ACCEPTED MANUSCRIPT

Brine–Oil Interfacial Tension Modeling: Assessment of Machine Learning Techniques Combined with Molecular Dynamics

A Kirch, YM Celaschi, JM de Almeida, CR Miranda

ACS Appl. Mater. Interfaces 12, 13, 15837-15843 (2020)

DOI: https://doi.org/10.1021/acsami.9b22189

This document is the Accepted Manuscript version of a Published Work that appeared in final form in ACS Applied Materials & Interfaces, copyright © 2020 American Chemical Society, after peer review and technical editing by the publisher. To access the final edited and published work see https://doi.org/10.1021/acsami.9b22189.

Brine-oil interfacial tension modeling: assessment of machine learning techniques combined with molecular dynamics

Alexsandro Kirch,[†] Yuri M. Celaschi,[†,‡] James M. de Almeida,[†,‡] and Caetano R. Miranda[*,†]

†*Instituto de Física, DFMT, Universidade de São Paulo, CP 66318, 05315-970, São Paulo, SP, Brazil*

‡*Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, 09210-580, Santo André, SP, Brazil*

E-mail: crmiranda@usp.br

Abstract

The physical chemistry mechanisms behind the oil-brine interface phenomena is not yet fully clarified. The knowledge of the relation between brine composition and concentration for a given oil may lead to the ionic-tuning of the injected solution on geochemical and enhanced oil recovery processes. Thus, it is worth examining the parameters influencing the interfacial properties. In this context, we have combined machine learning (ML) techniques with classical molecular dynamics simulations (MD) to predict oil/brine interfacial tensions (IFT) effectively, and compared to a linear regression (LR) method. To diversify our data set, we have introduced a new atomistic crude oil model (medium) with 36 different types of hydrocarbon molecules. The MD simulations were performed for mono and multi-component (toluene, heptane, heptol,

light, and medium) oil systems interfaced with sulfate and chloride brines with varying cations (Na^+, K^+, Ca^{2+}, and Mg^{2+}) and salinity concentration. Thus, a consistent IFT dataset was built for the ML training and LR fitting at room temperature and pressure conditions, over the feature space considering oil density, oil composition, salinity, and ionic concentrations. Based on gradient boosted (GB) algorithms, we have observed that the dominant quantities affecting the IFT are related to the oil attributes and the salinity concentration, and no specific ion dominates the IFT changes. When the obtained LR model was validated against MD and experimental data from the literature, the error varied up to 2% and 9%, respectively, showing a robust and consistent transferability. The combination of MD simulations and ML techniques may provide a fast and cost-effective IFT determination over multiple and complex fluid-fluid and fluid-solid interfaces.

Keywords

Machine Learning; Molecular Dynamics simulations; Oil/Brine Interface; Medium Oil Model; Linear Regression Model; Feature Importance Analysis

Introduction

Oil extraction processes can be evolved into three stages[1]. The primary extraction happens naturally, as there is a pressure gradient between the oil reservoir and the atmosphere after a well is drilled. It is followed by a secondary stage, which consists mainly of fluid injection, using them as a front wave to remove part of the remaining oil. The last one is the so-called enhanced oil recovery (EOR), which usually involves the injection of chemical species, such as surfactant-polymers, and/or brine, to reduce the wettability or to increase mobility. With the standard techniques, usually more than 60% of the total oil volume remains in the reservoir, trapped in the rock's porous media[2]. Therefore, it is essential to improve the technology of

the current EOR processes. Despite surfactants being regarded as the ideal candidates for EOR, such sophisticated synthetic methods, and logistic are not cost-effective.[3]

Nowadays, the improvements in the amount of extracted oil can be achieved by chemically manipulating the properties of complex and heterogeneous interfaces, which involves the reservoir's rock surface, the oil, and the injected fluid. In order to get optimum recipes, the water chemistry can be controlled by both ionic strength and composition modification which has been observed to enhanced the oil recovery[4]. This is known as the low-salinity brine injection, which among other mechanisms, reduces the oil/rock/brine interfacial tension (IFT)[5]. The IFT reduction leads to an increase in fluid diffusion and the miscibility/wettability[6]. Reducing interfacial tension allows the production of significant amounts of oil compared to usual water flooding methods[7]. In this context, Rotondi[8] et al. reported an optimal salinity ranging from 1000 to 2000 ppm (0.10 - 0.20 wt%), albeit the beneficial effect was observed even until 6000 ppm (0.60 wt%). However, the ideal brine content can vary for a given oil composition and reservoir conditions.

In the last decade, several mechanisms have been suggested to explain the low-salinity enhanced oil recovery phenomena[9–16]. Among them, the interface tension alteration by chemical manipulation was addressed in both experimental[17–20] and theoretical studies[21–25]. Buckley and Fan[18] measured the interfacial tension on the oil/brine interface. The authors observed the relation between the IFT, the acid number, and the viscosity of the oil phase. Employing molecular dynamics simulations, Underwood and Greenwell[22] studied the IFT of water-vapor, decane-vapor, and decane-water interfaces, considering a set of force fields. Also, the authors compared the results with the parameterization proposed by Smith and Dang[26], and with the reported experimental data, the results achieved with the selected force fields are in agreement, albeit the Smith and Dang[26] parameterization tends to overestimate the IFT values in higher salinities. Lara et al.[27] studied the effect of nanoparticles on the interface formed between a light-weight oil model and brine. They observed the interfacial tension of the studied systems increases for higher salinities. Besides these studies, Jian

and coauthors[20] observed the high concentration of asphaltenes displayed at the oil/water interface contributes to the IFT reduction.

High temperatures and pressures are the typical of oil reservoir thermodynamical, which are harsh and dangerous conditions to perform in the laboratory due to high costs and the possibility of serious injuries. In this context, the molecular simulations method is an efficient alternative to experiments. It may provide molecular scale information into the structure, and interaction properties of interfacial systems at extreme thermodynamical conditions.

The understanding of the physical and chemical processes at atomic resolution can lead to the ionic-tuning of the injected solution for the EOR[8]. Thus, it is worth examining the parameters influencing the interfacial phenomena. The relation between the IFT, the fluids composition, and the thermodynamic conditions may involve several variables. In this context, machine learning can be a suitable approach to correlate physical and chemical properties in a single and robust model.

