Energy Proceedings
Vol 39, 2024
ISSN 2004-2965

Liquid synthetic fuels design guided by chemical structure: A machine learning perspective

Rodolfo S. M. Freitas[1*], Cheng Chen[1], Xi Jiang[1*]

1 School of Engineering and Materials Science, Queen Mary University of London, Mile End Road, London E1 4NS, UK

(*Corresponding Author: rodolfo.dasilvamachadodefreitas@qmul.ac.uk, xi.jiang@qmul.ac.uk)

ABSTRACT

Physicochemical properties of synthetic fuels are important but difficult to measure/predict, especially when complex surrogate fuels are concerned. In the present work, machine learning (ML) models are constructed to discover intrinsic chemical structure-properties relationships. The models are trained using data from molecular dynamics (MD) simulations. The fuel structure is represented by molecular descriptors. Such a symbolic representation of the fuel molecule allows to link important features of the fuel composition with key properties of fuel utilization. The results show that the present approach can predict accurately the fuel properties of a wide range of pressure and temperature conditions.

Keywords: fuel properties, molecular dynamics simulations, molecular descriptor, machine learning models

1. INTRODUCTION

Fossil fuels still play a key role in energy supply, especially in difficult-to-decarbonize transport applications such as shipping, road freight, and aviation transport. Overall, they are responsible to emit more than 50% CO2 of the entire transport sector [1]. With the need to take a step towards net zero emissions and sustainable energy utilization, renewable fuels including biofuels are becoming increasingly important. Among efforts on developing low-emission fuels, liquid synthetic fuels like Oxymethylene Dimethyl Ethers (OMEx) have shown high potential for low-carbon transport applications, since they can be integrated into the current transportation system using existing infrastructure and be burned in existing engines (such as diesel engines for optimal fuel economy) with minor adjustments as drop-in fuels [2,3].

For the rapid integration of synthetic fuels into current infrastructures for storage, transport, and direct injection in combustion engines the physicochemical properties associated with fuel composition must be known. This represents a significant challenge since synthetic fuels are often composed of complex mixtures and the physicochemical properties depend on fuel composition variability linked with production source and process. To address this challenge, accurate information on the physicochemical properties of complex mixtures over the engine operational ranges is mandatory to adapt the system operation to alternative fuels, but this is not readily available.

In order to pursue this goal, MD simulations have been used to predict the physicochemical properties of practical fuels including transport properties at supercritical conditions [4]. However, MD simulations are generally expensive in terms of computational costs (CPU time and memory). Hence, although those simulations provide molecular details that can be potentially used to accurately predict fuel properties, it is not feasible to establish complete and detailed fuel property databases using MD simulations.

Recently, machine-learning models have gained attention to predict physicochemical properties from molecular structures [5]. Also, ML can be a powerful tool to predict the physicochemical properties of fuels from the chemical structures [6,7,8]. Freitas et al [9] proposed a methodology to explore the thermodynamic properties of practical fuels by combining MD simulations and ML models. The results show that ML models can yield accurate predictions of fundamental fuel properties from the chemical compositions of the fuels by using databases from MD simulations.

The present work aims to characterize the physicochemical properties of synthetic fuels dependence on thermodynamic state variables

