

Predicting the properties of bitumen using machine learning models trained with force field atom types and molecular dynamics simulations

Assaf, Eli I.; Liu, Xueyan; Lin, Peng; Ren, Shisong; Erkens, Sandra

DOI

[10.1016/j.matdes.2024.113327](https://doi.org/10.1016/j.matdes.2024.113327)

Publication date

2024

Document Version

Final published version

Published in

Materials and Design

Citation (APA)

Assaf, E. I., Liu, X., Lin, P., Ren, S., & Erkens, S. (2024). Predicting the properties of bitumen using machine learning models trained with force field atom types and molecular dynamics simulations. *Materials and Design*, 246, Article 113327. <https://doi.org/10.1016/j.matdes.2024.113327>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Predicting the properties of bitumen using machine learning models trained with force field atom types and molecular dynamics simulations

Eli I. Assaf^{a,*}, Xueyan Liu^a, Peng Lin^b, Shisong Ren^a, Sandra Erkens^{a,b}

^a Delft University of Technology, Delft, the Netherlands

^b Ministry of Infrastructure and Water Management (Rijkswaterstaat), the Netherlands

ARTICLE INFO

Keywords:

Molecular Dynamics
Machine Learning
Bitumen Design
Chemical Descriptors

ABSTRACT

This study enhances the molecular analysis of bitumen by transitioning from traditional chemical descriptors, such as SARA (Saturates, Aromatics, Resins, and Asphaltenes) fractions and elemental compositions, to specific force field atom types in Molecular Dynamics (MD) models. This shift improves the precision in predicting material properties critical for bituminous material characterization. Machine Learning Models (MLMs) were developed to use these atom types as input features, inherently reflecting fundamental chemical characteristics. Trained on data from over 1,770 LAMMPS simulations of diverse bitumen types and conditions, these MLMs enable the prediction of properties like density, heat capacity, solubility parameters, and thermal expansion coefficients without the need for additional MD simulations. The models utilize 30 chemical descriptors corresponding to specific atom types in the PCFF force field, which collectively account for over 95% of the influence on these properties. By accurately predicting fundamental, thermodynamic, and kinetic properties, the use of MLMs and force field atom types allows researchers to efficiently tweak the chemical nature of organic molecules and mixtures to achieve desired properties. With near-instantaneous prediction times, these MLMs offer valuable insights for advancing bitumen research in the construction and petroleum industries, reducing the need for more intensive simulation techniques.

1. Introduction

Bitumen consists of a complex mixture of high molecular mass hydrocarbons, typically derived from the fractional distillation of petroleum at the bottom of distillation columns. Annually, over 100 million tons are produced, highlighting its industrial significance. Currently, bitumen attracts considerable research attention due to the demand for more sustainable and durable binders. Specifically, the development of bitumens and other heavy oils that remain manageable at lower temperatures is crucial for the oil, gas, and construction industry [1]. Traditionally, research on bituminous materials has been predominantly experimental. However, the need for more robust and fundamental characterization and design techniques has driven scientists toward molecular simulations [2,3]. Nonetheless, employing molecular methods to study bitumen presents its own set of challenges [4].

The molecular structure of bitumen comprises thousands of highly variable hydrocarbon molecules—predominantly aromatic, resinous, or asphaltenic—making bitumen a difficult material to study and characterize [2]. This in turn complicates the establishment of a consistent

molecular structure necessary for simulations [5]. Moreover, molecular simulations are limited to ensembles no more than a few hundred molecules in size and no more than a few nanoseconds in length due to their high computational cost [6]. This limits the use of molecular modelling techniques to explore material properties that are fundamental in nature, but quickly become complicated when trying to capture mechanical and rheological properties of relevance to large scale applications akin to those used by Engineers [7]. Nevertheless, the use of MD has been increasingly popular, as these fundamental properties can still yield tendencies that aid in the characterization and evaluation of newer bitumens without the need of extensive experimentation techniques [8].

Multiple efforts have been made to chemically characterize bitumen and establish correlations between its compositional elements and its physical properties [9]. These efforts include the fractionation of molecule groups based on solubility (such as SARA fractionation), performing direct elemental analysis, and employing gel permeation chromatography (GPC-MS) techniques to assess molecular masses and their distributions. These methods are often used in combination to

* Corresponding author.

E-mail address: e.i.assaf@tudelft.nl (E.I. Assaf).

<https://doi.org/10.1016/j.matdes.2024.113327>

Received 28 May 2024; Received in revised form 12 September 2024; Accepted 16 September 2024

Available online 19 September 2024

0264-1275/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

achieve a more fundamental characterization and differentiation of bitumen samples. Nonetheless, some of these methodologies, despite their prevalent use in the bitumen industry, are noted for their lack of reproducibility, accuracy, or are simply too generalized to effectively capture the intricate chemical nature of bitumens as required for detailed molecular simulations. For instance, the Resinous and Asphaltenic fractions often contain molecules that are highly aromatic, sometimes more so than those in the Aromatics fraction, which renders the definition of the Aromatic fraction in SARA fractionation techniques somewhat ambiguous when applied to strict chemical applications [10].

Nonetheless, research studies have been conducted to link the chemical characteristics of bitumen with its measured properties, such as the impact of SARA fractions on the heat capacity of bitumen, aiming to better understand its behavior in relation to compositional differences. These studies often struggle to identify relevant or clear trends, largely due to the ambiguous nature of bitumen characterization. This challenge is amplified by the fact that molecular models are highly sensitive to the chemical structures of bituminous molecules and their complex interactions [11]. Additionally, the high variability of bituminous materials, along with the increasing need to evaluate a wider range of types—such as recycled, synthetic, and modified bitumens—further complicates the identification of behavioral trends [12]. As a result, research has entered a continuous cycle of compositional testing to identify and correlate chemical properties with more general features using time- and resource-intensive simulations. Their use is necessary, as lighter models, such as those using empirical correlations like those used in Equations of State [13], fluid packages, or QSAR/QSPR methods [14], are unable to handle the complexity of bituminous molecules [15,16].

There is a significant need to move beyond the parameters obtained from conventional bitumen characterization tests to enhance the way fundamental chemical features influence the trends and insights derived from molecular-level simulations. Additionally, there is a growing requirement to reduce the reliance on complex molecular simulations for evaluating changes in material composition, as well as other properties such as temperature or pressure, to obtain fundamental properties through computational techniques. Furthermore, there is a strong need to rapidly assess the impact of various properties on the compositional and conditional attributes of bitumen [17]. Tackling these needs should enable researchers working with bitumens to characterize, prepare, and study the chemomechanics of bituminous materials with minimized use of MD simulations, while still comprehending how fundamental chemical characteristics affect material properties.

In this paper, we address three specific challenges by implementing a series of methodological innovations. First, we utilize force field atom types to describe the fundamental chemical nature of the bituminous materials modeled, rather than relying on conventional parameters such as SARA fractions and elemental compositions. This approach aims to enhance the quality of the insights and trends derived from MD simulations. Second, we develop a series of Machine Learning Models (MLMs) that capture these fundamental chemical features and predict material properties, thus eliminating the need to rerun molecular simulations. Third, we examine how the properties predicted by the MLMs are affected by changes in the inputs, resulting in a series of insights that allow scientists to finely tune bituminous molecules based on the material properties required.

The outcome provides other scientists with a clear understanding of how fundamental chemical characteristics influence a series of physical properties bituminous materials, without the necessity of performing new simulations.

This manuscript begins by detailing the characteristics of the molecular modeling simulations used in this study. It covers the selection of molecules, sample selection, conditioning, and preparation, followed by a description of the simulation routines and the list of material properties measured during the simulations. The manuscript then transitions to a discussion of the machine learning methods employed to develop these the MLMs of this study. This section includes the selection of features (i.

e., control parameters such as force field atom types) and labels (i.e., physical properties to be predicted such as heat capacities), along with a detailed description of the chosen ML model and its parameters.

The results section presents all properties measured, both computationally and experimentally, and offers a comprehensive evaluation of the predictive capabilities of the MLMs on both observed and unobserved data. It also establishes clear trends between features and labels that are relevant of the study and characterization of bituminous materials.

The discussion section analyzes the results to elucidate the impact, from an engineering perspective, that the features have on the properties discussed and explores how this information can be used in the design of bituminous materials. Additionally, it assesses the effectiveness of the methods used in this study in enhancing bitumen design practices. The manuscript concludes by providing some demonstrative examples on how to utilize the MLMs to predict bitumen-related properties and by summarizing the key insights and trends identified throughout the study, and it outlines potential areas for improvement to be considered in future iterations of this research line.

2. Methodology

This section delineates the methodologies employed to fulfill the objectives of this research. Initially, the **2.1 Molecules Selection** and **2.2 Molecular Models Preparation** sections discuss the collection and characterization of various bitumen samples and their conversion into computational models using a predefined library of molecules. The **2.3 Simulations** section details the MD simulations performed. This section includes a description of the selected force field, input parameters, and foundational assumptions, along with the initialization process of models into their final MD configurations. The Simulations section concludes with an overview of the MD routines used to determine the material properties selected for the study. The **2.4 Machine Learning** section introduces the integration of machine learning techniques to construct models that predict the study's properties using input features such as force field atom types, molecular masses, and temperatures.

2.1. Molecules selection

The molecular set employed in this study to construct bitumen models includes 33 molecules, extending the original 12 molecules identified by Greenfield [18] based on the work of Shisong et al [19]. This expansion incorporates structural modifications that align with the SARA fractionation and elemental analysis, reflecting compositional changes observed in bitumen samples tested by our research team subjected to various aging conditions. Aimed at representing a comprehensive range of hydrocarbon classes typically found in bituminous materials—such as alkanes, alkenes, polycyclic aromatics, and non-aromatics—the collection includes functional groups like phenols, oxanes, pyridines, thiophenes, and sulfoxides. The selected molecules, characterized by high molecular masses ranging from 200 to 1000 g/mol, pronounced aromaticity, and significant planarity, mirror the complex molecular structures prevalent in bituminous materials.

The molecules in this study are systematically categorized by both their SARA fraction classification and their degree of aging (0 to 4, where 0 represents non-aged molecules and 4 signifies those that have undergone the most aging. Aging is a chemical process involving oxidation and sulfoxidation due to environmental exposure, crucially altering the properties of bitumen. This dual categorization, while introducing complexities into this study, aim to use the extensive research available on bitumen aging to ensure that the MD simulations accurately reflect known physical behaviors, especially those relevant to civil engineering applications [20].

Within the SARA categorization of molecules, the Saturates group comprises 2 molecules characterized by low molecular mass and minimal aging, typical of light oils and bitumen rejuvenators. The Aromatics

group contains 6 medium molecular mass molecules with at least 2 aromatic rings, functional groups like aldehydes and ketones through aging. The Resins group includes 15 modest mass hydrocarbons with heterocyclic functional groups such as oxanes, pyridines, and thiophenes, which undergo extensive oxidative changes, resulting in highly functionalized, amphiphilic molecules. Lastly, the Asphaltenes group consists of 10 high molecular mass, highly planar hydrocarbons that experience substantial aging, leading to increased polarity and influencing the microstructural morphology of bitumen [17]. The skeletal representation of the molecules, along with their chemical formula, molecular mass, and estimated density are presented in Table 1. Their corresponding SMILES notation [21] can be found in file */molecules_selection.docx* in the [Supplementary Information](#).

2.2. Molecular models Preparation

Four groups of molecular models are established for this research. The first group consists of 38 molecular models, each involving mixtures of a single molecule type to simulate a pure solution of a specific chemical compound, using the molecules listed in Table 1 (SARA) and Table 2 (Rejuvenators). The molecular systems in groups (2) through (4) are designed by selecting different molecules from Table 1 to replicate the composition of real bituminous blends, differentiated by four characteristics: bitumen source, aging condition, rejuvenator type, and rejuvenator dosage. To facilitate this, real bituminous samples are collected, conditioned, and characterized and subsequently, molecular models are constructed by combining molecules from Table 1 to reflect the properties observed in the characterized real bitumen samples. The details of this process are elaborated next.

2.2.1. Sample collection and Preparation

2.2.1.1. Bitumen sources. For this study, three bitumen sources, all with a 70/100 penetration grade, were selected: types “T”, “N”, and “F”. These samples were specifically chosen from a wider selection array based on their differing sulfur content by mass—type “N” at 0.8 %, “T” at 2.8 %, and “F” at 4.5 %. The differentiation in sulfur content is critical as it distinctly influences the properties of the bitumens, providing clear discretization among the samples even when bitumens with differing carbon, hydrogen, and oxygen contents may exhibit similar properties [22]. Additionally, the sulfur content in bitumens is gaining increased attention in research, especially as stricter regulations on sulfur levels in fuels are implemented. These regulations are driving the development of new bitumens with higher and more varied sulfur contents [23].

2.2.1.2. Aging conditions. Five aging conditions, categorized as 0 through 4, are selected to represent distinct stages in the life cycle of bitumen: condition 0 is fully fresh from its source (non-aged), condition (1) is short-term aged, similar to bitumen’s state when exposed to short but high thermal loads during road construction, and conditions (2), 3, and 4 simulate long-term aging, reflecting the gradual transformations that bitumen undergoes after roughly 3, 7, and 12 years of service life, respectively [24]. To prepare the aging conditions in the laboratory, non-aged samples are converted to their aged counterparts since obtaining real-life aged bitumen samples is remarkably challenging [25]. Level 1 aging is achieved by exposing 1 mm-thick fresh bitumen samples to 5 h in a Thin-film Oven at 160 °C, while Levels 2 through 4 are achieved by further aging Level 1 bitumens in a Pressurized Aging Vessel for 20, 40, and 80 h respectively at 100 °C and 2.1 MPa.

2.2.1.3. Rejuvenator types. Rejuvenators are chemical compounds that often mimic the molecular structure of specific SARA groups and are used as additives to modify the properties of aged bitumens. Although rejuvenation primarily involves a physical alteration rather than a direct chemical reversal of aging, these compounds are employed in an effort

to revert the effects of aging by restoring the changes in the SARA composition caused by aging to their original state when the bitumen was fresh [26].

For this study, four types of rejuvenators have been selected, each resembling the molecular structures found in four popular rejuvenator agents used in the bitumen industry: Vegetable Oil (V), Engine Oil (E), Naphthenic Oil (N), and Aromatic Oil (A) [27]. While rejuvenators typically consist of a mixture of many molecules—for instance, Engine Oil is known to contain thousands of different compounds—the approach here is to use a single molecule type that includes most (if not all) relevant functional groups present in these compounds. This simplification aids in better estimating their impact on the bitumen blend, as shown in multiple research studies involving the study of rejuvenators [27]. The skeletal representation of the rejuvenator molecules, along with their chemical formula, molecular mass, and estimated density are presented in Table 2. Their corresponding SMILES notation can be found in file */molecules_selection.docx* in the [Supplementary Information](#).

2.2.1.4. Rejuvenator dosages. In this study, four rejuvenator dosages are investigated: 0 %, 5 %, 10 %, and 15 % by mass, labelled as dosage 0, 1, 2, and 3 respectively. These percentages align with those typically used in practical applications and experimental research [26]. This dosage range is selected to demonstrate measurable differences in the properties of the bitumen, observable both experimentally and in MD simulations.

2.2.2. Model fitting

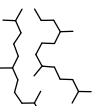
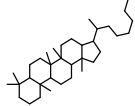
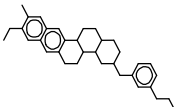
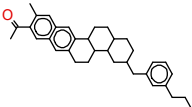
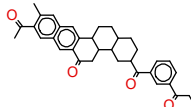
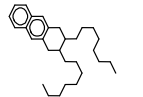
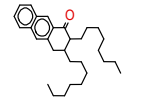
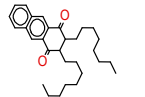
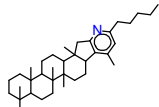
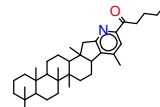
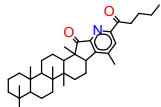
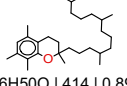
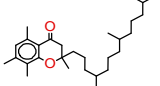
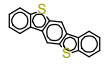
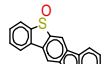
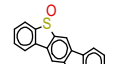
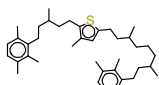
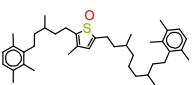
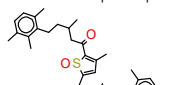
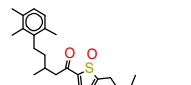
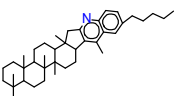
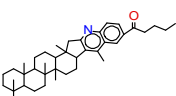
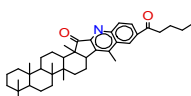
The molecular models, particularly those from Groups (2) through (4), are developed by fitting specific chemo-physical properties derived from the experimental characterization of the prepared samples (see the Sample Collection and Preparation section). These properties include average molecular mass of the mixture, estimated density by mass and number, dispersity, SARA composition by mass, elemental composition by Carbon, Hydrogen, Oxygen, Nitrogen, and Sulfur, Colloidal Index, and Saturation Degree. The properties are obtained through three distinct chemical characterization tests: SARA fractionation, Elemental Analysis, and GPC-MS. A summary of the characterization results obtained for Group (3) samples (“type N”) across all aging conditions is presented in Table 3.

