Machine learning models for characterizing macroscopic properties of diesel/ biodiesel surrogate fuels

Rodolfo S. M. Freitas[1], Ágatha P. F. Lima[1], Fernando A. Rochinha[1]

[1]*COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Rio de Janeiro 21941-598, Brazil*
rodolfosmfreitas@gmail.com

Abstract. Accurate determination of fuel properties of complex mixtures over a wide range of pressure and temperature conditions is essential to utilizing alternative fuels. Obtaining thermophysical properties of complex fuels is important for design and analysis but difficult to measure/predict, especially in the range of extreme conditions operations. Molecular dynamics (MD) simulations have been widely used to characterize physicochemical properties of fuels, including transport properties at supercritical conditions. Although MD simulations provide molecular details that can be potentially be used to predict fuel properties accurately, they are generally too expensive in terms of computational costs. In addition, MD predictions also need to be validated against experimental measurements, which can be even more costly, especially in extreme conditions. Accordingly, it is not feasible to establish complete and detailed fuel property databases consisting of a wide range of pressure and temperature conditions using MD simulations or experimental measurements. Machine learning (ML) has great potentials to discover from data the relation between inputs and outputs of complex systems. ML can be a powerful tool to predict fuel properties from chemical compositions of the fuel mixture and/or chemical structures of the fuel molecules. Those models can be trained using the database from MD simulations and/or experimental measurements in a data-fusion-fidelity approach. The present work aims to construct cheap-to-compute ML models to act as closure equations for predicting the physical properties of diesel/biodiesel surrogate fuels. Here, Gaussian Process (GP) and probabilistic generative models are adopted. GP is a popular non-parametric Bayesian approach to build surrogate models mainly due to its capacity to handle the aleatory and epistemic uncertainties. Generative models have shown the ability of deep neural networks employed with the same intent. In this work, machine learning analysis is focused on a particular property, the fuel density, but it can also be extended to other physicochemical properties.

Keywords: Molecular Dynamics, Data-fusion-fidelity machine learning models, Material properties

1 Introduction

Future environment legislation will require both drastically reduce pollutant emissions and more sustainable energy utilization. Following these concerns, alternative fuels will be increasingly more important in the transport industry, and especially for sectors where fossil fuels are the major energy supply. A significant challenge for the utilization of alternative fuels in combustion engines falls in an accurate determination of their physicochemical properties. That is the case of synthetic diesel and biodiesel, which are promise substitutes to petroleum diesel. Another challenge is due to the fact that such alternative fuels are often complex chemical compositions, and simplified surrogate fuels are often used to describe the chemical compositions of these fuels [1–3]. Also, modern combustion devices operating in supercritical conditions have been improving energy conservation efficiency and diminish pollutant emissions [4]. Thus, fuel properties at extreme conditions such as high pressure and temperature are very important to accurately describe the combustion characteristics. However, the physicochemical properties of the fuels at those conditions are extremely difficult to measure, falling to another major challenge.

Also, the design of combustion systems in supercritical conditions becomes more challenging due to the limitations of replicating flow and combustion in controlled laboratory environments. In this regard, computational simulations can provide tools for leveraging the use of modern combustion engines in extreme conditions. From a computational fluid dynamics (CFD) perspective, combustion models must combine the capability to predict the relevant flow features with the ability to solve physical phenomena involved in the combustion process, typically

describing govern quantities of interest such as physicochemical properties of the fuels and the mixing, which are obtained through approaches that can often lead uncertainties. That is for instance the case found within the Engine Combustion like the Spray-A injector [30], where the simulations encompass a variety of quite different conditions. More specifically, the operating conditions can range from 2,000 bar to even 0 bar, and similarly spread is also found with temperatures. Furthermore, such simulations can be more challenging for complex fuels mixtures that fall in the use of synthetic diesel and biodiesel, and simplified surrogate fuels are employed to predict the physicochemical properties [3, 5]. This allows the design of experiments, and also Molecular Dynamics (MD) simulations [6, 7].