ML has already been combined with molecular dynamics (MD) simulations in previous studies[28–31]. Wang and Ferguson[29] surveyed non-linear ML techniques to investigate biological materials, focusing on obtaining an optimized collective variable that can identify and discriminate meta-stable states. Jamal et al.[30] studied the effect of insulin-degrading enzymes. The authors got two active compounds with potential applications in the combat of Alzheimer's disease. Glazer et al.[31] proposed a new approach to improve the recognition of protein function. Riniker[28] obtained free energies of solvation for various solvents, and also partition coefficients for hydrocarbons and water. These works reveal the MD-ML coupling is an interesting toolbox to access the properties of the complex and heterogeneous systems such as oil/brine interfaces.

In this work, we have used MD to create a consistent dataset for the IFT at room thermodynamical conditions. In particular, we learned how different brine compositions and concentrations modify the IFT for a given oil model. Also, a new complex digital oil model has been proposed. With the obtained database, we have employed machine learning techniques to train IFT models, than, we compared with a model obtained from a linear

regression (LR). The obtained ML models had slightly better performance than the simple LR. Although, the LR model was still able to predict IFTs with errors from to 2% to 9% errors. In addition a ML method showed the driving features for the changes in IFT are related to the oil characteristics and brine salinity.

Methodology

Molecular Models

Crude oil is a multiphase and a high multiplicity components mixture involving mainly hydrocarbons[32]. In the context of atomistic simulations, a suitable molecular model is essential to mimic the real system complexity. The existent oil models are composed mainly by mono or bi-components[21,22,24,33–35]. Although a multicomponent light oil model (0.80 g/cm^3) has been previously proposed[6,27,36], it is limited to 8 distinguished molecules with 6 to 9 carbons in their chains.

To study more complex oils and diversify our dataset, we are proposing here a new molecular model with 36 components, having up to 20 carbon atoms per molecule and 0.84 g/cm^3 bulk density. The oil model was developed following the PNA approach, in which the organic molecules are divided into three separate fractions: (i) P–paraffinic (alkyl and branched alkanes), (ii) N–naphthenic (cycloalkanes), and (iii) A–aromatic (see **Tables S 1-3**). The chosen weight proportions were 69.08 (P), 3.86 (N), and 27.06 wt% (A) to represent the oil properties found in a pre-salt reservoir in Brazil.

Besides the proposed oil model, we investigated the toluene (100 wt% A), heptane (100 wt% P), heptol (50 wt% P and 50 wt% A), and the light oils. The last one has eight oil components distributed in PNA fractions with 61.80 P, 21.00 N, and 17.12 wt% A (see **Table S 4**).

We investigated a set of distinct brines with a variety of compositions and salinities. The seawater model with 3.5 wt% salinity being Na^+ (30.59%), K^+ (1.11%), Ca^{2+} (1.28%),

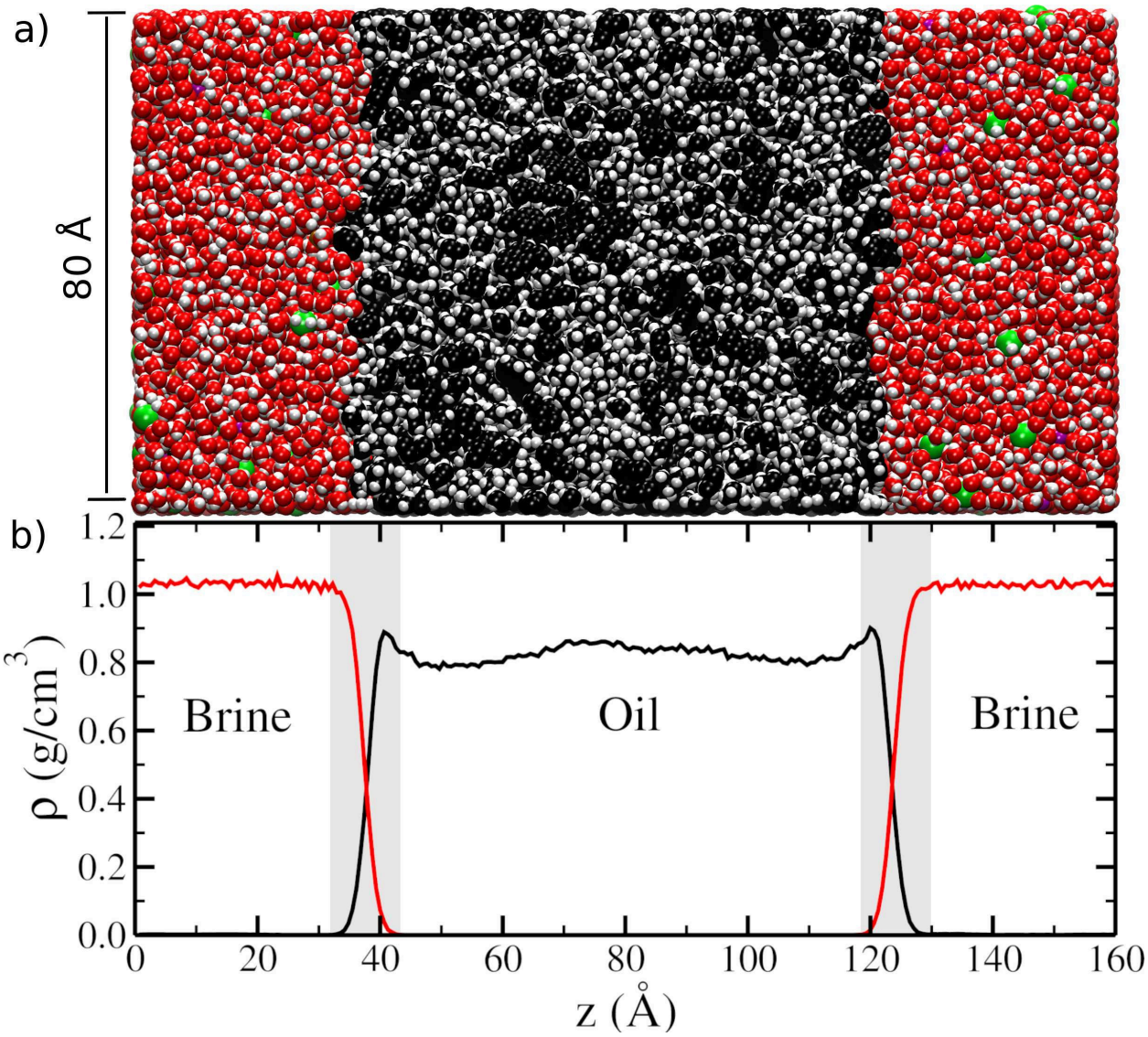


Figure 1: a) Molecular representation of the oil/brine interface and b) the corresponding density profiles for seawater (red) and the medium-weight oil (black line) models, highlighting the interfaces. Green, red, white, and black colors in a) represent the ions, oxygen, hydrogen, and carbon atoms, respectively.

