos (Componente 1 + Componente 2).	70
Tabela 4.4	Tempos computacionais para calcular as derivadas implícitas e numéricas para os sistemas testados.	71
Tabela 4.5	Valores da FO e do R^2 para os Testes 0 ao 4 antes e após as estimações de parâmetros.	77
Tabela 4.6	Valores da FO e do R^2 para o Teste 0 antes e após as estimações de parâmetros com firentes valores de r_{eff}	78
Tabela A.1	Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 0.	89
Tabela A.2	Parâmetros de interação binária para ligações de hidrogênio utilizados no Teste 0.	90
Tabela A.3	Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 1. Observação: apenas a área positiva do grupo CH ₃ OH foi multiplicada por 1,05 para não exceder o <i>bound</i> da carga negativa ($\sigma_k^- \geq -\sigma_k^{max}$).	91
Tabela A.4	Parâmetros de interação binária para ligações de hidrogênio utilizados no Teste 1.	92
Tabela A.5	Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 2. Observação: as áreas positivas dos grupos CH ₂ OH e N(CH ₂) ₃ não foram modificadas de modo que suas áreas sejam maiores do que a área de ligação de hidrogênio ($Q_k^+ \geq n_{acc} \cdot a_{eff}$).	92
Tabela A.6	Parâmetros de interação binária para ligações de hidrogênio utilizados no Teste 2.	93
Tabela A.7	Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 3.	94
Tabela A.8	Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 4.	95
Tabela B.1	Valores dos parâmetros dos grupos resultantes do Teste 0.	97
Tabela B.2	Valores dos parâmetros dos subgrupos resultantes do Teste 0.	97

Tabela B.3	Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 0.	98
Tabela B.4	Valores dos parâmetros dos grupos resultantes do Teste 1.	100
Tabela B.5	Valores dos parâmetros dos subgrupos resultantes do Teste 1.	100
Tabela B.6	Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 1.	101
Tabela B.7	Valores dos parâmetros dos grupos resultantes do Teste 2.	102
Tabela B.8	Valores dos parâmetros dos subgrupos resultantes do Teste 2.	102
Tabela B.9	Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 2.	103
Tabela B.10	Valores dos parâmetros dos grupos resultantes do Teste 3.	104
Tabela B.11	Valores dos parâmetros dos subgrupos resultantes do Teste 3.	104
Tabela B.12	Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 3.	105
Tabela B.13	Valores dos parâmetros dos grupos resultantes do Teste 4.	106
Tabela B.14	Valores dos parâmetros dos subgrupos resultantes do Teste 4.	107
Tabela B.15	Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 4.	108

Capítulo 1

Introdução

1.1 Motivação e Objetivos

A concepção, operação eficiente e otimização de um processo industrial, frequentemente, demanda dos engenheiros projetistas o cálculo ou estimativa do equilíbrio de fases de substâncias em mistura (FREDENSLUND et al., 1975). Por exemplo, a informação sobre o quanto de uma dada espécie A está presente em *cada fase em equilíbrio*, numa mistura de N componentes, é de suma importância.

De maneira geral, estas informações podem ser obtidas por via experimental, entretanto é conveniente que seja possível obtê-las a partir da predição de modelos matemáticos, por questões de tempo e custo envolvidos (HSIEH et al., 2014; GRENSEMANN; GMEHLING, 2005a). Quanto ao tipo de modelo, em princípio o equilíbrio de fases poderia ser calculado diretamente utilizando equações de estado, porém em muitos casos, o uso da quantidade conhecida como coeficiente de atividade da fase líquida é mais interessante.

O coeficiente de atividade representa a não idealidade de uma dada substância quando numa mistura, em uma temperatura e composição específicas no líquido. Ou seja, esta quantidade *mede* o efeito da interação entre uma molécula do tipo A quando posta em contato (misturada) com uma molécula do tipo B , por exemplo. Desta forma, quando esta abordagem é utilizada, o projeto de algum processo industrial dependerá fundamentalmente da existência de um modelo adequado que possa calcular o coeficiente de atividade de cada uma das substâncias que formam a mistura.

Os primeiros modelos de coeficiente de atividade foram desenvolvidos há bastante tempo (Margules é de 1895 e van Laar de 1910), sendo, de maneira geral, muito simples. Estes modelos podem ser totalmente empíricos, por exemplo Margules (dois ou três sufixos), ou apresentar alguma fundamentação teórica, como van Laar, Wilson (WILSON, 1964), NRTL (RENON; PRAUSNITZ, 1968) e UNIQUAC (ABRAMS; PRAUSNITZ, 1975). Todos os modelos citados anteriormente são ditos *não preditivos*, ou seja, devem ter seus parâmetros ajustados com base em dados experimentais de cada par de substâncias.

Em contrapartida, um modelo é dito *preditivo* quando pode ser extrapolado para misturas não incluídas no seu ajuste. Dentre os modelos preditivos de coeficiente de atividade, o modelo UNIFAC (UNIQUAC *Functional-group Activity Coefficients*) (FREDENSLUND et al., 1975) e suas variantes (GREENSEMANN; GMEHLING, 2005b; XUE et al., 2012) têm sido os mais aplicados na prática, inclusive até os dias de hoje. Entretanto, estes modelos continuam dependendo de parâmetros binários, que são ajustados utilizando dados experimentais para cada par de grupos, o que gera uma tabela de parâmetros com enormes proporções.

Desta forma, outra classe de modelos preditivos que tem se tornado uma alternativa interessante para cálculos de equilíbrio de fases são os modelos baseados na metodologia COSMO (*COnductor-like Screening MOdel*), como COSMO-RS (KLAMT, 1995) e suas variações – e.g. COSMO-SAC (LIN; SANDLER, 2002), COSMO-RS(OI) (GREENSEMANN; GMEHLING, 2005b), F-SAC (SOARES; GERBER, 2013; SOARES et al., 2013) e $\sigma - MTC$ (COSTA et al., 2016a; COSTA et al., 2016b).

A teoria COSMO foi formulada por Klamt e Schuurmann (1993) com o objetivo de descrever matematicamente os efeitos de diferentes solventes e suas energias de solvatação. A partir desta metodologia, Klamt (1995) desenvolveu o primeiro modelo de coeficiente de atividade baseado em COSMO, conhecido por COSMO-RS (*Realistic Solvation*)

Nos modelos baseados em COSMO, geralmente apenas um pequeno conjunto de parâmetros universais são ajustados, então é possível dizer que tais modelos dependem apenas do cálculo COSMO de cada molécula pura. Entretanto, apesar de apresentarem uma formulação mais elaborada, os modelos baseados em COSMO pre-

sentes na literatura recente ainda não apresentam a precisão necessária para cálculos de engenharia (KLAMT et al., 2010), sendo geralmente necessária a introdução de parâmetros empíricos para melhor descrever os sistemas (KLAMT; ECKERT, 2000; GRENSEMANN; GMEHLING, 2005b; KLAMT et al., 2010; GERBER; SOARES, 2013).

Com base no COSMO-SAC de Lin e Sandler (2002), o modelo preditivo F-SAC (*Functional Segment Activity Coefficient*) foi recentemente desenvolvido em nosso grupo de pesquisa (GERBER, 2012; SOARES; GERBER, 2013; SOARES et al., 2013). Este modelo combina os conceitos de contribuição de grupos e teoria de superfícies de contatos. As moléculas são divididas em grupos e subgrupos, de forma bastante similar ao UNIFAC, entretanto a formulação utilizada para o cálculo de interação entre os grupos é dada por um equacionamento semelhante ao do modelo COSMO-SAC. Além disso, os parâmetros para os grupos são estimados a partir de dados experimentais.

Em sequência ao seu desenvolvimento, o modelo F-SAC também foi estudado em diversos trabalhos no nosso grupo de pesquisa, como por exemplo Possani et al. (2014), Possani et al. (2014), Possani (2014), Flôres et al. (2016), Possani et al. (2016). Em todos estes trabalhos, a *estimação de parâmetros* foi uma etapa crucial, sendo sempre realizada com o mesmo algoritmo de otimização Nelder-Mead, conforme descrito em Soares e Gerber (2013).

O Nelder-Mead foi escolhido por ser um método simples e que não requer derivadas, uma vez que o modelo F-SAC possui uma formulação não linear e apresenta um método iterativo interno. As restrições do modelo F-SAC foram contornadas através da utilização de uma “penalidade”, de modo que o método retornava um valor alto de função objetivo num ponto em que houvesse a violação de alguma restrição. Entretanto, observou-se que, à medida que o número de parâmetros a serem estimados aumenta, o algoritmo tende a apresentar dificuldades na obtenção da solução, não satisfazendo o critério de convergência tipicamente utilizado, mesmo permitindo-se um máximo número de iterações na ordem de milhares.

Por exemplo, em Soares e Gerber (2013), Soares et al. (2013) um total de 142 parâmetros foram estimados; na Dissertação de Mestrado do autor deste trabalho (POSSANI, 2014), estimou-se 42 parâmetros. Em ambos trabalhos o Nelder-Mead apresentou dificuldades na resolução do problema de otimização (lentidão na convergência).

Além disto, testou-se o algoritmo sem e com a normalização dos parâmetros e a mesma lentidão foi observada em todos os testes. Desta forma, não foi possível estimar todos os parâmetros com todos os dados experimentais disponíveis de uma só vez.

As possíveis causas para as dificuldades de convergência observadas são o próprio algoritmo utilizado, que é pouco eficiente para resolver problemas muito grandes, além da complexidade do problema. A ineficiência do Nelder-Mead para resolver problemas muito grandes é conhecida. Han e Neumann (2006), por exemplo, estudaram o efeito da dimensionalidade do Nelder-Mead para uma função quadrática. Os autores demonstraram, matematicamente e através de experimentos numéricos, que, com o aumento do número de variáveis de otimização, o algoritmo tende a progredir cada vez mais lentamente.

Han e Neumann (2006) não especificaram qual o tamanho do problema de otimização que seria considerado “grande”, pois este tamanho vai depender do problema apresentado. Entretanto, os autores sugerem que, para problemas mais complexos, espera-se uma maior lentidão na convergência. Tipicamente, para a estimação de parâmetros do modelo F-SAC, o Nelder-Mead começa a apresentar elevada lentidão a partir de 10 parâmetros a serem estimados. Para superar as dificuldades encontradas, a estimação de parâmetros geralmente é realizada em partes. Ou seja, para um conjunto menor de dados, estima-se uma parte dos parâmetros, mantendo-se os outros fixos. Depois fixa-se os valores encontrados destes parâmetros e estima-se os parâmetros restantes, com outro conjunto de dados experimentais, e segue-se refinando o modelo.

Após a obtenção dos parâmetros com a metodologia descrita, foi testada a estimação de todos os parâmetros simultaneamente. Entretanto, mesmo partindo-se de um resultado já refinado, o Nelder-Mead não convergiu. Em todos os trabalhos citados anteriormente, esta metodologia foi utilizada, apesar de muitas vezes não ser explicitada. Assim, o objetivo principal do trabalho de Doutorado é estudar o uso de um algoritmo de otimização alternativo ao Nelder-Mead, também do tipo local, para estimação dos parâmetros do modelo F-SAC.

Assim, neste trabalho propõe-se a utilização de um método de otimização baseado em derivadas da função objetivo, o que demanda, em última instância, a obtenção

de valores confiáveis das derivadas do modelo F-SAC com relação aos seus parâmetros. Com o uso de um método de otimização deste tipo, espera-se simplificar a tarefa de estimação de parâmetros do modelo, atualmente trabalhosa, bem como possibilitar a ampliação da tabela atual de parâmetros, de maneira simples.

Para tanto, um esforço foi colocado na *derivação analítica* das equações que compõem o modelo F-SAC, de modo a se obter os valores exatos das derivadas do modelo com respeito aos seus parâmetros. Quanto ao método iterativo inerente ao F-SAC, a aplicação do *Teorema da Função Implícita* permitiu a obtenção de suas derivadas, denominadas neste trabalho de *derivadas implícitas*. A vantagem deste método é a obtenção das derivadas de maneira fácil e eficiente, sem a necessidade de se utilizar de perturbação numérica da função objetivo e também sem necessitar da introdução de mais um método numérico para resolver o modelo com derivadas, conforme será detalhado nos próximos capítulos.

1.2 Estrutura do Trabalho

Este trabalho encontra-se organizado da seguinte maneira:

Neste capítulo (**Capítulo 1**) foram introduzidos os assuntos que serão estudados ao longo do trabalho, bem como os conceitos utilizados.

No **Capítulo 2** serão retomados alguns conceitos básicos da Termodinâmica. O equacionamento do modelo F-SAC será apresentado, com uma nomenclatura mais simplificada do que a nomenclatura comumente encontrada na literatura. Além disto, serão apresentados os conceitos necessários para a estimação de parâmetros e os métodos de otimização estudados. Será também introduzido, de maneira simplificada, o Teorema da Função Implícita e como este pode ser aplicado para obtenção de derivadas do modelo F-SAC, sem a necessidade de perturbações numéricas.

No **Capítulo 3**, será apresentada a Função Objetivo a ser utilizada, bem como o equacionamento utilizado na obtenção das derivadas do modelo F-SAC. No **Capítulo 4** serão apresentados e discutidos os principais resultados obtidos.

O trabalho será sumarizado no Capítulo 5, juntamente com as principais conclusões e sugestões de trabalhos futuros.

Capítulo 2

Revisão Bibliográfica

2.1 Termodinâmica das Soluções

2.1.1 Propriedades Parciais Molares e Propriedades de Mistura

A regra das fases de Gibbs¹ estabelece que, fixadas duas propriedades intensivas² quaisquer para uma substância pura em apenas uma fase, todas as demais propriedades intensivas ficam especificadas. Se, adicionalmente, o número total de mols (ou seja, o tamanho) do sistema também for especificado, é possível a obtenção das propriedades extensivas.

Analogamente, estendendo-se a regra das fases para misturas, se, além de especificar duas propriedades intensivas independentes, o número de mols de cada espécie na mistura também for fixado, todas as outras propriedades extensivas do sistema também estarão determinadas. Desta forma, uma propriedade extensiva qualquer, K , pode ser escrita de maneira geral da seguinte forma em função, por exemplo, da temperatura (T), pressão (P) e do vetor de número de mols (\mathbf{n}):

$$K = f(T, P, \mathbf{n}) \quad (2.1)$$

onde $\mathbf{n} = [n_1, n_2, \dots, n_N]$, numa mistura de N componentes. Nota-se que poderiam ter sido fixadas quaisquer propriedades, entretanto, por tratar-se de equilíbrio (térmico,

¹Regra das fases de Gibbs (KORETSKY, 2012): $\tau = N - \pi + 2$, onde τ é o número de graus de liberdade, N o número de substâncias e π o número de fases presentes.

²Propriedades intensivas são aquelas que não dependem do *tamanho* do sistema considerado. Propriedades extensivas dependem do tamanho do sistema.

mecânico e químico)³ é conveniente escrever-se K como função de T e P ⁴:

O diferencial total de K pode ser escrito como a soma das derivadas parciais com respeito a cada uma das variáveis independentes, da seguinte forma:

$$dK = \left(\frac{\partial K}{\partial T}\right)_{P,n} dT + \left(\frac{\partial K}{\partial P}\right)_{T,n} dP + \sum_{i=1}^N \left(\frac{\partial K}{\partial n_i}\right)_{P,T,n_{j \neq i}} dn_i \quad (2.2)$$

Onde as propriedades que aparecem nos subscritos das derivadas indicam que as respectivas derivações foram realizadas mantendo-se tais propriedades *constantes*. Em T e P constantes, tem-se:

$$dK = \sum_{i=1}^N \left(\frac{\partial K}{\partial n_i}\right)_{P,T,n_{j \neq i}} dn_i \quad (2.3)$$

Neste ponto, é interessante introduzir o conceito de propriedade parcial molar \bar{K}_i :

$$\bar{K}_i \equiv \left(\frac{\partial K}{\partial n_i}\right)_{P,T,n_{j \neq i}} \quad (2.4)$$

Desta forma, uma propriedade parcial molar \bar{K}_i é definida como a derivada da propriedade K com respeito ao número de mols de um determinado componente i , mantendo-se constantes a temperatura, pressão e o número de mols dos outros componentes na mistura.

Substituindo esta definição na equação [Equação 2.3](#):

$$dK = \sum_{i=1}^N \bar{K}_i dn_i \quad (2.5)$$

Fixando-se ainda a composição total da mistura e integrando, com a constante de integração igual a zero⁵, chega-se em:

$$K = \sum_{i=1}^N \bar{K}_i n_i \quad (2.6)$$

³É bastante conveniente a escolha de T e P , tendo-se em vista os equilíbrios térmico e mecânico. Quanto ao equilíbrio químico, este está relacionado ao mínimo da energia de Gibbs, ou ainda às igualdades de potenciais químicos das espécies nas fases em equilíbrio, conforme será visto a seguir.

⁴Os critérios de equilíbrio de fases são trabalhados de maneira objetiva na introdução do Capítulo 6 do [Koretsky \(2012\)](#).

⁵No capítulo 6 do [Koretsky \(2012\)](#) é dada uma interpretação física junto a um desenvolvimento matemático rigoroso provando que esta quantidade é, de fato, igual a zero.

Fisicamente, pode-se interpretar a propriedade parcial molar \bar{K}_i como sendo a contribuição da espécie i em solução para a propriedade total da mistura, K .

Observa-se que a Equação 2.6 também pode ser escrita para uma propriedade na sua forma intensiva, somente mudando-se o número de mols pela fração molar (x_i), da seguinte forma: $k = \sum_{i=1}^N \bar{K}_i x_i$. Entretanto, é importante salientar que a derivada parcial nas Equações 2.2 à 2.4 deve ser *sempre* escrita com respeito ao número de mols das substâncias, tendo em vista que as frações molares não são independentes entre si. Por exemplo, ao alterar -se a fração molar de um dos componentes, as outras também se alteram de modo que a relação $\sum_{i=1}^N x_i = 1$ deve ser respeitada. Além disso, um limite interessante ocorre quando a substância i está pura, neste caso: $\lim_{x_i \rightarrow 1} \bar{K}_i = k_i$.

Derivando-se a equação Equação 2.6, à T e P constantes, tem-se:

$$dK = \sum_{i=1}^N \bar{K}_i dn_i + \sum_{i=1}^N n_i d\bar{K}_i \quad (2.7)$$

E, finalmente, para que as Equações 2.5 e 2.7 sejam equivalentes chega-se em:

$$\sum_{i=1}^N n_i d\bar{K}_i = 0 \quad (T \text{ e } P \text{ constantes}) \quad (2.8)$$

Esta equação é denominada equação de Gibbs-Duhem. Apesar da sua simplicidade, a equação de Gibbs-Duhem fornece uma poderosa ferramenta no estudo das relações e limitações entre as propriedades parciais molares de diferentes espécies numa mistura⁶.

A partir do que foi apresentado anteriormente, é fácil verificar que, de maneira geral, as propriedades de uma mistura não são calculadas simplesmente como as somas das propriedades individuais de cada componente puro. Desta forma, surgem as propriedades de mistura ΔK^M , que *medem* a diferença entre a propriedade total da mistura e a soma das propriedades das espécies puras multiplicadas pelos respectivos números de mols, ou seja:

$$\Delta K^M = K - \sum_{i=1}^N K_i = K - \sum_{i=1}^N k_i n_i \quad (2.9)$$

⁶Para um exemplo, veja o cenário apresentado em Koretsky (2012) p. 341), no qual se calcula o volume parcial molar de um componente numa mistura binária conhecendo-se o volume parcial molar do outro componente.

Substituindo a [Equação 2.6](#) nesta equação, tem-se:

$$\Delta K^M = \sum_{i=1}^N n_i (\bar{K}_i - k_i) \quad (2.10)$$

Analogamente, para propriedades intensivas:

$$\Delta k^M = \sum_{i=1}^N x_i (\bar{K}_i - k_i) \quad (2.11)$$

Desta forma, uma propriedade qualquer de mistura pode ser interpretada como *o quanto* esta propriedade se altera como resultado de um caminho hipotético em que as espécies puras passam por um processo de mistura isotérmico e isobárico.

Em nível molecular, a mudança de propriedade da mistura reflete em como as interações entre as espécies diferentes na mistura se comparam às interações semelhantes das espécies puras que substituíram. No caso especial em que as propriedades parciais molares de cada substância se igualam às propriedades das substâncias puras, $\Delta K^M = 0$.

2.1.2 Potencial Químico e Coeficiente de Fugacidade

Partindo-se da definição de propriedade parcial molar, o potencial químico (μ_i) é definido como a energia de Gibbs parcial molar de uma substância i :

$$\mu_i \equiv \left(\frac{\partial n^T g}{\partial n_i} \right)_{T,P,n \neq n_i} \quad (2.12)$$

onde $n^T = \sum_{i=1}^N n_i$ é o número total de mols e g é a energia de Gibbs na sua forma intensiva.

Classicamente, a condição *suficiente* de equilíbrio para um sistema fechado contendo m componentes, numa determinada temperatura e pressão, dá-se no mínimo global da energia de Gibbs sobre todos os possíveis estados. Aplicando-se esta definição na [Equação 2.3](#) (considerando equilíbrio térmico e mecânico, T e P constantes) para um sistema fechado com duas fases quaisquer em equilíbrio α e β , tem-se:

$$dG_{T,P} = \sum_{i=1}^m [\bar{G}_i dn_i]^\alpha + [\bar{G}_i dn_i]^\beta = 0 \quad (2.13)$$

Utilizando a definição de potencial químico:

$$\sum_{i=1}^m \mu_i^\alpha dn_i^\alpha + \mu_i^\beta dn_i^\beta = 0 \quad (2.14)$$

Entretanto, num sistema fechado o número de mols que sai de α vai para β , de modo que $dn_i^\alpha = -dn_i^\beta$, logo:

$$\sum_{i=1}^m [\mu_i^\alpha - \mu_i^\beta] dn_i^\alpha = 0 \quad (2.15)$$

Finalmente, para que a [Equação 2.15](#) seja satisfeita, chega-se que, no equilíbrio de fases, os potenciais químicos de cada espécie em cada fase em equilíbrio são iguais. Generalizando para um sistema contendo π fases em equilíbrio:

$$\mu_i^\alpha = \mu_i^\beta = \dots = \mu_i^\pi \quad (i = 1, 2, \dots, N) \quad (2.16)$$

Para uma temperatura constante, pode-se definir a fugacidade de uma espécie i em solução, \hat{f}_i , em termos de uma referência arbitrária de potencial químico μ_i° , da seguinte forma:

$$\mu_i - \mu_i^\circ = RT \ln \frac{\hat{f}_i}{\hat{f}_i^\circ} \quad (2.17)$$

onde \hat{f}_i° é a fugacidade no estado de referência e R é a constante dos gases. Nota-se que se a substância for pura, a fugacidade em mistura \hat{f}_i se torna simplesmente f_i e o potencial químico μ_i reduz-se a g_i , conforme a sua própria definição.

Tomando-se como referência o Gás Ideal (GI) na mesma temperatura e pressão da mistura, tem-se⁷:

$$\mu_i - \mu_i^{\text{GI}} = RT \ln \frac{\hat{f}_i}{\hat{f}_i^{\text{GI}}} = RT \ln \frac{\hat{f}_i}{P y_i} \quad (2.18)$$

onde y_i é a fração molar de i na fase vapor⁸.

A razão dentro do logaritmo natural aparece com tanta frequência na Termodinâmica de Misturas que esta ganhou um nome especial: coeficiente de fugacidade $\hat{\phi}_i$

⁷Esta equação foi obtida por G. N. Lewis ao integrar a equação que define a energia de Gibbs ($dg = vdP - sdT$), à T constante para o Gás Ideal. Apesar de ser obtida para o GI, esta equação é geral, ou seja, também é válida para casos não ideais, podendo ser utilizada, inclusive, para as fases líquida e sólida. A dedução e demais explicações podem ser encontradas em [Koretsky \(2012\)](#), pags. 366–367 e 392–394.

⁸Embora tenha sido utilizado x_i anteriormente para representar fração molar, geralmente y_i aparece quando nos referimos à fase vapor. Já x_i aparece mais comumente nas definições envolvendo a fase líquida.

de uma substância i em uma mistura, definido como a razão entre a fugacidade da substância e a fugacidade desta se ela fosse considerada um gás ideal ($\hat{f}_i^{\text{GI}} = Py_i$):

$$\hat{\phi}_i \equiv \frac{\hat{f}_i}{\hat{f}_i^{\text{GI}}} = \frac{\hat{f}_i}{Py_i} \quad (2.19)$$

2.1.3 Propriedades em Excesso e Coeficiente de Atividade

Na seção anterior, foram introduzidos os conceitos de fugacidade e coeficiente de fugacidade, sendo estes utilizados geralmente em combinação com uma equação de estado, tendo como estado de referência um GI. Para a fase líquida, muitas vezes é interessante a introdução de um outro estado de referência, que seja análogo ao desenvolvimento que utiliza o modelo do GI como referência. Entretanto, não é possível extrapolar a fase líquida para um estado ideal no qual não há interações entre as moléculas, visto que, para que haja fase condensada, as moléculas precisam necessariamente interagir entre si. Desta forma, como referência para a fase líquida, geralmente utiliza-se o conceito de *solução ideal* (ID).

Para utilizar-se a Equação 2.17, é interessante definir, primeiro, o conceito de propriedade em excesso. Uma propriedade em excesso, k^{E} , é definida como a diferença entre a propriedade termodinâmica numa solução real (isto é, não ideal) e ideal nas mesmas condições de temperatura, pressão e composição. Ou seja, para o caso da energia de Gibbs:

$$g^{\text{E}} \equiv g - g^{\text{ID}} \quad (2.20)$$

onde g^{E} e g^{ID} são as energias de Gibbs em excesso e de uma solução ideal nas suas formas intensivas, respectivamente.

Tomando-se como referência a solução ideal, na mesma temperatura e pressão da mistura, tem-se:

$$\mu_i - \mu_i^{\text{ID}} = \mu^{\text{E}} = RT \ln \frac{\hat{f}_i}{\hat{f}_i^{\text{ID}}} \quad (2.21)$$

Desta forma, define-se o coeficiente de atividade γ_i de um determinado componente i numa mistura como a razão entre a sua fugacidade numa solução *real* e em uma

solução *ideal* ($\hat{f}_i^{\text{ID}} = x_i f_i$)⁹:

$$\gamma_i \equiv \frac{\hat{f}_i}{\hat{f}_i^{\text{ID}}} = \frac{\hat{f}_i}{x_i f_i} \quad (2.22)$$

onde x_i é a fração molar da espécie i na solução; \hat{f}_i e \hat{f}_i^{ID} são as fugacidades de i numa solução real e ideal, respectivamente; e f_i é a fugacidade de i puro.

Com isso, o coeficiente de atividade de um componente i relaciona-se com a sua energia de Gibbs parcial molar em excesso (\bar{G}_i^{E}) por:

$$\bar{G}_i^{\text{E}} = \left(\frac{\partial n^{\text{T}} g^{\text{E}}}{\partial n_i} \right)_{T, P, \mathbf{n} \neq n_i} = RT \ln \gamma_i \quad (2.23)$$

Utilizando a [Equação 2.4](#):

$$G^{\text{E}} = RT \sum_{i=1}^N n_i \ln \gamma_i \quad (\text{na forma extensiva}) \quad (2.24a)$$

$$g^{\text{E}} = RT \sum_{i=1}^N x_i \ln \gamma_i \quad (\text{na forma intensiva}) \quad (2.24b)$$

Para obtenção de outras propriedades termodinâmicas em excesso, pode-se partir da relação clássica da energia de Gibbs para o caso em excesso, $dg^{\text{E}} = v^{\text{E}} dP - s^{\text{E}} dT$. Diferenciando-se expressão anterior com relação à temperatura, mantendo-se constantes a pressão e o número de mols:

$$\left(\frac{\partial g^{\text{E}}}{\partial T} \right)_{P, \mathbf{n}} = -s^{\text{E}} \quad (2.25)$$

Mas a energia de Gibbs pode ser calculada pela [Equação 2.24b](#), logo diferenciando-se esta expressão com relação à temperatura (P e \mathbf{n} constantes):

$$\left(\frac{\partial g^{\text{E}}}{\partial T} \right)_{P, \mathbf{n}} = R \sum_{i=1}^N x_i \ln \gamma_i + RT \sum_{i=1}^N x_i \frac{\ln \gamma_i}{\partial T} = -s^{\text{E}} \quad (2.26)$$

Além disso:

$$g^{\text{E}} = h^{\text{E}} - T s^{\text{E}} \quad (2.27)$$

⁹Esta equação é conhecida como *regra de Lewis/Randall* e aplica-se para cada espécie numa solução ideal, em todas condições de temperatura, pressão e composição (SMITH et al., 2007). Observe a óbvia semelhança entre as Equações [2.19](#) e [2.22](#), onde a única diferença é o estado de referência.

Substituindo-se a Equação 2.26 na Equação 2.27:

$$\frac{h^E}{RT} = -T \sum_{i=1}^N x_i \frac{\partial \ln \gamma_i}{\partial T} \quad (2.28)$$

Desta forma, ao diferenciar-se o coeficiente de atividade com respeito à temperatura pode-se obter a entalpia de excesso do sistema em estudo.

2.1.4 Modelos de Coeficiente de Atividade

Os modelos de coeficiente de atividade têm sido muito utilizados até hoje no projeto de processos industriais envolvendo misturas na fase líquida. De maneira geral, o uso de Equações de Estado (EoS) seria suficiente, entretanto, na prática, estas geralmente aparecem mais na estimativa das propriedades de substâncias puras e envolvendo a fase vapor¹⁰.

Mesmo quando a fase vapor está presente, como por exemplo em colunas de destilação, muitas vezes os modelos de atividade são preferidos (GMEHLING, 2009) às EoS, podendo ser aplicados na clássica relação de Raoult modificada¹¹:

$$Py_i = \gamma_i x_i P_i^{\text{sat}} \quad (2.29)$$

Observa-se que neste caso o coeficiente de atividade funciona como uma correção para o comportamento real da solução.

Além disso, já existem regras de mistura próprias para o acoplamento de EoS com modelos de atividade, como a Regra de Mistura Auto-Consistente (*Self Consistent Mixing Rule* – SCMR) (STAUDT; SOARES, 2012) desenvolvida neste grupo de pesquisa. Portanto, o estudo de modelos de Gibbs em excesso continua sendo relevante no dimensionamento de operações unitárias que envolvam misturas na fase líquida.

Conforme já foi citado anteriormente no Capítulo 1, os modelos clássicos de coeficiente de atividade começaram a ser desenvolvidos desde o final do século XIX. Tais

¹⁰Tendo em vista que as Equações de Estado ainda não avançaram suficientemente de modo a prever, com confiança, as propriedades das fases condensadas, os modelos de atividade são preferidos em detrimento destas para líquidos (e sólidos) (SANDLER, 1993; KORETSKY, 2012).

¹¹Em pressões baixas, é comum aproximar-se a fase vapor como sendo um gás ideal. Além disso, desprezando-se o fator de *Poynting* (ver Smith et al. (2007), p. 297), a fugacidade do componente puro i pode ser aproximada por sua pressão de saturação (P_i^{sat}).

modelos abrangem, desde o simples modelo de Margules, até as últimas versões do UNIFAC (GREENSEMANN; GMEHLING, 2005b; XUE et al., 2012) e os modelos baseados em COSMO, como o COSMO-RS (KLAMT, 1995), COSMO-SAC (LIN; SANDLER, 2002) e o F-SAC (SOARES; GERBER, 2013, 2013). Neste trabalho, os modelos clássicos não serão apresentados. Em vez disto, na próxima seção, será dado foco às diferenças dos modelos baseados em COSMO com relação a estes. Especificamente, o equacionamento do modelo F-SAC será apresentado em detalhes.

2.2 Modelos de Coeficiente de Atividade Baseados em COSMO

2.2.1 Introdução

Os modelos de atividade baseados na metodologia COSMO foram desenvolvidos como alternativas aos modelos clássicos de coeficiente de atividade. Inicialmente, a teoria COSMO (*Conductor-like Screening Model*) foi elaborada por Klamt e Schuurmann (1993) para calcular os efeitos de solvente e energias de solvatação. Neste modelo, cada molécula é considerada como imersa num condutor perfeito, sendo determinadas, então, as cargas induzidas na superfície da molécula, conforme mostrado para as moléculas de benzeno e acetona na Figura 2.1. As representações das cargas foram geradas utilizando o programa JCOSMO (GERBER; SOARES, 2010).

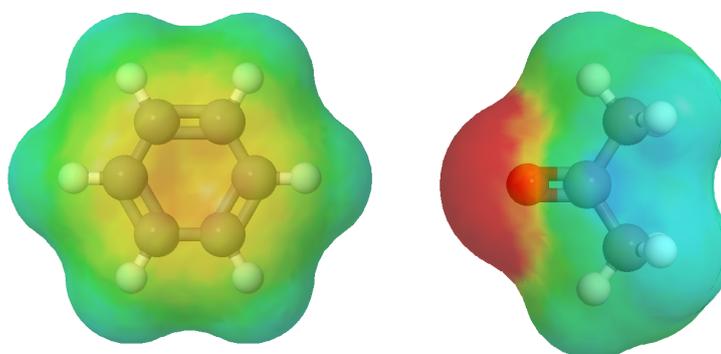


Figura 2.1: Cargas induzidas nas superfícies das moléculas de benzeno e acetona, de acordo com a teoria COSMO.