Molecular Dynamics has been successfully used to measure the physicochemical properties of diesel surrogate fuels [5]. Despite MD simulations provide essential molecular details that enable to accurately predict the fuel properties, such simulations require a high computational burden. Also, the predictions need to be validated against experimental measurements, which entails a more challenge, especially at supercritical conditions. Consequently, build physicochemical properties databases of fuels for a wide range of temperatures and pressures through MD simulations or experiment measurements becomes unavailable. In this work, we aim to leverage the MD simulations for predicting fuel physicochemical properties using Machine Learning (ML) models. Physicochemical properties are generally expressed as functions of local pressure and temperature, which motivates to refer to the closure models as Equations of State (EoS), that have to be embedded in complex CDF codes, which in turn makes not convenient the use of traditional schemes based on tabular and interpolation schemes. Thus, we aim to build models able to characterize the physicochemical properties of combustion devices over a wide range of operating conditions.

In the present work, we build two different machine learning models. First, we build a Gaussian Process (GP) model, a popular approach due to its success in being a proxy for high-fidelity models in different applications [8–13]. Also, we build a probabilistic conditional generative approach that explores existing low-dimensional structures capable of explaining high-dimensional data introducing probabilistic latent variables [14, 15]. Here, we train both in a data-driven learning approach where expensive MD simulations are used to produce the training dataset. Therefore, we rely on their ability to learn from small data regime and their capacity for extrapolation. Furthermore, we also want to take into consideration the unavoidable uncertainties arising from limited information (epistemic) and from noisy data (aleatoric).

The rest of this paper is organized as follows. Section 2 presents the machine learning models. Section 3 describes the ML results for typical fuel surrogates of diesel and biodiesel. Finally, section 4 concludes the study with recommendation for further investigations.

2 Machine Learning Models to Describe Physicochemical Properties

Here, we present a generic property as a function of pressure and temperature described by the ML models. In the training process, we assume the availability of, potentially expensive, dataset comprising input/output pairs $\{(p,T)_i, \gamma_i \quad i = 1, ..., n\}$ generated by an implicit mapping g characterizing the macroscopic thermodynamic relation between the property and the state variables:

$$\gamma = g(p, T; \xi). \tag{1}$$

Here, the role of g is played by upscaling MD simulations or, to a less extent, by experimental available data. The vector ξ represents potential noisy and is often considered a random. In order to keep a compact notation, we refer to the above dataset as $\mathcal{D} = (\mathbf{x}, \mathbf{y})$, with $\mathbf{x} \in R^{2n}$ and $\mathbf{y} \in R^n$ vectors containing inputs and outputs.

2.1 Gaussian Process regression

A GP represents an infinite collection of random variables, in which any finite number of such variables depict a joint Gaussian distribution [8]. Following the Bayesian perspective, to approximate g we assign a GP zero mean prior $f(\mathbf{x})$, i.e., $f \sim GP(f|\mathbf{0}, k(\mathbf{x}, \mathbf{x}^{'}; \theta))$, where k is a kernel parametrized by a vector of hyper-parameters θ to be learned from $\mathcal{D}$ and engenders a symmetric positive-definite $n \times n$ covariance matrix $K_{ij} = k(x_i, x_j; \theta)$. Also, instead of choosing the squared exponential form of the kernel as usual [8], here, we use the Matern3/2 covariance matrix belonging to the Matern family

$$k(\mathbf{r}, \theta) = \sigma^2 \left(1 + \sqrt{6}\frac{|\mathbf{r}|}{l}\right) exp\left(-\sqrt{6}\frac{|\mathbf{r}|}{l}\right) \tag{2}$$

, where $\mathbf{r} = \mathbf{x} - \mathbf{x}^{'}$ denoting the distance between different inputs, and the hyper-parameters are the standard

deviation σ, and the correlation lengths $\mathbf{l} = \{l_1, l_2, \ldots, l_{n_\mathbf{r}}\}$, and $n_\mathbf{r}$ denotes the dimension of input $\mathbf{r}$. Hence, the hyper-parameters vector reduces to $\theta = \{\mathbf{l}, \sigma\}$.