Mg^{2+} (3.88%), Cl^- (55.25%), and SO_4^{2-} (7.89%), based on the most abundant ions in the oceans. The extended brine model follows the same composition and salinity, but with the SO_4^{2-} absence resulting in Na^+ (33.63%), K^+ (0.78%), Ca^{2+} (3.22%), Mg^{2+} (1.10%), and Cl^- (61.26%). We also considered brine models by combining the above cations and anions within 1.0, 3.5, and 5.0 wt% saline solutions.

The oil/brine interface model was created by putting together the aqueous and oil phases into $80 \times 80 \times 160$ (Å^3) simulation boxes, as depicted in **Figure 1**a. By complying bulk density values (see **Figure 1**b), all molecules and ions were added at random positions and orientations using the PACKMOL package[37]. The total number of atoms in the systems ranged from 95000 to 110000.

Molecular Dynamics simulations

Molecular dynamics simulations were performed with the Large Atomic/Molecular Massively Parallel Simulator (LAMMPS) software[38] with accelerator packages for Graphics Processing Unit (GPU)[39–41]. We have employed force fields available in the literature, as follows: (I) SPCE-FH[42] for water, Na^+ and Cl^-; (II) CGenFF[43] for all the studied oil molecules; (III) Gavryushov[44] for Mg^{2+} and Ca^{2+}; (IV) Lee et al.[45] for K^+; and (V) Williams and Carbone[46] for SO_4^{2-}. The force fields validation for the new oil model was performed by single oil component benchmarks. The bulk densities and the oil/water interfacial tensions were compared against experimental data, as shown in Supporting Information **Table S** 5.

A 10 Å cutoff radius was applied for the Lennard-Jones (LJ) potential truncation in inorganic species. For the organic molecules, internal (8.5 Å) and external (10 Å) cutoff radii have been ascribed for the CHARMM energy switching function. The LJ parameters for two different non-bonded atoms were obtained by the Lorentz-Berthelot combination rule. Long-range electrostatic interactions were solved within the PPPM method[47,48]. The Velocity-Verlet algorithm[49], with a 0.5 fs time step, was employed to integrate the motion equations. MD simulations were performed under periodic boundary conditions over all

directions.

The Nosé-Hoover[50,51] thermostat and the Parrinelo-Rahman[52] barostat with the corrections proposed by Martyna et al.[53] were employed for the temperature and the pressure control, respectively. Since this is an exploratory work, we considered the room pressure and temperature (1 atm and 300 K) to obtain a consistent ML dataset through MD simulations. Thus the interface properties can be compared against the existing literature. Besides this work focus on ambient conditions, namely, an extrapolation to typical oil reservoir temperatures and pressures can be performed. The employed potentials have already been tested at reservoir conditions, with good agreement with experimental data.[6,54]

To reach the thermodynamic equilibrium state, we performed a 10.0 ps in the NVT, followed by a 3.0 ns MD run in the NPT ensemble. We have observed the thermodynamic quantities (temperature, pressure, volume, and energy), and the interfacial tension values were converged (see SI **Figure S 1**). The production phase runs over 5.0 ns under the NPT ensemble, where 2000 snapshots were collected for the IFT calculations. Therefore standard errors in the IFT are always lower than 0.2 mN/m for all the simulated cases.

We calculate the IFT through the equation proposed by Kirkwood and Buff[55]

$$\gamma = \int_0^{\mathrm{L}} \Big(P_{\mathrm{N}}(z) - P_{\mathrm{T}}(z) \Big) \mathrm{d}z, \tag{1}$$

where P_{N} and P_{T} are the normal and tangential pressures, respectively. Here, L is the integration interval in the z-axis, the perpendicular direction to the oil/brine interface (see **Figure 1**a). The normal (P_{N}) and tangential (P_{T}) pressures come from the per-atom stress tensor[48,56]. The z-axis was divided into 1 Å bins to integrate the normal and tangential components for every atom in that region. Only the diagonal stress per-atom tensor terms were considered in the IFT calculations, *i.e.*, $P_{\mathrm{T}} = 0.5(p_{\mathrm{xx}} + p_{\mathrm{yy}})$ and $P_{\mathrm{N}} = p_{\mathrm{zz}}$.

Machine Learning

The obtained IFTs fed the ML dataset with 130 entries. The oil fraction (P, N, and A) and density[57]; brine salinity and their ionic composition (Na^+, Ca^{2+}, Mg^{2+}, Cl^-, K^+, and SO_4^{2-}) are the feature space. We used the Scikit-learn package[58] considering five estimators, *i.e.* (I) Linear Regression (LR), (II) Random Forest (RF), (III) Extra Trees (ET), (IV) Gradient Boosted (GB), and the (V) ElasticNet Regression (EN).

The linear fitting of an equation

$$h_\theta = \theta_0 + \sum_{i=1}^{n} \theta_i X \tag{2}$$

to a dataset by a regression method is given by finding the θ parameters that minimize the error function $J(\theta)$:

$$J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (h_\theta^{(j)} - y^{(j)})^2 \tag{3}$$

where X is a $n \times m$ data matrix with n and m being the number of features and entries, respectively; $h_\theta^{(j)}$ is the linear equation result for the j-th X entry; $y^{(j)}$ is j-th known result from the dataset; θ_0 and θ_i are the intercept and slope coefficients. For the LR method, the Scikit-Learn implements the ordinary least squares.

In addition to the LR, the EN approach utilizes a second (*L1*) and third (*L2*) error regularization terms:

$$J_{EN}(\theta) = J(\theta) + L1 + L2 \tag{4}$$

given by

$$L1 = \alpha \textstyle\sum_{j=1}^{n} \theta_j, \tag{5}$$

$$L2 = \tfrac{\alpha}{2}(1-\alpha) \textstyle\sum_{j=1}^{m} \theta_j^2, \tag{6}$$

where α is the regularization parameter.