Baseado nesta metodologia, Klamt (1995) propôs o modelo COSMO-RS (*Realistic Solvation*). O COSMO-RS calcula a diferença de energia entre uma molécula em estado ideal (completamente solvatada num condutor perfeito) e em estado real, ou seja, quando posta em contato com outras moléculas em solução. Neste modelo, cada molécula é considerada como uma coleção de *segmentos superficiais* e o potencial químico de cada segmento de superfície é calculado utilizando-se as energias de interação entre todos os segmentos em solução. Finalmente, o potencial químico de cada molécula é calculado pela soma dos potenciais de cada segmento.

Uma característica interessante do modelo é que este utiliza-se das informações unicamente dos componentes puros (obtidas pelo COSMO) em substituição aos dados experimentais. Deste modo, apenas um pequeno conjunto de parâmetros deve ser estimado, sendo o modelo considerado (quase) totalmente preditivo e podendo, a princípio, ser utilizado para determinar o coeficiente de atividade de qualquer espécie em qualquer mistura.

A ideia básica considera, por exemplo, uma molécula *A* e uma *B*, inicialmente completamente envoltas em um condutor perfeito e longe o suficiente para que não haja interação entre elas. Cada uma delas possui sua própria energia COSMO, que é considerada como referência. As moléculas são progressivamente aproximadas até se tocarem em alguma posição. A única mudança de energia ocorre apenas no instante em que as moléculas se tocam, pois neste ponto o condutor é excluído.

Considere, por exemplo, um possível contato entre as moléculas de benzeno e acetona, conforme mostrado na Figura 2.2. Para cada contato entre as moléculas, o condutor será parcialmente excluído. Após, outras moléculas vão se aproximando até que o condutor seja completamente excluído e tenha-se uma *solução real*.

Observa-se pela Figura 2.2 que há infinitas possibilidades das moléculas entrarem em contato, tanto com moléculas de mesmo *tipo* quanto com diferentes. Desta forma, para calcular as probabilidades de todos os contatos, Klamt (1995) propôs um tratamento de termodinâmica estatística. Neste tratamento, representam-se as frações de áreas superficiais *vs.* as densidades de carga aparentes das moléculas em histogramas, denominados perfis- σ , conforme será mostrado.

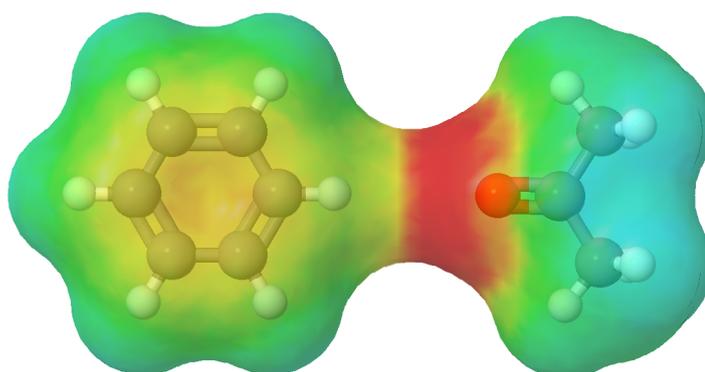


Figura 2.2: Possível contato entre as moléculas de benzeno e acetona, de acordo com a teoria COSMO.

Segundo a notação utilizada por [Lin e Sandler \(2002\)](#), o coeficiente de atividade de um segmento Γ_m , obtido a partir do tratamento de termodinâmica estatística, num líquido puro ou numa mistura¹², é dado genericamente por:

$$\ln \Gamma_m = - \ln \left[\sum_n p_n \Gamma_n \exp \left(\frac{-\Delta W_{m,n}}{RT} \right) \right] \quad (2.30)$$

onde p_n é a probabilidade de se encontrar um segmento n ou no líquido puro ou na mistura e $\Delta W_{m,n}$ é a energia de interação entre os segmentos m e n . Esta equação é denominada equação auto-consistente (*self-consistency equation*).

Resolver a [Equação 2.30](#) representa uma diferença particular entre os modelos do tipo COSMO e os modelos clássicos de coeficiente de atividade (Wilson, NRTL, UNIQUAC e UNIFAC e suas variantes). Enquanto os modelos clássicos apresentam equações explícitas, os modelos baseados em COSMO não apresentam uma equação na forma fechada (explícita), pois conforme pode-se observar na [Equação 2.30](#) o coeficiente de atividade de cada segmento depende dos valores de coeficiente de atividade de todos os segmentos (incluindo ele mesmo, veja o Γ_n dentro do somatório). Isto significa que estes modelos sempre necessitarão de um método numérico para serem resolvidos.

Nota-se que a [Equação 2.30](#) é, na verdade, um sistema de equações, consistindo

¹²A nomenclatura do COSMO-SAC ([LIN; SANDLER, 2002](#)), herdada pelo F-SAC, utiliza os sobrescritos s e i para denotar a mistura e o componente puro, respectivamente, conforme será detalhado mais adiante. Todavia, de maneira a enfatizar apenas as *diferenças* entre os modelos clássicos de coeficiente de atividade e os baseados em COSMO, optou-se, nesta seção, por apresentar a [Equação 2.30](#) neste formato genérico.

em uma equação para cada segmento m , seja em mistura ou no líquido puro. Cada molécula pode apresentar centenas de segmentos superficiais, levando a um sistema de equações muito grande a ser resolvido para misturas de muitos componentes. Além disso, o cálculo dos coeficientes de atividade requer a solução de sistemas de equações como o mostrado na [Equação 2.30](#) para *cada componente puro* e para a *mistura*.

Considere a [Figura 2.3](#), na qual são mostrados os centros de cada segmento (representados pelas pequenas esferas) para algumas moléculas típicas, quando utilizando-se o programa JCOSMO ([GERBER; SOARES, 2010](#)). Geralmente, os programas que implementam os modelos baseados em COSMO fixam o número máximo de segmentos de uma molécula. Desta forma, os segmentos com diferenças de cargas menores que um certo $\Delta\sigma$ (definido na implementação) são mesclados.

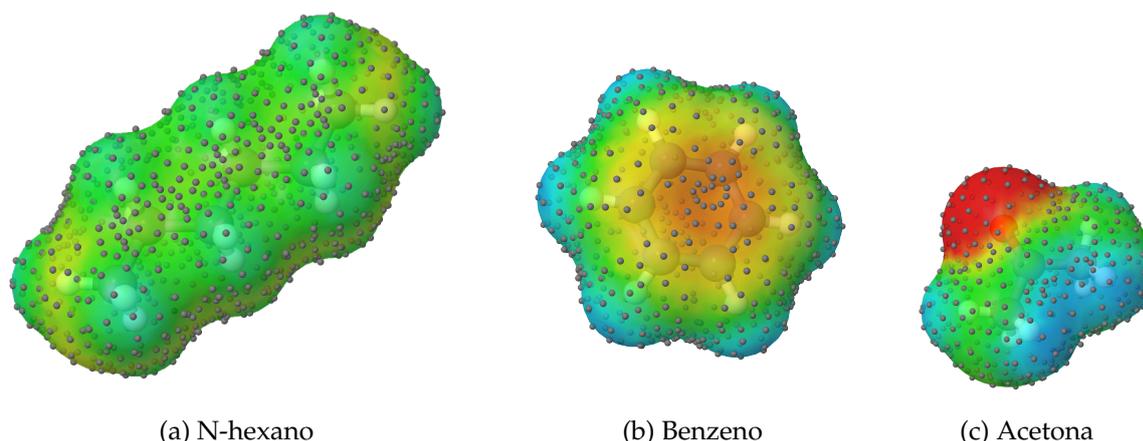


Figura 2.3: Centros dos segmentos de carga superficial (esferas) para algumas moléculas, de acordo com a metodologia COSMO.

Por exemplo, no programa COSMO-SAC FORTRAN ([Virginia Tech, 2015](#)), as moléculas são sempre consideradas como tendo 51 segmentos (por molécula), com cargas superficiais σ igualmente distribuídas entre $-0,025$ e $0,025$ e/Å². Ou seja, neste programa, segmentos com diferenças de cargas menores que $\Delta\sigma=0,001$ e/Å² são mescladas em um único segmento de área equivalente, não sendo distinguidos entre si. Caso o número de segmentos fosse reduzido para apenas 11, teríamos, neste caso, $\Delta\sigma^{11}=0,005$ e/Å².

De maneira geral, estas informações são omitidas nos trabalhos encontrados na literatura, de modo que um histograma contínuo, conforme mostrado na [Figura 2.4](#), é apresentado. Na realidade, a representação dos perfis- σ nos programas que implementam os modelos do tipo COSMO são discretos, conforme mostrado na [Figura 2.5](#) para $\Delta\sigma^{51}=0,001 e/\text{\AA}^2$ e $\Delta\sigma^{11}=0,005 e/\text{\AA}^2$. Estes histogramas foram gerados no presente trabalho utilizando o programa JCOSMO de [Gerber e Soares \(2010\)](#). Ou seja, na construção dos histogramas da [Figura 2.5](#), toda a densidade de carga que estiver em uma determinada faixa $\Delta\sigma$ tem sua área acumulada em um único valor.

Conforme discutido anteriormente, o tamanho dos sistemas a serem resolvidos dependerá fundamentalmente do número de segmentos de cada molécula. Desta forma, teoricamente, quanto mais segmentos forem considerados (menor $\Delta\sigma$), maior será o sistema de equações a ser resolvido (o problema fica mais complexo), entretanto, menos informação é perdida na etapa de mescla dos segmentos. Obviamente, quanto mais segmentos são considerados, menos informação do COSMO é perdida, porém, o esforço computacional é maior.

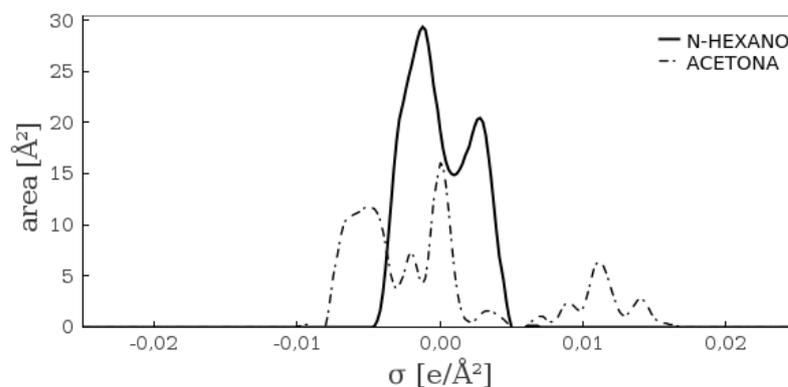


Figura 2.4: Perfis- σ contínuos das moléculas de n-hexano e acetona.

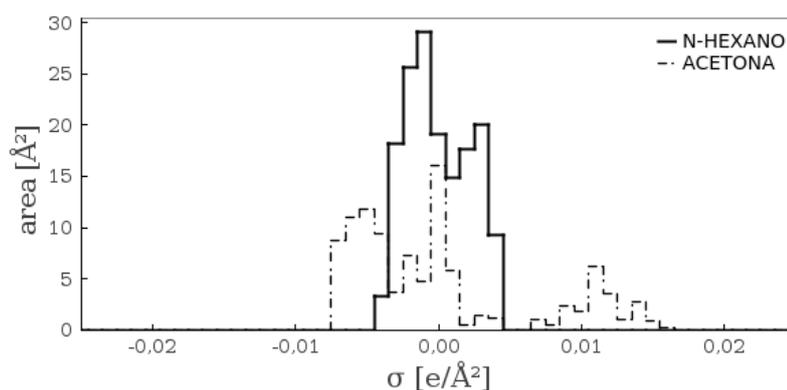
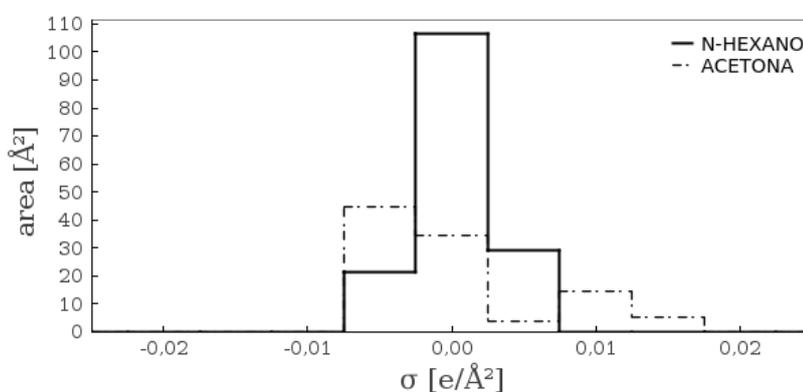
(a) $\Delta\sigma^{51}=0,001\text{ e}/\text{Å}^2$ (51 seg.)(b) $\Delta\sigma^{11}=0,005\text{ e}/\text{Å}^2$ (11 seg.)

Figura 2.5: Perfis- σ das moléculas de n-hexano e acetona com $\Delta\sigma^{51}=0,001\text{ e}/\text{Å}^2$ e $\Delta\sigma^{11}=0,005\text{ e}/\text{Å}^2$.

2.2.2 Equacionamento do Modelo F-SAC

No modelo F-SAC, bem como em todas as variantes do UNIFAC e nos modelos baseados em COSMO, o coeficiente de atividade da fase líquida, γ_i , é calculado a partir da soma de duas contribuições:

$$\ln \gamma_i = \ln \gamma_i^{\text{comb}} + \ln \gamma_i^{\text{res}} \quad (2.31)$$

onde $\ln \gamma_i^{\text{comb}}$ e $\ln \gamma_i^{\text{res}}$ são as contribuições combinatorial e residual, respectivamente.

A contribuição combinatorial é calculada pelo termo de Staverman-Guggenheim (SG), que aprimora os cálculos da energia livre de formação da

cavidade:

$$\ln \gamma_i^{\text{comb}} = \ln \phi'_i + 1 - \phi'_i - \frac{z}{2} q_i \left[\ln \left(\frac{\phi_i}{\theta_i} \right) + 1 - \frac{\phi_i}{\theta_i} \right] \quad (2.32)$$

$$\phi'_i = \frac{(V^i)^{3/4}}{\sum_j x_j (V^j)^{3/4}} \quad (2.33)$$

$$\phi_i = \frac{V^i}{\sum_j x_j V^j} \quad (2.34)$$

$$\theta_i = \frac{Q^i}{\sum_j x_j Q^j} \quad (2.35)$$

$$q_i = \frac{Q^i}{q} \quad (2.36)$$

onde z é o número de coordenação, usualmente fixado em 10; x_i é a fração molar do componente i ; e q é um parâmetro universal do modelo (área de normalização), fixado no modelo F-SAC (SOARES; GERBER, 2013; SOARES et al., 2013) em 50 \AA^2 .

O volume V^i e a área superficial Q^i da cavidade da molécula i são calculados por:

$$V^i = \sum_{m \in i} V_m^i \quad (2.37a)$$

$$Q^i = \sum_{m \in i} Q_m^i \quad (2.37b)$$

onde V_m^i e Q_m^i são o volume e a área superficial do segmento m da molécula i .

Enquanto que a contribuição combinatorial é fácil de ser entendida, a contribuição residual é particularmente mais complexa. Neste trabalho, a formulação apresentada terá uma notação um pouco diferente do que aquela classicamente apresentada nos trabalhos encontrados na literatura¹³. Tal modificação foi feita apenas de maneira a facilitar o entendimento geral do F-SAC (e também de toda a classe de modelos baseados em COSMO), tendo em vista que geralmente a formulação é difícil de ser entendida, principalmente para pesquisadores que não estejam familiarizados com os mesmos.

¹³Note que não foi realizada nenhuma modificação no modelo, apenas na notação das equações.

A contribuição residual é calculada como a diferença entre os coeficientes de atividade de cada segmento quando na solução, s , e no líquido puro, i :

$$\ln \gamma_i^{\text{res}} = \sum_{m \in i} \frac{Q_m^i}{a_{\text{eff}}} \cdot (\ln \Gamma_m^s - \ln \Gamma_m^i) \quad (2.38)$$

onde $a_{\text{eff}} = \pi \cdot r_{\text{eff}}^2$ é a área de contato padrão, sendo r_{eff} um dos parâmetros universais do modelo, fixado em $1,07 \text{ \AA}$ (SOARES; GERBER, 2013; SOARES et al., 2013); $\ln \Gamma_m^s$ e $\ln \Gamma_m^i$ são, respectivamente, o logaritmo natural do coeficiente de atividade de um segmento m na solução e no líquido puro, calculados pela equação auto-consistente (“self-consistency” equation):

$$\ln \Gamma_m^s = - \ln \left[\sum_{n \in s} p_n^s \Gamma_n^s \exp \left(\frac{-\Delta W_{m,n}}{RT} \right) \right] \quad (2.39a)$$

$$\ln \Gamma_m^i = - \ln \left[\sum_{n \in i} p_n^i \Gamma_n^i \exp \left(\frac{-\Delta W_{m,n}}{RT} \right) \right] \quad (2.39b)$$

Onde são p_m^s e p_m^i são as probabilidades de se encontrar um segmento m na mistura s e no líquido puro i , respectivamente; e $\Delta W_{m,n}$ é a energia de interação para cada contato entre os segmentos m e n .

A probabilidade de se encontrar um segmento m na mistura s é:

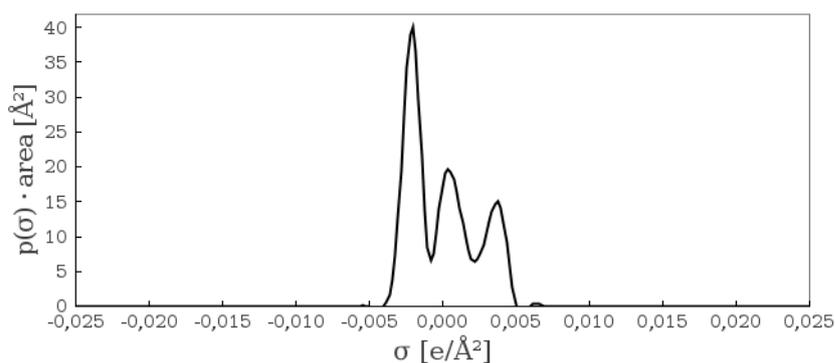
$$p_m^s = \frac{x_i Q_m^i}{\sum_j x_j Q_j^j} \quad (2.40)$$

A probabilidade de se encontrar um segmento m no líquido puro i é:

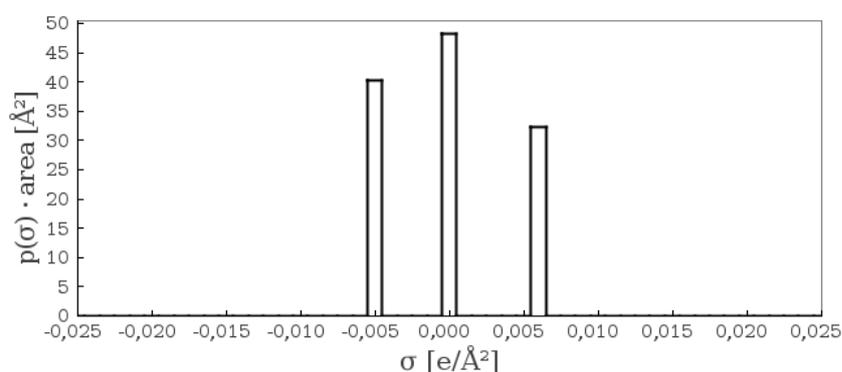
$$p_m^i = \frac{Q_m^i}{Q^i} \quad (2.41)$$

A principal diferença entre o F-SAC e os modelos do tipo COSMO reside no perfil- σ . No F-SAC propõe-se que cada grupo funcional tenha o seu próprio perfil- σ obtido empiricamente, em substituição aos cálculos de química quântica (COSMO). Ainda, as moléculas são consideradas como sendo formadas por um ou mais *grupos funcionais* (contribuição de grupos) e que cada grupo funcional tenha o seu próprio perfil- σ , composto por 3 segmentos (1 positivo, 1 neutro e 1 negativo).

A [Figura 2.6](#) ilustra o perfil- σ da molécula de benzeno calculado com o COSMO e o ajustado empiricamente pelo modelo F-SAC. A utilização deste exemplo é conveniente, pois esta molécula contém apenas um tipo de grupo funcional, o CH aromático.



(a) COSMO



(b) F-SAC

Figura 2.6: Perfil- σ para o benzeno calculado com o COSMO e ajustado para o F-SAC.

O perfil- σ de cada grupo funcional k pode ser representado por três parâmetros empíricos: a área absoluta do grupo funcional com carga positiva, Q_k^+ (segmento positivo), a área absoluta do grupo funcional com carga negativa, Q_k^- (segmento negativo), e a densidade de carga do segmento positivo (σ_k^+). A partir destas definições, a área neutra, Q_k° (segmento neutro), é dada por $Q_k^\circ = Q_k - Q_k^+ - Q_k^-$, onde Q_k é a área superficial total. Fazendo-se um balanço de cargas, de modo a manter o grupo neutro, a densidade de carga negativa é calculada por $\sigma_k^- = -\sigma_k^+ Q_k^+ / Q_k^-$. Então, o perfil- σ do grupo k é dado por:

$$\{(\sigma_k^-, Q_k^-); (0, Q_k^\circ); (\sigma_k^+, Q_k^+)\} \quad (2.42)$$

E finalmente, para molécula i , o perfil- σ é dado pela soma dos perfis- σ dos grupos funcionais que a compõem:

$$\sum_k \nu_k \cdot \{(\sigma_k^-, Q_k^-); (0, Q_k^\circ); (\sigma_k^+, Q_k^+)\} \quad (2.43)$$

onde ν_k é o número de vezes que o grupo k aparece na molécula. Os parâmetros Q_k^+ , Q_k^- , σ_k^+ e (Q_k) podem ser estimados partir de dados experimentais¹⁴.

Os métodos de contribuição de grupos geralmente utilizam diferentes grupos para representar a mesma molécula. Além disso, podem surgir ambiguidades que fazem a identificação destes grupos ser não única. No modelo F-SAC segue-se as seguintes regras:

1. a geometria do grupo funcional (ou seja, ângulo de ligação, etc.) deve ser a mesma, independentemente da molécula em que o mesmo se encontra;
2. cada átomo no grupo funcional deve ter aproximadamente a mesma carga em todas as moléculas em que o grupo ocorra, e este deve ser aproximadamente neutro;
3. cada grupo funcional deve ser a menor entidade na qual a molécula pode ser dividida.

Assim, os grupos funcionais são definidos a partir da inspeção visual das figuras tridimensionais que representam as cargas induzidas na superfície da molécula (geradas pelo COSMO). Desta inspeção, localiza-se o centro do grupo, que geralmente é um heteroátomo (diferente de carbono e hidrogênio), e expande-se em todas as direções de modo a englobar uma área neutra da molécula. Com isso, de modo a representar as moléculas, diferentes grupos devem ser criados.

Pela necessidade dos grupos serem neutros, um grande número deles deveria ser criado, o que resultaria em mais parâmetros. A fim de evitar um número elevado de parâmetros a ser estimado, grupos parecidos química e estruturalmente foram considerados como subgrupos de um mesmo grupo. Esta estratégia é similar à utilizada nos modelos tipo UNIFAC e uma explicação mais detalhada pode ser encontrada no trabalho original do F-SAC (SOARES; GERBER, 2013).

Assim, é assumido que os parâmetros eletrostáticos (Q_k^+ , Q_k^- e σ_k^+) são idênticos dentro de um mesmo grupo e os subgrupos diferem apenas na área (Q_k) e volume

¹⁴Equilíbrio líquido-líquido (LLE) binário e/ou ternário, coeficiente de atividade à diluição infinita (IDAC), equilíbrio líquido-vapor (VLE), entalpia de mistura, etc.

(V_k), sendo esta diferença englobada pela região neutra do subgrupo. Portanto, os parâmetros a serem estimados, sem considerar as misturas com associação (ligações de hidrogênio), são: Q_k^+ , Q_k^- e σ_k^+ para os grupos e Q_k para os subgrupos.

A energia de interação para cada possível contato entre dois segmentos m e n qualquer, da mistura ou do componente puro, depende da versão do modelo F-SAC que está sendo utilizada. Por exemplo, na Dissertação de Mestrado do autor deste trabalho (POSSANI, 2014), esta contribuição foi modificada de modo a levar em consideração dados de LLE, junto com os dados de IDAC e VLE em amplos intervalos de temperatura. Já no trabalho de Flôres et al. (2016), uma nova contribuição foi introduzida para levar em consideração as forças de dispersão.

Optou-se por estudar neste trabalho a versão original do F-SAC, por ser a versão mais completa do modelo, pois possui o maior número parâmetros para os grupos e subgrupos. No trabalho original de Soares e Gerber (2013), Soares et al. (2013), a energia de interação é dada por:

$$\Delta W_{m,n} = \frac{\alpha'(\sigma_m + \sigma_n)^2}{2} - \frac{E_{m,n}^{\text{HB}}}{2} \quad (2.44)$$

onde α' é a constante de desajuste de carga igual a $35\,750 \text{ kJ } \text{Å}^4 / \text{mol e}^2$; e $E_{m,n}^{\text{HB}}$ é o parâmetro binário de ligação de hidrogênio.

Considera-se que cada grupo capaz de realizar ligação de hidrogênio possui sítios doadores ou receptores (de ligação de hidrogênio) e que cada um destes sítios tem uma porção de área idêntica à área de contato padrão do modelo (calculada com r_{eff}). Deve-se, então, enumerar a quantidade de sítios receptores/doadores de ligação de hidrogênio para determinado grupo. Tendo-se este número definido, o parâmetro binário $E_{m,n}^{\text{HB}}$ pode ser estimado a partir de dados experimentais¹⁵.

A Figura 2.7 mostra a distribuição superficial de densidades de cargas para as moléculas de acetona, clorofórmio e etanol. Observa-se que a acetona possui sítios receptores de ligação de hidrogênio (área vermelha); já o clorofórmio tem um sítio doador (área azul). Isto significa que, numa mistura contendo acetona e clorofórmio,

¹⁵Equilíbrio líquido-líquido (LLE) binário e/ou ternário, coeficiente de atividade à diluição infinita (IDAC), equilíbrio líquido-vapor (VLE), entalpia de mistura, etc, para misturas de componentes que fazem ligação de hidrogênio

haverá a associação cruzada entre estas duas moléculas. Por outro lado, a molécula de etanol possui ambos sítios doadores e receptores; ou seja, ela pode tanto associar-se com outras moléculas que tenham um doador/receptor quanto se auto-associar.

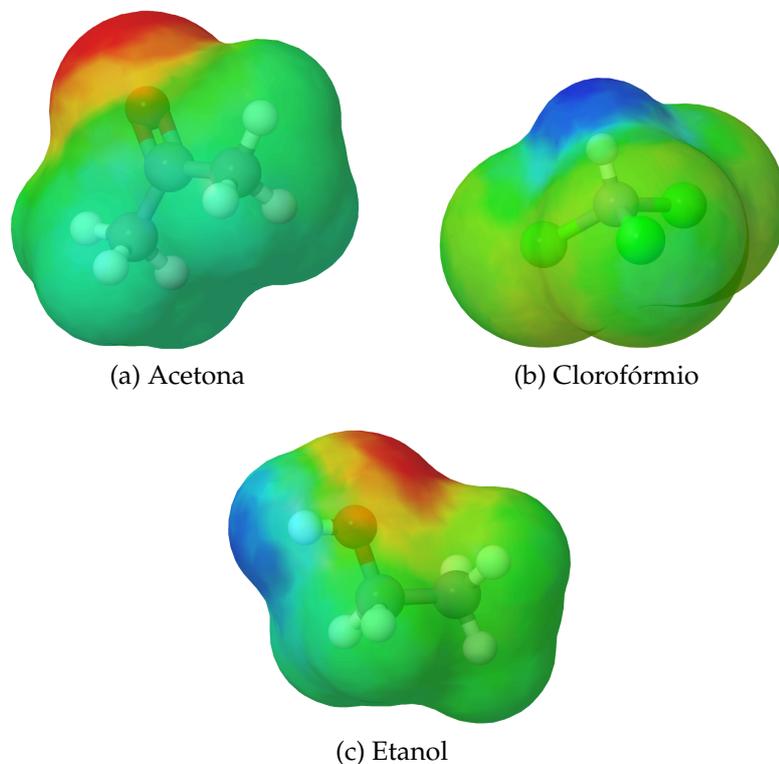


Figura 2.7: Representação COSMO para acetona, clorofórmio e etanol.

2.2.3 Resolução da Equação Auto-Consistente

Percebe-se que, para a resolução do modelo F-SAC, o sistema dado pela [Equação 2.39a](#) (para a mistura) deverá ser resolvido iterativamente. Além disso, o tamanho do sistema (número de variáveis) será igual à soma de todos os segmentos presentes em cada molécula. Da mesma forma, a [Equação 2.39b](#) deverá ser resolvida, de modo iterativo, individualmente para *cada molécula* presente na mistura (n vezes, numa mistura de n componentes). O tamanho de cada sistema dependerá do número de segmentos de cada molécula. Assim, para uma mistura qualquer de n componentes, o cálculo de $\ln \gamma_i$ requer $n + 1$ vezes a resolução de sistemas de equações no formato da [Equação 2.30](#).

Segundo os trabalhos de Klamt (1995), Klamt et al. (1998), os sistemas representados pelas Equações 2.39a e 2.39b (ou genericamente pela Equação 2.30) são resolvidos pelo clássico método da substituição sucessiva. De acordo com os autores, o método se inicia com um valor inicial unitário para o coeficiente de atividade de cada segmento e este valor é atualizado iterativamente. Os valores utilizados nas atualizações correspondem à média entre os valores antigos e novos, para evitar oscilações. Este processo tende a tornar a convergência mais lenta, entretanto, a convergência ocorre em milissegundos nos computadores modernos.

No trabalho de Lin e Sandler (2002) são encontrados poucos detalhes sobre os procedimentos adotados para resolver as Equações 2.39a e 2.39b. Supõe-se, entretanto, que o mesmo método das substituições sucessivas seja utilizado. Além disso, no código em FORTRAN do COSMO-SAC, desenvolvido pela Virginia Tech (2015) e disponibilizado livremente na internet, a substituição sucessiva é utilizada.

Nos códigos em Java do COSMO-SAC e do F-SAC do nosso grupo de pesquisa, uma implementação do método das substituições sucessivas similar ao código em FORTRAN da Virginia Tech (2015) é utilizado. Na Figura 2.8 é apresentado um fluxograma simplificado do algoritmo de substituição sucessiva conforme implementado no programa JCOSMO (GERBER; SOARES, 2010). O pseudo-código do método é mostrado no Algoritmo 1.

Conforme descrito no fluxograma da Figura 2.8, o algoritmo inicia-se com os valores de coeficiente de atividade para todos os segmentos com valor unitário (estimativa inicial). Após, para cada segmento um novo valor de Γ_m é calculado a partir da Equação 2.45, obtida eliminando-se os logaritmos nos dois lados da equação Equação 2.30:

$$\Gamma_m = \left[\sum_n p_n \Gamma_n \exp \left(\frac{-\Delta W_{m,n}}{RT} \right) \right]^{-1} \quad (2.45)$$

Os valores atualizados são reutilizados na mesma iteração (para os segmentos seguintes) e o método continua até que a diferença entre os valor novo e o atual da norma Euclidiana dos vetores de segmentos seja menor que uma dada tolerância (ponto de parada) ou até atingir o número máximo de iterações. No F-SAC, utiliza-se $\epsilon = 10^{-8}$ e $maxiter = 200$.