Molecules from Table 1 and Table 2 are carefully selected and combined to create molecular mixtures that match the experimentally obtained chemo-physical properties. This matching is accomplished by using SciPy’s Minimize function [28] to optimize the selection and proportion of molecules to minimize the Mean Squared Error (MSE) between the experimental values and those computed from the virtual mixture. The fitting process involves the following steps:

- 1 Loading all molecules listed in Table 1 and Table 2 into a Python environment using the RDKit module [29] to retrieve the chemical descriptors (e.g., molecular mass) from their SMILES notation.
- 2 Computing the chemo-physical properties for each molecule, as per Table 3.
- 3 Importing the experimental data for each sample to serve as reference values.
- 4 Creating an initial array representing the target mixture of molecules, initially containing 3 molecules of each type.
- 5 Calculating the mixture-wide chemo-physical properties given the initial array of molecules.
- 6 Employing SciPy’s Minimize function to iteratively adjust the number of molecules in the mixture and minimize the MSE between the estimated mixture properties (y_i^{est}) and the reference values obtained using experimental characterization tests (y_i^{exp}). The MSE is calculated using Eq. (1), as follows:

Table 1

Skeletal representation of the molecules employed in constructing the molecular models for this study. Within the Saturates category, (1) squalane and (2) hopane are depicted. The Aromatics group includes (3) dioctylcyclohexane naphthalene and (4) perhydrophenanthrene naphthalene. In the Resins category, (5) quinolinohopane, (6) thioisorenieratane, (7) benzobisbenzothiophene, (8) pyridinohopane, and (9) trimethylbenzeneoxane are shown. Lastly, the Asphaltenes category comprises (10) phenolic asphaltene, (11) pyrrolic asphaltene, and (12) thiophenic asphaltene.

SARA	ID	Aging Degree				
		0	1	2	3	4
Saturates	1	 C30H62 422 0.803 C30H62 422 0.803	—	—	—	—
	2	 C35H62 482 0.913 C35H62 482 0.913	—	—	—	—
Aromatics	3	 C35H44 464 1.030 C35H44 464 1.030	 C35H42O 478 1.064 C35H42O 478 1.064	 C35H36O4 1.173 C35H36O4 1.173	—	—
	4	 C30H46 406 0.916 C30H46 406 0.916	 C30H44O 420 0.955 C30H44O 420 0.955	 C30H42O2 434 0.995 C30H42O2 434 0.995	—	—
Resins	5	 C36H57N 503 0.977 C36H57N 503 0.977	 C36H55NO 517 1.006 C36H55NO 517 1.006	 C36H53NO2 531 1.040 C36H53NO2 531 1.040	—	—
	6	 C26H50O 414 0.893 C26H50O 414 0.893	 C26H48O2 428 0.930 C26H48O2 428 0.930	—	—	—
	7	 C18H10S2 290 1.417 C18H10S2 290 1.417	 C18H10S2 306 1.540 C18H10S2 306 1.540	 C18H10O2S2 322 1.68 C18H10O2S2 322 1.68	—	—
	8	 C40H60S 572 0.962 C40H60S 572 0.962	 C40H60OS 588 1.010 C40H60OS 588 1.010	 C40H58O2S 602 1.040 C40H58O2S 602 1.040	 C40H56O3S 616 1.071 C40H56O3S 616 1.071	—
	9	 C40H59N 553 1.007 C40H59N 553 1.007	 C40H57NO 567 1.035 C40H57NO 567 1.035	 C40H55NO2 581 1.067 C40H55NO2 581 1.067	—	—

(continued on next page)

Table 1 (continued)

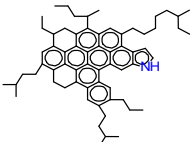
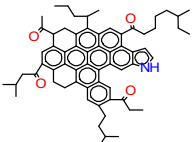
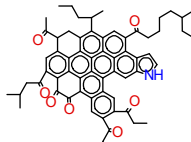
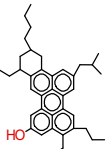
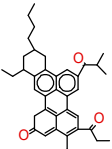
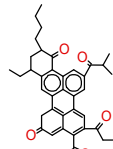
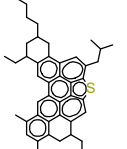
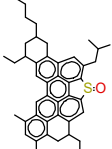
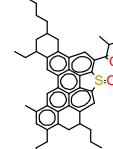
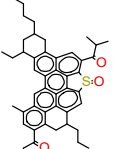
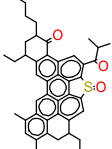
SARA	ID	Aging Degree				
		0	1	2	3	4
Asphaltenes	10	 C66H81N 888 1.104 C66H81N 888 1.104	 C66H73NO4 944 1.188 C66H73NO4 944 1.188	 C66H67NO7 986 1.254 C66H67NO7 986 1.254	—	—
	11	 C42H54O 574 1.049 C42H54O 574 1.049	 C42H50O3 602 1.130 C42H50O3 602 1.130	 C42H46O5 630 1.201 C42H46O5 630 1.201	—	—
	12	 C51H62S 707 1.100 C51H62S 707 1.100	 C51H62OS 723 1.160 C51H62OS 723 1.160	 C51H60O2S 737 1.189 C51H60O2S 737 1.189	 C51H56O3S 751 1.220 C51H56O3S 751 1.220	 C51H56O4S 765 1.252 C51H56O4S 765 1.252

Table 2
Skeletal representation of the rejuvenator molecules employed in this study, accompanied by their chemical formula, molecular mass, and estimated density.

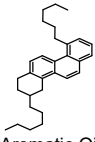
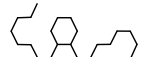
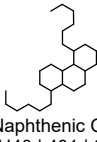
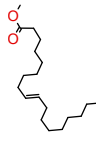
Rejuvenator				
				
Aromatic Oil C26H48 360 0.863 Aromatic Oil C26H48 360 0.863	Engine Oil C22H44 308 0.812 Engine Oil C22H44 308 0.812	Naphthenic Oil C30H40 401 0.984 Naphthenic Oil C30H40 401 0.984	Vegetable Oil C19H36O2 297 0.873 Vegetable Oil C19H36O2 297 0.873	

Table 3
Summary of the characterization results obtained for Group (3) (type “N”) bitumens across all aging conditions. These serve as the basis for constructing molecular models that represent each bitumen type and aging conditions examined in this study.

Test type		Aging Degree				
		0	1	2	3	4
SARA Fractionation	Saturates	5.8	5.88	5.80	5.80	5.71
	Aromatics	63.1	61.88	55.09	49.82	44.28
	Resins	26.2	26.47	30.94	34.89	37.73
	Asphaltenes	4.90	5.64	7.64	8.59	10.90
	Colloidal Index	0.12	0.14	0.19	0.22	0.28
Elemental Analysis	Carbon [m%]	85.5	85.2	84.7	84.4	83.7
	Nitrogen [m%]	0.66	0.67	0.67	0.65	0.67
	Hydrogen [m%]	10.5	10.4	10.3	10.0	9.77
	Oxygen [m%]	2.00	2.39	2.99	3.53	4.58
	Sulfur [m%]	1.28	1.28	1.27	1.29	1.29
GPC	Mass-avg. molar mass [g/mol]	1481	1779	2012	2194	2258
	Number-avg. molar mass [g/mol]	889	929	979	1015	1030
	Dispersity [:]	1.67	1.91	2.06	2.16	2.19

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i^{est}(x_1, x_2, \dots, x_n) - y_i^{exp})^2 \quad (1)$$

where m is the number of chemo-physical properties to fit, with equal penalty weights, n is the number of different molecule types in the mixture, and x represents the count of each molecule type.

- 7 The process is repeated for all the samples required, essentially those of Group (2) through (4), with varying aging conditions.

Four main constraints are enforced when generating the plausible molecular models:

1. The total number of atoms in each mixture must remain within $7000 \pm 10\%$, reducing the relative impact of size differences and boundary conditions and ensuring similar simulation performance across all models to facilitate high-throughput operations. This limitation inherently sets a limitation on the total number of molecules, from 73 to 76, comparing well with other models built in the literature [11,18].
2. Only molecules that are within the aging level above or below of the required mixture's aging condition are selected. For instance, a mixture representing an aging condition of 3 can only be comprised of molecules from aging conditions (2) through (4), unless such a selection cannot produce a valid molecular model. Models of aging conditions 0 and 4 are forced to use molecules of aging degree 0 or 4, respectively.
3. The properties of the obtained molecular mixtures must not overlap within the measured error intervals of other mixtures from different aging conditions. This condition is especially enforced on the oxygen content, which varies from 1 % to 5 %, ensuring that each mixture accurately represents its specified condition and maintains sufficient discretization to yield statistically valid comparisons.
4. The range of molecule types used in each generated mixture must be between 8 and 14. Moreover, there must be at least one molecule from each SARA category. This ensures sufficient variability in the molecules included while aligning with the typical systems used in other research articles that model bituminous materials, which commonly feature an average of 12 molecule types per mixture [18].

Table 4 lists the chemo-physical properties and Table 5 lists the type and number of molecules from Table 1 required to construct each of the mixtures for all samples of Groups (4) (of source type "F"). Similarly, File */model_construction.xlsx* in the Supplementary Information provides the same information for Groups (2) and (3) (from types "T" and "N"). While the molecular models from Group (1) do not require fitting, they are still bound to an atom count limit of $7000 \pm 10\%$, similar to the models from Groups (2), 3, and 4.

Table 4
Characteristics of the molecular systems built to represent Group (4)'s bitumens (of type "F") at 5 aging degrees.

Property	Aging Degree				
	0	1	2	3	4
Total # of molecules	76.0	75.0	74.0	74.0	75.0
Avg. molecular mass (Mw) [g/mol]	431.7	439.3	446.1	460.8	487.1
Density [g/cm ³]	1.019	1.042	1.041	1.051	1.070
Saturates [m/m %]	5.2	5.5	5.1	5.0	4.6
Aromatics [m/m %]	56.5	54.3	48.8	41.1	33.5
Resins [m/m %]	21.9	23.3	25.1	27.3	28.1
Asphaltenes [m/m %]	16.4	16.9	21.0	26.6	33.8
C [m/m %]	83.7	83.2	82.7	82.6	82.0
H [m/m %]	9.7	9.4	9.5	9.4	9.1
O [m/m %]	2.0	2.8	3.0	3.5	4.3
N [m/m %]	0.4	0.4	0.5	0.4	0.4
S [m/m %]	4.1	4.2	4.4	4.1	4.1

Rejuvenated samples are produced subsequent to the creation of non-rejuvenated samples: for each non-rejuvenated sample formulated, rejuvenator molecules from Table 2 are added in accordance with the four rejuvenator dosages by mass as specified in the study (i.e., 0, 5, 10 and 15 % by mass).

Therefore, considering the variables involved in the study—three bitumen sources (of types "T", "N", and "F"), five aging degrees (0 through 4), four rejuvenator types (1 through 4), and four rejuvenator dosages (0 through 3)—there are 53 models per bitumen source. This results in a total of 159 different molecular models for Groups (2) through (4). When combined with the models of Group (1), which involve pure mixtures comprised of molecules from Table 1 and Table 2, the total comes to 197 different molecular models utilized to investigate the properties of bituminous materials in this study. Special labels have been created to identify these samples if needed. The label format "AXXX" is composed of four characters: the first indicates the bitumen type, the second denotes the aging degree, the third specifies the rejuvenator type, and the fourth represents the rejuvenator dosage. For example, the label "T4E3" corresponds to a bitumen of type "T," aged to degree 4, rejuvenated with Engine Oil at a dosage of 15 % by mass. The file named */md_samples_information.xlsx* presents a thorough compilation of the chemo-physical properties for the samples specifically constructed for this study.

2.3. Simulations

2.3.1. Force field selection

The forces governing the motion of atoms in MD simulations are represented as potential energy functions, where the force and energy relationship is expressed in Eq. (2),

$$\mathbf{F} = -\nabla E \quad (2)$$

The molecular simulations of this study employ the Polymer Consistent Force Field (PCFF) [30], where the potential energy (E_{PCFF}) is given by the sum of 12 interaction types, as shown in Eq. (3),

$$E_{PCFF} = \sum E^b + E^{ia} + E^{oa} + E^t + E^{bb} + E^{ba} + E^{bt} + E^{aa} + E^{at} + E^{tt} + E^{VDW} + E^{coul} \quad (3)$$

where each interaction term is described in Table 6. The PCFF force field was chosen for its ability to accurately model complex organic mixtures, including those present in aromatic, resinous, or asphaltenic mixtures. Moreover, the PCFF force field has been widely employed in simulating similar MD systems. All the simulations are performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [31].

In the Supplementary Information, Folder */pcff* comprises a simulated sample of a bituminous blend. This simulation includes all necessary atom types and input parameters for conducting the study's simulations utilizing the PCFF force field. Script */pcff/input.data* can be used to load */pcff/structure.data* into LAMMPS.

2.3.2. Initialization

The steps for preparing molecular systems in LAMMPS for subsequent simulations are described below:

1. The molecular systems are initialized using SMILES notations retrieved and processed into stable 3D conformers via the Rdkit Python module.
2. The initialized molecules are placed into a simulation box at a low initial density of 0.20 g/cm^3 , ensuring an even distribution and minimizing particle overlap by utilizing a low-discrepancy Sobol distribution method.
3. The PCFF force field parameters and charges are assigned to the molecules using Rdkit atom descriptor functions (e.g., `atom.GetIsAromatic()`). The atomic positions are optimized to minimize E_{PCFF} .

Table 5

Type and number of molecules (from Table 1) needed to build each of the molecular models of Group (4)'s (type "F") bitumens at 5 aging degrees.

Category	Molecule	Aging (mol)	Aging (sample)				
			0	1	2	3	4
Saturates	Hopane	0	0	2	0	0	0
	Squalane	0	4	2	4	4	4
Aromatics	Diethylcyclohexane naphthalene	0	40	37	34	30	22
		1	1	1	1	1	0
		2	1	0	1	1	2
	Perhydrophenanthrene naphthalene	0	1	1	0	0	1
		1	2	3	3	2	3
		2	0	1	0	0	1
Resins	Benzobisbenzothiophene	0	11	1	1	2	0
		1	4	15	14	12	13
		0	5	4	6	5	4
	Pyridinohopane	1	0	1	0	0	0
		2	0	0	0	0	2
		0	0	0	0	0	0
	Quinolinhopane	1	0	0	0	0	0
		2	0	0	0	0	0
		0	0	0	0	0	0
	Thioisorenieratane	1	0	0	0	0	0
		2	0	0	0	0	0
		0	0	0	0	0	0
Asphaltenes	Phenolic Asphaltene	1	0	0	0	0	0
		2	0	0	0	0	0
		0	0	0	0	0	1
	Pyrrolic Asphaltene	1	0	1	0	0	0
		2	0	0	0	0	0
		0	0	0	0	0	0
	Thiophenic Asphaltene	1	5	5	6	3	6
		2	0	0	0	0	2
		3	0	0	0	9	1
		4	2	1	3	0	6
		0	0	0	0	0	0
		0	0	0	0	0	0

Table 6

The PCFF potential energy terms involve symbols i, j, k, l , and m for atom groups, with variables representing interatomic distances (r_{ij}), in-plane angles (θ_{ijk}), out-of-plane angles (χ_{ijk}), and dihedral angles (φ_{ijkl}). Multiplicity (M), force field constants ($k, r^0, \theta^0, \chi^0, \varphi^0, s_0, \epsilon_{ij}$ and ϵ_0), and partial atomic charges (q_i) are also included.

