

Predicting the diffusion coefficients of rejuvenators into bitumens using molecular dynamics, machine learning, and force field atom types

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

DOI

[10.1016/j.matdes.2024.113502](https://doi.org/10.1016/j.matdes.2024.113502)

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 diffusion coefficients of rejuvenators into bitumens using molecular dynamics, machine learning, and force field atom types. *Materials and Design*, 248, Article 113502. <https://doi.org/10.1016/j.matdes.2024.113502>

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 diffusion coefficients of rejuvenators into bitumens using molecular dynamics, machine learning, and force field atom types

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:

Bitumen
Rejuvenator
Fickian Diffusion
Molecular Dynamics
Machine Learning
Chemical Descriptors

ABSTRACT

This study explores the use of chemical descriptors derived from force field atom types to predict Fickian diffusion coefficients of rejuvenators in bitumen, utilizing machine learning models trained on data from 240 non-equilibrium molecular dynamics simulations. The simulations cover three bitumen types (NO, TO, FO), five aging degrees, and four temperatures (60 °C, 120 °C, 160 °C, 200 °C), capturing diffusion coefficients ranging from 0.0068e-10 m²/s in highly aged bitumens at 60 °C to 4.35e-10 m²/s in fresher samples at 200 °C. The MLM, built with 18 chemical descriptors for bitumen and rejuvenator sides, achieves an R² of 0.97, accurately predicting diffusion across varied conditions. This approach abstracts away from the need for repeated MD simulations, enabling diffusion predictions even for systems outside the original dataset. The manuscript presents three case studies to illustrate how the model can be used for the iterative design of rejuvenators by optimizing molecular structures based on critical chemical features, such as rejuvenator oxygen content, bitumen sulfur content, and molecular weights. It also demonstrates how the model offers a practical framework for understanding the diffusion and performance of rejuvenators by linking time-dependent factors—such as concentration, depth, and rejuvenation time—with the bulk properties of bitumen-rejuvenator systems, facilitating industrial applications.

1. Introduction

Roads constitute the backbone of a functional transport network, and maintaining them in optimal condition is crucial for fostering economic growth and stability. Consequently, significant efforts have been dedicated to improving the performance and durability of asphalt mixtures. Central to these efforts is enhancing the mechanical and rheological properties of bitumen, given its critical role in dictating the physical response of asphalt mixtures to operational demands [1].

Describing bitumen's structure on a molecular scale has proven remarkably challenging [2]. The chemical composition of bitumen varies significantly based on its source, refining processes, and characterization methods, resulting in a material with an uncertain formal definition [3]. Furthermore, the atomic structures of many molecules within bitumen remain disputed, as bitumen comprises millions of diverse organic molecules varying widely in aromaticity, saturation, polarity, and size [4].

Experimental attempts to chemically characterize bitumen often involve fractionating its molecules based on solubility [5]. This

approach has evolved into the widely-used SARA (Saturates, Aromatics, Resins, Asphaltenes) fractionation technique, which describes the chemical composition of bitumen samples without requiring fundamental molecular identification [6]. Understanding the interactions between these chemically distinct fractions has revealed that they form physically separated regions, developing into complex molecular arrangements. These arrangements impart unique morphologies to each bitumen sample, significantly affecting their mechanical and rheological responses, even when chemical compositions remain constant [7]. These structural variations span multiple size domains, presenting significant multiscale characterization challenges not commonly found in materials such as metals or polymers [8].

Bitumen's intricate structure, both at the molecular and microstructural levels, is also heavily influenced by its lifecycle [9]. From production, where bitumen is subjected to temperatures exceeding 200 °C, to its application in pavement exposed to weathering conditions for an average of 15 years, bitumen undergoes significant chemical changes, such as oxidation. These changes alter the proportions of its chemical fractions and morphology, leading to a gradual stiffening and

* Corresponding author.

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

loss of optimal mechanical properties in a process known as bitumen aging.

Recent research has focused on designing bituminous materials with intrinsic resistance to aging. Concurrently, another research branch is dedicated to developing state-of-the-art chemical compounds that can be applied to existing bitumen to reverse aging effects, aiming to restore the material's original mechanical and rheological properties [10]. This process, known as rejuvenation, employs various techniques to return aged bitumen to a fresh condition [11]. Simpler rejuvenation methods aim to restore the original proportions of SARA fractions by adding rejuvenators that resemble specific SARA fractions. More complex techniques seek to chemically reverse aging effects, though this remains a challenging task.

All rejuvenation techniques, whether simple or complex, face a common challenge: ensuring the rejuvenator reaches and uniformly diffuses throughout the applied bitumen layer [12]. This has spurred numerous research projects exploring bitumen's surface chemistry to understand the mechanisms that enhance rejuvenator diffusion and how the SARA fractions and their microscale arrangements influence this process [13–15]. Some rejuvenators, although effective within the bulk structure of bitumen, do not diffuse readily into a bitumen layer, rendering them ineffective when applied to bitumen surfaces [16].

Most studies on bituminous materials have focused on engineering-scale behaviors [17]. However, the advancement of computational models and the need for a fundamental understanding of bitumen mechanisms have driven scientists to utilize molecular modeling techniques [2,18]. Multiple articles have addressed various challenges, though scalability issues and limited computational resources still constrain atomistic modeling of bitumen at the microscale [19]. Nonetheless, MD simulations have proven effective in predicting surface chemistry-related phenomena, driven by fundamental mechanisms known to scale well throughout numerous spatiotemporal domains [20,21]. These simulations enable scientists to understand rejuvenators' interactions with bitumen's SARA fractions at an atomic level, enhancing the design of chemically tuned rejuvenators with minimal reliance on resource-intensive experimental techniques [22].

Research articles have explored rejuvenator diffusion into different bitumen samples using Non-Equilibrium MD (NEMD) techniques to study surface-to-surface interactions and diffusivity potentials [13–15,23]. NEMD is fundamentally a subset of Molecular Dynamics (MD), distinguished by its focus on systems where time-dependent changes in properties are explicitly studied [24]. In NEMD, the system is either initialized in a non-equilibrium state, deliberately perturbed (e.g., by applying external forces or gradients such as temperature or concentration) or observed as it evolves naturally toward equilibrium. This contrasts with standard (equilibrium) MD simulations, where the system is assumed to remain in thermodynamic equilibrium throughout the simulation. In equilibrium MD, properties like energy, pressure, or diffusivity are typically expected to remain constant over time, reflecting a steady-state balance, whereas in NEMD, these properties are inherently time-dependent and provide insights into transient responses or transport phenomena [25]. When studying Fickian diffusion coefficients, NEMD setups are necessary to observe how materials diffuse over time, studying the impact that time, concentration, and depth have on the diffusive potential of a rejuvenator into bitumen [26].

While existing studies using NEMD simulations have successfully computed Fickian diffusion coefficients to quantify rejuvenator diffusion into bitumen, these computations are often limited in scope, addressing only a narrow range of specific bitumen and rejuvenator combinations. These studies typically rely on tracking molecular displacements over time, such as mean square displacement [17], or monitoring concentration gradients under controlled conditions to derive diffusion coefficients using Fick's laws [20,23]. However, the methods and observations are frequently tied to the particular materials and system conditions used in the simulations, which makes them highly system-specific. This specificity necessitates new simulations whenever

the methods are applied to different rejuvenator-bitumen systems or when their results are contested or require reproduction [27]. As a result, despite significant progress in refining simulation protocols to incorporate factors such as temperature, molecular interactions, and material variety, there remains a gap in developing generalized models or transferable insights that can reliably predict diffusion behavior across diverse materials. This limitation hampers the broader applicability of these findings and underscores the need for approaches that are both predictive and rooted in fundamental chemical and structural descriptors of the systems under study [28].

This paper aims to address five key challenges by implementing a series of methodological innovations. First, we perform a comprehensive bitumen-rejuvenator sweep test, exploring a broad range of chemical compositions, aging conditions, and temperatures to identify trends across various bitumen types and rejuvenator formulations. Second, we use force field atom types to describe the chemical nature of the samples, moving beyond conventional parameters like SARA fractions and elemental compositions, thereby improving the accuracy of insights from MD and NEMD simulations. Third, we develop a Machine Learning Model (MLM) to predict the Fickian Diffusion coefficient of bitumen-rejuvenator systems, enabling accurate predictions without the need for additional molecular simulations, even for systems beyond the training dataset. Fourth, we analyze the effect of changes in input properties on predicted diffusion coefficients, enabling scientists to fine-tune rejuvenator molecules based on their diffusion potential. Finally, we integrate the predictive capabilities of the MLMs developed in previous research [29] with the diffusion predictions and simulations presented, allowing for the evaluation of bulk thermophysical properties as a function of rejuvenator concentration, depth, and time. This comprehensive framework enhances understanding of how chemical characteristics influence rejuvenator diffusion and offers a practical tool for designing rejuvenation strategies without the need for additional simulations. The research contributes to more effective maintenance strategies for asphalt pavements, promoting the durability and reliability of transportation infrastructure.

2. Methodology

2.1. Sample collection and preparation

2.1.1. Bitumen sources

A total of 13 bitumens from various sources were collected and characterized through elemental analysis and SARA fractionation. Many of these samples exhibited similar properties; therefore, the sample size was reduced while maintaining a diverse range of chemophysical properties to ensure sufficient variation. For this study, three bitumen sources, all with a 70/100 penetration grade, were selected: “JB”, “LB”, and “UB” bitumens. These samples were specifically chosen based on their differing sulfur content by mass—type **JB** at 0.8 %, **LB** at 2.8 %, and **UB** at 4.5 %. The sulfur content is crucial in distinguishing bitumens, as it significantly affects their properties, even when samples with varying carbon, hydrogen, and oxygen levels show similar characteristics [30]. Additionally, this focus on sulfur is growing in research, driven by stricter fuel regulations that are encouraging the development of bitumens with higher and more diverse sulfur content [31].

The selection of bitumens within the penetration grade of 70/100 was intentional and does not compromise the diversity of the chemical characteristics investigated. Penetration grade, while indicative of mechanical hardness, is not a reliable descriptor of chemical variability, as bitumens with differing penetration grades can exhibit similar chemical traits. Instead, the sulfur content was employed as a more robust criterion for sample selection, as it better reflects variations in functional groups and molecular composition, which are critical for understanding the chemical diversity of bituminous systems. This approach ensures that the study captures a wide range of fundamental chemical descriptors with a minimal number of samples, aligning with its objective

to comprehensively investigate chemical diversity through meaningful and scientifically relevant parameters.

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: degree 0 is fully fresh from its source (non-aged), degree 1 is short-term aged, similar to bitumen's state when exposed to short but high thermal loads during road construction, and degree 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 [32]. 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 [33]. Degree 1 aging is achieved by performing Thin-Film Oven Tests (TFOTs) exposing 1 mm-thick fresh bitumen samples to 5 h in a Thin-film Oven at 160 °C – following the standards described in [34] – while Degree 2 through 4 are achieved by further aging Degree 1 bitumens in a Pressurized Aging Vessel (PAV) for 20, 40, and 80 h respectively at 100 °C and 2.1 MPa, following PAV testing standards detailed in [34].

2.1.3. Rejuvenator types

Rejuvenators are chemical additives that mimic the molecular structure of specific SARA groups to modify the properties of aged bitumens. While rejuvenation mainly induces physical changes rather than reversing the aging chemically, these compounds aim to restore the SARA composition to its original state, as seen in fresh bitumen [35]. For this study, four types of rejuvenators have been selected, each resembling the molecular structures found in four popular rejuvenator compounds used in the bitumen industry: Vegetable Oil (VO), Engine Oil (EO), Naphthenic Oil (NO), and Aromatic Oil (AO) [36]. For illustration purposes, **VO** is depicted in **Gold**, **EO** in **Orange**, **NO** in **Green**, and **AO** in **Brown** throughout the manuscript.

2.2. Chemical characterization methods

Chemophysical properties for all samples were obtained to establish a preliminary fingerprint for each bitumen type across all aging degrees, which are later utilized in the **Model Fitting** section. These properties include average molecular mass and dispersity of the mixture, estimated density by mass and number, 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, detailed in the following subsections. A summary of the characterization results obtained for all the samples (type **JB**, **LB**, and **UB**) across all aging conditions is presented in **Table 1**.

2.2.1. Thin-Layer Chromatography with Flame Ionization detection (TLC-FID)

In this study, the mass fractions of Saturates, Aromatics, Resins, and Asphaltenes (or SARA fractions) of all bituminous samples are quantified using the Thin-Layer Chromatography coupled with Flame Ionization Detection (TLC-FID) method, as described by Khan et al. [37]. The experimental setup involved a controlled flow of air and hydrogen at rates of 2 l/min and 160 ml/min, respectively, to ensure optimal detection conditions.

To prepare the samples, approximately 0.1 g of bitumen are dissolved in 10 ml of toluene to create a homogenous bitumen/toluene solution. This solution is then subjected to chromatographic separation using a silica-coated chromatographic rod. The separation and detection of the individual fractions were carried out in a stepwise manner based on solvent polarity. The Saturates fraction is eluted using n-heptane, allowing its isolation and subsequent quantification. The Aromatics and Resins fractions were separated and measured using solvent mixtures of toluene/n-heptane (80:20 v/v) and dichloromethane/methanol (95:5 v/v), respectively. The Asphaltenes fraction, being insoluble in the aforementioned solvents, remained at the initial sampling spot on the chromatographic rod.

Table 1

Summary of the characterization results obtained for all the bitumen types (**JB**, **LB**, and **UB**) across all aging conditions. The values have been grouped by property in **JB**, **LB**, and **UB** order. These physicochemical properties are used in the **Model Fitting** section to build the molecular models for use in MD simulations.

Test type		0	1	2	3	4	
SARA Fractionation (2.2.1)	Saturates [m%]	5.8	5.88	5.80	5.80	5.71	
		3.6	3.6	3.6	3.7	3.7	
	<hr/>		5.2	5.48	5.5	5.69	5.8
	Aromatics [m%]	63.1	61.88	55.09	49.82	44.28	
		53.357.2	51.655.02	43.947.31	38.841.79	32.534.85	
	Resins [m%]	26.2	26.47	30.94	34.89	37.73	
		30.322	30.221.80	33.924.76	36.826.91	38.427.80	
Asphaltenes [m%]	4.90	5.64	7.64	8.59	10.90		
	12.815.6	14.617.78	18.622.43	20.725.61	25.431.55		
Colloidal Index [·]	0.12	0.14	0.19	0.22	0.28		
	<hr/>		0.1960.261	0.2220.293	0.2850.376	0.3230.428	0.4100.541
Elemental Analysis (2.2.2)	Carbon [m%]	85.5	85.2	84.7	84.4	83.7	
		84.0682.53	83.7282.21	83.2681.78	83.0281.45	82.1480.66	
	Nitrogen [m%]	0.66	0.67	0.67	0.65	0.67	
		0.90.60	0.910.61	0.920.61	0.890.59	0.910.61	
	Hydrogen [m%]	10.5	10.4	10.3	10.0	9.77	
		10.9010.36	10.8610.31	10.7510.21	10.449.92	10.109.56	
	Oxygen [m%]	2.00	2.39	2.99	3.53	4.58	
0.6182.22		1.0042.59	1.583.13	2.1133.72	3.3124.85		
Sulfur [m%]	1.28	1.28	1.27	1.29	1.29		
	3.524.30	3.514.29	3.494.26	3.534.31	3.544.32		
GPC (2.2.3)	Mass-avg. molar mass[g/mol]	1481	1779	2012	2194	2258	
		38,791,886	40,492,117	47,732,441	58,022,807	82,103,434	
	Number-avg. molar mass[g/mol]	889	929	979	1015	1030	
		1,290,647	1,310,667	1,346,694	1,400,720	1,488,748	
Molecular Dispersivity[·]	1.673.012.91	1.913.093.18	2.063.43.52	2.164.153.90	2.195.524.59		

2.2.2. Elemental analysis

The elemental compositions of the samples, namely the mass fractions of Carbon (C%), Hydrogen (H%), Sulfur (S%), Nitrogen (N%), and Oxygen (O%), are analyzed using an elemental analyzer (**Vario EL III**) manufactured by **Elementar Analysensysteme GmbH**, in Germany. In depth details about this method can be found in Sieper et al. [38]. The elemental analyzer is calibrated prior to the measurements using a standard reference substance, Sulfanilamide, to ensure accurate determination of elemental distributions. Approximately 10 mg of each bituminous sample are encapsulated in a thin capsule and placed into the sample tank. The samples undergo complete oxidation and combustion under controlled conditions, producing gaseous products that are subsequently separated and detected to quantify the elemental contents. Due to the inability of elemental analyzers to directly detect oxygen as a unique gaseous product during combustion, the oxygen content is calculated indirectly by subtracting the measured percentages of other elements from 100 %, as shown in [Eq. \(2.1\)](#),

$$O\% = 100 - (C\% + H\% + S\% + N\%) \quad (2.1)$$

assuming that no other elemental types are present in the rejuvenators. This approach provides a reliable characterization of the rejuvenators' elemental composition, enabling further investigation of their chemical properties. Obtaining the elemental composition for each sample is crucial in the **Model Fitting** section for constructing accurate molecular models of bitumen, especially that these metrics are highly accurate and capture S% and O% contents which are instrumental for discretizing systems according to bitumen type and aging degree.

2.2.3. Gel permeation chromatography (GPC)

The characterization of all bitumen samples, including their number-average molecular weight (M_n), weight-average molecular weight (M_w), and z-average molecular weight (M_z), was conducted using Gel Permeation Chromatography (GPC) as outlined in [39]. These metrics provide insights into the molecular size, variety, and distribution within each sample. Combined with the derived Polydispersity Index (D), they are instrumental in the **Model Fitting** section for constructing accurate molecular models of bitumen. Tetrahydrofuran is used as the mobile phase, with a flow rate set at 1 ml/min. To prepare the sample, approximately 0.05 g of bitumen is dissolved in 10 mL of THF at room temperature. The resulting solution is filtered through a syringe filter to remove any particulates before being injected into the GPC system.