Also, we do not follow a fully Bayesian approach, and obtain the vector of hyper-parameters θ by maximizing the marginal log-likelihood of the model, i.e.

$$log\, p(\mathbf{y}|\mathbf{x}, \theta) = -\frac{1}{2} log|\mathbf{K}| - \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{n}{2} log 2\pi. \tag{3}$$

using a conjugate gradient descend method. The final aim of the GP regression is to return a predictive model for γ, which means to compute its value for untested state variables $\mathbf{x}_*$ [9]

$$\mu_*(\mathbf{x}_*) = k_{*n}\mathbf{K}^{-1}\mathbf{y} \tag{4}$$

and

$$\sigma_*^2(\mathbf{x}_*) = k_{**} - k_{*n}\mathbf{K}^{-1}k_{*n}^T \tag{5}$$

where $k_{*n} = [k(\mathbf{x}_*, \mathbf{x}_1), \ldots, k(\mathbf{x}_*, \mathbf{x}_n)]$ and $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$. The predictions are computed using the posterior mean μ_*, and the uncertainty associated with that predictions is quantified through the posterior variance σ_*^2. It is worth to mention that in absence of noisy in the training data, the later represents epistemic uncertainty due to lack of data.

2.2 Probabilistic Conditional Generative model

Here, a probabilistic conditional generative approach proposed by [14, 15] is explored to describe physicochemical properties. That model integrates variational auto-encoders (VAE) [16], and generative adversarial networks (GANs) [17] and employ, from the beginning, a probabilistic perspective that enables to take into consideration noisy and limited data. The final goal is to build a data-driven probabilistic neural network that follow a conditional probability density function $p(\gamma|(p, T), \mathcal{D})$. Thus, it can provide accurate values for the property γ by estimating the expectation values $E_p(\gamma|(p, T), \mathcal{D})$, and, also, to quantify the uncertainty associated with the predictions.

More specifically, a vector of latent random variables is introduced aiming at seeking a hidden low dimensional structure for explaining the data structure. Such latent variables allow us to express the conditional probability associate to the data $\mathcal{D}$, $p(\mathbf{y}|\mathbf{x})$, as an infinite mixture model through

$$p(\mathbf{y}|\mathbf{x}) = \int p(\mathbf{y}, \mathbf{z}|\mathbf{x}) d\mathbf{z} = \int p(\mathbf{y}|\mathbf{x}, \mathbf{z}) p(\mathbf{z}|\mathbf{x}) d\mathbf{z} \tag{6}$$

, where $p(\mathbf{z}|\mathbf{x})$ is a prior distribution on the latent variables. Therefore, the first step is to learn the latent variables. A detailed description of the training approach can be seen in [14, 15].

The statistics of the outputs $\mathbf{y}$ can be characterized by sampling the latent variables from the prior $p(\mathbf{z})$ and passing them through the generator to yield conditional samples $\mathbf{y} = f_\theta(\mathbf{x}, \mathbf{z})$ that are distributed according to the predictive model distribution $p_\theta(\mathbf{y}|\mathbf{x})$. So, the expected value and variance of the predictive distribution at an untested point $\mathbf{x}_*$ are computed as.

$$\mu_\mathbf{y}(\mathbf{x}_*) = E_{p_\theta}[\mathbf{y}|\mathbf{x}_*, \mathbf{z}] \approx \frac{1}{N_s}\sum_{i=1}^{N_s}[f_\theta(\mathbf{x}_*, \mathbf{z}_i)] \tag{7}$$

$$\sigma_\mathbf{y}^2(\mathbf{x}_*) = Var_{p_\theta}[\mathbf{y}|\mathbf{x}_*, \mathbf{z}] \approx \frac{1}{N_s}\sum_{i=1}^{N_s}\left[f_\theta(\mathbf{x}_*, \mathbf{z}_i) - \mu_\mathbf{y}(\mathbf{x}_*)\right]^2, \tag{8}$$

where $\mathbf{z}_i \sim p(\mathbf{z})$, $i = 1, \ldots, N_s$, and N_s corresponds to the total number of Monte Carlo samples.

3 Results

Now, we present a demonstration to showcase the applicability of the proposed machine learning models. More specifically, we construct machine learning models to characterize the diesel/biodiesel surrogate fuels densities dependent on the chemical compositions, temperature, and pressure. Moreover, it is worth mention here that we consider as the input that characterizes the chemical compositions the number of atoms of carbon in the fuel molecule and it is considered a continuous variable.

Table 1. Accuracy metrics of machine learning models to predict the diesel densities.

(a) Gaussian Process training accuracy.

Train data	L_{2-MRE}	**R^2-score**
10 %	6.2805×10^{-2}	0.8538
50 %	4.7438×10^{-2}	0.9976
80 %	2.7272×10^{-2}	0.9991

(b) Generative model training accuracy.

