Demonstration of a Machine Learning-Based Approach to Predict Thermophysical Properties of Species Relevant to Aviation Fuels

Bhattacharya, Chandrachur ; Poblador Ibanez, J.; Han, Austin ; Dasgupta, Debolina; Nocivelli, Lorenzo

DOI
10.2514/6.2025-0743

Publication date
2025

Document Version
Final published version

Published in
Proceedings of the AIAA SCITECH 2025 Forum

AIAA SciTech Forum
6-10 January 2025, Orlando, FL
AIAA SCITECH 2025 Forum

10.2514/6.2025-0743

Demonstration of a Machine Learning-based Approach to Predict Thermophysical Properties of Species Relevant to Aviation Fuels

Chandrachur Bhattacharya *
Transportation and Power Systems, Argonne National Laboratory, Lemont, IL, 60439

Jordi Poblador-Ibanez †
Department of Maritime and Transport Technology, Delft University of Technology, Delft, 2628 CD, The Netherlands

Austin Han‡, Debolina Dasgupta §, and Lorenzo Nocivelli ¶
Transportation and Power Systems, Argonne National Laboratory, Lemont, IL, 60439

Sustainable Aviation Fuels (SAF) are being considered to replace current fuels, such as Jet A, to support the effort of industry and regulatory agencies to target the decarbonization of the aviation sector by 2050. Strict regulations on fuel properties, both in terms of applicability in current engines and in emission improvements (i.e., particulate matter control towards the reduction of contrails), require extensive analysis on the fuel thermophysical and chemical characteristics. The current lack of experimental data at engine-relevant pressure and temperatures for SAF candidates, motivates the exploration of accurate and robust models to capture the behavior of hydrocarbon mixtures at engine relevant conditions to support the development and deployment of net-zero carbon propulsion.

This work showcases a data-driven approach based on a novel encoder-Gaussian process, which is designed to guarantee smoothness, comes with uncertainty quantification, and can incorporate physics-guided understanding as required. These capabilities are utilized for the modeling of thermophysical properties of pure species, including transcritical regimes while reducing the need for access to the critical properties. This effort arises from the shortcomings of both input properties availability and overall performance of previously investigated cubic equations of state. This paper introduces MeGS-RFM, a machine-learning based real-fluid modeling approach, and compares its performance with available databases and a volume-translated Soave-Redlich-Kwong equation of state. MeGS-RFM uses a generative modeling approach to generalize across species not available in the training datasets. Finally, we use this to demonstrate improved characterization of iso-paraffins relevant to aviation fuels, showing better agreement with the sparse datasets in open literature.

I. Nomenclature

p	=	Pressure, bar
T	=	Temperature, K
ω	=	Acentric factor
μ	=	Dynamic Viscosity, mPa-s
ρ	=	Density, kg/m^3
u	=	Internal Energy, kJ/kg
C_p	=	Isobaric Specific Heat, kJ/kg-K
k	=	Thermal Conductivity, W/m-K
c	=	Critical Properties (subscript)

*Postdoctoral Appointee, Argonne National Laboratory, cbhattacharya@anl.gov, Young Professional Member AIAA.
†Postdoctoral Researcher, TU Delft, Young Professional Member AIAA.
‡Postdoctoral Appointee, Argonne National Laboratory, ahan@anl.gov, Young Professional Member AIAA.
§Senior Research Scientist, Argonne National Laboratory, ddasgupta@anl.gov, Senior Member AIAA.
¶Senior Research Scientist, Argonne National Laboratory, lnocivelli@anl.gov

II. Introduction

As a consequence of the growing electrification of light-duty vehicles, aviation is expected to become one of the major sources of greenhouse gas emissions for the transportation sector in the U.S.*. Coupled with the projected passenger growth†, this will lead to a three-fold increase in the greenhouse gas (GHG) emissions from the aviation sector by 2050, if no action is taken. In this scenario, improvements in airframe and propulsion efficiency need to be complemented by the introduction of alternative fuels to achieve net-zero GHG emissions [1]. For safety and technological constraints, Sustainable Aviation Fuels (SAF) represent the most viable option to reduce the carbon footprint of long-haul airplanes, for which they are expected to replace fossil-fuel based kerosene, relying on compatible propulsion systems. SAF are bio-fuels that offer net-zero carbon emissions during their life-cycle‡ [2], which includes fuel processing, distribution and consumption during flight. Some SAF, labeled as drop-in SAF, are designed to directly replace Jet A in current jet engines, but report deviations in their behavior, especially at extreme operating conditions [3]. This limits their current deployment as blends with conventional kerosene, i.e., Jet A and Jet A-1 up to 50%, as specified by the ASTM D4054 standard [4]. This scenario determines the need of extensive research efforts, including the SAF Grand Challenge§ program for fuel production scale-up and SAF end-use program for a possible 100% replacement of Jet A in engines.

SAF are synthesized bio-fuels with a simpler composition spectrum than fossil based jet fuels [3], dependent on the specific processing technique. Some of these processes include alcohol-to-jet (ATJ), Fischer-Tropsch (FT), hydroprocessed esters and fatty acids (HEFA) or power-to-liquid (PtL) processes. From this point of view, SAF ultimately differ from conventional jet fuels, whose composition includes complex mixtures of hydrocarbons including n-paraffins, iso-paraffins, aromatics and cycloparaffins [3], and offer a fertile ground to develop detailed models of their thermophysical properties. In fact, the characterization of Jet A often relies on simplified surrogates and chemical mechanisms developed to fit specific experimental data [5–8], which are affordable to implement in Computational Fluid Dynamics (CFD) workflows [9]. Instead, SAF present the possibility of characterizing their properties with accurate and manageable real-fluid models that can be directly integrated with CFD software aimed at modeling jet engine combustors.

In the present work, we propose a data-driven approach to model non-ideal fluid behavior of hydrocarbons and other species relevant for the fuel injection and combustion environment through a machine learning framework. This machine learning model uses generative methods and aims to predict the thermophysical properties of SAF-relevant species at realistic jet engine operating conditions, from very low pressures and temperatures typical of high-altitude relight, to temperature and pressures that determine near-critical or super-critical behavior, as reported at cruise, climb, and take-off [3, 10]. The goal of the model is to overcome two major shortcomings of standard real-fluid frameworks built upon cubic equation on state (EoS) and generalized correlations to predict thermophysical and transport properties in a multiphase environment: (1) lack of accuracy in specific conditions (e.g., density or viscosity of liquid hydrocarbons at ambient conditions), and (2) large uncertainties in modeling poorly characterized species (e.g., highly branched iso-paraffins) that are predominant in SAF. In fact, despite their extensive use in modeling liquid fuel injection at elevated temperatures and pressures [11–15], our recent findings [16, 17], highlight the limitations of classic cubic EoS in modeling realistic biofuels for aerospace propulsion. Some recent efforts have explored the use of neural network methods for modeling supercritical carbon-dioxide [18], pressure–temperature (PT) flash calculations [19], and even solving the transcritical mixing layer in liquid-oxygen and gaseous-hydrogen streams [20]. However, these models are purpose built and do not aim at creating a generalized data-driven real-fluid-modeling framework that can extend beyond the training data.

This paper first reports a brief description of the cubic EoS approach we previously implemented to model SAF-relevant species [16]. Next, we present details of the data-driven generative approach and frameworks, and analyze their performance on well-characterized species. This model is trained with the extensive REFPROP pure fluids database, comprising 157 species, of which 45 are hydrocarbons. Lastly, we display the performance of the frameworks on two poorly characterized iso-paraffins (iso-dodecane and iso-hexadecane) that are instrumental in the composition of SAF [9], and compare the results to the cubic EoS baseline. We then conclude the paper with current on-going and future planned work.

*https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions (EPA-420-F-22-018, May 2022)

†Data from the Federal Aviation Administration https://www.faa.gov/dataresearch/aviation

‡https://www.energy.gov/eere/bioenergy/sustainable-aviation-fuels

§https://www.energy.gov/eere/bioenergy/sustainable-aviation-fuel-grand-challenge

III. Methodology

A. Cubic Equation of State and Generalized Correlations to Model SAF

A recent work by the authors [16] screened the most commonly used cubic equations of state, i.e., the Soave-Redlich-Kwong (SRK) EoS [21] and the Peng Robinson (PR) EoS [22], focusing on their performance in predicting characteristic normal-, iso- and cyclo-paraffins found in SAF. These equations are part of the Van der Waals-type equations, and they have been equipped with volume translation methods [23–25] to overcome the limitations associated with two-parameters cubic EoS in predicting density of compressed fluids [26, 27]. The generalized formulation of a cubic equation of state is reported in Eq. (1), where R_u is the universal gas constant, δ_1, δ_2 are constants specific to the selected EoS, a, b are functions of critical properties and acentric factor specific of the species in the mixture, and c represents the volume translation coefficient calculated according to approaches available in literature [23–25]. Classic linear and quadratic mixing rules are implemented.

$$p = \frac{R_u T}{\overline{v} + c - b} - \frac{a}{(\overline{v} + c + \delta_1 b)(\overline{v} + c + \delta_2 b)} \tag{1}$$

Following the approach by Poling at al. [28], the determination of the properties through the EoS can be represented by means of a departure function from the ideal state. Eq. (2) acts as an example of the approach for specific enthalpy h, where h^* is the ideal-gas specific enthalpy, MW is the molecular weight and Z the compressibility factor. Similarly, we can derive the expressions for internal energy, entropy and speed of sound [28].

$$h = h^* + \frac{R_u T}{MW}\left[(Z-1) + \int_{\infty}^{\overline{v}} T\left(\frac{\partial Z}{\partial T}\right)_{\overline{v}} \frac{d\overline{v}}{\overline{v}}\right] \tag{2}$$

Alongside the aforementioned properties, additional models can be built on top of the EoS to predict transport properties of the species, such as dynamic viscosity μ and thermal conductivity k. These models use local pressure and temperature, and rely on the density values obtained through the EoS, to define correlations that can be used for both liquid and gas fluid states. Viscosity and thermal conductivity are calculated according to the generalized multi-parameter correlation by Chung et al. [29]. The scope of this work is limited to the modeling of single species in a specified condition, so any description of mixing strategies will be neglected, as well as any treatment of the vapor-liquid equilibrium and resulting surface tension evaluation. More detailed implementation of the comprehensive modeling framework can be found in [16].