2.3. 11 Molecules selection

2.3.1. 12 Bitumens

The molecular set used in this study to construct bitumen models comprises 33 molecules, expanding on the original 12 identified by Greenfield [40] based on Shisong et al. [41]. This expansion incorporates structural modifications consistent with SARA fractionation and elemental analysis of the samples. The set aims to represent typical hydrocarbon classes found in bitumen—such as alkanes, alkenes, polycyclic aromatics, and non-aromatics—and includes functional groups like phenols, oxanes, pyridines, thioketones, and sulfoxides. These molecules, with molecular masses ranging from 200 to 1000 g/mol, high aromaticity, and significant planarity, reflect the complex structures typically found in bitumen [27]. The molecules are categorized by SARA fraction and aging degree to align with research on bitumen aging, ensuring the MD simulations accurately capture physical behaviors relevant to civil engineering [9]. The skeletal structures, chemical formulas, molecular masses, and estimated densities are provided in [Table 2](#). Corresponding SMILES [42] notation can be found in [Table S-4](#) in the [Supplementary Information](#). The molecules are included as **ACD/ChemSketch** objects for direct use in chemistry drawing software.

2.3.2. Rejuvenators

Rejuvenators are typically composed of a complex mixture of molecules—for example, **EO** contains thousands of different compounds. In this study, a single representative molecule is used for each rejuvenator (**VO**, **EO**, **NO**, and **AO**), incorporating the key functional groups relevant to each. This simplification facilitates a more focused estimation of their impact on the bitumen blend, as demonstrated in several studies on rejuvenators [36]. The skeletal representation of the rejuvenator molecules, along with their chemical formula, molecular mass, and estimated density are presented in [Table 3](#). Their corresponding SMILES notation can be found in [Table S-5](#) in the [Supplementary Information](#). The molecules in this table are inserted as **ACD/ChemSketch** objects that can be directly imported and used in most widely utilized chemistry drawing software.

2.4. 14 Molecular models

2.4.1. 15 Model Fitting

Molecules from [Table 2](#) are selected and combined to create bitumen mixtures that match the experimentally obtained chemophysical properties from [Table 1](#). This matching is accomplished by using SciPy's **Minimize** function [43] to optimize the selection and proportion of molecules to minimize the Mean Squared Error (*MSE*) between the experimental values and those computed from the virtually-created mixture. The fitting process involves the following steps:

1. Loading all molecules listed in [Table 2](#) and [Table 3](#) into a Python environment using the RDKit module [44] to retrieve the chemical descriptors (e.g., molecular weight) from their SMILES notation.
2. Computing the chemophysical properties for each molecule in [Table 2](#) (i.e., as in pure form).
3. Creating an array representing the target mixture of molecules, initially containing 3 molecules of each type.
4. Calculating the mixture-wide chemophysical properties given the initial array of molecules using the properties from Step 2.
5. 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}) from [Table 1](#). The *MSE* is calculated using [Eq. \(2.2\)](#), as follows:

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i^{est}(x_1, x_2, \dots, x_n) - y_i^{exp})^2 \quad (2.2)$$

where m is the number of chemophysical 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.

6. The process is repeated for all the samples required until all the samples are obtained.

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

1. The total number of atoms in each mixture must remain within $7000 \pm 2\%$, 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 [4,40].
2. Only molecules that are within the aging degree 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, 3, and/or 4, unless such a selection cannot produce a valid molecular model. Models of aging

Table 2

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	—	—	—	—
	2	 C35H62 482 0.913	—	—	—	—
Aromatics	3	 C35H44 464	 C35H42O 478	 C35H36O4 1.173	—	—
	4	1.030 C30H46 406 0.916	1.064 C30H44O 420 0.955	 C30H42O2 434 0.995	—	—
Resins	5	 C36H57N 503	 C36H55NO 517	 C36H53NO2 531	—	—
	6	0.977 C26H50O 414 0.893	1.006 C26H48O2 428	1.040 —	—	—
	7	 C18H10S2 290 1.417	0.930 C18H10S2 306 1.540	 C18H10O2S2 322 1.68	—	—
	8	 C40H60S 572	 C40H60OS 588	 C40H58O2S 602	 C40H56O3S 616	—
	9	0.962 C40H59N 553	1.010 C40H57NO 567	1.040 C40H55NO2 581	1.071 —	—
	10	1.007 C40H59N 553	1.035 C40H57NO 567	1.067 C40H55NO2 581	—	—

(continued on next page)

Table 2 (continued)

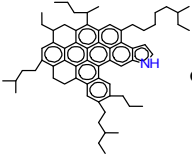
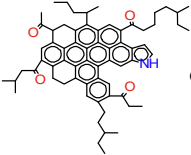
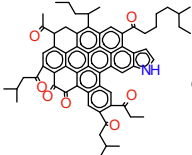
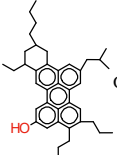
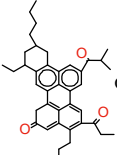
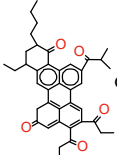
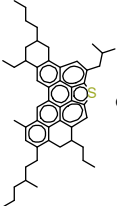
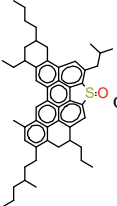
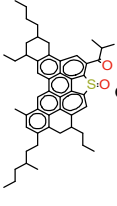
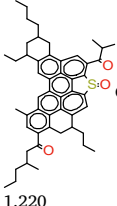
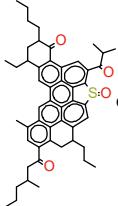
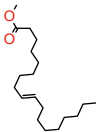
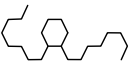
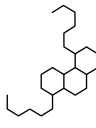
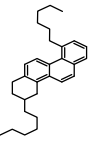
SARA	ID	Aging degree				
		0	1	2	3	4
Asphaltenes	10	 C66H81N 888 1.104	 C66H73NO4 944 1.188	 C66H67NO7 986 1.254	–	–
	11	 C42H54O 574 1.049	 C42H50O3 602 1.130	 C42H46O5 630 1.201	–	–
	12	 C51H62S 707 1.100	 C51H62OS 723 1.160	 C51H60O2S 737 1.189	 C51H56O3S 751 1.220	 C51H56O4S 765 1.252

Table 3

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

Rejuvenators				
	VOC19H36O2 297 0.873	EOC22H44 308 0.812	NOC30H40 401 0.984	AOC26H48 360 0.863

conditions 0 and 4 are forced to use molecules of aging degree 0 and 4, respectively. This approach ensures that “aged mixtures”—which are the primary focus—contain minimal contributions from molecules in aging categories that are too distant. Saturates are not known to undergo aging, so a single category is used to represent them across all aging states.

- 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 % across all samples, ensuring that each mixture accurately represents its specified aging condition and maintains sufficient discretization to yield statistically valid comparisons, especially when considering the impact of oxidative aging.
- 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 9 to 14 molecule types per mixture [40].

Table S-1 in the [Supplementary Information](#) lists the chemophysical properties and Table S-2 lists the type and number of molecules from Table 2 required to construct each of the mixtures for all bitumen types (**JB**, **LB**, **UB**) and aging degrees (0 through 4).

The rejuvenator mixtures, however, consist of a single molecule type (from Table 3) and do not undergo any Model Fitting process. Four molecular mixtures are created, with the only constraint being that the total number of atoms is comparable to that of the bitumen systems, set at $7000 \pm 2\%$, resulting in approximately 92 rejuvenator molecules per system for each type: **VO**, **EO**, **NO**, and **AO**.

2.4.2. 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. (2.3),

$$\mathbf{F} = -\nabla E \quad (2.3)$$

The molecular simulations of this study employ the Polymer Consistent Force Field (PCFF) [45], where the potential energy (E_{PCFF}) is given by the sum of 12 interaction types, as shown in Eq. (2.4),

$$E_{PCFF} = \sum E^b + E^{ia} + E^{oa} + E^t + E^{bb} + E^{ba} + E^{bt} + E^{aa} + E^{at} + E^{tt} + E^{VDW} + E^{coul} \quad (2.4)$$

where each interaction term and its respective potential is described in Table S-3 in the [Supplementary Information](#). 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) [46]. The non-bonded interaction distance is established at 12 Å, beyond which all interactions are considered negligible and set to zero.

The parameters required to run the simulations are differentiated by the atomic types used to represent each atom in the system. The PCFF, being a Class2 force field [47], differentiates atoms through the use of several criteria, among them the chemical element of the parent atom, number of attached hydrogens, hybridization state, coordination number, presence in a ring, involvement in an aromatic system, and the nature of the list of neighboring atoms. [45]. Table 4 contains a list of the atom types and their description used to model all the samples (bitumens and rejuvenators alike) in this study. These atom types are explicitly addressed in this manuscript as they serve as key chemical descriptors for defining the input features used in training the MLM, as outlined in the [Features selection](#) section.

In the [Supplementary Information](#), Folder */pcff* comprises a simulated sample of a bitumen-rejuvenator blend containing all the molecules from Table 2 and Table 3. 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. Additionally, **PDB2DAT**, a software package developed by our research group, can assign PCFF force field atom types to a PDB system when topologies for LAMMPS' [structure data files](#) are necessary.

Table 4

Atom types and their corresponding description used to represent the atoms in the molecular simulations involving the use of all-atom PCFF force field dynamics.

Atom type	Description	Atom type	Description
hc	Hydrogen bonded to carbon	hn	Hydrogen bonded to nitrogen
o=	Oxygen double bonded to O, N, C, S, P	c3	Sp ³ carbon with 3H's
ho	Hydrogen bonded to oxygen	oh	Oxygen in hydroxide ion (OH ⁻)
c_1	Amide, acid and ester carbonyl carbon	cpc	Alpha/ipso carbon in aromatic ethers (-C-O-C-)
c2	Sp ³ carbon with 2H's	c = 1	Non-aromatic, next to end doubly bonded carbon
c5h	Sp ³ carbon in 5-membered ring	c3oe	Alpha carbon in methyl containing ethers (-C-O-CH3)
c0oe	Alpha carbon in ether containing tertiary alkyl group (-C-O-C-R3)	nh	Sp ² nitrogen in 5- or 6-membered ring
sp	Sulfur in an aromatic ring (e.g. thiophene)	c0	Sp ³ carbon with 0H's
na1	Sp ³ nitrogen in secondary aliphatic amines	c5	Sp ² aromatic carbon in 5-membered ring
s'	S in thioketone group	cp	Sp ² aromatic carbon with partial double bond
oc	Sp ³ oxygen in ether or acetals	c_0	Aldehydes and ketones carbonyl carbon
cs	Sp ² aromatic carbon in 5 membered rings next to S	c=	Non-aromatic end doubly bonded carbon
np	Sp ² nitrogen in 5- or 6-membered ring	c5h1	Sp ² aromatic carbon in 5-membered ring
o_1	Carbonyl oxygen	c1	Sp ³ carbon with 1H
c = 2	Non-aromatic doubly bonded carbon	o_2	Ester oxygen

2.4.3. Individual models

The procedures for preparing molecular systems in LAMMPS for subsequent simulations are outlined in detail below:

1. The structures of the molecules of Table 2 and Table 3 are initialized using their SMILES notations, processed into metastable 3D conformers via the RDKit Python module [44].
2. The initialized molecules are placed into a simulation box at a low initial density of 0.20 g/cm³, ensuring an even distribution and minimizing particle overlap by utilizing a low-discrepancy Sobol distribution method. Molecules are placed 25 Angstroms away from the walls along the z-axis to ensure that the atoms remain confined within the boundaries. Moreover, as the systems are expected to be merged into surface-to-surface systems (i.e., bitumen surfaces facing rejuvenator surfaces) later, the initial x and y dimensions are set to match the x and y dimensions of the equilibrium density of the material, resulting in an elongated simulation box along the z-axis.
3. The PCFF force field parameters and charges are assigned to the molecules using RDKit atomic descriptor functions (e.g., `atom.GetIsAromatic()` to identify if an atom is a member of an aromatic ring). The atomic positions are optimized to minimize E_{PCFF} in Eq. (2.4) using the conjugate gradients method until both energies and forces reach a threshold precise to the nearest integer [48].
4. The systems undergo uniaxial compression along the z-axis (keeping the dimensions of x and y constant) to achieve the target density. This is performed under NVT conditions over 5 ns, using a true strain rate of 1 %, and applying uniaxial deformations to the simulation box every 1 picosecond elapsed. The molecules are confined along the z dimension by two walls of LAMMPS `fix indent` type, located at the edges of the simulation box (i.e., at $z = 0$ and $z = l_z$). Compression is stopped when the dimensions of the simulation box reach approximately equal magnitudes, and the system density matches the reported density estimate of the samples in Table 2 and Table 3. At this stage, subsequent NPT simulations will enable the box dimensions to adjust gradually, facilitating the system's convergence to a more accurate equilibrium volume and density.
5. The systems are subjected to 50 isobaric NPT annealing cycles to remove residual conformational strain, where the temperature is oscillated sinusoidally with an amplitude of ± 25 % of the set equilibrium temperature. Each cycle period varies between 0.2 and 1 ns. The systems are allowed to expand only along the z-axis, while the x and y dimensions are fixed at constant values. Pressure is controlled in the z-direction, with the walls along z acting as flexible, movable boundaries that maintain a pressure of 1 atm while still confining the particles. This setup ensures that the system can relax strain by adjusting the volume in the z-direction while preventing particles from escaping through the walls [49].
6. 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. Just as in Step 5, the systems are allowed to expand along the z-axis only [50].
7. An additional step is implemented when simulations do not require confinement along the z-axis. This involves executing Step 6 without walls along the z-axis and applying full periodic boundary conditions. This configuration enables the molecules to freely cross boundaries, effectively transforming confined systems into representations of bulk conditions in all directions. Although not strictly necessary for the surface-to-surface simulations in this study, this step allows for the generation of models whose properties can be more easily evaluated without the influence of confining conditions.
8. System stability is evaluated by conducting dynamics stages under NVT and NVE conditions for 50 ns each, without a barostat or thermostat. Stability is confirmed if the potential and kinetic

energies stay within 5 % of their initial average. Systems failing to meet this criterion are discarded, and the process is restarted from Step 1.

The modified Nose-Hoover integration algorithm [51] 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. The procedure is run separately for the same sample at 60, 120, 135, and 200 °C. All LAMMPS routines are run at 101325 Pa with a step size of 0.5 fs/step. Periodic boundary conditions are implemented only along the x and y directions. The resulting systems from Step 6 and Step 7 are used for later simulations performed in this study. Steps 1 and 2 entail the utilization of SMI2PDB [52], whereas Step 3 involves the use of PDB2DAT [53]. These tools have been created by our research team to execute the respective tasks rapidly and effectively. Figure S-1 in the Supplementary Information illustrates all the steps performed to initialize the molecular models of this study.

2.4.4. Surface-to-Surface models

2.4.4.1. Initialization. The models developed in Step 6 of the **Individual Models** section serve as the basis for all surface-to-surface models simulated in this study, including those of same composition but different temperature. The subsequent steps detail the creation of these bitumen-rejuvenator models are presented next, with Fig. 1 summarizing them through a series of illustrations.

1. Merging both bitumen and rejuvenator systems into a single one, translating the rejuvenator system along the z-axis so that the two systems are facing each other with an inter-surface distance equal to 5 Angstroms [14]. The dimension of the z-axis in the merged systems is therefore given by:

$$L_z = L_z^{bit} + L_z^{rej} + 5 \quad (2.5)$$

2. Walls are introduced at three locations: at the edges of the simulation box, where $z = 0$ and $z = l_z$, and between the two layers (at $L_z/2$). The walls at the extremes are modeled using LAMMPS' `wall/table type`, which utilizes tabulated values of $E(r)$ and $F(r)$ to calculate the interactions between the atoms and the walls. This approach improves the quality of the boundary conditions by providing interaction potentials that closely resemble the bulk behavior of the material near the wall. A more detailed description of this approach can be found in **Boundary Conditions**. The wall in the middle utilizes LAMMPS' `harmonic wall type`, which softly pushes particles back, preventing the particles from diffusing from one layer to another.
3. An isothermal-isobaric NPT routine is run for 5 ns to initialize particle interactions within the merged setup. During this stage, the system is allowed to interact across the inner wall (at $L_z/2$), while limiting periodic interactions across the outer walls at $z = 0$ and $z = l_z$. These outer walls are movable under the pressure control process, still applying 1 atm along the z-axis as defined previously. This pressure control ensures that while interactions occur within the merged system, the boundaries continue to reflect the influence of the external pressure equal to 101325 Pa. Allowing the system to expand or contract uniaxially until reaching equilibrium helps release residual stresses, which can significantly impact the diffusive behavior of rejuvenators into bitumen. In condensed systems, even small changes in linear volume can result in significant variations in internal stresses and, consequently, system pressure. Under conditions of high variation, this pressure is recognized as a key driver of diffusion, and as such, these variations must be minimized.

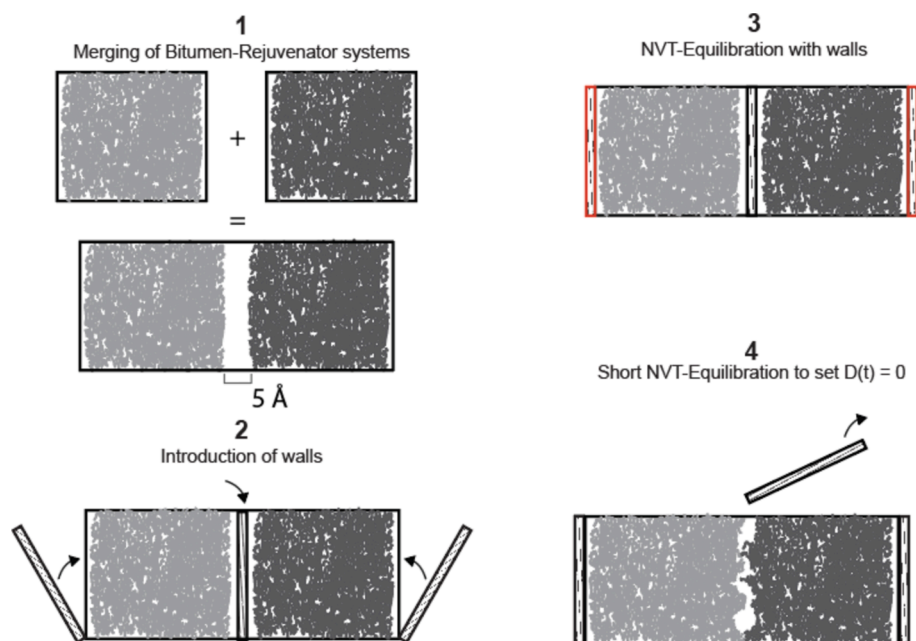


Fig. 1. Molecular models from the Initialization section are stacked along the z-axis to form surface-to-surface configurations, with their faces oriented towards each other.