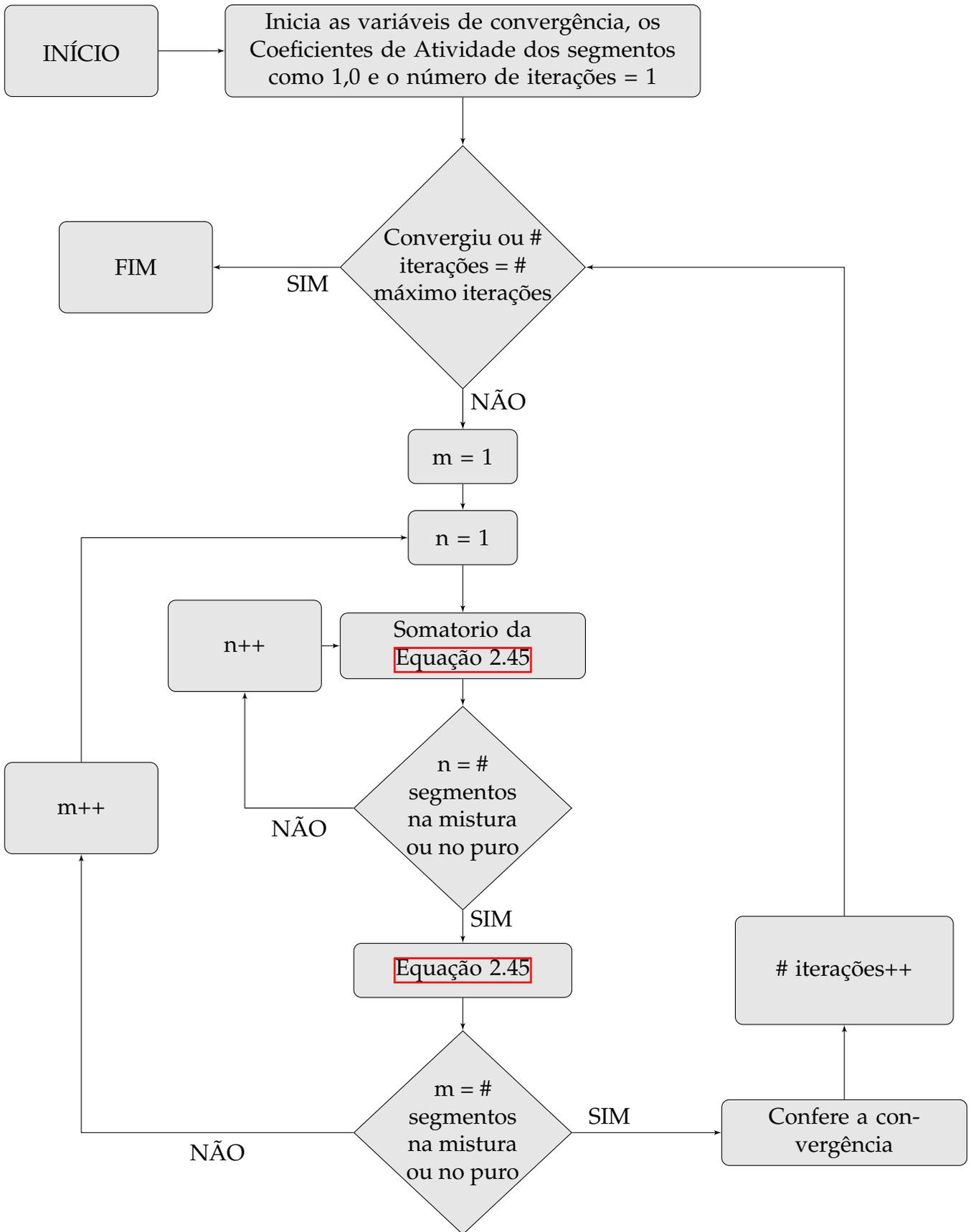


Figura 2.8: Fluxograma da substituição sucessiva.

Algoritmo 1 Resolve a equação auto-consistente pelo método da substituição sucessiva, dados a tolerância ϵ e número máximo de iterações *maxiter*.

```

1: niter  $\leftarrow$  1 {número de iterações}
2: convergiu  $\leftarrow$  falso
3: norm  $\leftarrow$  -1
4: para m = 1 até m = nseg faça
5:    $\Gamma_m \leftarrow$  1; {valor inicial}
6: fim laço para
7: enquanto convergiu = falso ou niter < maxiter faça
8:   newnorm  $\leftarrow$  0
9:   para m = 1 até m = nseg faça
10:     $\sum \leftarrow$  0
11:    para n = 1 até n = nseg faça
12:      $\sum \leftarrow \sum + p_n \Gamma_n \exp\left(\frac{-\Delta W_{m,n}}{RT}\right)$ ; {somatório na Equação 2.45}
13:    fim laço para
14:     $\Gamma_m \leftarrow (\Gamma_m + \sum^{-1})/2$  {média entre os valores atual e antigo}
15:    newnorm  $\leftarrow$  newnorm +  $(\Gamma_m)^2$ 
16:    fim laço para
17:    newnorm  $\leftarrow$   $\sqrt{\text{newnorm}}$ 
18:    se  $|newnorm - norm| / newnorm < \epsilon$  então
19:     convergiu  $\leftarrow$  verdadeiro
20:    fim laço se
21:    norm  $\leftarrow$  newnorm
22:    niter  $\leftarrow$  niter + 1
23: fim laço enquanto

```

O método das substituições sucessivas é apontado como o método escolhido pelos principais autores para a resolução da equação auto-consistente. Entretanto, estes sistemas, em princípio, poderiam ser resolvidos por *qualquer* outro método numérico conhecido, como, por exemplo, o método de Newton.

O método de Newton, também chamado método de Newton-Raphson, é bastante conhecido e utilizado na Engenharia. Geralmente, este método apresenta uma convergência rápida na resolução de problemas representados por funções contínuas, desde que dado um valor inicial confiável (PRESS et al., 1992).

Durante a realização deste Trabalho, a possibilidade de substituir o método descrito anteriormente por uma implementação do método de Newton foi estudada. O resultado deste estudo foi publicado em Possani e Soares (2019). Neste estudo, concluiu-se que, apesar de apresentar um número muito menor de iterações, o método de New-

ton foi mais lento, para o caso estudado, pois necessita da inversão da matriz Jacobiana da seguinte equação:

$$\Gamma_m \cdot \left[\sum_n p_n \Gamma_n \exp \left(\frac{-\Delta W_{m,n}}{RT} \right) \right] - 1 = 0 \quad (2.46)$$

A [Equação 2.2.3](#) foi obtida rearranjando-se os termos da [Equação 2.45](#). Esta equação será mostrada novamente no [Capítulo 3](#) e será utilizada na obtenção das derivadas do modelo F-SAC. Conforme discutido em [Possani e Soares \(2019\)](#), matriz Jacobiana da será sempre densa de modo que a sua inversão é lenta. Naquele trabalho, a inversão desta matriz foi realizada por decomposição LU.

2.3 Estimação de Parâmetros

2.3.1 Problema de Otimização

Independentemente do grau de empirismo ou do rigor utilizado na modelagem, a estimação de parâmetros constitui uma etapa crucial ([HIMMELBLAU et al., 2001](#)) na representação de dados experimentais, pois em geral uma série de parâmetros (do modelo) deve ser determinada (ou estimada).

A estimação de parâmetros é representada por um problema de otimização. Genericamente, um problema de otimização não linear¹⁶ pode ser escrito da seguinte forma¹⁷:

$$\begin{aligned} \min_{\mathcal{P} \in \mathbb{R}^n} \quad & \text{FO}(\mathcal{P}) \\ \text{sujeito a :} \quad & h(\mathcal{P}) = 0 \\ & g(\mathcal{P}) \leq 0 \end{aligned} \quad (2.47)$$

Onde FO é a função objetivo, \mathcal{P} são as variáveis de otimização (parâmetros) e h e g são as restrições de igualdade e de desigualdade (equações do modelo), respectivamente.

Assim, a estimação dos parâmetros é realizada através da otimização (minimização) da função objetivo, devendo ser assumida alguma métrica de diferença entre os valores experimentais e os valores calculados pelo modelo. Em geral, nos trabalhos encontrados na literatura que utilizam modelos de energia de Gibbs em excesso, a forma

¹⁶Nonlinear Program (NLP), ou seja, exclui o caso de variáveis inteiras.

¹⁷Adaptado da equação descrita em [Biegler \(2010\)](#).

desta função pode variar bastante de autor para autor. Ela depende basicamente do tipo de dados que estão sendo utilizados para a estimação, das dificuldades numéricas e também da intenção do respectivo autor.

Por exemplo, Malanowzki et al. (1981) representaram a FO como uma soma dos desvios entre o valores experimentais e os valores calculados pelo modelo UNIQUAC para, simultaneamente, dados de VLE, LLE e entalpia de excesso (h^E):

$$FO = FO_{VLE} + FO_{LLE} + FO_{h^E} \quad (2.48)$$

ou, de uma maneira mais genérica:

$$FO = \sum_k FO_k \quad (2.49)$$

onde um desvio qualquer (FO_k) é dado pelo somatório das diferenças quadráticas entre os dados experimentais ($X_{i,k}^{exp}$) e calculados ($X_{i,k}^{calc}$), ponderados pelas variâncias ($\sigma_{i,k}^2$) dos respectivos experimentos, considerando-se que os erros experimentais seguem uma distribuição gaussiana e independente (método da máxima verossimilhança):

$$FO_k = \sum_{i,k} \frac{(X_{i,k}^{exp} - X_{i,k}^{calc})^2}{\sigma_{i,k}^2} \quad (2.50)$$

Nos trabalhos do UNIFAC, tanto o original quanto as suas modificações, os autores utilizaram funções objetivo semelhantes às Equações 2.49 e 2.50. Em Fredenslund et al. (1975), primeiro trabalho do UNIFAC, e em Skjold-Jorgensen et al. (1979), primeira revisão do modelo, por exemplo, os autores utilizaram a diferença quadrática entre o logaritmo natural dos coeficientes de atividade para dados de VLE. Já nas revisões seguintes do UNIFAC (GMEHLING et al., 1982; MACEDO et al., 1983; TIEGS et al., 1987), utilizou-se a diferença quadrática para dados de composições do líquido e do vapor, temperatura e pressão de VLE. Semelhantemente, Hansen et al. (1991) e Wittig et al. (2003) construíram uma FO baseada nas diferenças quadráticas entre a composição do vapor e a pressão.

Da mesma forma, no trabalho de Weidlich e Gmehling (1987), no qual foi introduzido o UNIFAC (Do)¹⁸, a mais famosa modificação do UNIFAC, os autores se utilizaram simultaneamente de dados de IDAC, VLE e h^E . Entretanto, os detalhes quanto

¹⁸UNIFAC modificado de Dortmund, em alusão à cidade em que o modelo foi desenvolvido.

ao formato da FO não constam naquele trabalho. Já nos trabalhos que se seguiram ao UNIFAC (Do) (GMEHLING et al., 1993; GMEHLING et al., 1998; GMEHLING et al., 2002; JAKOB et al., 2006), os autores utilizaram, além de IDAC, VLE e h^E , também dados de azeotropia, equilíbrios líquido-líquido (LLE) e sólido-líquido (SLE) e capacidade térmica em excesso (c_p^E), sempre que disponíveis. As funções objetivo utilizadas nesses trabalhos foram semelhantes à Equação 2.49, entretanto a FO_k foi escrita da seguinte forma (mínimos desvios relativos):

$$FO_k = \sum_{i,k} \left(\frac{X_{i,k}^{\text{exp}} - X_{i,k}^{\text{calc}}}{X_{i,k}^{\text{exp}}} \right)^2 \quad (2.51)$$

O formato de função objetivo adotado na Equação 2.49, além de ser genérico, é interessante, principalmente quando deseja-se representar simultaneamente diferentes tipos de dados experimentais, conforme consta nos trabalhos do UNIFAC e UNIFAC (Do). Observando-se atentamente a Equação 2.50, percebe-se que, além da existência de dados experimentais, ainda é necessário saber a informação sobre as variâncias das medidas experimentais. Entretanto, nem sempre há a disponibilidade de tais informações principalmente nas coletâneas de dados experimentais encontrados na literatura ou em bancos de dados como, por exemplo, o próprio DDB¹⁹ utilizado nas parametrizações do UNIFAC e do UNIFAC (Do).

Nestes casos, é comum que os autores se utilizem de variâncias constantes para cada tipo de dado experimental, como foi feito nos trabalhos do UNIFAC. Ou ainda, que se utilizem da Equação 2.51, como foi no caso dos trabalhos do UNIFAC (Do).

Além disso, a utilização de pesos é bastante comum, tendo sido encontrada inclusive na maioria dos trabalhos citados anteriormente. Um peso é um fator empírico, particular para cada tipo de dado experimental, que multiplica as diferenças entre os resultados experimentais e calculados, de modo que as contribuições destes sejam adequadamente relevantes para a função objetivo.

Por exemplo, no trabalho de Gmehling (2009), que faz um apanhado geral do UNIFAC, suas modificações e extensões, é citado de maneira clara que a estimação de parâmetros, na maioria dos casos, deve ser repetida várias vezes com pesos ajustados.

¹⁹Dortmund Data Bank, contendo 6,7 milhões de dados termodinâmicos para aproximadamente 44000 substâncias (DDBST, 2014)

Desta forma, a [Equação 2.50](#) poderia ser ainda escrita considerando-se *pesos* ($w_{i,k}$), em vez das variâncias desconhecidas, para cada dado experimental:

$$FO_k = \sum_{i,k} w_{i,k} \cdot (X_{i,k}^{\text{exp}} - X_{i,k}^{\text{calc}})^2 \quad (2.52)$$

Ou, ainda, assumindo que os *pesos* são constantes para cada tipo de dado experimental (w_k):

$$FO_k = w_k \cdot \sum_{i,k} (X_{i,k}^{\text{exp}} - X_{i,k}^{\text{calc}})^2 \quad (2.53)$$

2.3.2 Dados Experimentais

Além de definir a função objetivo, é necessária a existência de dados experimentais para serem utilizados no processo de estimação de parâmetros. De maneira geral, tais dados podem ser obtidos pelo próprio autor, no caso de serem realizados experimentos para este fim, ou também podem ser consultados em alguma bibliografia, como artigos de outros autores, bancos de dados (por exemplo o DDB) ou as compilações da IUPAC (*Data Series*).

A compilação de dados experimentais presentes na literatura exige um olhar crítico do compilador sobre a confiabilidade dos experimentos. Dados de VLE, por exemplo, possuem testes de consistência termodinâmica para a sua verificação, ou seja, conjuntos que apresentem inconsistências podem ser identificados e descartados.

No presente trabalho, serão utilizados os dados de IDAC existente no banco de dados experimentais do nosso grupo de pesquisa. Este banco de dados contém 2979 pontos experimentais de coeficiente de atividade à diluição infinita para 172 substâncias, conforme descrito em trabalhos anteriores ([GERBER; SOARES, 2010](#); [SOARES, 2011](#); [GERBER; SOARES, 2013](#); [SOARES; GERBER, 2013](#); [SOARES et al., 2013](#)).

2.4 Métodos de Otimização

2.4.1 Introdução

Além da escolha de uma função objetivo adequada e do uso de dados experimentais, é necessária a escolha de algum algoritmo de otimização para a estimação de parâmetros. Os autores de otimização – e.g. Himmelblau et al. (2001), Biegler (2010) – costumam escrever que não há um único método de otimização aplicável eficientemente para todos os problemas existentes.

Na verdade, a performance do método escolhido para um caso particular dependerá das características do problema de otimização, seja o formato da função objetivo e das restrições, como também número de variáveis do problema. Desta forma, é necessário testar alguns métodos diferentes até encontrar aquele que seja mais adequado ao problema em particular.

Os métodos de otimização podem ser classificados em:

- Métodos que não usam derivadas, também denominados métodos de busca direta.
- Métodos que usam derivadas, que englobam todos os métodos que utilizam a primeira e/ou a segunda derivada da função objetivo.

De modo geral, os métodos de busca são de mais fácil implementação e aplicação numa grande variedade de problemas de otimização. Além disso, estes métodos requerem menor conhecimento do problema de otimização, sendo particularmente interessantes na exploração da otimização para novas aplicações, antes de investir tempo e esforço em estratégias mais sofisticadas (BIEGLER, 2010).

Segundo Himmelblau et al. (2001), métodos que não utilizam derivadas, em geral são mais eficientes para resolução de problemas de otimização com funções objetivo “mal comportadas”²⁰. Entretanto, a performance destes métodos não raro se deteriora

²⁰Funções “mal comportadas”, ou conforme o Himmelblau et al. (2001), *nonsmooth functions*, são aquelas que apresentam descontinuidades (e.g. função módulo, máximo, mínimo,...) em algum ponto, ou cujas derivadas apresentam estas descontinuidades.

rapidamente com o aumento do número de variáveis de otimização. Mesmo que a performance possa ser melhorada com o uso de computação paralela, estes métodos são raramente aplicados a problemas maiores (com algumas dúzias de variáveis de otimização).

Por outro lado, para funções “bem comportadas”²¹, métodos que utilizam derivadas podem ser mais eficientes e rápidos, se tornando mais interessantes principalmente quando o número de variáveis aumenta (BIEGLER, 2010). Entretanto, estes métodos, muitas vezes, falham ao resolver funções com mudanças abruptas ou com descontinuidades.

Nos trabalhos do UNIFAC e UNIFAC (Do) não é usual os autores apresentarem detalhes quanto aos métodos de otimização utilizados na estimação de parâmetros. Parece ser uma tendência os autores destes trabalhos se preocuparem mais em mostrar os dados experimentais utilizados e as respectivas tabelas de parâmetros obtidos. Além disto, quando o método é citado, em geral a citação é restrita a curtos comentários como os do trabalho de Hansen et al. (1991)²² e Weidlich e Gmehling (1987)²³.

Por exemplo, Macedo et al. (1983), Tiegs et al. (1987), Wittig et al. (2003), Gmehling et al. (1998), Gmehling et al. (2002), Jakob et al. (2006), não mencionam o método de otimização utilizado nos seus trabalhos. Já Fredenslund et al. (1975), Gmehling et al. (1982), Weidlich e Gmehling (1987) utilizaram o Nelder-Mead (método de busca) e Skjold-Jorgensen et al. (1979), Hansen et al. (1991) o método de Levenberg-Marquardt (método que utiliza derivada). Em Gmehling et al. (1993), os autores utilizaram o Nelder-Mead no início da otimização e o Levenberg-Marquardt para refinar o resultado perto do ótimo.

Conforme citado anteriormente, nos trabalhos referentes ao F-SAC realizados no nosso grupo de pesquisa até o presente momento (SOARES; GERBER, 2013; SOARES et al., 2013; POSSANI et al., 2014; POSSANI et al., 2014; FLÔRES et al., 2016; POSSANI et al., 2016), um algoritmo Nelder-Mead, conforme descrito na Subseção 2.4.2 foi utilizado, por sua simplicidade. Entretanto, em nenhum destes trabalhos foi pos-

²¹Funções contínuas, ou cujas derivadas sejam contínuas, num determinado intervalo.

²²“The Levenberg-Marquardt method was used for the minimization.”

²³“Optimizations were carried out by using the Simplex-Nelder-Mead method...”.

sível estimar todos os parâmetros com todos os dados experimentais disponíveis de uma só vez, devido à não convergência do algoritmo. Para superar as dificuldades encontradas, os parâmetros são geralmente estimados aos poucos.

Para facilitar o processo de estimação de parâmetros do modelo e verificação dos resultados, uma interface gráfica, denominada *F-SAC Optimizer* e escrita na linguagem computacional JAVA, foi desenvolvida pelo autor durante o Mestrado (POSSANI, 2014). Esta interface gráfica permite escolher quais os dados experimentais que deseja-se utilizar na estimação de parâmetros e quais parâmetros serão estimados, bem como a visualização dos resultados. A tela principal do *F-SAC Optimizer* é mostrada na Figura 2.9.

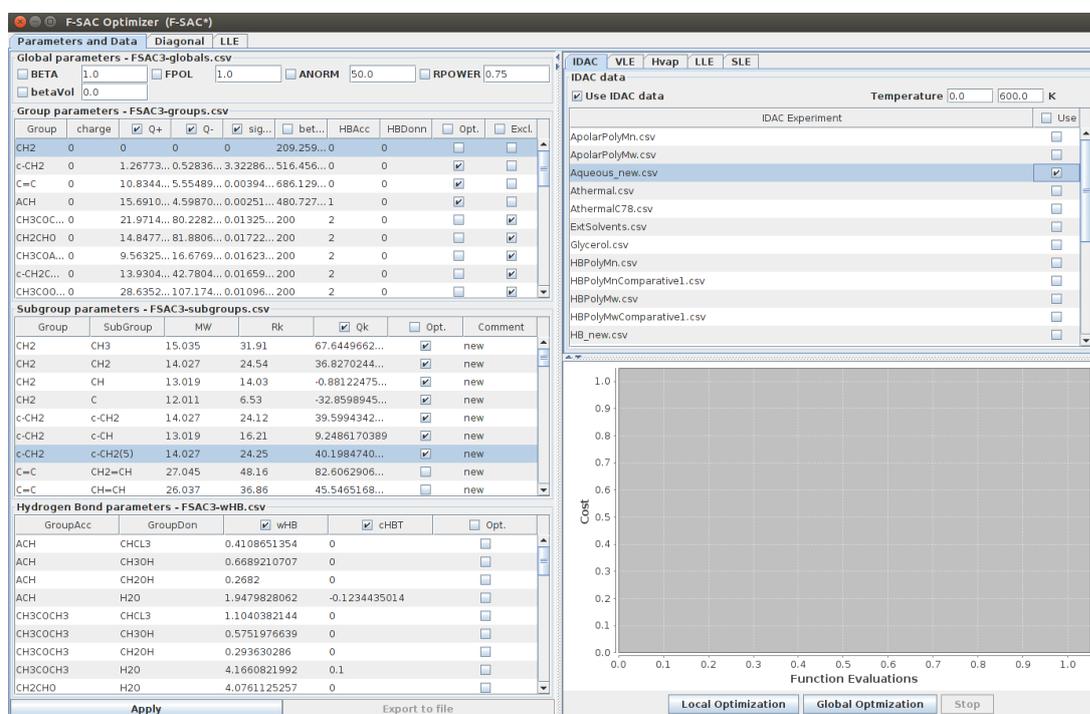


Figura 2.9: Tela inicial do *F-SAC Optimizer*. Extraído de POSSANI (2014).

Neste trabalho, buscou-se um algoritmo de otimização alternativo ao Nelder-Mead. Além disto, propõe-se a utilização de um método de otimização baseado em derivadas da função objetivo, devido ao número elevado de parâmetros a serem estimados. Assim, foram testados dois algoritmos de otimização que utilizam derivadas e que servem para problemas com restrição, o Ipopt e o MMA, descritos nas subseções

[2.4.3](#) e [2.4.4](#). Conforme será detalhado a seguir, o Ipopt não foi uma boa escolha, devido a algumas características particulares do método utilizado. Desta forma, testou-se também o MMA, disponível dentro do pacote NLOpt.

2.4.2 Nelder-Mead

O método de Nelder-Mead ([NELDER; MEAD, 1965](#)) é um dos métodos de otimização mais conhecidos, sendo bastante popular em vários ramos da ciência por sua simplicidade. O método não requer as derivadas da função objetivo e foi originalmente concebido para resolver problemas de otimização sem restrições, do seguinte formato:

$$\min_{x \in \mathcal{R}^n} f(x) \quad (2.54)$$

O Nelder-Mead baseia-se na construção de uma figura geométrica, *simplex*, com $n + 1$ vértices. Os valores da função objetivo nestes vértices são comparados em cada iteração e o algoritmo busca substituir o vértice com o pior valor da função objetivo, através de 4 tipos de transformação: reflexão; expansão; contração; e, caso as três anteriores falhem, encolhimento.

O método pode ser resumido nos seguintes passos:

- Passo 1: define os $n + 1$ vértices do *simplex* P_0, P_1, \dots, P_n . Dentre estes pontos, calcula o P_h e o P_l , que correspondem aos vértices de maior (y_h) e de menor (y_l) valores de função objetivo, respectivamente. Também, calcula o centróide desta figura, excluindo o P_h .
- Passo 2 (reflexão): calcula um novo ponto (P^*) a partir da reflexão de P_h por uma linha que passa pelo centróide da figura. Caso P^* gere um valor de função objetivo (y^*) tal que $y^* < y_l$, vai para o passo 3; caso $y^* > y_i$, onde $i = 0, \dots, n; i \neq h$, vai para o passo 4; caso y^* seja um valor entre y_h e y_l e pelo menos algum y_i seja maior que P^* , substitui P_h por P^* e vai para o Passo 6.
- Passo 3 (expansão): calcula um novo ponto (P^{**}) a partir da expansão de P^* por uma linha que passa pelo centróide da figura. Caso P^{**} gere um valor

de função objetivo (y^{**}) tal que $y^{**} < y_l$, substitui P_h por P^{**} ; caso contrário, a expansão falhou, então substitui P_h por P^* . Vai para o Passo 6.

- Passo 4 (contração): caso $y^* > y_h$ (P^* foi rejeitado), substitui P^* por P_h . Calcula um novo ponto (P^{**}) a partir da contração de P^* por uma linha que passa pelo centróide da figura. Caso $y^{**} < y^*$, substitui P_h por P^{**} e vai para o Passo 6; caso contrário, a contração falhou, então vai para o passo 5.
- Passo 5 (encolhimento): encolhe todo o *simplex* em direção ao P_l .
- Passo 6: Verifica os critérios de convergência. Caso sejam satisfeitos (convergiu), o método termina; caso não sejam satisfeitos e o número de iterações máximas não foi atingido, recalcula P_h e P_l , incrementa o número de iterações e volta para o Passo 2; caso contrário, divergiu e o método termina.

O critério de convergência utilizado por [Nelder e Mead \(1965\)](#) é escrito da seguinte forma:

$$\sqrt{\frac{\sum_i (y_i - \bar{y})^2}{n}} < \epsilon \quad (2.55)$$

onde \bar{y} é o valor da função objetivo no centróide do *simplex* e ϵ é a tolerância do método.

Para problemas com restrição, os autores sugerem a utilização de uma “penalidade”, como, por exemplo, retornar um valor alto para a função objetivo num ponto em que haja a violação de alguma restrição. Sugere-se a leitura do trabalho original ([NELDER; MEAD, 1965](#)), em especial o fluxograma da “Fig. 1”, para maiores informações sobre o método descrito.

2.4.3 Interior Point OPTimizer (Ipopt)

O Ipopt (sigla de *Interior Point OPTimizer* do inglês) é um pacote computacional de código livre, escrito na linguagem de programação C++ e distribuído sob a licença EPL (*Eclipse Public License*). O código foi inicialmente desenvolvido por Andreas Wächter e Carl Laird na linguagem FORTRAN77, entretanto, a partir de 2005 foi lançada a versão em C++ que é mantida até hoje (a versão em FORTRAN77 ainda está dis-

ponível em <https://github.com/coin-or/Ipopt/tree/stable/2.3> mas não recebe manutenção).

Atualmente, o projeto do Ipopt (<https://github.com/coin-or/Ipopt>) é mantido por Andreas Wächter e Stefan Vigerske, possuindo uma longa lista de colaboradores em todo o mundo. De maneira geral, o pacote pode ser utilizado nos sistemas operacionais Linux/UNIX, Mac OS X e Windows, podendo ser integrado em códigos em FORTRAN, C, C++, Java e MATLAB.

O algoritmo implementado dentro do Ipopt consiste num método de otimização *local*, tendo sido desenvolvido para tratar de problemas não-lineares em larga escala, do seguinte formato:

$$\begin{aligned} \min_{x \in \mathcal{R}^n} \quad & f(x) \\ \text{sujeito a :} \quad & g_L \leq g(x) \leq g_U \\ & x_L \leq x \leq x_U \end{aligned} \tag{2.56}$$

onde $f(x)$ é a função objetivo; x representa o vetor de n variáveis de otimização; os vetores g_L e g_U representam os limites (*bounds*) inferiores e superiores das restrições $g(x)$, respectivamente; e x_L e x_U são os limites inferiores e superiores de x , respectivamente²⁴.

O Ipopt utiliza um método de ponto interior, ou método de barreira (*barrier method*) utilizando como filtro uma busca em linha. Não é o objetivo do presente trabalho descrever de maneira detalhada o equacionamento proposto, de modo que o funcionamento deste método será explicado simplificadamente nos próximos parágrafos. Recomenda-se, para maiores detalhes de como funciona o método implementado dentro do Ipopt, a leitura do trabalho de Wächter e Biegler (2006) e a documentação do algoritmo em <https://coin-or.github.io/Ipopt/index.html>.

Internamente, o Ipopt não resolve o problema original fornecido pelo usuário na forma da [Equação 2.56](#). Na verdade, o algoritmo substitui as restrições de desigualdade por restrições de igualdade, com a introdução de “variáveis de folga” (*slack variables*), de modo que as únicas desigualdades presentes sejam os limites. A partir

²⁴Segundo consta no *site* do projeto (<https://projects.coin-or.org/Ipopt>), não há condicionante nenhuma quanto à convexidade ou à linearidade tanto de $f(x)$ quanto de $g(x)$. Impõe-se, apenas, que ambas sejam pelo menos duas vezes diferenciáveis. Além do mais, restrições de igualdade podem ser facilmente inseridas escrevendo-se o mesmo valor de g_L e g_U .

daí, os *bounds* são adicionados na função objetivo como um termo de barreira (*barrier*) logarítmico, de modo que a estratégia adotada seja resolver uma série de problemas de barreira.

O algoritmo inicia resolvendo o problema de forma relaxada e com valores moderados para o parâmetro de barreira. Então, este parâmetro é reduzido e resolve-se novamente o problema, utilizando como estimativa inicial a solução anterior. A partir daí, o método é repetido até que haja uma solução que atenda às condições de otimalidade de primeira ordem, com o grau de tolerância necessário. Dentro da resolução das condições de otimalidade do problema de barreira (conforme descrito anteriormente), é necessária a utilização de um método de Newton. No método de Newton, uma busca em linha é utilizada até que seja obtido um tamanho de passo que resulte em um progresso na resolução do problema de barreira.

O Ipopt utiliza a derivada primeira (gradiente) da função objetivo, a derivada primeira (matriz Jacobiana) do vetor de restrições e a derivada segunda (matriz Hessiana) da função Lagrangeana. O vetor gradiente da função objetivo deve ser fornecido pelo usuário, entretanto, se a matriz Hessiana não for fornecida, o algoritmo possibilita a aproximação desta matriz através do método *quasi-Newton* L-BFGS²⁵. Neste trabalho, serão fornecidas apenas as derivadas primeiras da função objetivo e do vetor de restrições.

Segundo destacado por Wächter e Biegler (2006), o conceito principal da aplicação do Ipopt é que um ponto qualquer, testado como solução do problema de otimização, pode ser aceito pelo método, caso ele diminua o valor da função objetivo ou melhore a violação das restrições, em vez de se utilizar uma combinação destas medidas através de uma função mérito. Ainda, por se tratar de um método de barreira, o Ipopt, enquanto ainda não convergiu, gera uma seqüência de pontos que podem não satisfazer todas as restrições. Ou seja, este algoritmo é bastante permissivo com relação à violação das restrições, pois ele permite que a restrição seja violada, em um determinado passo, desde que a FO diminua e somente depois tenta trazer os parâmetros de volta para um local que não viole a restrição.

²⁵*Limited-memory Broyden-Fletcher-Goldfarb-Shanno*. Para mais explicações, veja os capítulos 6 e 7 do livro do Nocedal (2006), em especial o algoritmo 7.5 descrito na pág. 179.

O modelo F-SAC é muito sensível a violações das suas restrições. O modelo pode divergir completamente quando suas restrições são minimamente violadas. Este é um aspecto crucial do Ipopt que inviabilizou o seu uso no caso específico tratado no presente trabalho. Desta forma, foi necessária a busca por outro método, o qual será descrito a seguir.

2.4.4 NonLinear optimization (NLOpt)

O NLOpt (sigla de *NonLinear optimization*) é uma biblioteca de código livre, escrita na linguagem C e distribuída sob a licença GNU LGPL (*Lesser General Public License*), que implementa diversos métodos de otimização não lineares. Os códigos foram desenvolvidos pelo grupo de pesquisa do Professor Steven G. Johnson do Instituto de Física e Matemática Aplicada do MIT e podem ser integrados em programas escritos em C, C++, FORTRAN, MATLAB (ou outros semelhantes), Python, entre outros²⁶.

As implementações dentro do NLOpt podem ser simplesmente traduções de códigos já existentes ou códigos originais de métodos descritos na literatura²⁷. Todos os algoritmos foram implementados de forma a resolver um problema de otimização no seguinte formato:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{sujeito a :} \quad & lb \leq x \leq ub \\ & fc(x) \leq 0 \\ & h(x) = 0 \end{aligned} \tag{2.57}$$

onde $f(x)$ é a função objetivo; x representa o vetor de n variáveis de otimização; os vetores lb e ub representam os limites inferiores e superiores de x , respectivamente; fc é o vetor de restrições de desigualdade; e h é o vetor de restrições de igualdade.

Dentre os algoritmos implementados, há métodos de otimização local e global, com e sem restrições, que se utilizam e que não se utilizam das derivadas da função objetivo, além de algoritmos para otimização em larga escala (milhões de pa-

²⁶O NLOpt inclui apenas implementações que não assumem convexidade nem da função, tampouco das restrições.

²⁷O próprio Professor Steven G. Johnson se responsabiliza por todas as implementações e possíveis *bugs* inseridos.

râmetros e centenas de restrições). Sugere-se a leitura da página <https://NLOpt.readthedocs.io/en/latest/> para mais informações sobre o NLOpt.

Conforme citado anteriormente, busca-se neste trabalho um algoritmo de otimização local e que se utilize das derivadas da função objetivo. Nesta categoria de métodos, o NLOpt possui pelo menos 5 algoritmos diferentes. Dentre estes, somente o MMA e o SLSQP²⁸ suportam restrições de desigualdade; os demais só podem ser utilizados para problemas sem restrição ou somente com *bounds*. Na verdade, qualquer um dos métodos do NLOpt (global, local, com derivada, sem derivada ...) poderia ser utilizado para problemas com restrições; neste caso, é necessário combinar o método escolhido com um algoritmo denominado AUGLAG (*Augmented Lagrangian*), que “transforma” a função objetivo e as restrições numa única função.