As mentioned in the introduction, the necessary inputs for cubic EoS, i.e., critical pressure and temperature, and acentric factor, are often not available for hydrocarbons that are relevant for SAF. In fact, if we focus on category C alternative fuels characterized within the National Jet Fuel Combustion Program (NJFCP) [9] to be representative of SAF, one notices that scarcely characterized iso-paraffins are a relevant fraction of the fuel composition. A source of uncertainty in the modeling of SAF is the selection of a single iso-paraffin for each carbon number, among a multitude of potential isomers. On top of this uncertainty, often the necessary input data for the EoS has not been characterized for commercially available iso-paraffins and an estimate of the properties of the species needs to be generated through group contribution methods by assuming its molecular structure. Critical pressure p_c, temperature T_c and specific volume v_c, are obtained from the work by Nannoolal et al. [30]. T_c is a function of normal boiling point T_b, which is available in NIST[¶] databases or can be determined with methods available in the literature [31]. Acentric factor is determined according to the approach by Tahami et al. [32] and the dipole moment for viscosity and thermal conductivity correlations is defined as specified by Müller et al. [33].

After the extensive screening performed by the authors [16], the best performance for SAF-relevant species, including air species, is obtained by the Soave-Redlich-Kwong EoS coupled with the volume translation method by Lin et al. [24] (VT-SRK). In the current work, we will consider this formulation as the reference model to assess the performance of the data-driven approach on iso-paraffins.

B. Data-driven Surrogate Framework: MeGS-RFM

The machine learning-based real-fluid modeling (RFM) surrogate is based on an encoder-Gaussian process (GP) approach which utilizes the Multi-capability encoder Gaussian process Surrogate (MeGS) framework, developed in-house at Argonne National Laboratory. MeGS is aimed at general purpose surrogate modeling, with an added

[¶]National Institute of Standards and Technology, U.S. Department of Commerce.

focus on sparse datasets as typically available in engineering problems [34]. It is built keeping in mind certain desired properties in machine learning based surrogate modeling such as smoothness, continuity and differentiability of the learned function, while including reliable uncertainty quantification. In addition, MeGS can perform in a regression only mode, a classifier only mode or in a dual regressor-classifier mode, and allows scalability to large datasets utilizing high-performance computing (HPC) acceleration. In this research, we leverage the two different operating modes and Graphics Processing Unit (GPU) acceleration to learn these smooth manifolds that represent the thermodynamic phase-space.

The choice of the machine learning framework strongly governs the performance of a data-driven model. Thus, it is essential to understand the dataset at hand. The equations of state indicate a smooth, continuous, and sometimes highly non-linear thermodynamic surface with a discontinuity in the fluid properties along the phase envelope. We utilize this understanding when creating the data-driven real-fluid modeling framework as described in the following subsections. First, we introduce the mathematical details of MeGS and then, the surrogate real fluid modeling framework, named MeGS-RFM. The sub-section ends with details of how the training dataset is acquired to train and build MeGS-RFM.

1. Mathematical Background : MeGS

In the domain of machine learning, there are plenty of options when it comes to choosing data-driven regression frameworks. A popular choice is via neural-networks, also known as deep learning [35]. This class of methods form an extremely flexible learning approach that has the ability to be a "universal function approximator". However, every approach has its own drawbacks - neural networks are notoriously difficult to optimize when it comes to choosing network architecture and ensuring good training without overfitting. Additionally, there is no inherent notion of function smoothness or model uncertainty in traditional deep learning. Recent advancements have tried to fill in these gaps via Bayesian networks or ensemble models [36]. However, as detailed in [36], these have their own caveats including the fact that the model uncertainties are sampled, i.e., only approximate.

Gaussian processes (GPs) [37] form another regression modeling approach, which come built on a rigorous mathematical foundation. GPs are stochastic processes defined as a collection of random variables, any subset of which has a jointly Gaussian distribution. A very concise summary of Gaussian processes is presented here, refer to [38] for a detailed study.

GPs assume a regression where the true output $\mathbf{y}$ of a function f at input point $\mathbf{x}$ can be modeled by a signal term $f(\mathbf{x})$ and a noise term ϵ as $\mathbf{y} = f(\mathbf{x}) + \epsilon$. Following the assumption of the random variable having joint Gaussian distribution, the signal term is taken to be a random variable following a GP distribution of its own given by $f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$.

The mean $m(\mathbf{x})$, defined as $m(\mathbf{x}) = E[f(\mathbf{x})]$, is the expected value of $f(\mathbf{x})$ at $\mathbf{x}$. The covariance matrix $k(\mathbf{x}, \mathbf{x}')$ models dependence between the values of the function $f(\mathbf{x})$ at the two different inputs $\mathbf{x}$ and $\mathbf{x}'$ and is defined as $k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$. This covariance matrix, also known as the kernel of the GP forms the "basis functions" used to fit the data, while allowing for noise in the data. Thus, kernels can be guided by knowledge about the data [39], and are generally chosen to be smooth and differentiable.

While having a structured approach, Gaussian processes do not scale well, owing to their $O(n^3)$ complexity [37] where n is the number of training data-points. Further, scalability issues arise when there are multiple inputs or outputs. Among the several methods that have been developed to overcome this issue, we leverage a software package, GPyTorch, which is a modular implementation of GPs built upon a PyTorch back-end [40]. Within the available modules in GPyTorch, MeGS (and subsequently MeGS-RFM) uses the deep GP-based inducing points methods [41, 42]. This allows scalability by reducing the complexity to $O(n)$ [43, 44] making training and inference for large multi-dimensional models tractable. In addition, traditional (single-layer) GPs find it difficult to fit data with discontinuities or having highly non-linear nature, hampered by the structured kernel approach. Deep GPs aim to alleviate that issue to some extent. However, we do not utilize deep GPs directly in our framework, as we found it to be substantially slow. Instead, we enhance the GP with a neural network based encoder.

Encoders (and their common companions, decoders) are essentially neural network blocks. In an encoder-decoder structure, the encoding network is tasked with extracting information, typically from a large input dataset to a smaller 'latent space'. The decoder is the downstream block that in turn reconstructs the original information from the reduced latent space encoding. This aims to act as a method to extract a reduced (or maybe even slightly increased) latent subspace that can accurately encode all the pertinent information in the data.

In MeGS, the encoder serves a slightly modified purpose, sitting between the input data and the Gaussian processes shown in the schematic in Figure 1. This is achieved leveraging the fact that GPyTorch is built upon PyTorch, which is a

capable deep learning library. This allows the DeepGP module of GPyTorch to be appropriately combined with deep neural network encoders to create powerful models. In the final MeGS framework (Figure 1) the input data passes through the encoder first and then the GP, while training the combined model simultaneously. This allows the framework to learn an encoder-GP model where the encoder learns a non-linear transformation for the inputs that allows the GP to fit the multi-dimensional correlated manifolds with more ease. A parallel class of approaches have been demonstrated using the ExactGP approach [44] and ApproximateGP modules [45] within GPyTorch. These approaches, however, attempt to learn the GP kernel function in a single layered GP, known as Deep Kernel Learning. The methodology in MeGS, on the other hand, utilizes the DeepGP module, and uses the deep net encoder to encode the data directly, making the approach more scalable and flexible. The variational evidence lower bound (ELBO) loss [46] is utilized to train the overall model. Regularization is provided by adding the log mean-squared error loss.

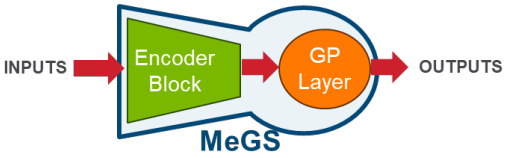


Fig. 1 A schematic representing the encoder-Gaussian process architecture in MeGS

2. Structure of MeGS-RFM

To build MeGS-RFM effectively we first look at the data we are aiming to model. The thermodynamic surfaces are generally smooth with discontinuities at the phase boundaries, and large non-linear variations of some properties, e.g., density or isobaric specific heat, in the vicinity of the critical point. In this research, we utilize a divide and conquer approach to fit and learn generalized thermodynamic surfaces for individual phases.

