

DAN NI LIN

**DESVENDANDO INTERFACES FLUÍDO-FLUÍDO ATRAVÉS DE
SIMULAÇÕES MOLECULARES E APRENDIZADO DE MÁQUINA**

**Trabalho de Conclusão de Curso
apresentado à Escola Politécnica da
Universidade de São Paulo para obtenção
do diploma de Engenharia de Petróleo.**

SANTOS

2022

DAN NI LIN

**DESVENDANDO INTERFACES FLUÍDO-FLUÍDO ATRAVÉS DE
SIMULAÇÕES MOLECULARES E APRENDIZADO DE MÁQUINA**

**Trabalho de Conclusão de Curso
apresentado à Escola Politécnica da
Universidade de São Paulo para obtenção
do diploma de Engenharia de Petróleo.**

**Área de concentração: Aprendizado de
máquina e análise de dados**

**Orientador: Prof. Dr. Caetano Rodrigues
Miranda**

Colaborador: Dr. Alexsandro Kirch

SANTOS

2022

FICHA CATALOGRÁFICA

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo-na-publicação

Lin, Dan Ni
Desvendando Interfaces Fluido-Fluido Através de Simulações Moleculares
e Aprendizado de Máquina / D. N. Lin, A. Kirch, C. Rodrigues Miranda -- São
Paulo, 2022.
42 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São
Paulo. Departamento de Engenharia de Minas e de Petróleo.

1.Dinâmica molecular 2.Aprendizado de máquina 3.Recuperação avançada
de petróleo I.Universidade de São Paulo. Escola Politécnica. Departamento de
Engenharia de Minas e de Petróleo II.t. III.Kirch, Alexsandro IV.Rodrigues
Miranda, Caetano

AGRADECIMENTOS

Este trabalho teve apoio da Fapesp – Processo (2021/ 00953-4) e do RCGI – Research Center for Gas Innovation, localizado na Universidade de São Paulo (USP) e financiado pela FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo (2014/ 50279-4 e 2020/1239-5) e Shell Brasil, e a importância estratégica do apoio dado pela ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis) através do incentivo regulatório de P&D.

Gostaria de agradecer ao Prof. Dr. Caetano Rodrigues Miranda pela oportunidade de pesquisa e por todo compartilhamento de conhecimento durante o desenvolvimento do trabalho.

Gostaria de agradecer também ao Dr. Alexsandro Kirch pelo acompanhamento e apoio obtido, principalmente na interpretação do dado oriundo do modelo de aprendizado de máquina.

Por fim, agradeço também ao Dr. Paulo Henrique Ribeiro Amaral pelas sugestões de conteúdo para melhorar o modelo de aprendizado de máquina deste projeto.

RESUMO

Dentre os desafios atuais na indústria de petróleo e gás está o controle das propriedades interfaciais com o objetivo de promover o aumento da produção de poços maduros. Recentemente, a injeção de salmoura de baixa salinidade vêm se destacando dentre as técnicas de recuperação melhorada de petróleo (EOR), tanto pelo apelo econômico quanto ambiental. Neste trabalho investigou-se a relação entre propriedades estruturais dos fluidos e a tensão interfacial do sistema óleo/salmoura. Esse estudo é baseado em pós-processamento de simulações de dinâmica molecular e análises de dados usando aprendizado de máquina a fim de extrair tendências e correlações. A partir das trajetórias obtidas através de simulações prévias de dinâmica molecular, aplicou-se um código próprio de pós-processamento para analisar as propriedades estruturais da interface, como a curtose dos ângulos de orientação das moléculas e os comparamos com a os resultados da fase bulk. Com base no ranqueamento das características estruturais que mais influenciam a tensão interfacial, percebeu-se que dentre as famílias de moléculas, a densidade de aromáticos e parafínicos ficaram no topo da lista de classificação. Além disso, percebeu-se que o sistema possui uma anisotropia, sendo que os anéis aromáticos apresentam um alto ordenamento na interface com orientação preferencial paralela ao plano. Também mostramos que a tensão interfacial e a largura da interface óleo-salmoura podem ser representadas por um modelo linear. A compreensão das características estruturais dos fluidos e suas relações com a tensão interfacial pode levar ao desenvolvimento de fluidos otimizados para a recuperação melhorada de petróleo.

Palavras-chave: Aprendizado de Máquina, dinâmica molecular, tensão interfacial, óleo, salmoura.

ABSTRACT

Among the current challenges in the oil and gas industry is the control of interfacial properties to increase production from mature wells. Recently, the low salinity brine injection is one of the most preferred techniques of improved oil recovery (EOR), because of its economic and environmental appeal. In this work, the relationship between fluids structural properties and the oil/brine system interfacial tension was investigated. This study is based on post-processing of molecular dynamics simulations and data analysis using machine learning to extract trends and correlations. From previous molecular dynamics simulations data, a code was applied to analyze interface parameters, such as molecules kurtosis orientation angles and compared them with the bulk phase results. Based on the structural characteristics ranking that most influence the interfacial tension are the density of aromatics and paraffinic which were at the top of the classification list. In addition, it was noticed that in the oil/ brine system interface has an anisotropy, which means that the aromatic rings present a high ordering at the interface with a parallel preferential orientation to the interface plane. Also, it was demonstrated that the interfacial tension and the oil-brine interface width can be represented by a linear model. Understanding the structural characteristics of fluids and their relationships to interfacial tension can lead to the development of optimized fluids for improved oil recovery techniques.

Keywords: Machine Learning, molecular dynamics, interfacial tension, oil, brine.

LISTA DE FIGURAS

Figura 1 - Quantidade de descoberta de novos reservatórios de 2015 a 2021.	11
Figura 2 - Comparativo da eficiência do EOR com técnicas convencionais e injeção de salmoura de baixa salinidade para a recuperação secundária e terciária ao longo dos anos.....	12
Figura 3 - Escala de investigação realizadas para entender os mecanismos da EOR com salmoura de baixa salinidade.....	15
Figura 4 - (a) Representação molecular da interface óleo/salmoura e (b) os correspondentes perfis de densidade para modelos de água do mar (vermelho) e óleo de peso médio (linha preta), destacando as interfaces.	16
Figura 5 - Representação do modelo matemático adotado para as interfaces.....	17
Figura 6 - Esquema do protocolo computacional utilizado no projeto.....	18
Figura 7 - Metodologia do Gradient Descent. (Modificado).....	20
Figura 8 - Ranqueamento das propriedades que mais influenciam o valor da tensão interfacial.....	24
Figura 9 - Densidade dos modelos de óleo presente no sistema.	25
Figura 10 - Representação do “weak hydrogen bonds”.	25
Figura 11 - Figura explicativa da curtose e entropia.	26
Figura 12 - Valores da curtose da distribuição dos ângulos de orientação em relação aos eixos y e z. Os valores extraídos da interface foram comparados com os da fase bulk.....	27
Figura 13 - Esquema explicativo da anisotropia do sistema com relação ao ordenamento das moléculas aromáticas.	27
Figura 14 - Esquema ilustrativo do ordenamento das moléculas na interface salmoura/óleo.	28

Figura 15 - Relação linear entre a tensão interfacial e a largura da interface, após a utilização da regressão linear simples entre a largura da interface e da tensão interfacial.....	28
Figura 16 - Imagem explicativa da relação "weak hydrogens bonds" com a tensão interfacial.....	30
Figura 17 - Figura comparativa de previsão das técnicas de aprendizado de máquina utilizados.	31

SUMÁRIO

1	INTRODUÇÃO	8
2	OBJETIVOS	10
3	REVISÃO BIBLIOGRÁFICA	11
4	METODOLOGIA.....	16
5	RESULTADOS	24
6	CONCLUSÕES	32
7	REFERÊNCIAS	34
8	ANEXOS.....	37

1 INTRODUÇÃO

Dentre os desafios atuais na indústria de petróleo e gás está o controle das propriedades interfaciais, para aumentar o fator de recuperação de óleo. Técnicas de recuperação avançada de petróleo (EOR) visam a aumentar a produção de petróleo bruto dos reservatórios. Tais métodos são importantes para a segurança energética, pois as reservas são limitadas.[1] Dentre as técnicas de EOR, em especial a injeção de salmoura de baixa salinidade busca a redução da tensão interfacial entre salmoura e óleo e a mudança da molhabilidade da superfície das rochas.[2] Já como principais vantagens temos o seu fácil controle, baixo custo operacional e pelo fato de ser ambientalmente amigável. [3,4]

A tensão interfacial tem origem no contato entre fluidos imiscíveis. No caso da interface óleo/água, a tensão interfacial pode estar relacionada a estrutura da rede de ligações de hidrogênio. Uma ligação de hidrogênio é uma interação altamente direcional envolvendo hidrogênio e uma molécula polar. As propriedades anômalas da água e suas diferentes fases, por exemplo, podem ser explicadas com base nos detalhes da estrutura da rede de ligações de hidrogênio [5].

A tensão interfacial óleo/salmoura pode ser influenciada pela composição e concentração da salmoura e pelo modelo do óleo. Portanto, as variações na tensão interfacial observadas nessas interfaces têm uma relação complexa com as características moleculares dos fluidos imiscíveis que ainda não é totalmente compreendida.[6] Além disso, já foram realizados experimentos em que se prova que de fato a utilização da salmoura de baixa salinidade aumenta o fator de recuperação dos reservatório, porém é necessário realizar mais investigações em escala molecular a fim de entender as interações físico-químicas entre as interfaces que até no momento são pouco compreendidas.[7, 8] Nesse contexto, o aprendizado de máquina pode ser uma abordagem adequada para correlacionar propriedades físicas e químicas em um único modelo robusto.

Um interesse da indústria do petróleo é maximizar a produção de poços e isso pode ser alcançado através da manipulação das propriedades interfaciais [6] e dentre as possibilidades, tem-se a injeção de salmoura com baixo teor de sal. Portanto, um ponto crítico nesta área de pesquisa é melhorar o entendimento das relações entre as

propriedades interfaciais e a composição das fases em contato (óleo e salmoura) por meio da aplicação de métodos de aprendizado de máquina.

2 OBJETIVOS

O presente trabalho tem como base o estudo precedente [6] em que foi investigada a relação entre a tensão interfacial e a composição dos fluidos. Utilizando os dados de tensão interfacial e a composição dos fluidos deste estudo, agora busca-se compreender como a tensão interfacial está relacionada com características estruturais a nível molecular dos fluidos na interface. Em particular, procura-se relacionar quantitativamente o ordenamento molecular através das propriedades físico-químicas com a tensão interfacial a fim de compreender melhor os mecanismos da recuperação avançada de petróleo.