Escolheu-se inicialmente o algoritmo denominado MMA (*Method of Moving Asymptotes*), por atender aos requisitos anteriores e ser de simples integração. Percebe-se que os outros métodos disponíveis na categoria de “métodos que utilizam derivada”, também poderiam ser utilizados, como por exemplo o SLSQP, que na verdade é similar, na sua concepção, com o MMA. Ainda, poderiam ser utilizados outros métodos desta categoria, combinados com o AUGLAG, conforme observado anteriormente.

Todavia, não é o objetivo deste trabalho estudar vários métodos de otimização e compará-los; procura-se, apenas, um método que seja útil na resolução do problema apresentado²⁹. Fica, portanto, como sugestão para trabalhos futuros, uma análise mais detalhada dos outros métodos implementados no NLOpt, podendo ser feita uma comparação entre os mesmos, em termos de resultados e performance computacional.

²⁸Veja a documentação do NLOpt (https://NLOpt.readthedocs.io/en/latest/NLOpt_Algorithms/) para verificar os significados das siglas.

²⁹Lembrando que um método alternativo ao Ipopt só foi buscado depois que se verificou que o mesmo não funcionava.

O MMA é um método local, globalmente convergente³⁰, que se utiliza do gradiente da função objetivo e da matriz Jacobiana das restrições de desigualdade (restrições de igualdade não são suportadas), tendo sido escrito baseado no trabalho de Svanberg (2002). Não é objetivo do presente trabalho explicar detalhadamente todo o equacionamento do MMA, de modo que nos próximos parágrafos, apenas uma explicação geral do método será fornecida.

O algoritmo do MMA possui dois laços de iteração, um externo e um interno. Partindo-se de um ponto viável x^k , em cada laço interno o MMA forma uma aproximação local (subproblema) do problema de otimização, utilizando as derivadas da função objetivo (gradiente) e das restrições (matriz Jacobiana) com relação às variáveis de otimização, somadas a um termo quadrático de penalidade. Esta aproximação forma uma função convexa e separável, que é resolvida por um método dual.

A resolução da aproximação quadrática leva a um novo ponto candidato, no qual a função objetivo e as restrições são calculados. Caso a aproximação tenha sido de fato conservativa, a solução é aceita, o ponto calculado torna-se x^{k+1} e a iteração externa é finalizada, sendo o processo repetido com o novo ponto. Caso contrário, as aproximações são modificadas, de forma a serem mais restritivas através do aumento da penalidade, e o algoritmo repete o processo anterior com o novo subproblema.

2.5 Obtenção das Derivadas e o Teorema da Função Implícita

A maior parte dos algoritmos para resolver problemas de otimização não lineares requerem, pelo menos, a derivada primeira da função objetivo (HIMMELBLAU et al., 2001). Se as aproximações das derivadas forem pouco confiáveis, o algoritmo utilizado pode se tornar lento, ir em direções equivocadas ou terminar por falta de movimento, podendo até convergir para um ponto qualquer que não seja um ótimo do problema.

³⁰Nota-se que “globalmente convergente” não é igual a “global”, ou seja, não significa que o algoritmo vai convergir para um ótimo global; significa apenas que o método, garantidamente, convergirá para algum mínimo local, desde que o ponto inicial seja viável (SVANBERG, 2002)

Conforme já justificado anteriormente, neste trabalho foi escolhido utilizar-se um algoritmo com derivadas, por causa, dentre outros motivos, do tamanho do problema a ser resolvido. Desta forma, é imprescindível: (i) a obtenção das derivadas da função objetivo utilizada³¹; (ii) que os valores calculados das derivadas sejam *confiáveis*.

Para uma função facilmente diferenciável, é interessante investir na sua derivação analítica, pois geralmente obtém-se derivadas confiáveis e em um menor tempo computacional. Por outro lado, para funções mais complicadas, ou que não apresentem uma solução explícita, ou seja, uma forma fechada (envolvem um método iterativo), a derivação analítica já não é interessante, de modo que, outros métodos são necessários, como por exemplo, a derivação por perturbação numérica.

Os métodos de perturbação numérica são, na prática, os mais utilizados nos códigos computacionais, inclusive para funções explícitas. Entretanto, estes métodos podem gerar derivadas pouco precisas, além de demandar, em geral, maior esforço computacional (HIMMELBLAU et al., 2001).

Outra alternativa seria o emprego dos chamados métodos de derivação algorítmica (automática). O livro do Griewank (Andreas Griewank, 2008) explica de maneira detalhada, do ponto de vista da construção dos algoritmos, este tipo de metodologia. Para aplicação da derivação algorítmica, seria necessário reescrever todos os códigos desenvolvidos no nosso grupo de pesquisa, o que geraria um esforço provavelmente muito exagerado para este trabalho.

Conforme destacado na descrição dos modelos baseados em COSMO (em especial neste caso o F-SAC) a resolução da equação auto-consistente dada por Equação 2.39 exige um método numérico. Desta forma, o uso da derivação analítica seria prejudicado, pois levaria a mais um sistema de equações a ser resolvido. Para evitar o uso de perturbações numéricas ou a reescrita de maior parte dos nosso códigos, propõe-se, neste trabalho, o uso do Teorema da Função Implícita, conforme será descrito a seguir.

O Teorema da Função Implícita pode ser enunciado da seguinte forma, conforme adaptado livremente do livro de Krantz e Parks (2002):

Teorema da Função Implícita. *Seja $F : R^N \rightarrow R^M$ e $N > M$ definidas na vizinhança de*

³¹O que leva, em última instância, à necessidade da obtenção das derivadas do modelo F-SAC.

(x_0, y_0) , escreve-se $R^N = R^{N-M} \times R^M$ e denota-se $N - M$ coordenadas para o vetor x e M coordenadas para o vetor y . Supõe-se:

- F é continuamente diferenciável;
- $F(x_0, y_0) = 0$;
- $\det \left(\frac{\partial F}{\partial y} \right) (x_0, y_0) \neq 0$.

Onde a matriz Jacobiana de F com respeito à y é definida por:

$$J(\bar{y}) = \left(\frac{\partial F}{\partial y} \right) \equiv \begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_M}{\partial y_1} & \cdots & \frac{\partial F_M}{\partial y_M} \end{bmatrix} \quad (2.58)$$

Então existem intervalos abertos $U \subseteq R^{N-M}$, $V \subseteq R^M$ que satisfazem $x_0 \in U$, $y_0 \in V$ e:

- Para cada $x \in U$, $F(x, y) = 0$ possui uma solução única $Y = y(x) \in V$;
- $Y(x_0) = y_0$;
- Para $x \in U$,

$$\frac{\partial Y}{\partial x} = - \left(\frac{\partial F}{\partial Y} \right)^{-1} \cdot \frac{\partial F}{\partial x} \quad (2.59)$$

A completa descrição desta teoria não foi apresentada aqui, nem é a intenção deste trabalho. O formato desta apenas foi rearranjado para que pudesse ser utilizada nos desenvolvimentos aqui apresentados. Para mais informações históricas sobre o famoso Teorema da Função Implícita, sugere-se ao leitor o excelente livro de [Krantz e Parks \(2002\)](#), que trata apenas deste Teorema e os resultados que podem ser obtidos pela sua aplicação.

A Teoria da Função Implícita aparece cedo nos cursos de Engenharia, na disciplina de Cálculo I. Entretanto, frequentemente ela não é denominada desta maneira, pois pode ser obtida na aplicação da Regra da Cadeia. Por exemplo, se a equação vetorial $F(x, y) = 0$ e o vetor $y = y(x)$, então a derivada total de F é dada por:

$$dF = \left(\frac{\partial F}{\partial x} \right)_y \cdot dx + \left(\frac{\partial F}{\partial y} \right)_x \cdot dy = 0 \quad (2.60)$$

Mas $\mathbf{y} = \mathbf{y}(\mathbf{x})$, logo:

$$d\mathbf{y} = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right) \cdot d\mathbf{x} \quad (2.61)$$

Aplicando na [Equação 2.60](#):

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = - \left(\frac{\partial \mathbf{F}}{\partial \mathbf{y}} \right)_x^{-1} \cdot \left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}} \right)_y \quad (2.62)$$

Com o uso do Teorema da Função Implícita, é possível obter as derivadas do modelo F-SAC, conforme será devidamente explicado no [Capítulo 3](#).

Capítulo 3

Metodologia

3.1 Descrição do Problema de Otimização

Para estimar os parâmetros do modelo, a função objetivo (FO) deve ser escrita como a soma dos desvios entre os valores experimentais e os valores calculados, levando em consideração todos os dados de equilíbrio de fases disponíveis, conforme mostrado na [Equação 2.49](#). Neste trabalho, serão utilizados apenas dados de coeficiente de atividade a diluição infinita (IDAC), logo a FO se reduz a:

$$FO = FO_{IDAC} \quad (3.1)$$

Os desvios com relação aos IDACs serão escritos como os desvios *quadráticos* entre os pontos experimentais e calculados, conforme mostrado na [Equação 2.53](#). Observe que, neste caso, por se tratar de apenas um tipo de dado experimental (IDAC), os pesos podem ser negligenciados (considerados iguais a 1).

$$FO_{IDAC} = \frac{1}{NE} \sum_i (\ln \gamma_{i,exp}^{\infty} - \ln \gamma_{i,model}^{\infty})^2 \quad (3.2)$$

onde NE é o número de dados experimentais; $\ln \gamma_{i,exp}^{\infty}$ e $\ln \gamma_{i,calc}^{\infty}$ são os logaritmos naturais dos coeficientes de atividade à diluição infinita experimentais e calculados, respectivamente.

Os limites superiores e inferiores dos parâmetros do modelo F-SAC são:

- $Q_k^+ \geq n_{acc} \cdot a_{eff}$;
- $Q_k^- \geq n_{don} \cdot a_{eff}$;

- $0 \leq \sigma_k^+ \leq \sigma_k^{max}$;

E a restrição do modelo é dada por:

$$-\sigma_k^{max} \leq -\frac{\sigma_k^+ Q_k^+}{Q_k^-} \leq 0 \quad (3.3)$$

onde $\sigma_k^{max} = 0,025 \text{ e}/\text{\AA}^2$; n_{acc} e n_{don} são os números de sítios receptores e doadores de ligação de hidrogênio, respectivamente. Observações:

- Os parâmetros $+Q_k^+$ e $+Q_k^-$ não possuem limites superiores (os limites são $+\infty$).
- Se determinado grupo não faz ligação de hidrogênio, então n_{acc} e n_{don} são zero.
- A restrição [3.3](#) é obtida a partir dos limites de σ_k^- , $-\sigma_k^{max} \leq \sigma_k^- \leq 0$, e do balanço de cargas para cada grupo, $\sigma_k^- = -\sigma_k^+ Q_k^+ / Q_k^-$.
- Caso os limites inferiores de $+Q_k^+$, $+Q_k^-$ e $+\sigma_k^-$ sejam respeitados, $\sigma_k^- \leq 0$ também será, de modo que não é necessário colocar esta restrição.

Finalmente, o problema de otimização é representado da seguinte forma:

$$\begin{aligned} \min_{\mathcal{P} \in \mathcal{R}^n} \quad & \frac{1}{NE} \sum_i (\ln \gamma_{i,exp}^\infty - \ln \gamma_{i,model}^\infty)^2 \\ \text{sujeito a :} \quad & n_{acc} \cdot a_{eff} \leq Q_k^+ \leq +\infty \\ & n_{don} \cdot a_{eff} \leq Q_k^- \leq +\infty \\ & 0 \leq \sigma_k^+ \leq \sigma_k^{max} \\ & \sigma_k^+ Q_k^+ / Q_k^- - \sigma_k^{max} \leq 0 \end{aligned} \quad (3.4)$$

Onde \mathcal{P} é o vetor de parâmetros do modelo, Q_k^+ , Q_k^- e σ_k^+ para os grupos, Q_k para os subgrupos e $E_{m,n}^{HB}$ para os pares de grupos doador–receptor de ligação de hidrogênio.

Optou-se por utilizar uma estratégia de otimização do tipo sequencial, ou *feasible path*. Isto significa que o modelo F-SAC, utilizado para calcular $\ln \gamma_i^{\infty,calc}$ na [Equação 3.2](#), não é resolvido simultaneamente ao problema de otimização. A [Figura 3.1](#) exemplifica de maneira esquemática o funcionamento do programa implementado. Conforme ilustrado, o algoritmo de otimização (item 1 da [Figura 3.1](#)), enquanto numa

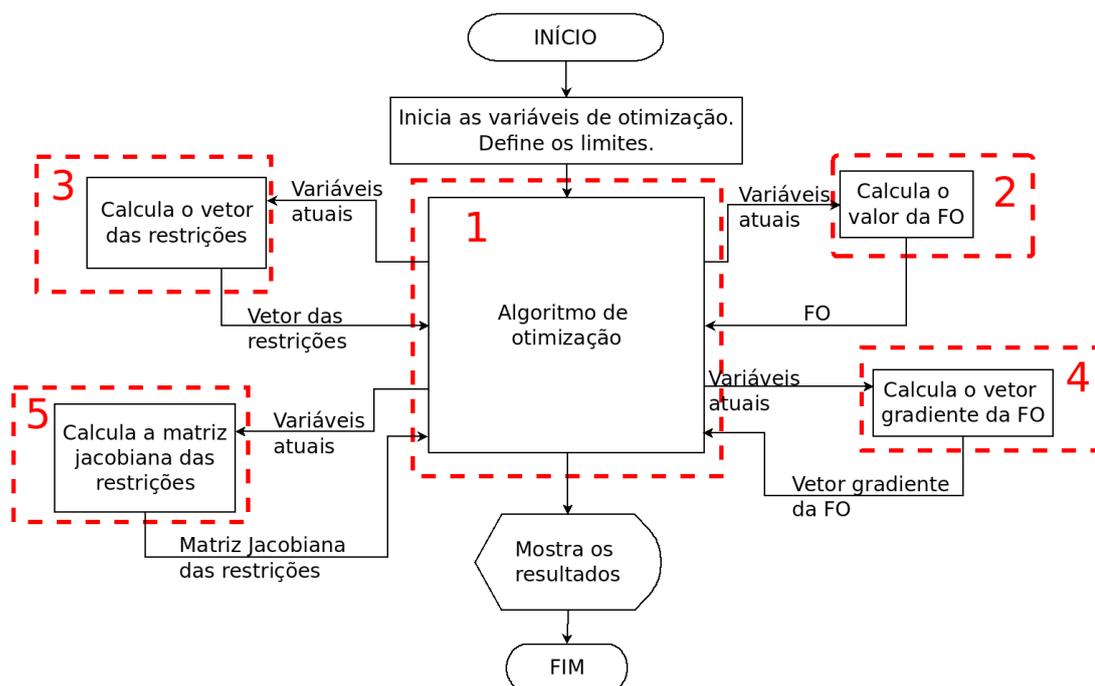


Figura 3.1: Esquema do programa implementado.

determinada iteração, requer o cálculo dos valores da função objetivo (2), das restrições (3) e das derivadas (4 e 5), os quais são realizados em trechos de código externos ao algoritmo de otimização. A estratégia sequencial foi adotada, em detrimento da estratégia de resolução simultânea, devido ao número muito elevado de restrições adicionadas ao problema de otimização, caso o método simultâneo fosse utilizado.

Por exemplo, considere um dado experimental de diluição infinita de etanol em água, a uma temperatura qualquer. As duas moléculas possuem, cada uma, 05 segmentos, resultando em uma mistura com 10 segmentos. Logo, para este ponto experimental, seriam adicionados mais 20 restrições de igualdade somente para o cálculo da equação auto-consistente (5 de cada molécula pura e 10 da mistura¹). Como cada ponto experimental utilizado na estimação de parâmetros corresponde a uma mistura binária, se considerarmos uma média de 20 restrições de igualdade para cada experimento, caso a estratégia simultânea fosse utilizada, seriam adicionadas cerca de 60000 restrições de igualdade².

¹Veja o primeiro parágrafo da Subseção 2.2.3 para mais detalhes.

²2979 (dados experimentais, veja a Subseção 2.3.2) \times 20 = 59580.

3.2 Derivada da Função Objetivo

A função objetivo descrita na [Equação 3.2](#) pode ser facilmente derivada com relação a um parâmetro qualquer do modelo (\mathcal{P}):

$$\frac{\partial}{\partial \mathcal{P}} \text{FO}_{IDAC} = -\frac{2}{\text{NE}} \sum_i (\ln \gamma_{i,exp}^{\infty} - \ln \gamma_{i,model}^{\infty}) \cdot \frac{\partial}{\partial \mathcal{P}} \ln \gamma_{i,model}^{\infty} \quad (3.5)$$

Nota-se, nesta equação, algumas particularidades interessantes:

- Não existe mais um termo quadrático dentro dos parênteses, logo, os valores do somatório podem ser positivos e negativos, enquanto que a função objetivo é sempre positiva.
- O “2” multiplicando o somatório não deve ser esquecido, pois caso o seja pode atrapalhar completamente o método de otimização.
- O operador negativo na frente da equação é proveniente do $-\ln \gamma_{i,model}^{\infty}$ e ocorre devido ao formato escolhido da [Equação 3.2](#).
- Não haveria diferença nos valores das derivadas da função objetivo caso os termos dentro dos parênteses estivessem invertidos, pois, nesta hipótese, o valor do somatório também teria sinal invertido e não haveria o operador negativo na frente da equação.

É interessante ressaltar que o termo de derivada parcial no lado direito da [Equação 3.5](#) indica a necessidade das derivadas do IDAC com relação aos parâmetros do modelo. Logo, em última instância, é necessário derivar o próprio F-SAC com relação a cada parâmetro, conforme mostrado a seguir.

3.3 Derivada do Modelo F-SAC

3.3.1 Introdução

Conforme explicado anteriormente ([Subseção 2.2.2](#)), os parâmetros do modelo F-SAC a serem estimados são:

- Os perfis- σ (parâmetros eletrostáticos: Q_k^+ , Q_k^- e σ_k^+) dos grupos.
- As áreas superficiais dos subgrupos (Q_k).
- Os parâmetros binários para cada par de grupos doador-receptor de ligação de hidrogênio ($E_{m,n}^{\text{HB}}$).

A estimação de parâmetros do modelo F-SAC envolve a obtenção destes para os grupos existentes. Desta forma, para realizar o uso de um método de otimização baseado em derivadas, conforme observado anteriormente, é necessário derivar o modelo F-SAC, com relação aos seus parâmetros, para cada grupo, subgrupo e par de grupos doador-receptor de ligação de hidrogênio.

A obtenção de tais derivadas será demonstrada nesta seção. Todavia, antes de serem mostradas as derivadas do modelo F-SAC com relação aos parâmetros propriamente ditos, serão apresentadas as suas derivadas com relação à temperatura e ao número de mols de um determinado componente da mistura.

É conveniente mostrar primeiro tais equacionamentos, de modo a facilitar o entendimento do leitor. Além do mais, foi importante no início deste trabalho que fossem realizadas tais derivações para testar a metodologia proposta. Por fim, os resultados obtidos destas derivadas apresentam alguns significados físicos interessantes, conforme apontado no [Capítulo 2](#):

- Ao diferenciar-se o coeficiente de atividade com respeito à temperatura, pode-se obter a entalpia de excesso do sistema em estudo.
- Ao diferenciar-se o coeficiente de atividade com respeito ao número de mols dos componentes da mistura, as relações entre as propriedades parciais molares das diferentes substâncias em solução podem ser estudadas a partir do uso da equação de Gibbs-Duhem.

Observe que em alguns casos os desenvolvimentos apresentados aqui serão muito similares entre si, de modo que estes serão suprimidos para evitar repetições.

3.3.2 Equacionamento Geral

Neste trabalho, a formulação do modelo F-SAC, conforme descrito na [Subseção 2.2.2](#), será utilizada. Assim, a derivada com respeito a alguma variável independente \mathcal{P} pode ser escrita da seguinte forma:

$$\frac{\partial \ln \gamma_i}{\partial \mathcal{P}} = \frac{\partial \ln \gamma_i^{comb}}{\partial \mathcal{P}} + \frac{\partial \ln \gamma_i^{res}}{\partial \mathcal{P}} \quad (3.6)$$

A contribuição combinatorial é uma equação explícita, geralmente escrita na forma do termo de Staverman-Guggenheim ([LIN; SANDLER, 2002](#)). De maneira geral, a derivação desta equação é bastante simples, como segue:

$$\begin{aligned} \frac{\partial \ln \gamma_i^{comb}}{\partial \mathcal{P}} &= \left(\frac{1}{\phi_i} - 1 \right) \cdot \left(\frac{\partial \phi_i}{\partial \mathcal{P}} \right) \\ -\frac{z}{2} q_i &\left\{ \frac{1}{\phi_i} \cdot \left(\frac{\partial \phi_i}{\partial \mathcal{P}} \right) + \frac{1}{\theta_i} \cdot \left[\left(\frac{\phi_i}{\theta_i} - 1 \right) \cdot \left(\frac{\partial \theta_i}{\partial \mathcal{P}} \right) - \left(\frac{\partial \phi_i}{\partial \mathcal{P}} \right) \right] \right\} \\ &-\frac{z}{2} \left(\ln \left(\frac{\phi_i}{\theta_i} \right) + 1 - \frac{\phi_i}{\theta_i} \right) \frac{\partial q_i}{\partial \mathcal{P}} \end{aligned} \quad (3.7)$$

Além disto, esta contribuição depende basicamente das frações molares (ou, em última instância, dos números de mols) de cada componente na mistura e das áreas totais dos subgrupos, de modo que suas derivadas serão não nulas somente na derivação do F-SAC com relação a eles, conforme será mostrado adiante.

O termo residual pode ser derivado da seguinte forma:

$$\frac{\partial \ln \gamma_i^{res}}{\partial \mathcal{P}} = \sum_{m \in i} \frac{1}{a_{eff}} \cdot \left[(\ln \Gamma_m^s - \ln \Gamma_m^i) \cdot \frac{\partial Q_m^i}{\partial \mathcal{P}} + Q_m^i \cdot \left(\frac{\partial \ln \Gamma_m^s}{\partial \mathcal{P}} - \frac{\partial \ln \Gamma_m^i}{\partial \mathcal{P}} \right) \right] \quad (3.8)$$

Percebe-se, no segundo termo dentro do somatório do lado direito da equação anterior, que esta necessita dos valores das derivadas dos coeficientes de atividade de cada segmento, $\ln \Gamma_m^s$ e $\ln \Gamma_m^i$. Conforme já mencionado, estes coeficientes são calculados, numericamente, a partir da resolução da equação auto-consistente. Logo, as derivadas de $\ln \Gamma_m^s$ e $\ln \Gamma_m^i$ poderiam ser obtidas derivando-se a [Equação 2.39](#), o que também levaria a equações algébricas sem solução analítica.

Para evitar a necessidade de um outro método numérico, poderia-se, alternativamente, obter a derivada anterior a partir de perturbações numéricas do modelo

já convergido ($\ln \Gamma_m^s$ e $\ln \Gamma_m^s$ já calculados), mudando infimamente cada um dos parâmetros por vez. Esta alternativa é interessante, entretanto, tende a aumentar muito o tempo computacional necessário, pois a cada perturbação de um parâmetro, a [Equação 2.39](#) tem que ser resolvida novamente, para a mistura e para os puros ($n + 1$ vezes). Ainda poderia-se reestruturar todo o código para possibilitar o uso da derivação algorítmica, conforme é estudado no livro do Griewank ([Andreas Griewank, 2008](#)).

Neste trabalho, escolheu-se uma solução alternativa, que é a aplicação do Teorema da Função Implícita. Para utilizar este teorema, podemos escrever genericamente a equação auto-consistente³ num formato residual:

$$\begin{aligned} \ln \Gamma_m &= -\ln \left[\sum_n p_n \Gamma_n \exp \left(\frac{-\Delta W_{m,n}}{RT} \right) \right] \Rightarrow \\ \Gamma_m &= \frac{1}{\left[\sum_n p_n \Gamma_n \exp \left(\frac{-\Delta W_{m,n}}{RT} \right) \right]} \Rightarrow \\ F_m &= \Gamma_m \cdot \left[\sum_n p_n \Gamma_n \exp \left(\frac{-\Delta W_{m,n}}{RT} \right) \right] - 1 = 0 \end{aligned} \quad (3.9)$$

É importante notar que, apesar de a função F_m na [Equação 3.9](#) ter sido escrita num formato escalar, na verdade trata-se de um *vetor*, que será chamado neste trabalho de \mathbf{F} , cujo tamanho é o número de segmentos da mistura ou do componente puro.

Desta forma, pode-se obter as derivadas a partir da aplicação direta do Teorema da Função Implícita, conforme segue:

$$\frac{\partial \Gamma}{\partial \mathcal{P}} = - \left(\frac{\partial \mathbf{F}}{\partial \Gamma} \right)_P^{-1} \cdot \left(\frac{\partial \mathbf{F}}{\partial \mathcal{P}} \right)_\Gamma \quad (3.10)$$

Onde Γ é o vetor dos coeficientes de atividade de cada segmento, na mistura ou no puro. Se compararmos a [Equação 3.10](#) com a equação do Teorema da Função Implícita (descrita na [Seção 2.5](#)), é fácil perceber que a variável independente (o “x”) é o parâmetro \mathcal{P} ; e a variável dependente (o “y”) é o vetor Γ , pois ele depende tanto de \mathcal{P} quanto dele mesmo (veja a equação auto-consistente).

Assim, a [Equação 3.10](#) é na verdade uma equação *matricial*, na qual a derivada do vetor \mathbf{F} com relação ao vetor de segmentos (da mistura ou do puro) resulta numa

³Para a mistura ou para cada um dos componentes puros.

matriz quadrada, cujos números de linhas e colunas são iguais ao número de segmentos. De forma semelhante, as derivadas dos vetores \mathbf{F} e $\mathbf{\Gamma}$ com relação ao parâmetro escalar \mathcal{P} resultam em vetores cujos tamanhos são o número de segmentos.

Derivando-se [Equação 3.9](#) com respeito a \mathcal{P} e mantendo-se todos os Γ constantes:

$$\left(\frac{\partial F_m}{\partial \mathcal{P}}\right)_{\mathbf{\Gamma}} = \Gamma_m \cdot \sum_n B_{m,n} \cdot \left[\frac{\partial p_n}{\partial \mathcal{P}} + \frac{\partial}{\partial \mathcal{P}} \left(\frac{-\Delta W_{m,n}}{RT}\right) \cdot p_n\right] \quad (3.11)$$

onde $B_{m,n} = \Gamma_n \cdot \exp\left[\frac{-\Delta W_{m,n}}{RT}\right]$. As duas derivadas do lado direito da [Equação 3.11](#) são agora facilmente obtidas com respeito a *qualquer* variável independente, como a temperatura, número de mols de algum componente ou os parâmetros do modelo.

Derivando-se [Equação 3.9](#) com respeito ao coeficiente de atividade de um segmento qualquer, Γ_j , e mantendo-se \mathcal{P} e todos os outros Γ constantes:

$$\frac{\partial F_m}{\partial \Gamma_j} = \Gamma_m \cdot A_{m,j} + \frac{\partial \Gamma_m}{\partial \Gamma_j} \sum_n A_{m,n} \cdot \Gamma_n \quad (3.12)$$

onde $A_{m,n} = p_n \exp(-\Delta W_{m,n}/RT)$; e a derivada parcial $\partial \Gamma_m / \partial \Gamma_j = 1$ para $m = j$ e $\partial \Gamma_m / \partial \Gamma_j = 0$ para $m \neq j$.

Finalmente, pode-se reescrever a [Equação 3.8](#) de modo que esta esteja em função das derivadas de Γ_m^s e Γ_m^i em vez dos seus logaritmos:

$$\frac{\partial \ln \gamma_i^{\text{res}}}{\partial \mathcal{P}} = \sum_{m \in i} \frac{1}{a_{\text{eff}}} \cdot \left[(\ln \Gamma_m^s - \ln \Gamma_m^i) \cdot \frac{\partial Q_m^i}{\partial \mathcal{P}} + Q_m^i \cdot \left(\frac{1}{\Gamma_m^s} \cdot \frac{\partial \Gamma_m^s}{\partial \mathcal{P}} - \frac{1}{\Gamma_m^i} \cdot \frac{\partial \Gamma_m^i}{\partial \mathcal{P}} \right) \right] \quad (3.13)$$

3.3.3 Limites e Restrições

Os limites superiores e inferiores do modelo F-SAC são tratados internamente ao algoritmo MMA, devendo-se apenas informá-los nos códigos computacionais dentro da chamada do método de otimização. A restrição da [Equação 3.3](#) depende apenas dos parâmetros eletrostáticos (Q_k^+ , Q_k^- e σ_k^+). As suas derivadas com relação a estes parâmetros serão mostradas nas próximas seções e bastam ser implementadas na função do MMA que chama as derivadas das restrições.

3.3.4 Derivada com Relação à Temperatura

Substituindo-se a temperatura (T), no lugar do parâmetro \mathcal{P} , na derivada do modelo F-SAC, temos:

$$\frac{\partial \ln \gamma_i}{\partial T} = \frac{\partial \ln \gamma_i^{comb}}{\partial T} + \frac{\partial \ln \gamma_i^{res}}{\partial T} \quad (3.14)$$

Conforme observado anteriormente, a contribuição combinatorial *não depende* da temperatura (veja a [Equação 2.32](#)), portanto, a sua derivada é zero:

$$\frac{\partial \ln \gamma_i^{comb}}{\partial T} = 0 \quad (3.15)$$

Para derivação da contribuição residual, substitui-se a temperatura no lugar do parâmetro \mathcal{P} da [Equação 3.13](#), e, sabendo-se que $\partial Q_m^i / \partial T = 0$ ⁴ tem-se:

$$\frac{\partial \ln \gamma_i^{res}}{\partial T} = \sum_{m \in i} \frac{Q_m^i}{a_{eff}} \cdot \left(\frac{1}{\Gamma_m^s} \cdot \frac{\partial \Gamma_m^s}{\partial T} - \frac{1}{\Gamma_m^i} \cdot \frac{\partial \Gamma_m^i}{\partial T} \right) \quad (3.16)$$

De forma semelhante, aplicando-se a temperatura no lugar do parâmetro \mathcal{P} da [Equação 3.11](#) e sabendo-se que $\partial p_n / \partial T = 0$ ⁵:

$$\left(\frac{\partial F_m}{\partial T} \right)_{\Gamma} = \Gamma_m \cdot \sum_n B_{m,n} \cdot p_n \cdot \frac{\partial}{\partial T} \left(\frac{-\Delta W_{m,n}}{RT} \right) \quad (3.17)$$

A derivada que aparece no lado direito da [Equação 3.17](#) é dada por⁶:

$$\frac{\partial}{\partial T} \left(\frac{-\Delta W_{m,n}}{RT} \right) = \left(\frac{\Delta W_{m,n}}{RT^2} \right) \quad (3.18)$$

Finalmente, a equação [Equação 3.17](#) pode ser aplicada, juntamente com a [Equação 3.12](#), no Teorema da Função Implícita ([Equação 3.10](#)), resultando num vetor de derivadas de todos os segmentos (da mistura ou dos componentes puros) com relação à temperatura. Este vetor, se utilizado na [Equação 3.14](#) resulta na derivada do modelo F-SAC com relação à temperatura.

⁴Pois Q_m^i é a área superficial do segmento m da molécula i , ou seja, um parâmetro independente da temperatura.

⁵Pois p_n é a probabilidade de se encontrar um segmento n na mistura ou no líquido puro, ou seja, é um parâmetro também independente da temperatura.

⁶ Pois $\partial \Delta W_{m,n} / \partial T = 0$.