This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

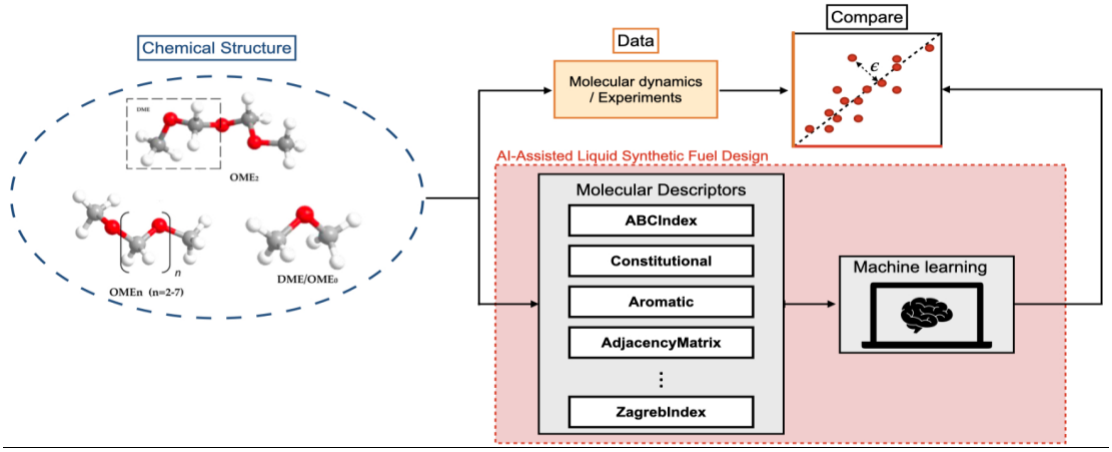

Fig. 1. Overview of AI-assisted liquid synthetic fuel design methodology

(temperature and pressure) and fuel chemical composition/structure using machine learning techniques to leverage data obtained through MD computations and/or experiments. Here, the chemical structure of such complex fuels may be described using chemical descriptors [7,10]. Descriptors allow us to correlate the chemical structure of the fuels with thermophysical properties. Such an approach allows an understanding of the underlying physics which links the chemical structure with the physical properties, as well as helps to design novel blends.

2. METHODOLOGY

In this section, we briefly introduced the AI-assisted liquid synthetic fuel design methodology that can potentially lead to fully sustainable combustion, i.e., the engines are 100% powered by sustainable/renewable synthetic fuels. Figure 1 shows an overview of the present methodology, whose further details are given ahead.

2.1 Molecular Descriptors

A molecular descriptor is a mathematical characterization of a chemical structure. The main idea is transforming chemical information embedded within a symbolic representation of a molecule to a set of features useful to represent the chemical composition [10]. Such a description may be used to connect important features of the fuel composition with key properties of fuel utilization, so allowing a step toward the development of interpretable machine-learning models. Going further, revealing the dependence of physicochemical properties of liquid synthetic fuels on fuel mixture chemical composition/chemical structure may lead to new information about the property, and provide a better understanding of the underlying physics of the relation between physicochemical properties and molecular structure.

2.2 Building machine learning models

In this section, we present a brief description of ML models for a generic property γ function of the chemical structure/composition and state variables, pressure (P) and temperature (T). In particular, the aim is to learn a mapping f characterizing the macroscopic thermodynamic relation between the physicochemical property and the chemical structure:

$$\gamma = f(\mathbf{\Phi}, P, T, \boldsymbol{\xi}). \quad (1)$$

Here, f is a nonlinear map that acts as a surrogate model for the costly MD simulation. $\mathbf{\Phi}$ is the vector of molecular descriptors that characterize the chemical structure of fuel. The vector $\boldsymbol{\xi}$ denotes potential noise and is often considered a random.

In the present study, we use a simple fully connected neural network (FCNN) to discover the relationship. between chemical structure and properties. Neural networks (NNs) are universal function approximators that can detect and decode intrinsic relations from data. NN models have been shown to be an effective tool for accurately predicting several physicochemical properties [11,12]. Also, such models are simple to implement with several end-to-end open-source machine learning platforms available.

A Theory section should extend, not repeat, the background to the article already dealt with in the Introduction and lay the foundation for further work. In contrast, a 'Calculation' section represents a practical development from a theoretical basis.

3. RESULTS AND DISCUSSION

In the present section, we demonstrate the performance of the proposed methodology. Here, we

consider single-component alkanes C_nH_{2n+2}, so reliable data for model assessment and validation can be used. The compounds considered are n-octane, n-nonane, n-decane, n-dodecane, and n-hexadecane. The dataset used to build the ML models consists of 1200 density values. In particular, 240 values of the density for each compound are considered, computed at a regular temperature grid within $T \in [320, 900]$ K, varying by 20 K, and at the specific pressure values: P = {3, 4, 6, 8, 10, 20, 100, 150} MPa.