Besides regressions, we employ decision tree techniques to obtain the feature importance rank[59]. This kind of methodology splits the input data into groups called nodes. Each node is derived from a parent (except the root node), and it may originate two nodes. If the node does not have any children when a particular condition is met, this is called termination node or leaf.

Among the decision tree algorithms, the Bagging approach utilizes "weak" trees to ensemble a robust final forest by adding a base function to each interaction[60]. In addition to the Bagging approach, the Random Forests technique uses random sub-samples to reduce overfitting[61]

For each node, the algorithms calculate an error or an information gain. The information gain is given by the Gini index or cross-entropy, while the error is calculated using the the Mean Absolute Error (MAE) given by:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{m} \left(y - y_i\right)}{m}; \tag{7}$$

and the Root Mean Square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m} \left(y - y_i\right)^2}{m}}; \tag{8}$$

where y and y_i are predicted and observed values, respectively. The total number m of dataset entries was divided into training (75%) and test (25%) sets by using randomly selected data.

The prediction score was evaluated by the r^2 parameter, which correlates the residual (S_{res}) and the total sum of squares (S_{tot}) as given by:

$$r^2 = 1 - \frac{S_{\mathrm{res}}}{S_{\mathrm{tot}}}, \tag{9}$$

where r^2 ranges from 0 (none data was predicted) to 1 (all data was correctly predicted).

Results

Figure 2 shows the calculated IFTs distribution and summarizes the values in **Tables S** 6-7. As revealed by the Pearson correlation analysis (see **Figure S** 2 in SI), the oil/brine IFT displays a strong linear correlation and opposite trend with paraffinic (P) and aromatic (A) fractions. For instance, toluene (100% A and 0% P) and heptane (0% A and 100% P) have the lowest and highest IFT values, respectively. Such a result may be a consequence of the "weak hydrogen bond" established between the aromatic rings and the water hydrogen atom, influencing the free energy at the oil/brine interface[36].

Four ML algorithms were employed to get more insights into the IFT trends, in addition to the LR algorithm. The corresponding MAE and MRSE errors obtained for the test set were compared against the Naive baseline, which was calculated with the equations (7) and (8), but considering the whole database (test plus training set). As shown in **Figure 3** a) and b), the ML approaches produce similar errors and lower than the reference value in both metrics. Also, the ML models applied to our dataset display a bias-variance trade-off, as shown in the Supporting Information **Figure S** 3.

All the employed ML methods had similar errors. The linear regression (LR) method showed larger errors, besides, they are similar to the ML methods. One can see that the ML algorithms could be very well applied to describe the IFT database, although, the gain is not very large when comparing to the much more convenient LR model. Thus, summarizing the model in a linear equation is interesting both for easy prediction and for reproducibility. The obtained LR model is as follows:

$$
\begin{aligned}
IFT = {} & 33.4689 \times \tfrac{mN}{m} + 0.0680 \times \tfrac{mN}{m} \times \mathrm{P} - 0.0068 \times \tfrac{mN}{m} \times \mathrm{N} - 0.0612 \times \tfrac{mN}{m} \times \mathrm{A} \\
& + 14.7967 \times \tfrac{nN \times m^2}{g} \times \mathrm{DensityOil} + 0.1139 \times \tfrac{dN}{m \times wt\%} \times \mathrm{Salinity} - 376.3637 \times \tfrac{dN}{m \times wt\%} \times \mathrm{C_{Cl^-}} \\
& + 672.3570 \times \tfrac{dN}{m \times wt\%} \times \mathrm{C_{Na^+}} + 1240.1167 \times \tfrac{dN}{m \times wt\%} \times \mathrm{C_{Mg^{2+}}} + 739.2522 \times \tfrac{dN}{m \times wt\%} \times \mathrm{C_{Ca^{2+}}} \\
& - 297.4796 \times \tfrac{dN}{m \times wt\%} \times \mathrm{C_{SO_4^{2-}}} + 380.6060 \times \tfrac{dN}{m \times wt\%} \times \mathrm{C_{K^+}} \qquad (10)
\end{aligned}
$$

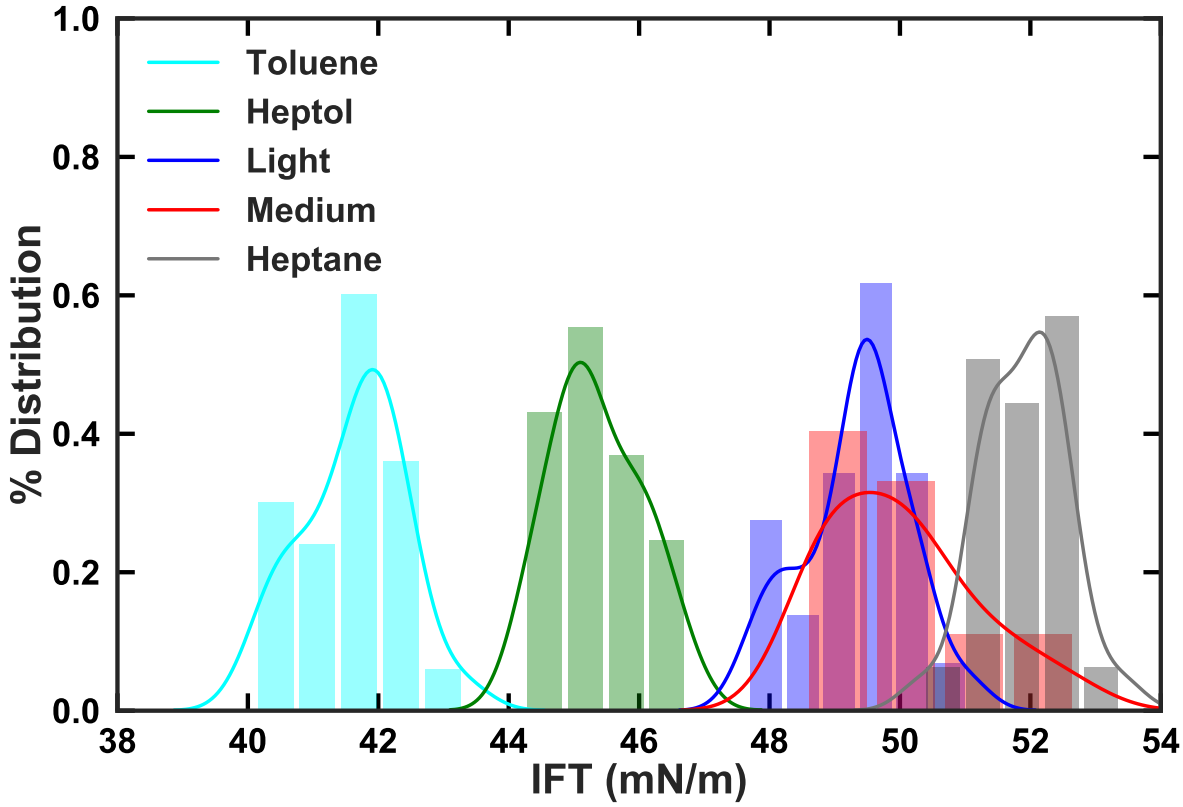


Figure 2: IFT distribution obtained for each oil model considering all brine compositions and concentrations. The solid colored lines are just a guide to the eye.