3.3.5 Derivada com Relação ao Número de Mols

A derivada da contribuição combinatorial com relação ao número de mols de um dos componentes da mistura (n_l) é obtida substituindo-se o parâmetro \mathcal{P} na [Equação 3.7](#), sabendo que $\partial q_i / \partial n_l = 0$:

$$\frac{\partial \ln \gamma_i^{\text{comb}}}{\partial n_l} = \left(\frac{1}{\phi'_i} - 1 \right) \cdot \left(\frac{\partial \phi'_i}{\partial n_l} \right) - \frac{z}{2} q_i \left\{ \frac{1}{\phi_i} \cdot \left(\frac{\partial \phi_i}{\partial n_l} \right) + \frac{1}{\theta_i} \cdot \left[\left(\frac{\phi_i}{\theta_i} - 1 \right) \cdot \left(\frac{\partial \theta_i}{\partial n_l} \right) - \left(\frac{\partial \phi_i}{\partial n_l} \right) \right] \right\} \quad (3.19)$$

As expressões que aparecem nesta derivada são:

$$\frac{\partial \phi'_i}{\partial n_l} = \frac{\partial}{\partial n_l} \left(\frac{(V^i)^{3/4}}{\sum_j x_j (V^j)^{3/4}} \right) = \frac{\partial}{\partial n_l} \left(\frac{(V^i)^{3/4}}{\sum_j (n_j/n^T) (V^j)^{3/4}} \right) \quad (3.20)$$

$$\frac{\partial \phi_i}{\partial n_l} = \frac{\partial}{\partial n_l} \left(\frac{V^i}{\sum_j x_j V^j} \right) = \frac{\partial}{\partial n_l} \left(\frac{V^i}{\sum_j (n_j/n^T) V^j} \right) \quad (3.21)$$

$$\frac{\partial \theta_i}{\partial n_l} = \frac{\partial}{\partial n_l} \left(\frac{Q^i}{\sum_j x_j Q^j} \right) = \frac{\partial}{\partial n_l} \left(\frac{Q^i}{\sum_j (n_j/n^T) Q^j} \right) \quad (3.22)$$

Onde $n^T = \sum_j n_j$ é o número total de mols na mistura. Para a expressão de ϕ'_i , tem-se:

$$\frac{\partial}{\partial n_l} \left(\frac{(V^i)^{3/4}}{\sum_j (n_j/n^T) (V^j)^{3/4}} \right) = - \frac{(V^i)^{3/4}}{(\sum_j x_j (V^j)^{3/4})^2} \cdot \left[\sum_j \frac{\partial}{\partial n_l} \left(\frac{n_j}{n^T} \right) \cdot (V^j)^{3/4} \right] \quad (3.23)$$

A derivada que aparece no lado direito da expressão anterior é dada por:

$$\frac{\partial}{\partial n_l} \left(\frac{n_j}{n^T} \right) = \left(\frac{1}{n^T} \right) \left(\frac{\partial n_j}{\partial n_l} \right) - \left(\frac{n_j}{n^{T2}} \right) \left(\frac{\partial n^T}{\partial n_l} \right) \quad (3.24)$$

Finalmente, substituindo tudo na expressão original e com alguns ajustes algébricos⁸, temos:

$$\frac{\partial \phi'_i}{\partial n_l} = - \frac{(V^i)^{3/4}}{(\sum_j x_j (V^j)^{3/4})^2} \cdot \frac{1}{n^T} \left[\sum_j \left(\frac{\partial n_j}{\partial n_l} \right) \cdot (V^j)^{3/4} - \sum_j x_j (V^j)^{3/4} \right] \quad (3.25)$$

⁷ Pois $q_i = Q^i/q$ e Q^i é a área superficial total da molécula i , ou seja, um parâmetro independente do número de mols de qualquer componente na mistura.

⁸ Lembrando que $\partial n^T / \partial n_l = 1$, da termodinâmica básica.

Para as expressões de ϕ_i e θ_i os desenvolvimentos são muito semelhantes, resultando em:

$$\frac{\partial \phi_i}{\partial n_l} = -\frac{V^i}{(\sum_j x_j V^j)^2} \cdot \frac{1}{n^T} \left[\sum_j \left(\frac{\partial n_j}{\partial n_l} \right) \cdot V^j - \sum_j x_j V^j \right] \quad (3.26)$$

$$\frac{\partial \theta_i}{\partial n_l} = -\frac{Q^i}{(\sum_j x_j Q^j)^2} \cdot \frac{1}{n^T} \left[\sum_j \left(\frac{\partial n_j}{\partial n_l} \right) \cdot Q^j - \sum_j x_j Q^j \right] \quad (3.27)$$

Para derivação da contribuição residual, substitui-se n_l no lugar do parâmetro \mathcal{P} da [Equação 3.13](#), e, sabendo-se que $\partial Q_m^i / \partial n_l = 0$ ⁹, obtém-se uma expressão quase idêntica àquela apresentada para a temperatura ([Equação 3.16](#)), com a única diferença de que em vez de T , a variável de derivação é n_l .

Aplicando-se n_l no lugar de \mathcal{P} da [Equação 3.11](#) e sabendo-se que $\Delta W_{m,n}$ não depende do número de mols de nenhum dos componentes da mistura (veja a [Equação 2.44](#)):

$$\left(\frac{\partial F_m}{\partial n_l} \right)_\Gamma = \Gamma_m \cdot \left(\sum_n B_{m,n} \right) \cdot \frac{\partial p_n}{\partial n_l} \quad (3.28)$$

É interessante observar que, no caso das substâncias puras, a derivada que aparece no lado direito da [Equação 3.28](#) é zero, pois $p_m^i = Q_m^i / Q^i$, ou seja, não depende de n_l . Esta conclusão é óbvia e coerente, pois a derivada dos perfis- σ dos componentes puros com relação à composição da mistura deve ser nula. Tal conclusão resulta em $\partial \Gamma_m^i / \partial n_l = 0$, o que também é coerente com o modelo estudado, pois os coeficientes de atividade de cada segmento nas substâncias puras também não podem depender da composição da mistura.

Para a mistura, tem-se:

$$\frac{\partial p_m^s}{\partial n_l} = \frac{\partial}{\partial n_l} \left(\frac{(n_i / n^T) \cdot Q_m^i}{\sum_j (n_j / n^T) \cdot Q^j} \right) \quad (3.29)$$

Esta equação é facilmente diferenciável, de modo que, com algum trabalho de derivação e de álgebra, chega-se em:

$$\frac{\partial p_m^s}{\partial n_l} = \frac{1}{n^T} \left\{ \frac{Q_m^i}{\sum_j x_j Q^j} \left(\frac{\partial n_i}{\partial n_l} - x_i \right) - \frac{x_i Q_m^i}{(\sum_j x_j Q^j)^2} \left[\sum_j \left(\frac{\partial n_j}{\partial n_l} - x_j \right) \cdot Q^j \right] \right\} \quad (3.30)$$

⁹Conforme observado anteriormente.

Finalmente, a derivada parcial $\partial n_{i,j}/\partial n_l = 1$ para $i, j = l$ e $\partial n_{i,j}/\partial n_l = 0$ para $i, j \neq l$.

3.3.6 Derivada com Relação à Área Total de um Subgrupo

Substituindo a área total do subgrupo k de um determinado componente i (Q_k^i) na [Equação 3.7](#):

$$\frac{\partial \ln \gamma_i^{\text{comb}}}{\partial Q_k^i} = -\frac{z}{2} q_i \left(\frac{1}{\theta_i} \cdot \left(\frac{\phi_i}{\theta_i} - 1 \right) \cdot \frac{\partial \theta_i}{\partial Q_k^i} \right) - \frac{z}{2} \left(\ln \left(\frac{\phi_i}{\theta_i} \right) + 1 - \frac{\phi_i}{\theta_i} \right) \frac{\partial q_i}{\partial Q_k^i} \quad (3.31)$$

$$\frac{\partial \theta_i}{\partial Q_k^i} = \frac{1}{\sum_j x_j Q^j} \cdot \frac{\partial Q^i}{\partial Q_k^i} - \frac{Q^i}{(\sum_j x_j Q^j)^2} \cdot \sum_j x_j \frac{\partial Q^j}{\partial Q_k^i} \quad (3.32)$$

$$\frac{\partial q_i}{\partial Q_k^i} = \frac{1}{q} \frac{\partial Q^i}{\partial Q_k^i} \quad (3.33)$$

A área total da molécula i é dada pela soma das áreas de todos os segmentos pertencentes a i : $Q^i = \sum_{m \in i} Q_m^i$. Entretanto, esta área também pode ser calculada como a soma de todas as áreas totais dos subgrupos que compõe i , cada uma destas áreas multiplicadas pelo respectivo número de vezes que cada um dos subgrupos aparece na molécula, ou seja: $Q^i = \sum_{k \in i} \nu_k \cdot Q_k^i$, onde ν_k é o número de vezes que o subgrupo k aparece na molécula i conforme a expressão dada em [2.43](#). Logo a derivada no lado direito da [Equação 3.33](#) pode ser escrita como segue:

$$\frac{\partial Q^i}{\partial Q_k^i} = \nu_k \quad (3.34)$$

Observe que as derivadas da contribuição combinatorial de i serão zero com relação às áreas dos subgrupos que não compõem tal molécula. Este mesmo comportamento é observado na derivada do termo de coeficiente de atividade dos segmentos do componente puro (Γ_m^i) na contribuição residual.

Para a contribuição residual, basta substituir \mathcal{P} por Q_k^i na [Equação 3.13](#). É interessante observar que, para o modelo F-SAC, algum subgrupo k que forma uma determinada molécula i possui 3 segmentos (1 positivo - Q_k^+ , 1 negativo - Q_k^- e 1 neutro - Q_k^0). Desta forma, caso o segmento m seja positivo ou negativo, sua derivada com

relação à área total do subgrupo será zero, pois Q_k^+ e Q_k^- também são parâmetros do modelo e, portanto, são variáveis *independentes*. Caso o segmento seja neutro, então deve-se lembrar que $Q_k^o = Q_k - Q_k^+ - Q_k^-$, portanto¹⁰:

$$\frac{\partial Q_m^i}{\partial Q_k^i} = \nu_k \quad (3.35)$$

onde m é o segmento neutro pertencente ao subgrupo k e o ν_k aparece seguindo-se a mesma lógica anterior.

Aplicando-se Q_k^i na [Equação 3.11](#) e lembrando que $\partial(-\Delta W_{m,n}/RT)/\partial Q_m^i = 0$ ¹¹:

$$\left(\frac{\partial F_m}{\partial Q_k^i}\right)_\Gamma = \Gamma_m \cdot \sum_{\sigma_n} B_{m,n} \cdot \frac{\partial p_n}{\partial Q_k^i} \quad (3.36)$$

Para o componente puro:

$$\frac{\partial p_m^i}{\partial Q_k^i} = \frac{1}{Q^i} \cdot \left(\frac{\partial Q_m^i}{\partial Q_k^i} - \frac{Q_m^i}{Q^i} \cdot \frac{\partial Q^i}{\partial Q_k^i}\right) \quad (3.37)$$

Para a mistura:

$$\frac{\partial p_m^s}{\partial Q_k^i} = \frac{x_i}{\sum_j x_j Q^j} \cdot \frac{\partial Q_m^i}{\partial Q_k^i} - \frac{x_i Q_m^i}{(\sum_j x_j Q^j)^2} \cdot \sum_j x_j \frac{\partial Q^j}{\partial Q_k^i} \quad (3.38)$$

onde $\partial Q_m^i/\partial Q_k^i = \nu_k$, caso m seja neutro e $\partial Q_m^i/\partial Q_k^i = 0$ caso m seja positivo ou negativo.

3.3.7 Derivada com Relação às Áreas Positivas e Negativas de um Grupo

Nesta seção, para evitar a repetição de equações e facilitar a nomenclatura, adotou-se o seguinte símbolo $Q_k^{+/-}$ para representar a área positiva (Q_k^+) ou a área negativa (Q_k^-) de um determinado grupo k .

A derivada da contribuição combinatorial com relação à $Q_k^{+/-}$ é zero, pois as áreas positiva e negativa não aparecem no cálculo desta contribuição. Um observador não totalmente atento poderia argumentar que tais parâmetros aparecem sim

¹⁰Logo $\partial Q_k^o/\partial Q_k^i = 1$.

¹¹Veja a [Equação 2.44](#).

no cálculo da área total do subgrupo k , tendo em vista que $Q^i = \sum_{m \in i} Q_m^i$. Entretanto, conforme mostrado anteriormente, a área total do subgrupo também é um parâmetro do modelo, assim como as áreas positivas e negativas do grupo ao qual o subgrupo pertence, de modo que a sua diferença é englobada pela parte neutra do subgrupo ($Q_k^\circ = Q_k - Q_k^+ - Q_k^-$).

Para a contribuição residual, basta substituir \mathcal{P} por $Q_k^{+/-}$ na [Equação 3.13](#). Aplicando-se $Q_k^{+/-}$ na [Equação 3.11](#), tem-se:

$$\left(\frac{\partial F_m}{\partial Q_k^{+/-}} \right)_\Gamma = \Gamma_m \cdot \sum_n B_{m,n} \cdot \left[\frac{\partial p_n}{\partial Q_k^{+/-}} - \frac{1}{RT} \frac{\partial \Delta W_{m,n}}{\partial Q_k^{+/-}} \cdot p_n \right] \quad (3.39)$$

A obtenção de $\partial p_n / \partial Q_k^{+/-}$ é semelhante ao mostrado na seção anterior, para os componentes puros e para a mistura, com a diferença que as derivadas de Q^i são nulas. As derivadas de Q_m^i dependem se o segmento m é positivo, negativo ou neutro, conforme a [Tabela 3.1](#):

Tabela 3.1: Valores da derivada da área do segmento m com relação à $Q_k^{+/-}$.

	$\partial/\partial Q_k^+$	$\partial/\partial Q_k^-$
Q_k^+	ν	0
Q_k^-	0	ν
Q_k°	$-\nu$	$-\nu$

A derivada de $\Delta W_{m,n}$ é obtida derivando-se a [Equação 2.44](#) com respeito a $Q_k^{+/-}$ ¹²:

$$\frac{\partial \Delta W_{m,n}}{\partial Q_k^{+/-}} = \alpha'(\sigma_m + \sigma_n) \cdot \left(\frac{\partial \sigma_m}{\partial Q_k^{+/-}} + \frac{\partial \sigma_n}{\partial Q_k^{+/-}} \right) \quad (3.40)$$

De maneira semelhante às áreas, a derivada de σ_m (e de σ_n) depende se o segmento é positivo, negativo ou neutro. É fácil perceber que, se o segmento for positivo, então a sua derivada será zero, pois este é um dos parâmetros do modelo (portanto independente). Para o segmento neutro (carga nula), a derivada também será zero, pois a sua carga é constante (e a derivada de uma constante é zero). Por outro lado, se o

¹²Observando-se a [Equação 2.44](#), tem-se que $\partial E_{m,n}^{\text{HB}} / \partial Q_k^{+/-} = 0$.

segmento for negativo, deve-se derivar a expressão $\sigma_k^- = -\sigma_k^+ Q_k^+ / Q_k^-$. Tais resultados são sumarizados na [Tabela 3.2](#):

Tabela 3.2: Valores da derivada da carga do segmento m caso este seja positivo, negativo ou neutro.

	$\partial/\partial Q_k^+$	$\partial/\partial Q_k^-$
σ_m^+	0	0
σ_m^-	σ_m^+ / Q_k^-	$-\sigma_m^+ / (Q_k^-)^2$
σ_m^o	0	0

3.3.8 Derivada com Relação à Carga Positiva de um Grupo

A derivada da contribuição combinatorial com relação à σ_k^+ é zero, pois a carga positiva não aparece no cálculo desta contribuição.

Para a contribuição residual, basta substituir \mathcal{P} por σ_k^+ na [Equação 3.13](#). Aplicando-se σ_k^+ na [Equação 3.11](#) e lembrando que $\partial p_n / \partial \sigma_k^+ = 0$ ¹³, tem-se:

$$\left(\frac{\partial F_m}{\partial \sigma_k^+} \right)_\Gamma = \Gamma_m \cdot \sum_n B_{m,n} \cdot \left[-\frac{1}{RT} \frac{\partial \Delta W_{m,n}}{\partial \sigma_k^+} \cdot p_n \right] \quad (3.41)$$

A derivada de $\Delta W_{m,n}$ com respeito a σ_k^+ é¹⁴:

$$\frac{\partial \Delta W_{m,n}}{\partial \sigma_k^+} = \alpha'(\sigma_m + \sigma_n) \cdot \left(\frac{\partial \sigma_m}{\partial \sigma_k^+} + \frac{\partial \sigma_n}{\partial \sigma_k^+} \right) \quad (3.42)$$

De maneira semelhante ao realizado na seção anterior, a derivada de σ_m (e de σ_n) depende se o segmento é positivo, negativo ou neutro. É fácil perceber que, se o segmento for neutro (carga nula), então a sua derivada será zero, pois a sua carga é constante (e a derivada de uma constante é zero). Para o segmento positivo, a derivada será igual a 1, caso $m = k$ e igual a zero caso $m \neq k$. Se o segmento for negativo, deve-se derivar a expressão $\sigma_k^- = -\sigma_k^+ Q_k^+ / Q_k^-$, resultando em $\partial \sigma_m^- / \partial \sigma_k^+ = -Q_m^+ / Q_m^-$, caso $m = k$ e $\partial \sigma_m^- / \partial \sigma_k^+ = 0$ caso $m \neq n$.

¹³Pois p_n não depende da carga.

¹⁴Observando-se a [Equação 2.44](#), tem-se que $\partial E_{m,n}^{\text{HB}} / \partial \sigma_k^+ = 0$.

3.3.9 Derivada com Relação à Ligação de Hidrogênio

A derivada da contribuição combinatorial com relação ao parâmetro binário de um par de grupos doador–receptor ($k-k'$) de ligação de hidrogênio, $E_{k,k'}^{\text{HB}}$, é zero, pois tal parâmetro não aparece no cálculo desta contribuição.

Para a contribuição residual, basta substituir \mathcal{P} por $E_{k,k'}^{\text{HB}}$ na [Equação 3.13](#), sabendo que a derivada de Q_m^i com relação a este parâmetro é zero. Aplicando-se $E_{k,k'}^{\text{HB}}$ na [Equação 3.11](#), tem-se (a derivada de p_n também é zero):

$$\left(\frac{\partial F_m}{\partial E_{k,k'}^{\text{HB}}} \right)_{\mathbf{r}} = \Gamma_m \cdot \sum_n B_{m,n} \cdot \left(-\frac{1}{RT} \frac{\partial \Delta W_{m,n}}{\partial E_{k,k'}^{\text{HB}}} \cdot p_n \right) \quad (3.43)$$

onde:

$$\frac{\partial \Delta W_{m,n}}{\partial E_{k,k'}^{\text{HB}}} = -\frac{1}{2} \cdot \frac{\partial E_{m,n}^{\text{HB}}}{\partial E_{k,k'}^{\text{HB}}} \quad (3.44)$$

Finalmente, a derivada parcial $\partial E_{m,n}^{\text{HB}} / \partial E_{k,k'}^{\text{HB}} = 1$ se $m, n = k, k'$; e $\partial E_{m,n}^{\text{HB}} / \partial E_{k,k'}^{\text{HB}} = 0$ se $m, n \neq k, k'$.

3.3.10 Obtenção dos Intervalos de Confiança dos Parâmetros Estimados

Os intervalos de confiança dos parâmetros θ_i foram obtidos a partir da [Equação 3.45](#).

$$\theta_i = \hat{\theta}_i \pm s_{\theta_i} \cdot t_{NE-NP}^{1-\alpha/2} \quad (3.45)$$

onde $\hat{\theta}_i$ é um parâmetro estimado i , s_{θ_i} é o desvio padrão de $\hat{\theta}_i$, t é a distribuição t-student para uma probabilidade $1 - \alpha/2$ e $NE - NP$ graus de liberdade, $NE = 2979$ é o número de dados experimentais, $NP = 142$ é o número total de parâmetros e α é definida para um nível de confiança de $1 - \alpha$. Neste trabalho foi arbitrado um nível de confiança de 95%, de modo que $t_{2979-142}^{1-0,05/2} = 1,961$.

O desvio padrão de cada parâmetro é calculado como a raiz quadrada da sua variância ($s_{\theta_i}^2$). As variâncias dos parâmetros são obtidas a partir da diagonal principal

da matriz de covariância (\mathbf{V}_θ) dos parâmetros.

$$\mathbf{V}_\theta = \begin{bmatrix} s_{\theta_1}^2 & s_{\theta_1, \theta_2} & \dots & s_{\theta_1, \theta_{NP}} \\ s_{\theta_1, \theta_2} & s_{\theta_2}^2 & \dots & s_{\theta_2, \theta_{NP}} \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_{NP}, \theta_1} & s_{\theta_{NP}, \theta_2} & \dots & s_{\theta_{NP}}^2 \end{bmatrix} \quad (3.46)$$

A matriz de covariância pode ser calculada de forma aproximada da seguinte forma (DRAPER; HIMMELBLAU, 2006):

$$\mathbf{V}_\theta = (\mathbf{B}^T \mathbf{V}_y^{-1} \mathbf{B})^{-1} \quad (3.47)$$

onde \mathbf{V}_y é a matriz de variância dos dados experimentais e \mathbf{B} é a matriz de NE linhas e NP colunas formada pela derivada primeira do modelo (\mathbf{y}^m) com relação aos parâmetros (θ), aplicada em cada ponto experimental, conforme mostrado na Equação 3.48.

$$\mathbf{B} = \frac{\partial \mathbf{y}^m}{\partial \theta} = B = \begin{bmatrix} \frac{\partial y_{1,1}^m}{\partial \theta_1} & \dots & \frac{\partial y_{1,NP}^m}{\partial \theta_{NP}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{NE,1}^m}{\partial \theta_1} & \dots & \frac{\partial y_{NE,NP}^m}{\partial \theta_{NP}} \end{bmatrix} \quad (3.48)$$

Conforme observado na Subseção 2.3.2, os dados experimentais utilizados neste trabalho foram obtidos a partir de um banco de dados coletado de diversos autores, os quais não informam os desvios dos experimentos realizados. Assim, adotou-se um valor constante de variância experimental de 0,01. Semelhantemente, arbitrou-se que os dados experimentais não estão correlacionados entre si, logo a covariância entre os experimentos é 0,00. Ou seja, neste trabalho \mathbf{V}_y é uma matriz diagonal, cujos valores da sua diagonal principal são 0,01.

Capítulo 4

Resultados e Discussão

Neste capítulo serão apresentados os resultados obtidos no presente trabalho. Na [Seção 4.1](#) serão mostrados os resultados das derivadas do modelo F-SAC e as medições de tempos computacionais. Na [Seção 4.2](#) serão mostrados resultados de estimações de parâmetros utilizando o algoritmo MMA. Na [Seção 4.3](#) será mostrado um estudo de caso simplificado que demonstra o efeito do raio de contato padrão (parâmetro universal do modelo) no F-SAC.

4.1 Resultados da Derivação

4.1.1 Introdução

Nesta seção, serão mostrados os resultados das derivadas do modelo F-SAC com relação à temperatura, ao número de mols e aos parâmetros do modelo. Os resultados das derivadas obtidas utilizando a metodologia proposta neste trabalho foram comparados com os resultados de derivadas numéricas e com dados obtidos na literatura, quando possível. Além disto, serão apresentados resultados de tempos computacionais necessários para o cálculo das derivadas.

Convencionou-se denominar de derivadas *implícitas* aquelas obtidas via Teorema da Função Implícita, conforme descrito nos Capítulos [2-3](#). As derivadas *numéricas*, por sua vez, são aquelas obtidas por perturbação numérica do modelo F-SAC através do método das diferenças finitas. Optou-se por utilizar derivadas numéricas

de primeira ordem (diferenças progressivas) em todos os testes, em vez de derivadas com maior ordem por serem de mais fácil implementação. É conhecido o fato de que as derivadas por diferenças progressivas serão menos precisas, entretanto não é o objetivo deste trabalho obter derivadas numéricas com alta precisão, apenas obtê-las para fins de comparação com as derivadas implícitas.

4.1.2 Derivadas com Relação à Temperatura

As entalpias de excesso foram calculadas utilizando as derivadas do modelo F-SAC com respeito à temperatura, conforme descrito na [Equação 2.28](#):

$$\frac{h^E}{RT} = -T \sum x_i \frac{\partial \ln \gamma_i}{\partial T} \quad (4.1)$$

As derivadas numéricas foram obtidas da seguinte forma:

$$\left(\frac{\partial \ln \gamma_i}{\partial T} \right)^{\text{numérico}} = \frac{\ln \gamma_i|_{T+\delta T} - \ln \gamma_i|_T}{\delta T} + \mathcal{O}(\delta T) \quad (4.2)$$

onde escolheu-se $\delta T = 0,001 \text{ K}$.

A [Tabela 4.1](#) mostra os desvios médios absolutos (AAD_{h^E}) entre as entalpias de excesso para a mistura binária de etanol e água em 04 temperaturas distintas. Os AAD_{h^E} foram calculados, para cada temperatura considerada, como segue:

$$\text{AAD}_{h^E} = \frac{1}{N_y} \sum_{i=1}^{N_y} |h_{i,\text{implcito}}^E - h_{i,\text{numérico}}^E| \quad (4.3)$$

onde $N_y = 40$ é o número de pontos calculados.

Tabela 4.1: Desvios Médios Absolutos (AAD_{h^E}) e Maiores Desvios ($|h_{i,\text{implcito}}^E - h_{i,\text{numérico}}^E|$).

Temperatura (K)	AAD_{h^E}	Maior desvio
323,15	$1,87 \times 10^{-5}$	$3,61 \times 10^{-5}$
343,15	$1,41 \times 10^{-5}$	$2,77 \times 10^{-5}$
363,15	$1,08 \times 10^{-5}$	$2,11 \times 10^{-5}$
383,15	$8,93 \times 10^{-6}$	$1,76 \times 10^{-5}$

Conforme pode ser observado, os desvios foram na ordem de 10^{-5} . De maneira semelhante, foram realizados outros testes, com diversos sistemas e temperaturas, que

levaram a resultados semelhantes aos apresentados. Entretanto, estes resultados não serão reproduzidos aqui pois seriam repetitivos.

4.1.3 Derivadas com Relação ao Número de Mols – Aplicação da Equação de Gibbs-Duhem

Conforme mostrado na [Seção 2.1](#), a equação de Gibbs-Duhem é dada pela [Equação 2.8](#). Se substituirmos o \bar{K}_i desta equação pela energia de Gibbs parcial molar em excesso, conforme a [Equação 2.23](#), tem-se:

$$\sum_i x_i \cdot \frac{\partial \ln \gamma_i}{\partial n_j} = 0 \quad (4.4)$$

onde n_j é o número de mols de um determinado componente j pertencente à mistura.

É óbvio que o modelo F-SAC é consistente termodinamicamente, pois seu equacionamento é baseado no modelo COSMO-SAC que teve a sua consistência termodinâmica provada por [Lin e Sandler \(2002\)](#). No desenvolvimento do COSMO-SAC, os autores, inclusive, criticam o modelo COSMO-RS de [Klamt \(1995\)](#), pois este, originalmente, não era consistente termodinamicamente. [Lin e Sandler \(2002\)](#) descobriram que o termo combinatorial proposto por [Klamt \(1995\)](#) não respeitava Gibbs-Duhem; tal equívoco foi corrigido nos trabalhos futuros do COSMO-RS

Com relação ao F-SAC, em nenhum momento do seu desenvolvimento esta propriedade foi realmente testada. Portanto, é interessante observar o comportamento do F-SAC, tendo em vista o resultado que já é esperado. Os resultados para o mesmo sistema testado por [Lin e Sandler \(2002\)](#) são dados na [Tabela 4.2](#), na qual as derivadas foram calculadas com respeito ao número de mols do acetato de metila (x_1).

As derivadas numéricas do modelo F-SAC foram obtidas por uma aproximação por diferenças finitas progressivas de primeira ordem:

$$\left(\frac{\partial \ln \gamma_i}{\partial n_i} \right)^{\text{numérico}} = \frac{\ln \gamma_i|_{n_i+\delta n_i} - \ln \gamma_i|_{n_i}}{\delta n_i} + \mathcal{O}(\delta n_i) \quad (4.5)$$

onde escolheu-se $\delta n_i = 0,0001$.

As derivadas obtidas por [Lin e Sandler \(2002\)](#) também foram calculadas por uma aproximação por diferenças finitas progressivas de primeira ordem, de maneira

Tabela 4.2: Resultados de Gibbs-Duhem, calculado pela [Equação 4.4](#), para o sistema acetato de metila + água, a 330,5 K.

x_1	COSMO-SAC	F-SAC	
		Implícito	Numérico
$2,40 \times 10^{-9}$	0	$5,08 \times 10^{-6}$	0,00126
$2,40 \times 10^{-6}$	-0,002	$5,17 \times 10^{-6}$	0,00126
$2,40 \times 10^{-4}$	0,001	$5,19 \times 10^{-6}$	0,00126
0,0025	-0,003	$5,19 \times 10^{-6}$	0,00123
0,0049	0	$5,10 \times 10^{-6}$	0,00121
0,013	-0,001	$4,76 \times 10^{-6}$	0,00112
0,026	-0,002	$4,10 \times 10^{-6}$	0,00100
0,041	-0,002	$3,67 \times 10^{-6}$	$8,80 \times 10^{-4}$
0,057	-0,001	$3,28 \times 10^{-6}$	$7,70 \times 10^{-4}$
0,075	-0,002	$2,97 \times 10^{-6}$	$6,65 \times 10^{-4}$
0,094	-0,002	$2,51 \times 10^{-6}$	$5,73 \times 10^{-4}$
0,116	-0,002	$2,20 \times 10^{-6}$	$4,84 \times 10^{-4}$
0,14	-0,001	$1,77 \times 10^{-6}$	$4,05 \times 10^{-4}$
0,166	-0,002	$1,59 \times 10^{-6}$	$3,37 \times 10^{-4}$
0,196	-0,001	$1,26 \times 10^{-6}$	$2,73 \times 10^{-4}$
0,229	-0,001	$9,28 \times 10^{-7}$	$2,18 \times 10^{-4}$
0,267	-0,001	$6,68 \times 10^{-7}$	$1,70 \times 10^{-4}$
0,311	-0,001	$5,30 \times 10^{-7}$	$1,28 \times 10^{-4}$
0,362	-0,001	$3,28 \times 10^{-7}$	$9,27 \times 10^{-5}$
0,422	-0,001	$2,25 \times 10^{-7}$	$6,38 \times 10^{-5}$
0,493	0	$1,41 \times 10^{-7}$	$4,05 \times 10^{-5}$
0,579	0,002	$9,97 \times 10^{-8}$	$2,29 \times 10^{-5}$
0,686	0,012	$4,21 \times 10^{-8}$	$1,03 \times 10^{-5}$
0,822	0,026	$2,13 \times 10^{-8}$	$2,76 \times 10^{-6}$
0,923	0,013	$6,57 \times 10^{-9}$	$5,06 \times 10^{-7}$
0,96	0,062	$4,05 \times 10^{-9}$	$1,76 \times 10^{-7}$
0,996	0,006	$5,17 \times 10^{-10}$	$1,14 \times 10^{-8}$
1	0,001	0	0

semelhante à [Equação 4.5](#), entretanto, os autores não reportaram no seu trabalho qual o valor de δn_j adotado. Conforme pode-se observar, os valores de Gibbs-Duhem de [Lin e Sandler \(2002\)](#) estão na ordem de $10^{-2} \sim 10^{-3}$. Semelhantemente, os valores de Gibbs-Duhem do F-SAC obtidos numericamente estão na ordem de 10^{-2} ou menores. Como estas derivadas foram obtidas numericamente, pequenas flutuações são aceitas. Por outro lado, os resultados implícitos de Gibbs-Duhem foram menores que 10^{-5} , indicando valores de derivadas mais precisos.

4.1.4 Derivadas com Relação aos Parâmetros do Modelo F-SAC

Nas seções anteriores, a metodologia aplicada neste trabalho para obtenção das derivadas do modelo F-SAC, sem a necessidade de perturbações numéricas, foi testada. Conforme observado, a metodologia funcionou de maneira correta para a temperatura e número de mols. Com esta premissa, foi realizada a implementação das derivadas do F-SAC com relação aos seus parâmetros, utilizando o equacionamento mostrado no [Capítulo 3](#)

Em geral, a derivação analítica, conforme realizada neste trabalho, pode apresentar vários problemas, como erros de derivação, erros de implementação, etc, principalmente para modelos com equações mais complicadas ([BIEGLER, 2010](#); [HIMMELBLAU et al., 2001](#); [PRESS et al., 1992](#)). Neste caso, é muito importante ter *certeza* de que as derivadas implementadas estão totalmente corretas. Inclusive, se houver algum erro nas derivadas, o método de otimização poderá não convergir, prejudicando toda a estimação de parâmetros, que é o objetivo final deste trabalho.

Durante o desenvolvimento das derivadas, a cada implementação, vários testes simples (com derivadas numéricas) foram realizados, de modo a garantir que tudo estivesse correto. De modo a sumarizar, em um resultado consistente, que tais derivadas estão de fato corretas, tendo em vista o grande esforço empregado nesta etapa do trabalho, realizou-se a comparação entre as derivadas implícitas e numéricas, através do cálculo do desvio médio, conforme a [Equação 4.6](#).

$$AAD_{\text{Derivada}} = \frac{1}{N_y} \sum_{i=1}^{N_y} |\text{Derivada}_{i,\text{implícita}} - \text{Derivada}_{i,\text{numérica}}| \quad (4.6)$$

onde $N_y = 44441$ é o número total de pontos calculados.

As derivadas foram realizadas para diferentes misturas binárias, em várias temperaturas, com relação aos seguintes parâmetros do modelo (para cada grupo e subgrupo): Q_k^+ , Q_k^- , σ_k^+ , Q_k e $E_{m,n}^{\text{HB}}$ (somente para pares que fazem ligação de hidrogênio). As misturas escolhidas estão descritas na [Tabela 4.3](#). As temperaturas testadas foram 273,15 K, 303,15 K, 333,15 K, 363,15 K, 393,15 K e 423,15 K.