Here, the molecular descriptors are computed using Mordred software [10]. Mordred is an open-source library that can generate more than 1800 descriptors. Without loss of generality, we use just the number of carbon and molecular weight as input descriptors for the ML model. However, a feature selection technique can be used to identify important features with meaningful property relationships in the data [7]. Also, a simple one-hidden layer FCNN with 16 neurons is constructed for mapping the fuel composition to the property. The model is trained for 300 epochs using the Adam optimizer with a learning rate of 10^{-4}. The proposed model was implemented in PyTorch [13], and computations were performed in single precision arithmetic on a single NVIDIA GeForce RTX 2060 GPU card.

In the training process, 80% of the data points are selected randomly to train the ML models. The remaining 20% is used to test. Figure 2 shows the parity plots between the predicted densities by the ML model and computed by MD simulations for the test dataset. As we can see the ML model returns excellent predictions with a coefficient of determination (R2-score) very near 1.0.

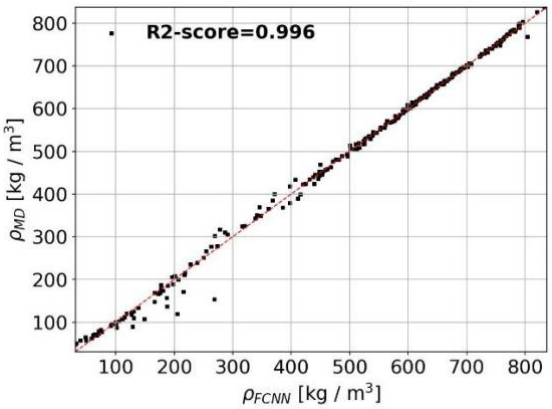

Fig. 2. Parity plots showing test set and predicted values of density

As a further illustration of the performance of such approaches to predict the density, we validate how the proposed ML technology performs in an extrapolation scenario. We validate them for the n-heptane, a fuel not used for building the models. To pursue this goal, instead of employing data provided by MD computations, we use an experimental database furnished by the National Institute of Standards and Technology (NIST). Figure 3 shows that the ML model can predict satisfactorily well the density at different pressures. However, at the lowest pressure at supercritical conditions (Tc = 540.13 K), where abrupt decay of the density occurs, the ML model returns density predictions far from satisfactory. This might be partially solved by adding more data at the supercritical region during the training process.

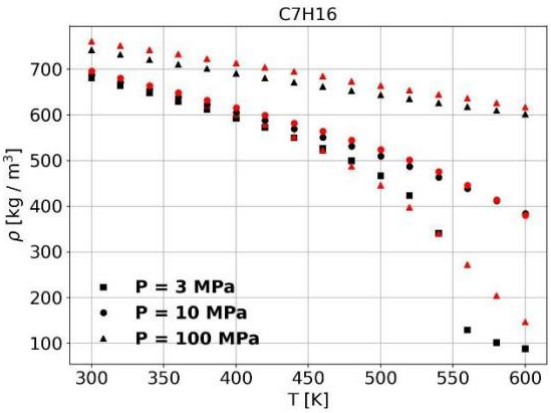

Fig. 3. n-Heptane predictions with the machine learning model at the pressures 3, 10, and 100 MPa. MD simulations (**black points**) and ML model (**red points**).

4. CONCLUSIONS

In this work, we propose a computational methodology based on the use of ML with Molecular Dynamics simulations to the mapping between the fuel composition and key properties of fuel utilization. The ML model has been demonstrated to be a powerful tool to reveal the dependence of physicochemical properties of liquid synthetic fuels on fuel mixture chemical composition / chemical structure. Furthermore, such a methodology allows the design of novel liquid synthetic fuel blends that can potentially lead to fully sustainable combustion.