4. The inner wall is then removed, and the particles are then allowed to accommodate lower energy positions in a subsequent NVT routine 0.25 ns long. While the separation between the two phases may be disrupted (some molecules are expected to come together), the diffusion expected is short enough to still consider the last point in this step to be the condition where $t = 0$ when computing the diffusion coefficient.

2.4.4.2. Boundary conditions. Given the relatively small size of the systems under study—where the diffusive behavior of both bitumens and rejuvenators is expected to influence the interaction potential up to the boundary walls—it is crucial that the walls interact with the molecules in a manner that accurately represents the bulk of the materials at the surface. In many surface-to-surface diffusion simulations found in the literature, walls are typically modeled as hard, reflective boundaries that fail to replicate the bulk material at the interface, resulting in “artificial” interactions between the molecules and the boundary

[14,15,28]. Some studies address this limitation by introducing an additional layer of material at the boundary, where the atoms are fixed to create a barrier that more closely mimics the bulk interactions [23]. However, this approach requires a layer almost as thick as the simulated material itself, effectively doubling the number of particles and significantly impacting computational performance. Since the diffusive simulations in this study are expected to run for tens of nanoseconds, such material-like barriers are computationally intractable.

To overcome this, the recently introduced `fix wall/table` command in LAMMPS is utilized, allowing for the construction of walls that interact with particles using tabulated potentials. These potentials relate the distance between the atoms and the wall to their corresponding potential energy and force values ($E(r)$ and $F(r)$), enabling a more realistic interaction with minimal additional computational cost.

To develop the tabulated potentials, simulations are conducted by initially positioning two surfaces of the same material very close together and then sequentially separating them until they reach a

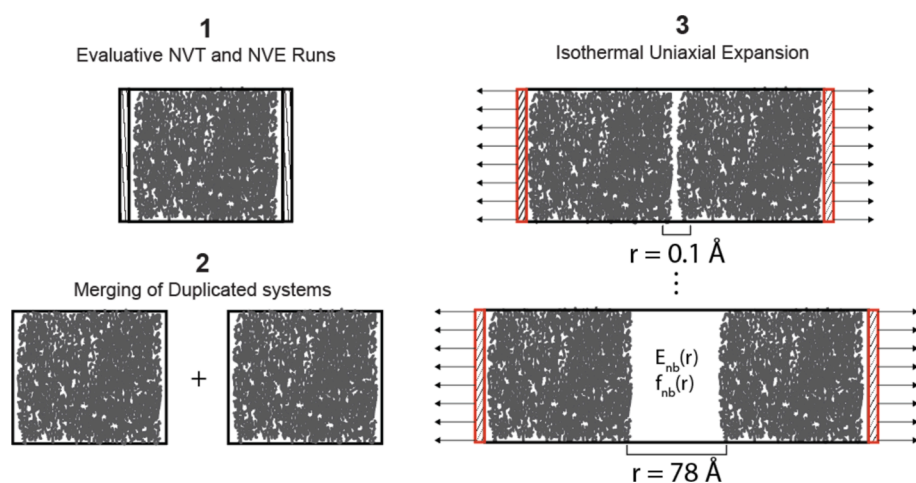


Fig. 2. Tabulated potentials are generated for all the bulk models created in this study to represent more realistic walls. Using these tabulated potentials as “walls” aims to minimize artificial artifacts during particle interactions with the boundaries in the simulations, as they more accurately reflect the material’s behavior in its bulk phase.

specified distance, equal to 75 Å or more to ensure that long-range electrostatic interactions described by the Coulomb term in the PCFF force field (see Table S-3 in the Supplementary Information) are minimized. The detailed procedure for these simulations is described below; a sequence of the steps is illustrated in Fig. 2.

1. Each individual system undergoes an NVT dynamics stage for 1 ns to initialize the system's dynamics to a steady state. This process is repeated separately for all four temperatures tested in this study (60 °C, 120 °C, 160 °C, and 200 °C). During this stage, confinement is applied to prevent atoms from crossing the z-axis.
2. A duplicate of the equilibrated system is created and positioned adjacent to the original, facing it along the z-axis, with a separation distance of 1 Å.
3. The two systems then undergo uniaxial expansion along the z-axis (while keeping the x and y dimensions constant) until the distance between the two regions equals to the original length of the simulation box (~78 Å). This expansion is performed under NVT conditions over 1 ns, using a true strain rate of 1 %, with uniaxial deformations applied to the simulation box every 1 picosecond. A true strain rate is used instead of a linear rate to ensure denser sampling at shorter distances, where the energetic potential is expected to exhibit the most significant variations.
4. Intermolecular energies corresponding to non-bonded interactions, along with their corresponding forces, are recorded. These values are obtained using the LAMMPS' `compute group/group` command, with the energies and forces averaged over the last 1000 steps before the next deformation step. This approach ensures that any intramolecular strain is dissipated before measurements are taken. The non-bonded interaction energy is calculated as shown in Eq. (2.6).

$$E_{nb} = E^{vdW} + E^{Coul} + E^{long} \quad (2.6)$$

Tabulated potentials are generated for each system examined in this study, encompassing 19 models—comprising 3 types of bitumen across 5 aging degree and 4 rejuvenators—at 4 different temperatures (60 °C, 120 °C, 160 °C, and 200 °C). This results in a total of 76 distinct potentials.

2.4.4.3. Sample labels. Considering there are 3 types of bitumen by source, 5 degrees of aging, and 4 types of rejuvenators, a total of 60 surface-to-surface models are created. Since the models are tested at four different temperatures, a total of 240 surface-to-surface simulations are conducted. The label format “AAA”, depicted in Fig. 3, is composed of three characters: the first indicates the bitumen type, the second denotes the aging degree, and the third specifies the rejuvenator type. For example, the label “l4e” corresponds to a bitumen of type **LB** aged to degree 4, rejuvenated with **EO**. A character represented by the letter “X” signifies “all members within its category”, used to identify multiple samples using one label.

2.4.5. Fickian diffusion simulations

The diffusion simulations extend the systems generated in Step 4 of **Initialization** by continuing the simulations for an additional 30 ns,

while maintaining all previously established simulation parameters. The choice of 30 ns is a compromise between ensuring measurable diffusion—particularly at lower temperatures and for incompatible bitumen-rejuvenator combinations—and avoiding excessive diffusion at higher temperatures, where bitumen-rejuvenator mixtures readily combine, rendering the calculation of the diffusion coefficient, D , via Eq. (2-7) statistically insignificant. Each simulation outputs 300 trajectories, sampled every 0.167 ns, capturing atomic positions, velocities, accelerations, and stress tensors. These trajectories are divided into three files, each containing data for Bitumen-only, Rejuvenator-only, and Bitumen-rejuvenator molecules. Additionally, system-wide properties, such as potential energies and densities, are sampled at the same rate, with the values calculated as averages of the instantaneous measurements taken at each time step within the sampling interval. Fig. 4 depicts the expected progression of the simulations over time.

2.5. Benchmarking properties

2.5.1. Thermophysical properties

Numerous thermophysical properties can be extracted from the simulations and can be categorized into time-dependent or steady-state measurements. The properties measured in the **Fickian Diffusion Simulations** are time-dependent and track the evolution of these properties over time, ensuring that the observed phenomena follow thermodynamically and kinetically favorable trends. Some properties measured are compared to those in their analogous systems from Step 6 and 7 in the **Individual Models** section and Step 4 of the **Initialization** of surface-to-surface models. This is performed to ensure that the properties observed during the **Fickian Diffusion Simulations** adhere to realistic trends and are not influenced by artificial numerical artifacts or complexities that may arise during these simulations. These properties, along with the methods used for their computation, are summarized in Table 5. Among the thermophysical properties computed, the Cohesive Energy Density (CED) is determined using a specific set of LAMMPS routines, building upon the systems outlined in Step 7 in the **Individual Models** section. Further details of these simulations are provided in the file Section S-1.5 in the **Supplementary Information**.

2.5.2. Fickian diffusion coefficient

The diffusion coefficient D for a Fickian diffusion process across one dimension can be calculated from the density profiles using Fick's Second Law of diffusion, as presented in Eq. (2.7):

$$C(z, t) = C_0 \left(\frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{z}{2\sqrt{Dt}} \right) \right) \right) \quad (2.7)$$

where $C(x, t)$ is the concentration of bitumen or rejuvenator along z (i.e., depth) and at a time t , and C_0 is the concentration of the material in its bulk phase. In this approach, a fitting process is employed to extract the value of D from the simulations in the Production section. This method is consistent with that used in previous studies [13,23,54], and the specific details of the procedure are provided next.

The trajectories for all the samples obtained from the **Fickian**

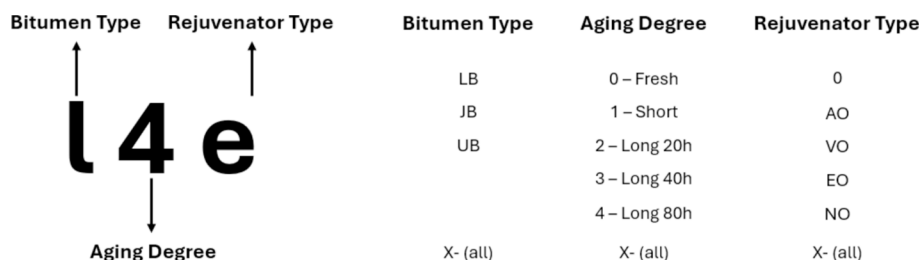


Fig. 3. The labeling system used to identify all surface-to-surface models in this study, categorized by bitumen type, aging degree, and rejuvenator type.

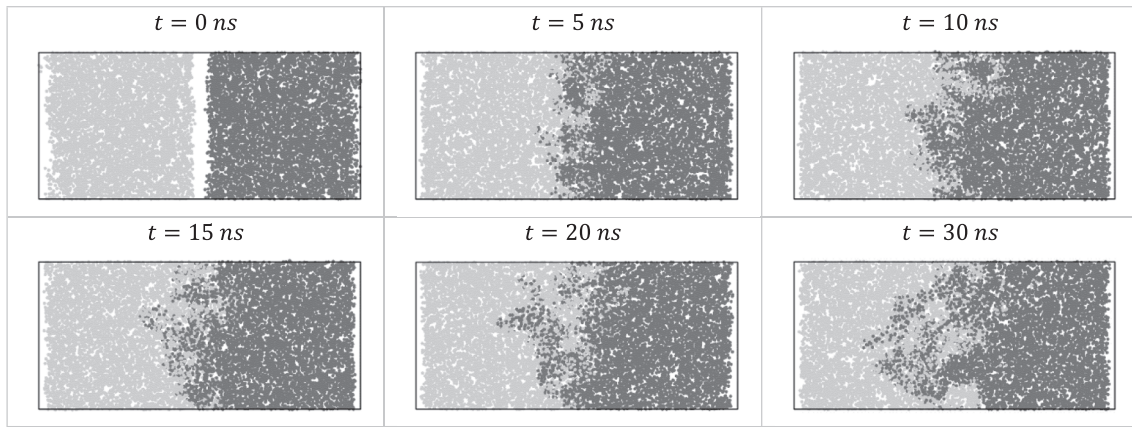


Fig. 4. Expected progression of the diffusion simulations over time where the bitumen and rejuvenator molecules are compatible.

Table 5

Thermophysical properties computed for evaluative purposes, encompassing both Step 6 and 7 in the **Individual Models** section (steady state) and **Fickian Diffusion Simulations** (time-dependent) sections.

Quantity	Time dependency	Expression	Notes
Potential Energy (P_E) [kJ/kg]	Both	E_{PCFF}	
Non-bond Energy (E_{nb}) [kJ/kg]	Both	$E_{nb} = E^{vdW} + E^{Coul} + E^{long}$	
Bonded Energy (E_b) [kJ/kg]	Both	$E_b = E_p - E_{nb}$	
Enthalpy (H) [kJ/kg]	Steady state	$H = E_p + E_k + PV$	
Molar Volume (V_m) [m ³]	Steady state	$V_m = \frac{M_w}{\rho}$	
Density (ρ) [kg/m ³]	Steady state	$\rho = \frac{m}{V}$	
Self-diffusion coefficient (D_{self}) [m ² /s]	Steady state	$D = \lim_{n \rightarrow \infty} \frac{ \mathbf{r}(t) - \mathbf{r}(0) ^2}{6t}$	
Enthalpy of Vaporization (H_v) [kJ/kg]	Steady state	$H_v = V_m CED + RT$	
Solubility Parameter (δ_{sol}) [kJ ^{0.5} /m ^{1.5}]	Time dependent	$\delta_{sol} = \sqrt{\frac{CED}{V_m}}$	
Surface Tension (T) [N/m]	Time dependent	$T = \frac{l_z}{2} \left(\frac{\sigma_{xx} + \sigma_{yy}}{2} - \sigma_{zz} \right)$	
Isochoric Heat Capacity (C_v) [kJ/kg/K]	Steady state	$C_v = \left(\frac{\delta U}{\delta T} \right)_v$	
Thermal Expansion Coefficient (β) [1/K]	Steady state	$\beta = \frac{1}{V} \left(\frac{\delta V}{\delta T} \right)_p$	
Fickian diffusion coefficient (D)	Time dependent	$\rho(x, t) = \rho_0 \text{erf} \left(\frac{x}{2\sqrt{Dt}} \right)$	

Diffusion Simulations section are divided into 50 slices along the z-axis of the simulation box. The position along the z-axis is normalized by applying a shift, calculated as the mean of the box boundaries in the z-dimension. The volume of each slice is determined by multiplying the slice thickness (obtained from the slice's z-boundaries) by the cross-sectional area in the x and y dimensions. Atoms are assigned to their corresponding slices based on their z-coordinates, with the atomic mass and position used to allocate them appropriately. The density for each slice, ρ_i , is then calculated as shown in [Eq. \(2.8\)](#):

$$\rho_i = \frac{\sum m_j}{V_{slice}} \quad (2.8)$$

where m_j is the mass of atom j within slice i , and V_{slice} is the volume of the slice. The density of each slice is stored alongside the mean z-position of the slice. To fit the data to the diffusion equation, the mean position of each slice is scaled by the simulation time t , using the transformation $z t^{-0.5}$. This transformation is commonly applied in Fickian diffusion

analysis to account for the time-dependent spreading of the density profile, allowing all the data points in different locations and times to be grouped into a single, “sigmoidal looking” continuous curve.

The transformed data is subsequently fitted to a sigmoidal curve using the error function, which represents the solution to the diffusion equation in one dimension. The error function describes the spatial and temporal evolution of the density distribution, given in [Eq. \(2.9\)](#) by:

$$\rho(x, t) = \rho_0 \text{erf} \left(\frac{z}{2\sqrt{Dt}} \right) \quad (2.9)$$

where ρ_0 corresponds to the bulk density of the material, z is the position of the particle in the z dimension, t is the simulation time elapsed, and D is the diffusion coefficient. The curve fitting process involved adjusting one parameter, B , related to the diffusion coefficient. The error function fit is performed using Scypy's [CurveFit](#) function to match the transformed data points with the density values. The relationship between the fitted parameter B and the diffusion coefficient D is obtained as shown in [Eq. \(2.10\)](#).

$$D = \frac{1}{4B^2} \quad (2.10)$$

2.6. Machine learning

This study introduces a MLM designed to predict the Fickian diffusion coefficient between two hydrocarbon blends—specifically, bitumen and rejuvenators—using a series of chemical descriptors and the system temperature as inputs. This model allows for the near-instantaneous estimation of diffusion coefficients, offering insights into how various chemical factors (e.g., changes in molecular structures) influence the diffusive behavior of both components. The details of this approach are presented in the following sections.

2.6.1. Model

The MLM used in this study is generated and trained through a Python script utilizing the Pandas [55] and Sklearn [56] libraries. This script implements a hybrid predictive model that combines the advantages of both tree-based and linear interpolation models. The core of the approach is encapsulated within a HybridModel class, which employs a Random Forest Regressor (RFR) [57] as the primary predictive model, alongside a Linear Regression model to handle interpolation. During the prediction process, the hybrid model first verifies whether the input features align with those in the training data. If they do, predictions are made directly using the tree-based model. If not, the model locates 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 output [58].

To assess the predictive potential of the machine learning models, 10 % of the properties measured through MD simulations are randomly selected and excluded from the training set. Specifically, out of a total of 240 samples (60 bitumen-rejuvenator combinations across a temperature sweep of 4 temperatures), 24 samples are used for evaluating the MLMs' unobserved predictive performance, while the remaining samples will serve as input for training the models. Although the 90/10 ratio is slightly higher than the commonly used 80/20 (or lower) ratio in MLM training, the limited number of available samples necessitates a more conservative approach, requiring a larger proportion of samples for model training [59].

The script follows these steps during execution:

- Data Reading and Preparation:** Load the dataset containing the x_1, x_1, \dots, x_{19} and y_i values for all the simulations of this study into Panda's DataFrame objects.
- Hybrid Model Initialization:** Initialize the hybrid model with a RFF and a linear interpolation model, incorporating a preprocessing pipeline for data standardization.
- Model Training:** Scale the training data using the preprocessing pipeline, train both the RFF and interpolation models, and construct a K-dimensional tree to enable nearest neighbor searches. Determine feature ranges for boundary checking.
- Prediction and Evaluation:** Predict the values of y_i using the hybrid model for 90 % of the training arrays of x_1, x_1, \dots, x_{19} values used. This ensures that 10 % 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 0.96 and the magnitudes of the differences between y_i^{MD} and y_i are below 5 % for both observed and unobserved datasets.

The MLM file to predict Fickian Diffusion, created by serializing the **HybridModel** objects along with all associated features using Python's **Pickle** module, is located in the */MLMs* directory within the **Supplementary Information**. The MLM can be loaded back into Python and be used to predict Diffusion coefficients by using a custom set input features.