Train data	L_{2-MRE}	**R^2-score**
10 %	4.9316×10^{-2}	0.9359
50 %	2.8989×10^{-3}	0.9983
80 %	2.1409×10^{-3}	0.9990

We consider as diesel and biodiesel surrogate fuels single-component alkanes C_nH_{2n+2} and fatty acid methyl esters (FAME - $C_nH_{2n}O_2$), respectively. These compounds have been widely used as surrogate models because of the availability of validated chemical mechanisms and experimental measurements. More specifically, the diesel surrogates fuels considered are n-octane,n-nonane, n-decane, n-dodecane, and n-hexadecane, and the biodiesel surrogate fuels are n-hexanoate, n-decanoate, n-myristate, n-palmitate and n-stearate. Also, the densities for the training dataset is computed using All-atom Optimised Potentials for Liquid Simulations (OPLS-AA) force field in equilibrium molecular dynamics (EMD) simulation with conditions ranging from high-pressure nozzle condition to supercritical chamber environment. Moreover, 80% of the dataset is selected randomly to training the ML models, and the remaining 20% is used to test. Moreover, the training data set is shared in three subsets with 10%, 50%, and 80% of data available to train the models. The aim here is to evaluate the impacts of construct the machine learning models in a small data regime.

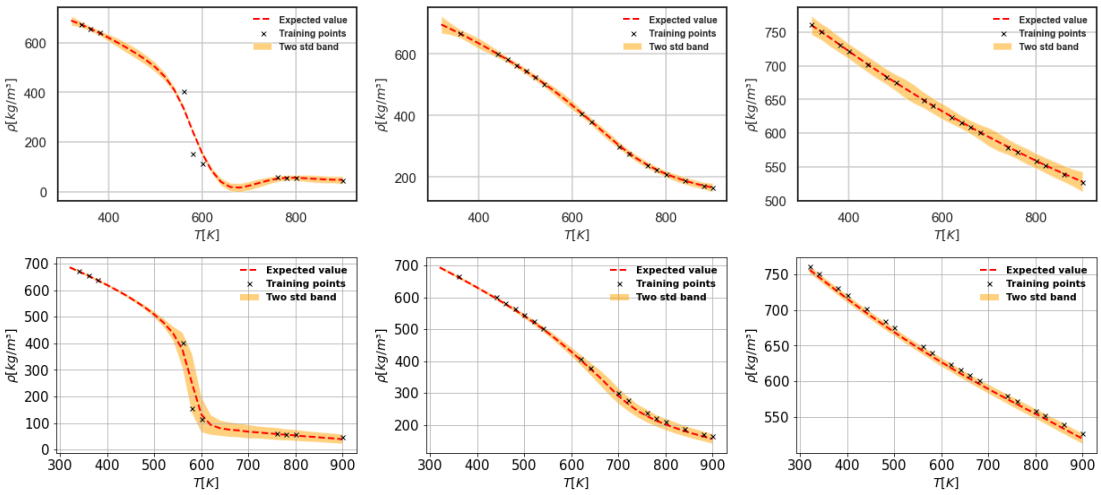
Here, accuracy is measured using the distance between the expected values predicted with the ML models and the predictions computed with the MD simulations. To evaluate the surrogate model, we check the accuracy computing the L_2 mean relative error (L_{2-MRE}) [18]. Also, we compute the coefficient of determination (R^2-scire) metric [19]. Also, we train the GP regression model of Eq. 1 via maximizing the marginal log-likelihood of Eq. 3 using the Mattern3/2 kernel function. The algorithm was implemented in GPy: Gaussian Process (GP) framework written in python [20]. Furthermore, the conditional generative model is constructed using fully connected feed-forward architectures for the encoder and generator networks with 4 hidden layers and 100 neurons per layer, while the discriminator architecture has 2 hidden layers with 100 neurons per layer. All activation uses a hyperbolic tangent non-linearity. The models are trained for 50,000 stochastic gradient descent steps using the Adam optimizer [21] with a learning rate of 10^{-4}, while fixing a two-to-one ratio for the discriminator versus generator updates. The proposed model was implemented in TensorFlow v2.1.0 [22], and computations were performed in single precision arithmetic on a single NVIDIA GeForce RTX 2060 GPU card.