The present work shows a successful prediction of fuel density guided by the chemical structure, that can also be extended to other physicochemical properties as well as more complex fuel molecules or multicomponent mixtures like dimethyl ethers or OMEx. The generation of reliable physicochemical properties of renewable fuels is an important step forward towards the generation of digital tools that can assist on the decarbonization using renewable fuels.

ACKNOWLEDGEMENT

The research leading to these results had received funding from UK EPSRC Engineering and Physical Sciences Research Council Grant No. EP/X019551/1.

DECLARATION OF INTEREST STATEMENT

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. All authors read and approved the final manuscript.

REFERENCE

[1] IEA, Global CO2 emissions from transport by subsector, 2000-2030, IEA, Paris. https://www.iea.org/data-and-statistics/charts/global-co2-emissions-from-transport-by-subsector-2000-2030, IEA. Licence: CC BY 4.0.

[2] Omari A, Heuser B, Pischinger S. Potential of oxymethylenether-diesel blends for ultra-low emission engines. Fuel 2017;209: 232–7. http://dx.doi.org/10.1016/j.fuel.2017.07.107.

[3] Pélerin D, Gaukel K, Härtl M, Jacob E, Wachtmeister G. Potentials to simplify the engine system using the alternative diesel fuels oxymethylene ether OME1 and OME3-6 on a heavy-duty engine. Fuel 2020;259:116231.http://dx.doi.org/10.1016/j.fuel.2019.116231.

[4] Chen C, Jiang X. Transport property prediction and inhomogeneity analysis of supercritical n-Dodecane by molecular dynamics simulation, Fuel, 2019, 244: 48-60, https://doi.org/10.1016/j.fuel.2019.01.181.

[5] Zhan H, Zhu X, Qiao Z, Hu J. Graph Neural Tree: A novel and interpretable deep learning-based framework for accurate molecular property predictions, Analytica Chimica Acta, 2023, 1244: 340558, https://doi.org/10.1016/j.aca.2022.340558.

[6] Li R, Herreros JM, Tsolakis A, Yang W. Machine learning-quantitative structure property relationship (ML-QSPR) method for fuel physicochemical properties prediction of multiple fuel types, Fuel, 2021, 304:121437, https://doi.org/10.1016/j.fuel.2021.121437.

[7] Comesana AE, Huntington TT, Scown CD, Niemeyer KE, Rapp ViH. A systematic method for selecting molecular descriptors as features when training models for predicting physiochemical properties, Fuel, 2022, 321: 123836, https://doi.org/10.1016/j.fuel.2022.123836.

[8] vom Lehn F, Brosius B, Broda R, Cai L, Pitsch H. Using machine learning with target-specific feature sets for structure-property relationship modeling of octane numbers and octane sensitivity, Fuel, 2020, 281: 118772, https://doi.org/10.1016/j.fuel.2020.118772.

[9] Freitas RSM, Lima ÁPF, Chen C, Rochinha FA, Mira D, Jiang X. Towards predicting liquid fuel physicochemical properties using molecular dynamics guided machine learning models, Fuel, 2022, 329: 125415, https://doi.org/10.1016/j.fuel.2022.125415.

[10] Moriwaki H, Tian Y-S, Kawashita N, Takagi T. Mordred: a molecular descriptor calculator. J Cheminform 10, 4 (2018). https://doi.org/10.1186/s13321-018-0258-y.

[11] Santak P, Conduit G. Predicting physical properties of alkanes with neural networks, Fluid Phase Equilibria, 2019, 501: 112259, https://doi.org/10.1016/j.fluid.2019.112259.

[12] Kessler T, John PCSt, Zhu J, McEnally CS, Pfefferle LD, Mack JH. A comparison of computational models for predicting yield sooting index, Proceedings of the Combustion Institute, 2021, 38: 1385-1393, https://doi.org/10.1016/j.proci.2020.07.009.

[13] Paszke A, Gros S, Massa F, Lerer A, Bradbury J, Chanan G, … Chintala S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Advances in Neural Information Processing Systems 2019, 32 (pp. 8024–8035). Curran Associates, Inc. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf