AIAA SciTech Forum
23-27 January 2023, National Harbor, MD & Online
AIAA SCITECH 2023 Forum

10.2514/6.2023-1283

Physics-integrated Segmented Gaussian Process (SegGP) learning for cost-efficient training of diesel engine control system with low cetane numbers

Sai Ranjeet Narayanan*, Zongxuan Sun†, and Suo Yang‡
Department of Mechanical Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA

Yi Ji§ and Simon Mak¶
Department of Statistical Science, Duke University, Durham, NC 27708, USA

Harsh Darshan Sapra‖ and Sage Kokjohn**
Department of Mechanical Engineering, University of Wisconsin Madison, Madison, WI 53706, USA

Kenneth Kim†† and Chol-Bum M. Kweon‡‡
Combat Capabilities Development Command Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA

Control model training is an essential step towards the development of an engine controls system. A robust controls strategy is required for engines to perform reliably and optimally under challenging conditions, such as using low cetane number fuels (vital to achieving a single fuel concept). Developing such a control strategy through physical experiments, however, can be very costly due to issues such as unexpected engine failure and manufacturing delays. One approach is to rely solely on CFD simulations for control model training, which can be accurate but places significant burden on computing resources to explore the desired control design space. Another approach is via purely data-driven machine learning models, but the training data needed to achieve desirable accuracy can also be prohibitively expensive to generate. To address this, we develop a novel physics-integrated Segmented Gaussian Process (SegGP) model, which integrates fundamental physics on the pressure curve within a flexible probabilistic learning framework. This integration of physics allows for accurate predictive modeling of pressure, heat release and thus control using limited training data, which greatly reduces computational burden. We demonstrate the effectiveness of this approach for quickening the control development training of diesel engines.

I. Nomenclature

$AHRR$ = Apparent Heat Release Rate
ARL = Army Research Laboratory
$ATDC$ = After Top Dead Center
CA = Crank Angle
$CA50$ = Crank Angle at 50% Total heat release
$CA90$ = Crank Angle at 90% Total heat release
CR = Connecting Rod
CN = Cetane Number

---

*Ph.D. Student, Student Member AIAA
†Professor
‡Richard and Barbara Nelson Assistant Professor, suo-yang@umn.edu (Corresponding Author), Senior Member AIAA
§Ph.D. Student
¶Assistant Professor
‖Post-Doctoral Researcher
**Phil and Jean Myers Associate Professor
††Mechanical Engineer, Weapons and Materials Research Directorate, Member AIAA
‡‡Program Manager, Computational and Information Sciences Directorate, Member AIAA

IA = Ignition Assistant
IHR = Integrated Heat Release
MIT = Main Injection Timing
ROM = Reduced Order Model
SOC = Start of Combustion
UAS = Unmanned Aircraft System

II. Introduction

Achieving a single fuel concept (SFC) for UAS engines necessitates the usage of fuels with low cetane number (CN), since it simplifies logistics [1]. Previous research indicates that it is possible to utilize fuels with low CN through variable energy assist with a glow plug [2]. There is a need to design and implement control systems to overcome the challenges associated with low CN fuels. Real time control hardware and software development is necessary to maintain robust and efficient UAS engine operation due to fuel property variation and change of ambient condition [3]. Current state control models/systems are typically trained by either experimental data and/or computational fluid dynamics (CFD) data. Experimental data acquisition may face challenges such as data inaccuracy, delays in manufacturing engine parts and other technical issues. On the other hand, CFD simulations, while yielding detailed engine relevant information, are computationally expensive and require significant computation hours to simulate even a single engine cycle.

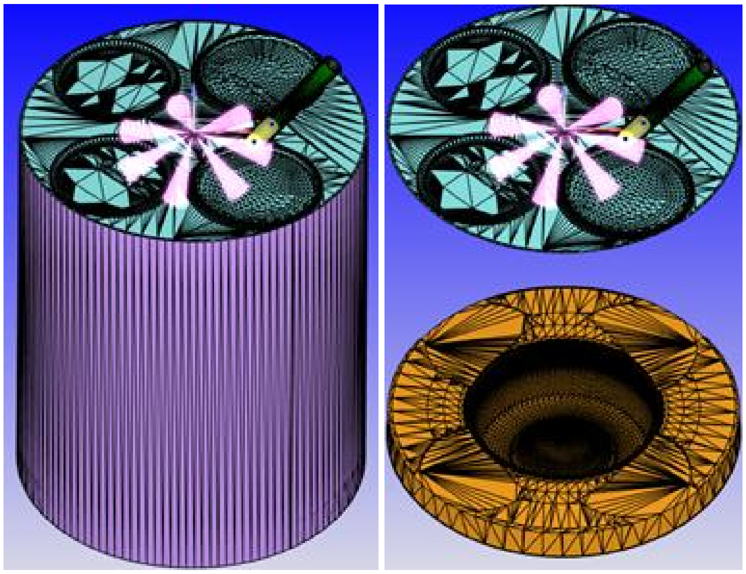
Machine Learning (ML) has widely been used in favor of CFD for developing models for internal combustion engine (ICE) applications [4–13] which include optimization, as well as for engine controls development [14–17]. Kodavasal et al. [5] used an ML random forest model to study cycle to cycle variations in spark ignition engines. Moiz et al. [18] developed a ML-Genetic Algorithm (ML-GA) for optimization of a gasoline compression ignition (GCI) engine operating with a low-octane gasoline-like fuel. Kavuri et al. [4] used a Gaussian process regression (GPR) model combined with genetic algorithm (GA) as a substitute for CFD to shorten the computational time and improving engine optimization for a compression ignition engine. They observed a 62% reduction in computational time when using the GPR-GA model when compared to pure CFD simulations using the same resources. Ravindran et al. [11] combined CFD with a GPR model for optimization of a gasoline direct injection spark ignition (DISI) engine at cold start conditions. Badra et al. [10] used a Machine Learning-Grid Gradient Ascent model for optimization of a light duty GCI engine in combination with CFD. Owoyele et al. [8] developed a surrogate scheme based ML optimization model (ActivO) for a compression ignition (CI) engine and was found to yield faster convergence speed to GA based approaches. Bertram et al. [6] used a Support Vector Regression models with the Particle Swarm Optimization routine (SVM-PSO) for optimization of diesel engine calibration. Bin et al. [19] found GPR to perform better than SVM and artificial neural networks (ANN) in their study of predicting thermal comfort index as function of environmental parameters such as air temperature and humidity. Aran et al. [14] used a GPR model for the feed-forward part of a diesel engine air-path controller. Erikkson and Nielsen [16] list physics-based regression models of embedded turbocharger maps in model-based feed-forward controller for air-path control system. Sastry et al. [17] compared the performance of supervised learning methods such as ANN and SVM for engine control development. Aliramezani et al. [15] have performed a comprehensive review of the usage of ML techniques for ICE modeling, optimization, diagnosis, and control challenges.

