1 Introduction

It is shown in the literature that complex combustion physics, such as auto ignition, stretched asymmetric flames, flame propagation, flame instabilities, and other flame dynamics can be captured in small scale configurations (nano and meso scales) [1, 2, 3, 4, 5, 6, 7, 8, 9]. For example, Li et al. [4] performed several numerical simulations at small scales in both cylindrical and parallel configurations and captured the effects of geometry and inlet velocity on the flame behavior in premixed hydrogen/air combustion. They were able to benchmark the important impacts of axial heat conduction on the flame dynamics through preheating the unburned mixture. Tsai [10] investigated the impact of channel size on flame morphology using isothermal planar and circular channels. He concluded that a sufficiently large channel diameter may result in the formation of an asymmetric flame. Ayoobi et al. illustrated that in channels with their width being of the same order of the flame thickness, asymmetric flames and more generally, flame instabilities are less likely to
appear [11, 8]. In another study, Kurdyumov et al. [12] used both steady and transient thermo-diffusive models and investigated premixed flames with the assumption of unity Lewis number in two-dimensional channels and step-wise wall temperature condition. They reproduced different flame modes observed in experimental studies, including stationary, oscillatory, symmetric, and asymmetric modes.

Combustion modeling at small scales is of interest in this work as a mean to generate data for developing and validating a Physics-Based Data Driven Proxy (PB-DDP) model to enable predicting the simulation results on each computational cell in the domain with significant reduction in computational time and memory usage. Such cases are most suitable for this study because a) the data driven model can be investigated against a wide range of different combustion phenomena and b) performing small-scale combustion simulations, in which computational expenses are not as restrictive as in those of practical scales, provides a luxury of employing detailed chemistry in simulations and investigating various operating conditions.

2 Methodology

2.1 Numerical Simulations

In order to generate the data required for developing the PB-DDP model, a set of small-scale combustion simulations are performed over different operating conditions. ANSYS Fluent® [13] is utilized to perform the simulations using high performance computing facilities in Wayne State University. Once these simulations are performed, flame characteristics are sought for each case following a previous study performed by Ayoobi and Schoegl [8]. Correspondingly, critical flame properties are used in developing and validating the PB-DDP model. More information on the specifics of the combustion simulations and the operating condition are presented in Section 2.3.

2.2 Physics-Based Machine Learning

Over the last few years, applications of Artificial Intelligence (AI), in general, and Machine Learning (ML), in particular, have gained broad popularity throughout various industries. This is due to the technological advancements that have created a new massive source of information, such as sensors and high-performance computing services, which require big-data acquisition and storage in different fields and also sophisticated methods of analysis.

There are three main types of ML algorithms, namely, supervised learning, unsupervised learning, and reinforcement learning. Unsupervised learning algorithms are mainly used in pattern detection and descriptive modeling, i.e. they are mainly applied to problems, where no output labels are set. They rather, detect patterns, group the data points, help in presenting the data and drawing conclusions. These algorithms are generally applied to non-physics based problems, such as social media and consumer relations and where explicit mathematical equations are not available to describe the relations between different parameters involved in the process. On the other hand, supervised learning algorithms learn the relationships between target outputs and input features so that one can predict the output values for new data based on the captured relationships in the training process. These algorithms are mostly applied in physics-based problems, where explicit
mathematical equations drive the input and output relationships.

Recently, there have been numerous research works on the application of ML algorithms in Computational Fluid Dynamics (CFD), most of which are limited to building interpretable reduced order models (ROMs). In ROMs, where the number of variables is reduced to simplify the governing equations and the relationships between inputs and outputs, some details are inevitably overlooked. The widespread success of ML-based predictive modeling in other disciplines, such as autonomous cars, suggests a great opportunity to advances in the state-of-the-art by combining conventional CFD simulation techniques with predictive capabilities of physics-based data-driven proxy (PB-DDP) models to truly capture the physics of the problem and enhance prediction capabilities of the simulations at a much lower cost. Previous application of AI in CFD problems has been limited to interpretable models from data [14, 15, 16], and predictive models have not been explored thoroughly in complex CFD problems. Such models, in essence, do not spatio-temporally reduce the order of the problem. They, however, possess the same numerical resolutions as their CFD equivalents do with the great advantage that their solutions can be achieved by significant reduction in computational costs (speed and memory). Ultimately, the predictive models should learn the nature of communications among grid cells and decode the spatial correlations between them (auto- and cross-correlations) in the entire computational domain. They will then predict solutions of completely new sets of simulation runs, from beginning to end.