The entire framework MeGS-RFM consists of four smaller MeGS blocks. Three of those act as regressors that are tasked with learning generalized (across species) thermodynamic surfaces for each of the three phases, namely liquid, vapor and supercritical. Thus, each sub-model is individually focused on a single phase that allows a continuous manifold to be learnt by each MeGS block. The fourth MeGS block acts in its classifier mode and is designed to identify the potential phase of the species under consideration at the desired operating parameters (here, temperature and pressure). Depending on the decision of the classifier, MeGS-RFM chooses which of three phase models will be used to predict the thermodynamic properties. A schematic of the MeGS-RFM block and the relevant model inputs and outputs are shown in Figure. **??**. Once trained, the combined framework has low inference times on either CPUs or GPUs, which is of paramount importance when considering it as an “alternate EoS" in a CFD environment.

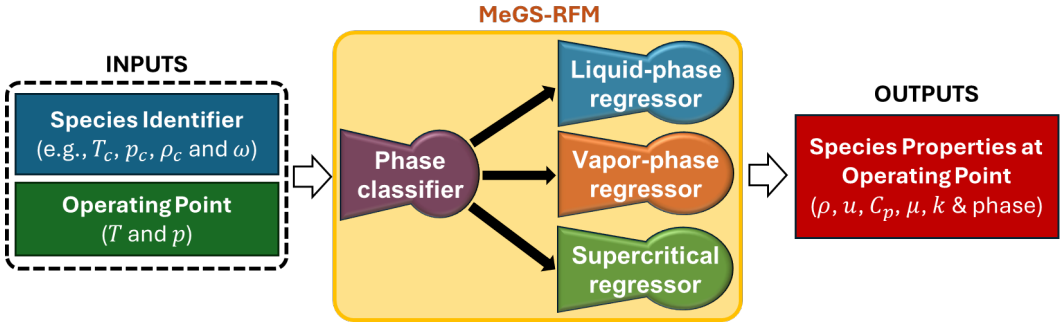


Fig. 2 A schematic representing the block-structure of MeGS-RFM, showing the inputs and outputs. MeGS-RFM comprises of 4 MeGS blocks, one in classifier-mode acting to determine the phase, followed by 3 phase-specific regressor-only blocks

3. Sampling the Thermodynamic Space

The thermodynamic surfaces described by the equation of state, as well as the correlations describing the thermodynamic quantities, are essentially high-dimensional manifolds. In the previous section we describe how this information guides the choice of model. In this section, we explain how we use our understanding of the thermodynamic space to build an appropriate sampling algorithm.

Like any machine learning approach, MeGS-RFM expects the data in an input-output format. For purposes of this initial demonstration, the primary operating parameters that have been varied are the pressure and temperature. Thus, these comprise two of our inputs. Additionally, the model has to be supplied with information on the species it is modeling to help distinguish multiple single-component molecules. In the physics-based equation of state, this is achieved by specifying the critical properties of the species, in addition to the acentric factor. These comprise the other model inputs. The outputs from the model are the thermodynamic properties of the specified species at the desired pressure and temperature. The design of MeGS-RFM allows the number of outputs to be flexible, and can model as many thermodynamic properties as desired. For demonstration, in this paper we choose six output quantities: density (ρ), internal energy (u), isobaric specific heat (C_p), dynamic viscosity (μ), thermal conductivity (k), as well as the phase of the species.

To train a model such as this, it is necessary to sample the operating space well, which in this case comprise of two variables, pressure (p) and temperature (T). Sobol sampling [47, 48] is used to sample this two-dimensional space. Sobol sampling is low-discrepancy, implying samples distributed evenly over the input T-p space. The temperature is bounded between 200K and 900K, and the pressure is between 0.01 and 60 bar [16]. It should be noted however, that the thermodynamic phase-space data is complex and often highly non-linear. Thus, traditional sampling methods, even Sobol sampling, might not sample appropriately in the non-linear regions (e.g., sharp density transition or specific heat across the pseudo-boiling line at supercritical conditions). Although not explored in this work, intelligent sampling, guided by model uncertainty could be utilized to build better training datasets. This is planned as future work.

In order to acquire the appropriate good-quality data for a large range of species, NIST REFPROP v10.0 [49] is utilized. REFPROP (acronym for REFerence fluid PROPerties) is a computer program developed and maintained by National Institute of Standards and Technology (NIST), which calculates the thermodynamic and transport properties of industrially important fluids and their mixtures. This choice is made to have a data-source that can provide thermodynamic data that is well validated and can be justifiably chosen as “ground truth". REFPROP v10.0 also allows the use of a Python based script to acquire and store data, which allows easy data sampling and organization for machine-learning.

Tabulated thermodynamic data for 44 hydrocarbon species is available in the REFPROP v10.0 database. The complete list is provided in the first 44 entries in Table 1. In order to create a dataset relevant to simulating reacting flows, we add 6 other species typically associated with combustion modeling. These are listed in the last 6 entries of Table 1 and are the major components of air (nitrogen and oxygen), the basic products of combustion (water, carbon dioxide, carbon monoxide) and hydrogen, another zero-carbon fuel of interest. This brings up a total of 50 species that we use to train and validate MeGS-RFM. In addition, these 6 non-hydrocarbon species provide insights into how generalization of the framework performs.

Internally, REFPROP uses state-of-the-art pure fluid and mixture models that are based on the most accurate equations of state for each species. This also means that the data from these models is bounded by the temperature within which they are valid. These limits are maintained when creating the dataset to ensure MeGS-RFM is trained on the highest quality of data available.

With our approach, we expect that by providing this large and diverse dataset for MeGS-RFM to learn from, the framework will learn appropriate mathematical correlations that would generalize to an even broader set of species - including those that are not a part of the training set. This is motivated by the need to characterize new fuel species that are pertinent to SAF, that often do not have much reported experimental data.

IV. Results and Discussion

A. Validating MeGS-RFM against NIST REFPROP v10.0 Data

MeGS-RFM aims to learn a generalized thermodynamic space that can model multiple thermodynamic properties of interest for various species over a range of operating pressures and temperatures. As described in the previous section, the data for training MeGS-RFM was obtained from the NIST software, REFPROP v10.0.

Following the practice in modeling the physics based equations of state using the VT-SRK method (see [16]), 50

Species	Carbon Number
Methane	1
Ethane	2
Ethylene	2
Propane	3
Propylene	3
Propyne	3
Isobutane	4
Butane	4
trans-Butene	4
cis-Butene	4
Butene	4
Isobutene	4
1,3-Butadiene	4
1-Butyne	4
Cyclobutene	4
Isopentane	5
Pentane	5
Neopentane	5
Pentene	5
1-Pentene	5
Cyclopentane	5
Isohexane	6
Hexane	6
2,2-Dimethylbutane	6
2,3-Dimethylbutane	6

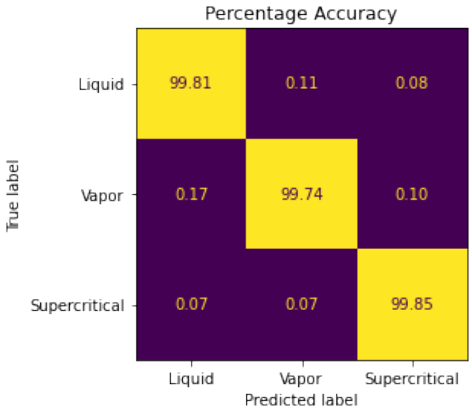
Species	Carbon Number
3-Methylpentane	6
Benzene	6
Cyclohexane	6
Heptane	7
Toluene	7
Methylcyclohexane	7
Isooctane	8
Octane	8
Ethylbenzene	8
m-Xylene	8
o-Xylene	8
p-Xylene	8
Nonane	9
Propylcyclohexane	9
Decane	10
Undecane	11
Dodecane	12
Hexadecane	16
Docosane	22
Carbon dioxide	-
Carbon monoxide	-
Oxygen	-
Nitrogen	-
Hydrogen	-
Water	-

Table 1 Listing of species used to train, test and validate the frameworks presented in this research

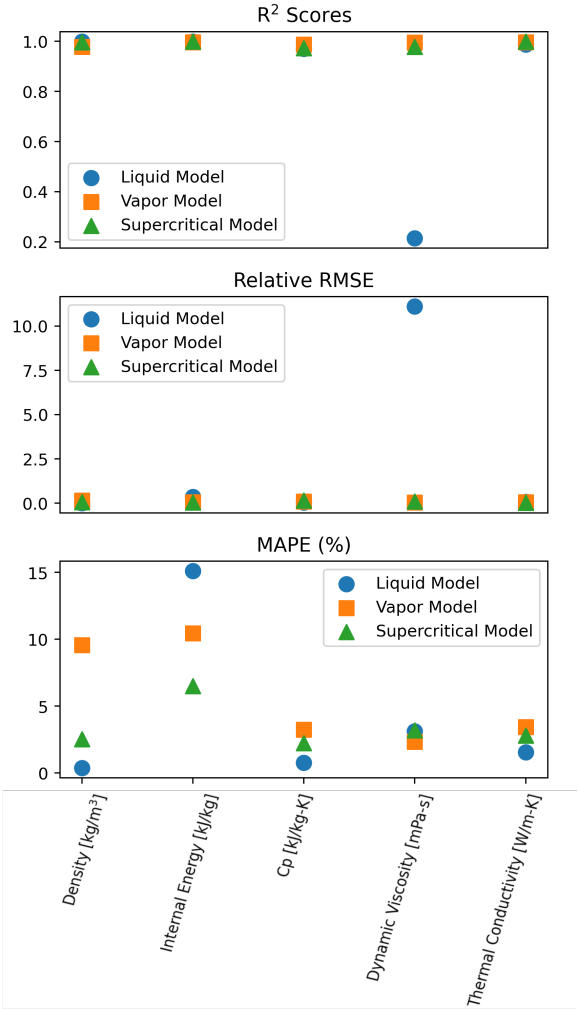
species are identified by their critical properties (ρ_c, p_c and T_c) and acentric factors ω. These form 4 out of the 6 inputs to MeGS-RFM, supplemented by the operating pressure and temperature. As described, the 6 model outputs are the phase and ρ, u, C_p, μ and k.