2.6.2. Features selection

The input parameters for training the MLMs in this study include the chemical composition of the samples, the molecular mass of bitumen and rejuvenator layers, and the system's temperature. Initially, all 30 atom types in the PCFF were considered, but this approach would require managing 60 chemical descriptors due to the bilayer nature of the systems (to describe the chemical nature of both the bitumen and rejuvenator layers separately). This would demand an excessively large data set for training [45]. To simplify, the PCFF atom types were combined into 8 descriptors that capture elemental and functional differences, resulting in 18 atomic type descriptors (9 for each layer). **Table 6** maps this simplification.

Although the number of types has been reduced from 30 to 18 (9 for bitumens and 9 for rejuvenators), this reduction remains sufficient to identify trends between diffusion potentials and more fundamental

Table 6

List of simplified 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.

x_n	Feature Name	Description	Color	Symbol	PCFF types
x_1	Saturation	Hydrogens bonded to carbon atoms, contributing to an increase in the degree of saturation.	Red	H	hc
x_2	Aliphatic	Sp ³ hybridized aliphatic carbons	Blue	Ch	c1 c0 c2 c3
x_3	Aromatic	Sp ² hybridized aromatic carbons	Green	ca	c5 cp
x_4	Cyclopentane	Sp ³ hybridized cyclopentane carbons	Olive	c5	c5h1 c5h
x_5	DoubleBond	Isolated doubly bonded carbons	Light Blue	c=	c= c=1 c=2
x_6	Oxygen	All atoms that are adjacent to or integrated within an oxidized functional group	Purple	Ox	oc o_1 c_0 c_1 c3oe c0oe cpc o= ho o_2 oh
x_7	Nitrogen	All atoms that are adjacent to or incorporated within a nitrogen-containing functional group	Olive	Nx	hn na1 nh np
x_8	Sulfur	All atoms that are adjacent to or incorporated within a sulfur-containing functional group	Pink	Sx	sp s' cs

chemical descriptors. The “atom-type” formula for one side (bitumen or rejuvenator) is obtained by applying Eq. (2.11). Fig. 5 depicts how atom types can be used to discretize molecules into functional blocks.

$$x_{i,1-8} = \frac{1}{N} \left(\sum a_1, \sum a_2, \dots, \sum a_8 \right) \quad (2.11)$$

where N is the total number of atoms, and a_i corresponds to the number of a certain atom type in the sample.

Fig. 6 provides an overview of how features are used to describe both bitumen and rejuvenator samples in a Fickian diffusion simulation. The MLM distinguishes between the two by separately analyzing their chemical descriptors. This approach helps the model capture the differences between bitumen and rejuvenators and better assess the driving force pushing diffusive interactions.

Therefore, the MLM is trained using 19 features. The first nine features (x_1 to x_9) represent the chemical descriptors and molecular weight of the bitumen side, while the next nine (x_{10} to x_{18}) describe the same properties for the rejuvenator side. The final feature (x_{19}) corresponds to the temperature of the system. This separation of features mirrors the distinction between bitumen and rejuvenator in surface-surface simulations, ensuring that each material is represented individually. Directory *MLMs/predictions* in the [Supplementary Information](#) includes the files containing the chemical features for all samples prepared in this study.

2.6.3. Labels selection

The properties, or labels, correspond to the properties that are to be predicted by the MLMs. In this study, only the Fickian Diffusion Coefficient described in **Fickian Diffusion Coefficient** is selected. This aims to establish a MLM that can, to a certain extent, abstract away from the use of MD simulations to predict the Fickian Diffusion Coefficient, D_i , of bitumen-rejuvenator systems almost as accurately as if they were measured using MD simulations (D^{MD}), as depicted in Eq. (2.12),

$$MLM = f(x_1, x_2, \dots, x_{19})_i = D_i \approx D_i^{MD} \leftarrow \rho_0 \operatorname{erf} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (2.12)$$

where x_n correspond to a combination of features describing the molecular system, i , for which the property, D_i , is required. By obtaining the value of D_i and knowing its relation to Eq. (2.12), the concentration of a rejuvenator at a specific depth and/or time can be estimated without having to set up experiments or perform additional MD simulations.

3. Results

3.1. Diffusion coefficients

Fig. 8 presents various simulation endpoints to demonstrate how the diffusion values discussed in this study relate to the extent of mixing achieved by the rejuvenators. Specifically, a diffusion coefficient below 0.2 indicates a “Slightly- to Non-diffused” state, while a D value between 0.2 and 1.5 corresponds to a “halfway” diffusion state. A D value exceeding 1.5 signifies a “Well diffused” state. This depiction follows Fick’s second law of diffusion in one dimension, in Eq. (2.7), which provides an estimate of how far, on average, particles have diffused over time. Conversion into equivalent millimeters per hour is provided as well.

Table 7, Table 8, and Table 9 summarize the diffusion rates of the samples based on rejuvenator type, bitumen type, and bitumen aging degree, respectively. Each table groups the samples into three categories—those with the highest, average, and lowest diffusion rates—across the studied temperature range, with the average diffusion coefficients provided in parentheses. Table 10 presents visual representations from simulations corresponding to the highest, average, and lowest diffusion coefficients measured across all four temperatures.

Fig. 9 displays visual representations of the diffusion simulation for sample “10v” at six distinct time points throughout the simulation.

3.2. Experimental validation

Before developing MLMs to predict diffusion coefficients and their application in pavement engineering, it is essential to validate the training data generated from MD simulations. This validation involves comparing MD-derived data with empirical observations, preferably for both individual steady-state systems and diffusion measurements in bitumen-rejuvenator combinations.

While it is not feasible to obtain experimental data for all the samples (19 for individual systems and 240 for diffusion systems), key physical properties—such as densities, heat capacities, thermal expansions, and solubility parameters—can still be obtained for certain hydrocarbon blends, presented in Table 11.

Careful attention must be paid to the Fickian diffusion coefficients from the time-dependent simulations in this study to ensure they match experimental data. Although individual experimental diffusion data is limited, and the experimental range does not fully match the variety

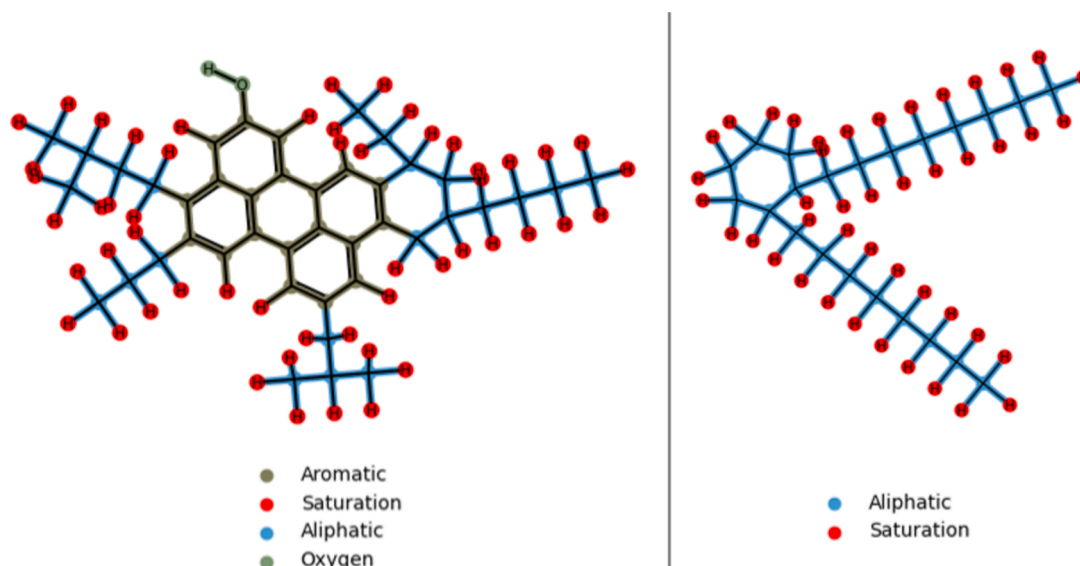


Fig. 5. Identification of chemical features described in Table 6 in bitumen-like molecules (Left – Phenolic Asphaltene) and rejuvenator-like molecules (Right – EO).

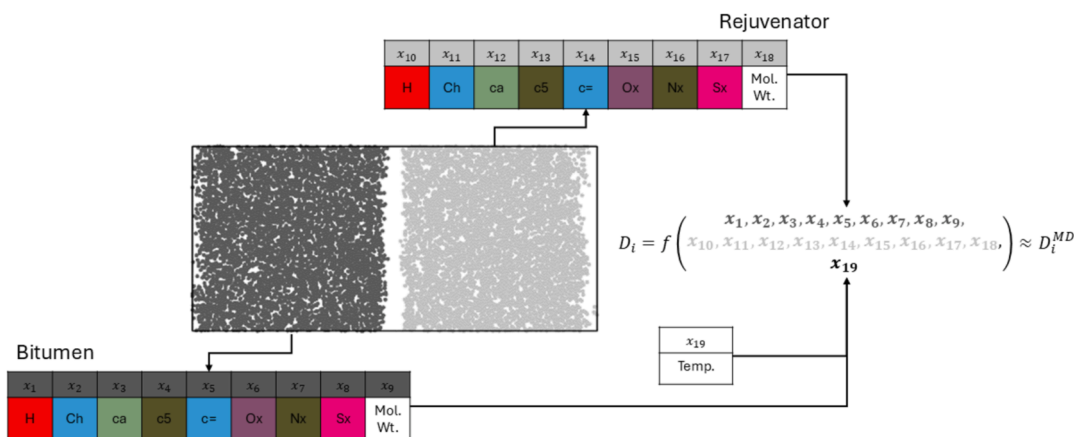


Fig. 6. The MLM trained to predict diffusion uses 19 features: the first 9 (x_1 – x_9) represent the chemical descriptors and molecular weight of the bitumen side, the next 9 (x_{10} – x_{18}) represent the same for the rejuvenator side, and the final feature (x_{19}) corresponds to the system-wide temperature. This segregation allows for the separate definition of the bitumen and rejuvenator sides, similar to how they would be distinguished in surface-surface simulations.

tested in the simulations, there is overlap in some samples and conditions, providing validation for the MD-based diffusion results [36,69]. Table 12 shows the diffusion coefficients from both experiments and simulations for samples where data is available.

3.3. Features range

Given that the study sweeps through an array of 240 different bitumen-rejuvenator simulations, there exists 240 unique arrays of features (x_1, x_2, \dots, x_{19}) used as inputs for training the MLM of this study. Therefore, the range of values (from minimum to maximum) covered by these simulations for each x_i are presented in Table 13. 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.

3.4. Features Importance and directionality

Table 14 displays the sorted and normalized importance scores for features that influence diffusion coefficients, organized from the most to the least influential.

Table 15 through 18 present the results for each temperature, displaying the measured diffusion coefficient in relation to the most influential features of the samples identified by the MLM. The tables categorize the systems into three groups: the top five systems with the highest diffusion coefficients, five systems with intermediate values, and five with the lowest diffusion coefficients.

3.5. Prediction potential

Table 19 displays scatter plots for D , comparing predicted values (D_i) with measured values (D_i^{MD}). These plots include data from both the observed and unobserved datasets used to assess the predictive performance of the MLMs. Ideally, the plots should display a 45° diagonal line, reflecting perfect agreement between the predicted and MD-measured values for both datasets, particularly when evaluating the model using unobserved samples. The complete list of simulated values and their corresponding predictions, based on the input features for each sample, is available in the tabulated files within the *MLMs/predictions* directory in the Supplementary Information.

4. Discussion

4.1. 36 diffusion coefficients

This analysis provides an initial interpretation of the results, focusing

on basic observations to generate insights. It does not explore the deeper correlations possible with machine learning or the effect of chemical descriptors on diffusion. Instead, it focuses on typical Pavement Engineering links between bitumen type, aging, and rejuvenator type, and their influence on diffusion potential.

4.1.1. Rejuvenator dependency

The performance of different rejuvenators in facilitating diffusion into bitumens is clearly temperature-dependent and varies significantly based on the type of rejuvenator used. From across all temperatures tested (60 °C, 120 °C, 160 °C, and 200 °C) (as seen in Fig. 7 and Table 7), **VO** exhibits the highest absolute and relative diffusion rates, consistently ranking the highest nominal values and in the top third of samples. At 160 °C, for instance, 10 out of 15 **VO** samples are found in the top third with an average diffusion coefficient of 1.28e-10 m²/s, as opposed to 0.57 and 0.32e-10 m²/s for the bottom thirds, and about 10 % higher than other rejuvenators' average. Even at lower temperatures, such as 60 °C, **VO** still performs relatively well, with 9 samples in the top category and an average diffusion coefficient of 0.284e-10 m²/s, almost ten times higher than the bottom thirds, and about 30 % higher than other rejuvenators' average in the top third. This trend suggests that **VO** is highly effective in promoting diffusion over a broad temperature range, maintaining strong relative performance even at lower temperatures, where diffusion rates diminish across all samples. Moreover, the performance of **VO** seems to improve with an increase in temperature, where both the magnitudes and distribution of scores become more apparent. Based on the visual reference depicted in Fig. 8, **VO** systems remain "slightly diffused" at 60 °C, an observation that quickly reverses at higher temperatures, where most of the systems pass the "halfway diffused" mixing point, even in cases that are grouped into the bottom thirds. As a rough estimate, the average diffusion coefficient for all **VO**-rejuvenated samples, encompassing all temperatures, bitumen types, and aging conditions, is 1.02e-10 m²/s, equivalent to 0.85 mm/h.

In contrast, **EO** displays a more balanced distribution across the top, middle, and bottom thirds, as seen in Table 7. In nominal terms, its diffusion coefficients also lie below that of **VO** across all temperatures but have a more intricate relationship with **NO** and **AO**. While **EOs** consistently score higher counts in the top thirds across all temperatures, the values of the diffusion coefficients do not dominate over these two, but rather compete very closely. This is especially true at higher temperature and top thirds, where **NO** and **AO**'s coefficients sometimes surpass that of **EO**. Moreover, while **EO**'s top third count aligns with that of **VO** at lower temperatures, they quickly diminish as temperatures increases (from a count of 7 in the top third when **VO** is 9, to a mere 4 when **VO** is 10). These observations suggest that **EO** has a relatively good performance, placing it below **VO**, but above **AO** and **NO**, and

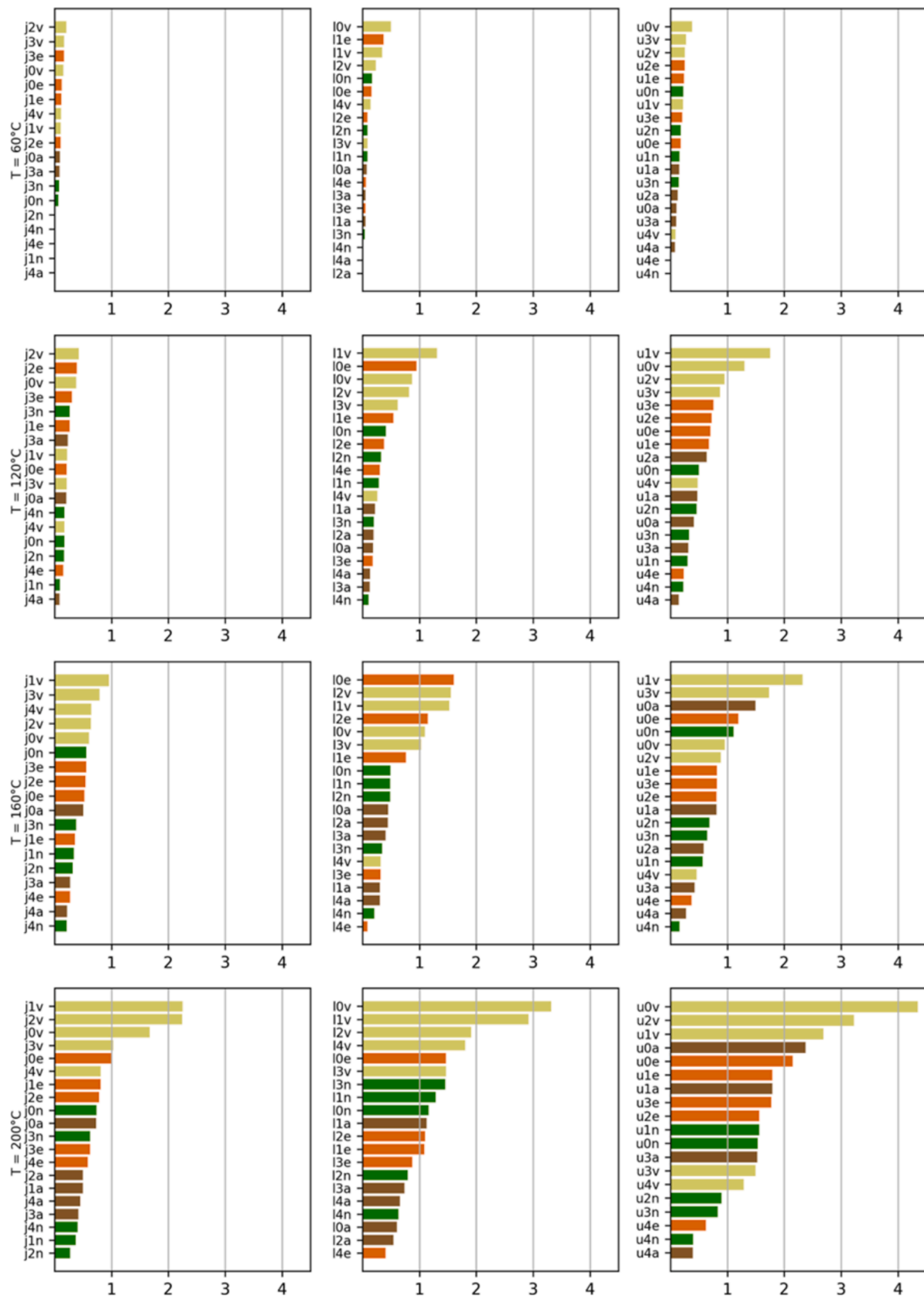


Fig. 7. Fickian Diffusion coefficients measured for every simulation case performed in this study. Light green, Brown, Orange, and Dark green correspond to **VO**, **AO**, **EO**, and **NO** respectively. Values reported are multiplied by a factor of 10^{10} , and are reported in units of m^2/s . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

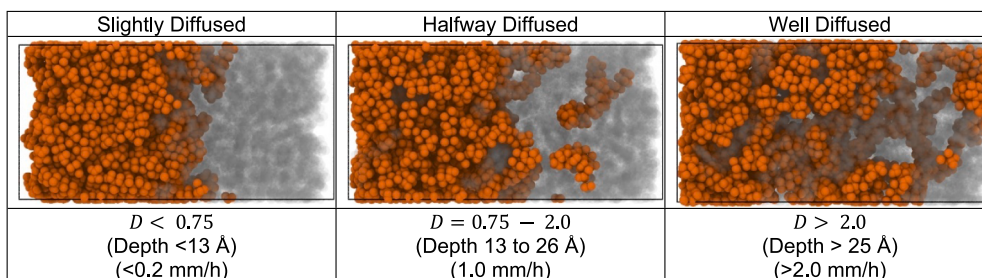


Fig. 8. Renders of the diffusion states of rejuvenators at the end of the simulations. These visualizations illustrate how the diffusion values translate into the extent of mixing achieved by the end of the simulations, aiding in the visualization and interpretation of the results presented in Fig. 7.