We aim to extend the above body of literature to develop a data-driven surrogate model for efficiently predicting the desired quantities for engine control input, such as the CA50 (crank angle at 50% of total fuel heat release) from the pressure and heat release curves. Given the expensive cost of CFD simulations, one is faced with limited training data for this surrogate model. A promising solution is the integration of prior system physics knowledge for surrogate modeling, which has seen recent success in improving surrogate modeling accuracy and interpretability with limited data [20–23]. In line with this literature, we will develop a physics-based Segmented Gaussian Process (SegGP) ROM, which integrates a physics-guided curve segmentation within the Gaussian process predictive model. The SegGP can serve as an efficient substitute for the expensive CFD simulations, thus enabling timely data generation for control model training.

The remaining of this paper proceeds as follows. Section III surveys the engine geometry and parameters used for our study, as well as the corresponding physical models used in the validated CFD case setup. The SegGP ROM methodology is then highlighted in detail, showing how the surrogate model stitches together the different predicted "segments" to produce curve predictions. Section IV reports the SegGP prediction results for the pressure, apparent heat release, and integrated heat release curves as a function of crank angle. The ROM predictions are compared and validated with the CFD and experimental results for the same cases. Here, the engine CA50 is chosen as the comparison metric for the validation. Section V concludes the paper by summarizing the research efforts and findings in this work.

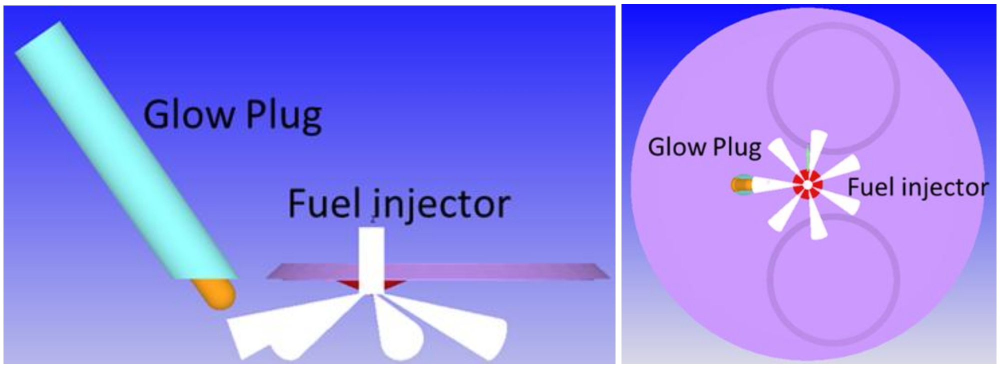
III. Methodology

The engine considered is a GM 2.0L Energy-Assisted Compression Ignition engine (EACI) containing an ignition assistant (IA, or glow plug) at the engine head [24, 25]. The engine model used in the CFD simulations which contains the IA and fuel injector is shown in Fig. 1 and Fig. 2. Simulations were conducted on Converge CFD 3.0 software [26]. These simulations were closed cycle, varying from -130 to +130 CA. Table 1 shows the model details and Table 2 shows the engine parameters and operating conditions for the cases used in this work. The two control relevant parameters varied in this study were the Cetane Number (CN) and Main Injection Timing (MIT). The fuel used in the simulations is a multi-component blend [27], and consists of decane, toluene and iso-octane as the fuel constituents species. The CN is varied by adjusting the mass fractions of these constituents.



(a) Engine liner (purple) shown **(b) Engine liner hidden**

Fig. 1 Metal Engine Model for CFD simulation. Fuel injection happens via seven nozzles, as seen from the pink spray cones. The glow plug is highlighted in the yellow cylindrical portion.



(a) Side view **(b) Bottom view**

Fig. 2 The IA (Glow plug) and Fuel injector.

Table 1 Engine model details.

PHYSICAL MODEL	NAME
Combustion	SAGE Chemical Kinetics Solver + Three point PDF method [28]
Chemical mechanism	Multicompoment (55 sp., 146 reactions) [27]
Turbulence model	RANS (RNG $k-\epsilon$) [29]
Droplet Evaporation model	Frossling model [30]
Spray Breakup model	KH-RT model [31]
Collision model	No-Time-Counter (NTC) collision model [32]

Table 2 Engine parameter details.

PARAMETER	VALUES [UNITS]
Engine speed	1200 [rpm]
Bore	82.3 [mm]
Stroke	76.2 [mm]
CR Length	147.7 [mm]
Nozzle holes	7
Nozzle diameter	141 [μm]
IA Temp/Power	1400 [K]/70 [Watt]
Inj. Duration	4.75 [CA]/0.66 [ms]
Inj. Mass	12 [mg]
IVC Temperature	320 [K]
IVC Pressure	88.87 [kPa]
Cone angle	18 [deg]
Tilt angle	70 [deg]

Our ROM modeling framework is inspired by the Treed Gaussian process (Treed GP) learning model [33], which builds a Bayesian nonparametric hierarchical model that first learns important partitions over the parameter space, then trains separate Gaussian process (GP) models within each region. Treed GPs have been successfully for many applications where the underlying response surface has non-stationary behavior over the parameter space. However, the partitions in the Treed GP are typically performed in a purely data-driven manner, which ignores the underlying physics implicit in the diesel engine system (described next). The integration of prior physical knowledge for reduced-order modeling has shown to be quite effective at improving predictive performance and interpretability (see Refs. [20, 22, 23]), and we aim to build a model which performs this physics integration for predicting pressure and IHR curves.

In our study, there are two important physical quantities: SOC (Start of Combustion) and CA90 (crank angle at 90% total heat release), which provide a useful physics-driven segmentation of the IHR and pressure curves. This is visualized in the left plot of Fig. 3. In particular, all pressure curves typically experience a slight incline in the crank angle range of [-50CA, SOC] due to natural compression of the engine, then a sudden sharp incline in the range of [SOC, CA90] due to fuel ignition, then finally a steady decline in the range of [CA90, +50CA] as the ignition ends and the engine expands. To integrate this physics, we propose the following Segmented Gaussian (SegGP) process modeling framework, which makes use of SOC and CA90 information for physics-informed partitions in Gaussian process learning. The SegGP model consists of four steps: (1) the pressure curves are partitioned using the SOC and CA90 cut-off points for training cases; (2) using this training data, separate Gaussian process models are fitted on each curve partition and used for prediction on validation cases; (3) using the same training data, separate Gaussian process models are fitted on the SOC and CA90 cut-offs and used for predicting the cut-offs for validation cases; (4) the curve and cut-off predictions from Steps (2) and (3) are then stitched together for the final prediction of the pressure curve. The heat release curves are then calculated from the pressure and volume data [34].