3.1 ML results for a diesel surrogate fuel

Tables 1a and 1b show the coefficient of determination (R^2-score) and L_2 mean relative error for GP and probabilistic conditional generative models, respectively. Here, the accuracy metrics are computed with the test samples. We observe that the accuracies of the machine learning models are not satisfactory in the small training data scenario with 10% of training data. For a more optimistic condition, with 50% of training data, we observe that the surrogate models return good predictions with R^2-score higher than 0.99. Finally, with 100% of the training data, we can see that the surrogate models return excellent predictions with R^2-score very near 1.0 and mean relative errors lower than 0.03%.

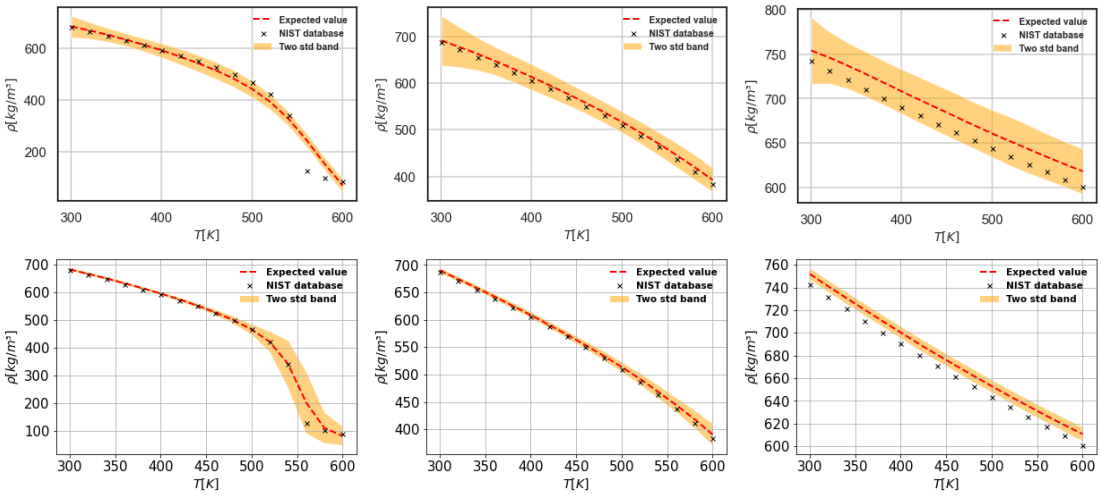
As a further illustration of the performance of such approaches to predict the physicochemical properties, we plot the octane density along with temperature for the ML models trained with 50% of the dataset randomly chosen, since this training scenario returns the best relation cost-accuracy. Figure 1 shows the octane density predictions along with the temperature at the pressures equal to 3, 10, and 100 MPa. We can observe that at 3MPa the GP model predicts well the density but near the critical points (transcritical conditions) return bad predictions. Also, we note that the probabilistic conditional generative model returns robust predictions with uncertainties bounds that capture the density even at transcritical conditions. Here, the predictive uncertainty of the proposed approaches reflects the uncertainty due to the inherent epistemic uncertainty in the GP and neural network approximations. We can also note that at 10 and 100 MPa the ML models return robust predictions with uncertainties bounds able to capture with accuracy the fuel density.

Also, we test the ability of the ML models to perform in an extrapolation scenario. Here, we compare the machine learning predictions for heptane (untested composition) density against them provide by the National Institute of Standards and Technology (NIST) database. Figure 2 shows that at 3 MPa and liquid conditions the ML model returns robust predictions of the heptane density with small uncertainties. However, at supercritical



(a) 3 MPa (b) 10 MPa (c) 100 MPa

Figure 1. Octane predictions with the GP (top) and probabilistic conditional generative models (bottom) at the pressures 3, 10, and 100 MPa.



(a) 3 MPa (b) 10 MPa (c) 100 MPa

Figure 2. Heptane predictions with the GP machine learning model (top) and conditional generative machine learning model (bottom) at the pressures 3, 10, and 100 MPa.

conditions, the GP model returns density predictions far from satisfactory. Also, we note that the generative model has uncertainty bounds able to capture the thermophysical property. The L_2 mean relative error between the NIST dataset and the expected values predicted by the GP and conditional generative models are 7.1697×10^{-2} and 2.0838×10^{-2}, respectively. We can also note that at higher pressure the learning models present better predictions with the GP model showing larger uncertainties bounds.