Table 7

The samples are categorized by rejuvenator type and temperature, then grouped into three categories based on diffusion rates: the top third with the highest rates, the middle third with average rates, and the bottom third with the lowest rates. The number in parentheses represents the average diffusion coefficient calculated from the sampled data.

T [°C]	Rejuvenator	Top	Middle	Bottom
60	<u>VO</u>	9 (0.284)	4 (0.109)	1 (0.0814)
	<u>EO</u>	7 (0.223)	4 (0.108)	4 (0.0319)
	<u>NO</u>	3 (0.188)	4 (0.118)	8 (0.0265)
	<u>AO</u>	0	6 (0.109)	7 (0.0382)
120	<u>VO</u>	9 (0.994)	2 (0.341)	3 (0.199)
	<u>EO</u>	6 (0.724)	6 (0.31)	3 (0.179)
	<u>NO</u>	2 (0.476)	7 (0.304)	6 (0.151)
	<u>AO</u>	2 (0.555)	3 (0.316)	8 (0.159)
160	<u>VO</u>	10 (1.28)	3 (0.577)	1 (0.319)
	<u>EO</u>	6 (1.06)	4 (0.594)	5 (0.279)
	<u>NO</u>	1 (1.1)	7 (0.556)	7 (0.278)
	<u>AO</u>	2 (1.15)	4 (0.493)	7 (0.314)
200	<u>VO</u>	9 (2.75)	5 (1.22)	0
	<u>EO</u>	4 (1.82)	7 (1.02)	4 (0.558)
	<u>NO</u>	2 (1.54)	7 (1.02)	6 (0.448)
	<u>AO</u>	3 (1.9)	1 (1.13)	10 (0.55)

Table 8

The samples are categorized by bitumen type and temperature, then grouped into three categories based on diffusion rates: the top third with the highest rates, the middle third with average rates, and the bottom third with the lowest rates. The number in parentheses represents the average diffusion coefficient calculated from the sampled data. Aging degree is not considered.

T [°C]	Bitumen type	Top	Middle	Bottom
60	<u>LB</u>	6 (0.293)	5 (0.0948)	9 (0.0371)
	<u>JB</u>	3 (0.177)	8 (0.111)	7 (0.0256)
	<u>UB</u>	10 (0.239)	6 (0.131)	4 (0.0438)
120	<u>LB</u>	6 (0.85)	6 (0.325)	8 (0.165)
	<u>JB</u>	0	7 (0.321)	11 (0.168)
	<u>UB</u>	13 (0.789)	6 (0.298)	1 (0.142)
160	<u>LB</u>	6 (1.32)	6 (0.517)	8 (0.285)
	<u>JB</u>	2 (0.871)	8 (0.569)	8 (0.294)
	<u>UB</u>	11 (1.17)	5 (0.584)	4 (0.306)
200	<u>LB</u>	4 (2.49)	10 (1.18)	6 (0.595)
	<u>JB</u>	3 (2.05)	6 (0.864)	11 (0.495)
	<u>UB</u>	12 (2.19)	4 (1.12)	3 (0.469)

whose detrimental temperature dependency may hinder its relative performance at higher temperatures. As a rough estimate, the average diffusion coefficient for all EO-rejuvenated samples, encompassing all

Table 9

The samples are categorized by bitumen aging degree and temperature, then grouped into three categories based on diffusion rates: the top third with the highest rates, the middle third with average rates, and the bottom third with the lowest rates. The number in parentheses represents the average diffusion coefficient calculated from the sampled data.

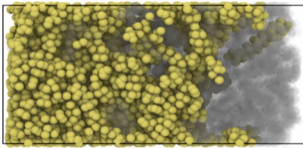
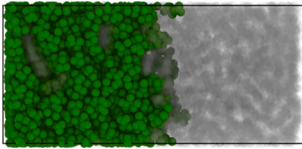
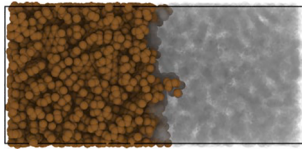
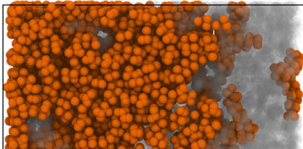
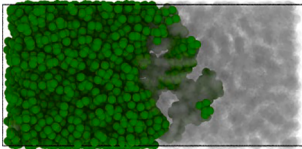
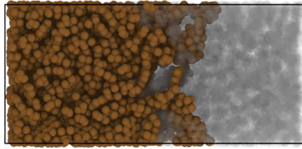
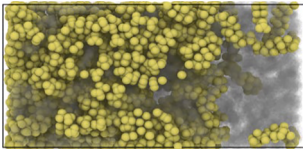
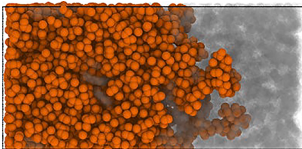
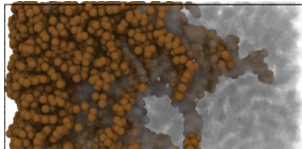
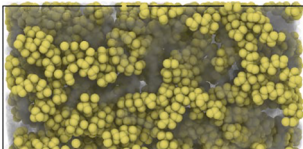
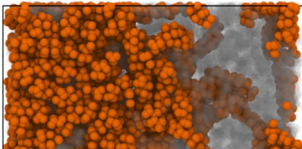
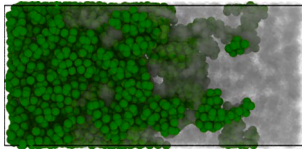
T [°C]	Aging Degree	Top	Middle	Bottom
60	0	6 (0.265)	4 (0.118)	2 (0.0667)
	1	4 (0.292)	5 (0.123)	2 (0.0282)
	2	5 (0.223)	4 (0.101)	2 (0.00741)
	3	4 (0.201)	4 (0.103)	4 (0.0548)
120	4	0	2 (0.124)	10 (0.0264)
	0	5 (0.864)	3 (0.397)	4 (0.192)
	1	5 (0.95)	3 (0.281)	3 (0.174)
	2	5 (0.715)	4 (0.377)	2 (0.178)
160	3	3 (0.747)	5 (0.286)	4 (0.176)
	4	1 (0.474)	4 (0.254)	7 (0.138)
	0	6 (1.24)	6 (0.519)	0 (N/A)
	1	5 (1.28)	3 (0.602)	3 (0.332)
200	2	4 (1.1)	6 (0.561)	1 (0.314)
	3	4 (1.09)	2 (0.596)	6 (0.357)
	4	0	2 (0.547)	10 (0.24)
	0	6 (2.56)	4 (1.1)	2 (0.665)
200	1	6 (2.17)	4 (1.08)	2 (0.431)
	2	4 (2.23)	4 (0.893)	3 (0.436)
	3	2 (1.65)	6 (1.19)	4 (0.599)
	4	1 (1.8)	2 (1.05)	9 (0.503)

temperatures, bitumen types, and aging conditions, is 0.58 e-10 m²/s, equivalent to about 0.64 mm/h.

NO and AO exhibit the poorest performance across all temperatures. Both types are heavily represented in the bottom third of samples, with NO, for example, placing 8 samples in the bottom third at 60 °C, and 6 samples at 200 °C. Similar observations can be said about AO. In nominal terms, the diffusion coefficients are, on average, 30 % lower than those of VO and EO across all observations. NO's performance at lower temperatures is slightly better (both nominally and relatively) than that of AO, an observation that reverses at higher temperatures in favor of AO. In fact, AO's diffusion coefficients experience a drastic increase at higher temperatures, which put it on par with the nominal scores of that of EO. The same cannot be said for NO, whose diffusion coefficients consistently rank the lowest in nominal values. Overall, while NO and AO consistently perform worse than VO and EO, relative temperature susceptibility (for worse in EO and for better in AO) indicate that while temperature may exponentially increase diffusion coefficients, the impact differs when considering relative values among rejuvenators. As a rough estimate, the average diffusion coefficients for NO- and AO-rejuvenated samples, across all temperatures, bitumen types, and aging conditions, are 0.42 and 0.43e-10 m²/s, respectively, corresponding to approximately 0.51 mm/h. While these averages provide a general

Table 10

Renders of the molecular systems at the conclusion of the simulations, illustrating the diffusion of rejuvenators into the bitumen surfaces. For clarity, the table presents only the highest (column 1), average (column 2), and lowest (column 3) diffusivity values across the four temperatures.

T	Highest	Average	Lowest
60 °C			
120 °C			
160 °C			
200 °C			

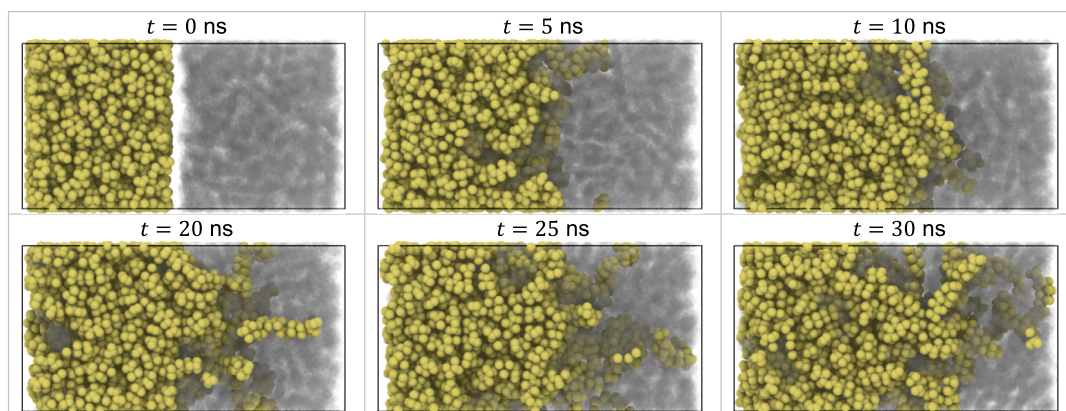


Fig. 9. Renders of sample “l2v” (**LB** bitumen aged for 20 h rejuvenated with **VO**) at six different elapsed simulation times at 120 °C.

Table 11

Averaged thermodynamic properties derived from the molecular models for each SARA fraction and each rejuvenator, alongside their experimental counterparts.

SARA	Experimental (y_i^{exp})					Computational (y_i^{MD})				
	ρ [kg/m ³]	ΔH_{vap} [kJ/kg]	δ_{sol} [kJ ^{0.5} /m ^{1.5}]	C_p [kJ/kg/K]	β [1/K] ($\bullet 10^4$)	ρ [kg/m ³]	ΔH_{vap} [kJ/kg]	δ_{sol} [kJ ^{0.5} /m ^{1.5}]	C_p [kJ/kg/K]	β [1/K] ($\bullet 10^4$)
	[60]	[61–63]	[60,63]	[64–67]	[66–68]					
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 12

Comparison of experimentally and computationally obtained Fickian diffusion coefficients for samples and temperatures where both datasets are available.

Sample	Temperature	$D[m^2/s]$ (10^{10})		
		Experimental	Computational (literature)	Computational (this study)
l3v	160 °C	0.969	3.82	1.027
l3e	160 °C	0.858	3.07	0.817
l3n	160 °C	0.540	2.78	0.341
l3a	160 °C	0.252	1.75	0.404

overview, they do not fully account for the influence of temperature; however, they assist in establishing a preliminary ranking among the rejuvenators.

However, when computing the average across all samples separately for each of the four temperatures (reported in Table 20), the performance order at lower temperatures is **VO**, **EO**, **NO**, and **AO**. At higher temperatures, **AO** slightly surpasses **NO**. In realistic terms, diffusion depths at high temperatures are, on average, an order of magnitude higher than those at lower temperatures.

4.1.2. Bitumen type dependency

The diffusion behavior of bitumen types (**LB**, **JB**, and **UB**) across varying temperatures shows distinct patterns in their ability to promote diffusion. **UB** bitumen consistently exhibits the highest count in the top third group (see Table 8) across all temperature regimes, where it scores above 10 in all observations. Moreover, **UB** also has the lowest bottom-third count, averaging only 4 counts and being the lowest of all bitumen types. Even though **UB** bitumen scores the highest top third counts (and the lowest bottom third counts), the magnitude of its diffusion coefficients in some cases is slightly below that of **LB** bitumens. This observation suggests that, while **UB**'s diffusion potential is superior for the majority of the samples, a localized phenomenon is favoring increased diffusion in a select few **LB** samples.

LB bitumen shows a more balanced distribution between the top, middle, and bottom thirds. The top thirds average a count of 6 (compared to 10 for **UB**, and a mere 2 for **JB**), while the bottom thirds a count of 6 and 8. Interestingly, the relative performance of **LB** bitumens improves at higher temperatures, where the count distribution becomes more skewed toward the top third, especially at 160 °C and beyond. Generally, the diffusion coefficients of **LB** bitumens rank between **UB** and **JB**, but in the top third cases, it surpasses that of other bitumen types across all temperatures. As mentioned earlier, this particularity may be attributed to a phenomenon that cannot be isolated and identified using bitumen types alone, but which is clearly favoring the magnitude of diffusion coefficients in a consistent number of **LB** bitumens samples, equal across all temperatures.

JB bitumen demonstrates the lowest diffusion rates, both relatively and nominally, where the top third counts average at 2, and the bottom third counts average at 9, and the diffusion coefficients are about 20 % lower than in other bitumen types across all observations. This places **JB**'s relative diffusive capacity as the weakest of the three types,

Table 13

Minimum and maximum values for all 19 features measured across all 60 simulations of this study.

x_i	Feature	x_i^{min}	x_i^{max}	x_i	Feature	x_i^{min}	x_i^{max}
Bitumen	1 Saturation	0.5198	0.5689	Rejuvenator	10 Saturation	0.5714	0.6667
	2 Aliphatic	0.219133	0.25364		11 Aliphatic	0.22856	0.35131
	3 Aromatic	0.14195	0.178823		12 Aromatic	0	0.2
	4 Cyclopentane	0.00068	0.0052		13 Cyclopentane	0	0
	5 DoubleBond	0	0.002763		14 DoubleBond	0	0.03509
	6 Oxygen	0.004839	0.067688		15 Oxygen	0	0.07016
	7 Nitrogen	0.001716	0.003279		16 Nitrogen	0	0
	8 Sulfur	0.002112	0.011563		17 Sulfur	0	0
	9 Mol. Weight [g/mol]	431.6466	508.0281		18 Mol. Weight [g/mol]	296.478	400.62
			19 Temperature [°C]	60	200		

consistently observed across all temperatures. In very few cases, especially at higher temperatures, **JB**'s diffusion scores surpass that of other bitumens (as evidenced in the bottom third at 200 °C) – which is too isolated of a case to be explainable analyzing the data available in Fig. 7 and Table 7, requiring a more detailed or fundamental approach. When computing the average diffusion coefficient by considering only bitumen type and temperature, the overall diffusion performance, ranked from high to low, is **UB**, **LB**, and **JB** bitumens across all temperatures, as depicted in Table 21.

4.1.3. Bitumen aging dependency

The data in Fig. 7 and Table 9 illustrate the diffusion behavior of rejuvenated bitumen samples across varying degrees of aging (0 to 4) and temperatures (60 °C to 200 °C). While the counts and diffusion magnitudes are more evenly distributed – making it more difficult to isolate trends – observations can still be made with relative ease. In general, fresher bitumens (those of age 0 and 1) score the highest count in the top third and consistently have the highest diffusion coefficients by magnitude across all temperatures, being, on average, 20 % higher than bitumens of age 2, 3, and 4. Heavily aged bitumens (3 and 4) are considerably underrepresented in the top third, in some cases with counts equal to 0, showcasing that beyond aging degree 2, the diffusive performance considerably lowers. This is especially the case for age 4

Table 14

The list shows the normalized importance scores for features influencing diffusion coefficients, ordered from most to least influential. The sign indicates whether an increase in the feature has a positive or negative impact on the diffusion coefficient. Scores near zero are represented as 0.

Most influential			Least influential		
x_i	Feature	Importance [%]	x_i	Feature	Importance [%]
x_i	Feature	Importance [%]	x_i	Feature	Importance [%]
x_{19}	Temperature	41.94	x_5	Bitumen DoubleBond	-1.29
x_9	Bitumen Mw	-19.04	x_{10}	Rejuvenator Saturation	1.11
x_{18}	Rejuvenator Mw	-8.51	x_1	Bitumen Saturation	1.11
x_{15}	Rejuvenator Oxygen	6.44	x_4	Bitumen Cyclopentane	1.10
x_{14}	Rejuvenator DoubleBond	6.12	x_2	Bitumen Aliphatic	-0.97
x_8	Bitumen Sulfur	5.66	x_{11}	Rejuvenator Aliphatic	0.70
x_6	Bitumen Oxygen	-2.71	x_{12}	Rejuvenator Aromatic	0.19
x_3	Bitumen Aromatic	1.56	x_{13}	Rejuvenator Cyclopentane	0.00
x_7	Bitumen Nitrogen	-1.53	x_{16}	Rejuvenator Nitrogen	0.00
			x_{17}	Rejuvenator Sulfur	0.00

Table 15

Top five, middle five, and bottom five systems based on diffusion coefficients, along with key influential features identified by the MLM. Results obtained at 60 °C.