We provide further details on the above modeling framework. Step (1) requires that we partition the training pressure

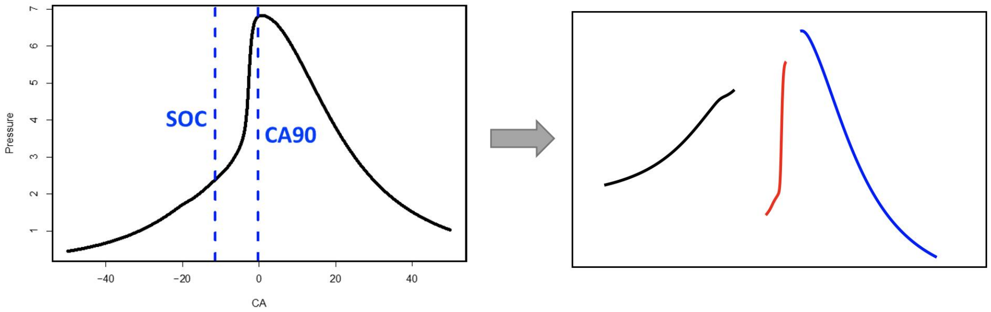


Fig. 3 Partition training pressure curves with SOC and CA90.

curves using available SOC and CA90 cut-off information, as shown in Fig. 3. Let us denote $y_{\mathrm{I}}(\mathbf{x}, t)$ as the pressure curve for $t \in [-50\mathrm{CA}, \mathrm{SOC}]$, $y_{\mathrm{II}}(\mathbf{x}, t)$ as the pressure curve for $t \in [\mathrm{SOC}, \mathrm{CA90}]$, and $y_{\mathrm{III}}(\mathbf{x}, t)$ as the pressure curve for $t \in [\mathrm{CA90}, +50\mathrm{CA}]$, where $\mathbf{x} = \{\mathrm{MIT}, \mathrm{CN}\}$ and t corresponds to crank angle.

Next, we train two separate Gaussian process models on the first and third segmented pressure curves (y_{I} and y_{III}):

$$y_{\mathrm{I}}(\mathbf{x}, t) \sim \mathrm{GP}(0, k_{\mathrm{I}}([\mathbf{x}, t], [\mathbf{x}', t'])), \tag{1}$$

$$y_{\mathrm{III}}(\mathbf{x}, t) \sim \mathrm{GP}(0, k_{\mathrm{III}}([\mathbf{x}, t], [\mathbf{x}', t'])). \tag{2}$$

In practice, we find that using the sum of a Matérn-3/2 kernel [35] and a linear kernel on t yields best predictive performance. We thus make use of the following composite covariance kernels for $k_{\mathrm{I}}(\cdot,\cdot)$ and $k_{\mathrm{III}}(\cdot,\cdot)$:

$$\begin{aligned} k([\mathbf{x}, t], [\mathbf{x}', t']) &= k_{\text{Matérn-3/2}}([\mathbf{x}, t], [\mathbf{x}', t']) + k_{\text{Linear}}(t, t'), \\ k_{\text{Matérn-3/2}}([\mathbf{x}, t], [\mathbf{x}', t']) &= \sigma^2_{\text{Matérn}} \cdot \left[\frac{1}{\Gamma(1.5) 2^{0.5}} \left(\frac{\sqrt{3}}{l} d([\mathbf{x}, t], [\mathbf{x}', t']) \right)^{1.5} \cdot K_{1.5} \left(\frac{\sqrt{3}}{l} d([\mathbf{x}, t], [\mathbf{x}', t']) \right) \right], \\ k_{\text{Linear}}(t, t') &= \sigma^2_{\text{Linear}} \cdot t \cdot t', \end{aligned}$$

where $\sigma^2_{\text{Matérn}}, \sigma^2_{\text{Linear}}$ are variance parameters, l is a length-scale parameter, $d(\cdot,\cdot)$ is the Euclidean distance, $K_{1.5}(\cdot)$ is the modified Bessel function of order 1.5, and $\Gamma(\cdot)$ is the gamma function. Note that, although $k_{\mathrm{I}}(\cdot,\cdot)$ and $k_{\mathrm{III}}(\cdot,\cdot)$ share the same form of covariance kernel $k(\cdot,\cdot)$, their variance and length-scale parameters are estimated on different segments of pressure curve data. For parameter estimation, we will make use of the widely-used maximum likelihood method (see Ref. [36]).

After the above Gaussian process models for y_{I} and y_{III} are trained (i.e., model parameters are estimated from training data), we predict $\hat{y}_{\mathrm{I}}$ and $\hat{y}_{\mathrm{III}}$ on the validation cases. With the kernel functions specified and trained, the prediction methodology for the SegGP follows directly from that for the standard GP model; see Ref. [36] for further details on predictive equations.

The second segment y_{II} with $t \in [\mathrm{SOC}, \mathrm{CA90}]$ is treated slightly differently from the first and third segments. As the values of SOC and CA90 vary for different values of control variables $\mathbf{x} = \{\mathrm{MIT}, \mathrm{CN}\}$, y_{II} can have large variation for different control variables at angle t. This makes it difficult for standard Gaussian processes to learn the relationship between the response of interest y_{II} and input parameters $(\mathbf{x}, t)$. One solution which we adopt is to standardize both the response variable y_{II} and the crank angle t prior to training the SegGP model, as shown in Fig. 4. The standardization is performed using the following formula:

$$t_{\mathrm{std}} = \frac{t - \mathrm{SOC}}{\mathrm{CA90} - \mathrm{SOC}}, \quad y_{\mathrm{II,std}} = \frac{y_{\mathrm{II}} - y_{\mathrm{SOC}}}{y_{\mathrm{CA90}} - y_{\mathrm{SOC}}}. \tag{3}$$

Similar to $\hat{y}_{\mathrm{I}}$ and $\hat{y}_{\mathrm{III}}$, we then predict standardized pressure curves $\hat{y}_{\mathrm{II,std}}$ on the standardized t for the validation cases.

With these three predicted curves, the last step is to then stitch the predictions $\hat{y}_{\mathrm{I}}, \hat{y}_{\mathrm{II,std}}, \hat{y}_{\mathrm{III}}$ together for a continuous and smooth pressure curve prediction for each validation case. To do so, we need an additional learning model which can predict the required cut-off points (SOC, CA90) for each validation case. We thus make use of two independent stationary Gaussian process models on SOC and CA90, using the control variables $\mathbf{x}$ as input parameters:

$$g_{\mathrm{SOC}}(\mathbf{x}) \sim \mathrm{GP}(0, k_{\mathrm{SOC}}(\mathbf{x}, \mathbf{x}')), \quad g_{\mathrm{CA90}}(\mathbf{x}) \sim \mathrm{GP}(0, k_{\mathrm{CA90}}(\mathbf{x}, \mathbf{x}')). \tag{4}$$

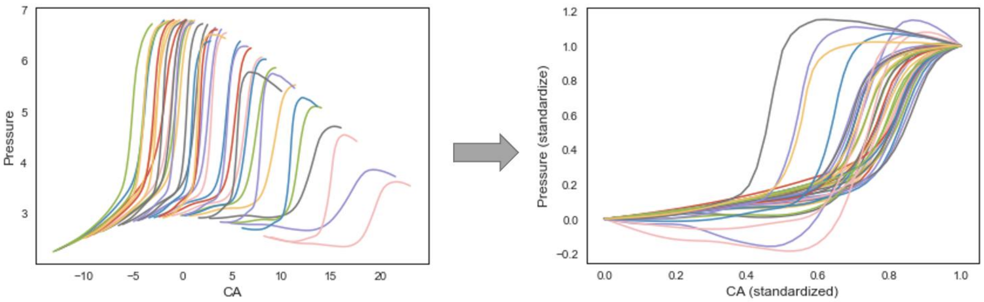


Fig. 4 Standardize both crank angle and pressure for y_{II} modeling. Each colored line represents a different simulation training case.