Of a total of 50 species, 48 were retained in the training set and 2 were completely held out during training. From the 48 species, only 70% of the available data-points sampled from REFPROP v10.0 are used to train the model. The remaining 30% acts to validate the model accuracy of the framework across the species that it has “seen" while training. Finally, the trained MeGS-RFM is asked to predict the thermodynamic properties of the 2 species that have been held out and are “unseen" by the model. High accuracy across this test set provides confidence that the model is not just fitting the data but is learning the underlying latent manifold we expect it to. Model training was performed on Swing, Argonne’s Laboratory Computing Resource Center (LCRC) GPU computing resources. Each Swing node comprises 8 NVIDIA A100 GPUs, of which only one was used at a time for training MeGS-RFM. Training the entire MeGS-RFM framework took slightly less than an hour when leveraging GPU acceleration, using $\sim$ 21,000 data-points for training (70% of 48 species data).

The encoder-Gaussian process structure requires some model tuning to get the best encoder architecture for each MeGS sub-module in MeGS-RFM. This is achieved by individually tuning each of the encoders until good training, validation and testing scores were obtained. We compared several metrics to “score" our models, including Maximum Absolute Percentage Error (MAE or MAPE) utilized also in [16] to assess the performance of physics-based EoS.



(a)



(b)

Fig. 3 Model scores for MeGS-RFM; using the critical properties and acentric factor to characterize the species (a) Phase classification accuracy, (b) Regression model scores (liquid, vapor and supercritical) across each input

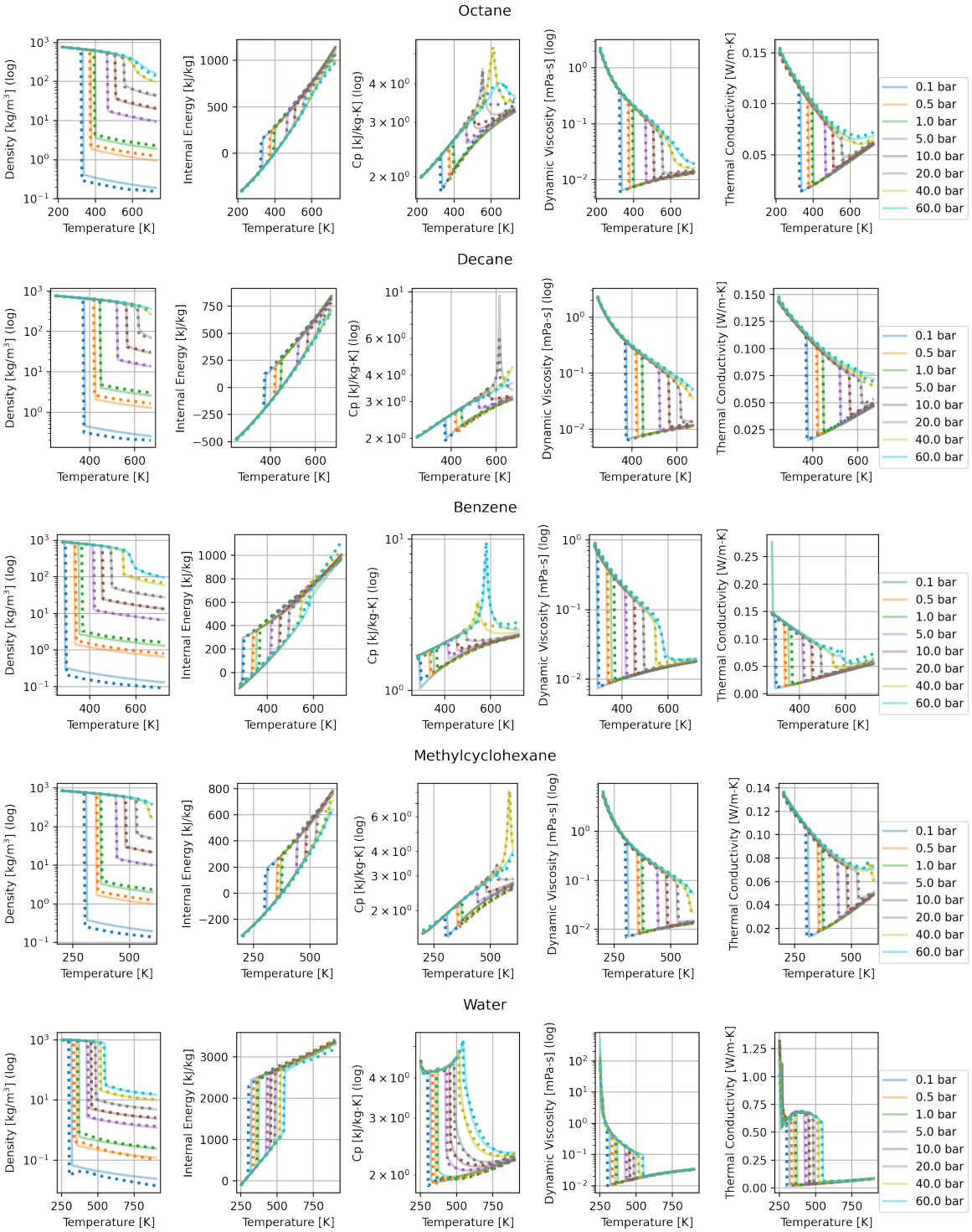


Fig. 4 Demonstrating the performance of MeGS-RFM on 5 representative species; using the critical properties and acentric factor to characterize the species - solid lines indicate ground truth (from REFPROP v10.0) and dashed lines are predicted by MeGS-RFM

However, in this data space, with target values spanning several orders of magnitude, MAPE alone was unreliable. To assess performance, we added the R^2 score and the relative root mean square error (RRMSE). For result interpretation, an R^2 score of 1 is a perfect model with 0 being a poor model, while for RRMSE, a score of 0 is a good model with higher values indicating poorer models.

We present some plots that demonstrate the accuracy of the tuned framework. The image in Figure 3a shows the confusion matrix for the classifier model in MeGS-RFM. The plots in Figure 3b show the R^2, RRMSE and MAPE scores for each output modeled with the different symbols representing the liquid, vapor and supercritical models. In this plot, the model errors are reported across the entire dataset. As a reminder, training was performed on 70% data of 48 species, the models were validated on the remaining 30%, and then generalizability was tested using the 2 held-out species.

The confusion matrix in Figure 3a indicates the percentage of samples per class of phase that MeGS-RFM can accurately identify. We see a very good model performance with $\sim$ 100% accuracy, indicated by the strong diagonal. From Figure 3b, we note that the vapor and supercritical models are performing very well across all metrics. The liquid model does well too, albeit with low R^2 and high RRMSE scores for modeling viscosity. In contrast, the MAPE of viscosity is low for all models, while internal energy has the poorest performance. Keeping in mind that each metric has its strengths, we consider a combined assessment score to understand model performance. Overall, these scores indicate that the model does well in learning the desired thermodynamic manifolds.

Figure 4 presents representative plots that demonstrate the performance of MeGS-RFM on 5 selected species. The species include an aromatic hydrocarbon (Benzene), a branched non-aromatic hydrocarbon (Methylcyclohexane), and Water in addition to common n-alkanes (Octane, Decane). This figure indicates that the model is capable of modeling the thermodynamic space in a generalized way with good accuracy. MeGS-RFM is capable of identifying the phase change temperature, identified by the sharp change in properties when changing between liquid and vapor phases (i.e., it has modeled a generalized saturation curve). Additionally, it accurately models transcritical behavior, identified by the peaks in the C_p values.

Minor errors are apparent, especially at the lowest pressure of 0.1 bar; however, the errors themselves are small (ordinate in log scale). This happens due to the sampling strategy. As detailed in the previous section, while Sobol sampling produces samples that are well-spread, it may not be the most data-efficient since it does not take into account the data complexity. Guided sampling via active learning will boost model performance while lowering training data size by sampling only the “required" data needed for improvement. For example, more sampling might be required near the phase boundaries to accurately capture the “edge" of the manifolds, or across the pseudo-boiling line at supercritical conditions where fluid properties vary sharply and non-linearly. MeGS already has an active learning capability built-in using uncertainty quantification [34] and an extension of this work will target smart sampling for training future variants of this framework. This will be especially useful when increasing the species in the training dataset to hundreds of species.

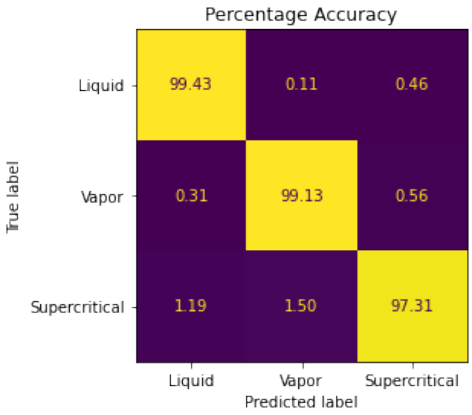
B. Modified Framework : MeGS-RFM$_{simple}$

The methodology described in the previous section uses a physics-driven understanding when choosing the critical properties and acentric factor as input properties used to identify the species of interest. However, for practical purposes, obtaining these properties is not straightforward. As we target new fuel blends, this often requires looking at species that are not always well characterized, and thus, all critical parameters or acentric factor might not be immediately available.