where the IFT is given in mN/m. P, N and A are the fraction of each of those componentes in the oil (from 0 to 1), DensityOil is the oil density in g/cm^3, Salinity is the total brine salinity in wt%/100. The other terms are ionic concentration in wt%/100, as follows: C_{Cl^-} - chlorine, $C_{SO_4^{2-}}$ - sulfate, C_{Na^+} - sodium, C_{K^+} - potassium, $C_{Mg^{2+}}$ - magnesium and $C_{Ca^{2+}}$ - calcium.

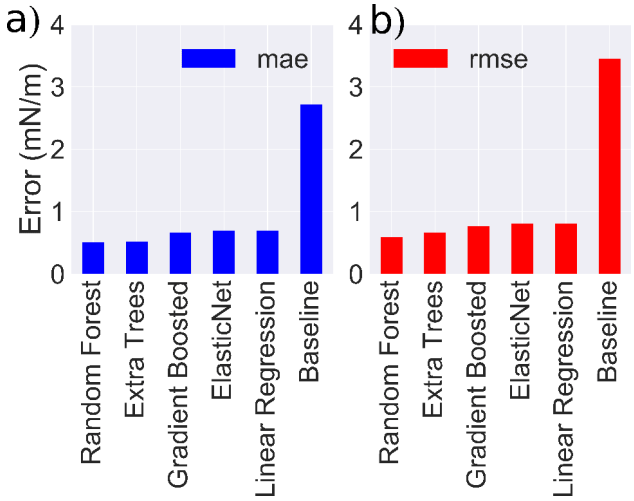


Figure 3: a) Mean Absolute Error (MAE) and b) Root Mean Square Error (RMSE) obtained for each estimator when applied to the test dataset. These values are lower than the Naive baseline.

The linear model was validated on the molecular dynamics simulations results from Remesal et al.[21], and experimental data from Kakati and Sangwai[62]. Remesal et al.[21] considered three types of oil and two saline solutions ($CaCl_2$ and NaCl) with five different concentrations (1.0, 2.0, 5.0, 11.0, and 22.0 wt%). Kakati and Sangwai[62] have measured the interfacial tension between several kinds of oil, and brines with salinities ranging from 0.05 to 3 wt%. It is important to stress that the reported works hold different salinities and oil models than the ones used for the LR model. The LR model could predict the reported IFTs with an average error of 3% for the Remesal et al.[21] data and 9% for the Kakati and Sangwai[62] data. The disagreement with the experimental data comes mainly because of an molecular dynamics overestimation of the toluene/brine interfacial tensions, thus, the trained model reproduces that. This difference may occur because the CGenFF potential is known to underestimate the toluene dipole moment[63], affecting the interfacial interactions.

Among the classification methods, we have employed the gradient boosted (GB) machine learning algorithm to analyze the feature importance of the proposed model. The main feature contributions are related to the oil phase characteristics, *i.e.*, the PNA fraction and oil density, in addition to the brine salinity, as shown in **Table 1**. The feature importance agrees with the Pearson correlation coefficients (shown in **Figure S 2**), where the PNA fraction and the oil density have the larger correlations with the IFT. The Cl^- has a larger correlation with the IFT than the salinity, not reflecting the GB result. Despite that, the Pearson correlation coefficient tells how linear is the correlation and says nothing about the angular coefficient. Thus, the correlation could be linear, but with a very low angular coefficient, that would lead to low feature importance.

Based on the GB results, we select the five most important features to simplify the LR model. The LR estimator considering just four features, provided a slightly lower score (0.93) in comparison with the full model (0.97) when applied to the data set reported by Remesal et al.[21]. Thus, we maintain all the feature set and do not disregard any feature to improve the model transferability.

Table 1: Feature importance rank calculated using the Gradient Boosted machine learning method.

Feature	**Value**
Paraffinic	0.716
Naphthenic	0.204
DensityOil	0.041
Aromatic	0.016
Salinity	0.012
Na^+	0.008
K^+	0.001
Ca^{2+}	0.001
Mg^{2+}	0.000
Cl^-	0.000
SO_4^{2-}	0.000

The simplified feature space (P, N, A, and Density oil) was tested with the Remesal et al.[21] and Kakati and Sangwai[62] datasets, the average error for the Remesal et al.[21] data

was 2%, and for the Kakati and Sangwai[62] data, the error was 9%, again, with the larger errors for the toluene. In **Figure 4** the model comparison with the literature data, is shown. As one can see, the model has a good transferability, apart from the toluene IFTs, that are on the lower left of the graph. It is important to note that the feature importance regarding the salt composition is very low, just the Na^+ that seems to have some relevancy. Thus, we cannot say that any kind of ion is dominant regarding the IFT changes.