Group	sample	x_8	x_{17}	x_{14}	x_{13}	x_7	x_5	x_2	x_6	x_4	$D[m^2/s] (10^{10})$
Highest	u3-v	461	296	0.0702	0.0351	0.0085	0.0188	0.166	0.00266	0.00177	0.272
	l1-v	453	296	0.0702	0.0351	0.00995	0.00532	0.173	0.00172	0.00206	0.341
	l1-e	453	309	0	0	0.00995	0.00532	0.173	0.00172	0.00206	0.369
	u0-v	432	296	0.0702	0.0351	0.0116	0.00958	0.161	0.00271	0.00181	0.374
	l0-v	448	296	0.0702	0.0351	0.0105	0.00484	0.171	0.00173	0.00207	0.497
Average	j2-e	475	309	0	0	0.0037	0.0225	0.148	0.00241	0.000161	0.103
	u0-a	432	401	0	0	0.0116	0.00958	0.161	0.00271	0.00181	0.103
	j1-v	474	296	0.0702	0.0351	0.00376	0.018	0.155	0.00245	0	0.106
	j4-v	489	296	0.0702	0.0351	0.00211	0.0349	0.142	0.0026	0	0.111
	j1-e	474	309	0	0	0.00376	0.018	0.155	0.00245	0	0.119
Lowest	l2-a	469	401	0	0	0.00544	0.0119	0.176	0.00238	0.00204	0.00291
	l4-a	508	401	0	0	0.00637	0.0677	0.179	0.0031	0.00276	0.00319
	l4-n	508	361	0	0	0.00637	0.0677	0.179	0.0031	0.00276	0.0051
	u4-n	487	361	0	0	0.00788	0.025	0.167	0.00268	0.00201	0.0058
	j4-a	489	401	0	0	0.00211	0.0349	0.142	0.0026	0	0.00688

Table 16

Top five, middle five, and bottom five systems based on diffusion coefficients, along with key influential features identified by the MLM. Results obtained at 120 °C.

Group	sample	x_8	x_{17}	x_{14}	x_{13}	x_7	x_5	x_2	x_6	x_4	$D[m^2/s] (10^{10})$
Highest	u2-v	446	296	0.0702	0.0351	0.00856	0.0142	0.165	0.00328	0.00219	0.949
	l0-e	448	309	0	0	0.0105	0.00484	0.171	0.00173	0.00207	0.95
	u0-v	432	296	0.0702	0.0351	0.0116	0.00958	0.161	0.00271	0.00181	1.3
	l1-v	453	296	0.0702	0.0351	0.00995	0.00532	0.173	0.00172	0.00206	1.31
	u1-v	439	296	0.0702	0.0351	0.00823	0.0135	0.169	0.00274	0.00183	1.75
Average	u1-n	439	361	0	0	0.00823	0.0135	0.169	0.00274	0.00183	0.296
	l4-e	508	309	0	0	0.00637	0.0677	0.179	0.0031	0.00276	0.301
	j3-e	487	309	0	0	0.00309	0.0234	0.142	0.00244	0.00276	0.304
	u3-a	461	401	0	0	0.0085	0.0188	0.166	0.00266	0.00177	0.308
	l2-n	469	361	0	0	0.00544	0.0119	0.176	0.00238	0.00204	0.323
Lowest	j4-a	489	401	0	0	0.00211	0.0349	0.142	0.0026	0	0.0862
	j1-n	474	361	0	0	0.00376	0.018	0.155	0.00245	0	0.0941
	l4-n	508	361	0	0	0.00637	0.0677	0.179	0.0031	0.00276	0.107
	l3-a	481	401	0	0	0.00584	0.0326	0.175	0.00258	0.00275	0.121
	l4-a	508	401	0	0	0.00637	0.0677	0.179	0.0031	0.00276	0.131

Table 17

Top five, middle five, and bottom five systems based on diffusion coefficients, along with key influential features identified by the MLM. Results obtained at 160 °C.

Group	sample	x_8	x_{17}	x_{14}	x_{13}	x_7	x_5	x_2	x_6	x_4	$D[m^2/s] (10^{10})$
Highest	l1-v	453	296	0.0702	0.0351	0.00995	0.00532	0.173	0.00172	0.00206	1.53
	l2-v	469	296	0.0702	0.0351	0.00544	0.0119	0.176	0.00238	0.00204	1.55
	l0-e	448	309	0	0	0.0105	0.00484	0.171	0.00173	0.00207	1.6
	u3-v	461	296	0.0702	0.0351	0.0085	0.0188	0.166	0.00266	0.00177	1.73
	u1-v	439	296	0.0702	0.0351	0.00823	0.0135	0.169	0.00274	0.00183	2.32
Average	j0-e	470	309	0	0	0.0047	0.0164	0.172	0.00244	0	0.52
	j2-e	475	309	0	0	0.0037	0.0225	0.148	0.00241	0.000161	0.542
	j3-e	487	309	0	0	0.00309	0.0234	0.142	0.00244	0.00276	0.553
	j0-n	470	361	0	0	0.0047	0.0164	0.172	0.00244	0	0.555
	u1-n	439	361	0	0	0.00823	0.0135	0.169	0.00274	0.00183	0.563
Lowest	l4-e	508	309	0	0	0.00637	0.0677	0.179	0.0031	0.00276	0.0844
	u4-n	487	361	0	0	0.00788	0.025	0.167	0.00268	0.00201	0.16
	l4-n	508	361	0	0	0.00637	0.0677	0.179	0.0031	0.00276	0.206
	j4-n	489	361	0	0	0.00211	0.0349	0.142	0.0026	0	0.211
	j4-a	489	401	0	0	0.00211	0.0349	0.142	0.0026	0	0.215

bitumens (the most aged), as they consistently score 0 in the top third, and 10 in the bottom third, regardless of temperature. The diffusion coefficient of heavily aged bitumens (age 3 and 4) are on average 30 % below those of other aging degrees, showcasing that this reduction in

diffusive performance is not relative to the group (i.e., the third they are grouped into) but also in nominal terms. Interestingly, bitumens aged to degree 1 (shortly aged) display a smaller count in top thirds, but those that do show, have higher diffusion magnitudes than those of bitumens

Table 18

Top five, middle five, and bottom five systems based on diffusion coefficients, along with key influential features identified by the MLM. Results obtained at 200 °C.

Group	sample	x ₈	x ₁₇	x ₁₄	x ₁₃	x ₇	x ₅	x ₂	x ₆	x ₄	D[m ² /s] (10 ¹⁰)
Highest	u1-v	439	296	0.0702	0.0351	0.00823	0.0135	0.169	0.00274	0.00183	2.69
	l1-v	453	296	0.0702	0.0351	0.00995	0.00532	0.173	0.00172	0.00206	2.92
	u2-v	446	296	0.0702	0.0351	0.00856	0.0142	0.165	0.00328	0.00219	3.23
	l0-v	448	296	0.0702	0.0351	0.0105	0.00484	0.171	0.00173	0.00207	3.32
	u0-v	432	296	0.0702	0.0351	0.0116	0.00958	0.161	0.00271	0.00181	4.35
Average	u2-n	446	361	0	0	0.00856	0.0142	0.165	0.00328	0.00219	0.892
	j0-e	470	309	0	0	0.0047	0.0164	0.172	0.00244	0	1.01
	j3-v	487	296	0.0702	0.0351	0.00309	0.0234	0.142	0.00244	0.00276	1.03
	l10e	453	309	0	0	0.00995	0.00532	0.173	0.00172	0.00206	1.09
	l2-e	469	309	0	0	0.00544	0.0119	0.176	0.00238	0.00204	1.1
Lowest	j2-n	475	361	0	0	0.0037	0.0225	0.148	0.00241	0.000161	0.269
	j1-n	474	361	0	0	0.00376	0.018	0.155	0.00245	0	0.369
	u4-a	487	401	0	0	0.00788	0.025	0.167	0.00268	0.00201	0.389
	u4-n	487	361	0	0	0.00788	0.025	0.167	0.00268	0.00201	0.394
	j4-n	489	361	0	0	0.00211	0.0349	0.142	0.0026	0	0.4

Table 19

Scatter plots of y₁ 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 MLM true prediction potential.

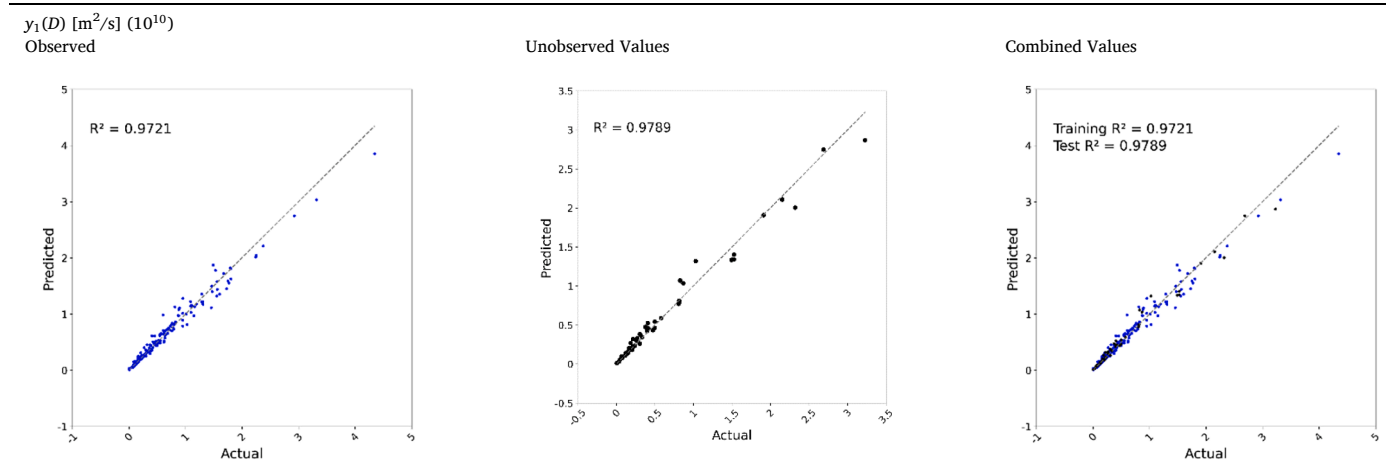


Table 20

Mean diffusion coefficients across all bitumen types and ages, grouped by both temperature and rejuvenator type.

Diffusion Coefficients [m ² /s] (Equivalent depth rate [mm/h])				
Temperature [°C]	60	120	160	200
<u>VO</u>	0.21 (0.39)	0.71 (0.71)	1.03 (0.86)	2.16 (1.25)
<u>EO</u>	0.14 (0.32)	0.45 (0.57)	0.68 (0.70)	1.11 (0.89)
<u>NO</u>	0.08 (0.24)	0.27 (0.44)	0.46 (0.58)	0.86 (0.79)
<u>AO</u>	0.07 (0.23)	0.26 (0.43)	0.50 (0.60)	0.88 (0.80)

in their fresh state. This could be attributed to the slight increase in polar centers that oxidized groups bring, increasing the diffusion potential to a certain extent, especially in bitumens with lower counts of polar groups (JB).

Table 9 indicates that aging reduces the susceptibility of bitumen to the effects of temperature increases on diffusion. Fresh and slightly aged bitumen (aging degrees 0, 1, and 2) exhibit significant improvements in diffusion coefficients with rising temperatures, as evidenced by fresh bitumen's increase from 0.265 at 60 °C to 2.56e-10 m²/s at 200 °C. However, as bitumen ages, the impact of temperature diminishes. Moderately aged bitumen (Aging degree 3) shows a less pronounced but still notable increase, from 0.201 at 60 °C to 1.65e-10 m²/s at 200 °C. In contrast, highly aged bitumen (Aging degree 4) demonstrates the

Table 21

Mean diffusion coefficients across all rejuvenator types and ages, grouped by both temperature and bitumen type.

Diffusion Coefficients [m ² /s] (Equivalent depth rate [mm/h])				
Temperature [°C]	60	120	160	200
<u>JB</u>	0.09–0.25	0.23–0.40	0.48–0.59	0.84–0.78
<u>LB</u>	0.13–0.30	0.42–0.55	0.67–0.69	1.27–0.96
<u>UB</u>	0.17–0.35	0.61–0.66	0.85–0.78	1.70–1.10

weakest response to temperature, with diffusion coefficients rising from negligible levels at 60 °C to only 1.8e-10 m²/s at 200 °C. This suggests that aging progressively limits the bitumen's ability to rejuvenate through thermal effects, resulting in diminished sensitivity to temperature increases in more aged samples.

The diffusion coefficient of highly aged bitumen (aging degree 4) at elevated temperatures bears resemblance to that of fresh bitumen (aging degree 0) at significantly lower temperatures. For instance, the diffusion coefficient of aged bitumen at 200 °C, which reaches an average of 1.00e-10 m²/s, mirrors the performance of fresh bitumen at much lower temperatures, akin to what might be observed at 120 °C. Similarly, the diffusion coefficient of moderately aged bitumen (aging degree 3) at 120 °C, with an average around 0.50e-10 m²/s, closely resembles that of

fresh bitumen at 60 °C, highlighting how aging effectively shifts the behavior of bitumen toward what would be observed under lower thermal conditions in fresher samples. The observation that higher aging degrees hinder diffusion potential is well-documented in the literature. Greater interest would be in analyzing the information to illustrate the same relationship while considering bitumen type. According to Fig. 10, on average, **UB** exhibits the best diffusion performance, followed by **LB**, with **JB** showing the lowest performance. Interestingly, **LB** performance at high temperatures and aging degrees (greater than 160 °C and at Ages 3 and 4) surpasses that of **UB**. Such behavioral shifts are not observed in other combinations of bitumen types. Generally, the ranking is **UB** > **LB** > **JB**; however, under higher aging degrees and temperatures, **LB** aligns with **UB**, and in extreme cases, surpasses it. Therefore, in these conditions, the ranking becomes **LB** > **UB** > **JB**.

4.2. Machine learning predictions

The objectives of using MLM predictions are threefold: First, to provide detailed insights into the chemical and conditional features that most significantly influence the Fickian diffusion coefficient. Second, to reduce reliance on continuous MD simulations for predicting diffusion coefficients by offering near-instantaneous predictions comparable to those obtained from MD simulations. Third, to equip researchers with a predictive tool (i.e., a model) capable of estimating the diffusion coefficients of bitumen-rejuvenator systems not included in this study but that sharing many chemical features, as is common with various hydrocarbon blends used in the pavement industry. The subsequent section will utilize these MLMs to identify the most important chemical features, examine their effects on the diffusive performance of bitumen and rejuvenators, and explore their applications in characterizing, designing, and improving future bitumen-rejuvenator systems.

4.2.1. 41 Features Importance

4.2.1.1. 42 Temperature. Temperature is observed to be the primary driver of diffusion, with a positive influence of 41.94 % as shown in Table 14, confirming the relationship between increased diffusion coefficients at elevated temperatures. This observation is supported by fundamental principles and is also reflected in experimentally obtained diffusion coefficients from comparable setups, where Fickian diffusion exhibits an exponential relationship with temperature.

4.2.1.2. Bitumen molecular weight. The molecular weight of bitumen negatively influences diffusion, with an influence factor of −19.04 %. Larger molecular masses correlate with reduced diffusion, which can be attributed to the increased size of the molecules, resulting in decreased molecular mobility, reduced accessible free volume, and the formation of more complex intermolecular structures. For example, the u0-, u1-, and even u2-x bitumens, with the lowest overall molecular weights (431–446 g/mol), display the highest diffusion, while 14- and j4-x

bitumens, which have the highest molecular weights (488–508 g/mol), show the lowest diffusion. This inverse relationship between molecular weight and diffusion aligns with the expected hindrance caused by larger molecular structures and their resulting decreased intermolecular mobility. A more detailed example shows that, at 60 °C, the average diffusion coefficient for u0-x bitumens is 0.218, for u4-x it is 0.043, for j0-x it is 0.107, for j4-x it is 0.033, for l0-x it is 0.222, and for l4-x it is 0.050e-10 m²/s. This suggests a ratio of five times between non-aged and fully aged bitumens across all types and rejuvenator categories. The ratio is similar but decreases when comparing aging degrees that are closer to one another, such as between j0-x and j2-x bitumens.

4.2.1.3. Rejuvenator molecular weight. The molecular weight of the rejuvenator also negatively impacts diffusion, with an influence factor of −8.51 %; however, this effect is less pronounced than that of bitumen molecular weight, which has an influence factor of −19.04 %. While the underlying reasoning for these effects is similar, the reduced impact of the rejuvenator can be attributed to the more complex structure of bitumen, which generally has higher average molecular weights. This complexity means that bitumen molecules are more sensitive to changes in molecular mobility and accessible space compared to rejuvenator molecules, which in turn enhances the sensitivity and influence of the former. For example, at 60 °C, fresh bitumens of all types show that **VO** systems (the rejuvenator with the lowest molecular weight) have a diffusion coefficient of 0.340, **EO** systems have 0.153, **NO** systems have 0.149, and **AO** systems have 0.0879e-10 m²/s. At 200 °C, the diffusion coefficients are as follows: **VO** at 3.11, **EO** at 1.54, **NO** at 1.14, and **AO** at 1.23e-10 m²/s. The smaller molecular structure of rejuvenators enhances mixing and mobility within the bitumen matrix. Notably, as indicated in **Rejuvenator dependency**, the temperature sensitivity observed between **AO** and **NO** is evident here as well; at lower temperatures, **NO** performs better, while **AO** quickly matches its performance at higher temperatures.

4.2.1.4. Rejuvenator oxygen content. Oxygen content in the rejuvenator is observed to positively affect diffusion, with an influence factor equal to 6.44 %, likely due to the increased polarity and better miscibility introduced by oxygen-containing functional groups. The presence of oxygen groups requires careful consideration, as an increase in oxygen features (x_{14}) can significantly increase the stiffness of hydrocarbons, as reported in the literature, thereby impeding their diffusion potential. Consequently, only modest increases in rejuvenator oxygen features appear to be advantageous, particularly when the molecular structure contributes to preventing stiffness, such as through a high number of aliphatic branches and/or minimal presence of aromatic rings.