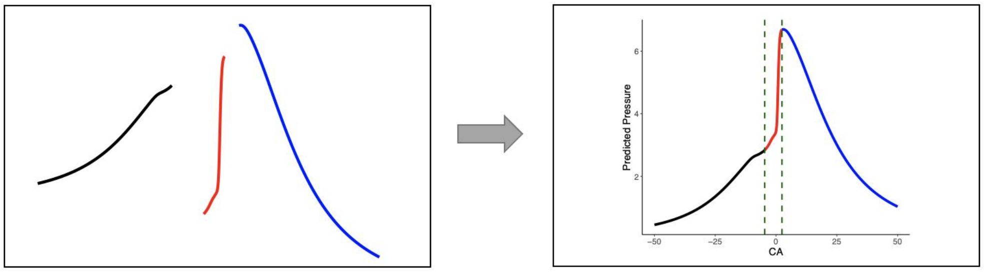


Fig. 5 Stitch $\hat{y}_{\mathrm{I}}$, $\hat{y}_{\mathrm{II}}$, $\hat{y}_{\mathrm{III}}$ together for final predicted pressure curve.

Here, k_{SOC} and k_{CA90} are squared exponential covariance kernel $k(\cdot,\cdot)$ of the form:

$$k(\mathbf{x},\mathbf{x}') = \sigma^2 \cdot \exp\left\{-\sum_{j=1}^{2} \frac{(x_j - x'_j)^2}{l_j^2}\right\},$$

with variance parameter σ^2 and length-scale parameters l_1, l_2. These parameters are again estimated via the maximum likelihood method [36].

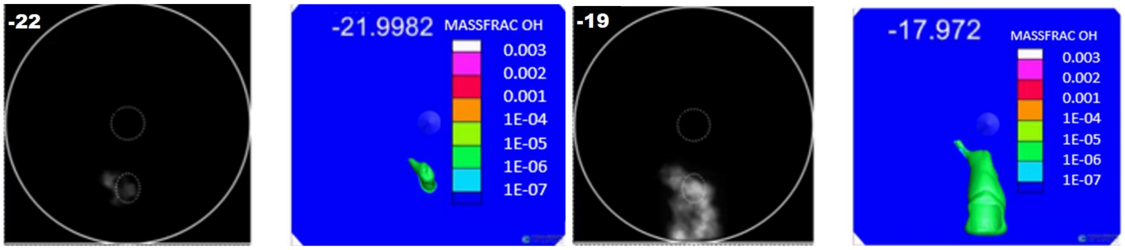
After the estimated curve cut-off points $\widehat{\mathrm{SOC}}$ and $\widehat{\mathrm{CA90}}$ are obtained for each validation case, we can then stitch the final predicted pressure curve as follows. The predicted curve is taken as $\hat{y}_{\mathrm{I}}$ from the crank angle range $[-50\mathrm{CA}, \widehat{\mathrm{SOC}}]$, and taken as $\hat{y}_{\mathrm{III}}$ from the range $[\widehat{\mathrm{CA90}}, +50\mathrm{CA}]$. For the second segment, we perform the following transformation to change back to the original scale:

$$\hat{y}_{\mathrm{II}}([\mathbf{x},t]) = \hat{y}_{\mathrm{II,std}}([\mathbf{x},t]) \cdot \left(\hat{y}_{\mathrm{III}}([\mathbf{x},\widehat{\mathrm{CA90}}] - \hat{y}_{\mathrm{I}}([\mathbf{x},\widehat{\mathrm{SOC}}]\right) + \hat{y}_{\mathrm{I}}([\mathbf{x},\widehat{\mathrm{SOC}}]). \tag{5}$$

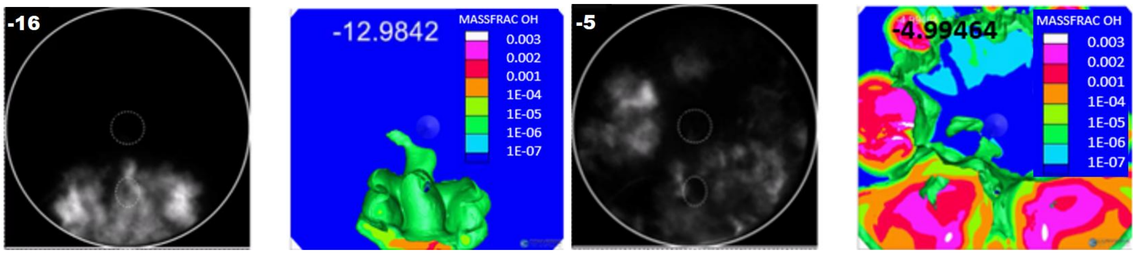
A visualization of this stitching process is given in Fig. 5.

IV. Results and Discussion

We now investigate the performance of the proposed SegGP surrogate model for our target diesel engine system. Figure 6 shows the OH emissions in the CFD simulation as the flame propagates (images captured via OH chemiluminescence) with CA in the actual engine (same GM 2.0 EACI single cylinder engine, but with a fitted piston window for optical access). It is noticed that the flame initiates near the IA (Fig. 6a), moves towards the piston bowl (Fig. 6b), spreads into two plumes (Fig. 6c) and then spreads further into the engine (Fig. 6d). The snapshots taken in the actual engine show similar behavior of the flame propagation at roughly the same CA. The CFD model was validated as such with the experiments at 1200 rpm (revolutions per minute), 1550 IA temperature, 3.63 CA injection duration, -24.75 MIT and 6.5 mg fuel injection mass. Further details about the optical engine can be obtained in previous studies [37, 38], and more CFD validation results can be found in the study conducted by Sapra et al [25].



(a) -22 CA ATDC (EXP left, CFD right) **(b) -18 CA ATDC (EXP left, CFD right)**



(c) -13 CA ATDC (EXP left, CFD right) **(d) -5 CA ATDC (EXP left, CFD right)**

Fig. 6 OH emissions (mass fraction) from CFD prediction, compared to the flame propagation in the actual engine experiment (images obtained through OH chemiluminescence), at different CA.