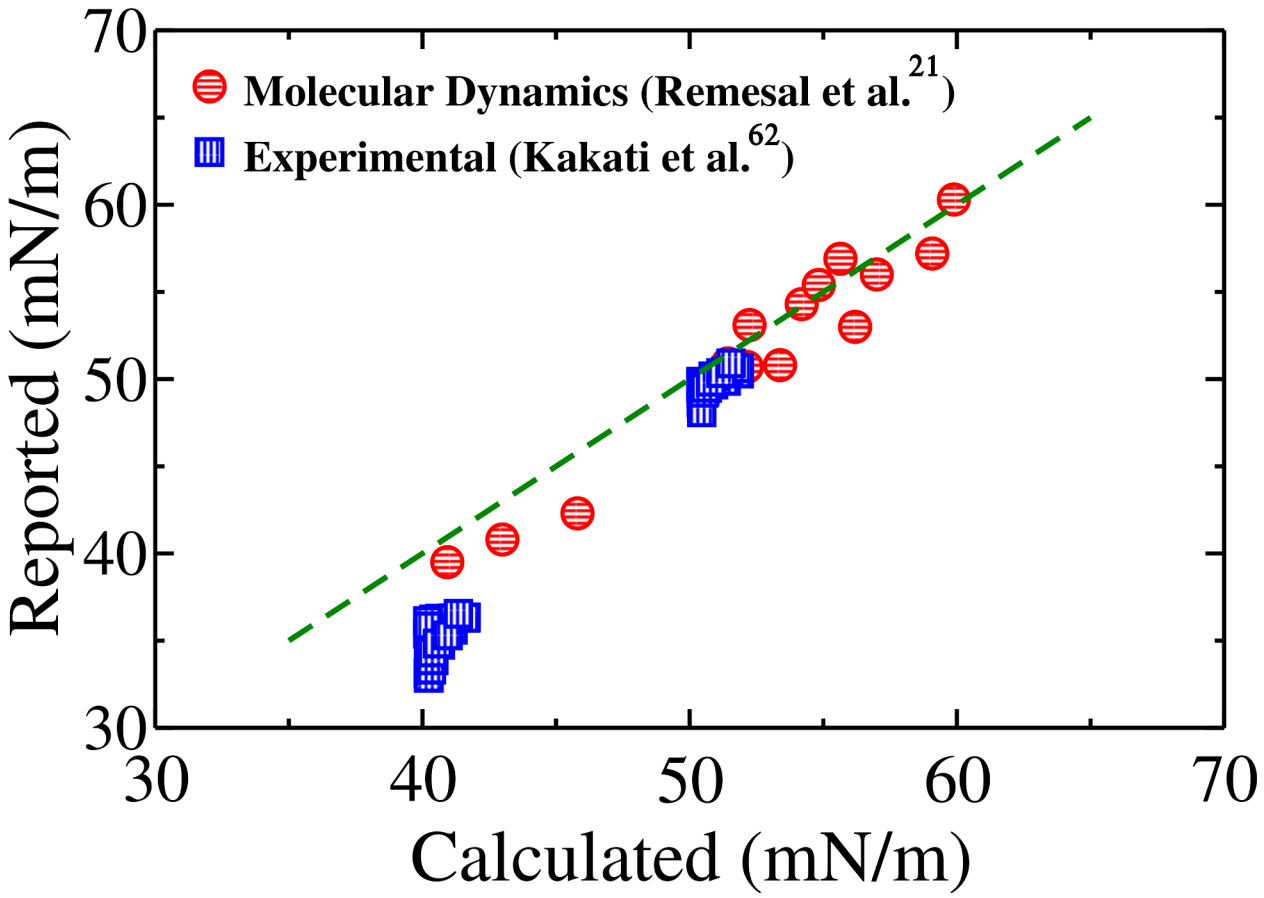


Figure 4: Comparison of the predicted IFT values obtained with our trained model against the ones reported by Remesal et al.[21] and Kakati and Sangwai[62]. Red circles are molecular dynamics data from Remesal et al.[21], and blue squares are experimental data from Kakati and Sangwai[62]. The dashed line represents $x = y$.

Conclusions

We have used classical molecular dynamics simulations to create a consistent oil/brine interfacial tensions (IFT) dataset for 300 K temperature and 1 atm pressure. Besides, we have proposed a new atomistic oil model, with 36 different types of hydrocarbon molecules, to diversify the database. With the obtained database, we have employed four different machine learning (ML) algorithms and compared them to a simple linear regression (LR) model. The ML models performed better, although, it was not by a large margin. For convenience and reproducibility, we have used the LR for further predictions. The model has been validated with literature data from both simulations and experiments, where average errors from 2% to 9% were observed. The feature importance rank obtained by the gradient boosted method revealed the oil fractions and its density were the most important features influencing the oil/brine IFT. Interestingly, as the oil features dominated the IFT changes, thus, concealing the influence from the ionic composition. Apart from the oil properties, only the total brine salinity had a significant contribution.

In summary, the generation of consistent IFT databases from molecular simulations can be an important tool to address the oil industry challenges. With organized available data, one can apply various models to further predict non-simulated situations with good transferability. Those databases have the potential to be expanded by considering the reservoir thermal conditions, a wider range of brine compositions, and more complex interfaces.

Acknowledgement

The authors acknowledge Petrobras for the funding. Also, thanks to the Brazilian agencies FAPESP, CNPq, and CAPES for the financial support and CENAPAD-SP, USP, and UFABC for the computing time. This research was developed with the support of HPC resources provided by the Superintendencia de Tecnologia da Informacao da Universidade de São Paulo.

Supporting Information Available

The following files are available free of charge.

- Supporing.pdf:

References

(1) Alagorni, A. H.; Yaacob, Z. B.; Nour, A. H. An Overview of Oil Production Stages: Enhanced Oil Recovery Techniques and Nitrogen Injection. *International Journal of Environmental Science and Development* **2015**, *6*, 693–701.

(2) Muggeridge, A.; Cockin, A.; Webb, K.; Frampton, H.; Collins, I.; Moulds, T.; Salino, P. Recovery Rates, Enhanced Oil Recovery and Technological Limits. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2014**, *372*, 20120320.

(3) Raffa, P.; Broekhuis, A. A.; Picchioni, F. Polymeric Surfactants for Enhanced Oil Recovery: a Review. *Journal of Petroleum Science and Engineering* **2016**, *145*, 723–733.

(4) da Silva, G. D.; de Freitas Martins, E.; Salvador, M. A.; Baptista, A. D. T.; de Almeida, J. M.; Miranda, C. R. From Atoms to Pre-salt Reservoirs: Multiscale Simulations of the Low-Salinity Enhanced Oil Recovery Mechanisms. *Polytechnica* **2019**, *2*, 30–50.

(5) Abdalla, A.; Sun, R.; Guiyun, T.; Huang, A.; Wang, M. Low Salinity as New Technique of Enhanced Oil Recovery. *International Journal of Chemical Engineering and Applications* **2017**, *8*, 117.

(6) de Lara, L. S.; Michelon, M. F.; Miranda, C. R. Molecular Dynamics Studies of Fluid/Oil Interfaces for Improved Oil Recovery Processes. *The Journal of Physical Chemistry B* **2012**, *116*, 14667–14676.

