

---

# Random forests for Accelerating Turbulent Combustion Simulations

---

**Wai Tong Chung \***  
Department of Mechanical Engineering  
Stanford University  
Stanford, CA 94305  
wtchung@stanford.edu

**Aashwin Ananda Mishra**  
SLAC National Accelerator Laboratory  
Menlo Park, CA 94025, USA  
aashwin@stanford.edu

**Nikolaos Perakis**  
Chair of Space Propulsion  
Technical University of Munich  
85748 Garching, Germany  
nikolaos.perakis@tum.de

**Matthias Ihme**  
Department of Mechanical Engineering  
Stanford University  
Stanford, CA 94305  
mihme@stanford.edu

## Abstract

High-fidelity combustion simulations are useful for optimizing engineering designs, and can result in reduced design costs, increased engineering performance, and lower emissions. However, these techniques are limited by high computational expense. In this investigation, we accelerate unsteady combustion simulations by employing random forests for dynamic combustion submodel assignment. Random forests, trained with local flow properties as input variables and combustion model errors as training labels, assign three different combustion models – finite-rate chemistry (FRC), flamelet progress variable (FPV), and inert mixing (IM) – with high classification accuracy in *a priori* tests. *A posteriori* simulations, integrating the machine learning model in the computational fluid dynamics solver, demonstrate that high-fidelity simulations can be performed with this approach at significantly reduced cost compared to detailed chemistry simulations and simultaneously achieving improved accuracy over low-order combustion models.

## 1 Introduction

Combustion processes are ubiquitous in engineering applications, such as in rockets, power plants, and automotive engines. Accurate combustion simulation techniques are useful for optimizing engineering designs, and can result in reduced design costs, increased engineering performance, and substantially lower greenhouse-gas emissions and pollutants. However, commonplace adoption of such high-fidelity simulation techniques is often limited by their computational expense. Hence, a significant portion of combustion research has been devoted to the development of cost-efficient physics-based models for representing the combustion chemistry and turbulent scales [1] in computationally expensive large-scale high-fidelity simulations.

Alternatively, data-driven methods can be employed for fast and accurate predictive modeling. In particular, artificial neural networks have been employed for regressing thermophysical quantities [2–6], and modeling turbulent terms [7, 8]. However, applications of regression models in flow-physics problems are still in their infancy, and face challenges when extrapolating beyond the training set – resulting in generalization errors that arise from numerical predictions that only match specific flow configurations represented by the training data [9].

---

\*Corresponding author.

This study ameliorates this issue by employing a machine learning classification algorithm that selects well-tested physics-based combustion submodels of varying fidelity and complexity, and assigns them to different regions of the simulation domain. Thus, the potential numerical errors made by the machine-learning algorithm are limited by the predictive capability of the lowest performing submodel. To this end, we examine the feasibility of employing random forests [10] for the purpose of local and dynamic model assignment in large-eddy simulations (LES) of a gaseous-oxygen/gaseous-methane (GOX/GCH4) rocket combustor [11, 12].

## 2 Mathematical models

Large-eddy simulations in the present study are performed by solving the Favre-filtered conservation equations for mass, momentum, energy, and chemical species:

$$\partial_t \bar{\rho} + \nabla \cdot (\bar{\rho} \tilde{\mathbf{u}}) = 0 \quad (1a)$$

$$\partial_t (\bar{\rho} \tilde{\mathbf{u}}) + \nabla \cdot (\bar{\rho} \tilde{\mathbf{u}} \tilde{\mathbf{u}}) = -\nabla \cdot (\bar{p} \mathbf{I}) + \nabla \cdot (\bar{\boldsymbol{\tau}}_v + \bar{\boldsymbol{\tau}}_t) \quad (1b)$$

$$\partial_t (\bar{\rho} \tilde{e}) + \nabla \cdot [\tilde{\mathbf{u}} (\bar{\rho} \tilde{e} + \bar{p})] = -\nabla \cdot (\bar{\mathbf{q}}_v + \bar{\mathbf{q}}_t) + \nabla \cdot [(\bar{\boldsymbol{\tau}}_v + \bar{\boldsymbol{\tau}}_t) \cdot \tilde{\mathbf{u}}] \quad (1c)$$

$$\partial_t (\bar{\rho} \tilde{\phi}) + \nabla \cdot (\bar{\rho} \tilde{\mathbf{u}} \tilde{\phi}) = -\nabla \cdot (\bar{\mathbf{J}}_v + \bar{\mathbf{J}}_t) + \bar{\mathbf{S}} \quad (1d)$$

with density  $\rho$ , velocity vector  $\mathbf{u}$ , specific total energy  $e$ , stress tensor  $\boldsymbol{\tau}$ , and heat flux vector  $\mathbf{q}$ ;  $\bar{\cdot}$  denotes a filtered quantity and  $\tilde{\cdot}$  is a Favre-filtered quantity. Subscripts  $v$  and  $t$  denote viscous and turbulent quantities, respectively. The pressure  $p$  is computed from the ideal gas equation of state.  $\phi$ ,  $\mathbf{J}$ , and  $\mathbf{S}$  are the transported scalars, scalar diffusive fluxes, and scalar source terms for the candidate combustion models. The dynamic Smagorinsky model [13] and dynamic thickened-flame model [14] are used to model closure in the turbulent terms. Simulations are performed by employing an unstructured compressible finite-volume solver [15–17].

In this work, we employ three different combustion submodels, namely an inert mixing (IM) model, the flamelet/progress variable (FPV) model [18, 19], and a finite-rate chemistry (FRC) model. The present framework couples the different combustion models with the approach developed by Wu et al. [17], which ensures the conservation of mass, momentum, and energy. The GRI-3.0 chemical mechanism [20], involving 33 chemical species, is used to describe combustion chemistry.

## 3 Data-driven methods

In this section, we describe the procedure for incorporating a supervised learning algorithm for combustion submodel assignment. Firstly, we use the instantaneous flow-field solutions from the FRC simulation of the combustor as the learning dataset. FRC data are then used to reconstruct FPV and IM quantities of interest  $\alpha \in Q$  by interpolating from generated flamelet tables [21].

$$\alpha^y \approx \alpha_{\text{table}}^y(\tilde{Z}_{\text{FRC}}, \tilde{C}_{\text{FRC}}) \text{ where } y \in \{\text{FPV}, \text{IM}\}. \quad (2)$$

Secondly, we assign labels  $\mathcal{Y} = \{\text{IM}, \text{FPV}, \text{FRC}\}$  to the training data. We consider FRC as combustion model of highest fidelity but at the expense of highest computational cost. Therefore, we assign labels in the training set based on the normalized combustion submodel error  $\epsilon_Q^y$  of quantities of interest  $\alpha \in Q$  between FRC and the models of lower fidelity [22]:

$$\epsilon_Q^y = \frac{1}{N} \sum_{\alpha \in Q} \frac{|\alpha^{\text{FRC}} - \alpha^y|}{\|\alpha^{\text{FRC}}\|_\infty} \text{ with } y \in \{\text{FPV}, \text{IM}\}, \quad (3)$$

where the error for considering  $N$  number of quantities of interest (QoIs) is a normalized linear combination of each individual submodel error. A model of higher fidelity is assigned when the QoI submodel error  $\epsilon_Q^y$  exceeds a user-defined threshold  $\theta_Q^y$ , with FRC chosen when all conditions for selecting FPV and IM are not met. Figure 1 shows the use of this labeling approach on the training data in mixture fraction-progress variable ( $\tilde{Z}$ - $\tilde{C}$ ) composition space for  $\theta_{\{T, \text{CO}\}} = 0.02$  and  $\theta_{\{T, \text{CO}\}} = 0.05$ , respectively. In both cases, IM is shown to be assigned at points where  $\tilde{C} \approx 0$ , FPV is assigned mostly to conditions near the equilibrium composition. The submodel assignment reverts back to FRC in regions dominated by non-equilibrium effects and heat-losses that are not

captured by the adiabatic steady-state flamelet formulation. Employing  $\theta_{\{T,CO\}} = 0.02$  is seen to be more stringent than employing  $\theta_{\{T,CO\}} = 0.05$ , with a 0.18 greater fraction of scatter data on the stable branch assigned as FRC, especially for fuel-rich mixtures. It should be noted that while most out-of-flamelet regions would be assigned FRC, some regions with low reactivity and far from stoichiometry (eg.  $\tilde{Z} = 0.7$ ) generate smaller errors which are then be assigned FPV.