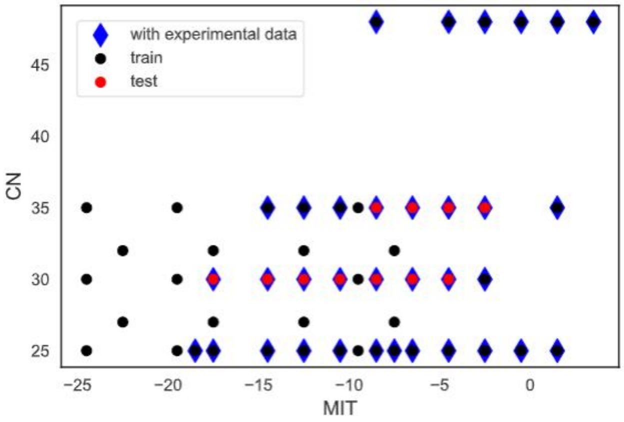
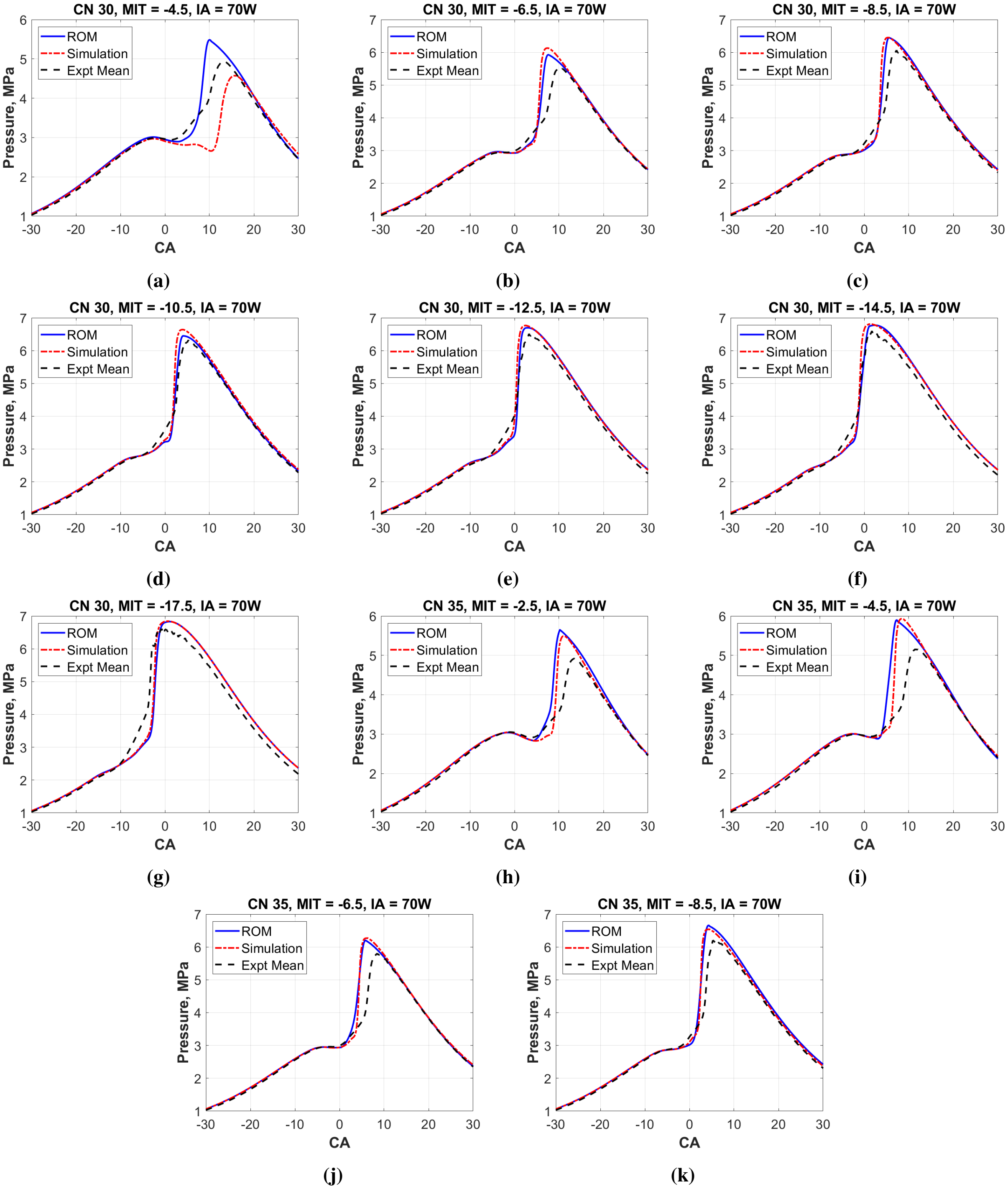


Fig. 7 The Cetane Number (25, 27, 30, 32, 35 and 48) and Main Injection Timing (varies) parametric space used for training and validating the ROM.

Figure 7 displays the parametric space for the ROM training and validation (testing). Blue diamond points have experimental data available for validation. Black dots are the training points (using CFD simulations). Red dots are the chosen validation points.

Figure 8 displays the comparisons of the mean engine in-cylinder pressure curves as a function of the crank angle, for the ROM, CFD simulations and experiments at chosen validation points (which were shown as red dots in Fig. 7). We see that the ROM predictions (blue solid curves) closely resemble the CFD simulations (red dotted curves) for most validation cases. First, we observe that our ROM predictions are almost identical to CFD simulations for regions of the pressure curves with shallow slopes, i.e., before SOC and after the peak of the pressure curves. This suggests that the proposed SegGP surrogate model can indeed accurately predict for the relationship between pressure and the three input parameters: CN, MIT and CA. Next, our ROM predictions can also capture the sharp increase of the pressure curves between SOC and CA90. This highlights the benefit of segmenting pressure curves into three physics-guided partitions and performing prediction on each segment; it allows us to model the trends of pressure curves locally. We note that the surrogate model curve predictions seem to differ most from the CFD simulations for the validation case with CN 30



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k)

Fig. 8 Mean Engine Pressure profile comparisons between the ROM, Simulations and Experiments at validation points.

and MIT -4.5 in Fig. 8. One reason might be that this validation case is close to the boundary of the design space (see Fig. 7), and it is known that GPR-based emulators may have trouble predicting at the design space boundaries due to extrapolation [36]. The CFD simulation curve also deviates more from the experimental curve in this case, which would impact the accuracy of the ROM training in this region. It should be noted that due to the limitations of experimental measurements, parameters such as T_{IVC} and P_{IVC} can only be sensibly estimated, and the output is highly sensitive to these parameters. The lower CN (25 and 30) and MIT closer to 0 CA conditions especially show large variance in the CFD results due to this sensitivity, implying that a more physics based approach is desired for estimating such parameters.