(7) Wu, Y.; Shuler, P. J.; Blanco, M.; Tang, Y.; Goddard, W. A. An Experimental Study of Wetting Behavior and Surfactant EOR in Carbonates With Model Compounds. *SPE Journal* **2008**, *13*, 26–34.

(8) Rotondi, M.; Callegaro, C.; Masserano, F.; Bartosek, M. Low Salinity Water Injection: eni's Experience. Abu Dhabi International Petroleum Exhibition and Conference. 2014.

(9) Tang, G.-q.; Morrow, N. R. Oil Recovery by Waterflooding and Imbibition-Invading Brine Cation Valency and Salinity. *Paper SCA9911* **1999**,

(10) McGuire, P.; Chatham, J.; Paskvan, F.; Sommer, D.; Carini, Low Salinity Oil Recovery: an Exciting New EOR Opportunity for Alaska's North Slope. SPE Western Regional Meeting. 2005.

(11) Lager, A.; Webb, K. J.; Black, C.; Singleton, M.; Sorbie, K. S. Low Salinity Oil Recovery: an Experimental Investigation. *Petrophysics* **2008**, *49*.

(12) RezaeiDoust, A.; Puntervold, T.; Strand, S.; Austad, T. Smart Water as Wettability Modifier in Carbonate and Sandstone: a Discussion of Similarities/Differences in the Chemical Mechanisms. *Energy & Fuels* **2009**, *23*, 4479–4485.

(13) Pu, H.; Xie, X.; Yin, P.; Morrow, N. R. Low-Salinity Waterflooding and Mineral Dissolution. SPE Annual Technical Conference and Exhibition. 2010.

(14) Austad, T.; RezaeiDoust, A.; Puntervold, T. Chemical Mechanism of Low Salinity Waterflooding in Sandstone Reservoirs. SPE Improved Oil Recovery Symposium. 2010.

(15) Emadi, A.; Sohrabi, M. Visual Investigation of Oil Recovery by Low Salinity Water Injection: Formation of Water Micro-Dispersions and Wettability Alteration. SPE Annual Technical Conference and Exhibition. 2013.

(16) da Silva, G. D.; de Freitas Martins, E.; Salvador, M. A.; Baptista, A. D. T.; de Almeida, J. M.; Miranda, C. R. From Atoms to Pre-salt Reservoirs: Multiscale

Simulations of the Low-Salinity Enhanced Oil Recovery Mechanisms. *Polytechnica* **2019**, *2*, 30–50.

(17) Lashkarbolooki, M.; Ayatollahi, S.; Riazi, M. Effect of Salinity, Resin, and Asphaltene on the Surface Properties of Acidic Crude Oil/Smart Water/Rock System. *Energy & Fuels* **2014**, *28*, 6820–6829.

(18) Buckley, J. S.; Fan, T. Crude Oil/Brine Interfacial Tensions. *Petrophysics* **2007**, *48*.

(19) Manshad, A. K.; Olad, M.; Taghipour, S. A.; Nowrouzi, I.; Mohammadi, A. H. Effects of Water Soluble Ions on Interfacial Tension (IFT) Between Oil and Brine in Smart and Carbonated Smart Water Injection Process in Oil Reservoirs. *Journal of Molecular Liquids* **2016**, *223*, 987–993.

(20) Jian, C.; Poopari, M. R.; Liu, Q.; Zerpa, N.; Zeng, H.; Tang, T. Reduction of Water/Oil Interfacial Tension by Model Asphaltenes: the Governing Role of Surface Concentration. *The Journal of Physical Chemistry B* **2016**, *120*, 5646–5654.

(21) Remesal, E. R.; Suárez, J. A.; Márquez, A. M.; Sanz, J. F.; Rincón, C.; Guitián, J. Molecular Dynamics Simulations of the Role of Salinity and Temperature on the Hydrocarbon/Water Interfacial Tension. *Theoretical Chemistry Accounts* **2017**, *136*, 66.

(22) Underwood, T. R.; Greenwell, H. C. The Water-Alkane Interface at Various Nacl Salt Concentrations: a Molecular Dynamics Study of the Readily Available Force Fields. *Scientific Reports* **2018**, *8*, 352.

(23) Sedghi, M.; Piri, M.; Goual, L. Atomistic Molecular Dynamics Simulations of Crude Oil/Brine Displacement in Calcite Mesopores. *Langmuir* **2016**, *32*, 3375–3384.

(24) Van Buuren, A. R.; Marrink, S. J.; Berendsen, H. J. A Molecular Dynamics Study of the Decane/Water Interface. *The Journal of Physical Chemistry* **1993**, *97*, 9206–9212.

(25) De Almeida, J. M.; Miranda, C. R. Improved Oil Recovery in Nanopores: NanoIOR. *Scientific Reports* **2016**, *6*, 28128.

(26) Smith, D. E.; Dang, L. X. Computer Simulations of NaCl Association in Polarizable Water. *The Journal of Chemical Physics* **1994**, *100*, 3757–3766.

(27) de Lara, L. S.; Rigo, V. A.; Miranda, C. R. Functionalized Silica Nanoparticles Within Multicomponent Oil/Brine Interfaces: a Study in Molecular Dynamics. *The Journal of Physical Chemistry C* **2016**, *120*, 6787–6795.

(28) Riniker, S. Molecular Dynamics Fingerprints (MDFP): Machine Learning from MD Data to Predict Free-Energy Differences. *Journal of Chemical Information and Modeling* **2017**, *57*, 726–741.

(29) Wang, J.; Ferguson, A. Nonlinear Machine Learning in Simulations of Soft and Biological Materials. *Molecular Simulation* **2018**, *44*, 1090–1107.

(30) Jamal, S.; Goyal, S.; Shanker, A.; Grover, A. Machine Learning and Molecular Dynamics Based Insights Into Mode of Actions of Insulin Degrading Enzyme Modulators. *Combinatorial Chemistry & High Throughput Screening* **2017**, *20*, 279–291.

(31) Glazer, D. S.; Radmer, R. J.; Altman, R. B. *Biocomputing 2008*; World Scientific, 2008; pp 332–343.

(32) Pedersen, K. S.; Christensen, P. L.; Shaikh, J. A.; Christensen, P. L. *Phase Behavior of Petroleum Reservoir Fluids*; CRC press, 2006.