Interaction term	Expression
Bond bending	$E^b = \sum_{ij} \sum_{M=2}^4 k_{(m)ij} (r_{ij} - r_{ij}^0)^m$
In-plane angle bending	$E^{ia} = \sum_{ijk} \sum_{M=2}^4 k_{a,ijk}^m (\theta_{ijk} - \theta_{(m),ijk}^0)^m$
Out-of-plane angle bending	$E^{oa} = \sum_{ijk} k_{ijk} \cdot (\chi_{ijk} - \chi_{ijk}^0)^2$
Symmetric torsional angle bending	$E^t = \sum_{ijkl} \sum_{m=1}^4 k_{(m)ijkl} (1 + \cos(m\varphi_{ijkl} - \varphi_{(m)ijkl}^0))$
Cross-coupling bond-bond	$E^{bb} = \sum_{ijkl} k_{ijkl} (s_{ij} - s_{0,ij}) (s_{kl} - s_{0,kl})$
Cross-coupling bond-angle	$E^{ba} = \sum_{ijk} k_{ijk} (r_{ij} - r_{0,ij}) (\theta_{ijk} - \theta_{0,ijk})$
Cross-coupling bond-torsion	$E^{bt} = \sum_{ijkl} (r_{ij} - r_{ij}^0) \sum_{m=1}^3 k_{(m)ijkl} \cos(m\varphi_{(m)ijkl})$
Cross-coupling angle-angle	$E^{aa} = \sum_{angleijkl} k_{ijkl} (\theta_{ijk} - \theta_{0,ijk}) (\theta_{jkl} - \theta_{0,jkl})$
Cross-coupling angle-torsion	$E^{at} = \sum_{ijkl} (\theta_{ijk} - \theta_{ijk}^0) \sum_{m=1}^3 k_{(m)ijkl} \cos(m\varphi_{(m)ijkl})$
Cross-coupling torsion-torsion	$E^{tt} = \sum_{ijklm} k_{ijklm} \cos(\varphi_{ijkl}) \cos(\varphi_{jklm})$
Van der Waals (Lennard-Jones)	$E^{vdW} = \sum_i \sum_{j \neq i} \epsilon_{ij} \left[2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^9 - 3 \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right]$
Electrostatic (Coulomb)	$E^{coul} = \frac{1}{4\pi\epsilon_0} \sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}}$

using the conjugate gradients method until both energies and forces reach a threshold precise to the nearest integer [32].

- The systems undergo compression to achieve the target density. This is performed under NVT conditions over 5 ns, using a true strain rate of 1 %, and applying isotropic deformations to the simulation box every 1 picosecond elapsed.
- The systems are subjected to 50 NPT annealing cycles to remove residual conformational strain, where the temperature is oscillated sinusoidally by ± 25 % of the set equilibrium temperature over a cycle period of 1 ns.
- The systems' densities are stabilized at the desired equilibrium temperature and pressure during two successive 50 ns NPT stages. The final density is calculated based on the average recorded in the latter NPT stage.
- To evaluate system stability, dynamics stages are conducted under NVT and NVE conditions for 50 ns each, with decreasing levels of control by a barostat or a thermostat. This step checks if the potential and kinetic energies remain within 5 % of their initial average. Only systems that meet these criteria are accepted; those that do not are discarded and the process is repeated from Step 1.

The modified Nose-Hoover integration algorithm [33] utilized to estimate the motion of the atoms includes a drag factor to reduce oscillatory effects on controlled temperatures and pressures. The damping factors for temperature and pressure are set at 500 steps, with a particle velocity drag coefficient maintained at 1.0. All LAMMPS procedures are conducted at a constant temperature of (-60 °C) and pressure of 101325 Pa, with a step size of 0.5 fs/step. Periodic boundary

conditions are implemented in all directions throughout every step. The resulting systems from Stages 6 and 7 are used for later simulations performed in this study. Steps 1 and 2 entail the utilization of SMI2PDB [34], whereas Step 3 involves the use of PDB2DAT [35]. These tools have been created by our research team to execute the respective tasks rapidly and effectively.

2.3.3. Production

The following simulation routines, extending from the Initialization stage, are utilized to measure various properties essential for study, as outlined in the Labels selection section. These are outlined as follows.

2.3.3.1. Temperature Sweep. In this study, nine different temperatures were selected for evaluation: -60 , -20 , 0 , 25 , 60 , 120 , 135 , 160 , and 200 °C. These temperatures are commonly utilized as benchmarks for testing bituminous materials in both laboratory and real-world settings. This range is especially significant as it covers most of the viscoelastic response envelope of most bitumens, providing a wide number of cases for assessing the practical applicability of the tested models [36]. To conduct the temperature sweep, each molecular model is heated and equilibrated at the specified temperatures. The process begins with the model initialized at -60 °C (from Step 1 in the Initialization section) and proceeds as follows:

1. An isothermal NPT dynamics step runs for 5 ns at the initial temperature (T_1) to reinitialize the models in LAMMPS.
2. The model is then heated to the next temperature (T_2) over 5 ns using an NPT dynamics step. The increase in temperature from T_1 to T_2 is set to complete in $3.5 \cdot 10^9$ steps with a drag factor of 0.5 to ensure a smooth temperature ramp under isobaric conditions within 5 ns.
3. At T_2 , the system's properties are stabilized during two successive 50-nanosecond isothermal-isobaric NPT steps.
4. Two additional dynamics steps, each lasting 5 ns under NVT and NVE conditions, are conducted without the influence of a barostat or thermostat to assess stability. Simulations that are not stable at higher temperatures require their respective model to be constructed again (Step 1 from Initialization section).

This process, Step 1 through 4, is repeated sequentially for each of the nine temperatures. During these steps, simulations are configured to dump LAMMPS-related properties, namely unwrapped and unscaled atomic positions, velocities, accelerations, forces, stress tensors, and potential energies. These properties are integral to computing the study's properties (detailed in the Labels selection section). Each step from 1 to 4 captures 1000 evenly distributed data points, ensuring a comprehensive collection of points throughout all iterations of the Temperature Sweep test.

2.3.3.2. Cohesive energy density. The Cohesive Energy Density (CED) quantifies the energy required to separate the molecules in a chemical system to an infinite distance, essentially measuring the energy necessary to vaporize a sample. It is one of the properties to be studied (and is essential to compute others) and requires a separate LAMMPS subroutine for its calculation. In bituminous research, the CED is often interpreted as a measure of a sample's viscosity, where higher CEDs correspond to stiffer (i.e., solid-like) bitumens [37]. The CED is calculated using Eq. (4),

$$CED = \frac{E_{nb}}{V} \quad (4)$$

where V is the volume of the molecular model and E_{nb} is the non-bonded energy. The non-bonded energy in the PCFF force field is given by all the potential energy contributions from non-bonded interaction terms in Eq. (3), and is given by Eq. (5),

$$E_{nb} = E_{vdw} + E_{Coul} + E_{long} \quad (5)$$

where E_{vdw} is the Van der Waals energy, E_{Coul} represents the electrostatic energy, and E_{long} accounts for long-range energetic corrections needed in MD simulations to address electrostatic forces that extend beyond the force field's cutoff distance and periodic boundaries. The CED is computed using MD simulations by using the autocorrelation function of E_{nb} as a function of time, as stated in Eq. (6):

$$CED = \frac{1}{V} \int_0^{t_f} \langle E_{nb}(0) - E_{nb}(t) \rangle dt \quad (6)$$

where V is the system's volume, t is the instantaneous time, and t_f is the total simulation time for the simulation. The computation involves two NVT dynamics routines, each one nanosecond long, following the NVT routine of Step 4 in the Temperature Sweep. The first routine measures the initial value $E_{nb}(0)$, and the second assesses $E_{nb}(t)$. Both stages output 1000 data points throughout the simulations.

Separate NVT stages from those run in the **2.3.3.1 Temperature Sweep** are necessary because certain LAMMPS parameters are specifically adjusted to enhance the measurement quality of E_{nb} and enable faster convergence of the autocorrelation function. These adjustments include extending the potential energy cutoff from 9.0 to 14 Angstroms and disabling the impact of Van der Waals and Coulombic forces on intramolecular atoms very close to each other (i.e., those within 1 and 2 bonds from the parent atom), thus producing a smoother potential energy surface more reflective of the effects of intermolecular energy interactions, rather than those that are intramolecular [37].

2.4. Machine Learning

2.4.1. Features selection

The input parameters, or features, used to train the MLMs in this study are based on the chemical composition of the samples (namely features x_1 through x_{30}), their molecular mass (x_{31}), and temperature (x_{32}). The chemical composition of the samples corresponds to an array of 30 values, each representing the fraction content (by number) of a single atom type present in the sample. The atom types are obtained by combining all the force field atom types present in all the samples tested in this study, numbered x_1 through x_{30} . The atom types and their descriptions are presented in Table 7. The PCFF force field differentiates atoms not only by their elemental symbol but also by their hybridization state, degree, ring presence, participation in an aromatic system, and in some cases (e.g., sulfur atoms), the list of neighboring atoms bonded to them [30].

Generating an "atom-type" composition formula for each model of this study (instead of a conventional chemical formula that groups atoms simply by their element symbol) aids in differentiating the impact that seemingly "equal" atoms have on the simulations based on their chemical function (e.g., a hydrogen in an alcohol behaves differently than one in a terminal carbon [38]). The "atom-type" formula is obtained by applying Eq. (7),

$$x_{i,1-30} = \frac{1}{N} \left(\sum a_1, \sum a_2, \dots, \sum a_{30} \right) \quad (7)$$

where N is the total number of atoms, and a_i corresponds to the number of a certain atom type in the sample. File */atom_type.formulas.xlsx* in the **Supplementary Information** displays the "atom-type" formula ($x_{1,\dots,30}$) for all the molecular models of this study.

Feature x_{31} , the sample's average molecular mass, serves to roughly account for the size of the molecules used in the molecular model. This feature introduces a size-dependency factor and prevents situations where the values of x_1 through x_{30} are technically within their bounds but correspond to molecules that are either too small or well beyond the size covered by the reference molecules used to train the MLM.

Table 7

List of atom types and their respective description derived from the PCFF force field. These atom types are instrumental in formulating the “atom type” formulas employed for chemical characterization of MD systems and for training the MLMs of this study. Colors have been preassigned to identify atom types when drawing molecules.

x_n	Atom type	Color	Description	Color	x_n	Atom type	Description
x_1	hc	Red	Hydrogen bonded to carbon	Black	x_{16}	hn	Hydrogen bonded to nitrogen
x_2	o=	Green	Oxygen double bonded to O, N, C, S, P	Purple	x_{17}	c3	Sp ³ carbon with 3 H's
x_3	ho	Blue	Hydrogen bonded to oxygen	Yellow	x_{18}	oh	Oxygen in hydroxide ion (OH ⁻)
x_4	c_1	Yellow	Amide, acid and ester carbonyl carbon	Blue	x_{19}	cpc	Alpha/ipso carbon in aromatic ethers (-C-O-C-)
x_5	c2	Cyan	Sp ³ carbon with 2 H's	Green	x_{20}	c=1	Non-aromatic, next to end doubly bonded carbon
x_6	c5h	Magenta	Sp ³ carbon in 5-membered ring	Light Green	x_{21}	c3oe	Alpha carbon in methyl containing ethers (-C-O-CH ₃)
x_7	c0oe	Grey	Alpha carbon in ether containing tertiary alkyl group (-C-O-C-R ₃)	Olive	x_{22}	nh	Sp ² nitrogen in 5-or 6-membered ring
x_8	sp	Brown	Sulfur in an aromatic ring (e.g. thiophene)	Light Blue	x_{23}	c0	Sp ³ carbon with 0 H's
x_9	na1	Dark Green	Sp ³ nitrogen in secondary aliphatic amines	Teal	x_{24}	c5	Sp ² aromatic carbon in 5-membered ring
x_{10}	s'	Dark Blue	S in thioketone group	Brown	x_{25}	cp	Sp ² aromatic carbon with partial double bond
x_{11}	oc	Orange	Sp ³ oxygen in ether or acetals	Light Green	x_{26}	c_0	Aldehydes and ketones carbonyl carbon
x_{12}	cs	Light Green	Sp ² aromatic carbon in 5 membered rings next to S	Purple	x_{27}	c=	Non-aromatic end doubly bonded carbon
x_{13}	np	Purple	Sp ² nitrogen in 5- or 6- membered ring	Red	x_{28}	c5h1	Sp ² aromatic carbon in 5-membered ring
x_{14}	o_1	Blue	carbonyl oxygen	Blue	x_{29}	c1	Carbon in CO
x_{15}	c=2	Magenta	Non-aromatic doubly bonded carbon	Magenta	x_{30}	o_2	Ester oxygen

Additionally, the use of molecular mass as a feature is beneficial because it is known to significantly affect the properties of organic mixtures, where higher molecular masses often correspond to thicker, more solid-like hydrocarbons. The range of values for x_{31} spans from 200 g/mol to 1000 g/mol, covering the molecular mass of most molecules present in bituminous materials [39]. Feature x_{32} corresponds to the equilibrium kinetic temperature of the sample. Table 11 in the Features Range

section displays the range of values (low to high) that each feature (x_1 through x_{32}) covers to train the MLM in this study.

2.4.2. Labels selection

The properties, or labels, correspond to the properties that are to be predicted by the MLMs. There are 12 properties in total, ranging from fundamental MD properties like potential energies (total, Van der Waals,

Table 8

List of 12 properties to be computed using MD simulations and to be predicted by the MLMs of this study.

Group	Label	Property	Expression	Notes
Fundamental	y_1	Potential Energy (E_p) [kJ/kg]	E_p	Obtained from the average of the second NPT run in Step 3 of the 2.3.3.1 Temperature Sweep run.
	y_2	Van der Waals Energy (E_{vdw}) [kJ/kg]	E_{vdw}	
	y_3	Electrostatic Energy (E_{Coul}) [kJ/kg]	E_{Coul}	
Volumetric	y_4	Molar Volume (V_m) [m ³ /mol]	$V_m = \frac{M_w}{\rho}$	
	y_5	Accessible Volume (V_a) [L]	$V_a = \frac{V_{free}}{V}$	
	y_6	Density (ρ) [kg/m ³]	$\rho = \frac{m}{V}$	
Kinetic	y_7	Self-diffusion coefficient (D) [m ² /s]	$D = \lim_{n \rightarrow \infty} \frac{ \mathbf{r}(t) - \mathbf{r}(0) ^2}{6t}$	
Thermodynamic	y_8	Cohesive Energy Density (CED) [kJ/m ³]	$CED = \frac{1}{V} \int_0^f (E_{nb}(0) - E_{nb}(t)) dt$	Obtained from the 2.3.3.2 Cohesive Energy Density run.
	y_9	Enthalpy of Vaporization (H_v) [kJ/kg]	$H_{vap} = V_m CED + RT$	
	y_{10}	Solubility Parameter (δ_{sol}) [kJ ^{0.5} /m ^{1.5}]	$\delta_{sol} = \sqrt{\frac{CED}{V_m}}$	
	y_{11}	Isobaric Heat Capacity (C_p) [kJ/kg/K]	$C_p = \left(\frac{\delta H}{\delta T} \right)_p$	Obtained by capturing internal energy fluctuations during the 2.3.3.1 Temperature Sweep run (Step 1 through 3).
	y_{12}	Thermal Expansion Coefficient (β) [1/K]	$\beta = \frac{1}{V} \left(\frac{\delta V}{\delta T} \right)_p$	

etc.) to more physically relevant properties (density, heat capacity, etc.). These properties aim to establish MLMs that can, to a certain extent, abstract away from the use of MD simulations to predict the properties of bituminous materials (y_i) almost as accurately as the properties often generated using MD simulations for bitumens or similar materials (y_i^{MD}). There is one MLM generated for each label or property y_i (for a total of 12), for which each MLM_i is given by Eq. (8) as

$$MLM_i = f(x_1, x_1, \dots, x_{32})_i = y_i \approx y_i^{MD} \quad (8)$$

where x_i correspond to a combination of features describing the molecular system for which the property, y_i , is required. The full list of labels, or properties, along with their expressions and descriptions to be predicted by the MLMs of this study are presented in Table 8.