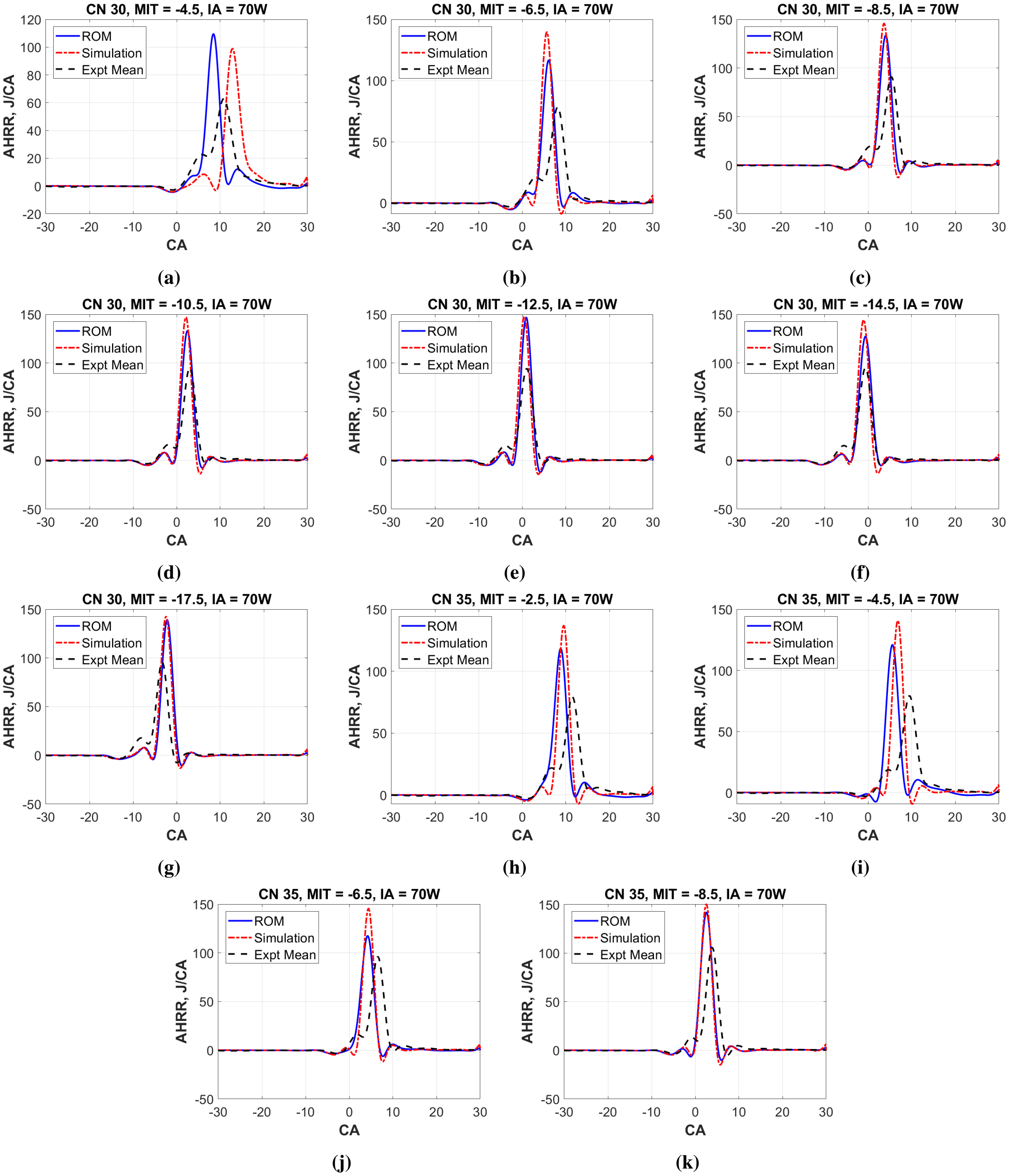
Similarly, Figure 9 compares the apparent heat release rate (AHRR) curves for the proposed ROM, CFD and Experiments. We note that the AHRR ROM predictions (blue solid curves) are derived from the pressure curves in Fig. 8. As before, the proposed ROM appears to well-capture the qualitative trends of the simulated AHRR curves from CFD (red dotted curves). The sharp gradient rise in AHRR due to start of ignition, the maximum heat release rate observed (the peak of the curve) as well as the duration of combustion (the width of the curve) all depend on the CN and MIT, and these appear to be predicted well by the SegGP (when compared to CFD). As before, the predicted AHRR curves appear quite accurate for most validation cases, with some discrepancies observed for the setting of CN 30 and MIT -4.5 (for reasons explained previously). Quantitatively, there are deviations in the peak AHRR values observed between experiments and CFD. This can be explained by the differences in measuring the pressure between CFD and experiments: The pressure transducer placed in the engine only measures the pressure in a small localized region around it. In the CFD simulations, the pressure measured is the spatially averaged "mean" engine pressure at the specified CA. Thus, this difference leads to the deviations in the pressure curve results (as seen in Fig. 8 and since the AHRR is calculated directly from the pressure, the deviations carry forward into these results as well. This is why we observe a difference in the peaks of the AHRR curves between CFD (hence ROM) and experiments.

Table 3 Absolute error in CA50 between ROM and Experiments for each validation case.

CASE	CN	MIT	CA50 ERROR (ABS)
a	30	-4.5	1.90
b	30	-6.5	1.90
c	30	-8.5	1.15
d	30	-10.5	0.50
e	30	-12.5	0.10
f	30	-14.5	0.25
g	30	-17.5	1.40
h	35	-2.5	2.70
i	35	-4.5	3.70
j	35	-6.5	2.20
k	35	-8.5	1.30

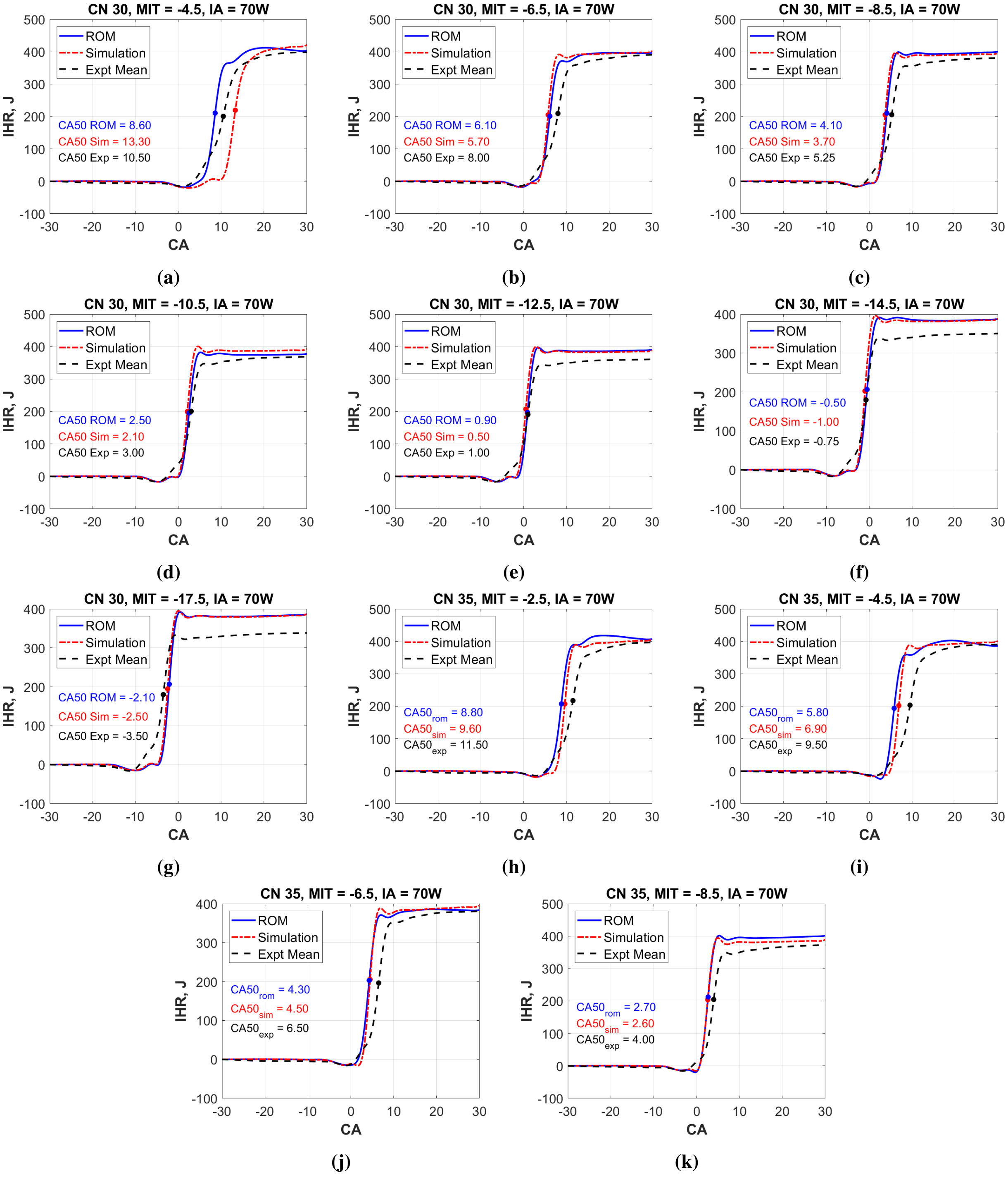
Figure 10 compares the integrated heat release (IHR) profiles between the three modes. These figures also have the calculated CA50 in-laid for the three modes, and Table 3 shows the absolute error in CA50 between the ROM and experiments based on the validation case. The CA50 is chosen as the variable of comparison since that is used in the study's engine controls unit to determine the current condition of the engine. Moreover, the CA50 is used to train our controls system's lookup tables (LUT) for feed-forward controls application. We can see a higher average absolute CA50 error for the CN 35 test points (average absolute CA50 error = 2.475) than the CN 30 test points (average absolute CA50 error = 1.028). This is because the neighboring training points for the CN 30 test points are closer and more clustered, having training data available at CN 25, 27, 30, 32 and 35 (Fig. 7). This gives a more accurate prediction by the ROM when trying to interpolate and predict the profiles at CN 30. On the other hand, the neighboring training points for the CN 35 test points are close by below and at CN 35, but are far apart at CN 48. This causes a less accurate interpolation by the ROM, hence leading to a larger error in CA50. As mentioned in the previous paragraph, the deviation in the pressure curves of CFD from experiments carries forward in to the AHRR curves, and by extension, the IHR curves as well. This is again evident by the differences in the CA50 values between CFD and experiments.