3 REVISÃO BIBLIOGRÁFICA

No contexto atual, a sociedade ainda é fortemente dependente do petróleo, em que 84,3% da geração de energia primária advém de combustíveis fósseis e mais especificamente 33,1% é oriundo do petróleo. [9]

Entretanto, segundo dados disponibilizados pelo Rystad Energy (figura 1), do período de 2015 a 2021, observa-se a queda gradual de descobertas de novos reservatórios de petróleo, sendo um indicativo da necessidade de intensificar a utilização de métodos de recuperação de petróleo com intuito de aumentar o fator de recuperação dos poços já em fase de produção dessa forma estendendo a vida útil da produção. [1]

Global discoveries for 2021 on course to lowest in decades /November volumes
Million barrels of oil equivalent

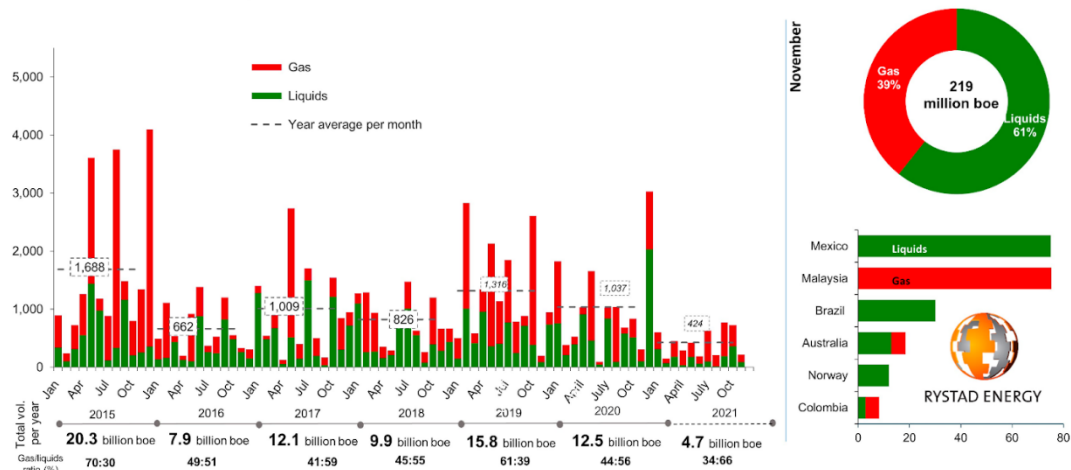


Figura 1 - Quantidade de descoberta de novos reservatórios de 2015 a 2021.

Fonte: *Rystad Energy*. (2021)

Uma das alternativas é a utilização das técnicas de recuperação avançada de petróleo (EOR) que visam a aumentar a produção de óleo bruto dos reservatórios. Dentre os desafios atuais na indústria de petróleo e gás está o entendimento e controle das propriedades interfaciais visando aumentar o fator de recuperação de óleo. [6]

De acordo com o estudo de Deng et al [10], as propriedades mais relevantes para as técnicas de EOR, são a tensão interfacial e a molhabilidade. Os autores

observaram que apenas com a redução do valor da tensão interfacial, já é possível aumentar o fator de recuperação consideravelmente.

Dentre as técnicas de EOR, a injeção da salmoura de baixa salinidade tem contribuído para alteração da molhabilidade das rochas, levando ao aumento do fator de recuperação de petróleo. A preferência pelas técnicas de EOR que são classificadas como “water-based”, dá-se pelo fácil controle e baixo custo operacional associado [9]. Além desses motivos, existe o fato de que tal técnica também é um método de baixo impacto ambiental. [4]

Quando se trata de técnicas de EOR de baixa salinidade, significa a redução ou alteração do conteúdo iônico no fluido que será injetado no reservatório. Na figura 2, observa-se um aumento no fator de recuperação em a) e c) e o *watercut* (indica a porcentagem de água que é produzida) em duas etapas em b) e a diminuição no corte de água em d) são todas uma resposta à injeção de salmoura de baixa salinidade. Sendo assim, observa-se que com a utilização da salmoura de baixa salinidade, há de fato um maior fator de recuperação e a redução da injeção de volume de fluido [8].

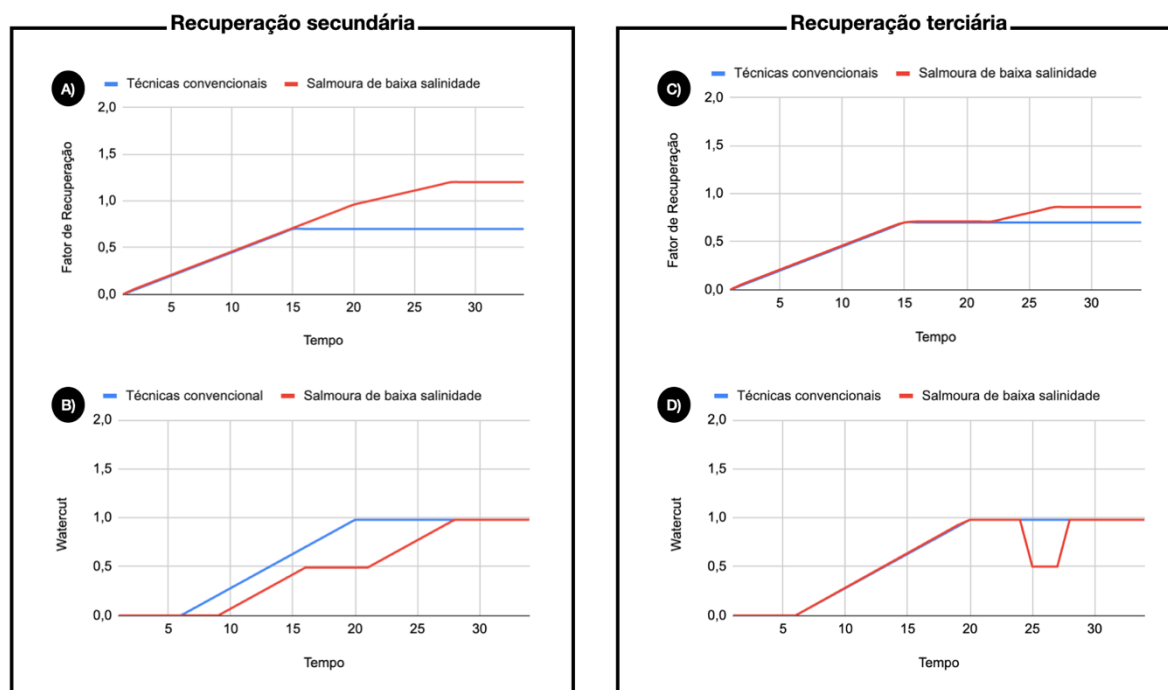


Figura 2 - Comparativo da eficiência do EOR com técnicas convencionais e injeção de salmoura de baixa salinidade para a recuperação secundária e terciária ao longo dos anos.

Fonte: Bartels, Mahani, Berg e Hassanizadeh (2019)

Em outro estudo feito por Makoto Kunieda et al. [6], foi investigado a interface água e óleo leve usando a dinâmica molecular. Verificou-se que a classe de moléculas aromáticas se concentra na região de interface, enquanto os demais hidrocarbonetos ficam distribuídos uniformemente por toda a fase oleosa. A pesquisa mostrou que a auto-acumulação de aromáticos na interface óleo-água é explicada pela formação dos assim chamados “weak hydrogen bonds” entre as moléculas aromáticas e a água, o que leva a obtenção de um valor menor de tensão interfacial (IFT do inglês – *Interfacial Tension*) quando comparada com outras interfaces água-óleo. As formações de camadas ricas em aromáticos na interface podem ser um fator determinante nos processos de EOR. [6]

Além disso, Derkani Maryam et al. [11] trazem uma outra visão das diferentes variáveis e os mecanismos interfaciais que influenciam o processo de recuperação avançada de petróleo com salmoura de baixa salinidade em reservatórios de carbonatos. Segundo o texto, os parâmetros dos reservatórios carbonáticos que influenciam no aumento do fator de recuperação ao utilizar a salmoura de baixa salinidade são a composição da água conata e seu pH, saturação inicial de água, composição do petróleo bruto, tempo de maturação, temperatura, pressão, a mineralogia da rocha, porosidade e permeabilidade. Já os parâmetros que mais influenciam no processo de EOR são a composição iônica, temperatura e a presença de surfactantes. O artigo mostra que há inconsistência da descrição dos processos entre a literatura e estudos experimentais, dentre os motivos são ao fato da técnica de recuperação avançada com salmoura de baixa salinidade depender de inúmeros parâmetros. De acordo com o artigo, a maioria dos estudos afirmam que tal método tem como resultado a alteração da molhabilidade dos reservatórios carbonáticos, o que possibilita o aumento do fator de recuperação. [11]

A aplicação da salmoura de baixa salinidade tem se mostrado efetivo em reservatórios carbonáticos e de arenitos. Com relação a concentração salina, para reservatórios de arenitos, a concentração ideal é menor que 5000 ppm, [2] mas há um consenso em outros estudos indicando que a concentração deve ser maior que 7000 ppm. Já em reservatórios carbonáticos, ainda não foi estipulado um valor ideal da concentração salina, mas normalmente se estipula uma concentração entre 10.000 e 20.000 ppm. Além da influência da salinidade, tem-se a composição química como

fator chave durante a recuperação avançada com salmoura de baixa salinidade. A remoção de íons divalente ou monovalente do fluido, acaba alterando o fator de recuperação tanto em rochas carbonáticas quanto em areníticos. Por exemplo, com a presença de íons Ca^{2+} , Mg^{2+} e SO_4^{2-} na salmoura de baixa salinidade, acaba por contribuir positivamente no processo de recuperação de petróleo. [2]

De acordo com a literatura [2], foi observado que a presença de grupos químicos com carga e grupos funcionais polares geram um campo potencial de interação, tendo como possibilidade a mudança da molhabilidade da superfície da rocha. Já quando há ausência desses elementos com carga ou polaridade, nenhum efeito positivo da salmoura de baixa salinidade foi observada no fator de produção dos reservatórios. Por fim, notou-se que a presença de óleo com componentes que possuem carga interage com os cátions presentes salmoura de baixa salinidade durante o processo de recuperação avançada, o que contribui na redução da tensão interfacial entre salmoura e óleo, facilitando o processo de recuperação, resultado da diminuição da resistência de capilaridade do sistema. [2]

No que tange a tecnologia para análises de dados, a inteligência artificial surge como uma poderosa ferramenta de predição, podendo auxiliar na otimização do processo produtivo e contribuir para o aumento da recuperação de hidrocarbonetos. [12]

Em estudo prévio foi combinado técnicas de aprendizado de máquina (ML) com simulações de dinâmica molecular (MD) para prever o valor da tensão interfacial entre óleo/salmoura (IFT), utilizando o Gradient Boosting (GB) e o método de regressão linear. Com base em algoritmos do GB, os autores observaram que o IFT está relacionado às propriedades físico-químicas e a concentração de salinidade da salmoura, e nenhum íon específico domina as alterações de IFT relativas as interações fluido-fluido. Quanto ao modelo de regressão linear obtido foi validado contra MD e dados experimentais da literatura, o erro variou até 2% e 9%, respectivamente, mostrando uma robusta e consistente transferência de dados. A combinação de simulações de MD e técnicas de ML pode fornecer uma determinação de IFT rápida e econômica ao longo múltiplas e complexas interfaces fluido – fluido e fluido – sólido. [7]

Expandindo o escopo da aplicação da inteligência artificial, mais especificamente tratando sobre as técnicas de aprendizado de máquina, no estudo feito por Erich A. Muller e Kezheng Zhu [13], foi utilizado as técnicas redes neurais artificiais (ANN) e regressão de processo gaussiana (GPR) para correlacionar e prever propriedades termodinâmicas como pressão e temperatura críticas, pressões de vapor e densidades de modelo puro fluido como alternativa para substituir as equações de estado para fluidos puros. Como resultado, foi constatado que ao comparar os modelos de previsão com os valores obtidos a partir das equações de estado, obteve-se boa aproximação, onde para ANN os coeficientes de performance foram R^2 (Coefficient of Determination) = 0,9974; MSE (Mean Squared Error) = $2,55 \times 10^{-3}$; AAD (Average Absolute Deviation) = 2,4%; já para GPR os coeficientes de performance foram R^2 = 0,9994; MSE = $4,87 \times 10^{-5}$; AAD = 2,4%. Logo, tais modelos de aprendizado de máquina construídos são uma nova opção para correlação e previsão das propriedades termo físicas de fluidos monofásicos. [13]

Para aplicação das técnicas de aprendizado de máquina, é necessário a construção de uma base de dados consistente, sendo a dinâmica molecular como uma opção para geração de dados. A dinâmica molecular trata-se da simulação computacional em nível atomístico que ao escolher um campo de força base, o sistema é capaz de simular e descrever todas as interações interatômicas por meio da integração das equações newtonianas de movimento para cada instante de tempo. [14] Estudos experimentais costumam ser custosos e as condições de reservatórios são muito difíceis de serem atingidas. Como vantagem, a dinâmica molecular surge como alternativa para atingir condições de reservatório e descrever as interfaces óleo/salmoura ainda pouco compreendidos.

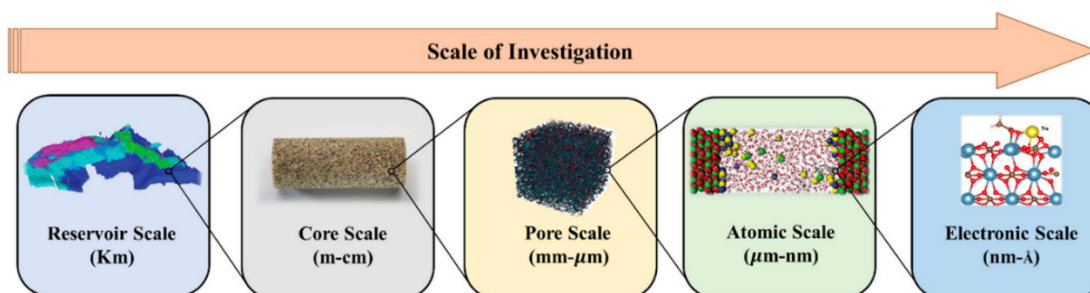


Figura 3 - Escala de investigação realizadas para entender os mecanismos da EOR com salmoura de baixa salinidade.

Fonte: Derkani, Fletcher, Abdallah, Sauerer, Anderson e Zhang (2018)

4 METODOLOGIA

Para o nosso estudo consideramos simulações prévias de dinâmica molecular (Alexsandro Kirch et al. [6]) envolvendo as condições normais de temperatura e pressão (300 K e 1 atm, respectivamente). O modelo de interface óleo/salmoura foi criado reunindo as fases aquosa e oleosa em caixas de simulação de $80 \times 80 \times 160$ (Å), conforme ilustrado na figura 4a), sob condições periódicas de contorno. Definidos os valores de densidade aparente (Figura 4b), todas as moléculas e íons foram adicionados em posições e orientações aleatórias usando o pacote PACKMOL. [15] O número total de átomos nos sistemas variou de 95.000 a 110.000.

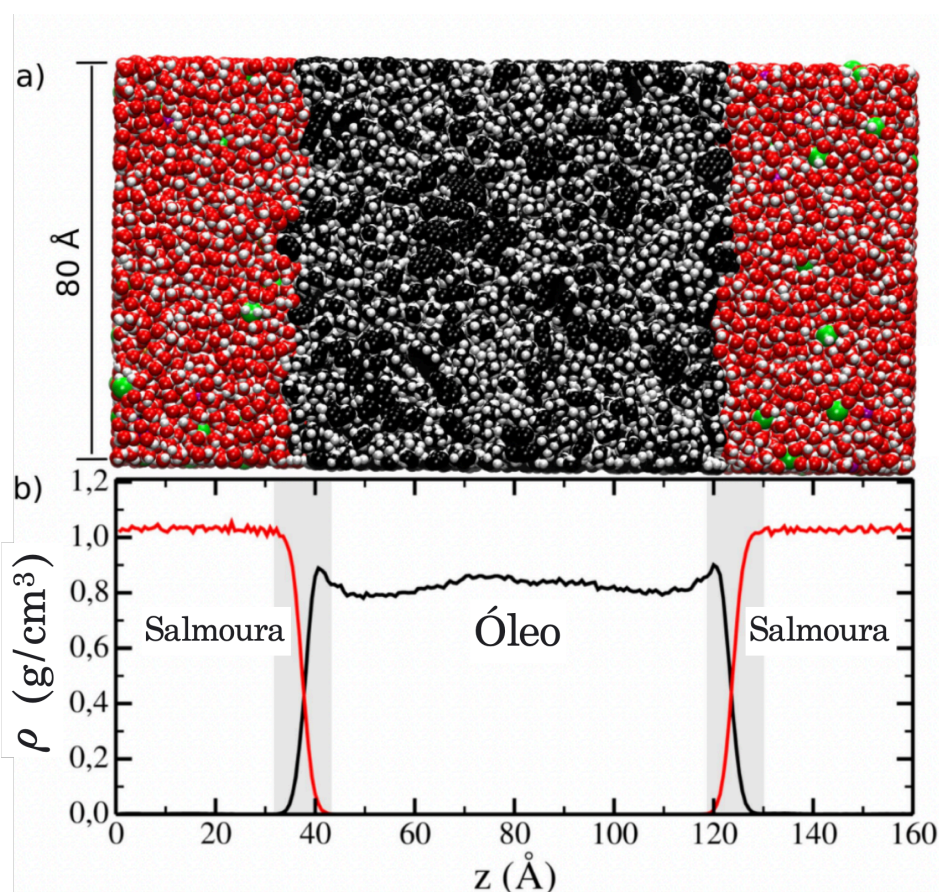


Figura 4 - (a) Representação molecular da interface óleo/salmoura e (b) os correspondentes perfis de densidade para modelos de água do mar (vermelho) e óleo de peso médio (linha preta), destacando as interfaces.

Fonte: KIRCH, Alexsandro; Brine–Oil Interfacial Tension Modeling: assessment of machine learning techniques combined with molecular dynamics.

Para a definição da região das interfaces, utilizou-se a equação contida na figura 5 e a equação conhecida como *Gibbs Dividing Surface*. Neste contexto considera uma linha imaginária que divide dois fluidos imiscíveis, sendo definida pelo cálculo do excesso de moléculas dos fluidos no outro em que a somatório seja zero, região em que não acontece adsorção das moléculas nos fluidos. [16]

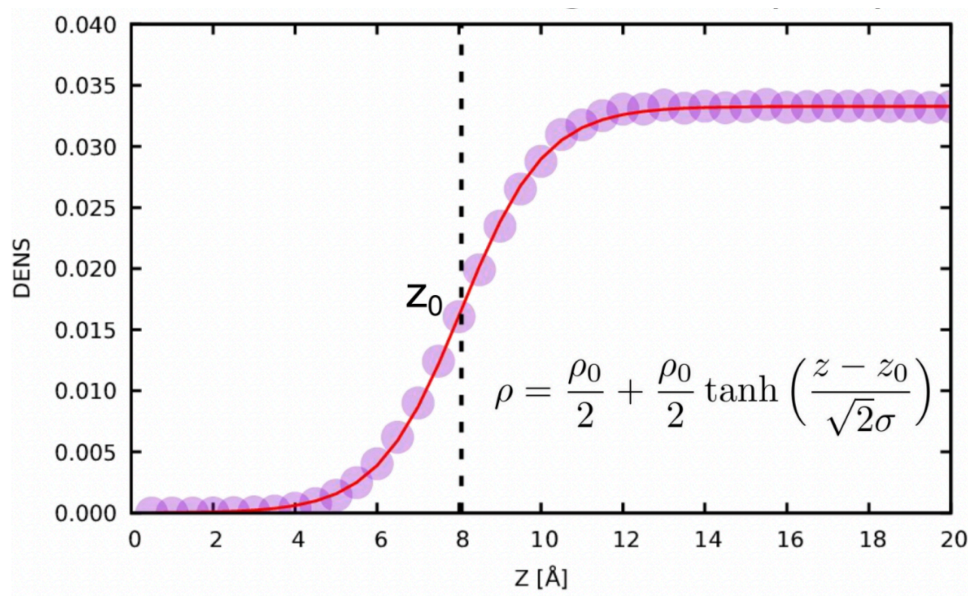


Figura 5 - Representação do modelo matemático adotado para as interfaces.

Fonte: Computer & Geoscience Vol. 14, No. 4, pp. 489-503, 1998

Neste estudo foram consideradas salmouras com diferentes composições, com a inclusão de íons (Na, K, Ca e Mg) nas concentrações 1%, 3,5% e 5%. Já os modelos de óleo utilizados são o heptano, o tolueno, heptol (mistura de heptano e tolueno), óleo leve e óleo médio que, por sua vez, apresentam densidades que variam de 0,72 a 0,95 g/cm³. Para mais detalhes da composição dos óleos, consulte as tabelas 1,2 e 3 na seção “anexos”.