2.4.3. Models

The MLMs of this study are generated and trained using a Python script, using Pandas [40] and Sklearn [41] libraires, that defines and implements a hybrid predictive model, combining the strengths of both tree-based and linear interpolation models. It first initializes a Hybrid-Model class, which utilizes a Random Forest Regressor (RFF) [42] as the primary predictive model and a Linear Regression model for interpolation capabilities. During prediction, the hybrid model first checks if the input features match those in the training data. If they do, predictions are directly made using the tree model. Otherwise, it identifies the two nearest neighbors using a k-dimensional tree, predicts their values with the tree model, and interpolates between them based on their distances to provide the final prediction [43].

To assess the predictive potential of the machine learning models, 20 % of the properties measured through MD simulations are randomly selected and excluded from the training set. Specifically, out of a total of 1,773 samples (197 samples across a temperature sweep of 9 temperatures), 354 samples are used for evaluating the MLMs' unobserved predictive performance, while the remaining samples will serve as input for training the models. The 20/80 ratio is widely adopted in the literature for similar MLMs and is expected to perform well with the available data in this study, given the abundance of data points and how similar the materials are with respect to the range of input features selected [44]. The script follows these steps during execution:

- 1. Data Reading and Preparation:** Load the dataset containing the x_1, x_1, \dots, x_{32} and y_i^{MD} values for all the simulations of this study into Panda's DataFrame objects.
- 2. Hybrid Model Initialization:** Initialize the hybrid model with a RFF and a linear interpolation model, incorporating a preprocessing pipeline for data standardization.
- 3. Model Training:** Train both the RFF and interpolation models and construct a K-dimensional tree to enable nearest neighbor searches. Determine feature ranges for boundary checking.
- 4. Prediction and Evaluation:** Predict the values of y_i using the hybrid model for 80 % of the training arrays of x_1, x_1, \dots, x_{32} values used. This ensures that 20 % of the computed data points remain unobserved by the MLMs, which are used in Step 6 to corroborate the capacity of the MLM to predict properties given a combination of features previously unknown to it.
- 5. Interpolation Smoothness:** The generated MLM is tested with a finer set of x_i values (20 values in between), to ensure that the predicted values are smooth and continuous ($y_i^{left} < y_i < y_i^{right}$).
- 6. Results Handling and Model Saving:** Save the values of x_i, y_i^{MD} , and y_i to a CSV file. The resulting model, MLM_i , is serialized and saved if the R-squared value is above 0.98 and the magnitudes of the differences between y_i^{MD} and y_i are below 5 % for both observed and unobserved datasets.

All 12 MLM files, created by serializing the HybridModel objects along with all associated features using Python's Pickle module, are located in the /MLMs directory within the [Supplementary Information](#). These MLM files can be reloaded back into Python and be used to predict properties by using custom set input features (x_1, \dots, x_{32}).

3. Results

This section offers a detailed summary of the study's results. [Section 3.1 Simulated Properties](#) compares the MD simulation results with experimental observations. [Section 3.2 Atom Type Occurrences](#) and [3.3 Features Range](#) explain the relationship between SARA fractions and force field atom types, and discuss the study's feature ranges, showing how atom types can be applied to real bitumens. [3.4 Prediction Potential](#) highlights the predictability of the generated MLMs, demonstrating their accuracy in forecasting material properties compared to MD simulations. [Section 3.5 Features Importance and Directionality](#) examines the influence and directionality of each feature on the measured properties, providing insights into manipulating inputs to obtain certain outputs. Lastly, [section 3.6 Max-Min Features Optimization](#) provides guidance on optimizing the measured properties, enabling scientists to use specific feature combinations to achieve desired material characteristics.

3.1. Simulated properties

File /subset_properties.xlsx in the [Supplementary Information](#) contains the measured properties (y_1 through y_{12}) for all pure rejuvenator molecular models (for all temperatures). For validation purposes, [Table 9](#) shows both experimental and computational measurements of 5 physically relevant properties: density, enthalpy of vaporization, solubility parameter, heat capacity, and thermal expansion coefficient. Due to the lack of experimental data on each sample, information is displayed for each SARA fraction and Rejuvenators instead.

3.2. Atom type Occurrences

[Table 10](#) presents the normalized occurrence of each atom type within the SARA fractions and rejuvenators group. This information helps clarify the dominant atom types in each SARA fraction, offering insights into the chemical nature of the material under study. For example, the table shows that SARA fractionation does not adhere strictly to chemical fingerprinting principles. Specifically, the aromatic content in Asphaltenes and Resins is equal to or greater than that of the Aromatics fraction, suggesting that SARA analysis may not reliably reflect the chemical composition of the material. Furthermore, the data in [Table 10](#) indicates that current rejuvenators closely resemble the characteristics of the Saturates fraction, implying that these compounds are designed to emulate the chemo-physical behavior of lighter bitumen fractions.

3.3. Features range

Given that the study sweeps through an array of 197 different bituminous samples, corresponding to those of Groups (1) through (4), and 9 different temperatures, there exists 1773 unique arrays of (x_1, x_1, \dots, x_{32}) features used as inputs for training the 12 MLMs of this study. Therefore, the range of values (from minimum to maximum) covered by the study's simulations for each x_i are presented in [Table 11](#). These serve to understand the boundaries of the features covered by this study, and whether the features of another material to be studied by other researchers lies within these ranges.

Table 9
Averaged thermodynamic properties obtained from the molecular models corresponding to each SARA fraction and their experimental counterpart.

	Experimental (y_i^{exp})					Computational (y_i^{MD})				
SARA	ρ [kg/m ³]	ΔH_{vap} [kJ/kg]	δ_{sol} [kJ ^{0.5} /m ^{1.5}]	C_p [kJ/kg/K]	β [1/K]	ρ [kg/m ³]	ΔH_{vap} [kJ/kg]	δ_{sol} [kJ ^{0.5} /m ^{1.5}]	C_p [kJ/kg/K]	β [1/K]
	[45]	[46–48]	[45,48]	[49–52]	(•10 ⁴)					(•10 ⁴)
					[51–53]					
Saturates	850	270–290	509.9	2.0–3.0	5–8	873.5	283.3	494.2	2.82	7.60
Aromatics	1000	300	631.6	1.5–2.0	4–7	1002	338.2	574.4	1.91	6.58
Resins	1050	300	618.9	1.5–2.0	1–4	1060	325.3	580.5	2.06	3.32
Asphaltenes	1070	350–450	575.8	1.0–1.5	0.1–1.5	1068	328.7	588.7	2.19	0.32
Engine Oil	750–900	300	500.2	1.8–2.2	5–10	833.7	330.4	524.8	2.95	7.37
Vegetable Oil	850–930	200–250	489.5	2.0–3.5	8–12	865.4	356.1	555.2	2.67	10.58
Aromatic Oil	950–1000	300	607.8	1.5–2.5	4–8	977.2	349.1	584.2	2.34	7.19
Naphthenic Oil	850–900	300–320	598.0	1.5–2.5	7–10	888.1	300.5	516.3	2.69	8.33

Table 10
Normalized occurrence of each atom type for each SARA category, including their standard deviation. These values aid in identifying a hydrocarbon mixture’s SARA category based on its composition.

x_n	Atom Type	Saturates	Aromatics	Resins	Asphaltenes	Rejuvenator
x_1	hc	0.6565 ± 0.0245	0.5558 ± 0.0438	0.5346 ± 0.111	0.518 ± 0.025	0.6296 ± 0.0413
x_2	o=	0	0	0.0083 ± 0.0173	0.0032 ± 0.0044	0
x_3	ho	0	0	0	0.0009 ± 0.0031	0
x_4	c_1	0	0	0	0	0.0044 ± 0.0088
x_5	c2	0.1591 ± 0.0209	0.1601 ± 0.0467	0.0968 ± 0.0521	0.1078 ± 0.0142	0.2368 ± 0.0366
x_6	c5h	0.0103 ± 0.0146	0	0.002 ± 0.0042	0	0
x_7	c0oe	0	0	0.0017 ± 0.0044	0	0
x_8	sp	0	0	0.0073 ± 0.0185	0.0008 ± 0.0026	0
x_9	na1	0	0	0.0002 ± 0.0002	0	0
x_{10}	s’	0	0	0.0083 ± 0.0173	0.0032 ± 0.0044	0
x_{11}	oc	0	0	0.0017 ± 0.0044	0	0
x_{12}	cs	0	0	0.0132 ± 0.0372	0.0016 ± 0.0053	0
x_{13}	np	0	0	0.0042 ± 0.0053	0	0
x_{14}	o_1	0	0.0178 ± 0.0201	0.007 ± 0.0085	0.0196 ± 0.02	0.0044 ± 0.0088
x_{15}	c = 2	0	0	0.0027 ± 0.007	0.0019 ± 0.0043	0
x_{16}	hn	0	0	0	0.0019 ± 0.0032	0
x_{17}	c3	0.0847 ± 0.0032	0.0327 ± 0.0067	0.0732 ± 0.0387	0.071 ± 0.0021	0.0259 ± 0.0057
x_{18}	oh	0	0	0	0.0009 ± 0.0031	0
x_{19}	cpc	0	0	0	0.001 ± 0.0032	0
x_{20}	c = 1	0	0	0.0026 ± 0.0058	0	0
x_{21}	c3oe	0	0	0	0	0.0044 ± 0.0088
x_{22}	nh	0	0	0	0.0019 ± 0.0032	0
x_{23}	c0	0.0258 ± 0.0365	0	0.0209 ± 0.0265	0	0
x_{24}	c5	0	0	0.0132 ± 0.0372	0.0201 ± 0.0146	0
x_{25}	cp	0	0.1701 ± 0.0405	0.1631 ± 0.1569	0.1855 ± 0.0137	0.05 ± 0.1
x_{26}	c_0	0	0.0178 ± 0.0201	0.007 ± 0.0085	0.0196 ± 0.02	0
x_{27}	c=	0	0	0.0026 ± 0.0058	0	0.0088 ± 0.0175
x_{28}	c5h1	0.0103 ± 0.0146	0	0.0028 ± 0.0048	0	0
x_{29}	c1	0.0532 ± 0.017	0.0457 ± 0.0209	0.0269 ± 0.0146	0.0409 ± 0.0042	0.0314 ± 0.0353
x_{30}	o_2	0	0	0	0	0.0044 ± 0.0088

Table 11
Minimum and maximum values for all 32 features measured across all 197 samples of this study.

x_i	Feature	x_i^{min}	x_i^{max}	x_i	Feature	x_i^{min}	x_i^{max}
x_1	hc	0.3125	0.674	x_{17}	c3	0	0.101
x_2	o=	0	0.063	x_{18}	oh	0	0.010
x_3	ho	0	0.010	x_{19}	cpc	0	0.011
x_4	c_1	0	0.018	x_{20}	c = 1	0	0.010
x_5	c2	0	0.273	x_{21}	c3oe	0	0.018
x_6	c5h	0	0.021	x_{22}	nh	0	0.007
x_7	c0oe	0	0.013	x_{23}	c0	0	0.054
x_8	sp	0	0.067	x_{24}	c5	0	0.133
x_9	na1	0	0.001	x_{25}	cp	0	0.563
x_{10}	s’	0	0.063	x_{26}	c_0	0	0.054
x_{11}	oc	0	0.013	x_{27}	c=	0	0.035
x_{12}	cs	0	0.133	x_{28}	c5h1	0	0.021
x_{13}	np	0	0.011	x_{29}	c1	0	0.081
x_{14}	o_1	0	0.054	x_{30}	o_2	0	0.018
x_{15}	c = 2	0	0.020	x_{31}	Molecular mass [g/mol]	290.4	986.2
x_{16}	hn	0	0.007	x_{32}	Temperature [K (C°)]	213.15 (−60)	473.15 (200)

3.4. Prediction potential

Table 13 presents scatter plots for properties y_1 through y_{12} , showing predicted values (y_i) against measured values (y_i^{MD}). The plots include data from both the observed and unobserved datasets used to evaluate the predictive performance of the MLMs. Ideally, the plots should exhibit a 45° diagonal line, indicating perfect agreement between predicted and measured values in both observed and unobserved datasets. The complete list of simulated values along with their respective predictions, based on the input features for each sample analyzed in this investigation, is available in the tabulated files located within the /predictions directory in the [Supplementary Information](#).

3.5. Features importance and directionality

This section elucidates the significance of each feature (x_1 through x_{32}) on the predicted properties (y_1 through y_{12}) by the MLMs generated in this study. Table 12 presents the influential features' scores across all 12 MLMs, normalized across all 32 features, with the top five most influential features highlighted in bold. Additionally, Table 12 illustrates the directional influence of each feature on an increase in property y_i , denoted by plus and minus signs. This assessment involves testing intermediary values within the feature range (see Table 11) to determine whether the influence is ascending or descending. Identifying the most influential features affecting each property aids in selecting relevant properties for fine-tuning when predicting the properties of new bitumens, particularly when understanding whether increasing or decreasing a certain feature x_i would increase or decrease the value of y_i . File /features_importance.xlsx contains a comprehensive list of influence factors, including those whose importance nears zero ($< 10^{-3}$).

3.6. Max-Min features optimization

This section focuses on optimizing combinations of input features, x_1 through x_{32} , to ascertain the conditions under which each property (y_1 through y_{12}) achieves its maximum and minimum values. The employed algorithm assesses various feature combinations using a differential evolution strategy, aiming to approach absolute maximum and minimum values. Utilizing SciPy's differential_evolution [28] function for optimization, the algorithm imposes penalties for out-of-bounds values to guide the search within specified property constraints, ensuring effective optimization of feature combinations for maximizing and minimizing property values. This analysis serves to provide a set of precomputed feature combinations to guide researchers in the process of maximizing or minimizing material properties. For instance, researchers may be interested in the combination of features needed to design bitumens whose heat capacity is maximized.

Considering that the impact of temperature (x_{32}) is generally understood (e.g., higher temperature decreases material density), the analysis is performed solely at 25 °C for simplicity, focusing on features corresponding only to chemical traits. File /max-min_optimization.xlsx in the [Supplementary Information](#) presents a comprehensive list of different feature combinations, from x_1 through x_{32} , yielding the computed maximum and minimum values for each property y_i . Table 14 summarizes these results by including only the top 5 influential features from the combinations attained, which often account for 95 % of the cumulative influence over y_i .

4. Discussion

In this study, prior to assessing the potential of MLMs to predict material properties and their application in designing and adjusting bituminous systems, it is crucial to validate the training data derived from MD simulations. This validation is conducted by comparing the MD-generated data with empirical observations. Although it is

challenging to secure experimental data for each of the 1773 samples individually, it is possible to compare key physical properties such as densities, heat capacities, thermal expansions, and solubility parameters for complex mixtures like heavy oils and bitumens, based on their SARA fractions and common compounds like vegetable oil. This approach does not allow for individual model validation but is sufficient to confirm broader trends, thereby supporting the applicability of the findings using MD.

The results affirm that the MD simulations accurately represent the magnitude and trends of the observed properties, as detailed in Table 9. This substantiates the relevance of using this data to train the MLMs. Notably, the MLMs developed in this study can predict all the MD-derived properties with an accuracy within 5 %, maintaining consistency with the observed trends, as observed in Table 13.

Given these findings, the discussion section of this paper will not reiterate the predictive capability of the MLMs, or the quality of the simulation results, which are already established. Instead, it will focus on analyzing how specific features (x_1 to x_{32}) influence the properties (y_1 to y_{12}) of bitumens. This analysis will guide the optimization of bituminous material design by varying parameters (force field atom types, molecular mass, and temperature) to enhance desired properties, thus benefiting from the predictive potential of the MLMs to design new bitumens. A brief discussion is presented for each property y_i below.

4.1. Potential energy (E_p)

Interpreting the physical implications of variations in potential energy is challenging due to the multiple interaction terms that contribute in contrasting ways to the system's overall potential energy. Additionally, the relationship between potential energy and more relevant physical properties can be non-trivial. Generally, non-zero potential energies indicate instability, suggesting non-equilibrium positions that reflect strain in bonds, angles, or dihedrals due to either over-compression or under-compression [38]. In systems with multiple molecules and atom types, a perfect equilibrium position for every interaction type is unattainable [32].