On the whole, the ROM provides good qualitative agreement of the important trends, as well as good quantitative



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Fig. 9 Engine Apparent Heat Release Rate (AHRR) profile comparisons between the ROM, Simulations and Experiments at validation points.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k)

Fig. 10 Engine Integrated Heat Release (IHR) profile comparisons between the ROM, Simulations and Experiments at validation points.

Table 4 Time required to finish a single run/case using the ROM and CFD.

METHOD	TIME REQUIRED
CFD simulation	18 to 24 hours
ROM prediction	1 to 2 minutes

agreements for some of the cases. It provides close agreements with CFD, and an improvement in the CFD model results would automatically lead to a better ROM prediction. There is a significant amount of time saved for generating the results using the ROM when compared to CFD. Table 4 compares the time required to run one simulation (one case) for the ROM and CFD. Note that approximately one hour is needed (once only) to train the ROM model. All of the ROM predictions were completed in a matter of minutes (~100 seconds per prediction), as compared to the hours (~80,000 seconds) required for even a single CFD simulation (run on an AMD Milan 7763 64-core processor). The ROM performs almost 800 times faster than CFD, while being just as accurate. Thus, it can serve as a viable alternative to CFD for fast generation of data for initial control system training.

V. Conclusion

A physics informed data driven ROM for predicting the temporal profile of engine variables (pressure, AHRR and IHR) required for controls systems training is presented. CFD simulations using a validated model were conducted to train the model, where the control input variable (MIT) and fuel CN were varied. Simulations were also conducted for validating the ROM predictions, and the results show good preliminary agreements, with a significant gain in computational time. Further testing of the ROM at other splits of training & validation points is needed to draw a comprehensive conclusion about the ROM's viability for data generation for control system training. Other glow plug powers and operating conditions can also be explored, and the ROM can be developed to be more robust and adaptable over a diverse range of conditions. Other control inputs can also be tested/varied for parametric study, such as the injection duration and injection fuel mass.

Acknowledgments

This work is jointly funded by the Army Research Laboratory (ARL) under Cooperative Agreement Number W911NF2020161, and by the National Science Foundation (NSF) under grant NSF DMS 221072. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. The authors would like to thank the Engine Research Center (ERC) at the University of Wisconsin for providing the experimental data for validation. The authors also gratefully acknowledge the Minnesota Supercomputing Institute (MSI) for the computational resources provided to conduct the CFD simulations. The authors would finally like to thank Convergent Science for providing CONVERGE 3.0 licenses and technical support for this work.

References

[1] Edwards, J. T., "Reference jet fuels for combustion testing," *55th AIAA aerospace sciences meeting*, 2017, p. 0146.

[2] Amezcua Cuellar, E. R., Rothamer, D., Kim, K., and Kweon, C.-B., "Optical Engine Study of Variable Energy Assisted Compression Ignition using a Glow Plug for Unmanned Aircraft Propulsion Systems," *AIAA Scitech 2020 Forum*, 2020, p. 2281.

[3] Sun, Z., and Zhu, G. G., *Design and control of automotive propulsion systems*, CRC press, 2014.

[4] Kavuri, C., and Kokjohn, S. L., "Exploring the potential of machine learning in reducing the computational time/expense and improving the reliability of engine optimization studies," *International Journal of Engine Research*, Vol. 21, No. 7, 2020, pp. 1251–1270.

[5] Kodavasal, J., Abdul Moiz, A., Ameen, M., and Som, S., "Using machine learning to analyze factors determining cycle-to-cycle variation in a spark-ignited gasoline engine," *Journal of Energy Resources Technology*, Vol. 140, No. 10, 2018.

[6] Bertram, A. M., and Kong, S.-C., "Computational optimization of a diesel engine calibration using a novel SVM-PSO method," Tech. rep., SAE Technical Paper, 2019.

[7] Probst, D. M., Raju, M., Senecal, P. K., Kodavasal, J., Pal, P., Som, S., Moiz, A. A., and Pei, Y., "Evaluating optimization strategies for engine simulations using machine learning emulators," *Journal of Engineering for Gas Turbines and Power*, Vol. 141, No. 9, 2019.

[8] Owoyele, O., and Pal, P., "A novel machine learning-based optimization algorithm (ActivO) for accelerating simulation-driven engine design," *Applied Energy*, Vol. 285, 2021, p. 116455.

[9] Zhang, Q., Ogren, R. M., and Kong, S.-C., "Application of improved artificial bee colony algorithm to the parameter optimization of a diesel engine with pilot fuel injections," *Journal of Engineering for Gas Turbines and Power*, Vol. 139, No. 11, 2017.