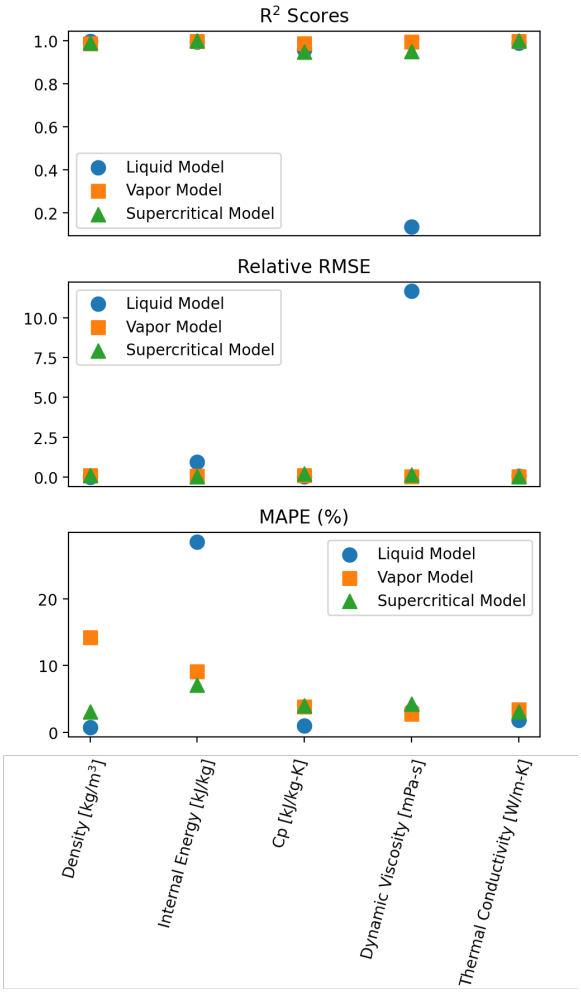
To find an alternate approach that may speed up fuel characterization, we once again leverage the structure of MeGS-RFM and the ability of the framework to learn in a data-driven manner. We do this by replacing the species parameters (i.e., acentric factor and critical properties) with physical properties that are more easily obtainable, the molecular weight (MW) and normal boiling point (T_b) of the species. This modified variant of MeGS-RFM (referred to as MeGS-RFM$_{simple}$ henceforth) is expected to make the task of fuel characterization simpler.

Beyond data availability, the choice of MW and T_b as input parameters is rooted in group-contribution methods calculating fluid properties based on the molecular structure of each species, e.g., [31], which may include the calculation of critical properties and acentric factor as well. Effectively, MeGS-RFM$_{simple}$ is tasked with learning the molecular details of each species given MW and T_b, which are understood as the the basis upon which any fluid model can be built, and used in the evaluation of other properties.

MeGS-RFM$_{simple}$ has 4 inputs (reduced from the 6 inputs in MeGS-RFM) comprising the operating pressure and temperature with the species being identified by its MW and T_b. The outputs, as described in the previous section, are the phase prediction and 5 chosen thermodynamic properties (ρ, u, C_p, μ, k).



(a)



(b)

Fig. 5 Model scores for MeGS-RFM$_{simple}$; using the molecular weight and normal boiling point to characterize the species (a) Phase classification accuracy, (b) Regression model scores (liquid, vapor and supercritical) across each input

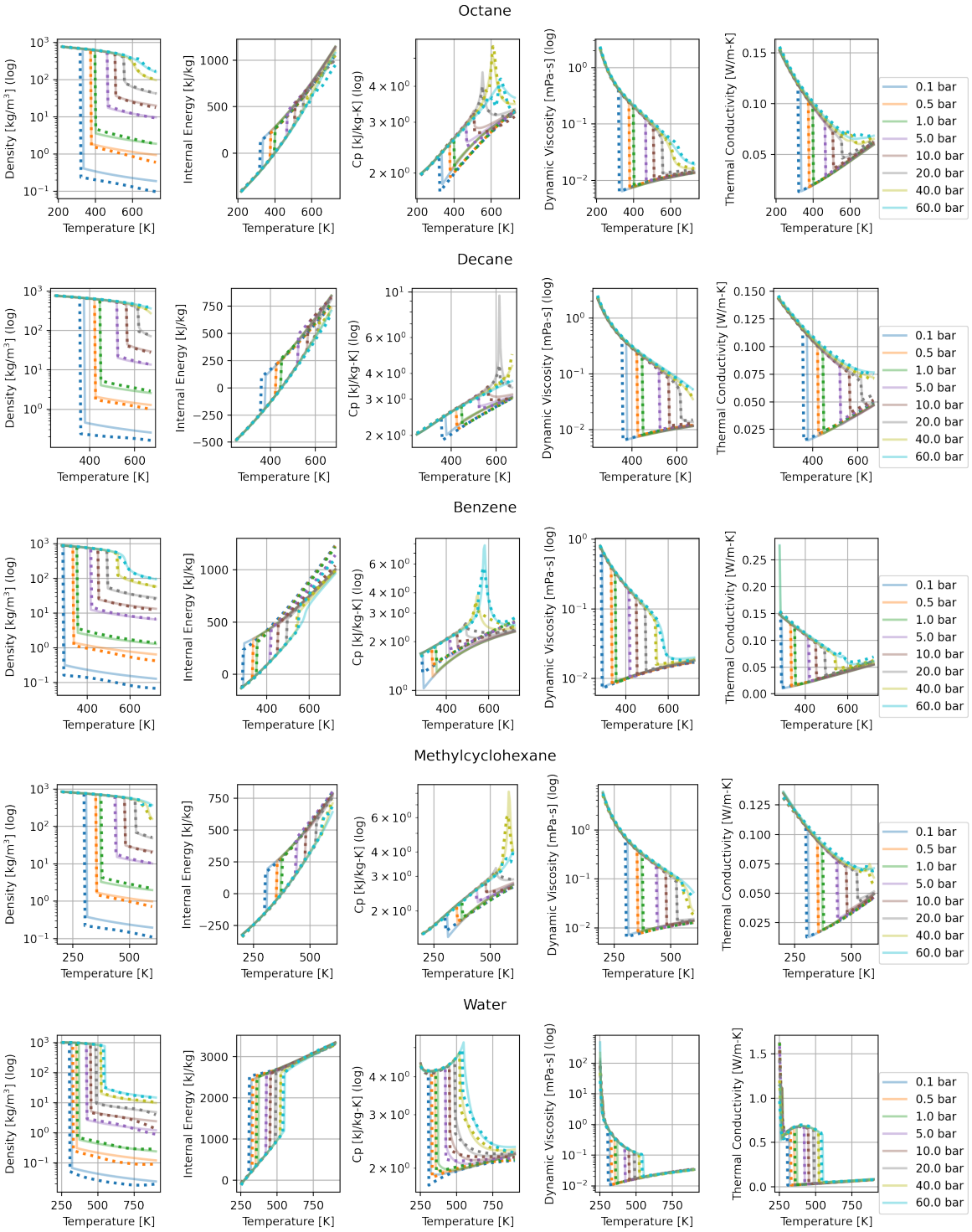


Fig. 6 Demonstrating the performance of MeGS-RFM$_{simple}$ on 5 representative species; using the molecular weight and normal boiling point to characterize the species - solid lines indicate ground truth (from REFPROP v10.0) and dashed lines are predicted by MeGS-RFM$_{simple}$

Once again, we report model scores and results similar to the previous section. Figure 5a reports the confusion matrix showing good classification accuracy across all phases. Similarly, Figure 5b also indicates that MeGS-RFM$_{simple}$ is also adept at learning the thermodynamic manifolds, showing low errors overall. In fact, MeGS-RFM$_{simple}$ has a comparable performance to MeGS-RFM, indicating that this data-driven framework can model the complex correlations between MW and T_b and the thermodynamic state properties. This alleviates the need for the critical properties and acentric factor when modeling and characterizing new species.

C. Obtaining Critical Properties and Acentric Factor directly from Boiling Point and Molecular Weight

The success of MeGS-RFM$_{simple}$ also implies that the framework can build a correlation between the T_b and MW and the critical properties and ω. To test this, we build a small standalone model, once again using a single MeGS regressor block, that takes in the T_b and MW as inputs to the model and attempts to predict the critical properties and ω for those species. Once again, we revert to our 50-species dataset (Table 1) as our training and testing set. A 50-50 train-test data split is used for the purposes of this demonstration.