3.2 ML results for a biodiesel surrogate fuel

Now, we depict the performance of ML models to leverage the predictions of biodiesel surrogate fuels densities. We observe that the accuracies of the machine learning models are satisfactory even in the small training data scenario with 10% of the dataset, as shown in Tables 2a and 2b. The models return good predictions with R^2-score higher than 0.98 and mean relative errors lower than 0.8%. Moreover, with 80% of dataset for training, we can see that the GP model return excellent predictions with R^2-score very near 1.0 and 4.9196×10^{-9} mean relative error.

Similarly to section 3.1, we plot the decanoate densities along with temperature for the ML models trained with 50% of the dataset randomly chosen. Figure 3 shows the decanoate density predictions along with the temperature at the pressures equal to 4, 10, and 100 MPa. We can observe that the GP model returns robust predictions

Table 2. Accuracy metrics of machine learning models to predict the biodiesel densities.

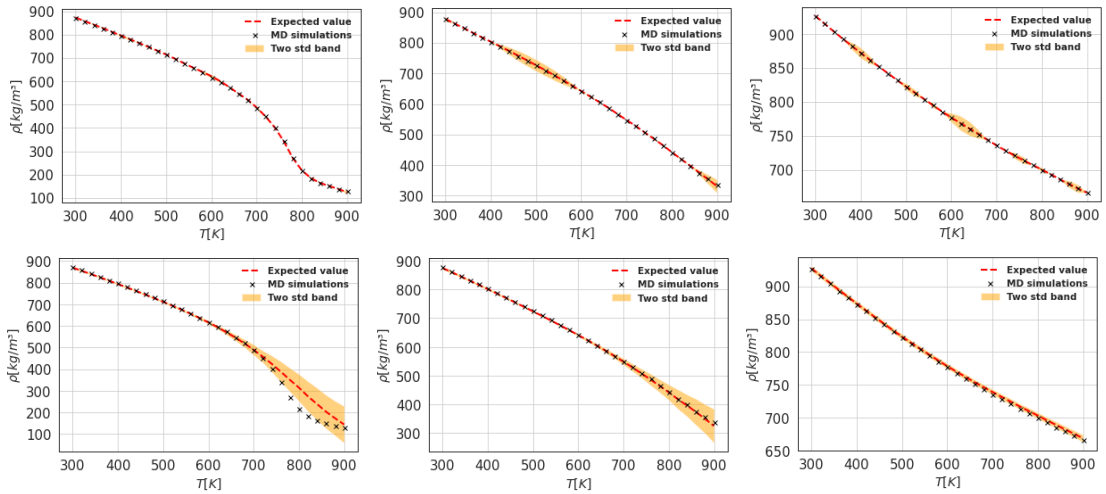
(a) Gaussian Process training accuracy

Train data	L_{2-MRE}	R^2-score
10 %	1.9056×10^{-2}	0.9888
50 %	7.3092×10^{-1}	0.9994
80 %	4.9196×10^{-9}	0.9999

(b) Generative model training accuracy

Train data	L_{2-MRE}	R^2-score
10 %	1.5836×10^{-2}	0.9884
50 %	7.1665×10^{-3}	0.9917
80 %	1.2659×10^{-2}	0.9877

of the density even at transcritical conditions. Also, we note that the probabilistic conditional generative model predicts well with uncertainties bounds that capture the density except for a few points under transcritical conditions. Here, the predictive uncertainty of the models are associated to the inherent epistemic uncertainty in the GP and neural network approximations.



(a) 4 MPa (b) 10 MPa (c) 100 MPa

Figure 3. Decanoate predictions with the GP machine learning model (top) and conditional generative machine learning model (bottom) at the pressures 4, 10, and 100 MPa.

4 Conclusions

We propose machine learning models to leverage Molecular Dynamics simulations for obtaining physicochemical properties, like density or heat diffusion coefficient. The ML models have been revealed powerful tools to predict the fuel properties from the chemical compositions of the fuel mixture and/or chemical structures of the fuel molecules, as demonstrated by our numerical examples. Also, such ML models can be employed efficiently in the context of UQ many-query tasks.

We place our contribution in the emerging area of physics-aware machine learning, where the final model, in many different ways, blends two main components: availability of experimental data and/or often expensive computational models relying on first principles and phenomenological closure equations, and machine learning data-driven models. Such combination allows describing physicochemical properties over a wide range of flow conditions at relatively low cost, and also offers a broad spectrum of opportunities to enhance CFD codes.