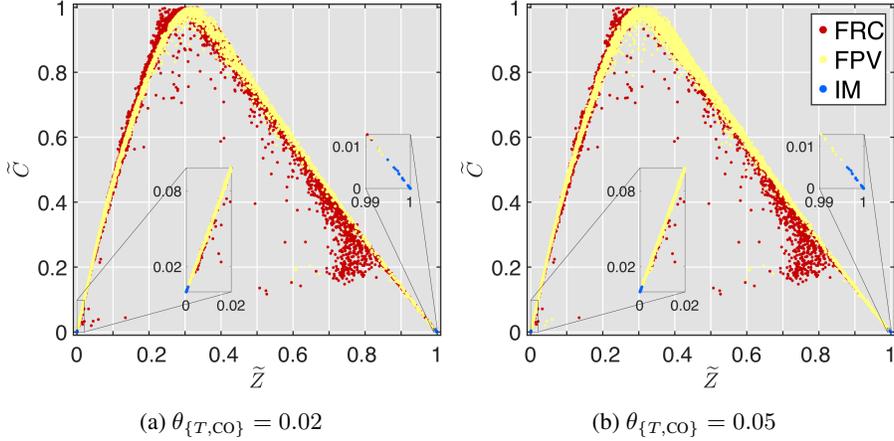


Figure 1: Training data for two different combustion submodel error thresholds  $\theta_{\{T,CO\}}$ .

Thirdly, we construct the feature vector  $\mathbf{x} \in \mathcal{X}$ . To this end, we applied the Maximal Information Coefficient (MIC) [23] to identify the top six (out of fifteen) thermophysical quantities with the strongest relationships with the local combustion submodel error. These six features, namely mixture fraction, progress variable, density, local Prandtl number, and Euclidean norm of the mixture fraction gradient, *viz.*,  $\mathbf{x} = [\tilde{Z}, \tilde{C}, \tilde{\rho}, \tilde{T}, Pr_{\Delta}, \|\nabla \tilde{Z}\|_2]$  are then selected for constructing the feature set.

Lastly, we train, validate, and test the classification algorithm. In the present investigation, the random forest classifier from the `OPENCV` library [24] is used. Classification cost scales with the number of trees, tree depth and the number of training points [25]. Hence, a random forest consisting of twenty decision trees, and maximum depth of ten nodes is employed. Additionally,  $1 \times 10^4$  training points have been randomly sampled from a single simulation snapshot consisting of  $2 \times 10^5$  cells. A similar approach is used in other supervised learning problems [9]. We note that the flow in the present configuration is statistically stationary, and thus training data from a single snapshot was found to be sufficient for representing the thermophysical behavior of the combustor. The number of trees, tree depth, and the number of training points are determined *a priori* by ensuring that the classification performance remains unchanged on a validation set. Training is performed once *a priori*, and requires 530 ms of walltime with 1 CPU. In *a posteriori* simulations, random forest evaluations for  $2 \times 10^5$  cells at each timestep require 1 ms of wall time with 600 CPUs.

## 4 Experimental configuration and computational setup

We perform simulations of the gaseous oxygen-gaseous methane rocket combustor setup by Silvestri et al. [11, 12] using an axisymmetric domain. Inlet fuel and oxidizer mass flow rates and temperature, along with chamber and nozzle wall temperatures are prescribed following experimental measurements [11, 12, 26]. All remaining boundaries are defined as adiabatic non-slip walls with the exception of the exhaust, which is modeled as a pressure outlet. The computational domain is discretized by a block-structured mesh consisting of  $2 \times 10^5$  cells. The wall-normal direction is resolved down to  $30 \mu\text{m}$ , and a wall model [27] is employed for the viscous sublayer. A typical timestep is 25 ns, corresponding to a convective CFL number of 1.0.