A escolha do heptano e tolueno foi pelo fato de ambos os hidrocarbonetos possuírem a mesma quantidade de carbono, porém com diferentes funções orgânicas, já com relação a inclusão do heptol neste estudo, foi para averiguar o comportamento da mistura heptano/tolueno. Também incluímos modelos multicomponente para os modelos de óleo leve e médio e comparar os resultados com os modelos de óleos mais simples (heptano, tolueno e heptol). Para mais detalhes da composição do óleo leve e médio, consulte as tabelas 2 e 3 na seção “anexos” no fim deste trabalho.

Com base no arquivo de trajetórias do estudo precedente, criamos um banco de dados contendo informações sobre a estruturação das moléculas na interface, dentre elas, tem-se a orientação média das moléculas em relação à interface, curtose, tortuosidade, entropia e a densidade das moléculas do sistema. Esses dados foram usados para criar nosso modelo de aprendizado de máquina que relaciona as características estruturais do sistema com a tensão interfacial na região da interface salmoura/ óleo.

Posteriormente, criou-se 3 modelos de aprendizado de máquina (*XG Boost*, Regressão linear de múltiplas variáveis e regressão linear simples) que foi obtida por meio da criação de um código utilizando a linguagem de programação Python e finalmente com os dados processados, foi possível correlacionar os dados estruturais com a tensão interfacial (veja Figura 6).

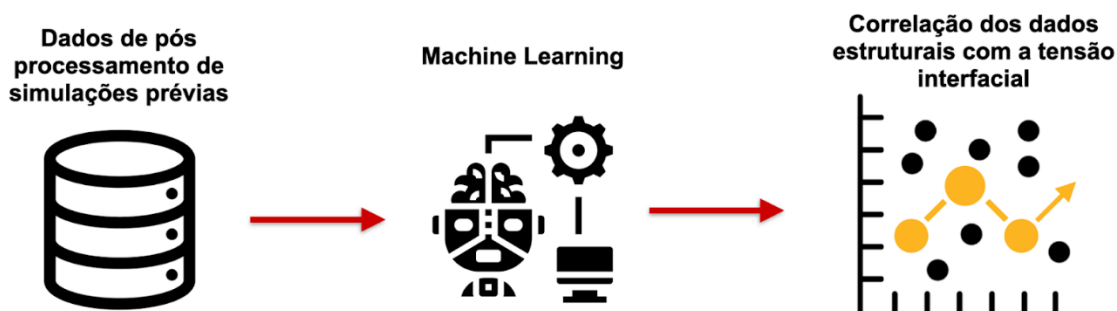


Figura 6 - Esquema do protocolo computacional utilizado no projeto.

Para entender o algoritmo por trás da técnica *XG Boost*, é necessário compreender primeiramente o funcionamento do *Gradient Boosting*. O *Gradient Boosting* é resultado de um conjunto de modelos fracos de árvores de decisão (também conhecido como “*weak learner*”) que são adicionados ao modelo gradualmente para a diminuição dos erros das árvores anteriores até quando se atinge um valor aceitável de erro entre o valor previsto e esperado ou quando não é mais possível minimizar o erro com a utilização do *Loss Function* ao adotar o *Gradient Descent*. O modelo preditivo final é obtido pela soma de todos os “*weak learners*” apresentado na equação 1. [17]

$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{j,m} I(x \in R_{j,m}) \quad (1)$$

, onde F_m é o valor a ser previsto,

$F_{m-1}(x)$ é o valor anterior previsto,

v é o learning rate que varia de 0 a 1,

$\sum_{j=1}^{J_m} \gamma_{j,m} I(x \in R_{j,m})$ refere – se a soma residual de todas os folhas de cada árvore de decisão criado.

Dando mais detalhes sobre o *Loss Function*, trata-se da escolha de uma função para minimizar a função erro com a utilização do *Gradient Descent* que é uma metodologia computacional que tem como objetivo em achar o ponto mínimo da função *Loss Function*. Geralmente a função utilizada como *Loss Function* é a função conhecida como soma residual dos quadrados (equação 2). [17]

$$L = \sum_i (y_i - \hat{y}_i)^2 \quad (2)$$

, onde y_i é o valor esperado do dado de treino

e $\hat{y}_i$ é o valor previsto a partir do dado de teste.

Na figura 7 tem-se a esquematização da metodologia do Gradient Descent com um exemplo de regressão linear simples, em que à medida que o modelo preditivo vai chutando valores de coeficiente lineares e angulares, espera-se que o erro também vá diminuindo ao achar o mínimo da derivada em função dos coeficientes linear e angular da equação 2 do exemplo dada na figura 7 e tal ideia é representada pela equação 3 e 4. [17]

$$L = \sum_i (y_i - \hat{y}_i)^2 \quad (3)$$

, onde $\hat{y}_i = mx_i + b$ e m é o coeficiente angular e b é o coeficiente linear,

no caso do exemplo da regressão linear simples.

$$\text{Min}(L) \rightarrow \nabla L \sim 0 \quad (4)$$

Entretanto *Gradient Descent* tem como por objetivo de dar pequenos passos ao longo da função para achar o ponto mínimo, desse modo inclui-se o *learning rate* (um fator que é definido pelo próprio algoritmo de 0 a 1) como mostrado na equação 5 e ela também é chamado como “*step size*”. Tal parâmetro é útil para quando se tem

funções em quando não é possível resolver a derivada igual a zero, mas sim se aproximar de um valor mais pequeno possível. [17, 18]

$$\text{Step size}_n = \text{learningRate} * \nabla L \quad (5)$$

, onde "n" refere – se a quantidade de coeficientes que será necessário para o modelo preditivo.

$$\begin{bmatrix} a_1 \\ \vdots \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \end{bmatrix} - \text{Step size}_n \quad (6)$$

, onde "a" são os coeficientes para a construção da equação matemática para predição

e "n" refere – se a quantidade de coeficientes necessários para o modelo preditivo.

No caso deste exemplo, n=2 e para ter o comportamento iterativo, implementa-se o seguinte protocolo computacional: [17, 19]

Input: Data $\{(x_i, y_i)\}_{i=1}^n$ e uma Loss Function diferenciável $L(y_i, F(x))$

Passo 1: Escolher inicialmente os valores dos coeficientes: $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$

Passo 2: Para $n = 1, \dots, n$:

(A) Calcule $\text{Step size}_n = \text{learningRate} * \nabla L$

(B) Calcule $\begin{bmatrix} a_1 \\ \vdots \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \end{bmatrix} - \text{Step size}_n$ até $\text{Min}(L) \rightarrow \nabla L \sim 0$

Passo 3: saída $\begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}$

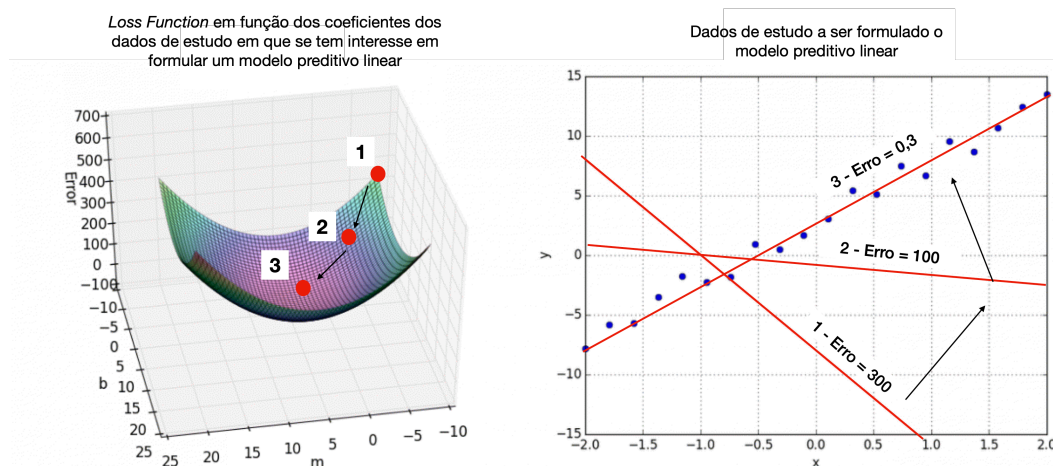


Figura 7 - Metodologia do Gradient Descent. (Modificado)

Fonte: <https://medium.com/analytics-vidhya/what-makes-xgboost-so-extreme-e1544a4433bb>

<Acessado em 19 de novembro de 2022>

Agora voltando para o algoritmo do Gradient Boosting, ele funciona da seguinte maneira [18]:

Input: Data $\{(x_i, y_i)\}_{i=1}^n$ e uma Loss Function diferenciável $L(y_i, F(x))$

Passo 1: inicialize o modelo com um valor constante:

$$F_0(x) = \operatorname{argmin} \text{ of } \gamma \sum_{i=1}^n L(y_i, \gamma)$$

Passo 2: para $m = 1$ até M :

(A) Calcule $r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x_i)=F_{m-1}(x_i)}$ para $i = 1, \dots, n$

(B) Ajuste as árvores de regressão para os valores de r_{im} e crie uma região terminal R_{jm}

para $j = 1, \dots, J_m$

(C) Para $j = 1, \dots, J_m$, calcule:

$$\gamma_{jm} = \operatorname{argmin} \text{ of } \gamma \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$$

(D) Atualize $F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{j,m} I(x \in R_{j,m})$

Passo 3: Saída $F_M(x)$

No protocolo computacional do *Gradient Boosting*, inicializa-se a técnica com a entrada de dados estruturados/ tabulados e a definição de uma *Loss Function* diferenciável que na maioria dos casos utiliza-se a soma residual dos quadrados multiplicada pelo fator $\frac{1}{2}$ para fins de simplificação das contas matemáticas (equação 7). [18]

$$L = \frac{1}{2} \sum_i (y_i - \hat{y}_i)^2 \quad (7)$$

, onde y_i é o valor esperado do dado de treino

e $\hat{y}_i$ é o valor previsto a partir do dado de teste.

No passo 1 do protocolo, deve-se obter o valor de $F_0(x)$ que é irá inicializar o código. A expressão $\operatorname{argmin} \text{ of } \gamma \sum_{i=1}^n L(y_i, \gamma)$, refere-se em calcular o valor de γ quando a soma das derivadas parciais da equação 7 é igual a zero, sendo assim ao obter o valor de γ também se tem o valor de $F_0(x)$. Já no passo 2, tem-se o momento da

iteração (loop) e a criação de M árvores de regressão. No passo 2(A), calcula-se o residual que é representada pela derivada do *Loss Function* (equação 8) e é neste momento em que o *Gradient Descent* é aplicado, quer-se achar os resíduos que tendam a zero. No fim desse passo, será calculado i resíduos, ou seja, dependendo de quantos dados para treino forem fornecidos, para cada dado de treino, ter-se-á um valor residual associado.