Physics-based (PB) DDP models are being extensively used in oil and gas industry as well as many other industries for design optimization and uncertainty quantification. In a recent study, a supervised learning algorithm was successfully used in a CFD problem, where a DDP model was developed to predict the characteristics of a fluidized bed problem [17].

To develop and validate a PB-DDP model in this study, (i) new features are defined by performing tier-ing system; (ii) high-fidelity DNS runs are performed and the amount of the data that should be collected from each simulation run are determined. (iii) a set of features, such as the mixture composition, mixture temperature, and geometry index, as inputs and flame properties, such as flame temperature and heat release rate, as outputs are identified; and (iv) grid search is performed to select the optimum parameters for the selected algorithm to result in the most accurate mod-
els. In the application of machine learning (ML) in the physics-based problems, subject-matter expertise (i.e., advanced understanding of the physics of the problem) plays a prominent role in performing and implementing a successful ML project. Combining this expertise and knowledge of data science will result in the best outcome through efficient data gathering and pre-processing. In essence, a physics-based ML model should preserve the physical properties of the system [18].

In order for the ML algorithms to effectively capture the special auto- and cross-correlations, a 2D tier system has been employed in which each cell under investigation is in contact with 4 surrounding cells. Depending on the complexity of the problem, tier I, II or III can be used. For example, if in a two-dimensional case, 10 features are assigned to the original cell, the values of each feature in neighboring cells in the tie-ring system will also be included as new features in the data set and will result in $(1 + 4) \times 10 = 50$ features (if just using the tier I cells). Figure 1 represents the schematic of tier-ing system for a 2-D case, where a cell with its tier I-, tier II- and tier III-associated cells are illustrated.

Figure 1: Schematic of tier-ing system for a 2-D case, where a cell with its tier I-, tier II- and tier III-associated cells are illustrated.

2.3 Case Study

The required data to develop and validate the PB-DDP model are generated from conducting 2D simulations for premixed combustion at small-scale channels with different operating conditions. Figure 2 illustrates a schematic of the 2D domain that is used for numerical simulations in this work. Here, $w = 2 \text{ mm}$ and $L = 20 \text{ mm}$ designate the width and length of the channel, respectively. It is noted that the domain that is used in simulations has the width of $1 \text{ mm}$ with an axis-symmetric boundary condition imposed on the lower wall. Also, the width of the channel is selected to be of the same order of magnitude of the flame thickness to restrict the flame behavior to stationary, symmetric flames for this study. In these simulations, syngas with two different fuel compositions, consisting of hydrogen ($\text{H}_2$), methane ($\text{CH}_4$), and carbon monoxide ($\text{CO}$), is selected to be premixed with air at stoichiometric level ($\phi = 1.0$) and enter the channel at a specified inlet velocity with a detailed chemical mechanism (46 species and 235 reactions) [19]. Details of the operating conditions for these cases are presented in Table 1. In order to ensure the occurrence of ignition and existence of a flame, a fixed temperature gradient is imposed on the upper wall of the channel, i.e. the upper wall temperature is at $300 \text{ K}$ at the inlet and increases linearly to $1500 \text{ K}$ at the outlet.