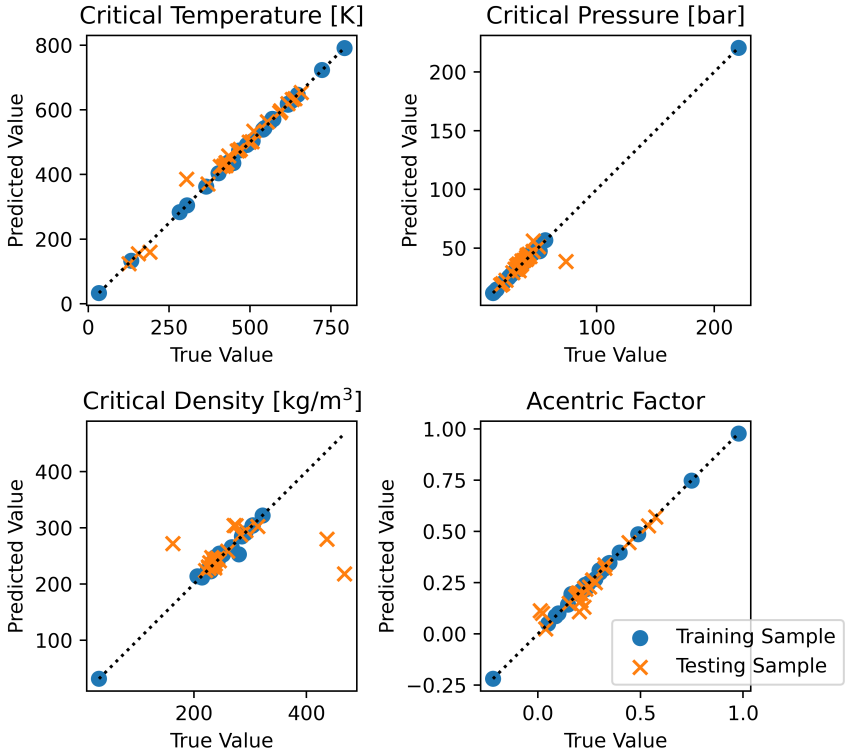


Fig. 7 Estimating critical properties and acentric value from molecular weight and boiling point: True vs. predicted across 50-50 train and test data split - ideal prediction line shown in dotted black

Figure 7 reports the result for this model, showing the model predictions across 25 training and 25 held out test species. The four sub-plots show the predicted vs. true value for the four outputs (P_c, T_c, ρ_c and the acentric factor). A good model would have predictions along the diagonal line (indicated by the black dotted line). We see that our MeGS based model is able to learn the appropriate correlations, giving us further insight into why MeGS-RFM$_{simple}$ has performance similar to that of the regular MeGS-RFM.

D. Model Performance on n-alkanes compared to VT-SRK EoS

We first compare the MeGS-RFM results to those obtained using the VT-SRK EoS [16]. This is illustrated in Figures 8 and 9, where the ground truth (REFPROP) is shown by a solid line, the MeGS-RFM predictions via dashed lines and the VT-SRK EoS using dotted lines. Predictions are shown over a range of temperatures at various pressures. It is clear that the deviation of MeGS-RFM is concentrated at low pressures, and in general low in magnitude. MeGS-RFM does a

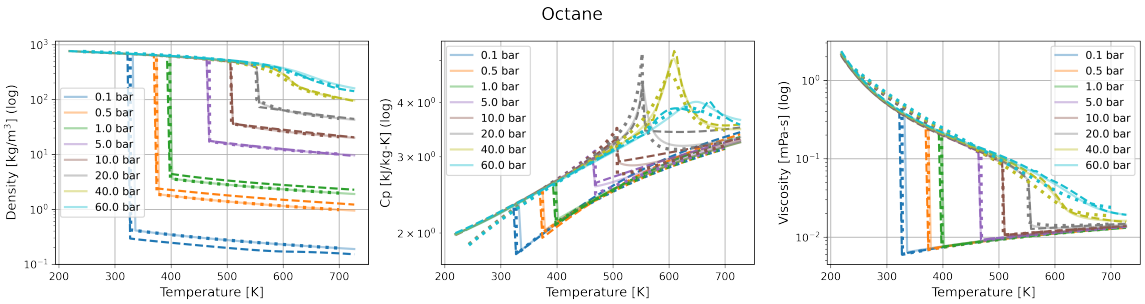


Fig. 8 Comparing REFPROP ground truth (solid line), MeGS-RFM prediction (dashed line) and VT-SRK EoS predictions (dotted line) at various pressures and temperatures for the n-paraffin Octane

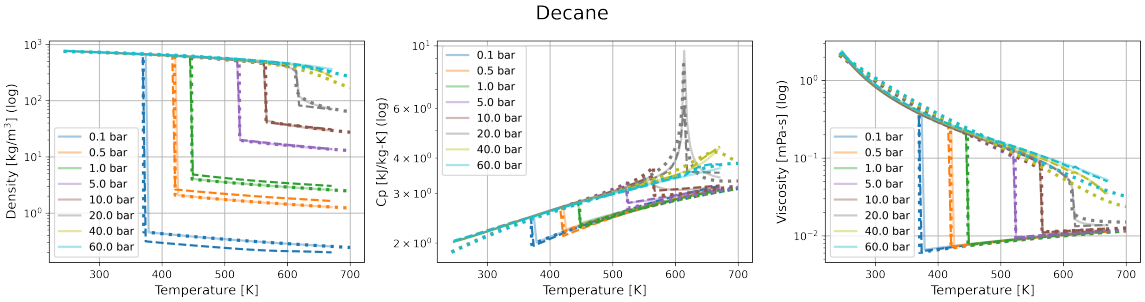


Fig. 9 Comparing REFPROP ground truth (solid line), MeGS-RFM prediction (dashed line) and VT-SRK EoS predictions (dotted line) at various pressures and temperatures for the n-paraffin Decane

relatively good job for density, C_p, and viscosity. The data-driven MeGS-RFM is also able to capture trans-critical behavior with similar accuracy to the EoS. As mentioned, intelligent active learning type sampling will improve the model performance by acquiring more training data at discontinuities and non-linear regions of the thermodynamic manifolds.

E. Model Performance on Iso-paraffins

The essential motivation for this work is to build a generalized real fluid model that uses a data-driven approach as compared to physics-based equations of state such as the VT-SRK [16]. This, in turn, is driven by the need to understand fuel characteristics for SAF, often composed of component species that may be poorly characterized, such as highly branched iso-paraffins largely present in NJFCP's category C surrogates fuels [3].

We propose that a well-trained data-driven real-fluid modeling framework such as MeGS-RFM or MeGS-RFM$_{simple}$ can substitute the physics-based EoS methods to model the poorly characterized fuel species. To serve as proof-of-concept, we demonstrate this on two iso-paraffins of relevance to SAF, iso-dodecane and iso-hexadecane [3].

The plots in Figures 10 and 11 present results on predicting density and dynamic viscosity for these fuels. The ground truth here is provided by limited low-P, low-T experimental data [50], and we compare model performance of the VT-SRK physics based equation of state [16], MeGS-RFM and MeGS-RFM$_{simple}$.

Clearly, we see that both MeGS-RFM and MeGS-RFM$_{simple}$ can accurately predict the density and dynamic viscosity for these iso-paraffins that are not part of the training set (or even REFPROP v10.0, our training data-source). Both data-driven frameworks are accurately able to model the density variation with better accuracy than the VT-SRK approach, with MeGS-RFM seeming to do a better job than MeGS-RFM$_{simple}$ for iso-hexadecane. Moreover, MeGS-RFM recovers the nearly linear behavior of density at low temperatures while MeGS-RFM$_{simple}$ still seemingly suffers from similar non-linear density variations as VT-SRK. The prediction for dynamic viscosity is also vastly improved over VT-SRK. The values for critical properties, acentric factor, molecular weight and boiling point for these iso-paraffins used as inputs are listed in [16].

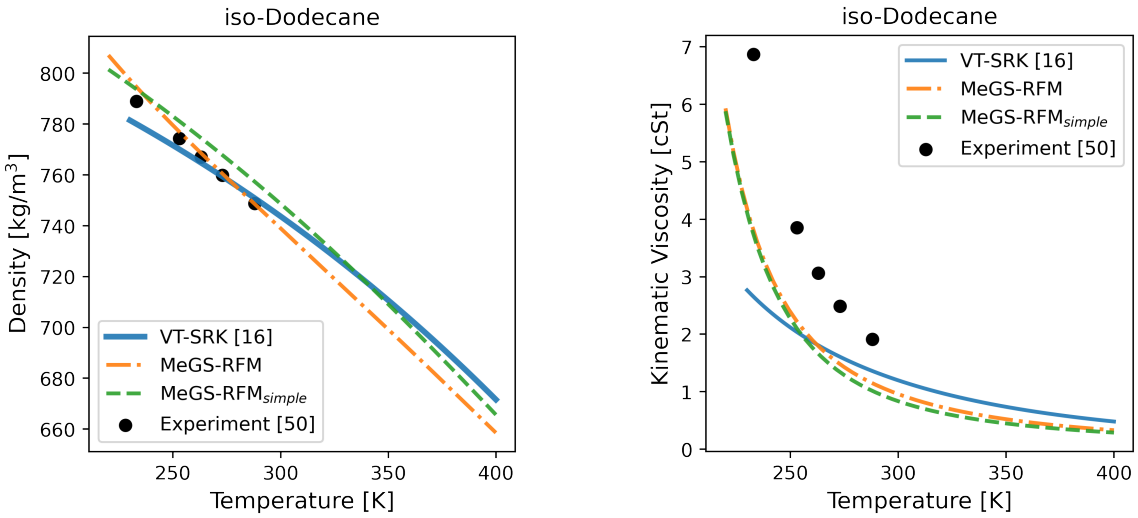


Fig. 10 Characterizing iso-dodecane : Comparing predictions from the VT-SRK EoS [16], MeGS-RFM, and MeGS-RFM$_{simple}$ to experimental data [50]