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x_i)=F_{m-1}(x_i)} = (y_i - F_{m-1}(x_i)) \quad (8)$$

, onde r refere – se ao cálculo do resíduo,

i indica a quantidade de dados para treino,

m refere – se a quantidade de árvores criadas,

y_i é o valor esperado (dados de treino)

e $F_{m-1}(x_i)$ é o valor constante obtido no passo 1

ou o valor previsto após a primeira iteração.

Com os cálculos dos i resíduos, no passo 2(B) será criada as árvores de regressão. Para cada iteração, será criada uma árvore de regressão com j quantidade de folhas que é representada pela letra $R_{j,m}$, lê-se folha j da árvore m. As folhas de cada árvores são criadas seguindo critérios pré-estabelecidas e os resíduos serão agrupados por folhas. Indo para o passo 2(C), para cada folha j da árvore de regressão m, deve-se obter o valor de γ ao calcular a derivada de $L(y_i, F_{m-1}(x_i) + \gamma)$ igual a zero. Note que é semelhante ao passo 1, só que é adicionado o valor de F_{m-1} e cada folha de cada árvore terá um γ associado. Prosseguindo para o passo 2(D), será obtido um novo valor de previsão do dado de interesse, sendo representado pela letra $F_m(x)$ por meio da soma da função de previsão obtida anteriormente $F_{m-1}(x)$ com a soma da parcela $\nu \sum_{j=1}^J \gamma_{j,m} I(x \in R_{j,m})$, onde ν é o *learning rate* e $\sum_{j=1}^J \gamma_{j,m} I(x \in R_{j,m})$ é o somatório de todos os γ de cada folha presente na árvore criado durante a iteração m. E por fim, o código só irá parar, quando se atingir a quantidade máxima M de árvores criadas ou quando a resíduo atingir o valor mínimo estipulado no passo 2(A), obtendo assim a que será o modelo preditivo $F_M(x)$ do conjunto de dados de interesse (passo3).

A técnica de aprendizado de máquina *XG Boost* é uma versão mais sofisticada do *Gradient Boosting*. Tal metodologia baseia-se em um conjunto de modelos (*ensemble*) de previsão fracos (geralmente é utilizado as árvores de decisão, sendo possível capturar correlações não lineares) com processos de otimização do modelo de previsão. A otimização é feita com a utilização do *Loss Function* e da função de regularização, onde o primeiro tem como objetivo de minimizar os erros da previsão para cada nova iteração do modelo e o segundo é responsável por evitar *overfitting* do modelo com a introdução de um *bias*. Entretanto, diferente do *Gradient Boosting*, o *XG Boost* é eficiente na substituição de dados faltantes do dataset, o *ensemble* é feito simultaneamente ao invés de sequencial, o processamento de dados é mais rápido quando comparado com outras técnicas de aprendizado de máquina, pois utiliza todos os processadores disponíveis do computador e no próprio algoritmo foram incluído várias melhorias, dentre elas tem-se a utilização de modelos mais complexos para a regularização do modelo, como *Lasso regression* e *Ridge regression*) e a inclusão do *cross-validation* dentro do próprio modelo. [20, 21, 22, 23]

A escolha da utilização do *XG Boost* neste trabalho é motivada, além de ser uma versão otimizada que deriva de várias outras técnicas de aprendizado de máquina, tal metodologia também tem sido uma das mais utilizadas atualmente para resolver problemas de regressão preditiva com dados estruturados/ tabulados. [22, 23]. Neste trabalho, foi utilizada a função regressão e o *Feature Importance*, que é uma função disponível no *XG Boost* desenvolvido com base da biblioteca Sklearn. O *Feature importance* é uma metodologia que busca ranquear as características estruturais que mais influenciam a tensão interfacial com o intuito de entender quais das 841 propriedades estruturais que mais influencia na tensão interfacial. Uma vez obtido o ranqueamento das propriedades, adotamos técnicas de aprendizados de máquina mais simples, como regressão linear e de múltiplas variáveis para previsão do valor da tensão interfacial e comparação da performance entre as técnicas de aprendizado de máquina utilizados. Além disso, também utilizamos as técnicas de regressão linear simples e de múltiplas variáveis, com o intuito de empregar modelos mais simples de aprendizado de máquina para estudar a relação entre o IFT e as propriedades físico-químicas.

5 RESULTADOS

Com a utilização do *XG Boost* para processar e correlacionar as propriedades estruturais com a tensão interfacial, obtivemos uma tabela (ver figura 8) que visa ranquear as propriedades que mais influenciam a tensão interfacial. Como principais resultados, nota-se que as características estruturais que mais influenciam a tensão interfacial são o tipo de densidade de carbono dos hidrocarbonetos, a entropia, a largura da interface, a curtose e pôr fim o “*weak hydrogen bonds*”. Para mais detalhes de quais tipos hidrocarbonetos são considerados as densidades dos carbonos e hidrogênio, veja a tabela 4 da seção “anexos”.

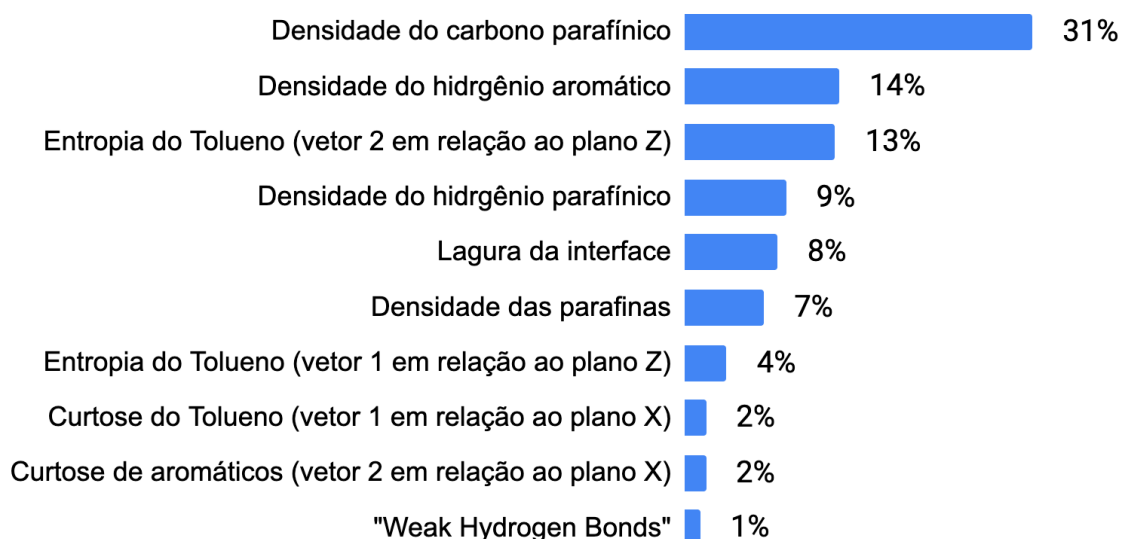


Figura 8 - Ranqueamento das propriedades que mais influenciam o valor da tensão interfacial.

A tabela presente na Figura 8 mostra que o tipo de molécula de óleo que está na interface influencia significativamente a tensão interfacial. Em particular, modelos de óleo que possuem aromáticos na sua composição (vide figura 9), como é o caso do tolueno, apresentam um acúmulo dessas moléculas na interface, favorecendo a interação com a água através de um tipo especial de ligações de hidrogênio conhecidas como “*weak hydrogen bonds*” [6]. Define-se “*weak hydrogen bonds*” como atração de natureza eletrostático entre moléculas que possuam hidrogênio e um outro elemento com moderado ou baixa eletronegatividade, onde a energia envolvida é menor que 4 kcal/mol. [24]

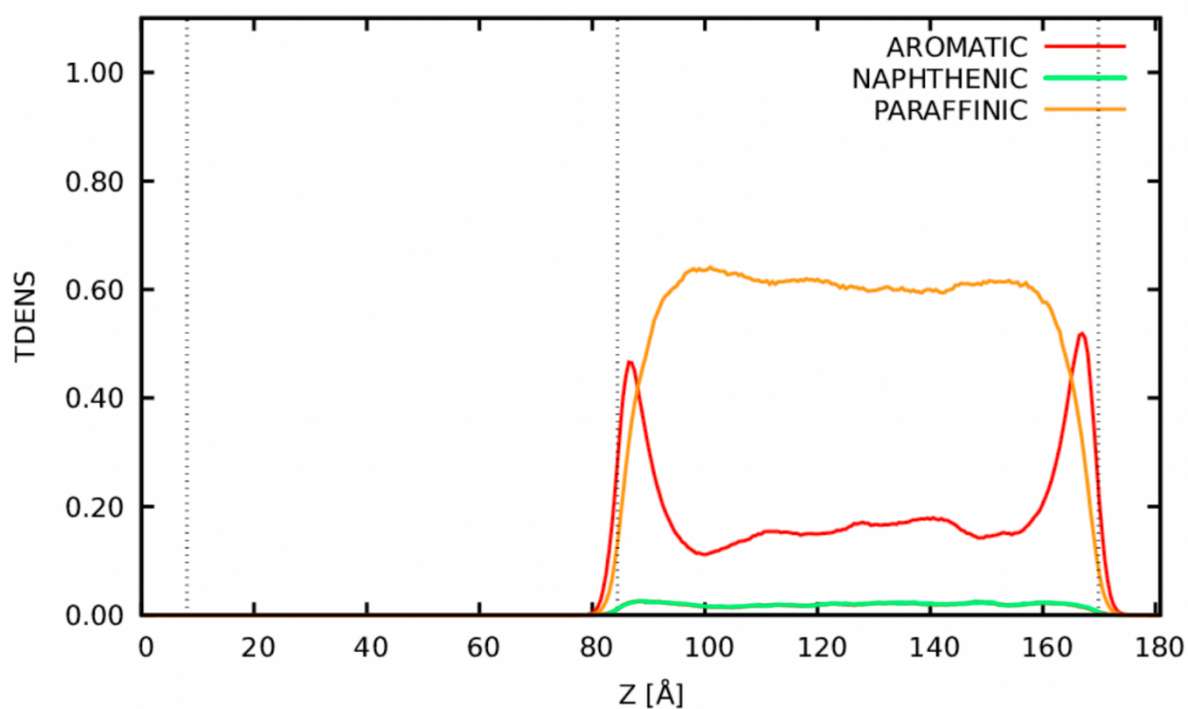


Figura 9 - Densidade dos modelos de óleo presente no sistema.