A escolha dos sistemas e das temperaturas foi arbitrária, apenas foram escolhi-

Tabela 4.3: Sistemas escolhidos (Componente 1 + Componente 2).

Mistura	Mistura	Mistura
Etanol+Água	Água+Ciclohexano	Ciclohexano+N-octano
Etanol+Acetona	Água+N-octano	Ciclohexano+Benzeno
Etanol+Ciclohexano	Água+Benzeno	N-octano+Benzeno
Etanol+N-octano	Acetona+Ciclohexano	Ciclohexano+N-octano
Etanol+Benzeno	Acetona+N-octano	
Água+Acetona	Acetona+Benzeno	

das misturas de substâncias comuns encontradas na indústria e num intervalo de temperatura usual. As composições utilizadas foram obtidas por uma função de números aleatórios (valores entre 0 e 1), chamada dentro do código computacional implementado.

As derivadas numéricas foram calculadas por uma expressão semelhante à [Equação 4.2](#). Como os parâmetros do modelo variam bastante em ordem de grandeza¹, adotou-se uma perturbação de 10^{-6} vezes cada um dos parâmetros (perturbação de 0,0001%)². O desvio médio calculado foi de $AAD_{Derivada} = 1,083 \times 10^{-5}$ e o valor máximo de $|Derivada_{i,implícita} - Derivada_{i,numérica}|$ foi $2,5 \times 10^{-3}$. A partir destes resultados, considerou-se que as derivadas do modelo F-SAC com relação aos seus parâmetros estão corretas, tanto no seu desenvolvimento quanto na implementação.

4.1.5 Tempos Computacionais

Conforme observado no [Capítulo 2](#), os métodos de perturbação numérica podem gerar derivadas pouco precisas, além de demandar, em geral, maior esforço computacional. Para evitar o uso de perturbações, propôs-se neste trabalho, o uso do Teorema da Função Implícita. Desta forma, nesta seção, os tempos computacionais necessários para obtenção das derivadas implícitas e numéricas foram comparados.

A [Tabela 4.4](#) mostra os tempos computacionais médios para a obtenção das de-

¹Por exemplo, as áreas superficiais variam da ordem 1 a 300 \AA^2 enquanto que as cargas variam no intervalo de $-0,025$ a $0,025 \text{ e/\AA}^2$.

²Foram testados diversos valores de perturbação e percebeu-se que, abaixo de 10^{-6} vezes os parâmetros, os resultados das derivadas numéricas permanecem inalterados.

derivadas implícitas e numéricas com relação às áreas dos subgrupos que formam 4 misturas de substâncias comuns na indústria química, em diferentes temperaturas e composições. Os números entre parênteses na primeira coluna representam os número de subgrupos do modelo F-SAC para cada substância e x_1 é a fração molar do primeiro componente.

Tabela 4.4: Tempos computacionais para calcular as derivadas implícitas e numéricas para os sistemas testados.

Mistura	x_1	T (K)	Tempo de derivação (ms)	
			Implícita	Numérica
Dimetil éter (1) + 1-buteno (3)	0,2	283,15	0,222730	0,75109
Clorofórmio (1) + Acetona (1)	0,5	303,15	0,106288	0,255738
Etanol (2) + N-octano (2)	0,8	343,15	0,387796	1,657136
Água (1) + N-hexano (2)	1,0	425,15	0,321874	1,153468

Os cálculos foram realizados em um computador pessoal com processador Intel[®] Core[™] 2 Duo E6300, com 8GB de RAM, executando o Sistema Operacional Ubuntu 18.04 LTS. Como alguns cálculos levam apenas frações de milissegundos para serem realizados, a metodologia para medição de tais tempos foi a seguinte: os cálculos das derivadas foram repetidos um determinado número de vezes e o tempo necessário foi medido; o tempo obtido foi dividido pelo número de repetições. Neste trabalho, optou-se por realizar cada cálculo 500.000 vezes, pois observou-se que a partir deste número, os resultados de tempos computacionais médios não variavam mais.

Conforme pode-se observar da [Tabela 4.4](#), as derivadas implícitas foram mais rápidas para todos os sistemas testados. Além disto, é possível inferir que, conforme o número de parâmetros aumenta, as diferenças entre os tempos implícitos e numéricos tendem a aumentar. Por exemplo: para o sistema clorofórmio + acetona (2 subgrupos apenas), a relação foi $0,106288/0,255738 = 0,4156$; para o sistema etanol + n-octano (4 subgrupos), esta relação foi de $0,387796/1,657136 = 0,2344$, evidenciando que as diferenças entre os tempos computacionais necessários (implícitos e numéricos) tendem a ser maiores. Portanto, com base nos resultados obtidos nesta seção do trabalho, entende-se que a metodologia proposta demanda menor esforço computacional, além de ser mais precisa, pois não utiliza nenhuma aproximação.

4.2 Resultados das Otimizações

4.2.1 Introdução

Nesta seção serão mostrados os resultados das estimações de parâmetros realizadas neste trabalho. Escolheu-se a versão original do F-SAC de Soares e Gerber (2013), Soares et al. (2013) por ser a versão mais completa do modelo. Para testar o algoritmo utilizado, elaborou-se um pacote contendo 5 testes, os quais serão descritos na Subseção 4.2.2. As respectivas tabelas de parâmetros antes e depois das estimações são mostradas nos Apêndices A-B. Nestas tabelas, são mostrados, também, estimativas dos intervalos de confiança de cada parâmetro estimado neste trabalho.

4.2.2 Descrição dos Testes

Denominou-se Teste 0 a estimação de parâmetros partindo da parametrização original do F-SAC (SOARES; GERBER, 2013; SOARES et al., 2013). Conforme observado anteriormente, apesar dos resultados publicados apresentarem apenas pequenos desvios com relação aos dados experimentais utilizados, tais resultados não representam um mínimo local da função objetivo utilizada. Além do mais, é interessante observar, a partir dos parâmetros originais do modelo, se ainda é possível um maior refinamento dos parâmetros, bem como inferir o comportamento do algoritmo utilizado.

Tendo em vista que o MMA aparentemente funcionou de forma correta para o Teste 0 (conforme será visto), elaborou-se os testes 1 ao 4 para verificar mais profundamente a capacidade deste método em resolver o problema proposto. No Teste 1, todos os parâmetros originais (utilizados no Teste 0) foram multiplicados por 1,1; no Teste 2, os parâmetros originais do Teste 0 foram multiplicados por 0,9.

No Teste 3 os parâmetros dos grupos e de ligação de hidrogênio foram mantidos iguais aos do Teste 0 e utilizou-se os parâmetros dos subgrupos oriundos diretamente dos cálculos COSMO das moléculas envolvidas. A obtenção de tais parâmetros é descrita na seção de resultados do trabalho de Soares e Gerber (2013) e foi realizada

utilizando-se programa JCOSMO (GERBER; SOARES, 2010). No teste 4, manteve-se apenas os parâmetros originais (Teste 0) de ligação de hidrogênio. Os parâmetros dos subgrupos foram os mesmos do Teste 3 e os parâmetros dos grupos foram obtidos dos cálculos COSMO das moléculas envolvidas utilizando-se o JCOSMO.

4.2.3 Resultados dos Testes

As Figuras 4.1-4.5 mostram os resultados originais e os resultados obtidos após as parametrizações realizadas para os Testes 0 ao 4. As linhas tracejadas acima e abaixo das diagonais representam, respectivamente, os desvios positivos e negativos de 1 unidade de logaritmo natural. A Tabela 4.5 mostra os resultados das FO (calculados pela Equação 3.2) e dos coeficientes de correlação (R^2) obtidos para todos os testes. O Apêndice B mostra os parâmetros estimados em cada teste e os respectivos intervalos de confiança dos parâmetros. As matrizes de covariância entre os parâmetros não foram mostradas pois são matrizes com 142 linhas e 142 colunas, totalizando 20164 elementos em cada um dos 5 testes, de modo que seriam, caso mostrados, resultados de difícil visualização.

Todas as estimações de parâmetros foram conduzidas em um computador pessoal com processador Intel[®] Core[™] 2 Duo E6300, com 8GB de RAM, executando o Sistema Operacional Ubuntu 18.04 LTS. Os tempos computacionais necessários não excederam 10 min em cada teste.

Conforme pode-se observar da Figura 4.1, ambos os resultados mostraram, em geral, uma boa concordância com os dados experimentais. Além disto, percebe-se visualmente que os pontos da Figura 4.1-b parecem estar mais próximos da diagonal, indicando que houve um refinamento do resultado original. Isto ocorre porque o resultado original, apesar de não ser um mínimo local da função objetivo, já estava muito próximo de um mínimo.

Observa-se visualmente das Figuras 4.2-4.5 que os pontos ficaram mais próximos das diagonais após as estimações de parâmetros. Semelhantemente, os valores da FO diminuíram e de R^2 se aproximaram de 1 para todos os testes. Tais resultados demonstram que o algoritmo selecionado pode ser empregado na resolução do

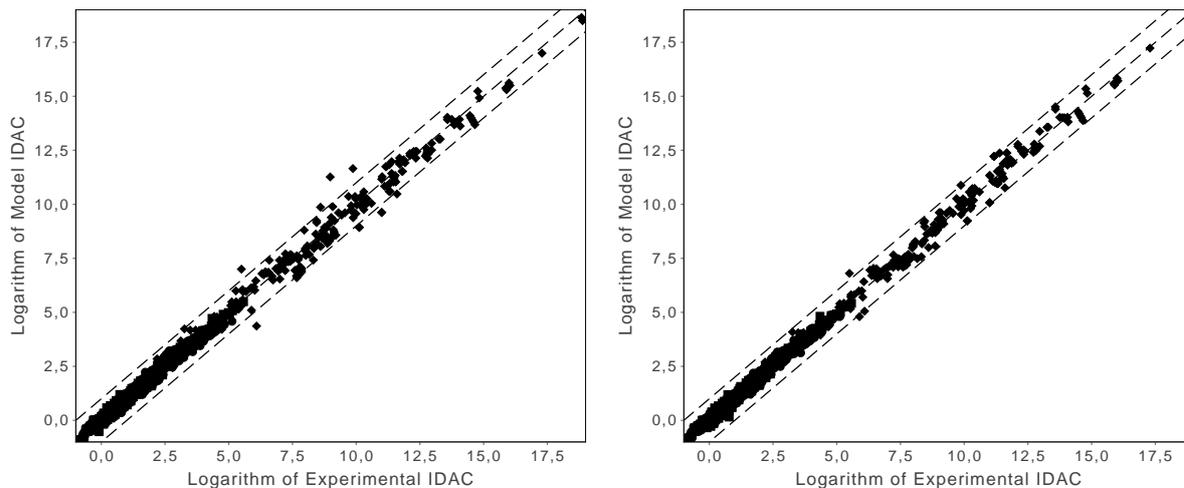
problema de otimização proposto, partindo-se de pontos diferentes. Entretanto, as soluções alcançadas não foram iguais em nenhum dos testes realizados, conforme pode ser demonstrado pelos valores diferentes de FO e dos parâmetros obtidos.

Por exemplo, foi possível chegar em valores próximos de FO (entretanto diferentes parâmetros) nos testes 0 ao 3. Por outro lado, a diagonal da [Figura 4.5](#) e o valor da FO depois da estimação de parâmetros mostram que o algoritmo para o teste 4 não chegou a resultados tão bons quanto os dos outros testes. Isto demonstra que o algoritmo utilizado também funcionou neste caso, contudo, chegou-se num outro mínimo local, de valor mais alto.

Observa-se que os intervalos de confiança, em geral, foram relativamente largos para os parâmetros estimados. Isto indica uma alta correlação entre os parâmetros estimados em todos os testes. Destaca-se os intervalos obtidos para o parâmetros eletrostáticos do grupo *GBLactone*, que são absurdos. Isto ocorre devido ao fato de que, no banco de dados utilizado, há somente um dado experimental com uma substância que utiliza este grupo.

O tipo de algoritmo estudado neste trabalho (local), em nenhum momento garante que vai se chegar num mínimo global. O algoritmo garante que vai se chegar num mínimo local, de forma eficiente. O problema de otimização utilizado é altamente não linear, com 142 parâmetros a serem estimados e 2979 dados experimentais. A partir dos testes executados, foi demonstrado que a FO apresenta diversos mínimos locais. Além disto, fica evidente que o resultado da estimação de parâmetros depende da condição inicial.

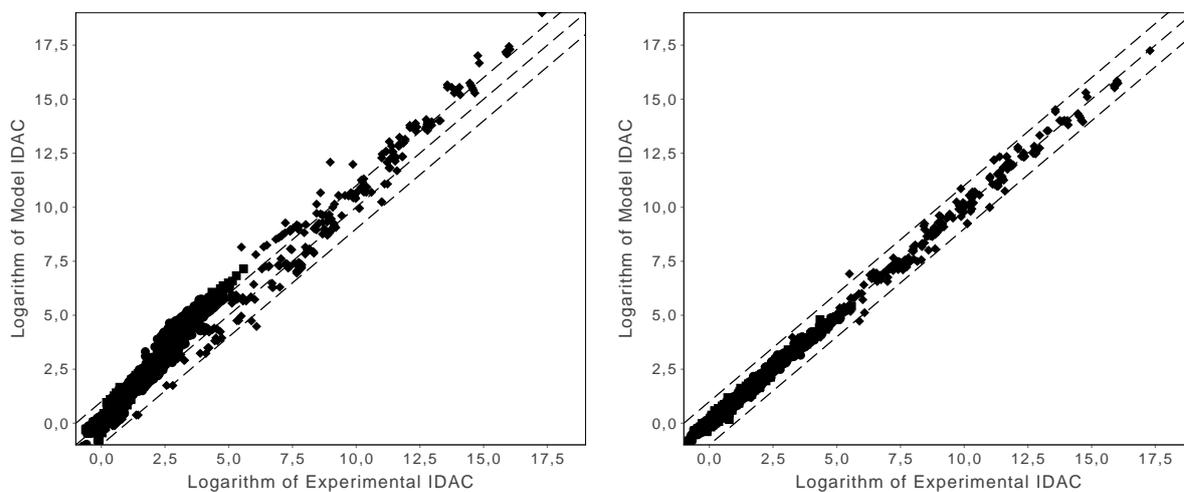
Desta forma, a metodologia proposta poderá ser utilizada na exploração dos mínimos da FO (partindo-se de diversos pontos iniciais), para se chegar-se na parametrização mais adequada ou que leve a um menor mínimo dentre os obtidos. Sugere-se, a partir dos resultados observados neste trabalho, a utilização de um algoritmo “híbrido”. Neste algoritmo, um método global poderia ser utilizado para se chegar num ponto inicial para o uso de um método local (por exemplo, o MMA) para refinar o resultado.



(a) F-SAC com parâmetros originais de Soares e Gerber (2013), Soares et al. (2013).

(b) Parametrização do Teste 0.

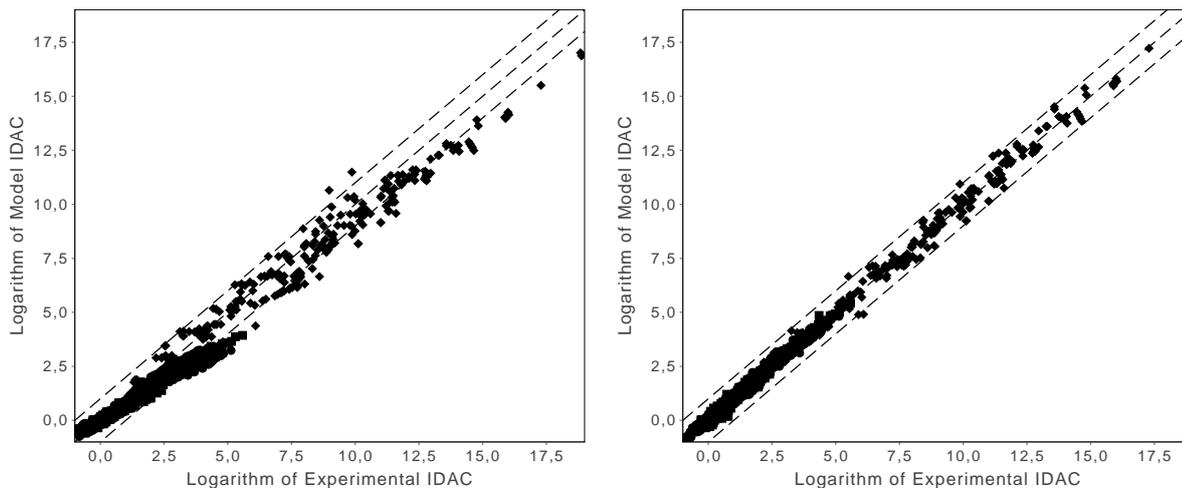
Figura 4.1: Logaritmo natural do IDAC experimental *versus* o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.



(a) F-SAC com os parâmetros originais multiplicados por 1,1.

(b) Parametrização do Teste 1.

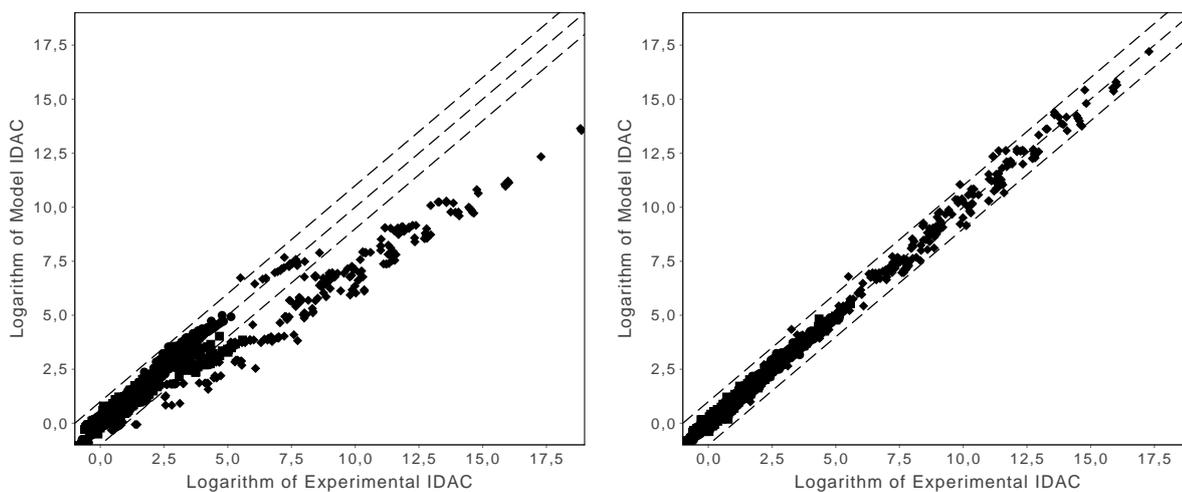
Figura 4.2: Logaritmo natural do IDAC experimental *versus* o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.



(a) F-SAC com os parâmetros originais multiplicados por 0,9.

(b) Parametrização do Teste 2.

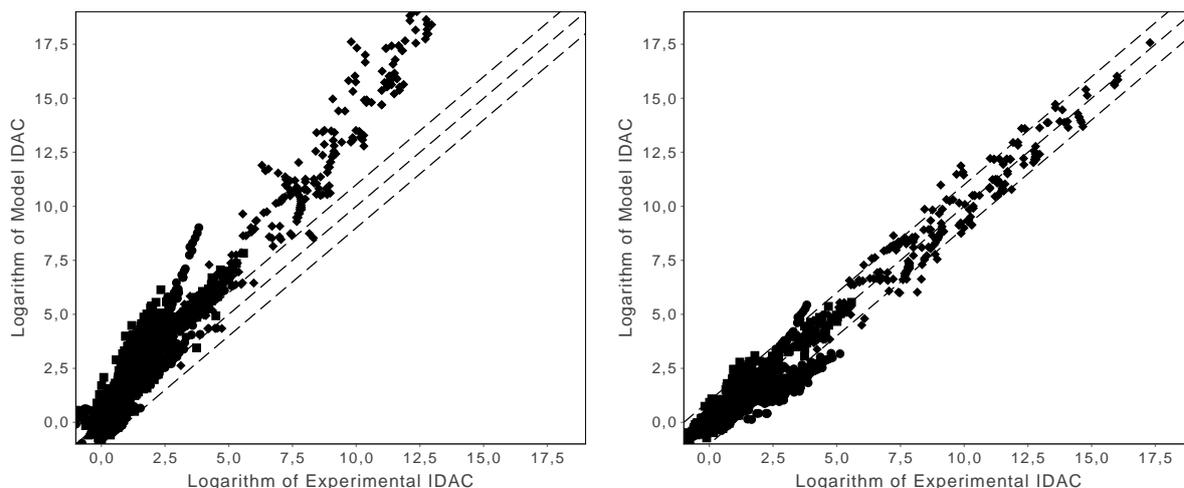
Figura 4.3: Logaritmo natural do IDAC experimental *versus* o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.



(a) F-SAC com os parâmetros dos subgrupos calculados pelo COSMO e os demais parâmetros originais.

(b) Parametrização do Teste 3.

Figura 4.4: Logaritmo natural do IDAC experimental *versus* o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.



(a) F-SAC com os parâmetros dos grupos e subgrupos calculados pelo COSMO e os demais parâmetros originais.

(b) Parametrização do Teste 4.

Figura 4.5: Logaritmo natural do IDAC experimental *versus* o logaritmo natural do IDAC calculado para todas as misturas utilizadas na parametrização do modelo.

Tabela 4.5: Valores da FO e do R^2 para os Testes 0 ao 4 antes e após as estimações de parâmetros.

Teste	Antes da estimaco		Aps a estimaco	
	FO	R^2	FO	R^2
0	0,03945	0,99408	0,02389	0,99651
1	0,37394	0,95528	0,02314	0,99662
2	0,22738	0,96039	0,02564	0,99625
3	0,73796	0,81460	0,030779	0,99551
4	2,45162	0,82662	0,32422	0,95157

4.3 Efeito do Raio de Contato Padro no Modelo F-SAC

Nas estimaces de parmetros realizadas na [Subseo 4.2.3](#), o valor do parmetro universal raio de contato padro (r_{eff}) foi fixado em 1,07 Å. De maneira semelhante, os demais parmetros universais do modelo (veja a descrio do F-SAC na [Subseo 2.2.2](#)) tambm foram fixados. Desta forma, com a disponibilidade da ferramenta desenvolvida no presente trabalho,  interessante estudar a influncia dos

valores destes parâmetros no modelo F-SAC.

Contudo, um estudo da influência dos parâmetros universais no F-SAC não está englobado no escopo do presente trabalho. Assim, nesta seção será mostrada apenas um estudo simplificado de como o r_{eff} influencia nos resultados das estimações de parâmetros. Objetiva-se, com isto, demonstrar a utilidade dos algoritmos desenvolvidos em estudos a serem realizados posteriormente no grupo de pesquisa onde este trabalho está inserido.

Os testes realizados foram muito simples: o Teste 0 foi repetido com diferentes valores de r_{eff} . A Tabela 4.6 mostra os valores da função objetivo (Equação 3.2) e do R^2 , antes e depois da estimação de parâmetros, para alguns valores de r_{eff} . Na Figura 4.6 são mostrados os valores da função objetivo depois da estimação de parâmetros para diferentes valores de r_{eff} . Novamente, todas as estimações de parâmetros foram conduzidas em um computador pessoal com processador Intel[®] Core[™] 2 Duo E6300, com 8GB de RAM, executando o Sistema Operacional Ubuntu 18.04 LTS e os tempos computacionais necessários não excederam 10 min em cada teste. Os valores dos parâmetros estimados não são mostrados, pois não são relevantes nesta seção.

Conforme pode-se observar, o menor valor da função objetivo foi obtido para $r_{eff} = 1,10 \text{ \AA}$, que não é o valor normalmente adotado nos trabalhos do F-SAC. Fica clara a possibilidade de se refinar um pouco mais os resultados do F-SAC. Inclusive, o intervalo de valores de r_{eff} poderia ser expandido para se visualizar o comportamento da curva mostrada na Figura 4.6 num intervalo maior.

Tabela 4.6: Valores da FO e do R^2 para o Teste 0 antes e após as estimações de parâmetros com firentes valores de r_{eff} .

r_{eff}	Antes da estimação		Após a estimação	
	FO	R^2	FO	R^2
1,00	0,10197	0,98430	0,02505	0,99633
1,05	0,03068	0,9906	0,02414	0,99647
1,07 (original)	0,03945	0,99408	0,02389	0,99651
1,10	0,07001	0,98976	0,02347	0,99656
1,15	0,09389	0,98610	0,02435	0,99644

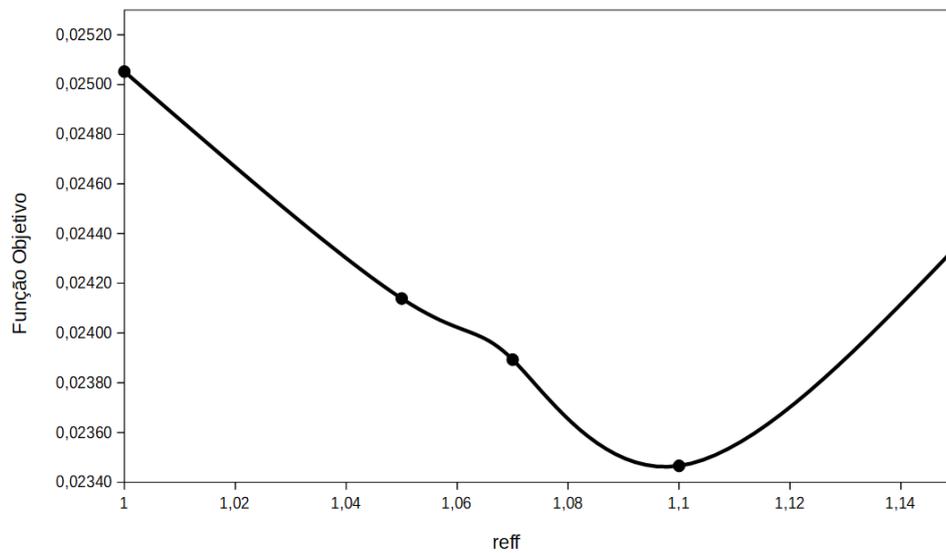


Figura 4.6: Valores da função objetivo depois da estimação de parâmetros *versus* r_{eff} .

Conforme relatado, todos os tempos computacionais foram menores do que 10 min para cada teste em um computador pessoal comum, de modo que fica demonstrada a utilidade deste trabalho. Além disto, fica como sugestão para trabalhos futuros um estudo mais aprofundado da influência de cada um dos parâmetros universais no modelo F-SAC.

Capítulo 5

Conclusões e Trabalhos Futuros

5.1 Conclusões

Neste trabalho estudou-se o uso de um algoritmo de otimização do tipo local e baseado em derivadas para a estimação dos parâmetros do modelo F-SAC. Foi proposta a aplicação do Teorema da Função Implícita para a obtenção das derivadas implícitas do modelo com relação a qualquer variável independente. O uso deste teorema foi sugerido para contornar o método iterativo inerente ao F-SAC, que dificulta a sua derivação de forma analítica.

Foram apresentados os equacionamentos para o cálculo das derivadas do modelo com relação à temperatura, ao número de mols de algum componente na mistura e aos seus parâmetros. As derivadas implícitas com respeito à temperatura e ao número de mols foram realizadas neste trabalho para testar a metodologia proposta.

As derivadas implícitas e numéricas com respeito à temperatura foram utilizadas para calcular as entalpias de excesso do sistema etanol + água, nas temperaturas de 323,15 K, 343,15 K, 363,15 K e 383,15 K. Os resultados implícitos e numéricos apresentaram desvios médios na ordem de 10^{-5} , com maior desvio de $1,87 \times 10^{-5}$ e maior desvio individual de $3,61 \times 10^{-5}$.

As derivadas implícitas e numéricas com respeito ao número de mols foram utilizadas num teste de consistência de Gibbs-Duhem, para o sistema acetato de metila + água, a 330,5 K. Os resultados obtidos neste trabalho foram comparados com os resultados de [Lin e Sandler \(2002\)](#). Conforme observou-se, os valores de Gibbs-Duhem

calculados por Lin e Sandler (2002) ficaram na ordem de $10^{-2} \sim 10^{-3}$. Os valores numéricos de Gibbs-Duhem, calculados neste trabalho, ficaram na ordem de 10^{-2} , ou menores, enquanto os valores implícitos foram menores do que 10^{-5} .

A partir dos resultados anteriores, observou-se que a metodologia proposta funcionou de maneira correta para a temperatura e número de mols. Assim, utilizou-se o Teorema da Função Implícita para derivar o modelo F-SAC com relação aos seus parâmetros. Os valores implícitos e numéricos das derivadas do F-SAC com relação aos seus parâmetros foram comparados para 16 misturas binárias de substâncias comuns encontradas na indústria, com diferentes composições, em um intervalo de 273,15 K a 423,15 K. Observou-se um desvio médio de $1,083 \times 10^{-5}$.

Os tempos computacionais para a obtenção das derivadas exatas e numéricas, com relação às áreas dos subgrupos de 4 misturas de substâncias comuns na indústria química, foram comparados. As derivadas implícitas foram mais rápidas para todos os sistemas. Além disto, observou-se que, conforme o número de parâmetros aumenta, as diferenças entre os tempos exatos e numéricos tendem a aumentar. Isto demonstra que a metodologia proposta demanda menor esforço computacional, além de prover valores de derivadas mais confiáveis.

Com base nos resultados de derivadas obtidos, concluiu-se que o Teorema da Função Implícita poderia ser utilizado na obtenção das derivadas exatas do modelo F-SAC com relação a qualquer variável. Desta forma, as derivadas do modelo F-SAC com respeito aos seus parâmetros foram utilizadas no método de otimização denominado MMA, descrito em (https://nlopt.readthedocs.io/en/latest/NLopt_Algorithms/) e disponível dentro do pacote computacional NLopt.

Para testar o MMA, foram propostos 5 testes partindo-se de valores diferentes de parâmetros. Em todos os testes realizados, foi possível chegar-se num mínimo da função objetivo, entretanto, as soluções alcançadas não foram iguais em nenhum dos testes realizados. Isto ocorre pois o problema de otimização proposto apresenta diversos mínimos locais. Contudo, mesmo que o ponto inicial seja ruim (muito distante de um mínimo), o método garantidamente chega num mínimo local. Além disto, foi realizado um estudo simplificado da influência de um dos parâmetros universais do F-SAC, o raio de contato padrão (r_{eff}), no comportamento do modelo. Observou-se que

o menor valor da função objetivo foi obtido para $r_{eff} = 1,10 \text{ \AA}$, que não é o valor normalmente adotado nos trabalhos do F-SAC. Assim, ficou demonstrada a possibilidade de se refinar um pouco mais os resultados obtidos neste trabalho.

Por fim, os resultados das otimizações demonstraram que o algoritmo MMA pode ser empregado na estimação de parâmetros do modelo F-SAC. Desta forma, a metodologia proposta poderá ser utilizada na exploração dos mínimos da FO (partindo-se de diversos pontos iniciais), para se chegar-se na parametrização mais adequada ou que leve a um menor mínimo dentre os obtidos. Com esta metodologia, esta tarefa é facilitada, pois a estimação de cada conjunto de pontos pode ser obtida de maneira rápida e eficiente. Além disto, na introdução de novos parâmetros no modelo e no uso de mais dados experimentais ou outros tipos de experimentos (LLE, por exemplo), a reestimação de parâmetros também poderá ser facilitada.

5.2 Trabalhos Futuros

A partir das conclusões obtidas, sugere-se os seguintes trabalhos futuros:

- Expandir o trabalho para o uso de outros tipos de dados experimentais, como VLE, LLE, SLE, entalpia de excesso, etc. Neste caso, basta determinar o formato da função objetivo que será utilizada (IDAC + VLE + LLE) e derivá-la. O conceito de peso poderá ser utilizado para que os diferentes tipos de dados experimentais sejam considerados de maneira igualitária na função objetivo.
- Utilizar um algoritmo “híbrido”, no qual um método global poderia ser utilizado para se chegar num ponto inicial para o uso de um método local.
- Realizar um estudo detalhado da influência dos parâmetros universais do F-SAC. Inclusive, o modelo poderia ser diferenciado com relação a estes parâmetros e eles poderiam ser incluídos na estimação de parâmetros junto com os demais.

5.3 Principais Publicações

A presente tese de Doutorado resultou nas seguintes publicações:

- Artigo publicado na revista *Fluid Phase Equilibria*:
 - POSSANI, L. F. K.; STAUDT, P.B.; SOARES, R. de P. Prediction of water solubilities in hydrocarbons and oils using F-SAC coupled with SRK-EoS, v. 427, p. 394–405, 2016.
- Artigo publicado na revista *Brazilian Journal of Chemical Engineering*:
 - POSSANI, L. F. K.; SOARES, R. de P. Numerical and computational aspects of COSMO-based activity coefficient models, v. 36, n. 1, p. 587–598, 2019.