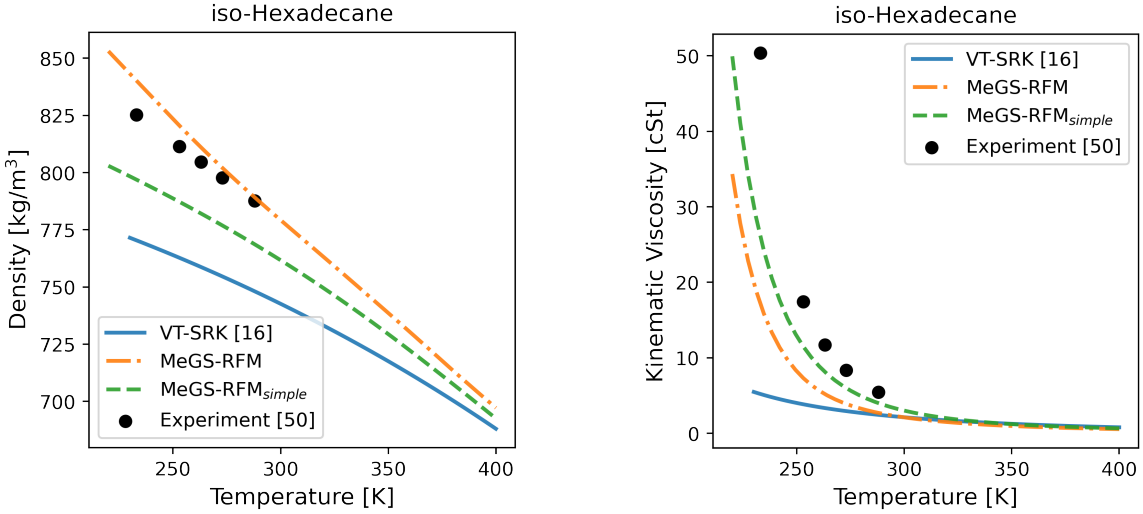


Fig. 11 Characterizing iso-hexadecane : Comparing predictions from the VT-SRK EoS [16], MeGS-RFM, and MeGS-RFM$_{simple}$ to experimental data [50]

V. Conclusions and Future Work

In this research paper, we present and demonstrate MeGS-RFM, and MeGS-RFM$_{simple}$, both generalizable data-driven real-fluid modeling frameworks using a generative approach. In its current form, the frameworks are able to model the thermodynamic states (ρ, u, C_p, μ, k and phase) for a variety of single-component species, including hydrocarbon fuels, air and combustion products with good accuracy. The inputs to MeGS-RFM are the same as those in physics-based equations of state: the critical properties and acentric factor of the species. Since obtaining these values is difficult for lesser characterized species, MeGS-RFM$_{simple}$ aims to simplify the task by using the molecular weight and normal boiling temperature instead. This would allow fast characterization of new fuel species relevant to SAF. Further, we show that these generalized frameworks are capable of predicting properties for lesser-characterized fuel species, focusing on 2 hydrocarbons of interest to SAF, iso-dodecane and iso-hexadecane. We show improvements compared to the VT-SRK physics-based equation of state on single-component fuels, validated by the authors in previous works [16, 17].

Since SAFs are mixtures of hydrocarbons, the eventual goal is to create a generalized data-driven framework to model multi-component mixtures. This is a larger problem, requiring a different approach, and efforts are still in their nascent stage. In on-going preliminary studies, a MeGS based mixture modeling framework shows its capability to model the interactions between multiple-species in a mixture, and obtain mixture properties, fugacity coefficients and mass fractions in each phase. This mixture model also requires a different machine learning architecture. The models presented in this research, MeGS-RFM and MeGS-RFM$_{simple}$, serve as initial proofs of concept leading towards this larger effort. Once trained and validated, frameworks like these have very low inference time, and in the long term can act as alternatives to equations of state formulations used in CFD. This will accelerate high-fidelity simulations and analyses of the multi-phase effects in multi-component fuels, such as SAF, under realistic combustion environments.

Acknowledgments

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. http://energy.gov/downloads/doe-public-access-plan

This work has been funded by the DOE Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office by technology manager, Kevin Stork and program manager Gurpreet Singh. The machine learning models were developed using the Swing GPU cluster, operated by the Laboratory Computing Resource Center (LCRC) at Argonne.

References

[1] Dasgupta, D., Goldsborough, S., Sforzo, B., Som, S., McNenly, M., Wagnon, S., Fioroni, G., Yellapantula, S., Phil, J., Szybist, J., et al., "Workshop on Sustainable Aviation Fuel-Use Research Opportunities," Tech. rep., Argonne National Laboratory (ANL), Argonne, IL (United States), 2023. https://doi.org/10.2172/1969251.

[2] Kurzawska, P., "Overview of Sustainable Aviation Fuels including emission of particulate matter and harmful gaseous exhaust gas compounds," *Transportation Research Procedia*, Vol. 59, 2021, pp. 38–45. https://doi.org/10.1016/j.trpro.2021.11.095.

[3] Colket, M., Heyne, J., Rumizen, M., Gupta, M., Edwards, T., Roquemore, W., Andac, G., Boehm, R., Lovett, J., Williams, R., et al., "Overview of the national jet fuels combustion program," *AIAA Journal*, Vol. 55, No. 4, 2017, pp. 1087–1104. https://doi.org/10.2514/1.J055361.

[4] Kallupalayam Ramasamy, K., Thorson, M., Billing, J., Holladay, J., Drennan, C., Hoffman, B., and Haq, Z., "Hydrothermal Liquefaction: Path to Sustainable Aviation Fuel," Tech. rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2021. https://doi.org/10.2172/1821809.

[5] Huber, M., Lemmon, E., and Bruno, T., "Surrogate mixture models for the thermophysical properties of aviation fuel Jet-A," *Energy & Fuels*, Vol. 24, No. 6, 2010, pp. 3565–3571. https://doi.org/10.1021/ef100208c.

[6] Dooley, S., Won, S., Chaos, M., Heyne, J., Ju, Y., Dryer, F., Kumar, K., Sung, C.-J., Wang, H., Oehlschlaeger, M., et al., "A jet fuel surrogate formulated by real fuel properties," *Combustion and Flame*, Vol. 157, No. 12, 2010, pp. 2333–2339. https://doi.org/10.1016/j.combustflame.2010.07.001.

[7] Jones, W., Marquis, A., and Vogiatzaki, K., "Large-eddy simulation of spray combustion in a gas turbine combustor," *Combustion and Flame*, Vol. 161, No. 1, 2014, pp. 222–239. https://doi.org/10.1016/j.combustflame.2013.07.016.

[8] Yu, J., Wang, Z., Zhuo, X., Wang, W., and Gou, X., "Surrogate definition and chemical kinetic modeling for two different jet aviation fuels," *Energy & Fuels*, Vol. 30, No. 2, 2016, pp. 1375–1382. https://doi.org/10.1021/acs.energyfuels.5b02414.

[9] Edwards, T., Colket, M., Cernansky, N., Dryer, F., Egolfopoulos, F., Friend, D., Law, E., Lenhert, D., Lindstedt, P., Pitsch, H., et al., "Development of an experimental database and kinetic models for surrogate jet fuels," *45th AIAA Aerospace Sciences Meeting and Exhibit*, 2007, p. 770. https://doi.org/10.2514/6.2007-770.

[10] Jofre, L., and Urzay, J., "Transcritical diffuse-interface hydrodynamics of propellants in high-pressure combustors of chemical propulsion systems," *Progress in Energy and Combustion Science*, Vol. 82, 2021, p. 100877. https://doi.org/10.1016/j.pecs.2020.100877.

[11] Wang, X., Wang, Y., and Yang, V., "Three-dimensional flow dynamics and mixing in a gas-centered liquid-swirl coaxial injector at supercritical pressure," *Physics of Fluids*, Vol. 31, No. 6, 2019. https://doi.org/10.1063/1.5097163.

[12] Sciacovelli, L., and Bellan, J., "The influence of the chemical composition representation according to the number of species during mixing in high-pressure turbulent flows," *Journal of Fluid Mechanics*, Vol. 863, 2019, pp. 293–340. https://doi.org/10.1017/jfm.2018.992.

[13] Poblador-Ibanez, J., and Sirignano, W., "Temporal atomization of a transcritical liquid n-decane jet into oxygen," *International Journal of Multiphase Flow*, Vol. 153, 2022, p. 104130. https://doi.org/10.1016/j.ijmultiphaseflow.2022.104130.

[14] Fathi, M., Hickel, S., and Roekaerts, D., "Large eddy simulations of reacting and non-reacting transcritical fuel sprays using multiphase thermodynamics," *Physics of Fluids*, Vol. 34, No. 8, 2022. https://doi.org/10.1063/5.0099154.

[15] Gaballa, H., Habchi, C., and de Hemptinne, J.-C., "Modeling and LES of high-pressure liquid injection under evaporating and non-evaporating conditions by a real fluid model and surface density approach," *International Journal of Multiphase Flow*, Vol. 160, 2023, p. 104372. https://doi.org/10.1016/j.ijmultiphaseflow.2022.104372.