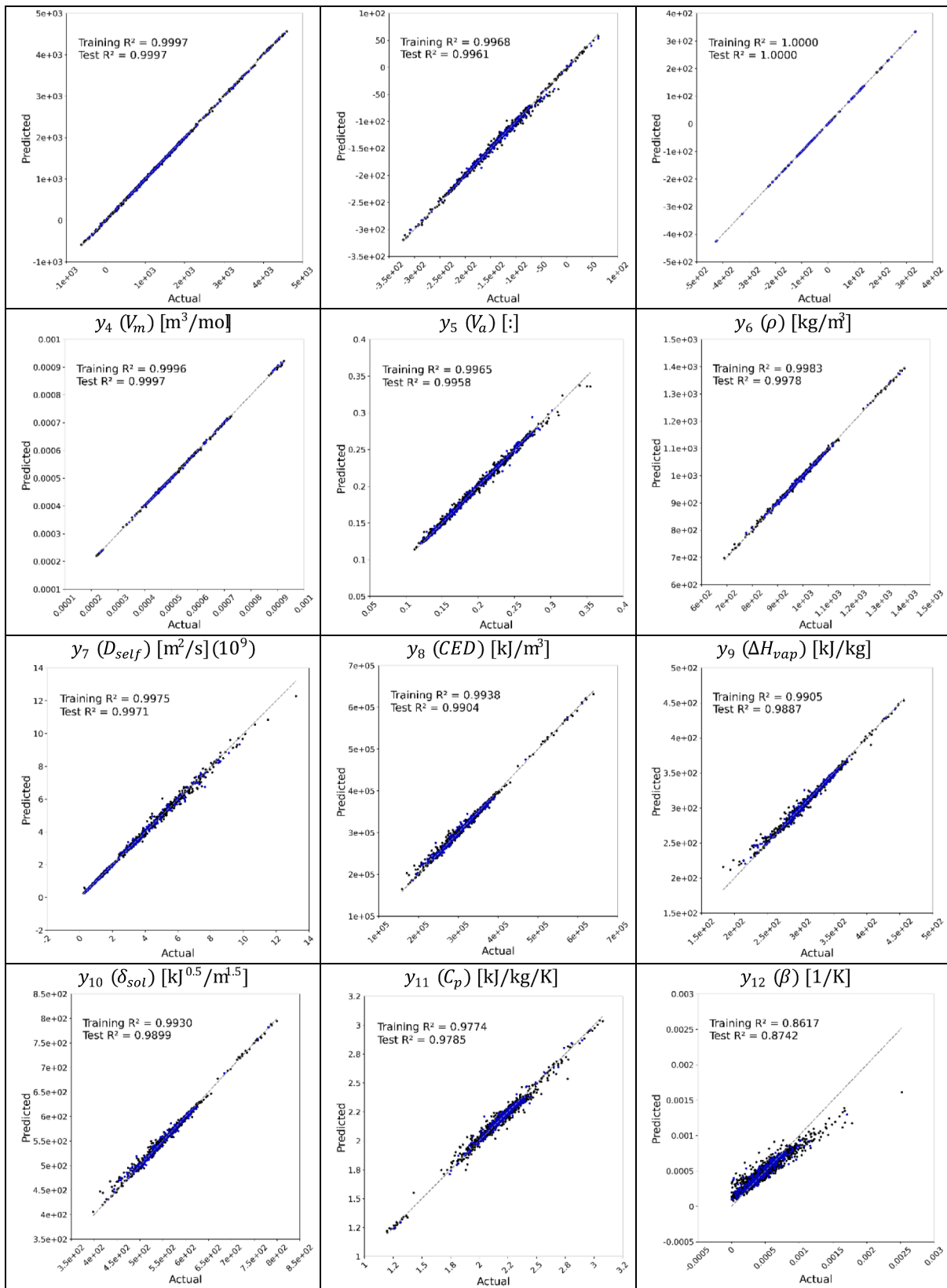
Nonetheless, lower overall potential energies are associated with more stable and thermodynamically favorable conformations [54]. Therefore, the objective in adjusting potential energy terms is to identify thermodynamically stable conformations where the system-wide potential energy is as close as possible to the theoretical absolute equilibrium under given conditions. Additionally, potential energies can serve as a critical design element in the construction of new molecules. Chemically unrealistic modifications to molecules can lead to artificially high potential energies, suggesting that such modifications either need to be adjusted or avoided. For example, when simulating progressively aged molecules, it is crucial to oxidize them in a manner that produces structures with progressively lower potential energies, ensuring that the oxidation process remains thermodynamically favorable.

The presence of multiple equilibrium positions and the lack of direct relatability to physical properties complicate the use of a directionality index. This complexity arises because the potential energy, given by E_{PCFF} , comprises numerous interaction potentials with various equilibrium positions as seen in Table 5. Consequently, a star (*) is added to reported values whose directionalities change throughout the feature range tested.

The most influential features affecting the potential energy of the molecular systems are features x_{24} (atom type "c5"), x_{25} ("cp"), x_{32} (temperature), x_5 ("c2"), and x_1 ("hc") with scores of -49 %, +23 %, 7 %, 3 %, and 3 %* respectively. While an increase in temperature predictably raises potential energy [54], its influence (7 %) is by far not the most significant factor. Over 70 % of the influence is attributed to the content of "c5" and "cp" atom types, corresponding to the backbones of 5-membered and 6-membered aromatic rings, respectively. Notably, an increase in 6-membered aromatic rings results in a more energetic configuration. However, the addition of 5-membered rings, typical of

Table 12

Scatter plots corresponding to all properties (y_1 through y_{12}) depicting predicted values against measured values. Blue dots represent combinations of input features used to train the MLMs (observed), while black dots denote combinations of features considered unknown (unobserved), thereby denoting the MLMs true prediction potential.



resinous and asphaltenic structures containing sulfur and sulfoxidic groups, results in the largest decrease in potential energy.

This observation is notable because, contrary to the general expectation in organic chemistry, 6-membered rings usually lead to lower energy (i.e., more stable) conformations. However, in certain featural compositions specific to bitumens, the addition of 5-membered rings is greatly favored over 6-membered rings. Unlike six-membered rings, which are typically planar and symmetrical, five-membered rings can adopt non-planar conformations, disrupting planarity and reducing overall strain in the molecule [55]. This flexibility leads to energetically more favorable configurations, thereby reducing potential energy in larger polycyclic frameworks like those in resins or asphaltenes. This observation aligns well with many resinous or asphaltenic structures reported in the literature, where the presence of 6-membered polycyclic aromatic regions is often accompanied by 5-membered aromatic rings to decrease intramolecular strains [56].

The remaining influential features pertain to the content of hydrogens connected to carbons (feature x_0 of atom type “hc”), indicating saturation degree, and the presence of aliphatic carbons (features x_5 , x_{17} , x_{29} of atom types “c2”, “c3”, and “c1” respectively) found in branches. An increase in these features mildly decreases the system’s potential energy. However, this observation can vary, and the addition of these chemical features should be studied on a case-by-case basis, as the addition of branches plays an important factor in the packing of the molecules and in the steric hindrance experienced by their conformations.

To generate systems with potential energies that are negative but close to zero (indicating near system-wide equilibrium compared to other bitumens), all features except those related to high saturation and Sp^3 hybridized carbons should be minimized. Specifically, features x_{24} , x_{25} , x_5 , and x_6 should be set to below 0.009, 0.012, 0.09, and 0.03, respectively, while ramping up features x_1 and x_{29} (types “hc” and “c1”) to above 0.65 and 0.02, respectively. However, the nature of such systems resembles saturates or rejuvenators rather than conventional bituminous blends.

The range of potential energies spans from -1000 kJ/kg to 5000 kJ/kg, with most bitumens having a potential energy around 1550 kJ/kg. Although this value lacks physical interpretation on its own, it aligns well with those found in simulated environments reported in the literature [57]. The values of the features (excluding temperature) necessary to construct bitumens nearing these extremes are detailed in Table 11.

4.2. Van der Waals energy (E_{vdw})

Van der Waals interactions involve a single equilibrium position in their calculations, facilitating the capture of clearer trends. These interactions are less affected by intramolecular strains from bonds, angles, dihedrals, or impropers, making them less dependent on intramolecular conformations. Unlike other physical properties, this energy consists of two components within the Lennard-Jones potential: a repulsive term with large positive values and an attractive term with smaller negative values. A system with a stable attractive regime is often preferred, as it indicates well-separated, equilibrated, and evenly distributed molecules, reflecting both physical stability and numerical stability of the simulations [54].

The top five most influential features affecting Van der Waals interactions are temperature (+33.5 %) (feature x_{32}), molecular mass (+11.2 %) (feature x_{31}), and features corresponding to aliphatic chains (+34 %) (x_5 “c2”, x_{17} “c3”, x_{23} “c0”, and x_{29} “c1”), all of which have a positive influence. Increases in temperature, molecular mass, and aliphatic chains are fundamentally known to raise Van der Waals energies in a bituminous system [38]. Larger, more positive values indicate that the repulsive term is dominating, signifying pronounced steric hindrance effects and generally less cohesion or compaction among the molecules. Similarly to potential energy, the use of Van der Waals energies as a design feature is mostly beneficial when testing the stability of molecular conformations on customized molecules.

To achieve lower Van der Waals energies in bitumen, all the aforementioned features should be minimized. Aliphatic features should be set to their minimum range values, and molecular mass should be kept below 350 g/mol. The range of Van der Waals energies for the molecular systems in this study spans from -350 to 100 kJ/kg, which compare well with other values obtained from the literature [15]. The values of the features (excluding temperature) necessary to construct bitumens nearing these extremes are detailed in Table 11.

4.3. Electrostatic energy (E_{Coul})

Electrostatic energies are approximated by a Coulombic potential in the PCFF force field, whose dependency on distance is significantly less aggressive than in the case of Van der Waals energies, and they continuously decrease with increasing interatomic distance. This characteristic means that these interactions can span much longer distances, requiring careful interpretation when dealing with systems containing large molecules and periodic boundary conditions [58]. Higher values of E_{Coul} indicate the presence of more polarizable features (e.g., heteroatomic functional groups) or a higher content of electron delocalization effects (e.g., polycyclic aromatic rings), though their impact on physical properties is non-trivial. Consequently, its use as a design feature is limited to measuring numerical stability and ensuring that the electrostatic influence over Van der Waals interactions remains below 5 %, as reported in the literature for most heavy oils, bitumens, and other petroleum derivatives [57]. This can be used when deciding on the location of highly polar characteristics in molecules as two highly polar features placed very close to each other may make E_{Coul} take precedence over E_{vdw} , a phenomenon not typical in bituminous molecules.

Unlike many other properties, temperature and molecular mass have negligible impacts on E_{Coul} , implying that only chemical features dictate it. Moreover, the influence cannot be reduced to a short list of most influential factors, as these are distributed among a much wider array, namely 15 chemical features, cumulatively accounting for over 95 % of the influence. This is expected, given the multiple functional groups and chemical traits significantly impacting the electrostatics of a molecular system present in this study. Among the most influential (positively) are atom types “o.1”, “c.0”, “o=”, “s’”, and “cp”, while negatively influential are types “c3”, “hc”, and “hn”. This indicates that a more aromatic, oxidized environment increases the influence of electrostatic forces, while in a more hydrogenated, reduced environment, these forces are lower. The range of electrostatic energies for the molecular models in this study spans from -500 to 400 kJ/kg. While experimentally obtained values are not available, these compare well to those obtained in other MD simulations performed in other studies [15]. The values of the features (excluding temperature) necessary to construct bitumens nearing these extremes are detailed in Table 11.

4.4. Density (ρ) and molar volume (V_m)

The most influential features affecting the density of the molecular system are features x_1 , x_{32} , and x_{25} , accounting for over 80 % of the influence. Feature x_1 corresponds to hydrogens connected to carbon atoms, of type “hc”, indicative of the system’s saturation degree. An increase in “hc” content results in a significant decrease in density, consistent with the low densities (700 – 900 kg/m³) of highly saturated hydrocarbons such as Saturates and Rejuvenator fractions in bitumens seen in Table 4.

Expectedly, feature x_{32} reflects the impact of temperature, where increased temperature decreases material density, ensuring that both the MD simulations and the predictive potential of the MLMs align with fundamental observations. Feature x_{25} pertains to the “cp” atom type, the backbone of 6-membered aromatic rings. An increase in “cp” atoms significantly increases material density, aligning with the higher densities (above 1000 kg/m³) of highly aromatic compounds seen in Table 4.

Table 13

List of normalized importance scores for features influencing predicted properties by the studied MLMs. Bold values indicate the top five scores. The sign denotes the direction of influence on the predicted property (y_i) with an increase in the feature (x_i). Importance factors nearing zero ($< 10^{-3}$) are shown as 0.

Feature		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
		E_p	E_{vdw}	E_{coul}	V_m	V_a	ρ	D_{self}	CED	ΔH_{vap}	δ_{sol}	C_p	β
x_1	hc	-0.029	-0.013	-0.08	0.013	0.05	-0.373	0.047	-0.138	-0.04	-0.137	0.678	0.065
x_2	o=	-0.003	0.038	0.119	0	0.002	0.001	-0.003	0.006	-0.007	-0.005	-0.003	0.011
x_3	ho	0	0	-0.006	0	0	0	0	0	0	0	0	0.009
x_4	c_1	0	0	0	0	0	0	0	0	-0.002	0	0	0.003
x_5	c2	-0.033	0.013	0.012	-0.011	0.011	-0.069	0.012	-0.044	-0.019	-0.051	0.044	0.073
x_6	c5h	-0.028	0.002	0	0	0	0	-0.001	-0.004	-0.007	0.006	-0.003	0.013
x_7	c0oe	0.004	0.002	-0.006	-0.001	0	0	0	-0.003	-0.008	-0.003	-0.002	0.008
x_8	sp	-0.007	-0.002	-0.003	0	0	0	0	-0.002	-0.003	-0.002	0.002	0.009
x_9	nal	0	-0.006	0	0	0	0	-0.132	-0.005	-0.008	-0.005	0	0.002
x_{10}	s'	-0.003	0.029	0.108	0	-0.001	0.001	-0.002	-0.006	-0.01	-0.005	-0.003	0.010
x_{11}	oc	0.003	0.001	-0.009	0	0	0	0	-0.003	-0.008	-0.003	-0.002	0.008
x_{12}	cs	0	-0.002	-0.001	0	0	0	0	-0.002	-0.002	-0.002	-0.002	0.007
x_{13}	np	-0.026	0.004	-0.004	0	0	0.002	-0.001	-0.001	-0.004	-0.001	-0.002	-0.016
x_{14}	o_1	-0.002	-0.003	0.233	0	0.001	0.001	-0.001	0.009	0.014	0.011	-0.006	0.024
x_{15}	c = 2	0	0	-0.009	0	0	0	-0.001	-0.002	-0.009	-0.002	-0.006	0.022
x_{16}	hn	-0.003	-0.007	-0.014	-0.027	0	0	0	-0.005	-0.006	-0.006	-0.002	0.010
x_{17}	c3	0.011	0.067	-0.123	0.079	0.033	0.055	-0.007	-0.044	-0.036	-0.05	-0.018	0.033
x_{18}	oh	0	0	-0.008	0	0	0	0	0	0	0	0	0.009
x_{19}	cpc	0	0	0	0	0	0	0	0	0	0	0	0.002
x_{20}	c = 1	0	0.004	0	0	0	0	0	-0.002	-0.004	-0.002	0.001	0.015
x_{21}	c3oe	0	0	0	0	0	0	0	0	-0.002	-0.001	0	0.004
x_{22}	nh	0	0.004	-0.002	-0.036	0.004	0	-0.003	-0.002	-0.006	-0.003	-0.002	0.023
x_{23}	c0	-0.013	0.281	-0.004	-0.008	0	0.002	-0.005	-0.002	-0.096	-0.002	-0.002	0.012
x_{24}	c5	-0.485	0.004	-0.005	0.004	0.001	0.001	-0.002	-0.008	-0.013	-0.008	-0.003	-0.018
x_{25}	cp	0.228	0.021	0.034	0.045	-0.016	0.157	0.007	0.27	0.16	0.22	-0.072	-0.040
x_{26}	c_0	0.002	0.003	0.168	0	0.002	0.004	0	0.007	0.011	0.008	-0.003	0.020
x_{27}	c=	-0.001	-0.002	0	0	0	0	-0.002	-0.002	-0.005	-0.003	0.003	0.026
x_{28}	c5h1	-0.002	0.009	-0.002	0.002	-0.001	0.003	-0.394	-0.004	-0.014	-0.005	0.004	0.023
x_{29}	c1	-0.025	0.034	-0.029	0.035	0.004	0.003	-0.002	-0.008	-0.043	-0.009	-0.026	0.041
x_{30}	o_2	0	0	0	0	0	0	0	0	-0.001	-0.001	0	0.005
x_{31}	Mw	0.025	0.112	0.018	0.68	-0.011	0.009	0.025	-0.005	-0.01	-0.006	-0.027	-0.045
x_{32}	T	0.066	0.335	0	0.054	0.854	-0.312	0.348	-0.411	-0.448	-0.442	0.08	0.392

Features x_5 , x_{17} , and x_{29} , corresponding to atom types “c2”, “c3”, and “c1” carbons in branches, have over 12 % influence. Their effects vary due to saturation degrees: higher saturation (e.g., “c3” carbons) generally decreases density, while lower saturation (e.g., “c1” carbons) increases it. This matches experimental observations where more reduced hydrocarbons have lower densities, provided their morphology remains amorphous. This is because the presence of aliphatic chains, especially rich in “c3” and “c2” carbons, introduce steric hindrance, reducing molecular compaction and thereby, the density of the material.