Referências Bibliográficas

- ABRAMS, D. S.; PRAUSNITZ, J. M. Statistical thermodynamics of liquid mixtures: A new expression for the excess gibbs energy of partly or completely miscible systems. *AIChE Journal*, v. 21, n. 1, p. 116–128, 1975.
- Andreas Griewank, A. W. *Evaluating Derivatives Principles and Techniques of Algorithmic Differentiation*. Second. [S.l.]: SIAM - Society for Industrial and Applied Mathematics, 2008. ISBN 978-0-898716-59-7.
- BIEGLER, L. T. *Nonlinear Programming: concepts, algorithms and applications to chemical processes*. First edition. [S.l.]: Society for Industrial and Applied Mathematics and the Mathematical Optimization, 2010. ISBN 978-0-898717-02-0.
- COSTA, C. T.; TAVARES, F. W.; SECCHI, A. R. Equation of state based on the hole-lattice theory and surface-charge density (COSMO): Part A - Pure compounds. *Fluid Phase Equilibria*, v. 409, p. 472–481, 2016.
- COSTA, C. T.; TAVARES, F. W.; SECCHI, A. R. Equation of state based on the hole-lattice theory and surface-charge density (COSMO): Part B - Vapor-liquid equilibrium for mixtures. *Fluid Phase Equilibria*, v. 419, p. 1–10, 2016.
- DDBST. *Dortmund Data Bank Software and Separation Technology*. 2014. Disponível em: <http://www.ddbst.com/>.
- DRAPER, N. R.; HIMMELBLAU, D. M. *Process Analysis by Statistical Methods*. [S.l.]: John Wiley & Sons Inc., 2006. 924 p. ISBN 9780471399858.
- FLÔRES, G. B.; STAUDT, P. B.; SOARES, R. D. P. Including dispersive interactions in the F-SAC model. *Fluid Phase Equilibria*, v. 426, p. 10–11, 2016.
- FREDENSLUND, A. A.; JONES, R. L.; PRAUSNITZ, J. M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, v. 21, n. 6, p. 1086–1099, 1975.
- GERBER, R. P. *NOVO MODELO DE COEFICIENTE DE ATIVIDADE: F-SAC*. Dissertação (Mestrado) — UFRGS, 2012.
- GERBER, R. P.; SOARES, R. de P. Prediction of infinite-dilution activity coefficients using UNIFAC and COSMO-SAC variants. *Ind. Eng. Chem. Res.*, v. 49, n. 16, p. 7488–7496, 2010.
- GERBER, R. P.; SOARES, R. de P. Assessing the reliability of predictive activity coefficient models for molecules consisting of several functional groups. *Braz. J. Chem. Eng.*, v. 30, n. 1, p. 1–11, 2013.

- GMEHLING, J. Present status and potential of group contribution methods for process development. *The Journal of Chemical Thermodynamics*, v. 41, n. 6, p. 731–747, 2009.
- GMEHLING, J.; LI, J.; SCHILLER, M. A modified UNIFAC model. 2. present parameter matrix and results for different thermodynamic properties. *Ind. Eng. Chem. Res.*, v. 32, n. 1, p. 178–193, 1993.
- GMEHLING, J. et al. A modified unifac (dortmund) model. 3. revision and extension. *Ind. Eng. Chem. Res.*, v. 37, n. 12, p. 4876–4882, 1998.
- GMEHLING, J.; RASMUSSEN, P.; FREDENSLUND, A. Vapor-liquid equilibria by UNIFAC group contribution. Revision and extension. 2. *Ind. Eng. Chem. Process Des. Dev.*, v. 21, n. 1, p. 118–127, 1982.
- GMEHLING, J. et al. A modified unifac (dortmund) model. 4. revision and extension. *Ind. Eng. Chem. Res.*, v. 41, n. 6, p. 1678–1688, 2002.
- GRENSEMANN, H.; GMEHLING, J. Performance of a Conductor-Like Screening Model for Real Solvents Model in Comparison to Classical Group Contribution Methods. *Industrial & Engineering Chemistry Research*, v. 44, n. 5, p. 1610–1624, mar 2005. ISSN 0888-5885.
- GRENSEMANN, H.; GMEHLING, J. Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. *Ind. Eng. Chem. Res.*, v. 44, n. 5, p. 1610–1624, 2005.
- HAN, L.; NEUMANN, M. Effect of dimensionality on the Nelder–Mead simplex method. *Optim. Methods. Softw.*, v. 21, n. 1, p. 1–16, 2006.
- HANSEN, H. K. et al. Vapor-liquid equilibria by UNIFAC group contribution. 5. Revision and extension. *Ind. Eng. Chem. Res.*, v. 30, n. 10, p. 2352–2355, 1991.
- HIMMELBLAU, D. M.; LASDON, L. S.; EDGAR, T. F. *Optimization of Chemical Processes*. Second edition. [S.l.]: McGraw-Hill Chem. Eng. series, 2001. ISBN 0-07-039359-1.
- HSIEH, C.-M.; LIN, S.-T.; VRABEC, J. Considering the dispersive interactions in the COSMO-SAC model for more accurate predictions of fluid phase behavior. *Fluid Phase Equilibria*, Elsevier B.V., v. 367, p. 109–116, 2014. ISSN 03783812.
- JAKOB, A. et al. Further development of modified UNIFAC (dortmund): Revision and extension 5. *Ind. Eng. Chem. Res.*, v. 45, n. 23, p. 7924–7933, 2006.
- KLAMT, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.*, v. 99, n. 7, p. 2224–2235, 1995.
- KLAMT, A.; ECKERT, F. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib.*, v. 172, n. 1, p. 43–72, 2000.
- KLAMT, A.; ECKERT, F.; ARLT, W. COSMO-RS: An alternative to simulation for calculating thermodynamic properties of liquid mixtures. *Annu. Rev. Chem. Biomol. Eng.*, v. 1, p. 101–122, 2010.
- KLAMT, A. et al. Refinement and parametrization of COSMO-RS. *The Journal of Physical Chemistry*, v. 102, n. 26, p. 5074–5085, 1998.

- KLAMT, A.; SCHUURMANN, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2*, n. 5, p. 799–805, 1993.
- KORETSKY, M. *Engineering and chemical thermodynamics*. 2. ed. [S.l.]: Wiley, 2012. ISBN 0470259612.
- KRANTZ, S.; PARKS, H. *The implicit function theorem: history, theory, and applications*. [S.l.]: Modern Birkhauser Classics, 2002. ISBN 2012952321.
- LIN, S. T.; SANDLER, S. I. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind. Eng. Chem. Res.*, v. 41, n. 5, p. 899–913, 2002.
- MACEDO, E. A. et al. Vapor-liquid equilibria by UNIFAC group contribution. Revision and extension. 3. *Ind. Eng. Chem. Process Des. Dev.*, v. 22, n. 4, p. 676–678, 1983.
- MALANOWSKI, S. et al. Simultaneous representation of binary VLE, LLE and h^E data using the UNIQUAC model. *Chem. Eng. Sci.*, v. 36, n. 10, p. 1727–1730, 1981.
- NELDER, J. A.; MEAD, R. A Simplex Method for Function Minimization. *The Computer Journal*, v. 7, n. 4, p. 308–313, 1965.
- NOCEDAL, S. J. W. *Numerical Optimization*. Second. [S.l.]: Springer, 2006. ISBN 9780387303031.
- POSSANI, L. et al. Simultaneous correlation of infinite dilution activity coefficient, vapor-liquid, and liquid-liquid equilibrium data with F-SAC. *Fluid Phase Equilibria*, v. 364, p. 31–41, feb 2014. ISSN 03783812.
- POSSANI, L. F. K. *Correlação Simultânea de IDAC, VLE e LLE com o Modelo F-SAC*. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul, 2014.
- POSSANI, L. F. K. et al. Mutual Solubilities of Hydrocarbon–Water Systems with F-SAC. *Fluid Phase Equilibria*, v. 384, p. 122–133, dec 2014.
- POSSANI, L. F. K.; SOARES, R. d. P. NUMERICAL AND COMPUTATIONAL ASPECTS OF COSMO-BASED ACTIVITY COEFFICIENT MODELS. *Brazilian Journal of Chemical Engineering*, v. 36, n. 1, p. 587–598, 2019.
- POSSANI, L. F. K.; STAUDT, P. B.; SOARES, R. d. P. Prediction of water solubilities in hydrocarbons and oils using F-SAC coupled with SRK-EoS. *Fluid Phase Equilibria*, Elsevier Ltd, v. 427, p. 394–405, 2016.
- PRESS, W. H. et al. *Numerical Recipes in C: The Art of Scientific Computing*. [S.l.]: Press Syndicate of the University of Cambridge, 1992. ISBN 0521431085.
- RENON, H.; PRAUSNITZ, J. M. Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE Journal*, v. 14, n. 1, p. 135–144, jan 1968. ISSN 0001-1541.
- SANDLER, S. I. *Models for Thermodynamic and Phase Equilibria Calculations*. First. [S.l.]: CRC Press, 1993. 704 p. (Chemical Industries). ISBN 9780824791308.
- SKJOLD-JORGENSEN, S. et al. Vapor-Liquid Equilibria by UNIFAC Group Contribution. Revision and Extension. *Ind. Eng. Chem. Process Des. Dev.*, v. 18, n. 4, p. 714–722, 1979.

- SMITH, J. M.; NESS, H. C. V.; ABBOTT, M. M. *Introdução à Termodinâmica da Engenharia Química*. Sétima edição. [S.l.]: LTC, 2007.
- SOARES, R. de P. The Combinatorial Term for COSMO-Based Activity Coefficient Models. *Ind. Eng. Chem. Res.*, v. 50, n. 5, p. 3060–3063, 2011.
- SOARES, R. de P.; GERBER, R. P. Functional-Segment Activity Coefficient Model. 1. Model Formulation. *Ind. Eng. Chem. Res.*, v. 52, n. 32, p. 11159–11171, 2013.
- SOARES, R. de P. et al. Functional-Segment Activity Coefficient Model. 2. Associating Mixtures. *Ind. Eng. Chem. Res.*, v. 52, n. 32, p. 11172–11181, 2013.
- STAUDT, P. B.; SOARES, R. de P. A self-consistent Gibbs excess mixing rule for cubic equations of state. *Fluid Phase Equilib.*, v. 334, p. 76–88, 2012.
- SVANBERG, K. A Class of Globally Convergent Optimization Methods Based on Conservative Convex Separable Approximations. *SIAM Journal on Optimization*, v. 12, n. 2, p. 555–573, 2002.
- TIEGS, D. et al. Vapor-liquid equilibria by UNIFAC group contribution. 4. Revision and extension. *Ind. Eng. Chem. Res.*, v. 26, n. 1, p. 159–161, 1987.
- Virginia Tech. COSMO-SAC-VT-2005. 2015. Disponível em: <http://www.design.che.vt.edu/VT-Databases.html>.
- WÄCHTER, A.; BIEGLER, L. T. On the Implementation of Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming. *Mathematical Programming*, v. 106, n. 1, p. 25–57, 2006.
- WEIDLICH, U.; GMEHLING, J. A modified UNIFAC model. 1. prediction of VLE, h^e , and γ^∞ . *Ind. Eng. Chem. Res.*, v. 26, n. 7, p. 1372–1381, 1987.
- WILSON, G. M. Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing. *Journal of the American Chemical Society*, v. 86, n. 2, p. 127–130, jan 1964. ISSN 0002-7863.
- WITTIG, R.; LOHMANN, J.; GMEHLING, J. Vapor-liquid equilibria by unifac group contribution. 6. revision and extension. *Ind. Eng. Chem. Res.*, v. 42, n. 1, p. 183–188, 2003.
- XUE, Z.; MU, T.; GMEHLING, J. Comparison of the a Priori COSMO-RS Models and Group Contribution Methods: Original UNIFAC, Modified UNIFAC(Do), and Modified UNIFAC(Do) Consortium. *Ind. Eng. Chem. Res.*, v. 51, p. 11809–11817, 2012.

Apêndice A

Tabelas utilizadas nas Estimações de Parâmetros

Tabela A.1: Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 0.

Grupo	$Q_k^+/\text{Å}^2$	$Q_k^-/\text{Å}^2$	$\sigma_k^+/\text{enm}^{-2}$	Subgrupo	$Q_k^0/\text{Å}^2$
CH2	0	0	0	CH3	46,69
				CH2	24,64
				CH	-0,53
				C	-20,88
				c-CH2	27,00
				c-CH	5,47
				C=C	6,16
ACH	5,38	6,71	0,0056	CH=CH	33,33
				CH2=C	33,38
				CH=C	11,59
				c-CH=CH	51,45
CH3COCH3	21,97	80,23	0,0133	ACH	20,11
				AC	1,76
CH2CHO	14,85	81,88	0,0172	CH3COCH3	122,37
				CH3COCH2	90,18
				CH3COCH	68,92
				CH2COCH2	66,27
				CH2COCH	27,33
				CH3CHO	90,53
CH3COAC3	9,56	16,68	0,0162	CH2CHO	67,25
				CH3COAC3	105,66
c-CH2COCH2	13,93	42,78	0,0166	c-CH2COCH2	77,88
Triacetin	34,68	62,83	0,0146	Triacetin	296,41
GBLactone	18,05	41,30	0,0161	GBLactone	120,60
CH3COOCH3	28,94	107,69	0,0110	CH3COOCH3	143,72
				CH3COOCH2	107,86
				CH3COOCH	95,89
				CH2COOCH2	96,93
				CH2COOCH	96,93
Phthalate	33,60	69,26	0,0132	AC6H4(CH2COO)2	207,91
CH2COOAC3	12,59	64,23	0,0158	CH2COOAC3	95,94
CH3OCH2	13,01	22,71	0,0091	CH3OCH3	103,58
				CH3OCH2	62,75
				CH3OC	31,42

Continua na página seguinte.

Tabela A.1 – Continuação da página anterior.

Grupo	$Q_k^+/\text{Å}^2$	$Q_k^-/\text{Å}^2$	$\sigma_k^+/\text{enm}^{-2}$	Subgrupo	$Q_k/\text{Å}^2$
				CH2OCH2	56,09
				CHOCH	10,12
c-CH2OCH2	8,12	27,20	0,0143	c-CH2OCH2	52,13
N(CH2)3	3,60	28,58	0,0175	N(CH2)3	35,64
DMSO	11,36	22,84	0,0188	DMSO	99,04
CH2Cl	9,30	9,07	0,0089	CH2Cl	58,08
CCl4	18,40	26,23	0,0062	CCl4	138,83
CH3CCl3	40,07	20,28	0,0039	CH3CCl3	146,92
ClAC3	37,41	12,13	0,0026	ClAC3	76,57
CHCl3	43,32	14,76	0,0035	CHCl3	114,25
CH3OH	8,31	4,72	0,0132	CH3OH	68,25
CH2OH	7,34	4,69	0,0145	CH2OH	43,84
				CHOH	23,02
H2O	8,84	12,16	0,0123	H2O	35,19

Tabela A.2: Parâmetros de interação binária para ligações de hidrogênio utilizados no Teste 0.

Doador	Receptor	$E_{m,n}^{\text{HB}}/\text{kJmol}^{-1}$
CH3OH	ACH	0,6185
CH3OH	CH3COCH3	0,5638
CH3OH	CH3OH	2,2111
CH3OH	CH2OH	2,6585
CH3OH	H2O	0,5942
CH2OH	ACH	0,2682
CH2OH	CH3COCH3	0,3128
CH2OH	CH3COOCH3	1,1003
CH2OH	CH3OCH2	1,8863
CH2OH	CH3OH	1,4941
CH2OH	CH2OH	1,9144
CH2OH	H2O	0,9985
CHCl3	ACH	0,4216
CHCl3	CH3COCH3	0,8704
CHCl3	CH3COOCH3	1,8350
CHCl3	CH3OCH2	2,9951
CHCl3	CH3OH	2,0616
CHCl3	CH2OH	1,9655
H2O	ACH	0,2146
H2O	CH3COCH3	4,6463
H2O	CH2CHO	4,0761
H2O	CH3COOCH3	5,2351
H2O	CH3OCH2	5,7129
H2O	CH3OH	5,7460
H2O	CH2OH	5,3976
H2O	H2O	5,2209

Tabela A.3: Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 1. Observação: apenas a área positiva do grupo CH₃OH foi multiplicada por 1,05 para não exceder o *bound* da carga negativa ($\sigma_k^- \geq -\sigma_k^{max}$).

Grupo	$Q_k^+/\text{\AA}^2$	$Q_k^-/\text{\AA}^2$	σ_k^+/enm^{-2}	Subgrupo	$Q_k^0/\text{\AA}^2$
CH ₂	0	0	0	CH ₃	51,36
				CH ₂	27,10
				CH	-0,58
				C	-22,97
				c-CH ₂	29,70
				c-CH	6,02
C=C	6,78	4,07	0,0055	CH ₂ =CH	65,47
				CH=CH	36,66
				CH ₂ =C	36,72
				CH=C	12,75
				c-CH=CH	56,60
				ACH	22,13
ACH	5,91	7,38	0,0062	AC	1,93
				ACH	22,13
CH ₃ COCH ₃	24,17	88,25	0,0146	CH ₃ COCH ₃	134,60
				CH ₃ COCH ₂	99,20
				CH ₃ COCH	75,82
				CH ₂ COCH ₂	72,89
				CH ₂ COCH	30,07
				CH ₃ CHO	99,58
CH ₂ CHO	16,33	90,07	0,0189	CH ₂ CHO	73,98
				CH ₃ CHO	99,58
CH ₃ COAC ₃	10,52	18,34	0,0179	CH ₃ COAC ₃	116,23
c-CH ₂ COCH ₂	15,32	47,06	0,0183	c-CH ₂ COCH ₂	85,67
Triacetin	38,15	69,11	0,0160	Triacetin	326,05
GBLactone	19,85	45,43	0,0177	GBLactone	132,66
CH ₃ COOCH ₃	31,83	118,46	0,0121	CH ₃ COOCH ₃	158,10
				CH ₃ COOCH ₂	118,65
				CH ₃ COOCH	105,48
				CH ₂ COOCH ₂	106,62
				AC ₆ H ₄ (CH ₂ COO) ₂	228,70
				CH ₂ COOAC ₃	105,53
Phthalate	36,96	76,18	0,0146	CH ₃ OCH ₃	113,94
				CH ₃ OCH ₂	69,02
				CH ₃ OC	34,57
				CH ₂ OCH ₂	61,70
				CHOCH	11,14
				c-CH ₂ OCH ₂	57,34
N(CH ₂) ₃	3,96	31,43	0,0192	N(CH ₂) ₃	39,20
DMSO	12,49	25,12	0,0207	DMSO	108,94
CH ₂ Cl	10,23	9,98	0,0098	CH ₂ Cl	63,89
CCl ₄	20,24	28,85	0,0068	CCl ₄	152,71
CH ₃ CCl ₃	44,08	22,31	0,0043	CH ₃ CCl ₃	161,61
ClAC ₃	41,15	13,34	0,0029	ClAC ₃	84,23
CHCl ₃	47,65	16,24	0,0038	CHCl ₃	125,68
CH ₃ OH	8,73	5,19	0,0145	CH ₃ OH	75,08
CH ₂ OH	8,07	5,16	0,0159	CH ₂ OH	48,22
				CHOH	25,32
H ₂ O	9,73	13,38	0,0135	H ₂ O	38,71

Tabela A.4: Parâmetros de interação binária para ligações de hidrogênio utilizados no Teste 1.

Doador	Receptor	$E_{m,n}^{HB} / \text{kJmol}^{-1}$
CH3OH	ACH	0,6804
CH3OH	CH3COCH3	0,6202
CH3OH	CH3OH	2,4322
CH3OH	CH2OH	2,9244
CH3OH	H2O	0,6536
CH2OH	ACH	0,2950
CH2OH	CH3COCH3	0,3440
CH2OH	CH3COOCH3	1,2103
CH2OH	CH3OCH2	2,0749
CH2OH	CH3OH	1,6435
CH2OH	CH2OH	2,1058
CH2OH	H2O	1,0983
CHCL3	ACH	0,4638
CHCL3	CH3COCH3	0,9575
CHCL3	CH3COOCH3	2,0185
CHCL3	CH3OCH2	3,2946
CHCL3	CH3OH	2,2678
CHCL3	CH2OH	2,1620
H2O	ACH	0,2360
H2O	CH3COCH3	5,1110
H2O	CH2CHO	4,4837
H2O	CH3COOCH3	5,7586
H2O	CH3OCH2	6,2842
H2O	CH3OH	6,3206
H2O	CH2OH	5,9373
H2O	H2O	5,7429

Tabela A.5: Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 2. Observação: as áreas positivas dos grupos CH2OH e N(CH2)3 não foram modificadas de modo que suas áreas sejam maiores do que a área de ligação de hidrogênio ($Q_k^+ \geq n_{acc} \cdot a_{eff}$).

Grupo	$Q_k^+ / \text{Å}^2$	$Q_k^- / \text{Å}^2$	$\sigma_k^+ / \text{enm}^{-2}$	Subgrupo	$Q_k^0 / \text{Å}^2$
CH2	0	0	0	CH3	42,02
				CH2	22,18
				CH	-0,48
				C	-18,79
				c-CH2	24,30
				c-CH	4,92
C=C	5,54	3,33	0,0045	CH2=CH	53,57
				CH=CH	30,00
				CH2=C	30,04
				CH=C	10,43
				c-CH=CH	46,31
ACH	4,84	6,04	0,0051	ACH	18,10
				AC	1,58
CH3COCH3	19,77	72,21	0,0119	CH3COCH3	110,13
				CH3COCH2	81,16
				CH3COCH	62,03

Continua na página seguinte.

Tabela A.5 – Continuação da página anterior.

Grupo	$Q_k^+/\text{Å}^2$	$Q_k^-/\text{Å}^2$	$\sigma_k^+/e\text{nm}^{-2}$	Subgrupo	$Q_k/\text{Å}^2$				
CH2CHO	13,36	73,69	0,0155	CH2COCH2	59,64				
				CH2COCH	24,60				
				CH3CHO	81,48				
				CH2CHO	60,53				
CH3COAC3	8,61	15,01	0,0146	CH3COAC3	95,09				
c-CH2COCH2	12,54	38,50	0,0149	c-CH2COCH2	70,09				
Triacetin	31,21	56,55	0,0131	Triacetin	266,77				
GBLactone	16,24	37,17	0,0145	GBLactone	108,54				
CH3COOCH3	26,05	96,92	0,0099	CH3COOCH3	129,35				
				CH3COOCH2	97,07				
				CH3COOCH	86,31				
				CH2COOCH2	87,24				
				Phthalate	30,24	62,33	0,0119	AC6H4(CH2COO)2	187,12
				CH2COOAC3	11,33	57,80	0,0142	CH2COOAC3	86,35
				CH3OCH2	11,71	20,44	0,0082	CH3OCH3	93,22
								CH3OCH2	56,47
								CH3OC	28,28
								CH2OCH2	50,48
CHOCH	9,11								
c-CH2OCH2	7,31	24,48	0,0129					c-CH2OCH2	46,91
N(CH2)3	3,60	25,72	0,0157					N(CH2)3	32,07
DMSO	10,22	20,55	0,0169	DMSO	89,14				
CH2Cl	8,37	8,16	0,0080	CH2Cl	52,27				
CCl4	16,56	23,61	0,0056	CCl4	124,95				
CH3CCl3	36,06	18,25	0,0036	CH3CCl3	132,23				
ClAC3	33,67	10,91	0,0024	ClAC3	68,91				
CHCl3	38,99	13,29	0,0031	CHCl3	102,83				
CH3OH	7,48	4,25	0,0119	CH3OH	61,43				
CH2OH	7,34	4,23	0,0130	CH2OH	39,45				
				CHOH	20,72				
H2O	7,96	10,95	0,0111	H2O	31,67				

Tabela A.6: Parâmetros de interação binária para ligações de hidrogênio utilizados no Teste 2.

Doador	Receptor	$E_{m,n}^{\text{HB}}/\text{kJmol}^{-1}$
CH3OH	ACH	0,5567
CH3OH	CH3COCH3	0,5074
CH3OH	CH3OH	1,9900
CH3OH	CH2OH	2,3927
CH3OH	H2O	0,5348
CH2OH	ACH	0,2414
CH2OH	CH3COCH3	0,2815
CH2OH	CH3COOCH3	0,9903
CH2OH	CH3OCH2	1,6977
CH2OH	CH3OH	1,3447
CH2OH	CH2OH	1,7229
CH2OH	H2O	0,8986
CHCl3	ACH	0,3795
CHCl3	CH3COCH3	0,7834
CHCl3	CH3COOCH3	1,6515
CHCl3	CH3OCH2	2,6956

Continua na página seguinte.

Tabela A.6 – Continuação da página anterior.

Doador	Receptor	$E_{m,n}^{HB} / \text{kJmol}^{-1}$
CHCL3	CH3OH	1,8555
CHCL3	CH2OH	1,7689
H2O	ACH	0,1931
H2O	CH3COCH3	4,1817
H2O	CH2CHO	3,6685
H2O	CH3COOCH3	4,7116
H2O	CH3OCH2	5,1416
H2O	CH3OH	5,1714
H2O	CH2OH	4,8578
H2O	H2O	4,6988

Tabela A.7: Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 3.

Grupo	$Q_k^+ / \text{Å}^2$	$Q_k^- / \text{Å}^2$	$\sigma_k^+ / \text{enm}^{-2}$	Subgrupo	$Q_k^0 / \text{Å}^2$
CH2	0	0	0	CH3	36,07
				CH2	19,54
				CH	-1,02
				C	-15,09
				c-CH2	21,25
C=C	6,16	3,70	0,0050	c-CH	4,68
				CH2=CH	53,88
				CH=CH	38,41
				CH2=C	34,96
				CH=C	19,74
ACH	5,38	6,71	0,0056	c-CH=CH	40,20
				ACH	18,87
CH3COCH3	21,97	80,23	0,0133	AC	4,40
				CH3COCH3	102,04
				CH3COCH2	82,85
				CH3COCH	62,64
				CH2COCH2	65,59
CH2CHO	14,85	81,88	0,0172	CH2COCH	47,11
				CH3CHO	82,41
				CH2CHO	61,26
CH3COAC3	9,56	16,68	0,0162	CH3COAC3	104,62
c-CH2COCH2	13,93	42,78	0,0166	c-CH2COCH2	69,75
Triacetin	34,68	62,83	0,0146	Triacetin	237,55
GBLactone	18,05	41,30	0,0161	GBLactone	116,25
CH3COOCH3	28,94	107,69	0,0110	CH3COOCH3	112,74
				CH3COOCH2	92,86
				CH3COOCH	76,76
				CH2COOCH2	77,43
				Phthalate	33,60
CH2COOAC3	12,59	64,23	0,0158	CH2COOAC3	99,47
CH3OCH2	13,01	22,71	0,0091	CH3OCH3	88,86
				CH3OCH2	72,06
				CH3OC	30,08
				CH2OCH2	53,99
				CHOCH	13,66
c-CH2OCH2	8,12	27,20	0,0143	c-CH2OCH2	65,69
N(CH2)3	3,60	28,58	0,0175	N(CH2)3	53,41
DMSO	11,36	22,84	0,0188	DMSO	111,02

Continua na página seguinte.

Tabela A.7 – Continuação da página anterior.

Grupo	$Q_k^+/\text{Å}^2$	$Q_k^-/\text{Å}^2$	$\sigma_k^+/\text{enm}^{-2}$	Subgrupo	$Q_k/\text{Å}^2$
CH2Cl	9,30	9,07	0,0089	CH2Cl	59,21
CCl4	18,40	26,23	0,0062	CCl4	132,94
CH3CCl3	40,07	20,28	0,0039	CH3CCl3	132,25
ClAC3	37,41	12,13	0,0026	ClAC3	83,55
CHCl3	43,32	14,76	0,0035	CHCl3	116,11
CH3OH	8,31	4,72	0,0132	CH3OH	66,88
CH2OH	7,34	4,69	0,0145	CH2OH	52,53
				CHOH	33,70
H2O	8,84	12,16	0,0123	H2O	42,93

Tabela A.8: Valores dos parâmetros dos grupos e subgrupos utilizados no Teste 4.

Grupo	$Q_k^+/\text{Å}^2$	$Q_k^-/\text{Å}^2$	$\sigma_k^+/\text{enm}^{-2}$	Subgrupo	$Q_k^0/\text{Å}^2$
CH2	0	0	0	CH3	36,07
				CH2	19,54
				CH	-1,02
				C	-15,09
				c-CH2	21,25
				c-CH	4,68
C=C	8,51	4,24	0,0072	CH2=CH	53,88
				CH=CH	38,41
				CH2=C	34,96
				CH=C	19,74
				c-CH=CH	40,20
ACH	5,00	4,72	0,0060	ACH	18,87
				AC	4,40
CH3COCH3	20,42	30,78	0,0129	CH3COCH3	102,04
				CH3COCH2	82,85
				CH3COCH	62,64
				CH2COCH2	65,59
				CH2COCH	47,11
CH2CHO	20,62	19,85	0,0123	CH3CHO	82,41
				CH2CHO	61,26
CH3COAC3	22,40	44,39	0,0112	CH3COAC3	104,62
c-CH2COCH2	20,61	18,88	0,0134	c-CH2COCH2	69,75
Triacetin	61,23	73,50	0,0109	Triacetin	237,55
GBLactone	30,19	48,74	0,0126	GBLactone	116,25
CH3COOCH3	26,09	33,42	0,0116	CH3COOCH3	112,74
				CH3COOCH2	92,86
				CH3COOCH	76,76
				CH2COOCH2	77,43
				AC6H4(CH2COO)2	183,29
Phthalate	39,56	52,86	0,0116	AC6H4(CH2COO)2	183,29
CH2COOAC3	24,31	37,89	0,0105	CH2COOAC3	99,47
CH3OCH2	20,07	58,21	0,0090	CH3OCH3	88,86
				CH3OCH2	72,06
				CH3OC	30,08
				CH2OCH2	53,99
				CHOCH	13,66
				c-CH2OCH2	65,69
c-CH2OCH2	13,52	46,59	0,0125	c-CH2OCH2	65,69
N(CH2)3	7,83	22,44	0,0133	N(CH2)3	53,41
DMSO	24,01	44,18	0,0163	DMSO	111,02
CH2Cl	22,88	13,63	0,0073	CH2Cl	59,21

Continua na página seguinte.

Tabela A.8 – Continuação da página anterior.

Grupo	$Q_k^+/\text{Å}^2$	$Q_k^-/\text{Å}^2$	$\sigma_k^+/e\text{nm}^{-2}$	Subgrupo	$Q_k/\text{Å}^2$
CCl4	73,27	16,87	0,0020	CCl4	132,94
CH3CCl3	40,04	39,00	0,0040	CH3CCl3	132,25
ClAC3	45,20	32,61	0,0046	ClAC3	83,55
CHCl3	37,46	15,32	0,0040	CHCl3	116,11
CH3OH	15,19	13,38	0,0131	CH3OH	66,88
CH2OH	15,33	10,01	0,0128	CH2OH	52,53
				CHOH	33,70
H2O	16,77	19,10	0,0122	H2O	42,93

Os parâmetros de interação binária para ligações de hidrogênio utilizados nos Testes 3 e 4 são iguais aos utilizados no Teste 0 (descritos na [Tabela B.15](#)).

Apêndice B

Tabelas Resultantes das Estimações de Parâmetros

Tabela B.1: Valores dos parâmetros dos grupos resultantes do Teste 0.

Grupo	$Q_k^+ / \text{Å}^2$		$Q_k^- / \text{Å}^2$		$\sigma_k^+ / \text{enm}^{-2}$	
CH2	0		0		0	
C=C	6,16	± 16,63	3,70	± 6,75	0,0050	± 0,0089
ACH	5,38	± 4,26	6,71	± 4,63	0,0056	± 0,0026
CH3COCH3	21,97	± 5,51	80,23	± 32,48	0,0133	± 0,0011
CH2CHO	14,85	± 8,55	81,88	± 65,21	0,0172	± 0,0048
CH3COAC3	9,56	± 4,28	16,68	± 8,97	0,0162	± 0,0049
c-CH2COCH2	13,93	± 5,04	42,78	± 22,15	0,0166	± 0,0028
CH3COOCH3	28,94	± 10,76	107,69	± 75,46	0,0110	± 0,0016
Triacetin	34,68	± 11,09	62,83	± 27,34	0,0146	± 0,0037
GBLactone	18,05	± $1,59 \times 10^7$	41,30	± $1,54 \times 10^8$	0,0161	± $1,76 \times 10^4$
Phthalate	33,60	± 27,29	69,26	± 55,51	0,0132	± 0,0071
CH2COOAC3	12,59	± 38,99	64,23	± 280,65	0,0158	± 0,0253
CH3OCH2	13,01	± 30,72	22,71	± 59,49	0,0091	± 0,0111
c-CH2OCH2	8,12	± 4,98	27,20	± 22,73	0,0143	± 0,0039
N(CH2)3	3,60	± 7,16	28,58	± 140,08	0,0175	± 0,0148
DMSO	11,36	± 1,95	22,84	± 8,11	0,0188	± 0,0026
CH2Cl	9,30	± 6,34	9,07	± 5,09	0,0089	± 0,0030
CCl4	18,40	± 16,69	26,23	± 34,48	0,0062	± 0,0023
CH3CCl3	40,07	± 190,53	20,28	± 63,70	0,0039	± 0,0119
ClAC3	37,41	± 181,60	12,13	± 28,27	0,0026	± 0,0091
CH3OH	8,31	± 0,75	4,72	± 0,11	0,0132	± 0,0010
CH2OH	7,34	± 0,70	4,69	± 0,15	0,0145	± 0,0012
CHCl3	43,32	± 37,71	14,76	± 9,73	0,0035	± 0,0020
H2O	8,84	± 0,87	12,16	± 2,20	0,0123	± 0,0015

Tabela B.2: Valores dos parâmetros dos subgrupos resultantes do Teste 0.