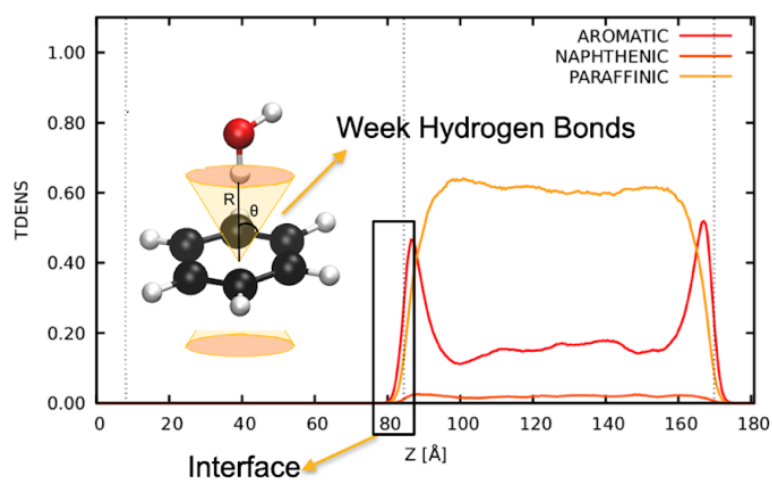


Figura 10 - Representação do “weak hydrogen bonds”.

O ranking apresentado na figura 8, mostra também que, além do tipo de molécula do óleo na interface, deve-se prestar atenção também em como moléculas na interface estão ordenadas. Neste contexto, a entropia configuracional de Shannon e a curtose são parâmetros relevantes para avaliar a relação da tensão interfacial com características estruturais da interface.

A entropia configuracional de Shannon e a curtose são parâmetros para quantificar o ordenamento das moléculas em um sistema. Tais propriedades são inversamente relacionadas: quanto maior o valor da entropia de Shannon, menor será o valor da curtose, indicando que as moléculas não estão paralelas (ordenados) em relação ao sistema de referência, já quando o valor da entropia de Shannon é pequeno, ter-se-á um valor maior da curtose e desta vez, as moléculas estarão paralelos (ordenados) em relação ao sistema referencial adotado.

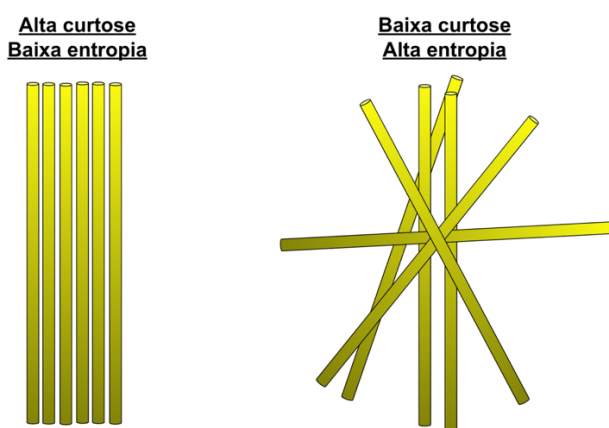


Figura 11 - Figura explicativa da curtose e entropia.

Neste trabalho enfatizou-se os resultados da curtose, sendo calculado pela equação 9. Observou-se que a curtose dos aromáticos calculada referente ao plano paralela a interface (plano z) possui um valor mais alto que a fase bulk, indicando um alto ordenamento das moléculas naquela direção como apresentado na figura 11. Já em relação ao plano perpendicular à interface (plano x e y), os anéis aromáticos possuem um valor similar à fase bulk (figura 12).

$$Curtose = \frac{\sum_i^N (x_i - \bar{x})^2}{N \sigma^4} \quad [graus] \quad (9)$$

, onde x_i refere – se ao ângulo das moléculas em relação a referência adotado

, $\bar{x}$ refere – se a média do ângulo calculado a partir do sistema

, N refere – se a quantidade de moléculas totais do sistema

, σ refere – se a variância.

Valores da curtose em relação a interface

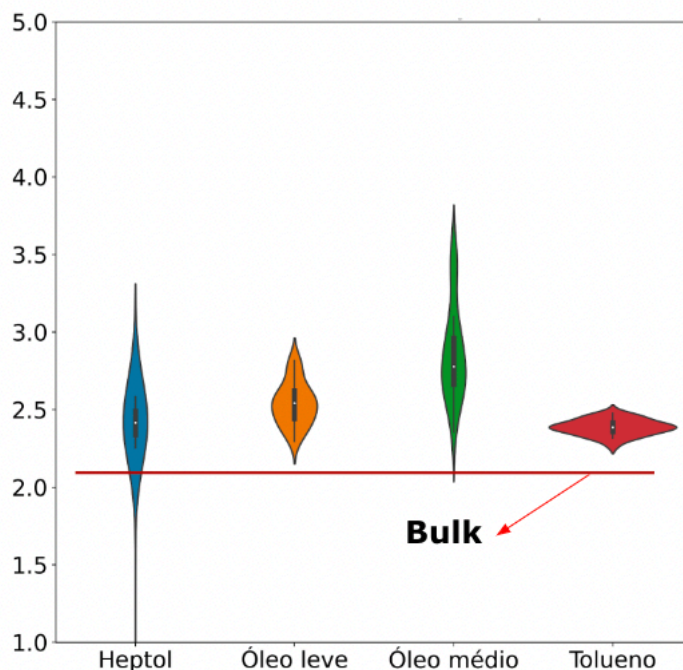


Figura 12 - Valores da curtose da distribuição dos ângulos de orientação em relação aos eixos y e z. Os valores extraídos da interface foram comparados com os da fase bulk.

Deste modo, com base na figura 12, percebe-se que existe uma anisotropia no sistema com relação ao ordenamento das moléculas aromáticas. De fato, a análise do ângulo médio de orientação dessas moléculas (figura 16) revelou que elas preferem dispor-se paralelamente ao plano da interface para aumentar a área de contato, ocasionando uma diminuição da tensão interfacial. Já os aromáticos em relação ao plano x e y estão dispostos perpendicular a eles (figura 14 e 15).

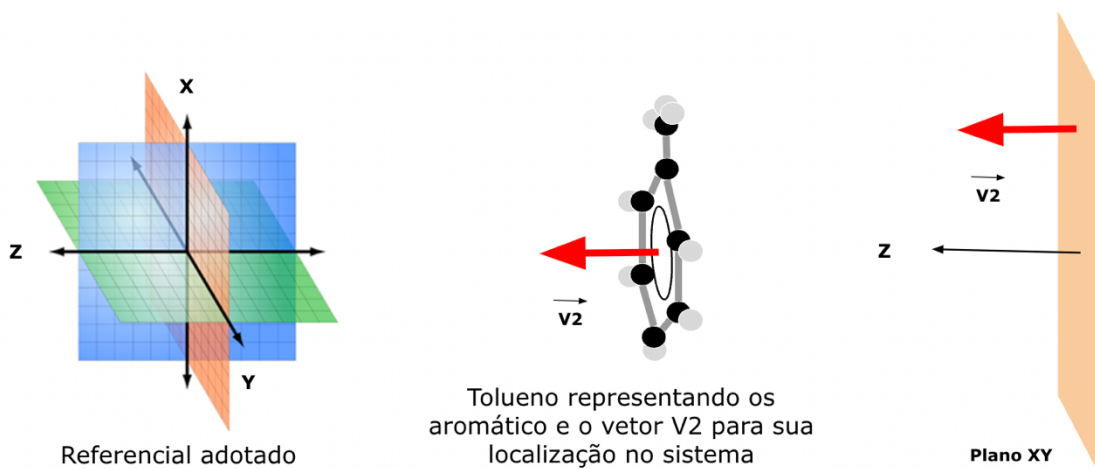


Figura 13 - Esquema explicativo da anisotropia do sistema com relação ao ordenamento das moléculas aromáticas.

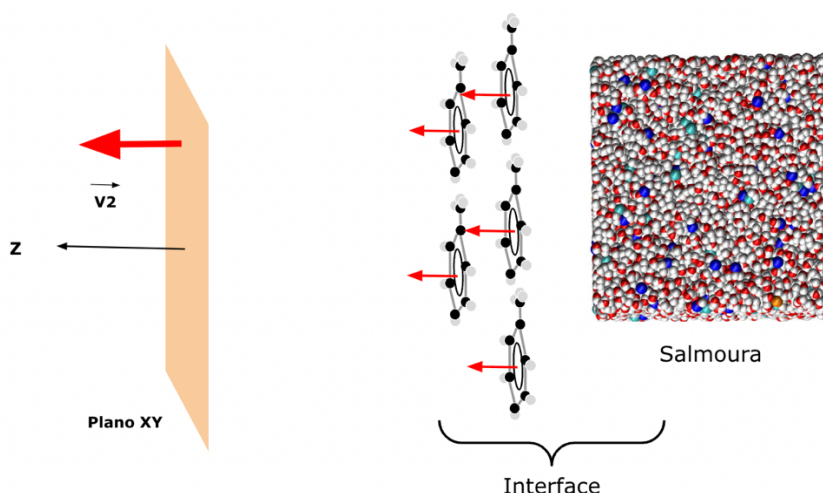


Figura 14 - Esquema ilustrativo do ordenamento das moléculas na interface salmoura/ óleo.

Na análise da classificação da importância dos atributos utilizados em nosso modelo de aprendizado de máquina (figura 8), observa-se que a largura da interface também é um atributo altamente relacionado com a tensão interfacial, como demonstrado na figura 15. Utilizou-se a técnica de regressão linear simples para relacionar a largura da interface com a tensão interfacial, observando que a largura da interface está inversamente relacionada com a tensão interfacial e de fato existe uma correlação linear entre essas duas propriedades. (equação 10)

$$IFT = -29,2 * \text{Largura da interface} + 109,3 \quad (10)$$

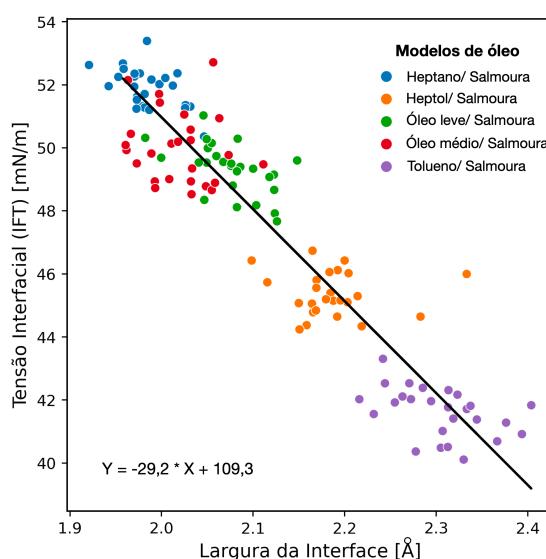


Figura 15 - Relação linear entre a tensão interfacial e a largura da interface, após a utilização da regressão linear simples entre a largura da interface e da tensão interfacial.