[16] Poblador-Ibanez, J., and Nocivelli, L., "Toward a Real-Fluid Modeling Framework for Sustainable Aviation Fuels," *Fuel Communications*, Vol. 18, 2024, p. 100100. https://doi.org/10.1016/j.jfueco.2023.100100.

[17] Poblador-Ibanez, J., Nocivelli, L., and Dasgupta, D., "Multi-component Evaporation of Sustainable Aviation Fuel Droplets," *AIAA SCITECH 2024 Forum*, 2024, p. 1997. https://doi.org/10.2514/6.2024-1997.

[18] Longmire, N. P., and Banuti, D., "Extension of SU2 using neural networks for thermo-fluids modeling," *AIAA Propulsion and Energy 2021 Forum*, 2021, p. 3593. https://doi.org/10.2514/6.2021-3593.

[19] Wu, Y., and Sun, S., "Removing the performance bottleneck of pressure–temperature flash calculations during both the online and offline stages by using physics-informed neural networks," *Physics of Fluids*, Vol. 35, No. 4, 2023. https://doi.org/10.1063/5.0150341.

[20] Sahranavardfard, N., Aubagnac-Karkar, D., Costante, G., Rahantamialisoa, F., Habchi, C., and Battistoni, M., "Computation of Real-Fluid Thermophysical Properties Using a Neural Network Approach Implemented in OpenFOAM," *Fluids*, Vol. 9, No. 3, 2024, p. 56. https://doi.org/10.3390/fluids9030056.

[21] Soave, G., "Equilibrium constants from a modified Redlich-Kwong equation of state," *Chemical Engineering Science*, Vol. 27, No. 6, 1972, pp. 1197–1203. https://doi.org/10.1016/0009-2509(72)80096-4.

[22] Peng, D.-Y., and Robinson, D., "A new two-constant equation of state," *Industrial & Engineering Chemistry Fundamentals*, Vol. 15, No. 1, 1976, pp. 59–64. https://doi.org/10.1021/i160057a011.

[23] Lin, H., and Duan, Y.-Y., "Empirical correction to the Peng–Robinson equation of state for the saturated region," *Fluid Phase Equilibria*, Vol. 233, No. 2, 2005, pp. 194–203. https://doi.org/10.1016/j.fluid.2005.05.008.

[24] Lin, H., Duan, Y.-Y., Zhang, T., and Huang, Z.-M., "Volumetric property improvement for the Soave- Redlich- Kwong equation of state," *Industrial & Engineering Chemistry Research*, Vol. 45, No. 5, 2006, pp. 1829–1839. https://doi.org/10.1021/ie051058v.

[25] Baled, H., Enick, R., Wu, Y., McHugh, M., Burgess, W., Tapriyal, D., and Morreale, B., "Prediction of hydrocarbon densities at extreme conditions using volume-translated SRK and PR equations of state fit to high temperature, high pressure PVT data," *Fluid Phase Equilibria*, Vol. 317, 2012, pp. 65–76. https://doi.org/10.1016/j.fluid.2011.12.027.

[26] Yang, V., "Modeling of supercritical vaporization, mixing, and combustion processes in liquid-fueled propulsion systems," *Proceedings of the Combustion Institute*, Vol. 28, No. 1, 2000, pp. 925–942. https://doi.org/10.1016/S0082-0784(00)80299-4.

[27] Prausnitz, J., and Tavares, F., "Thermodynamics of fluid-phase equilibria for standard chemical engineering operations," *AIChE Journal*, Vol. 50, No. 4, 2004, pp. 739–761. https://doi.org/10.1002/aic.10069.

[28] Poling, B., Prausnitz, J., O'connell, J., et al., *The Properties of Gases and Liquids*, Vol. 5, McGraw-Hill New York, 2001.

[29] Chung, T., Ajlan, M., Lee, L., and Starling, K., "Generalized multiparameter correlation for nonpolar and polar fluid transport properties," *Industrial & Engineering Chemistry Research*, Vol. 27, No. 4, 1988, pp. 671–679. https://doi.org/10.1021/ie00076a024.

[30] Nannoolal, Y., Rarey, J., and Ramjugernath, D., "Estimation of pure component properties: Part 2. Estimation of critical property data by group contribution," *Fluid Phase Equilibria*, Vol. 252, No. 1-2, 2007, pp. 1–27. https://doi.org/10.1016/j.fluid.2006.11.014.

[31] Joback, K., and Reid, R., "Estimation of pure-component properties from group-contributions," *Chemical Engineering Communications*, Vol. 57, No. 1-6, 1987, pp. 233–243. https://doi.org/10.1080/00986448708960487.

[32] Tahami, S., Ghasemitabar, H., and Movagharnejad, K., "Estimation of the acentric factor of organic compounds via a new group contribution method," *Fluid Phase Equilibria*, Vol. 499, 2019, p. 112246. https://doi.org/10.1016/j.fluid.2019.112246.

[33] Muller, K., Mokrushina, L., and Arlt, W., "Second-order group contribution method for the determination of the dipole moment," *Journal of Chemical & Engineering Data*, Vol. 57, No. 4, 2012, pp. 1231–1236. https://doi.org/10.1021/je2013395.

[34] Bhattacharya, C., Christopher, J., Thierry, D., Biruduganti, M., Supekar, S., and Dasgupta, D., "Data-Driven Surrogate Modeling of Microturbine Combustors Burning Hydrogen Blends," *Turbo Expo: Power for Land, Sea, and Air*, Vol. 87035, American Society of Mechanical Engineers, 2023, p. V009T18A011. https://doi.org/10.1115/GT2023-103229.

[35] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *Nature*, Vol. 521, No. 7553, 2015, pp. 436–444. https://doi.org/10.1038/nature14539.

[36] Nemani, V., Biggio, L., Huan, X., Hu, Z., Fink, O., Tran, A., Wang, Y., Zhang, X., and Hu, C., "Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial," *Mechanical Systems and Signal Processing*, Vol. 205, 2023, p. 110796. https://doi.org/10.1016/j.ymssp.2023.110796.

[37] Williams, C., and Rasmussen, C., "Gaussian processes for regression," *Advances in Neural Information Processing Systems*, Vol. 8, 1995. https://doi.org/https://doi.org/10.7551/mitpress/3206.001.0001.

[38] Schulz, E., Speekenbrink, M., and Krause, A., "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, Vol. 85, 2018, pp. 1–16. https://doi.org/10.1016/j.jmp.2018.03.001.

[39] Jäkel, F., Schölkopf, B., and Wichmann, F., "A tutorial on kernel methods for categorization," *Journal of Mathematical Psychology*, Vol. 51, No. 6, 2007, pp. 343–358. https://doi.org/10.1016/j.jmp.2007.06.002.

[40] Gardner, J., Pleiss, G., Weinberger, K., Bindel, D., and Wilson, A., "Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration," *Advances in Neural Information Processing Systems*, Vol. 31, 2018. https://doi.org/10.5555/3327757.3327857.

[41] Damianou, A., and Lawrence, N., "Deep gaussian processes," *Artificial Intelligence and Statistics*, PMLR, 2013, pp. 207–215. https://doi.org/10.48550/arXiv.2106.12135.

[42] Salimbeni, H., and Deisenroth, M., "Doubly stochastic variational inference for deep Gaussian processes," *Advances in Neural Information Processing Systems*, Vol. 30, 2017. https://doi.org/10.5555/3294996.3295212.

[43] Damianou, A., Titsias, M., and Lawrence, N., "Variational Gaussian process dynamical systems," *Advances in Neural Information Processing Systems*, Vol. 24, 2011. https://doi.org/10.5555/2986459.2986739.

[44] Wilson, A., Hu, Z., Salakhutdinov, R., and Xing, E., "Deep kernel learning," *Artificial Intelligence and Statistics*, PMLR, 2016, pp. 370–378. https://doi.org/10.48550/arXiv.1511.02222.

[45] Wilson, A., Hu, Z., Salakhutdinov, R., and Xing, E., "Stochastic variational deep kernel learning," *Advances in Neural Information Processing Systems*, Vol. 29, 2016. https://doi.org/10.5555/3157382.3157388.

[46] Hensman, J., Matthews, A., and Ghahramani, Z., "Scalable variational Gaussian process classification," *Artificial Intelligence and Statistics*, PMLR, 2015, pp. 351–360. https://doi.org/10.48550/arXiv.1411.2005.

[47] Tarantola, S., Becker, W., and Zeitz, D., "A comparison of two sampling methods for global sensitivity analysis," *Computer Physics Communications*, Vol. 183, No. 5, 2012, pp. 1061–1072. https://doi.org/10.1016/j.cpc.2011.12.015.

[48] Renardy, M., Joslyn, L., Millar, J., and Kirschner, D., "To Sobol or not to Sobol? The effects of sampling schemes in systems biology applications," *Mathematical Biosciences*, Vol. 337, 2021, p. 108593. https://doi.org/10.1016/j.mbs.2021.108593.

[49] Lemmon, E., Bell, I., Huber, M., and McLinden, M., "REFPROP Documentation, v10," , 2018.

[50] Boehm, R., Hauck, F., Yang, Z., Wanstall, C., and Heyne, J., "Error quantification of the Arrhenius blending rule for viscosity of hydrocarbon mixtures," *Frontiers in Energy Research*, Vol. 10, 2022, p. 1074699. https://doi.org/10.3389/fenrg.2022.1074699.