As shown in Table 1, a different inlet velocity is used for each of the two fuel mixtures. The simulation inputs and predicted results for a handful of time steps, spanning from initial to last time
Table 1: Variations of fuel mixture and inlet velocity for the selected cases

<table>
<thead>
<tr>
<th>Case Number</th>
<th>Fuel Mixture</th>
<th>Inlet Velocity (m/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1-1</td>
<td>30% CO, 5% CH₄, 65% H₂</td>
<td>1</td>
</tr>
<tr>
<td>Case 2-4</td>
<td>40% CO, 10% CH₄, 50% H₂</td>
<td>4</td>
</tr>
</tbody>
</table>

steps, are used to train an ANN model. The predictive capability of the trained neural network is evaluated using the blind tests, which have not been used during the training process of ANN model. The training data are sampled from the relevant distributions using modified Latin Hypercube Sampling (LHS) technique, which is a version of LHS in data partitioning step to guarantee that the model sees enough samples at different temperature regions. LHS guarantees that the samples are taken from the actual data-set and not from fitted distribution to the data set as traditional LHS suggests. With that, not only does the training model see the high frequency regions, but it also considers the extreme cases, such as regions with very high temperature and high gradients or regions with very low temperature or gradients.

Fuel Mixture 1

(a) Time variant total heat release rate in the domain against simulation time. It is noted that, in addition to cases 1-1 and 2-4, this figure includes 3 more inlet velocities for each mixture for the sake of comparison and to indicate the importance of inlet velocity on the flames behavior.

Fuel Mixture 2

(b) Local heat release rate contours (Top Figure: Case 1-1 and Bottom Figure: Case 2-4); contours are scaled from 0 W (blue) to 0.02 W (red).

Figure 3: Selected results from numerical simulations
3 Results and Discussion

In this section, the combustion process for the selected cases will be described, where the sensitivity of the results on the operating condition are emphasized. Then, the development and validation of the PB-DDP model based on the generated data will be discussed.

3.1 Flame Behavior

Figure 3a illustrates the variation (against time) of total heat release rate in the domain for the selected cases along with similar cases (different inlet velocities) to further illustrate the dependency of the transient combustion process on operating condition. In each case, total heat release rate in the domain increases until it reaches its peak value, which indicates the ignition process downstream of the channel. The flame is then formed and starts propagating upstream until it stabilizes at a certain location, where the total heat release rate remains stable. Depending on the operating condition, it is possible to see a flame not reaching a stable condition and experiencing oscillatory behavior or frequent ignitions and extinctions [8, 7]. However, the operating conditions for this work are carefully chosen to ensure that flames reach a stable and stationary condition. Comparing results of the two mixtures with the same inlet velocity (Figure 3a) also shows how the variation of the fuel mixture changes the combustion processes. Figure 3b illustrates local heat release rates inside the domain when the flame is stabilized for the selected cases (Case 1-1 and Case 2-4 in Table 1). It is shown that depending on the operating conditions, flames would experience different characteristics, i.e. flame length, flame speed, location of flame, and other flame properties change considerably as the inlet mixture or the inlet velocity change. Here, when ignition takes place, the subsequent increase in thermal expansion increases the bulk velocity and the burning rates. In Case 2-4, with a larger inlet velocity, the flame requires a larger surface area to be able to burn enough reactants and sustain; therefore, a more stretched flame is observed. While all generated time-variant data including grid information, flow velocity, local pressures, species mass fractions, reaction rates, etc. are used in training and validation of the PB-DDP model, temperature distribution in the domain is selected present the development and validation of the PB-DDP model.

3.2 PB-DDP Model Predictions

From the simulation data, multiple time steps from each case were used to develop and validate the PB-DDP model. The data generated by simulations were partitioned into training and test data sets for each mixture. The test-related time steps are used to validate the predictive capabilities of the PB-DDP model for each mixture. More specifically, the simulation data were divided into two sections: the first section with the largest amount of data (80% of the data) is used to train the model, from which 15% of the training data is randomly picked to check the accuracy of the training as calibration data set. The remaining 20% of the data is used as the test portion of the data and are only used to examine the predictive capabilities and the robustness of the PB-DDP model. The trained artificial neural network (ANN) is used to predict the simulation results (ex. temperature here) at each grid cell (total 20 × 400 cells in the domain).
As mentioned in Section 2.2, various features (pressure, velocities, mass fractions, reaction rates, etc.) are associated to each cell and its tire I cells in the domain (see Figure 1). More specifically, the DNS results from the time steps 0.01 s, 0.014 s, 0.018 s, and 0.022 s for Case 1-1 and 0.004 s, 0.007 s, 0.01 s, and 0.013 s for Case 2-4 are used for the training of the PB-DDP model. The selected time steps for Case 2-4, mostly represent the stationary flame or smooth flame propagation within a very small distance. However, in the selected time steps for Case 1-1, the flame is not yet stationary and is still propagating upstream so that the predictability of the model for a moving flame can be examined as well. The model predictions are tested by looking at temperature distributions for time steps 0.017 s and 0.008 s, for cases 1-1 and 2-4, respectively. Again, these test time steps are not seen by the model. Cross plots of predictions made by the PB-DDP model against those from DNS results are illustrated in Figure 4. The comparisons are made for the data in all cells in the domain. Also, the distribution of the data for both data driven and DNS results are provided in the same figure. These cross plots suggest that the overall predictability of the data driven model is acceptable, however there are discrepancies that need to be addressed. Especially, these discrepancies are more observable for Case 1-1 with the data being more scattered as compared to those in Case 2-4. It is noted that Predictions by this model are done at a significantly shorter time on a regular PC workstation.