Acknowledgements. The research leading to these results had received funding from the Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP) through Programa de Recursos Humanos (PRH) under the PRH 8 - Mechanical Engineering for the Efficient Use of Biofuels, grant agreement numbers F0A5.EDDE.B5C0.3BCB and 2B61.4F5C.A83B.A713.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the

property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] W. J. Pitz and C. J. Mueller. Recent progress in the development of diesel surrogate fuels. *Progress in Energy and Combustion Science*, vol. 37, n. 3, pp. 330–350, 2011.
[2] J. Y. Lai, K. C. Lin, and A. Violi. Biodiesel combustion: Advances in chemical kinetic modeling. *Progress in Energy and Combustion Science*, vol. 37, n. 1, pp. 1–14, 2011.
[3] Y. Shen, Y.-B. Liu, and B.-Y. Cao. C4+ surrogate models for thermophysical properties of aviation kerosene rp-3 at supercritical pressures. *Energy & Fuels*, vol. 35, n. 9, pp. 7858–7865, 2021.
[4] A. Karlsson. *Modeling Auto-Ignition, Flame Propagation and Combustion in Non-Stationary Turbulent Sprays*. PhD thesis, Chalmers University of Technology, 1995.
[5] C. Chen and X. Jiang. Transport property prediction and inhomogeneity analysis of supercritical n-dodecane by molecular dynamics simulation. *Fuel*, vol. 244, pp. 48–60, 2019.
[6] M. Razi, A. Narayan, R. Kirby, and D. Bedrov. Fast predictive models based on multi-fidelity sampling of properties in molecular dynamics simulations. *Computational Materials Science*, vol. 152, pp. 125–133, 2018.
[7] W. Xing, A. Shah, P. Wang, S. Zhe, Q. Fu, and R. Kirby. Residual gaussian process: A tractable nonparametric bayesian emulator for multi-fidelity simulations. *Applied Mathematical Modelling*, vol. 97, pp. 36–56, 2021.
[8] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
[9] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, n. 2198, pp. 20160751, 2017.
[10] G. Su, L. Peng, and L. Hu. A gaussian process-based dynamic surrogate model for complex engineering structural reliability analysis. *Structural Safety*, vol. 68, pp. 97–109, 2017.
[11] T. Chen, J. Morris, and E. Martin. Gaussian process regression for multivariate spectroscopic calibration. *Chemometrics and Intelligent Laboratory Systems*, vol. 87, n. 1, pp. 59–71. Selected papers presented at the Conferentia Chemometrica 2005 Hajdúszoboszló, Hungary 28-31 August 2005, 2007.
[12] J. Yuan, K. Wang, T. Yu, and M. Fang. Reliable multi-objective optimization of high-speed wedm process based on gaussian process regression. *International Journal of Machine Tools and Manufacture*, vol. 48, n. 1, pp. 47–60, 2008.
[13] G. M. Guerra, R. Freitas, and F. A. Rochinha. Constructing accurate phenomenological surrogate for fluid structure interaction models. In K. L. Cavalca and H. I. Weber, eds, *Proceedings of the 10th International Conference on Rotor Dynamics – IFToMM*, pp. 295–305, Cham. Springer International Publishing, 2019.
[14] Y. Yang and P. Perdikaris. Adversarial uncertainty quantification in physics-informed neural networks. *Journal of Computational Physics*, vol. 394, pp. 136–152, 2019a.
[15] Y. Yang and P. Perdikaris. Conditional deep surrogate models for stochastic, high-dimensional, and multi-fidelity systems. *Computational Mechanics*, vol. 64, pp. 417–434, 2019b.
[16] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2014.
[17] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014.
[18] R. S. M. Freitas, C. H. S. Barbosa, G. M. Guerra, A. L. G. A. Coutinho, and F. A. Rochinha. An encoder-decoder deep surrogate for reverse time migration in seismic imaging under uncertainty. *Comput Geosci*, vol. 25, pp. 1229–1250, 2021.
[19] S. Weisberg. *Applied Linear Regression*. John Wiley & Sons, Inc., 2005.
[20] GPy. GPy: A gaussian process framework in python. `http://github.com/SheffieldML/GPy`, since 2012.
[21] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2017.
[22] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org, 2015.