Densities range from $\sim 700 \text{ kg/m}^3$ to $\sim 1400 \text{ kg/m}^3$, mostly around 1000 kg/m^3 . Saturates and Rejuvenators have the lowest densities ($\sim 900 \text{ kg/m}^3$), while some resins can reach 1400 kg/m^3 , consistent with experimental and computational values in Table 4. To create bituminous materials with low densities (below 900 kg/m^3), resembling light oils typical of Saturates and Rejuvenators, the saturation degree must be maximized (feature x_1 of atom type “hc” at ~ 0.67), aromatic content minimized (x_{25} of atom type “cp” at 0), aliphatic nature maximized (x_5 at 0.20 and x_{17} at 0.06), and molecular mass kept low (x_{32} between 200 and 500 g/mol). If high molecular masses are required ($>1000 \text{ g/mol}$), then most of the mass contribution should come from “c2” atom types, a condition only attainable when aliphatic chains are long (e.g., longer than 5 carbons). Conversely, if high density bitumens are required (exceeding 1050 kg/m^3), x_1 should be around 0.33, x_{25} at 0.52, x_5 and x_{17} at 0, and molecular mass should be above 500 g/mol .

While other features have a low impact, accounting for less than 5 % influence, the presence of heteroatoms (e.g., atom of types “sp”, “o=”, “oc”, or “s”) increases density slightly. This aligns with observations that fresh bitumens have densities around 990 kg/m^3 , while sulfur-rich and/or aged bitumens reach upwards of 1070 kg/m^3 [27]. This can be attributed to the polarization/electron delocalization effects that these functional groups bring, especially when located in polycyclic aromatic regions of a molecule. The feature values (excluding temperature)

necessary to build bitumens nearing high or low densities are found in Table 11.

Per definition, the most influential feature on the molar volume is the molecular mass (x_{31}), accounting for over 60 % of the influence. Excluding the impact of temperature and molecular mass, the remaining 25 % is attributed to the same factors affecting density. The molar volumes range from $0.22 \cdot 10^{-3} \frac{\text{m}^3}{\text{mol}}$ to $0.92 \cdot 10^{-3} \frac{\text{m}^3}{\text{mol}}$, with an average of about $0.51 \cdot 10^{-3} \frac{\text{m}^3}{\text{mol}}$. These values are consistent with observations of heavy oil mixtures, bitumens, and other derivatives [59]. Therefore, to control the molar volume of bitumens, the same factors influencing density should be considered. However, some features must be carefully selected as they may result in contrasting influences (e.g., adding long branches increases molecular mass while decreasing density). The values of the features (excluding temperature) necessary to build bitumens nearing these extremes are found in Table 11.

4.5. Accessible volume (V_a)

While the results for Accessible Volume are closely related to those for Density, the order, magnitude, and directionality of the influential factors differ considerably. Over 85 % of the influence on Accessible Volume is attributed to temperature (feature x_{32}) alone. The next five most influential features account for 14 % of the influence, with feature x_1 (content of hydrogens connected to carbons, atom type “hc”) being the most significant among these. The other influential features are related to molecular mass and aromatic carbons. This indicates that an increase in temperature significantly increases the empty space accessible to other molecules, while higher molecular mass, highly saturated, and branched molecules result in lower accessible space.

The accessible volumes for the models in this study range from 9 % to 35 %, with the majority of the samples ranging between 12 % and 20 %.

Table 14Combinations of the top 5 most influential features (x^1 being the most influential) that yield a near maximum or a minimum of the properties y_1 through y_{12} .

Property (y_i)	Case	Value	x^1	x^2	x^3	x^4	x^5
y_1 E_p [kJ/kg]	max	4310	x_{23}	0.1039	x_{24}	0.3414	x_4
	min	−334	x_{23}	0.008684	x_{24}	0.01183	x_4
y_2 E_{vdw} [kJ/kg]	max	3.672	x_{22}	0.04157	x_{30}	903.8	x_{16}
	min	−305.2	x_{22}	0.02483	x_{30}	321.6	x_{16}
y_3 E_{coul} [kJ/kg]	max	333.4	x_{13}	0.04971	x_{25}	0.03143	x_{16}
	min	−425.6	x_{13}	0.01669	x_{25}	0.01734	x_{16}
y_4 V_m [m ³ /mol]	max	0.00092	x_{30}	878.4	x_{16}	0.008917	x_{24}
	min	0.000223	x_{30}	297.4	x_{16}	0.006336	x_{24}
y_5 V_a [:]	max	0.2118	x_0	0.6272	x_{16}	0.06815	x_{24}
	min	0.1337	x_0	0.4001	x_{16}	0.004513	x_{24}
y_6 ρ [kg/m ³]	max	1376	x_0	0.3147	x_{24}	0.5231	x_4
	min	811.8	x_0	0.673	x_{24}	0.05581	x_4
y_7 D_{self} [m ² /s] (10 ⁹)	max	5.69	x_{27}	0.001244	x_8	0.000115	x_0
	min	0.5979	x_{27}	0.001709	x_8	0.000996	x_0
y_8 CED [kJ/m ³]	max	6.12E+05	x_{24}	0.5299	x_0	0.313	x_4
	min	2.47E+05	x_{24}	0.01938	x_0	0.6714	x_4
y_9 ΔH_{vap} [kJ/kg]	max	440.5	x_{24}	0.5185	x_{22}	0.02756	x_{28}
	min	257.6	x_{24}	0.03386	x_{22}	0.05407	x_{28}
y_{10} δ_{sol} [kJ ^{0.5} /m ^{1.5}]	max	782.3	x_{24}	0.5128	x_0	0.3136	x_4
	min	504.6	x_{24}	0.05064	x_0	0.6724	x_4
y_{11} C_p [kJ/kg/K]	max	3.01	x_0	0.6732	x_{24}	0.002845	x_4
	min	1.82	x_0	0.313	x_{24}	0.5493	x_4
y_{12} β [1/K]	max	1.00E-03	x_4	0.0736	x_0	0.4142	x_{30}
	min	1.05E-04	x_4	0.2516	x_0	0.6723	x_{30}

To design bitumen with maximized accessible space (where V_r exceeds 20 %), excluding the effect of temperature, the content of hydrogens connected to carbons should be increased ($x_1 > 0.60$), the content of sp^3 hybridized carbons in chains maximized ($x_{17} > 0.06$, $x_5 > 0.20$), and molecular mass minimized ($x_{31} < 300$ g/mol). Conversely, to achieve low accessible space (where y_5 is below 10 %), the opposite adjustments are necessary. Tuning the accessible space is crucial in characterizing and designing bituminous materials, as phenomena like aging, weathering, or rejuvenation depend significantly on the Fickian diffusive coefficient of other chemical compounds, which increases with accessible space [27].

The impact of heteroatoms in functional groups, such as phenols, sulfoxides, pyridines, etc., on the accessible volume is negligible. This is counterintuitive since functional groups significantly influence the stability of molecular conformations and are expected to affect volumetric properties. Typically, the presence of these functional groups increases material density, suggesting a decrease in accessible volume. However, the increased polarity introduced by these groups can lead to complex interactions, such as the formation of soft structures (e.g., hydrogen bonds), which can mitigate the expected decrease in accessible volume. Therefore, the overall influence of functional groups on accessible volume is context-dependent, requiring detailed analysis of the specific chemical structure and physical conditions of the system. The values of the features (excluding temperature) necessary to build bitumens nearing maximum and minimum accessible volumes are found in Table 11.

4.6. Self-diffusion coefficient (D_{self})

The top five most important features affecting the self-diffusion coefficient (D_{self}) account for over 95 % of the influence. Notably, Feature x_{28} , corresponding to the content of 5-membered ring carbons (atom type “c5h”), has a significant negative impact on D_{self} , with a score of almost −40 %. This indicates that a small increase in the content of cyclopentanes or 5-membered aromatic rings can substantially decrease the relative motion of the molecules. The substantial decrease in potential energy that features “c5” and “c5h” atom types bring, as explained in 4.1 Potential Energy, suggests that these features stabilize the conformation of the molecules, reducing their mobility by potentially decreasing strain and vibratory effects associated with unstable ring conformations.

The second most influential feature, x_{32} (temperature), has a positive impact of + 35 %. This is expected, as an increase in temperature decreases density, increases accessible space between atoms, and provides higher kinetic energy to the particles. It is notable that temperature is not the most influential feature, indicating that certain chemical traits can take precedence over temperature in determining the self-diffusive potential of molecules. Feature x_9 , corresponding to atom type “na1”, has a similar effect to Feature x_{28} but to a lesser extent (−13.8 %). This suggests that diffusion is particularly sensitive to the presence of heteroatomic functional groups and the features corresponding to such atom types.

Additionally, molecular mass and x_0 (“hc” atom type content) play smaller, yet positive roles (+2.5 % and + 5 % respectively). While an increase in molecular mass is often associated with decreased mobility of molecules, in bitumens, lower molecular mass hydrocarbons often correspond to more branched molecules, which restrict their diffusive motion through other molecules. Consequently, the influence of feature x_0 is slightly higher and takes precedence over molecular mass when determining the self-diffusion of molecules rich in “hc” hydrogens (i.e., reduced, aliphatic, saturated, and branched characteristics in hydrocarbons).

The range of self-diffusion coefficients spans from $5.0 \cdot 10^{-10}$ to $1.4 \cdot 10^{-8} \text{ m}^2/\text{s}$, with most bitumens having a diffusion coefficient around $5.0 \cdot 10^{-9} \text{ m}^2/\text{s}$. These ranges compare well with those found in simulated and experimental environments in the literature [60].

To design a bitumen with low self-diffusive behavior (an order of magnitude lower than the average), excluding temperature effects, one would maximize the content of 5-membered rings (e.g., $x_{28} > 0.0017$), reduce quinolinic nitrogen (x_9) below 0.001, maintain hydrogen-carbon (x_1) below 0.60, have a molecular mass above 450 g/mol, and ensure a reduction of long aliphatic chains rich in $-\text{CH}_2-$ carbons (e.g., x_5 below 0.06). The values of the features (excluding temperature) necessary to construct bitumens nearing these extremes are detailed in Table 11.

The dependence of self-diffusion coefficients on temperature is known to follow an Arrhenius model, and thus, is fundamentally expected to increase exponentially with temperature [27]. With all other features unchanged, the predicted values of D_{self} from a temperature sweep test ranging from -60 to 200 °C in 10-degree steps also follow an exponential trend, confirming the realistic prediction potential of the MLM given the influence of temperature.

4.7. Cohesive energy density (CED) and enthalpy of vaporization (ΔH_{vap})

The top five most influential factors on the CED and Heat of Vaporization (ΔH_{vap}) are largely attributed to the same features: x_1 , x_5 , x_{17} , x_{23} , x_{25} , x_{29} , and x_{32} , corresponding to atom types “hc”, “c2”, “c3”, “c0”, “cp”, “c1”, and temperature respectively. Together, these factors account for over 90 % of the influence. Similar to other properties, these groups balance hydrogen content (x_1 “hc”), the presence of aliphatic branches (x_5 “c2”), x_{17} “c3”, x_{23} “c0”, x_{29} “c1”), aromaticity (x_{25} “cp”), and temperature (x_{32}) dependence. An increase in temperature (−41 %) and an increase in saturated features (−25 %) significantly decrease both CED and ΔH_{vap} , while the addition of aromatic features results in a substantial increase in both (~21 %). This is expected, as CED and ΔH_{vap} measure the energy required to separate molecules to an infinite distance (essentially their non-bonded potential energies). Increasing temperature, saturation features, and reducing intermolecular electron delocalization effects reduce intermolecular forces, their non-bonded energy, and thus the CED and ΔH_{vap} .

While both CED and ΔH_{vap} are influenced by mostly the same factors, their order and magnitude of influence differ. This difference arises because ΔH_{vap} considers the energy differences (as work) done by the barostat to maintain constant pressure throughout the simulation by adjusting the simulation’s volume. Consequently, features more influential on the material’s volumetrics gain precedence. Highly polar groups (such as those from heteroatoms) cumulatively impact ΔH_{vap} significantly, up to + 10 %, a trend not seen in the case of the CED. This trend aligns well with observations, as bitumens rich in heteroatomic polar groups result of severe aging conditions are stiffer at much higher temperatures.

The range of ΔH_{vap} extends from 150 kJ/kg to 500 kJ/kg, with most bitumens hovering around 300 kJ/kg. These values compare well to those observed for heavy crude oils, bitumens, and oil derivatives of similar molecular structures, as observed in Table 9. To design a bitumen with maximized CED, aiming for high cohesion at elevated temperatures (excluding the temperature feature), the aromatic content

should be increased (e.g., setting feature x_{25} to 0.52) and saturated aliphatic branches minimized (x_0 to < 0.3 , x_5 to < 0.015 , and x_{17} to < 0.0013). Additionally, maximizing the content of the heteroatom type “o_1” to 0.022 (compared to an average of 0.010) further enhances CED. The addition of oxidative properties (“o_1” type) to increase CED aligns well with observation in real bituminous samples that have been severely aged. This combination of features yields a molecular system with a CED three times higher than the average, as demonstrated in Table 11, and showcases how high aromaticity, high hybridization, and strong polarity and electron delocalization effects increase the CED. A similar approach can be applied when tuning ΔH_{vap} , as shown in Table 11.

4.8. Solubility parameter (δ_{sol})

The solubility parameter, fundamentally a function of the CED and V_m , provides valuable insight into the strength of intermolecular interactions and the ability of different systems to mix. This parameter is critical for assessing whether mixtures of molecules will form a homogeneous phase upon mixing [37] and thus can be used to tune the stability of a mixture of different compounds.

When examining bitumens, smaller differences in solubility parameters can be advantageous as it allows materials to be miscible, potentially leading to more stable mixtures. Conversely, higher differences in solubility parameters may be necessary when the goal is to maintain intricate morphological characteristics intentionally incorporated into the bitumen, such as those found in Styrene-Butadiene polymer-modified bitumens, where the dissolution of Styrene or Butadiene into its surrounding media (i.e., bitumen) can detrimentally affect the bitumen’s mechanical properties [12].

The most influential factor in the solubility parameter is the system’s temperature, which has a negative influence of −44.2 %. This is expected because an increase in temperature typically decreases the CED of a material, assuming a negligible decrease in molar volume, and thus increases the mixing potential of different compounds, as fundamentally expected [37].

Similar to other properties, much of the remaining influence is attributed to oxidative characteristics (e.g., atom of types “cp”, “o_1”, “c1”) typical of 6- and 5-membered aromatic rings, highly polar functional groups, and sp1 and sp0 hybridized carbons, collectively accounting for over + 25 % of influence. Conversely, highly reduced characteristics, such as atoms of types “hc”, “c2”, “c3”, “c5”, and “c5h”, contribute negatively with over −25 % influence.