Ainda no gráfico da figura 15, pode-se notar que os modelos de óleo que contém maior composição de aromáticos, apresentam um valor mais alto da largura da interface e à medida que os modelos vão transacionando para aqueles que apresentam uma mistura de outros hidrocarbonetos até o óleo que só contém heptano, percebe-se que possuem um valor da largura da interface menor. Com base nessas observações, é possível interpretar que devido as cadeias lineares dos hidrocarbonetos, infere-se que tais moléculas tem uma preferência em se orientarem paralelamente em relação a interface. Já sobre os aromáticos, mesmo sabendo que possuem também orientação preferencial paralela a interface, uma das explicações possíveis por apresentarem valor da interface maior é dada a sua relação com os “*weak hydrogen bonds*” formados entre o anel benzênico e os hidrogênios das moléculas de água presentes na interface. Vide a imagem 16 para acompanhar a explicação.

Com o ranqueamento resultado do *XG Boost*, aplicamos mais uma vez a técnica de regressão linear e conseguimos obter mais uma equação matemática para previsão do valor da tensão interfacial só que com mais parâmetros envolvidos apresentada na Equação 11.

$$\begin{aligned}
 IFT &= 1052,9 * \text{Densidade do carbono parafínico} \\
 &- 2234,0 * \text{Densidade do carbono aromático} \\
 &+ 1,8 * \text{Entropia do tolueno (vetor2, plano Z)} \\
 &- 150,9 * \text{Densidade do hidrogênio parafínico} \\
 &- 8,9 * \text{Largura da interface} \\
 &- 6,3 \text{ Densidade das parafinas} \\
 &- 1,1 * \text{Entropia do tolueno (vetor1, plano Z)} \\
 &- 0,1 * \text{Curtose do tolueno (vetor1, plano Y)} \\
 &- 0,7 * \text{Curtose de aromáticos (vetor2, plano X)} \\
 &+ 66,1 \left[\frac{mN}{m} \right] \quad (11)
 \end{aligned}$$

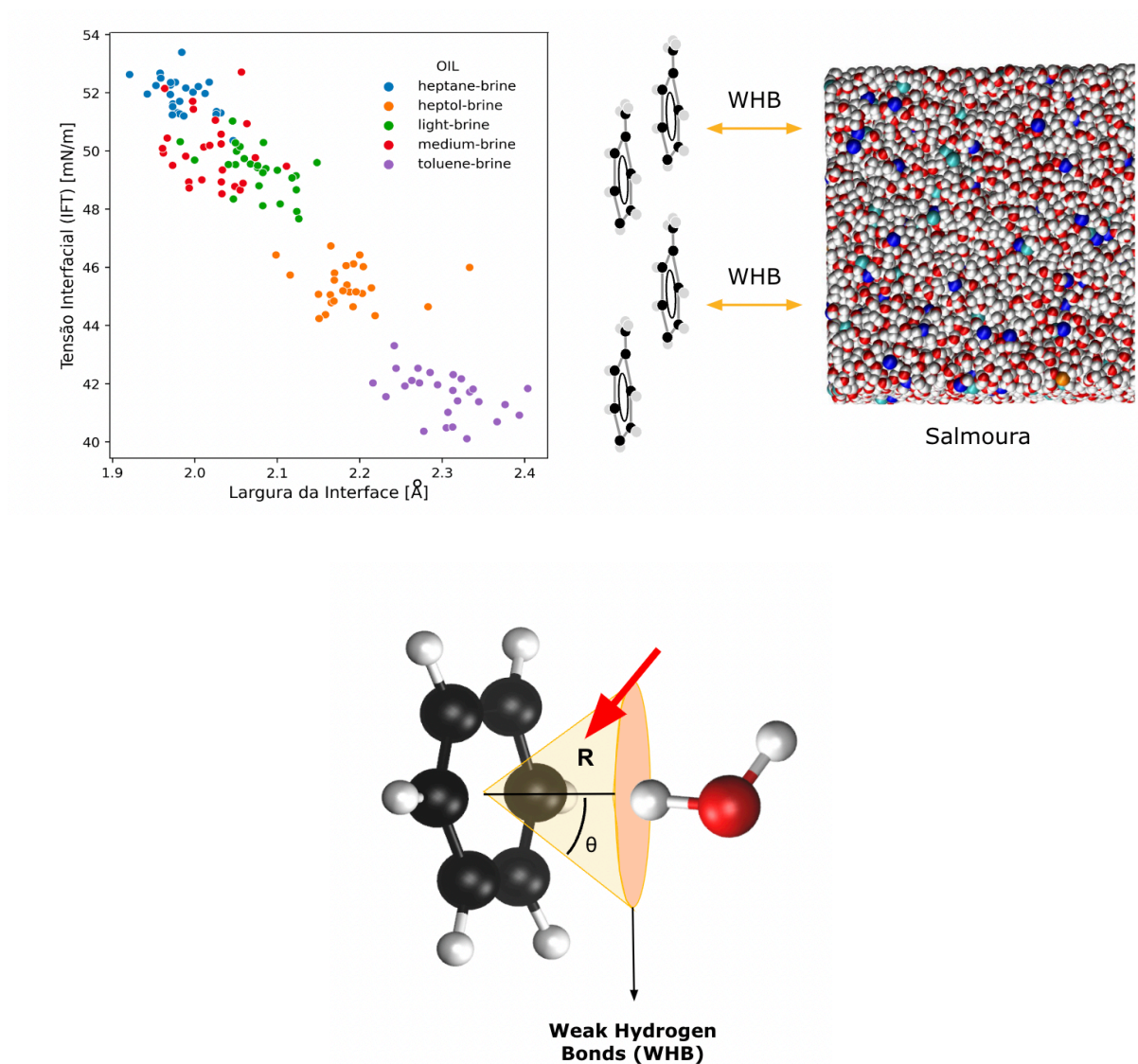


Figura 16 - Imagem explicativa da relação "weak hydrogens bonds" com a tensão interfacial.

Comparando o desempenho das técnicas de aprendizado de máquina usado para previsão da tensão interfacial como mostrado na Tabela 1, percebemos que o XG Boost apresenta um erro menor que o modelo de regressão de múltiplas variáveis e simples, cujo erro seria a diferença entre o valor esperado e calculado. Além disso, o XG Boost também apresentou a maior acurácia de previsão (R^2) da tensão interfacial.

Tabela 1 - Performance das técnicas de aprendizado de máquina utilizados.

Modelo de aprendizado de máquina	Erro [mN/m]	R2
XG Boost	0,38	0,98
Regressão linear de múltiplas variáveis	0,59	0,96
Regressão linear simples	1,1	0,83

Já na figura 18, percebe-se que à medida que o modelo de aprendizado de máquina é simplificado, observa-se o aumento da variância dos modelos preditivos, o que indica que outros parâmetros estruturais do banco de dados também influenciam significativamente no valor do IFT, mas que a princípio modelos mais simples já conseguem prever o valor do IFT com boa acurácia. No geral para todos os modelos utilizados, percebe-se que obtivemos erro consideravelmente baixo e R2 alto, indicando que a técnica de regressão conseguiu correlacionar satisfatoriamente as informações da dinâmica molecular. Além disso, para modelos mais simplificados é possível adotar a regressão linear como uma opção da obtenção do valor do IFT.

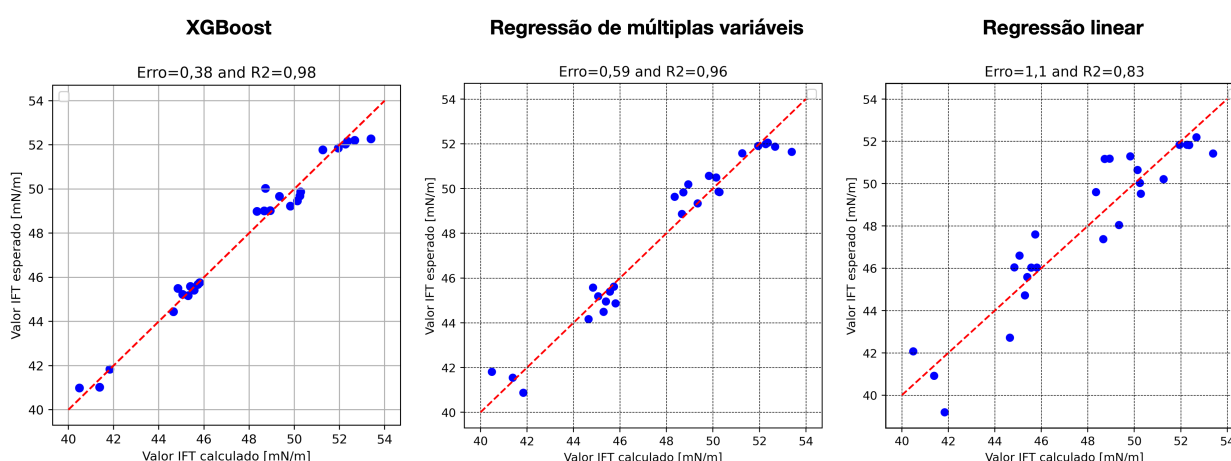


Figura 17 - Figura comparativa de previsão das técnicas de aprendizado de máquina utilizados.

6 CONCLUSÕES

Neste trabalho buscamos desvendar a interface óleo-salmoura através de técnicas de aprendizado de máquina aplicado a dados de simulações de dinâmica molecular. Obtivemos relações entre a tensão interfacial e as características estruturais do sistema que podem ser resumidas em cinco tópicos:

- *Feature Importance*

Através da técnica *XG Boost* ordenamos as características estruturais que mais influenciam a tensão interfacial. Mostrou-se que o tipo de molécula presente na interface (representado pelo perfil de densidade), bem como seu ordenamento são as características mais relevantes.

- Densidade de aromáticos e parafínicos

Percebeu-se que dentre as classes de moléculas, a densidade de aromáticos e parafínicos tem aparecido no topo do ranking mostrando que essas classes de moléculas são determinantes no resultado da tensão interfacial.

- Parâmetros de ordem

Mostrou-se que o sistema possui uma anisotropia, sendo que os anéis aromáticos apresentam um alto ordenamento na interface com orientação preferencial paralela ao plano da interface. Essa orientação busca maximizar a área de contato para diminuir a tensão interfacial. Esse resultado pode ajudar a propor novos fluidos de injeção para diminuir a tensão interfacial, uma vez que agora sabemos que os átomos de carbono do óleo são os que estão mais expostos à interface.

- Largura da interface x tensão interfacial

A tensão interfacial e a largura da interface óleo-salmoura podem ser representadas por um modelo linear. Com isso tem-se agora uma ferramenta que permite prever a tensão interfacial apenas com informações da largura da interface.