An interesting observation arises when comparing the oxygen content of rejuvenators to bitumen samples with varying levels of oxygen. The differences between the oxygen features of bitumen (x_5) and rejuvenators (x_{14}) are calculated and plotted against the measured diffusion coefficients. Initially, this analysis showed no clear trend across all

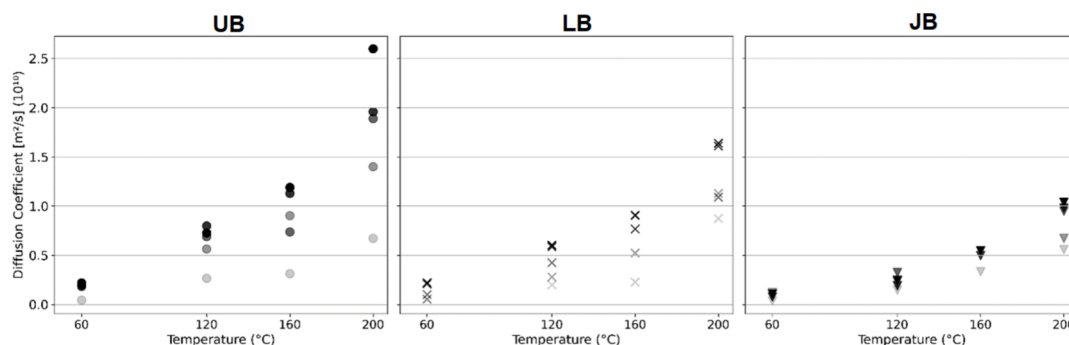


Fig. 10. Average diffusion coefficient of all samples, accounting for bitumen type, aging degree and temperature.

samples. However, previous insights suggest that increasing the oxygen content in the rejuvenator can enhance diffusion up to a certain point, after which it becomes detrimental. Taking this into account, the data was analysed by grouping bitumen samples according to their aging degree. For samples with aging degrees 0 and 1 (less aged bitumen), a clear trend emerged: smaller differences in oxygen content (smaller x_5-x_{14} values) significantly promote diffusion. In contrast, for samples with aging degrees of 2 and higher (more aged bitumen), this trend disappears, and no new patterns are observed. In practical terms, matching oxygen content between bitumens and rejuvenators can be advantageous up to a certain level of aging. Fundamentally, this may be due to the smaller differences in solubility parameters between bitumens and rejuvenators, as solubility is strongly influenced by the presence of polar groups, including oxygen. Compounds with similar solubility parameters tend to mix more effectively than those with larger differences, not accounting for the general stiffening that oxidation may bring to hydrocarbons of this nature.

4.2.1.5. Bitumen sulfur content. The sulfur content in bitumen demonstrates a moderate positive effect on diffusion, quantified at +5.2 %. For example, at 60 °C, u0-x bitumens, which have the highest sulfur content ($x_7 = 0.0115$), exhibit a diffusion coefficient of $0.218e-10 \text{ m}^2/\text{s}$. In contrast, 10-x bitumens ($x_7 = 0.0105$) show a coefficient of 0.155, while j0-x bitumens ($x_7 = 0.0046$) display a value of $0.107e-10 \text{ m}^2/\text{s}$. At 200 °C, the diffusion coefficients for these bitumens are 2.590, 1.638, and $1.037e-10 \text{ m}^2/\text{s}$, respectively.

This trend is consistent across all temperature regimes, indicating that the diffusion coefficient is highly sensitive to minor variations in sulfur content, particularly when the sulfur features fluctuate from just 1.1 % to 0.5 % by number and 1 to 5 % by weight. Similar patterns are observed in bitumens subjected to more extensive aging; however, the sensitivity diminishes significantly. For example, a reduction in sulfur content in aged bitumens, specifically comparing **UB** and **JB** bitumen, leads to a slight decrease in diffusion, from 0.673 to 0.559 at 200 °C. This trend is consistent across all temperatures and becomes more pronounced with increasing temperature.

4.2.1.6. Rejuvenator double bonds. Carbon DoubleBonds in the rejuvenator, though contributing positively to diffusion, have a moderate positive influence (measured at +6.12 %), similar to those of other factors such as molecular weight and oxygen content. For instance, **VO** contains small amounts of Double Bonds (as indicated by feature 14), which may slightly enhance diffusion, though this effect is secondary. Nevertheless, rejuvenators with more isolated double bonds could potentially see a marginal increase in diffusion, as these structural features may improve the reactivity or flexibility of the rejuvenator.

4.2.2. Maximize/Minimize diffusion coefficients

Given the insights obtained in the Features Importance section along with the summarized factors of **Table 14**, there are 9 most influential features that account for a little over 95 % of the influence, namely features x_8 , x_{17} , x_{14} , x_{13} , x_7 , x_5 , x_2 , x_6 , excluding temperature. In essence, to maximize diffusion coefficients, chemical systems must: 1) be maintained at the highest possible temperature; 2) utilize bitumens and rejuvenators with the lowest average molecular weight; 3) maximize the polarizable features of rejuvenators, while excluding those that significantly promote molecular stacking; 4) maximize sulfur features in bitumen while minimizing oxygen features; and 5) preferentially favor the presence of aliphatic features over aromatic features in bitumen, while avoiding excessive nitrogen content. Conversely, the opposite strategies should be employed to minimize diffusion. More formally, these features must be adjusted according to the directionality indices shown in **Table 14**.

Table 22 presents the averaged values of the top 5, middle 5, and bottom 5 samples from **Table 15** through **Table 18**, highlighting the

relationship between key chemical features and diffusion coefficients. By summarizing the data across these samples, **Table 22** illustrates how maximizing or minimizing the influential features directly impacts the diffusion coefficient. This table demonstrates that the diffusion coefficient can be reliably predicted by adjusting the values of the most significant features, as identified by the MLM. The directionality (maximize or minimize) and weights of these features, as detailed in **Table 14**, further clarify the influence of each feature on diffusion trends observed in the experimental samples. Notably, these results are based on actual experimental observations rather than predictions from the MLM, ensuring that there is no bias toward features deemed significant by the model. The trends remain consistent across all tested samples, with no deviations or shifts in behavior. Notably, the Diffusion Coefficient values reported in **Table 22** are extracted from MD simulations rather than predicting them using the MLM, ensuring that the reported values are accurate and free from potential bias introduced by the MLM.

5. Case studies

The scope of experimental observations in this field is limited, and the number of potential case studies that can be conceived is vast. Nonetheless, this section aims to introduce a few illustrating examples aimed at showcasing the potential use of MLMs to predict bitumen-rejuvenator diffusion potential while remaining relevant to Pavement Engineering applications.

5.1. Example 1

The Discussion section indicates that the diffusive potential of **EO** is occasionally comparable to that of **VO**, and in most instances, both are categorized as the highest performing rejuvenators among the four tested. This study aims to introduce simple chemical modifications to the structure of **EO** to assess whether its diffusive performance can be enhanced to match or exceed that of **VO**.

The section on **Rejuvenator Oxygen Content** indicates that to optimize diffusion, two key features of the rejuvenator should be modified: molecular weight, which should be reduced, and oxygen content, which should be increased. Contrastingly, altering the molecular structure by substituting Sp^3 carbons with oxygens typically leads to an increase in molecular weight. To address this, it is essential to maximize the number of carbon and hydrogen atoms while minimizing the number of oxygen atoms introduced.

One effective strategy is to incorporate carboxylic groups into the octyl branches of **EO**, similar to the case of **VO**. This results in two modified structures: Modified 1, which contains one carboxylic group in one branch, and Modified 2, which includes two carboxylic groups—one in each branch. The introduction of carboxylic groups serves a dual purpose: it increases the number of “DoubleBond” features, thereby enhancing diffusion, and it also adds “Oxygen” features. Consequently, this strategy effectively mitigates the negative impact associated with increased molecular weight. These changes are reflected in **Fig. 11**.

The comparative analysis between the **EO**, two modified structures, and **VO** highlights distinct differences in performance across the tested temperatures (see **Table 23**). At 60 °C and 120 °C, both modified structures exhibit slight improvements over the **EO**, with Modified 2 showing a clearer advantage, while the **VO** consistently outperforms all other structures, particularly at lower temperatures. As temperatures increase, Modified 2 begins to show a more significant improvement, particularly at 160 and 200 °C, where it surpasses both Modified 1 and **EO**, and approaching the performance of the **VO**. This indicates that Modified2 is the most promising candidate for improving **EO** rejuvenator performance at higher temperatures, offering clear advantages over the base molecule and Modified1, while Modified1 shows only marginal benefits overall. It is important to note that these modifications focus solely on diffusion performance and do not consider the potential effects of highly oxygenated compounds on the properties of the base

Table 22

Summary of chemical feature values and their relationship with diffusion coefficient across four temperature regimes (60, 120, 160, and 200 °C). 'Minimize' features should have lower values and 'Maximize' features higher values for increased diffusion. The table highlights how following these conditions impacts the resulting diffusion coefficient, with the 'top5' category aiming to maximize diffusion and the 'bottom5' to minimize it. Green indicates maximized features, and red indicates minimized features. Values in parenthesis display the overall feature range.

Temp. [°C]	Group	Most influential Features (Chemical only)						Diffusion [m ² /s] (10 ¹⁰)
		x_8	x_{17}	x_{14}	x_{13}	x_7	x_5	
		Minimize (431.6–508.0)	Minimize (296.4–400.6)	Maximize (0–0.0702)	Maximize (0–0.0351)	Maximize (0.002–0.011)	Minimize (0.0048–0.068)	
60	Top	449.4	298.6	0.0562	0.0281	0.0101	0.0088	0.371
	Middle	468.8	322.2	0.0281	0.0140	0.0050	0.0206	0.108
	Bottom	492.2	385.0	0.0000	0.0000	0.0056	0.0414	0.005
120	Top	443.6	298.6	0.0562	0.0281	0.0098	0.0095	1.250
	Middle	472.8	348.2	0.0000	0.0000	0.0063	0.0271	0.306
	Bottom	492.0	385.0	0.0000	0.0000	0.0049	0.0442	0.107
160	Top	454.0	298.6	0.0562	0.0281	0.0085	0.0109	1.746
	Middle	468.2	329.8	0.0000	0.0000	0.0049	0.0184	0.547
	Bottom	496.2	358.6	0.0000	0.0000	0.0050	0.0460	0.175
160	Top	443.6	296.0	0.0702	0.0351	0.0098	0.0095	3.302
	Middle	465.0	316.8	0.0140	0.0070	0.0063	0.0142	1.024
	Bottom	482.4	369.0	0.0000	0.0000	0.0051	0.0251	0.364

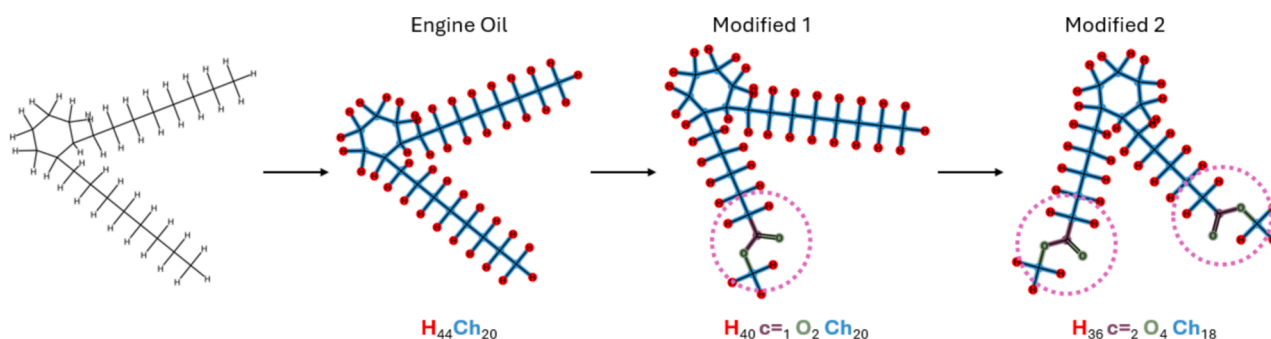


Fig. 11. Identification of chemical features in the molecular structures of **EO**, Modified 1, and Modified 2.

Table 23

Diffusion values predicted using the MLM to assess whether the modified molecular structures can surpass the diffusion performance of **EO**.

Temperature (°C)	MLM predicted diffusion [m ² /s] (10 ¹⁰)			
	EO (Base)	Modified1	Modified2	VO
60	0.22	0.24	0.25	0.36
120	0.75	0.74	0.83	1.22
160	1.18	1.17	1.31	1.28
200	2.11	2.14	3.27	3.85

bitumen as a bulk material, which could negatively impact other performance metrics.

5.2. Example 2

The **Results** section of this manuscript showed that **VO** shows the highest diffusive performance as a rejuvenator. However, this analysis does not consider the potential negative effects **VO** may have on other performance metrics. For example, while **VO** significantly softens bitumen, altering its rheological properties, this softening may not always be desirable, especially in regions of the bitumen that have not undergone aging. Although softening is an intended effect of rejuvenation, excessive softening in unaged areas could compromise the material's overall performance.

The goal of this example is to integrate the predictive models developed by Assaf et. Al. [29], which estimate 12 thermophysical properties of various bituminous materials, including bitumen-rejuvenator blends, with the diffusion insights obtained in this study. By combining these approaches, the effect of rejuvenator concentration on the bulk properties of a bitumen layer at different depths and times can be evaluated. This process is depicted in Fig. 12.

To implement this, the chemical features of both the bitumen and rejuvenator are used to generate a concentration profile where the rejuvenator concentration is progressively increased from 0 % to 100 %. These features are then input into the MLMs to predict the 12 thermo-physical properties for each system. Fick's second law of diffusion, together with the diffusion coefficient predicted by the MLMs in this manuscript, is applied to determine how concentration, depth, and time influence the bulk properties of the rejuvenated bitumen. This process is demonstrated using the "u3v" sample at 60 °C under the following conditions:

- Rejuvenator concentration (by wt.%) after 8 h at a depth of 1.0 mm and the bulk properties of the bitumen-rejuvenator under these conditions.
- Time required to reach a rejuvenator concentration of 5 % by weight at a depth of 1.0 mm and the bulk properties of the bitumen-rejuvenator under these conditions.

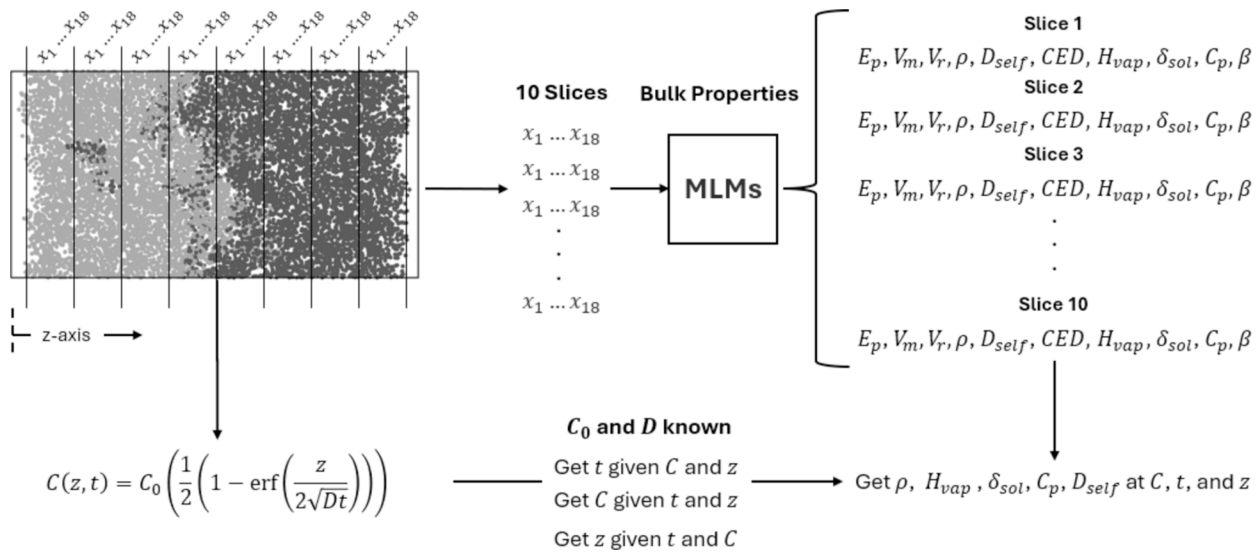


Fig. 12. The MLMs developed by Assaf et al. [29] can be used to predict thermophysical properties of regional slices cut along the diffusion direction (along the z -axis) to investigate the time- and concentration dependency of the material properties. Out of the 12 properties available for prediction, five are selected for computation due to their relevance at the engineering scale: Density, Heat of Vaporization, Heat Capacity, Solubility Parameter, and the Self-diffusion Coefficient. The bulk properties, for each “slice” system at different depths along z , are presented in Table 24.

- c) Depth at which the rejuvenator concentration reaches 10 % by weight after 8 h and the bulk properties of the bitumen-rejuvenator under these conditions.

5.2.1. Case A: Solving for concentration

Knowing z , t , and D , $C(z, t)$ is given by directly evaluating Eq. (2.7). Plugging $C_0 = 849.7 \text{ kg/m}^3$, $z = 1 \text{ mm}$, $t = 1 \text{ h}$, and the computed diffusion coefficient for sample “u3v”, $D = 0.272\text{e-}10 \text{ m}^2/\text{s}$, results in a rejuvenator concentration of 18.52 % by weight, and a mixture-wide density of 973.6 kg/m^3 . At such bitumen-rejuvenator proportions (refer to Table 24, “Slice7”), the values of H_v , c_p , δ_{sol} , and D_{self} are equal to 311.3 kJ/kg , 4.61 kJ/kg/K , $523.4 \text{ kJ}^{0.5}/\text{m}^{1.5}$, and $2.72\text{e-}9 \text{ m}^2/\text{s}$, respectively, showcasing how the thermophysical properties of the mixture change and can be obtained with respect to concentration changes, depth, and/or time.