Subgrupo	$Q_k / \text{Å}^2$	
CH3	46,61	± 5,77
CH2	24,59	± 3,51
CH	-0,53	± 1,09
C	-22,30	± 2,15

Continua na página seguinte.

Tabela B.2 – Continuação da página anterior.

Subgrupo	$Q_k/\text{Å}^2$	
c-CH ₂	27,48	± 3,72
c-CH	5,50	± 1,82
CH ₂ =CH	59,53	± 7,64
CH=CH	33,51	± 5,28
CH ₂ =C	33,37	± 5,15
CH=C	11,59	± 3,63
c-CH=CH	41,71	± 5,49
ACH	21,44	± 2,88
AC	1,73	± 0,94
CH ₃ COCH ₃	120,60	± 14,82
CH ₃ COCH ₂	93,05	± 12,56
CH ₃ COCH	68,28	± 10,70
CH ₂ COCH ₂	66,49	± 10,08
CH ₂ COCH	28,52	± 7,55
CH ₃ CHO	93,93	± 12,96
CH ₂ CHO	68,87	± 11,02
CH ₃ COAC ₃	104,54	± 19,46
c-CH ₂ COCH ₂	78,02	± 13,79
Triacetin	294,88	± 45,63
GBLactone	120,62	± 23,62
CH ₃ COOCH ₃	140,92	± 17,44
CH ₃ COOCH ₂	114,05	± 14,68
CH ₃ COOCH	91,08	± 12,79
CH ₂ COOCH ₂	87,59	± 12,30
AC ₆ H ₄ (CH ₂ COO) ₂	201,58	± 32,31
CH ₂ COOAC ₃	95,38	± 18,57
CH ₃ OCH ₃	104,43	± 13,86
CH ₃ OCH ₂	71,07	± 9,98
CH ₃ OC	31,48	± 6,51
CH ₂ OCH ₂	55,74	± 7,93
CHOCH	9,98	± 4,27
c-CH ₂ OCH ₂	52,30	± 8,04
N(CH ₂) ₃	35,66	± 11,32
DMSO	99,04	± 17,67
CH ₂ Cl	56,37	± 7,44
CCl ₄	140,43	± 19,19
CH ₃ CCl ₃	148,79	± 19,51
ClAC ₃	75,36	± 10,78
CHCl ₃	116,55	± 16,60
CH ₃ OH	70,02	± 8,04
CH ₂ OH	43,93	± 5,62
CHOH	22,47	± 3,85
H ₂ O	34,81	± 3,52

Tabela B.3: Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 0.

Doador	Receptor	$E_{m,n}^{\text{HB}}/\text{kJmol}^{-1}$	
ACH	CH ₃ OH	0,5512	± 0,7152
CH ₃ COCH ₃	CH ₃ OH	0,5733	± 1,6067
CH ₃ OH	CH ₃ OH	2,2484	± 0,8196

Continua na página seguinte.

Tabela B.3 – Continuação da página anterior.

Doador	Receptor	$E_{m,n}^{\text{HB}} / \text{kJmol}^{-1}$	
CH ₂ OH	CH ₃ OH	2,6647 ±	2,1159
H ₂ O	CH ₃ OH	0,5635 ±	1,5996
ACH	CH ₂ OH	0,1083 ±	0,5229
CH ₃ COCH ₃	CH ₂ OH	0,3189 ±	1,5285
CH ₃ COOCH ₃	CH ₂ OH	1,0986 ±	1,3540
CH ₃ OCH ₂	CH ₂ OH	1,9030 ±	4,3873
CH ₃ OH	CH ₂ OH	1,4975 ±	3,2130
CH ₂ OH	CH ₂ OH	1,8919 ±	0,7082
H ₂ O	CH ₂ OH	1,0047 ±	1,3093
ACH	CHCL ₃	0,4251 ±	0,3580
CH ₃ COCH ₃	CHCL ₃	0,8983 ±	1,7688
CH ₃ COOCH ₃	CHCL ₃	1,7859 ±	1,1972
CH ₃ OCH ₂	CHCL ₃	2,9884 ±	1,9489
CH ₃ OH	CHCL ₃	2,0590 ±	0,8782
CH ₂ OH	CHCL ₃	1,9668 ±	0,9379
ACH	H ₂ O	0,0717 ±	0,4808
CH ₃ COCH ₃	H ₂ O	4,9474 ±	0,6356
CH ₂ CHO	H ₂ O	4,0155 ±	0,9157
CH ₃ COOCH ₃	H ₂ O	5,4204 ±	0,5712
CH ₃ OCH ₂	H ₂ O	5,6866 ±	0,8502
CH ₃ OH	H ₂ O	5,5930 ±	0,6348
CH ₂ OH	H ₂ O	5,4537 ±	0,6835
H ₂ O	H ₂ O	5,1413 ±	0,6052

Tabela B.4: Valores dos parâmetros dos grupos resultantes do Teste 1.

Grupo	$Q_k^+/\text{Å}^2$		$Q_k^-/\text{Å}^2$		$\sigma_k^+/\text{enm}^{-2}$	
C=C	6,05	± 15,86	4,93	± 8,72	0,0055	± 0,0090
ACH	5,74	± 4,45	7,88	± 5,55	0,0059	± 0,0026
CH3COCH3	23,15	± 5,26	87,59	± 35,23	0,0133	± 0,0011
CH2CHO	14,75	± 7,44	89,95	± 64,16	0,0188	± 0,0049
CH3COAC3	9,38	± 3,72	18,90	± 9,79	0,0178	± 0,0052
c-CH2COCH2	13,65	± 4,30	47,39	± 21,86	0,0182	± 0,0029
CH3COOCH3	30,06	± 10,48	118,48	± 81,59	0,0111	± 0,0015
Triacetin	34,07	± 8,81	70,46	± 29,27	0,0159	± 0,0030
GBLactone	17,78	± $1,77 \times 10^7$	45,95	± $5,2 \times 10^7$	0,0176	± $7,76 \times 10^3$
Phthalate	32,10	± 20,79	78,36	± 48,27	0,0144	± 0,0063
CH2COOAC3	12,69	± 40,92	70,73	± 311,32	0,0173	± 0,0309
CH3OCH2	11,99	± 27,74	25,99	± 66,77	0,0098	± 0,0113
c-CH2OCH2	8,26	± 4,86	30,00	± 22,84	0,0155	± 0,0045
N(CH2)3	3,60	± 6,09	31,44	± 119,19	0,0191	± 0,0154
DMSO	11,14	± 2,11	25,71	± 9,27	0,0206	± 0,0025
CH2Cl	9,82	± 6,17	10,91	± 6,71	0,0099	± 0,0032
CCl4	17,74	± 14,29	29,03	± 34,91	0,0066	± 0,0021
CH3CCl3	42,01	± 181,20	24,28	± 68,33	0,0041	± 0,0111
ClAC3	40,81	± 212,55	13,99	± 36,44	0,0027	± 0,0097
CH3OH	7,90	± 0,71	4,72	± 0,10	0,0139	± 0,0010
CH2OH	7,19	± 0,70	4,75	± 0,15	0,0147	± 0,0011
CHCL3	46,59	± 46,49	17,20	± 12,33	0,0036	± 0,0023
H2O	9,18	± 1,33	14,39	± 3,04	0,0123	± 0,0020

Tabela B.5: Valores dos parâmetros dos subgrupos resultantes do Teste 1.

Subgrupo	$Q_k/\text{Å}^2$	
CH3	51,33	± 6,29
CH2	27,27	± 3,88
CH	-0,58	± 1,23
C	-24,79	± 2,27
c-CH2	30,39	± 4,09
c-CH	6,04	± 2,01
CH2=CH	65,32	± 8,37
CH=CH	36,74	± 5,79
CH2=C	36,55	± 5,66
CH=C	12,74	± 3,92
c-CH=CH	45,94	± 6,05
ACH	23,74	± 3,17
AC	1,85	± 1,05
CH3COCH3	134,16	± 16,27
CH3COCH2	102,33	± 13,71
CH3COCH	75,44	± 11,68
CH2COCH2	73,72	± 11,08
CH2COCH	31,65	± 8,27
CH3CHO	103,05	± 14,16
CH2CHO	75,55	± 12,10
CH3COAC3	116,39	± 21,78
c-CH2COCH2	87,46	± 15,35
Triacetin	335,75	± 49,60
GBLactone	134,17	± 24,92

Continua na página seguinte.

Tabela B.5 – Continuação da página anterior.

Subgrupo	$Q_k/\text{\AA}^2$	
CH ₃ COOCH ₃	154,84	± 19,00
CH ₃ COOCH ₂	125,82	± 16,08
CH ₃ COOCH	100,36	± 13,97
CH ₂ COOCH ₂	97,47	± 13,57
AC ₆ H ₄ (CH ₂ COO) ₂	228,56	± 35,93
CH ₂ COOAC ₃	108,59	± 20,31
CH ₃ OCH ₃	114,47	± 15,14
CH ₃ OCH ₂	79,78	± 11,07
CH ₃ OC	34,57	± 7,38
CH ₂ OCH ₂	61,89	± 8,82
CHOCH	11,09	± 4,85
c-CH ₂ OCH ₂	59,84	± 8,99
N(CH ₂) ₃	39,25	± 12,03
DMSO	110,02	± 18,83
CH ₂ Cl	61,48	± 8,31
CCl ₄	155,69	± 21,14
CH ₃ CCl ₃	165,41	± 21,43
CIAC ₃	84,00	± 11,88
CHCl ₃	129,84	± 18,20
CH ₃ OH	77,10	± 8,83
CH ₂ OH	50,21	± 6,29
CHOH	25,59	± 4,24
H ₂ O	36,66	± 3,79

Tabela B.6: Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 1.

Doador	Receptor	$E_{m,n}^{\text{HB}}/\text{kJmol}^{-1}$	
ACH	CH ₃ OH	0,7356	± 0,7470
CH ₃ COCH ₃	CH ₃ OH	0,7393	± 1,3920
CH ₃ OH	CH ₃ OH	2,3706	± 0,7190
CH ₂ OH	CH ₃ OH	2,9023	± 2,6950
H ₂ O	CH ₃ OH	0,6110	± 1,6671
ACH	CH ₂ OH	0,2647	± 0,5825
CH ₃ COCH ₃	CH ₂ OH	0,0383	± 1,6435
CH ₃ COOCH ₃	CH ₂ OH	1,1872	± 1,2695
CH ₃ OCH ₂	CH ₂ OH	2,0713	± 4,4385
CH ₃ OH	CH ₂ OH	1,6111	± 3,9874
CH ₂ OH	CH ₂ OH	2,0244	± 0,6250
H ₂ O	CH ₂ OH	1,0865	± 1,3800
ACH	CHCl ₃	0,4656	± 0,3838
CH ₃ COCH ₃	CHCl ₃	0,9662	± 1,8459
CH ₃ COOCH ₃	CHCl ₃	1,9167	± 1,1832
CH ₃ OCH ₂	CHCl ₃	3,2565	± 1,9471
CH ₃ OH	CHCl ₃	2,2673	± 0,9099
CH ₂ OH	CHCl ₃	2,1995	± 0,9547
ACH	H ₂ O	0,0651	± 0,5653
CH ₃ COCH ₃	H ₂ O	5,3532	± 0,7054
CH ₂ CHO	H ₂ O	4,4664	± 0,9907
CH ₃ COOCH ₃	H ₂ O	5,8552	± 0,6302
CH ₃ OCH ₂	H ₂ O	5,9619	± 0,6564

Continua na página seguinte.

Tabela B.6 – Continuação da página anterior.

Doador	Receptor	$E_{m,n}^{HB}/\text{kJmol}^{-1}$	
CH3OH	H2O	6,0276 ±	0,6368
CH2OH	H2O	5,8929 ±	0,6400
H2O	H2O	5,4722 ±	0,6441

Tabela B.7: Valores dos parâmetros dos grupos resultantes do Teste 2.

Grupo	$Q_k^+/\text{Å}^2$		$Q_k^-/\text{Å}^2$		$\sigma_k^+/\text{enm}^{-2}$	
CH2	0		0		0	
C=C	5,89 ±	18,28	3,12 ±	5,76	0,0046 ±	0,0095
ACH	5,09 ±	3,88	5,98 ±	3,95	0,0057 ±	0,0025
CH3COCH3	22,72 ±	6,10	71,99 ±	30,80	0,0123 ±	0,0013
CH2CHO	16,38 ±	10,84	73,21 ±	69,65	0,0154 ±	0,0049
CH3COAC3	10,04 ±	5,15	14,71 ±	9,13	0,0147 ±	0,0048
c-CH2COCH2	14,35 ±	5,94	38,23 ±	23,03	0,0152 ±	0,0029
CH3COOCH3	29,26 ±	12,42	96,68 ±	72,76	0,0104 ±	0,0018
Triacetin	34,51 ±	14,83	56,01 ±	27,81	0,0133 ±	0,0046
GBLactone	17,99 ±	$1,79 \times 10^7$	36,77 ±	$7,19 \times 10^7$	0,0146 ±	$5,63 \times 10^3$
Phthalate	32,83 ±	36,44	61,99 ±	65,13	0,0121 ±	0,0084
CH2COOAC3	12,75 ±	44,23	57,73 ±	307,56	0,0144 ±	0,0251
CH3OCH2	13,02 ±	32,11	20,18 ±	50,63	0,0086 ±	0,0108
c-CH2OCH2	8,36 ±	5,62	24,34 ±	23,88	0,0131 ±	0,0038
N(CH2)3	3,60 ±	9,45	25,72 ±	170,72	0,0157 ±	0,0166
DMSO	11,37 ±	1,86	20,33 ±	7,35	0,0170 ±	0,0030
CH2Cl	10,60 ±	6,53	7,55 ±	4,27	0,0083 ±	0,0027
CCl4	17,72 ±	20,73	23,27 ±	37,97	0,0056 ±	0,0027
CH3CCl3	36,15 ±	225,02	18,20 ±	70,39	0,0036 ±	0,0142
ClAC3	33,96 ±	172,06	10,54 ±	25,57	0,0026 ±	0,0093
CH3OH	8,93 ±	0,87	4,80 ±	0,13	0,0129 ±	0,0011
CH2OH	7,61 ±	0,78	4,55 ±	0,15	0,0140 ±	0,0013
CHCL3	39,80 ±	28,73	12,51 ±	6,64	0,0034 ±	0,0017
H2O	8,45 ±	0,56	10,51 ±	1,50	0,0128 ±	0,0012

Tabela B.8: Valores dos parâmetros dos subgrupos resultantes do Teste 2.

Subgrupo	$Q_k/\text{Å}^2$	
CH3	42,45 ±	5,58
CH2	22,52 ±	3,40
CH	-0,47 ±	1,03
C	-19,03 ±	2,00
c-CH2	25,15 ±	3,61
c-CH	4,95 ±	1,72
CH2=CH	54,00 ±	7,34
CH=CH	30,13 ±	5,02
CH2=C	30,00 ±	4,90
CH=C	10,43 ±	3,45
c-CH=CH	37,93 ±	5,27
ACH	19,48 ±	2,77
AC	1,57 ±	0,89
CH3COCH3	106,97 ±	14,26
CH3COCH2	83,31 ±	12,08
CH3COCH	61,32 ±	10,32
CH2COCH2	58,77 ±	9,64

Continua na página seguinte.

Tabela B.8 – Continuação da página anterior.

Subgrupo	$Q_k/\text{\AA}^2$	
CH ₂ COCH	25,23	± 7,19
CH ₃ CHO	82,80	± 12,33
CH ₂ CHO	61,24	± 10,43
CH ₃ COAC ₃	93,38	± 18,38
c-CH ₂ COCH ₂	68,39	± 13,01
Triacetin	256,92	± 43,90
GBLactone	106,42	± 23,44
CH ₃ COOCH ₃	125,48	± 16,74
CH ₃ COOCH ₂	100,57	± 13,98
CH ₃ COOCH	81,82	± 12,32
CH ₂ COOCH ₂	76,29	± 11,62
AC ₆ H ₄ (CH ₂ COO) ₂	177,47	± 30,64
CH ₂ COOAC ₃	82,86	± 17,77
CH ₃ OCH ₃	94,05	± 13,34
CH ₃ OCH ₂	61,03	± 9,30
CH ₃ OC	28,30	± 6,07
CH ₂ OCH ₂	49,23	± 7,49
CHOCH	9,04	± 4,01
c-CH ₂ OCH ₂	45,44	± 7,60
N(CH ₂) ₃	32,07	± 11,15
DMSO	87,64	± 17,24
CH ₂ Cl	51,17	± 6,92
CCl ₄	126,06	± 18,33
CH ₃ CCl ₃	132,85	± 18,64
ClAC ₃	65,84	± 10,05
CHCl ₃	103,99	± 15,98
CH ₃ OH	64,04	± 7,82
CH ₂ OH	38,84	± 5,37
CHOH	20,31	± 3,76
H ₂ O	32,77	± 3,41

Tabela B.9: Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 2.

Doador	Receptor	$E_{m,n}^{\text{HB}}/\text{kJmol}^{-1}$
ACH	CH ₃ OH	0,5333 ± 0,6904
CH ₃ COCH ₃	CH ₃ OH	0,4969 ± 1,7281
CH ₃ OH	CH ₃ OH	2,1612 ± 0,9123
CH ₂ OH	CH ₃ OH	2,4505 ± 2,0709
H ₂ O	CH ₃ OH	0,5230 ± 1,5682
ACH	CH ₂ OH	0,0229 ± 0,4908
CH ₃ COCH ₃	CH ₂ OH	0,2708 ± 1,6437
CH ₃ COOCH ₃	CH ₂ OH	0,9679 ± 1,5031
CH ₃ OCH ₂	CH ₂ OH	1,7040 ± 4,3961
CH ₃ OH	CH ₂ OH	1,3818 ± 3,0194
CH ₂ OH	CH ₂ OH	1,7615 ± 0,8370
H ₂ O	CH ₂ OH	0,9135 ± 1,4369
ACH	CHCl ₃	0,3869 ± 0,3414
CH ₃ COCH ₃	CHCl ₃	0,8131 ± 1,6919
CH ₃ COOCH ₃	CHCl ₃	1,6481 ± 1,1698
CH ₃ OCH ₂	CHCl ₃	2,7033 ± 1,9957

Continua na página seguinte.

Tabela B.9 – Continuação da página anterior.

Doador	Receptor	$E_{m,n}^{HB}/\text{kJmol}^{-1}$	
CH3OH	CHCL3	1,8237 ±	0,8514
CH2OH	CHCL3	1,7614 ±	0,9053
ACH	H2O	0,0583 ±	0,4314
CH3COCH3	H2O	4,4916 ±	0,5423
CH2CHO	H2O	3,6090 ±	0,8041
CH3COOCH3	H2O	4,9221 ±	0,4873
CH3OCH2	H2O	5,3106 ±	1,0985
CH3OH	H2O	5,1245 ±	0,6286
CH2OH	H2O	4,9078 ±	0,7453
H2O	H2O	4,7057 ±	0,5653

Tabela B.10: Valores dos parâmetros dos grupos resultantes do Teste 3.

Grupo	$Q_k^+/\text{Å}^2$		$Q_k^-/\text{Å}^2$		$\sigma_k^+/\text{enm}^{-2}$	
CH2	0		0		0	
C=C	6,01 ±	20,71	3,96 ±	10,29	0,0050 ±	0,0129
ACH	5,39 ±	9,46	6,72 ±	11,28	0,0056 ±	0,0081
CH3COCH3	23,22 ±	28,29	79,57 ±	108,97	0,0125 ±	0,0135
CH2CHO	14,27 ±	22,32	81,71 ±	148,71	0,0170 ±	0,0219
CH3COAC3	10,30 ±	13,87	16,65 ±	24,09	0,0163 ±	0,0200
c-CH2COCH2	13,82 ±	18,73	42,81 ±	65,20	0,0166 ±	0,0194
CH3COOCH3	29,97 ±	40,02	107,59 ±	172,04	0,0104 ±	0,0119
Triacetin	31,76 ±	49,31	63,58 ±	99,88	0,0145 ±	0,0201
GBLactone	18,64 ±	$1,86 \times 10^7$	41,15 ±	$4,11 \times 10^7$	0,0161 ±	$1,6 \times 10^4$
Phthalate	32,50 ±	59,74	70,10 ±	125,90	0,0132 ±	0,0205
CH2COOAC3	14,98 ±	49,90	64,04 ±	243,35	0,0160 ±	0,0374
CH3OCH2	12,41 ±	42,56	22,95 ±	82,85	0,0090 ±	0,0200
c-CH2OCH2	9,34 ±	13,82	27,04 ±	44,10	0,0145 ±	0,0178
N(CH2)3	3,67 ±	13,34	28,57 ±	221,11	0,0175 ±	0,0358
DMSO	13,05 ±	14,68	22,99 ±	28,89	0,0189 ±	0,0206
CH2Cl	10,13 ±	16,10	8,82 ±	13,73	0,0090 ±	0,0118
CCl4	17,41 ±	34,31	26,42 ±	63,42	0,0061 ±	0,0085
CH3CCl3	39,65 ±	193,91	20,73 ±	70,72	0,0039 ±	0,0136
ClAC3	37,94 ±	178,65	11,59 ±	31,33	0,0029 ±	0,0105
CH3OH	8,42 ±	9,19	4,83 ±	4,94	0,0132 ±	0,0142
CH2OH	7,47 ±	8,17	4,97 ±	5,11	0,0145 ±	0,0156
CHCL3	43,17 ±	80,98	15,46 ±	26,05	0,0035 ±	0,0055
H2O	9,61 ±	10,63	12,78 ±	14,97	0,0120 ±	0,0135

Tabela B.11: Valores dos parâmetros dos subgrupos resultantes do Teste 3.

Subgrupo	$Q_k/\text{Å}^2$	
CH3	41,15 ±	5,08
CH2	22,10 ±	3,10
CH	-1,02 ±	1,04
C	-15,18 ±	1,92
c-CH2	24,39 ±	3,26
c-CH	4,69 ±	1,72
CH2=CH	54,22 ±	6,81
CH=CH	38,28 ±	5,41
CH2=C	34,64 ±	4,93
CH=C	19,66 ±	3,76

Continua na página seguinte.

Tabela B.11 – Continuação da página anterior.

Subgrupo	$Q_k/\text{\AA}^2$	
c-CH=CH	39,36	± 5,01
ACH	19,45	± 2,56
AC	4,32	± 1,03
CH ₃ COCH ₃	109,04	± 13,17
CH ₃ COCH ₂	84,62	± 11,16
CH ₃ COCH	62,68	± 9,65
CH ₂ COCH ₂	64,76	± 9,28
CH ₂ COCH	47,05	± 8,03
CH ₃ CHO	86,13	± 11,92
CH ₂ CHO	63,36	± 10,50
CH ₃ COAC ₃	103,40	± 18,05
c-CH ₂ COCH ₂	70,50	± 12,28
Triacetin	248,26	± 37,26
GBLactone	115,99	± 22,29
CH ₃ COOCH ₃	117,51	± 14,97
CH ₃ COOCH ₂	100,98	± 12,94
CH ₃ COOCH	77,17	± 11,17
CH ₂ COOCH ₂	77,05	± 10,84
AC ₆ H ₄ (CH ₂ COO) ₂	184,28	± 29,59
CH ₂ COOAC ₃	96,24	± 17,88
CH ₃ OCH ₃	89,66	± 17,71
CH ₃ OCH ₂	74,10	± 9,62
CH ₃ OC	30,15	± 5,99
CH ₂ OCH ₂	53,85	± 7,22
CHOCH	13,63	± 4,09
c-CH ₂ OCH ₂	61,99	± 8,47
N(CH ₂) ₃	51,89	± 8,24
DMSO	108,94	± 16,74
CH ₂ Cl	57,09	± 7,07
CCl ₄	128,49	± 17,16
CH ₃ CCl ₃	133,00	± 16,95
ClAC ₃	81,79	± 10,26
CHCl ₃	113,88	± 15,60
CH ₃ OH	67,48	± 7,68
CH ₂ OH	47,90	± 5,64
CHOH	33,05	± 4,81
H ₂ O	33,80	± 3,24

Tabela B.12: Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 3.

Doador	Receptor	$E_{m,n}^{\text{HB}}/\text{kJmol}^{-1}$
ACH	CH ₃ OH	0,5886 ± 0,6502
CH ₃ COCH ₃	CH ₃ OH	0,5835 ± 1,3888
CH ₃ OH	CH ₃ OH	2,2981 ± 0,6973
CH ₂ OH	CH ₃ OH	2,6600 ± 1,8885
H ₂ O	CH ₃ OH	0,5663 ± 1,9507
ACH	CH ₂ OH	0,2064 ± 0,5012
CH ₃ COCH ₃	CH ₂ OH	0,2226 ± 1,4201
CH ₃ COOCH ₃	CH ₂ OH	1,0605 ± 1,2486
CH ₃ OCH ₂	CH ₂ OH	1,8836 ± 4,2341

Continua na página seguinte.

Tabela B.12 – Continuação da página anterior.

Doador	Receptor	$E_{m,n}^{HB}/\text{kJmol}^{-1}$
CH3OH	CH2OH	1,4947 ± 2,9111
CH2OH	CH2OH	1,8657 ± 0,6104
H2O	CH2OH	1,0196 ± 1,5962
ACH	CHCL3	0,4298 ± 0,3409
CH3COCH3	CHCL3	0,8713 ± 1,7002
CH3COOCH3	CHCL3	1,8003 ± 1,0967
CH3OCH2	CHCL3	2,9872 ± 1,8632
CH3OH	CHCL3	2,0518 ± 0,8607
CH2OH	CHCL3	1,9704 ± 0,9642
ACH	H2O	0,1703 ± 0,4046
CH3COCH3	H2O	4,6955 ± 0,5815
CH2CHO	H2O	3,9350 ± 0,8553
CH3COOCH3	H2O	5,0057 ± 0,5145
CH3OCH2	H2O	5,6831 ± 0,7374
CH3OH	H2O	5,5809 ± 0,6049
CH2OH	H2O	5,5476 ± 0,6676
H2O	H2O	5,0550 ± 0,5683

Tabela B.13: Valores dos parâmetros dos grupos resultantes do Teste 4.

Grupo	$Q_k^+/\text{Å}^2$		$Q_k^-/\text{Å}^2$		$\sigma_k^+/\text{enm}^{-2}$	
CH2	0		0		0	
C=C	6,90 ±	7,68	6,23 ±	5,73	0,0070 ±	0,0047
ACH	4,82 ±	5,56	4,84 ±	4,72	0,0054 ±	0,0036
CH3COCH3	15,93 ±	3,67	32,15 ±	12,42	0,0124 ±	0,0015
CH2CHO	18,51 ±	3,89	20,72 ±	6,48	0,0118 ±	0,0015
CH3COAC3	22,41 ±	31,53	44,38 ±	83,99	0,0112 ±	0,0078
c-CH2COCH2	18,09 ±	2,58	20,88 ±	4,56	0,0127 ±	0,0015
CH3COOCH3	19,08 ±	7,67	35,65 ±	22,24	0,0105 ±	0,0020
Triacetin	57,20 ±	39,07	74,70 ±	49,69	0,0105 ±	0,0047
GBLactone	27,99 ±	$2,79 \times 10^7$	49,47 ±	$4,94 \times 10^7$	0,0124 ±	$1,23 \times 10^3$
Phthalate	38,95 ±	18,83	53,04 ±	27,99	0,0115 ±	0,0034
CH2COOAC3	24,10 ±	41,00	37,94 ±	72,38	0,0105 ±	0,0097
CH3OCH2	20,07 ±	42,75	58,21 ±	206,33	0,0090 ±	0,0081
c-CH2OCH2	13,47 ±	9,40	46,59 ±	56,87	0,0125 ±	0,0036
N(CH2)3	7,81 ±	15,12	22,44 ±	67,02	0,0133 ±	0,0112
DMSO	21,16 ±	7,65	45,37 ±	14,07	0,0161 ±	0,0045
CH2Cl	22,27 ±	13,79	14,29 ±	7,44	0,0072 ±	0,0025
CCl4	73,04 ±	283,03	17,43 ±	26,19	0,0019 ±	0,0055
CH3CCl3	40,07 ±	275,54	38,99 ±	283,64	0,0040 ±	0,0139
ClAC3	45,09 ±	196,73	32,66 ±	94,03	0,0045 ±	0,0122
CH3OH	11,38 ±	1,70	14,25 ±	4,15	0,0124 ±	0,0019
CH2OH	11,38 ±	1,26	13,57 ±	2,46	0,0124 ±	0,0014
CHCL3	37,54 ±	43,11	15,24 ±	13,96	0,0040 ±	0,0030
H2O	15,13 ±	1,36	19,49 ±	2,67	0,0111 ±	0,0010

Tabela B.14: Valores dos parâmetros dos subgrupos resultantes do Teste 4.

Subgrupo	$Q_k/\text{Å}^2$	
CH3	35,47	± 5,31
CH2	19,00	± 3,31
CH	-1,02	± 1,50
C	-15,10	± 3,01
c-CH2	21,26	± 3,48
c-CH	4,68	± 3,10
CH2=CH	54,20	± 7,93
CH=CH	38,44	± 7,36
CH2=C	34,98	± 6,10
CH=C	19,74	± 5,38
c-CH=CH	40,26	± 6,04
ACH	18,47	± 2,85
AC	4,40	± 1,52
CH3COCH3	104,67	± 14,45
CH3COCH2	83,02	± 12,36
CH3COCH	62,62	± 13,88
CH2COCH2	65,43	± 11,68
CH2COCH	47,09	± 12,00
CH3CHO	82,38	± 12,96
CH2CHO	61,79	± 10,52
CH3COAC3	104,55	± 19,06
c-CH2COCH2	71,15	± 11,12
Triacetin	246,62	± 39,21
GBLactone	118,06	± 23,97
CH3COOCH3	114,12	± 16,44
CH3COOCH2	95,41	± 14,20
CH3COOCH	76,78	± 14,49
CH2COOCH2	77,55	± 13,11
AC6H4(CH2COO)2	184,32	± 30,42
CH2COOAC3	99,65	± 18,01
CH3OCH3	88,88	± 17,59
CH3OCH2	72,10	± 12,32
CH3OC	30,07	± 7,51
CH2OCH2	53,84	± 8,33
CHOCH	13,65	± 6,15
c-CH2OCH2	65,43	± 9,46
N(CH2)3	53,32	± 9,07
DMSO	111,71	± 18,50
CH2Cl	59,43	± 9,82
CCl4	131,40	± 19,98
CH3CCl3	132,12	± 21,33
ClAC3	83,46	± 14,97
CHCl3	114,91	± 17,80
CH3OH	71,57	± 8,95
CH2OH	53,82	± 6,88
CHOH	33,76	± 6,63
H2O	45,83	± 4,02

Tabela B.15: Parâmetros de interação binária para ligações de hidrogênio resultantes do Teste 4.

Doador	Receptor	$E_{m,n}^{HB}/\text{kJmol}^{-1}$
ACH	CH3OH	0,6165 ± 0,4179
CH3COCH3	CH3OH	0,5661 ± 1,4607
CH3OH	CH3OH	2,2494 ± 0,8927
CH2OH	CH3OH	2,6642 ± 2,3974
H2O	CH3OH	0,5868 ± 2,8290
ACH	CH2OH	0,2241 ± 0,3166
CH3COCH3	CH2OH	0,2999 ± 1,1150
CH3COOCH3	CH2OH	1,0926 ± 0,7619
CH3OCH2	CH2OH	1,8844 ± 1,4338
CH3OH	CH2OH	1,4983 ± 4,1075
CH2OH	CH2OH	2,1291 ± 0,5665
H2O	CH2OH	0,9863 ± 2,1054
ACH	CHCL3	0,4225 ± 0,4105
CH3COCH3	CHCL3	0,8731 ± 1,7432
CH3COOCH3	CHCL3	1,8391 ± 1,1639
CH3OCH2	CHCL3	2,9945 ± 1,4916
CH3OH	CHCL3	2,0561 ± 1,1806
CH2OH	CHCL3	1,9634 ± 1,1942
ACH	H2O	0,2058 ± 0,2767
CH3COCH3	H2O	4,6984 ± 0,2833
CH2CHO	H2O	4,0671 ± 0,2736
CH3COOCH3	H2O	5,2066 ± 0,2480
CH3OCH2	H2O	5,7170 ± 0,3510
CH3OH	H2O	5,6579 ± 0,5054
CH2OH	H2O	5,3100 ± 0,3724
H2O	H2O	5,2694 ± 0,3339