To design a system with a maximized or minimized δ_{sol} (excluding the temperature influence), one should consider the oxidative and reduced characteristics. Maximizing the content of aromatic rings and polar functional groups, while minimizing saturated aliphatic branches and highly reduced features, will significantly affect the solubility parameter. This aligns with fundamental chemistry rules, where compounds of similar polarity are miscible.

The solubility parameters measured in this study range from 350 to 800 $\text{kJ}^{0.5}/\text{m}^{1.5}$, which correspond well with experimentally obtained values of 474 to 758 $\text{kJ}^{0.5}/\text{m}^{1.5}$ for a variety of oil samples, reported in Table 9. Moreover, the trends presented align with the findings of Ramos et al., where oil with higher saturation degrees have a lower solubility parameter. The range described, while pertaining solely to bituminous materials, encapsulates nearly the entirety of solubility parameters observed across multiple petroleum blends. This encompasses a spectrum from lighter to heavier compounds, incorporating those distilled hydrocarbons absent in bitumen (e.g., benzene, ethanol, etc.) [61]. Thus, despite the apparent nominal variance, an increase from 350 to 800 $\text{kJ}^{0.5}/\text{m}^{1.5}$ captures the miscibility potential of a diverse array of hydrocarbons, ensuring that the range studied is well suited for studies involving bituminous materials and other hydrocarbon mixtures. As a reference, the solvent used in SARA fractionation tests to separate the Saturates fraction away from other fractions, n-heptane, has a δ_{sol} equal

to $471\text{kJ}^{0.5}\text{m}^{-1.5}$, further emphasizing how δ_{sol} can be used to tweak the miscibility potential of different organic compounds using the MD and the MLM data obtained in this study. Detailed influential design parameters for optimizing the solubility parameter, to a maximum or to a minimum, are provided in Table 11.

4.9. Specific heat (C_p)

Feature x_1 , which corresponds to the hydrogens bonded to carbon atoms (and to some degree, the saturation degree), is by far the most influential factor, with a score of + 67 %. The remaining four most influential features are considerably less impactful, each contributing around 5 %. These features are related to the observation in feature x_1 , where an increase in the presence of carbon atoms in highly saturated branches (features x_5 , x_{17} , and x_{29} of types “c2”, “c3”, and “c1” respectively) and a decrease in the presence of aromatic rings (features x_{24} and x_{25} for types “c5” and “cp”, respectively) result in a higher specific heat capacity. An increase in the system’s temperature (feature x_{32}) results in an increase in heat capacity, albeit its impact is around + 8.3 %, which is less than expected. However, this observation aligns well with the fundamentals of energetics, temperature, and heat capacities [6]. An increase in molecular mass results in a mild decrease in heat capacity, which is a generally accepted trend in organic chemistry, especially in bituminous molecules.

The higher specific heat capacities of more hydrogenated hydrocarbons are due to the increased degrees of freedom in their molecular motions. These hydrocarbons have more vibrational, rotational, and torsional modes, allowing them to distribute energy across additional modes of motion. In contrast, aromatic compounds have rigid molecular structures with fewer degrees of freedom. This rigidity restricts energy distribution to fewer modes, leading to greater temperature changes and thus lower specific heat capacities.

The heat capacities measured in this study range from as low as 1.1kJ/kg/K to as high as 3.4kJ/kg/K , while most bitumens heat capacity hover at around 2.2kJ/kg/K . These values compare well with previous findings, reported in Table 8. Moreover, these findings also indicate that heat capacity is primarily influenced by the presence of hydrogen-carbon and hydrogen-hydrogen interactions, which aligns well with the findings of this study, further validating the realism of the MD simulations and the prediction capacity of MLMs. This indicates that the Saturates and Rejuvenators fractions have the highest heat capacities, while the Aromatics and Resins fraction are expected to have the lowest heat capacities. In addition, most oil derivatives are observed to have heat capacities ranging from 2.0 to 4.5kJ/kg/K , ensuring that the range of heat capacities studied is well suited for use in other studies.

To maximize the heat capacity of bitumens, the saturation degree (and thus, all features heavily influencing this metric) must be maximized. Specifically, this involves setting x_1 above 0.67, x_5 above 0.10, and maintaining x_{31} around 650 g/mol. The most influential design parameters for creating a bitumen sample with either maximized or minimized heat capacity (excluding the influence of temperature) are detailed in Table 11.

4.10. Thermal expansion coefficient (β)

Similar to previous cases, the most influential factors on the thermal expansion coefficient (β) include temperature (+39.2 %), aliphatic carbons (+15 % for c1, c2, and c3), hydrogen bonded to carbons (+6.5 %), molecular mass (−5%), and aromatic features (−4%). The observed increase in β due to the addition of more reduced hydrocarbon structures can be explained by these features having less cohesive forces (favoring expansion) while capturing relatively large amounts of heat (as indicated by their presence bringing a significant increase in heat capacity), resulting in thermal expansion coefficients.

The thermal expansion coefficients in the molecular models of this study range from $2.25 \cdot 10^{-6}\text{ 1/K}$ to as high as $2.52 \cdot 10^{-3}\text{ 1/K}$, averaging around $5.04 \cdot 10^{-4}\text{ 1/K}$. These values compare well to those observed for heavy oils and conventional bitumens. To design bitumens with a relatively low thermal expansion coefficient ($< 1 \cdot 10^{-4}\text{ 1/K}$), aiming for a material that remains dimensionally stable under varying temperatures, the content of reduced hydrocarbon features should be minimized, and the presence of oxidized features should be increased, including polycyclic aromatics and oxidized functional groups (such as sulfoxides and carboxyl groups). In essence, this involves reducing features that lower densities while favoring an increase in heat capacity. The values of the features (excluding temperature) necessary to construct bitumens nearing these extremes are detailed in Table 11.

5. Case studies

5.1. Rejuvenator selection

In this example, the suitability of the four rejuvenators listed in Table 2 is evaluated for the rejuvenation of the most oxidized and sulfur-rich bitumen studied, specifically the “F” bitumen type aged to degree 4, referred to as F400. For simplicity, this evaluation is limited to a temperature of 60°C . The analysis involves predicting the physical properties of the bitumen sample and of each rejuvenator, and then selecting the rejuvenator that best matches the bitumen’s properties, with priority given to matching solubility parameters (δ_{sol}) and maximizing the self-diffusion coefficient (D_{self}) between the bitumen and rejuvenator.

Bitumen sample F400 has already been prepared, and its atom type formula is available in the [supplementary file atom_types_formula.xlsx](#). The same applies to the rejuvenators, whose atom type formulas are provided in the same file. Fig. 1 illustrates how Aromatic Oil appears when its atom types are assigned, and its atom type formula is generated. These are fed to the MLMs to predict the mixtures’ properties of Table 8. All 12 properties are generated, taking about 70 ms per chemical system on a personal computer’s CPU. The predicted properties for F400 and all four rejuvenators are presented in Table 15.

Aromatic Oil emerges as the most suitable rejuvenator for mixing with bitumen F400 due to its high value ($D_{\text{self}} = 4.15$) and a solubility parameter ($\delta_{\text{sol}} = 568.7$) that closely matches F400’s ($\delta_{\text{sol}} = 559.6$), ensuring both fast diffusion and miscibility. While Vegetable Oil has a higher diffusion coefficient ($D_{\text{self}} = 6.77$), which could facilitate faster mixing, its solubility parameter ($\delta_{\text{sol}} = 536.1$) is less aligned with F400, potentially leading to reduced miscibility and stability. Thus, for a rejuvenator that diffuses and mixes well, Aromatic Oil is the better choice, though Vegetable Oil might be considered if rapid diffusion is prioritized over miscibility.

Deepening the designing concept, small but artificial modifications can be introduced into the molecular structure of Aromatic Oil to increase the density of the rejuvenator (which is currently slightly below that of water and substantially lower than that of F400 at 60°C), while retaining the solubility parameter as close as possible to that of F400. Following the guidelines presented in Section 4.4 for Density and Section 4.8 for Solubility Parameter, a single modification can achieve the target properties needed – by converting the only cyclohexane ring present in the Aromatic Oil structure into an aromatic ring. This conversion is depicted in Fig. 1, and its changes are reflected in a new “atom type” formula.

Upon feeding the Modified Aromatic Oil’s features into the MLMs, most of the physical properties remain close if not closer to those of F400, while the density is increased to remain just above that of water – at 1004 kg/m^3 at 60°C – as seen in Table 15.

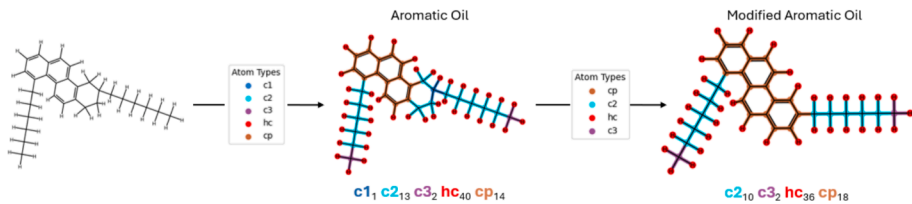


Fig. 1. Identification of the atomic types in both Aromatic and Modified Aromatic Oils used to generate the atom type formula required by the MLMs of this study to predict material properties.

Table 15
Predicted physical properties for F400 bitumen and all four rejuvenators, including the artificially modified Aromatic Oil.

Feature	Property	Bitumen	Rejuvenator Oil				
		F400	Engine	Vegetable	Aromatic	Modified Aromatic Oil	Naphthenic
y ₁	E _p [kJ/kg]	2361.0	−144.1	2.4	2302.9	2388.8	−121.5
y ₂	E _{vdw} [kJ/kg]	−161.6	−246.4	−250.4	−164.7	−161.5	−202.1
y ₃	E _{Coal} [kJ/kg]	8.02	3.30	23.32	43.56	35.6	−23.81
y ₄	V _m [m ³ /mol]	0.001	0.000	0.000	0.000	0.00039	0.000
y ₅	V _a [·]	0.188	0.223	0.219	0.177	0.169	0.211
y ₆	ρ [kg/m ³]	1021.5	805.05	849.6	968.5	1004.6	868.7
y ₇	D _{self} [m ² /s] (10 ⁹)	2.46	6.18	6.77	4.15	3.95	5.53
y ₈	CED [kJ/m ³]	3.14	2.51	2.90E	3.24E	322	2.48
y ₉	ΔH _{vap} [kJ/kg]	315.8	312.0	340.7	336.8	334.2	293.8
y ₁₀	δ _{sol} [kJ ^{0.5} /m ^{1.5}]	559.6	500.0	536.1	568.7	566.7	497.8
y ₁₁	C _p [kJ/kg/K]	1.91	2.70	2.51	2.15	1.87	2.38
y ₁₂	β [1/K]	0.00019	0.00098	0.00097	0.00036	0.00038	0.00074

5.2. Temperature Susceptibility

In this example, the impact of aging conditions—ranging from 0 (fresh) to 4 (fully aged)—and temperatures of 25 °C, 120 °C, and 200 °C on the density (ρ) and heat of vaporization (ΔH_{vap}) of bitumen of type “T” is calculated. This analysis highlights the ability of the MLMs to predict both chemical and environmental changes typical of bituminous experiments.

To perform this analysis, previously prepared bitumen models from this study—T000, T100, T200, T300, and T400, representing non-rejuvenated bitumens with increasing aging levels—are utilized. For each bitumen sample, the corresponding “atom type” formulas and molecular masses are input into the MLMs, and descriptor x_{32} – temperature – is set to 25 °C, 120 °C, and 200 °C. The predictions are summarized in Table 16.

The MLM predictions reveal that increasing levels of aging lead to a rise in bitumen density, where the values align well with experimental findings on aged bituminous samples. As anticipated, density decreases with increasing temperature, in agreement with fundamental observations. Similar trends are observed in the computed ΔH_{vap} , where more energy is required to vaporize highly aged samples. As temperature increases, the energy required for vaporization decreases, consistent with both experimental observations and fundamental principles.

Table 16
Predicted densities (ρ) and heat of vaporizations (ΔH_{vap}) for bitumens of type “T” with increasing levels of aging.

Bitumen Type	Property	ρ [kg/m ³]			ΔH_{vap} [kJ/kg]		
	Temperature [°C]	25	120	200	25	120	200
T000		1010.5	960.5	908.2	330.5	304.6	278.1
T100		1014.9	963.2	916.1	337.4	300.4	280.5
T200		1022.6	976.7	932.4	339.9	308.5	282.7
T300		1045.1	1011.7	977.01	342.1	317.6	297.0
T400		1088.4	1054.8	1020.9	347.7	329.1	305.8

5.3. Binary mixture

In this example, two different compounds—Squalane and Phenolic Asphaltene from Table 1—are combined to optimize two contrasting physical properties: density (ρ) and heat of vaporization (ΔH_{vap}). No molecular structure modifications are applied; instead, the objective is to determine the appropriate ratio of each molecule in the mixture to maximize the value of both ρ and ΔH_{vap} .

The SMILES strings for Squalane and Phenolic Asphaltene are provided in the file molecules_selection.docx in the Supplementary Information. These strings are used to generate the atom type formula for each molecule and aggregate them into a mixture-wide atomic formula. This process is depicted in Fig. 2. To this end, ten mixtures, limited to 100 molecules each, are created with increasing amounts of Squalane. At one extreme, mixture “S0” is composed entirely of 100 Phenolic Asphaltene molecules, while at the other extreme, mixture “S100” consists entirely of 100 Squalane molecules. Both densities and heat of vaporizations are reported for each mixture in Table 17.

Table 17 showcases that beyond Mixture “S40” (40 Squalane and 60 Asphaltene molecules), ΔH_{vap} no longer decreases. In contrast, the ρ continues to decrease with increasing Squalane content. As a result, mixtures ranging from “S40” to “S50” maintain both ρ and ΔH_{vap} at relatively high levels, indicating that a binary mixture with a similar number proportion of both components is effective. This translates to a mixture of approximately 30 g of Squalane and 70 g of Phenolic Asphaltene.

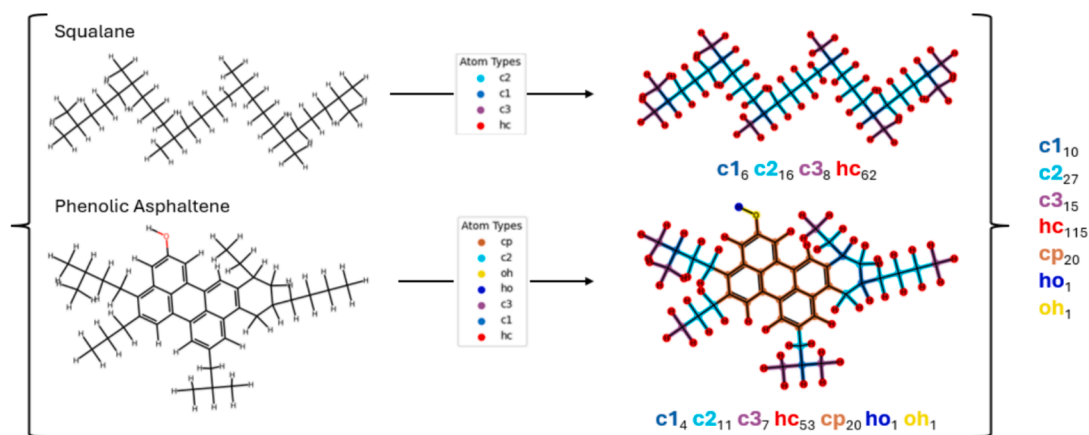


Fig. 2. When constructing a molecular mixture, the types of atoms within each molecule are quantified and consolidated into a single formula. This formula represents the overall chemical composition of the mixture.

Table 17

Densities and heat of vaporizations predicted for all 10 mixtures (“S0” through “S100”). Bolded are the mixtures whose densities and heat of vaporizations are both at a maximum.