To further understand the discrepancies mentioned above, a closer comparison between DNS and PB-DDP model results is provided for both cases, where temperatures across the domain are compared (Figure 5). Contour plots in this figure only represent half of the channel as the lower wall has axis-symmetric boundary condition. It is shown that the overall flame properties, such as flame shape and flame location are predicted well by the PB-DDP model. Here, the distribution of local errors (\( \frac{\Delta T_{DNS} - \Delta T_{PB-DDP}}{\Delta T_{DNS}} \times 100 \)) are provided as well (Figure 5). The error distribution in the domain provides insightful information about the model: a) how well the PB-DDP model can predict the simulations on a cell by cell basis, and b) where in the domain the model is stronger or weaker. Such information is useful in further improving the data driven model algorithms undertaken for these predictions.
When comparing cell-based values obtained from predictions, it is observed that PB-DDP predictions and DNS results are not in complete agreements. This discrepancy is especially more pronounced where there are sharper gradients. This is why maximum errors occur closer to the flame front/reaction zone, which has high temperature and correspondingly higher heat release rates. This suggests that the model parameters should be further optimized to address such sharp changes. Comparing the two cases, it is shown that the maximum error is Case 2-4 is less than 15%, which only occurs at the flame front in a location close to the axis-symmetric boundary and the errors in other locations are mainly around 10% or less. The errors in Case 1-1 are however
larger and even exceed 20% at some locations on the flame front. The overall error in this case is larger, which is attributed to the propagation of the flame within the times steps that are being examined here, i.e. with a more transient case, the model needs to be further refined to better capture the transient behavior of the flame.

This is an ongoing study by the authors at its initial stages. The following concerns that were aroused from this work should be addressed in the subsequent studies for more accurate applications of data driven approaches to reactive flow modeling: (a) the model in this work is designed for post-ignition because at the ignition time, temporal and spatial changes are considerably sharp and simulation data at smaller time steps are required to train a data driven model and predict such changes, (b) the input to the proxy model comes directly from DNS simulations at each deployment stage to allow for checking the accuracy of the PB-DDP outputs with the ground truth (DNS outputs at each time step), and (c) there are artificial discrepancies observed on the boundaries that are attributed to the tier-ing model that is used in this work. This model can to be further refined for the cells along boundaries to minimize such errors.

4 Conclusions

The multi-scale nature of reactive flows along with the interactions between different physical phenomena (ex. flow characteristics, chemical reaction, molecular diffusion, turbulence, etc.) increase the computational costs (time and memory) exponentially, such that direct numerical simulation of practical cases becomes impractical unless simplified models are introduced in simulations and some physics are overlooked. Hence, it is essential to seek new methodologies in order to perform precise simulations in a reasonable period of time while the accuracy is not sacrificed. This paper examined the plausibility of employing physics-based machine learning techniques to model reacting flows. Using premixed combustion at small scales, results confirmed that it is possible to replicate costly numerical simulations by physics-based data-driven proxy (PB-DDP) models in an insignificant computational expenses. It is noted the overall flame behavior from PB-DDP predictions were in good agreement with their corresponding DNS results. However, there were discrepancies mainly in areas with high gradients, which can be resolved by tuning the proxy model in future studies.

References