- Regressão para previsão da tensão interfacial

Por fim, criou-se três modelos de regressão, com erro inferior a 1,1 e R^2 superior a 0,85, indicando que os modelos de aprendizado de máquina construídos conseguiram capturar satisfatoriamente informações da dinâmica molecular, cujos dados foram testados experimentalmente, garantindo a qualidade dos resultados obtidos. Ademais, com modelos mais simplificados é possível adotar a regressão linear como uma opção da obtenção do valor do IFT.

Além dos aspectos fundamentais dos fenômenos interfaciais, este estudo pode contribuir para o desenvolvimento de fluidos de injeção mais eficazes no processo de EOR. De fato, o estudo revelou detalhes únicos sobre a interface óleo/salmoura que são difíceis de serem acessados através de técnicas experimentais.

7 REFERÊNCIAS

[1] PALZOR SHENGA. Rystad Energy. 2021 Global Oil And Gas Discoveries Projected To Sink To Lowest Level In 75 Years. 2021. Disponível em: <https://www.rystadenergy.com/newsevents/news/press-releases/2021-global-oil-and-gas-discoveries-projected-to-sink-to-lowest-level-in-75-years2/>. Acesso em: 19 jul. 2022.

[2] GHASEMI, Mehdi; SHAFIEI, Ali; FOROOZESH, Jalal. A systematic and critical review of application of molecular dynamics simulation in low salinity water injection. *Advances In Colloid And Interface Science*, [S.L.], v. 300, p. 102594, fev. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.cis.2021.102594>.

[3] MOHAMMADKHANI, Samira; SHAHVERDI, Hamidreza; ESFAHANY, Mohsen Nasr. Impact of salinity and connate water on low salinity water injection in secondary and tertiary stages for enhanced oil recovery in carbonate oil reservoirs. *Journal Of Geophysics And Engineering*, [S.L.], v. 15, n. 4, p. 1242-1254, 18 abr. 2018. Oxford University Press (OUP). <http://dx.doi.org/10.1088/1742-2140/aaae84>.

[4] LYU, Chaohui; ZHONG, Liguu; NING, Zhengfu; CHEN, Mingqiang; COLE, David R.. Review on Underlying Mechanisms of Low Salinity Waterflooding: comparisons between sandstone and carbonate. *Energy & Fuels*, [S.L.], v. 36, n. 5, p. 2407-2423, 21 fev. 2022. American Chemical Society (ACS). <http://dx.doi.org/10.1021/acs.energyfuels.1c04248>.

[5] MOORE, F. G.; RICHMOND, G. L.. Integration or Segregation: how do molecules behave at oil/water interfaces?. *Accounts Of Chemical Research*, [S.L.], v. 41, n. 6, p. 739-748, 29 maio 2008. American Chemical Society (ACS). <http://dx.doi.org/10.1021/ar7002732>.

[6] KIRCH, Alexsandro; CELASCHI, Yuri M.; ALMEIDA, James M. de; MIRANDA, Caetano R.. Brine–Oil Interfacial Tension Modeling: assessment of machine learning techniques combined with molecular dynamics. *Acs Applied Materials & Interfaces*, [S.L.], v. 12, n. 13, p. 15837-15843, 19 mar. 2020. American Chemical Society (ACS). <http://dx.doi.org/10.1021/acsami.9b22189>.

[7] KUNIEDA, Makoto; NAKAOKA, Kennichi; LIANG, Yunfeng; MIRANDA, Caetano R.; UEDA, Akira; TAKAHASHI, Satoru; OKABE, Hiroshi; MATSUOKA, Toshifumi. Self-Accumulation of Aromatics at the Oil–Water Interface through Weak Hydrogen Bonding. *Journal Of The American Chemical Society*, [S.L.], v. 132, n. 51, p. 18281-18286, 8 dez. 2010. American Chemical Society (ACS). <http://dx.doi.org/10.1021/ja107519d>.

[8] BARTELS, W.-B.; MAHANI, H.; BERG, S.; HASSANIZADEH, S.M.. Literature review of low salinity waterflooding from a length and time scale perspective. *Fuel*,

[S.L.], v. 236, p. 338-353, jan. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.fuel.2018.09.018>.

[9] RITCHIE, Hannah; ROSER, Max; ROSADO, Pablo. Energy mix. 2020. Disponível em: <https://ourworldindata.org/energy-mix>. Acesso em: 19 jul. 2022.

[10] DENG, Xiao; TARIQ, Zeeshan; MURTAZA, Mobeen; PATIL, Shirish; MAHMOUD, Mohamed; KAMAL, Muhammad Shahzad. Relative contribution of wettability Alteration and interfacial tension reduction in EOR: a critical review. Journal Of Molecular Liquids, [S.L.], v. 325, p. 115175, mar. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.molliq.2020.115175>.

[11] DERKANI, Maryam; FLETCHER, Ashleigh; ABDALLAH, Wael; SAUERER, Bastian; ANDERSON, James; ZHANG, Zhenyu. Low Salinity Waterflooding in Carbonate Reservoirs: review of interfacial mechanisms. Colloids And Interfaces, [S.L.], v. 2, n. 2, p. 20, 18 maio 2018. MDPI AG. <http://dx.doi.org/10.3390/colloids2020020>.

[12] SOLANKI, Parth; BALDANIYA, Dhruv; JOGANI, Dhruvikumar; CHAUDHARY, Bhavesh; SHAH, Manan; KSHIRSAGAR, Ameya. Artificial intelligence: new age of transformation in petroleum upstream. Petroleum Research, [S.L.], v. 7, n. 1, p. 106-114, mar. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.ptlrs.2021.07.002>.

[13] ZHU, Kezheng; MÜLLER, Erich A.. Generating a Machine-Learned Equation of State for Fluid Properties. The Journal Of Physical Chemistry B, [S.L.], v. 124, n. 39, p. 8628-8639, 1 set. 2020. American Chemical Society (ACS). <http://dx.doi.org/10.1021/acs.jpccb.0c05806>.

[14] NATURE. Molecular dynamics. Disponível em: <https://www.nature.com/subjects/molecular-dynamics#:~:text=Molecular%20dynamics%20is%20a%20method,materials%20chemistry%2C%20biochemistry%20and%20biophysics>. Acesso em 19 jul. 2022.

[15] Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: a Package for Building Initial Configurations for Molecular Dynamics Simulations. J. Comput. Chem. 2009, 30, 2157– 2164, DOI: 10.1002/jcc.21224

[16] Liquids and Solutions: Surfaces of Liquids: 2nd Year Michaelmas Term. Disponível em <http://rkt.chem.ox.ac.uk/lectures/liqsolns/liquid_surfaces.html> Acesso em 20 de nov. de 2022.

[17] What makes “XGBoost” so Extreme? By Eoghan Keany. Disponível em <<https://medium.com/analytics-vidhya/what-makes-xgboost-so-extreme-e1544a4433bb>> Acesso em 20 de nov. de 2022.

[18] Gradient Boost: Regression Details. Disponível em <<https://www.youtube.com/watch?v=2xudPOBz-vs>> Acesso em 20 de nov. de 2022.

[19] Gradient Descent, Step-by-Step. Disponível em <<https://www.youtube.com/watch?v=sDv4f4s2SB8&t=753s>> Acesso em 20 de nov. de 2022.

[20] Conheça o algoritmo XGBoost. Disponível em <<https://www.datageeks.com.br/xgboost/>> Acesso em 20 de nov. de 2022.

[21] XGBoost: Mathematical Details. Disponível em <<https://www.youtube.com/watch?v=ZVFeW798-2I>> Acesso em 20 de nov. de 2022.

[22] <https://www.youtube.com/watch?v=2xudPOBz-vs>

[22] About XG Boost. Disponível em < <https://www.madrasresearch.org/post/xgboost>> Acesso em 20 de nov. de 2022.

[23] Gradiente descendente, Laboratório de Estatística e Geoinformação - LEG/UFPR. Disponível em <http://cursos.leg.ufpr.br/ML4all/apoio/Gradiente.html#gradiente_descendente_boosting> Acesso em 20 de nov. de 2022.

[24] DESIRAJU, G.R.; STEINER, T. The Weak Hydrogen Bond in Structural Chemistry and Biology. New York: Oxford Univ. Press, Oxford And New York, 1999.

8 ANEXOS

Tabela 1 - Composição do óleo médio

Molécula	Porcentagem em peso	Molécula	Porcentagem em peso
Propano	0,6	Eicosano	20,7
Isobutano	0,3	Ciclohexano	0,5
Butano	1,1	Cicloheptano	0,9
Isopentano	0,8	Ciclooctano	1,1
Pentano	1,5	Benzeno	0,8
Hexano	1,9	o-xileno	0,6
Heptano	3,5	m-sileno	0,6
Octano	4,5	p-xileno	0,6
Nonano	4	123-trimetilbenzeno	0,9
Decano	3,6	124-trimetilbenzeno	0,9
Undecano	4,9	135-trimetilbenzeno	0,9
Dodecano	2,7	Naftaleno	2,4
Tridecano	4,5	Dimetilnaftaleno	1,8
Tetradecano	2,3	Tolueno	1,4
Petadecano	3,7	Antraceno	0,8
Hexadecano	1,7	Fenantreno	0,8
Heptadecano	2,5	Pireno	1,1
Octadecano	2,7	Benzopireno	14
Nonadecano	2,4		

Tabela 2 - Composição do óleo leve

Molécula	Porcentagem em peso
Hexano	11,14
Heptano	14,03
Octano	16
Nonano	20,71
Benzeno	4,21
Tolueno	12,91
Ciclohexano	7,25
Cicloheptano	13,75

Tabela 3 - Composição do heptol

Molécula	Porcentagem em peso
Heptano	50,00
Heptol	50,00

Tabela 4 - Amostra da presença de carbonos e hidrogênio presente em quais hidrocarbonetos

Tipos	Origem
Carbono parafínico	Isopentano, propano, butano, pentano, hexano, heptano, octano, nonano, decano, undecano, dodecano, tridecano, tetradecano, pentadecano, hexadecano, heptadecano, octadecano, nonadecano, eicosano
Hidrogênio aromático	Benzeno, tolueno, xileno, trimetilbenzeno, naftaleno, antraceno, pireno
Hidrogênio parafínico	Isopentano, propano, butano, pentano, hexano, heptano, octano, nonano, decano, undecano, dodecano, tridecano, tetradecano, pentadecano, hexadecano, heptadecano, octadecano, nonadecano, eicosano

Tabela 5 – Propriedades interfaciais

Numeração	Elemento e Função química
1	Curtose
2	Densidade de distribuição das moléculas na interface
3	Entropia
4	Hidrogênio aromático
5	Parâmetro de Ordem
6	Tortuosidade
7	<i>Weak Hydrogens Bonds</i>