[10] Badra, J., Sim, J., Pei, Y., Viollet, Y., Pal, P., Futterer, C., Brenner, M., Som, S., Farooq, A., Chang, J., et al., "Combustion system optimization of a light-duty GCI engine using CFD and machine learning," Tech. rep., SAE Technical Paper, 2020.

[11] Ravindran, A. C., and Kokjohn, S. L., "Combining machine learning with 3D-CFD modeling for optimizing a DISI engine performance during cold-start," *Energy and AI*, Vol. 5, 2021, p. 100072.

[12] Zhang, H., Bavandla, K. C., Gao, X., Gao, J., Yi, P., and Yang, S., "Optimization and uncertainty quantification of spray break-up submodel with regularized multi-task neural nets," *AIAA Scitech 2020 Forum*, 2020, p. 0909.

[13] Dikeman, H. E., Zhang, H., and Yang, S., "Stiffness-Reduced Neural ODE Models for Data-Driven Reduced-Order Modeling of Combustion Chemical Kinetics," *AIAA SCITECH 2022 Forum*, 2022, p. 0226.

[14] Aran, V., and Unel, M., "Gaussian process regression feedforward controller for diesel engine airpath," *International Journal of Automotive Technology*, Vol. 19, No. 4, 2018, pp. 635–642.

[15] Aliramezani, M., Koch, C. R., and Shahbakhti, M., "Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions," *Progress in Energy and Combustion Science*, Vol. 88, 2022, p. 100967.

[16] Eriksson, L., and Nielsen, L., *Modeling and control of engines and drivelines*, John Wiley & Sons, 2014.

[17] VM, S. S., Garg, P., Silvas, E., and Willems, F., "Systematic Comparison of Supervised Learning Methods to Reduce Calibration Effort in Engine Control Development," *IFAC-PapersOnLine*, Vol. 55, No. 20, 2022, pp. 37–42.

[18] Moiz, A. A., Pal, P., Probst, D., Pei, Y., Zhang, Y., Som, S., and Kodavasal, J., "A machine learning-genetic algorithm (ML-GA) approach for rapid optimization using high-performance computing," *Society of Automotive Engineers Technical Paper Series*, Vol. 11, No. 5, 2018.

[19] Bin, S., and Wenlai, Y., "Application of Gaussian process regression to prediction of thermal comfort index," *2013 IEEE 11th International Conference on Electronic Measurement & Instruments*, Vol. 2, IEEE, 2013, pp. 958–961.

[20] Mak, S., Sung, C. L., Wang, X., Yeh, S. T., Chang, Y. H., Joseph, V. R., Yang, V., and Wu, C. F. J., "An efficient surrogate model for emulation and physics extraction of large eddy simulations," *Journal of the American Statistical Association*, Vol. 113, No. 524, 2018, pp. 1443–1456.

[21] Chen, J., Mak, S, Joseph, V. R., and Zhang, C., "Function-on-function kriging, with applications to three-dimensional printing of aortic tissues," *Technometrics*, Vol. 63, 2020, pp. 384–395.

[22] Ji, Y., Mak, S., Soeder, D., Paquet, J., and Bass, S. A., "A graphical Gaussian process model for multi-fidelity emulation of expensive computer codes," *arXiv preprint arXiv:2108.00306*, 2021.

[23] Ji, Y., Yuchi, H. S., Soeder, D., Paquet, J.-F., Bass, S. A., Joseph, V. R., Wu, C., and Mak, S., "Multi-Stage Multi-Fidelity Gaussian Process Modeling, with Application to Heavy-Ion Collisions," *arXiv preprint arXiv:2209.13748*, 2022.

[24] Miganakallu, N., Stafford, J., Amezcua, E., Kim, K. S., Kweon, C.-B. M., and Rothamer, D. A., "Impact of Ignition Assistant on Combustion of Cetane 30 and 35 Jet-Fuel Blends in a Compression-Ignition Engine at Moderate Load and Speed," *Internal Combustion Engine Division Fall Technical Conference*, Vol. 86540, American Society of Mechanical Engineers, 2022, p. V001T03A011.

[25] Sapra, H. D., Hessel, R., Amezcua, E., Stafford, J., Miganakallu, N., Rothamer, D., Kim, K., Kweon, C.-B. M., and Kokjohn, S., "Numerical Modeling and Analysis of Energy-Assisted Compression Ignition of Varying Cetane Number Jet Fuels for High-Altitude Operation," *Internal Combustion Engine Division Fall Technical Conference*, Vol. 86540, American Society of Mechanical Engineers, 2022, p. V001T01A003.

[26] Richards, K., Senecal, P., and Pomraning, E., "CONVERGE (v3. 0), Convergent Science," *Madison, WI*, 2020.

[27] Ren, S., Kokjohn, S. L., Wang, Z., Liu, H., Wang, B., and Wang, J., "A multi-component wide distillation fuel (covering gasoline, jet fuel and diesel fuel) mechanism for combustion and PAH prediction," *Fuel*, Vol. 208, 2017, pp. 447–468.

[28] Senecal, P., Pomraning, E., Richards, K., Briggs, T., Choi, C., McDavid, R., and Patterson, M., "Multi-dimensional modeling of direct-injection diesel spray liquid length and flame lift-off length using CFD and parallel detailed chemistry," *SAE transactions*, 2003, pp. 1331–1351.

[29] Tsan-Hsing, S., William, W., Aamir, S., Zhigang, Y., and Jiang, Z., "A new k-ε eddy viscosity model for high reynolds number turbulent flows," *Computers & Fluids*, Vol. 24, No. 3, 1995, pp. 227–238.

[30] Amsden, A., "A computer program for chemically reactive flows with sprays," *Report of Los Alamos National Laboratory*, 1989.

[31] Reitz, R. D., "Mechanism of breakup of round liquid jets," *Encyclopedia of fluid mechanics*, Vol. 10, 1986.

[32] Schmidt, D. P., and Rutland, C., "A new droplet collision algorithm," *Journal of Computational Physics*, Vol. 164, No. 1, 2000, pp. 62–80.

[33] Gramacy, R. B., and Lee, H. K. H., "Bayesian treed Gaussian process models with an application to computer modeling," *Journal of the American Statistical Association*, Vol. 103, No. 483, 2008, pp. 1119–1130.

[34] Heywood, J. B., *Internal combustion engine fundamentals*, McGraw-Hill Education, 2018.

[35] Stein, M. L., *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Science & Business Media, 1999.

[36] Santner, T. J., Williams, B. J., Notz, W. I., and Williams, B. J., *The Design and Analysis of Computer Experiments*, Vol. 1, Springer, 2003.

[37] Amezcua, E. R., Kim, K., Rothamer, D., and Kweon, C.-B., "Ignition Sensitivity Analysis for Energy-Assisted Compression-Ignition Operation on Jet Fuels with Varying Cetane Number," Tech. rep., SAE Technical Paper, 2022.

[38] Groendyk, M., and Rothamer, D., "Establishing Thermal Stability in an Optically-Accessible CIDI Engine," *SAE International Journal of Advances and Current Practices in Mobility*, Vol. 2, No. 2020-01-0789, 2020, pp. 2650–2664.