Mixture	ρ [kg/m ³]	ΔH_{vap} [kJ/kg]	Mixture	ρ [kg/m ³]	ΔH_{vap} [kJ/kg]
S0	1004.9	325.8	S60	939.9	303.3
S10	1006.	324.2	S70	937.8	304.2
S20	993.5	323.6	S80	925.3	303.6
S30	980.7	320.3	S90	835.2	303.8
S40	974.4	306.4	S100	812.3	303.5
S50	945.1	304.0	—	—	—

These case studies, though simplistic and open to interpretation, showcase how the MLMs of this study can be used to near-instantaneously obtain preliminary property estimates, potentially reducing the time and resources required to arrive at similar conclusions.

6. Conclusion

The transition from using broad chemical descriptors, such as the SARA fractions and elemental compositions, to more fundamental chemical descriptors in MD models of bitumen significantly improves the ability to capture the impact of different chemical characteristics. This approach not only simplifies the process but also aids in the fundamental characterization, preparation, and customization of future bituminous models. Each feature’s impact on various properties—whether fundamental, thermodynamic, or kinetic—can be quantified and adjusted to achieve desired material characteristics. The use of force field atom types is straightforward, as MD methods already discretize atomistic systems based on relevant chemical characteristics, reducing ambiguity in feature selection due to their universal and fundamental nature.

In this study, 30 chemical descriptors, each corresponding to a unique atom type in the PCFF force field, were examined. It was found that most material properties are influenced by at most 10 features, which often overlap across properties, accounting for more than 95 % of the influence. These features include temperature, molecular mass, hydrogen content, aromaticity, the presence of 5-membered rings, and highly polar heteroatomic functional groups. Balancing these features carefully would have the most significant impact on tuning the properties studied.

Over 193 models and 1770 separate LAMMPS runs were performed to obtain the various properties of this study. These models were carefully crafted to represent different bitumen types, aging conditions, rejuvenator types, and rejuvenation dosages, resulting in a

comprehensive array of bituminous materials. This wide range of features and properties enhanced the predictive capacity of the MLMs developed, which were trained to predict all the properties presented in this study based on input features. This approach provides an effective and efficient method to abstract away from running continuous MD simulations with small changes in molecular structures, offering researchers quick insights into the impact of certain feature changes on a material’s response.

This study focused on 12 material properties, including fundamental MD properties (e.g., potential energy), volumetric properties (e.g., density), thermodynamic properties (e.g., heat capacity), and kinetic properties (e.g., diffusion coefficients). It was found that interpreting the impact of fundamental MD properties on material behavior is challenging and often non-intuitive, with multiple directionality trends. However, these properties are essential for evaluating the numerical stability and validity of the simulations and the conformational stability of designed molecules, making them valuable for use in molecular design fields. More relatable properties, such as density and heat capacity, provide easier paths to interpret physical trends and are easier to capture due to their continuous relationships with the study’s features, many of which have been experimentally assessed in related petroleum studies.

While the MLMs developed in this study can address the impact of specific molecular features on bitumen properties, establishing a new solid foundation for using MD simulations and experimental techniques in bitumen design, future work should capture more relevant properties, particularly those important in rheological or mechanical studies (e.g., dynamic properties and viscosities). Furthermore, using force field atom types as features requires consistent use of similar force fields between simulations, as different force fields may use varying rules for identifying atom types.

Additionally, this study employed a simple tree model with interpolative capabilities, which worked well given the extensive array of models covering a condensed range of organic chemistry features. Future work should include a wider array of compositional features, study more extreme cases (e.g., very high content of polar groups), and use MLMs better suited for predicting properties given a less comprehensive set of training cases.

Moreover, this study assumes all MD models are fully amorphous, with no distinct morphological features in the bituminous samples. Bitumens are known to undergo phase separation and form complex, highly heterogeneous intermolecular features, heavily affecting their mechanical and rheological properties. Future MLMs should include a more complex set of features to better characterize the true state of bituminous materials, given that newer, synthetic bitumens often involve the use of intermolecular features with highly heterogeneous morphologies.

7. Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used OpenAI's ChatGPT4 to shorten the length of certain sections. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Eli I. Assaf: Writing – original draft, Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xueyan Liu:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Peng Lin:** Project administration, Formal analysis, Data curation, Conceptualization. **Shisong Ren:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Sandra Erkens:** Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data related to the chemical characterization of bitumen samples and the properties assessed through MD simulations are currently restricted as they are integral to an active research project. The [Supplementary Information](#), however, provides a subset of this data, includes essential details sufficient to replicate the simulations conducted, and features the MLMs generated along with their predicted datasets, which are available for further research into new bituminous materials

Acknowledgements

This paper/article is created under the research program Knowledge-based Pavement Engineering (KPE). KPE is a cooperation between the Ministry of Infrastructure and Water Management (Rijkswaterstaat), TNO, and TU Delft in which scientific and applied knowledge is gained about asphalt pavements and which contributes to the aim of Rijkswaterstaat to be completely climate neutral and to work according to the circular principle by 2030. The opinions expressed in these papers are solely from the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.matdes.2024.113327>.

References

- [1] J.C. Nicholls, D. James, Literature review of lower temperature asphalt systems, *Proceedings of the Institution of Civil Engineers-Construction Materials* 166(5) (2013) 276–285.
- [2] L.W. Corbett, Composition of asphalt based on generic fractionation, using solvent deasphalting, elution-adsorption chromatography, and densimetric characterization, *Anal. Chem.* 41 (4) (1969) 576–579.
- [3] Z. Chen, J. Pei, R. Li, F. Xiao, Performance characteristics of asphalt materials based on molecular dynamics simulation—a review, *Constr. Build. Mater.* 189 (2018) 695–710.
- [4] K. Ohno, K. Esfarjani, Y. Kawazoe, *Computational materials science: from ab initio to Monte Carlo methods*, Springer, 2018.
- [5] M.L. Greenfield, Molecular modelling and simulation of asphaltenes and bituminous materials, *Int. J. Pavement Eng.* 12 (4) (2011) 325–341.
- [6] S.A. Hollingsworth, R.O. Dror, Molecular dynamics simulation for all, *Neuron* 99 (6) (2018) 1129–1143.
- [7] D.D. Li, M.L. Greenfield, Viscosity, relaxation time, and dynamics within a model asphalt of larger molecules, *J. Chem. Phys.* 140 (3) (2014).
- [8] D. Lesueur, The colloidal structure of bitumen: consequences on the rheology and on the mechanisms of bitumen modification, *Adv. Colloid Interface Sci.* 145 (1–2) (2009) 42–82.
- [9] S. Ren, X. Liu, P. Lin, Y. Gao, S. Erkens, Molecular dynamics simulation on bulk bitumen systems and its potential connections to macroscale performance: review and discussion, *Fuel* 328 (2022) 125382.
- [10] A.M. Kharat, J. Zacharia, V.J. Cherian, A. Anyatonwu, Issues with Comparing SARA Methodologies, *Energy Fuel* 21 (6) (2007) 3618–3621.
- [11] M.L. Greenfield, L. Zhang, Developing model asphalt systems using molecular simulation: final model, University of Rhode Island, Transportation Center, 2009.
- [12] W. Wu, M.C. Cavalli, W. Jiang, N. Kringos, Differing perspectives on the use of high-content SBS polymer-modified bitumen, *Constr. Build. Mater.* 411 (2024) 134433.
- [13] R.J. Angel, Equations of state, *Rev. Mineral. Geochem.* 41 (1) (2000) 35–59.
- [14] M. Karelson, V.S. Lobanov, A.R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR studies, *Chem. Rev.* 96 (3) (1996) 1027–1044.
- [15] F. Fallah, *Molecular Dynamics Modeling and Simulation of Bitumen Chemical Aging*, (2017).
- [16] S. Ren, X. Liu, Y. Zhang, P. Lin, P. Apostolidis, S. Erkens, M. Li, J. Xu, Multi-scale characterization of lignin modified bitumen using experimental and molecular dynamics simulation methods, *Constr. Build. Mater.* 287 (2021) 123058.
- [17] R. Tian, H. Luo, X. Huang, Y. Zheng, L. Zhu, F. Liu, Correlation Analysis between Mechanical Properties and Fractions Composition of Oil-Rejuvenated Asphalt, *Materials (basel)* 15 (5) (2022).
- [18] D.D. Li, M.L. Greenfield, Chemical compositions of improved model asphalt systems for molecular simulations, *Fuel* 115 (2014) 347–356.
- [19] S. Ren, X. Liu, P. Lin, S. Erkens, Y. Xiao, Chemo-physical characterization and molecular dynamics simulation of long-term aging behaviors of bitumen, *Constr. Build. Mater.* 302 (2021) 124437.
- [20] E. Proserpi, E. Bocci, A review on bitumen aging and rejuvenation chemistry: processes, materials and analyses, *Sustainability* 13 (12) (2021) 6523.
- [21] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci.* 28 (1) (1988) 31–36.
- [22] Q. Shi, J. Wu, Review on sulfur compounds in petroleum and its products: state-of-the-art and perspectives, *Energy Fuel* 35 (18) (2021) 14445–14461.
- [23] S. Vedachalam, N. Baquerizo, A.K. Dalai, Review on impacts of low sulfur regulations on marine fuels and compliance options, *Fuel* 310 (2022) 122243.
- [24] F. Migliori, J.-F. Corté, Comparative study of RTFOT and PAV aging simulation laboratory tests, *Transp. Res. Rec.* 1638 (1) (1998) 56–63.
- [25] H.U. Bahia¹, D.A. Anderson, The Pressure Aging Vessel (PAV): a test to simulate rheological changes due to field aging, *Physical properties of asphalt cement binders* 1241 (1995) 67.
- [26] S. Matolia, G. Guduru, B. Gottumukkala, K.K. Kuna, An investigation into the influence of aging and rejuvenation on surface free energy components and chemical composition of bitumen, *Constr. Build. Mater.* 245 (2020) 118378.
- [27] S. Ren, X. Liu, P. Lin, S. Erkens, Y. Gao, Chemical characterizations and molecular dynamics simulations on different rejuvenators for aged bitumen recycling, *Fuel* 324 (2022) 124550.
- [28] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nat. Methods* 17 (3) (2020) 261–272.
- [29] G. Landrum, Rdkit documentation, Release 1 (1–79) (2013) 4.
- [30] H. Sun, S.J. Mumby, J.R. Maple, A.T. Hagler, An ab initio CFF93 All-Atom Force Field for Polycarbonates, *J. Am. Chem. Soc.* 116 (7) (1994) 2978–2987.
- [31] A.P. Thompson, H.M. Aktulga, R. Berger, D.S. Bolintineanu, W.M. Brown, P.S. Crozier, P.J. In't Veld, A. Kohlmeyer, S.G. Moore, T.D. Nguyen, LAMMPS: a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, *Computer Physics Communications* 271 (2022) 108171.
- [32] M.C. Payne, M.P. Teter, D.C. Allan, T. Arias, a.J. Joannopoulos, Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients, *Reviews of modern physics* 64(4) (1992) 1045.
- [33] D.J. Evans, B.L. Holian, The nose–hoover thermostat, *J. Chem. Phys.* 83 (8) (1985) 4069–4074.
- [34] E.I. Assaf, X. Liu, P. Lin, S. Erkens, SMI2PDB: a self-contained Python tool to generate atomistic systems of organic molecules using their SMILES notations, *Software Impacts* 100655 (2024).
- [35] E.I. Assaf, X. Liu, P. Lin, S. Erkens, PDB2DAT: automating LAMMPS data file generation from PDB molecular systems using Python, Rdkit, and Pysimm, *Software Impacts* 100656 (2024).
- [36] R.O. Rasmussen, R.L. Lytton, G.K. Chang, Method to predict temperature susceptibility of an asphalt binder, *J. Mater. Civ. Eng.* 14 (3) (2002) 246–252.
- [37] D.W. Van Krevelen, K. Te Nijenhuis, Chapter 7 - Cohesive Properties and Solubility, in: D.W. Van Krevelen, K. Te Nijenhuis (Eds.), *Properties of Polymers (fourth Edition)*, Elsevier, Amsterdam, 2009, pp. 189–227.
- [38] F. Ercolessi, A molecular dynamics primer, *Spring college in computational physics, ICTP, Trieste* 19 (1997).
- [39] M. Domin, A. Herod, R. Kandiyoti, J.W. Larsen, M.J. Lazaro, S. Li, P. Rahimi, A Comparative Study of Bitumen Molecular-Weight Distributions, *Energy Fuel* 13 (3) (1999) 552–557.

- [40] W. McKinney, P. Team, Pandas-Powerful python data analysis toolkit, Pandas—Powerful Python Data Analysis Toolkit 1625 (2015).
- [41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: Machine learning in Python, the Journal of machine Learning research 12 (2011) 2825–2830.
- [42] M.R. Segal, Machine learning benchmarks and random forest regression, (2004).
- [43] R.F. Sproull, Refinements to nearest-neighbor searching in k-dimensional trees, *Algorithmica* 6 (1991) 579–589.
- [44] I. Muraina, Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts, 7th International Mardin Artuklu Scientific Research Conference, 2022, pp. 496–504.
- [45] A.C.d.S. Ramos, M.P. Rolemborg, L.G.M.d. Moura, E.L. Zilio, M.d.F.P.d. Santos, G. González, Determination of solubility parameters of oils and prediction of oil compatibility, *Journal of Petroleum Science and Engineering* 102 (2013) 36–40.
- [46] W. Fang, Q. Lei, R. Lin, Enthalpies of vaporization of petroleum fractions from vapor pressure measurements and their correlation along with pure hydrocarbons, *Fluid Phase Equilib.* 205 (1) (2003) 149–161.
- [47] R. Ceriani, R. Gani, Y.A. Liu, Prediction of vapor pressure and heats of vaporization of edible oil/fat compounds by group contribution, *Fluid Phase Equilib.* 337 (2013) 53–59.
- [48] J. Wu, Y. Xu, Effect of asphaltene content in bitumen on thermodynamic properties of light hydrocarbons dissolving in bitumen, *Fluid Phase Equilib.* 490 (2019) 22–32.
- [49] N. Durupt, A. Aoulmi, M. Bouroukba, M. Rogalski, Heat capacities of liquid polycyclic aromatic hydrocarbons, *Thermochim Acta* 260 (1995) 87–94.
- [50] O. Fasina, Z. Colley, Viscosity and specific heat of vegetable oils as a function of temperature: 35 C to 180 C, *Int. J. Food Prop.* 11 (4) (2008) 738–746.
- [51] W.R. Lindberg, R.R. Thomas, R.J. Christensen, Measurements of specific heat, thermal conductivity and thermal diffusivity of Utah tar sands, *Fuel* 64 (1) (1985) 80–85.
- [52] R. Cassis^o, N. Fuller^o, L.G. Hepler, R.J. McLean, A. Skauge, Specific Heat Capacities of Bitumens an Heavy Oils, Reservoir Minerals, Clays, Dehydrated Clays, Asphaltenes, and Cokes.
- [53] H.W. Bearce, E.L. Pepper, Density and thermal expansion of American petroleum oils, US Government Printing Office 1916.
- [54] P.V. Coveney, S. Wan, On the calculation of equilibrium thermodynamic properties from molecular dynamics, *PCCP* 18 (44) (2016) 30236–30240.
- [55] T.M. Krygowski, M.K. Cyrański, Structural aspects of aromaticity, *Chem. Rev.* 101 (5) (2001) 1385–1420.
- [56] O.C. Mullins, E.Y. Sheu, Structures and dynamics of asphaltenes, Springer Science & Business Media 2013.
- [57] E.I. Assaf, X. Liu, P. Lin, S. Erkens, S. Nahar, L.I. Mensink, Studying the impact of phase behavior in the morphology of molecular dynamics models of bitumen, *Mater. Des.* 230 (2023) 111943.
- [58] C. Sagui, T.A. Darden, Molecular dynamics simulations of biomolecules: long-range electrostatic effects, *Annu. Rev. Biophys. Biomol. Struct.* 28 (1) (1999) 155–179.
- [59] E.I. Assaf, X. Liu, P. Lin, S. Erkens, Introducing a force-matched united atom force field to explore larger spatiotemporal domains in molecular dynamics simulations of bitumen, *Mater. Des.* 240 (2024) 112831.
- [60] R. Karlsson, U. Isacsson, J. Ekblad, Rheological characterisation of bitumen diffusion, *J. Mater. Sci.* 42 (2007) 101–108.
- [61] A.F. Barton, CRC handbook of solubility parameters and other cohesion parameters, Routledge, 2017.