(33) Rivera, J. L.; McCabe, C.; Cummings, P. T. Molecular Simulations of Liquid-Liquid Interfacial Properties: Water—n-Alkane and Water-Methanol—n-Alkane Systems. *Physical Review E* **2003**, *67*, 011603.

(34) Kashefi, K.; Pereira, L. M.; Chapoy, A.; Burgass, R.; Tohidi, B. Measurement and

Modelling of Interfacial Tension in Methane/Water and Methane/Brine Systems at Reservoir Conditions. *Fluid Phase Equilibria* **2016**, *409*, 301–311.

(35) Herdes, C.; Ervik, Å.; Mejía, A.; Müller, E. A. Prediction of the Water/Oil Interfacial Tension from Molecular Simulations Using the Coarse-Grained SAFT-γ Mie Force Field. *Fluid Phase Equilibria* **2018**, *476*, 9–15.

(36) Kunieda, M.; Nakaoka, K.; Liang, Y.; Miranda, C. R.; Ueda, A.; Takahashi, S.; Okabe, H.; Matsuoka, T. Self-Accumulation of Aromatics at the Oil-Water Interface Through Weak Hydrogen Bonding. *Journal of the American Chemical Society* **2010**, *132*, 18281–18286.

(37) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: a Package for Building Initial Configurations for Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2009**, *30*, 2157–2164.

(38) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **1995**, *117*, 1–19.

(39) Brown, W. M.; Wang, P.; Plimpton, S. J.; Tharrington, A. N. Implementing Molecular Dynamics on Hybrid High Performance Computers–Short Range Forces. *Computer Physics Communications* **2011**, *182*, 898–911.

(40) Brown, W. M.; Kohlmeyer, A.; Plimpton, S. J.; Tharrington, A. N. Implementing Molecular Dynamics on Hybrid High Performance Computers–Particle–Particle Particle-Mesh. *Computer Physics Communications* **2012**, *183*, 449–459.

(41) Brown, W. M.; Yamada, M. Implementing Molecular Dynamics on Hybrid High Performance Computers-Three-Body Potentials. *Computer Physics Communications* **2013**, *184*, 2785–2793.

(42) Alejandre, J.; Chapela, G. A.; Bresme, F.; Hansen, J.-P. The Short Range Anion-H Interaction is the Driving Force for Crystal Formation of Ions in Water. *The Journal of Chemical Physics* **2009**, *130*, 174505.

(43) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **2010**, *31*, 671–690.

(44) Gavryushov, S. Dielectric Saturation of the Ion Hydration Shell and Interaction between Two Double Helices of DNA in Mono-and Multivalent Electrolyte Solutions: Foundations of the ε-Modified Poisson-Boltzmann Theory. *The Journal of Physical Chemistry B* **2007**, *111*, 5264–5276.

(45) Lee, S.-J.; Song, Y.; Baker, N. A. Molecular Dynamics Simulations of Asymmetric NaCl and KCl Solutions Separated by Phosphatidylcholine Bilayers: Potential Drops and Structural Changes Induced by Strong Na^+-Lipid Interactions and Finite Size Effects. *Biophysical Journal* **2008**, *94*, 3565–3576.

(46) Williams, C. D.; Carbone, P. A Classical Force Field for Tetrahedral Oxyanions Developed Using Hydration Properties: the Examples of Pertechnetate (TcO^{4-}) and Sulfate (SO_4^{2-}). *The Journal of Chemical Physics* **2015**, *143*, 174502.

(47) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; CRC Press, 1988.

(48) Thompson, A. P.; Plimpton, S. J.; Mattson, W. General Formulation of Pressure and Stress Tensor for Arbitrary Many-Body Interaction Potentials Under Periodic Boundary Conditions. *The Journal of Chemical Physics* **2009**, *131*, 154107.

(49) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford university press, 2017.

(50) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *The Journal of Chemical Physics* **1984**, *81*, 511–519.

(51) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Physical Review A* **1985**, *31*, 1695.

(52) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(53) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *The Journal of Chemical Physics* **1994**, *101*, 4177–4189.

(54) de Lara, L. S.; Wrobel, E. C.; Lazzarotto, M.; de Lázaro, S. R.; Camilo, A.; Wohnrath, K. An experimental and theoretical study of the aggregate structure of calix [6] arenes in Langmuir films at the water/air interface. *Physical Chemistry Chemical Physics* **2016**, *18*, 22906–22913.

(55) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Surface Tension. *The Journal of Chemical Physics* **1949**, *17*, 338–343.

(56) Heyes, D. M. Pressure Tensor of Partial-Charge and Point-Dipole Lattices with Bulk and Surface Geometries. *Physical Review B* **1994**, *49*, 755.

(57) Haynes, W. M. *CRC Handbook of Chemistry and Physics*; CRC press, 2014.

(58) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

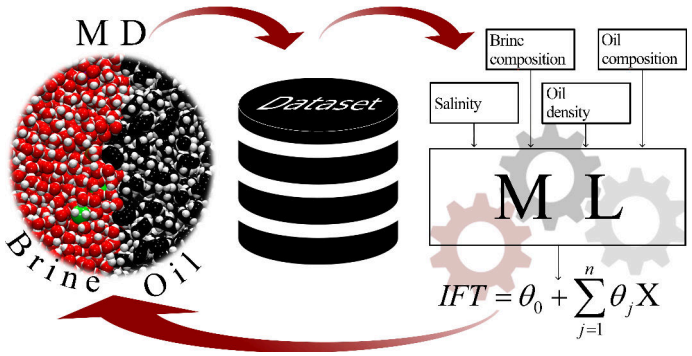
(59) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Science & Business Media, 2009.

(60) Breiman, L. Bagging Predictors. *Machine Learning* **1996**, *24*, 123–140.

(61) Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32.

(62) Kakati, A.; Sangwai, J. S. Effect of Monovalent and Divalent Salts on the Interfacial Tension of Pure Hydrocarbon-Brine Systems Relevant for Low Salinity Water Flooding. *Journal of Petroleum Science and Engineering* **2017**, *157*, 1106–1114.

(63) Lopes, P. E.; Lamoureux, G.; Roux, B.; MacKerell, A. D. Polarizable empirical force field for aromatic compounds based on the classical drude oscillator. *The Journal of Physical Chemistry B* **2007**, *111*, 2873–2885.



TOC Graphics