5.2.2. Case B: Solving for time

Knowing z , $C(z, t)$, and D , t is given by rearranging Eq. (2.7) into Eq. (5.1), below:

$$t = \frac{z^2}{4D(\operatorname{erf}^{-1}(\alpha))^2} \text{ where } \alpha = 1 - \frac{2C(z, t)}{C_0} \quad (5.1)$$

Plugging the same values reported in 5.2.1, the time required for the rejuvenator to reach a concentration of 5 % by weight at 1 mm, is equal

to about 2 h and 10 min. The bulk properties of such a mixture – ρ , H_v , C_p , δ_{sol} , and D_{self} – are equal to 997.8 kg/m^3 , 309.1 kJ/kg , 4.43 kJ/kg/K , $577.5 \text{ kJ}^{0.5}/\text{m}^{1.5}$, and $2.54 \text{ e-}9 \text{ m}^2/\text{s}$, respectively (by interpolation between “Slice 9” and “Slice 10” in Table 24).

5.2.3. Case C: Solving for depth

Knowing t , $C(z, t)$, and D , z is given by rearranging Eq. (2.7) into Eq. (5.2), below:

$$z = 2\sqrt{Dt} \operatorname{erf}^{-1}(\alpha) \text{ where } \alpha = 1 - \frac{2C(z, t)}{C_0} \quad (5.2)$$

Plugging the same values reported in 5.2.1, the depth at which the rejuvenator concentration reaches 10 % by weight after 8 h is 4.60 mm. The bulk properties of such a mixture – ρ , H_v , C_p , δ_{sol} , and D_{self} – are equal to 992.9 kg/m^3 , 309.2 kJ/kg , 4.45 kJ/kg/K , $575.6 \text{ kJ}^{0.5}/\text{m}^{1.5}$, $2.56\text{e-}9 \text{ m}^2/\text{s}$, respectively (by interpolation between “Slice 8” and “Slice 9” in Table 24).

These analyses can be customized and can be done in near-instantaneous times, as all the inputs are obtainable or predictable using the results and MLMs of this study across all samples and temperatures presented, and more importantly, on samples that include unobserved combinations in their chemical structures (such as obtaining the bulk properties of bitumen-rejuvenator combinations in “slice” regions). This is particularly powerful as most of these properties do not scale linearly and weighted averages cannot be used to estimate them given the initial concentration of both species.

Table 24

MLM-predicted values for Density, Heat of Vaporization, Heat Capacity, Solubility Parameter, and self-diffusion coefficient as a function of simulation box’s depth.

Slice Number	Rejuvenator[vol.%]	Rejuvenator[wt.%]	ρ [kg/m ³]	H_v [kJ/kg]	C_p [kJ/kg/K]	δ_{sol} [kJ ^{0.5} /m ^{1.5}]	D_{self} [m ² /s]
Slice0	100	100.0	849.7	342.2	5.507	495.9	6.774
Slice1	96.5	95.9	855.2	340.6	5.448	497.6	6.222
Slice2	92.3	91.0	861.8	331.1	5.353	498.8	5.844
Slice3	86.5	84.3	871.0	326.8	5.129	501.8	5.016
Slice4	63.9	59.9	906.5	321.4	4.998	508.9	4.540
Slice5	39.4	35.4	945.0	315.5	4.810	518.8	3.440
Slice6	25.9	22.8	966.3	314.9	4.674	520.3	2.887
Slice7	21.4	18.6	973.4	311.3	4.611	523.4	2.724
Slice8	14.2	12.3	984.6	310.6	4.497	544.3	2.720
Slice9	9.00	7.7	992.9	309.2	4.452	575.6	2.562
Slice10	0.00	0.0	1007	309.0	4.402	581.2	2.489

5.3. Example 3

More comprehensive analyses can be performed, involving the study of the thermophysical properties and Fickian diffusion dynamics across different rejuvenators – expanding **Example 2**'s focus on “u3v” to “u3x”. For instance, contrasting constraints may be required to be met— where rejuvenated bitumens must report a density above 1000 kg/m³ while needing to be “rejuvenated” in 4 h, at 1 mm, and at 60 °C, with a minimum rejuvenator concentration between 1 and 2 % and by weight. These conditions put **VO** at a disadvantage, as the density of the bitumen-rejuvenator in such case is lower than 1000 kg/m³. When performing the computations for all four rejuvenators, where **VO**, **EO**, **NO**, and **AO** have bulk densities of 849.7, 805.1, 868.7, and 968.4 kg/m³, and diffusion coefficients of 0.272, 0.206, 0.147, and 0.101e-9 m²/s, respectively, **AO** is the only candidate that can match the requirements, even though it has been reported to be the “worst” performing rejuvenator of all, as its concentration by weight at 1 mm., 60 °C, and after 4 h is just about 1.2 %, and the mixture density is at 1005 kg/m³. The next candidate would be **NO**, but at such conditions, even though the density still hovers around 1000 kg/m³, the concentration by weight is far above that required, at ~8 %. Such findings further corroborate that careful analysis must be performed to select rejuvenators, where higher diffusion coefficients do not necessarily translate into better performance. The use of MLMs, though, to evaluate them certainly allow this to be feasible in very short times. Also, such example elucidates how steady-state, bulk properties of individual materials predicted using other MLMs can be of use in time-dependent situations, such as those involving Fick's diffusion dynamics.

6. Conclusion

This manuscript presents an in-depth investigation into the diffusion behaviors of rejuvenator-bitumen systems using NEMD simulations, considering factors such as rejuvenator type, temperature, bitumen type, aging degree, and chemical descriptors. Three bitumen types, **UB**, **LB**, and **JB**, were chosen based on sulfur composition, conditioned at five aging degrees, and paired with four rejuvenators: **VO**, **EO**, **NO**, and **AO**, which are commonly found in commercially available rejuvenation blends. A total of 240 non-equilibrium MD simulations were performed across four temperature regimes (60 °C, 120 °C, 160 °C, and 200 °C), providing a comprehensive view of the Fickian diffusion coefficients for various bitumen-rejuvenator combinations.

The results show that diffusion coefficients range from as low as 0.0068e-10 m²/s in highly aged bitumens (degree 4) rejuvenated with **AO** and **NO** at the lowest temperature (60 °C) to as high as 4.35e-10 m²/s in fresher bitumens rejuvenated with **VO** and **EO** at 200 °C, with values in line with experimental observations. Diffusion is primarily influenced by temperature, followed by rejuvenator type (**VO** > **EO** > **NO** > **AO**), bitumen aging degree (higher aging results in reduced diffusion, especially beyond aging degree 2), and bitumen type, where sulfur-rich and lower molecular weight bitumens (**UB** > **LB** > **JB**) exhibit the highest diffusion rates. Interestingly, slightly aged bitumens show favorable diffusion characteristics in certain conditions, while heavily aged samples exhibit diminished sensitivity to temperature changes, likely due to structural changes that impede molecular mobility.

The MLM employed in this study achieved an R² of 0.97, demonstrating its efficacy in predicting diffusion coefficients with high accuracy. By utilizing 19 featural descriptors (nine each for the bitumen and rejuvenator sides and one for temperature), the MLM captures trends and correlations much more efficiently than conventional methods like SARA fractionation or Elemental Analysis. The most influential features affecting diffusion are temperature (42 %), bitumen molecular weight (-20 %), rejuvenator molecular weight (-9 %), and more fundamental chemical features such as rejuvenator oxygen content, double bonds, and bitumen sulfur content, all contributing positively to diffusion potential. Meanwhile, descriptors like bitumen oxygen (indicative of

aging), nitrogen, and aliphatic content negatively impact diffusion.

This detailed characterization provides a useful framework for tuning the most influential features to optimize diffusion performance. For example, selecting rejuvenators with slightly oxidized functional groups or bitumens with high sulfur content and aromatic groups enhances diffusion, whereas higher molecular weights and increased aging reduce it. The MLM predictions offer near-instantaneous predictions, making them valuable for the iterative study, design, and evaluation of rejuvenators, allowing adjustments to molecular structures based on the key chemical descriptors.

While higher diffusion rates—such as those observed with **VO**—can expedite the rejuvenation process, maximizing diffusion alone does not necessarily lead to optimal material properties. For example, **VO**'s high diffusion rate may cause significant softening of the bitumen, potentially compromising rheological stability, particularly in unaged sections or at greater depths. Therefore, diffusion performance must be carefully balanced with other key thermophysical properties such as density, heat capacity, and cohesiveness to meet industrial requirements. This study further highlights the value of integrating the Fickian diffusion coefficients obtained from MD simulations with previously published MLMs for bulk property predictions. This comprehensive approach facilitates the design and characterization of rejuvenators, ensuring that formulations optimize both diffusion and broader material properties across different temperatures, concentrations, depths, and time scales. By addressing these dual considerations, this methodology provides a powerful tool for tailoring rejuvenator formulations to practical industrial needs.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used OpenAI's ChatGPT4o (October 2024) 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: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Xueyan Liu**: Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Peng Lin**: Writing – review & editing, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Shisong Ren**: Writing – review & editing, Validation, Supervision, Methodology, Data curation, Conceptualization. **Sandra Erkens**: Writing – review & editing, Supervision, Software, Resources, Project administration, Investigation, 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.

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.2024.113502>.

Data availability

Most of the data required to reproduce the results of this study are already available either in the manuscript or in the SI. Once accepted, we aim to provide you with the actual LAMMPS inputs files.

References

- J.C. Nicholls, D. James, Literature review of lower temperature asphalt systems, *Proc. Inst. Civ. Eng.-Constr. Mater.* 166 (5) (2013) 276–285.
- 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.
- 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.
- M.L. Greenfield, L. Zhang, Developing model asphalt systems using molecular simulation: final model, University of Rhode Island. Transportation Center, 2009.
- A.M. Kharat, J. Zacharia, V.J. Cherian, A. Anyatonwu, Issues with comparing SARA methodologies, *Energy Fuel* 21 (6) (2007) 3618–3621.
- 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.
- E.I. Assaf, X. Liu, P. Lin, S. Erkens, L.I. Mensink, Studying the impact of phase behavior in the morphology of molecular dynamics models of bitumen, *Mater. Des.* 230 (2023) 111943.
- M.L. Greenfield, Molecular modelling and simulation of asphaltenes and bituminous materials, *Int. J. Pavement Eng.* 12 (4) (2011) 325–341.
- E. Prosperi, E. Bocci, A review on bitumen aging and rejuvenation chemistry: processes, materials and analyses, *Sustainability* 13 (12) (2021) 6523.
- Z. Yang, X. Zhang, Z. Zhang, B. Zou, Z. Zhu, G. Lu, W. Xu, J. Yu, H. Yu, Effect of aging on chemical and rheological properties of bitumen, *Polymers* 10 (12) (2018) 1345.
- V. Loise, P. Caputo, M. Porto, P. Calandra, R. Angelico, C. Oliviero Rossi, A review on bitumen rejuvenation: mechanisms, materials, methods and perspectives, *Appl. Sci.* 9 (20) (2019) 4316.
- D. Kuang, Z. Feng, J. Yu, X. Chen, B. Zhou, A new approach for evaluating rejuvenator diffusing into aged bitumen, *J. Wuhan Univ. Technol.-Mater. Sci. Ed.* 26 (1) (2011) 43–46.
- G. Xu, H. Wang, W. Sun, Molecular dynamics study of rejuvenator effect on RAP binder: diffusion behavior and molecular structure, *Constr. Build. Mater.* 158 (2018) 1046–1054.
- M. Xu, J. Yi, D. Feng, Y. Huang, Diffusion characteristics of asphalt rejuvenators based on molecular dynamics simulation, *Int. J. Pavement Eng.* 20 (5) (2019) 615–627.
- H. Ding, H. Wang, X. Qu, A. Varveri, J. Gao, Z. You, Towards an understanding of diffusion mechanism of bio-rejuvenators in aged asphalt binder through molecular dynamics simulation, *J. Clean. Prod.* 299 (2021) 126927.
- A.V. Chechkin, I.M. Zaid, M.A. Lomholt, I.M. Sokolov, R. Metzler, Bulk-mediated diffusion on a planar surface: full solution, *Phys. Rev. E—Statist. Nonlinear Soft Matter Phys.* 86 (4) (2012) 041101.
- D.D. Li, M.L. Greenfield, Viscosity, relaxation time, and dynamics within a model asphalt of larger molecules, *J. Chem. Phys.* 140 (3) (2014).
- 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.
- K. Ohno, K. Esfarjani, Y. Kawazoe, *Computational Materials Science: From Ab Initio to Monte Carlo Methods*, Springer, 2018.
- M. Tsige, G.S. Grest, Molecular dynamics simulation of solvent–polymer interdiffusion: Fickian diffusion, *J. Chem. Phys.* 120 (6) (2004) 2989–2995.
- D.J. Keffer, C.Y. Gao, B.J. Edwards, On the relationship between Fickian diffusivities at the continuum and molecular levels, *J. Phys. Chem. B* 109 (11) (2005) 5279–5288.
- 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.
- Y. Gao, Y. Zhang, C. Zhang, X. Liu, R. Jing, Quantifying oxygen diffusion in bitumen films using molecular dynamics simulations, *Constr. Build. Mater.* 331 (2022) 127325.
- A.Y. Kuksin, I. Morozov, G. Norman, V. Stegailov, I. Valuev, Standards for molecular dynamics modelling and simulation of relaxation, *Mol. Simul.* 31 (14–15) (2005) 1005–1017.
- F. Müller-Plathe, A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity, *J. Chem. Phys.* 106 (14) (1997) 6082–6085.
- G. Rehage, O. Ernst, J. Fuhrmann, Fickian and non-Fickian diffusion in high polymer systems, *Discuss. Faraday Soc.* 49 (1970) 208–221.
- 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).
- S. Ren, X. Liu, P. Lin, Y. Gao, S. Erkens, Review on the diffusive and interfacial performance of bituminous materials: from a perspective of molecular dynamics simulation, *J. Mol. Liq.* 366 (2022) 120363.
- E.I. Assaf, X. Liu, P. Lin, S. Ren, S. Erkens, Predicting the properties of bitumen using machine learning models trained with force field atom types and molecular dynamics simulations, *Mater. Des.* (2024) 113327.
- 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.
- S. Vedachalam, N. Baquerizo, A.K. Dalai, Review on impacts of low sulfur regulations on marine fuels and compliance options, *Fuel* 310 (2022) 122243.
- F. Migliori, J.-F. Corté, Comparative study of RTFOT and PAV aging simulation laboratory tests, *Transp. Res. Rec.* 1638 (1) (1998) 56–63.
- 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.
- C. EN, 14769: Bitumen and Bituminous Binders—Accelerated Long-Term Ageing Conditioning by a Pressure Ageing Vessel (PAV), European Committee for Standardization: Brussels, Belgium (2012).
- 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.
- 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.
- S.A. Khan, S. Sarfraz, D. Price, TLC-FID calibration and accurate weight determination of SARA fractions in heavy crude oil, *Pet. Sci. Technol.* 30 (23) (2012) 2401–2406.
- H.P. Sieper, H.J. Kupka, T. Williams, A. Rossmann, S. Rummel, N. Tanz, H. L. Schmidt, A measuring system for the fast simultaneous isotope ratio and elemental analysis of carbon, hydrogen, nitrogen and sulfur in food commodities and other biological material, *Rapid Commun. Mass Spectrometry* 20 (17) (2006) 2521–2527.
- H. Dettman, A. Inman, S. Salmon, K. Scott, B. Fuhr, Chemical characterization of GPC fractions of Athabasca bitumen asphaltenes isolated before and after thermal treatment, *Energy Fuel* 19 (4) (2005) 1399–1404.
- D.D. Li, M.L. Greenfield, Chemical compositions of improved model asphalt systems for molecular simulations, *Fuel* 115 (2014) 347–356.
- 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.
- 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.
- 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.
- G. Landrum, Rdkit documentation, Release 1 (1–79) (2013) 4.
- 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.
- 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, *Comput. Phys. Commun.* 271 (2022) 108171.
- J.R. Maple, M.J. Hwang, T.P. Stockfisch, U. Dinur, M. Waldman, C.S. Ewig, A. T. Hagler, Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules, *J. Comput. Chem.* 15 (2) (1994) 162–182.
- 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, *Rev. Mod. Phys.* 64 (4) (1992) 1045.
- K.F. Mansfield, D.N. Theodorou, Molecular dynamics simulation of a glassy polymer surface, *Macromolecules* 24 (23) (1991) 6283–6294.
- P. Van der Ploeg, H. Berendsen, Molecular dynamics simulation of a bilayer membrane, *J. Chem. Phys.* 76 (6) (1982) 3271–3276.
- D.J. Evans, B.L. Holian, The nose–hoover thermostat, *J. Chem. Phys.* 83 (8) (1985) 4069–4074.
- 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* (2024) 100655.
- 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* (2024) 100656.
- W. Sun, H. Wang, Molecular dynamics simulation of diffusion coefficients between different types of rejuvenator and aged asphalt binder, *Int. J. Pavement Eng.* 21 (8) (2020) 966–976.
- W. McKinney, P. Team, *Pandas—Powerful python data analysis toolkit, Pandas—Powerful Python Data Analysis Toolkit* 1625 (2015).
- 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, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.

- [57] M.R. Segal. Machine learning benchmarks and random forest regression, (2004).
- [58] R.F. Sproull, Refinements to nearest-neighbor searching in k-dimensional trees, *Algorithmica* 6 (1991) 579–589.
- [59] 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.
- [60] 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, *J. Pet. Sci. Eng.* 102 (2013) 36–40.
- [61] 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.
- [62] 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.
- [63] 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.
- [64] N. Durupt, A. Aoulmi, M. Bouroukba, M. Rogalski, Heat capacities of liquid polycyclic aromatic hydrocarbons, *Thermochim Acta* 260 (1995) 87–94.
- [65] 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.
- [66] 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.
- [67] R. Cassis, N. Fuller, L.G. Hepler, R.J. McLean, A. Skauge. Specific Heat Capacities of Bitumens in Heavy Oils, Reservoir Minerals, Clays, Dehydrated Clays, Asphaltenes, and Cokes.
- [68] H.W. Bearce, E.L. Peffer, Density and thermal expansion of American petroleum oils, US Government Printing Office 1916.
- [69] S. Ren, X. Liu, S. Erkens, Unraveling the critical indicators for evaluating the high-temperature performance of rejuvenator-aged bitumen blends, *Case Stud. Constr. Mater.* 19 (2023) e02522.