## 5 Results

We first perform an *a priori* assessment to determine the accuracy of random forest classification, as shown in Table 1, on a monolithic FRC simulation test dataset from an unseen timestep. Temperature and CO mass fraction fields from the test dataset are shown in fig. 2a. and 2b. Temperature  $\tilde{T}$  is chosen as a QoI to describe the combustion efficiency and engine performance. CO mass fraction,  $\tilde{Y}_{CO}$ , is chosen to challenge the deficiencies of FPV and IM in capturing intermediate species [17]. Throughout this study we explore cases that use the same threshold for both IM and FPV, *viz.*,  $\theta_Q^M = \theta_Q^{FPV} = \theta_Q$  for simplicity. Classification accuracy range from approximately 0.7 to 0.8, which is comparable to the use of random forests in other flow physics problems [28].

Table 1: Classification accuracy and submodel assignment of cases investigated.

Case,	$\theta_T=0.05$	$\theta_{CO}=0.05$	$\theta_T=0.02$	$\theta_{CO}=0.02$	$\theta_{\{T,CO\}}=0.05$	$\theta_{\{T,CO\}}=0.02$
Quantity-of-interest, $Q$	$\tilde{T}$	$\tilde{Y}_{CO}$	$\tilde{T}$	$\tilde{Y}_{CO}$	$\{\tilde{T}, \tilde{Y}_{CO}\}$	$\{\tilde{T}, \tilde{Y}_{CO}\}$
Classification accuracy	0.774	0.756	0.725	0.715	0.753	0.734
IM:FPV:FRC	5:67:28	18:48:34	5:33:62	18:35:47	6:63:31	6:42:52

Figure 2c. and 2d. demonstrates the *a priori* combustion submodel assignment on an unseen FRC-simulation snapshot for cases  $\theta_{\{T,CO\}} = 0.05$  and  $\theta_{\{T,CO\}} = 0.02$ . For both cases shown, inert mixing (IM) is assigned in 6% of the domain at the injector and the oxidizer core, where chemical processes are insignificant. In general, FRC is assigned at the near-wall and fuel-rich regions within the combustor where intermediate species reactions are not captured well by tabulated chemistry submodels.  $\theta_{\{T,CO\}} = 0.05$  results in 31% FRC assignment within the domain, while  $\theta_{\{T,CO\}} = 0.02$  results in 52% FRC assignment. We observe that model assignment in both cases is not spatially smooth, and that model assignment appears speckled. This is because the smoothness of classification boundaries formed within the 6-dimensional feature space is not translated when transformed to physical space. This is a common issue in classification problems involving spatial data, such as in medical imaging or image processing. Two strategies can be employed to improve spatial smoothness in classification problems [29, 22]: (i) applying the classification techniques to a neighborhood of cells, or (ii) applying a spatial filter on the predicted labels and discretizing the filtered labels. In the *a posteriori* simulations in fig. 3, we apply the latter strategy since it is better suited with the current framework that uses local quantities as QoIs and features.

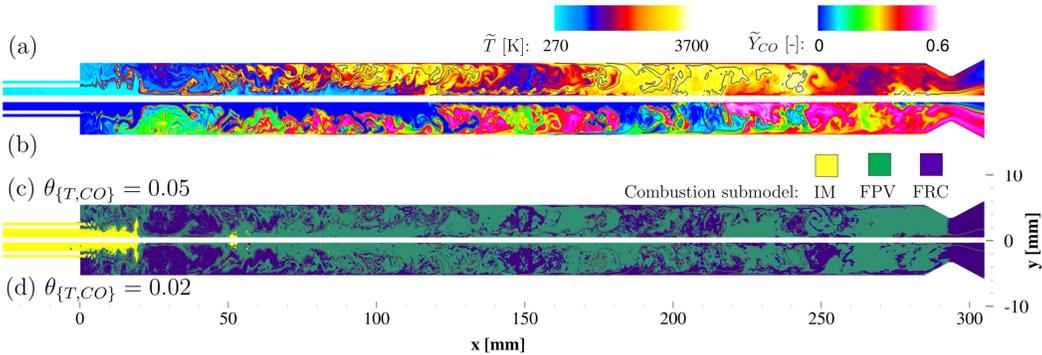


Figure 2: Instantaneous (a) temperature, (b) CO mass fraction, and (c,d) combustion submodel assignment from test set in the *a priori* assessment. The location of the stoichiometric mixture,  $\tilde{Z}_{st} = 0.2$ , is shown by black lines.

Figure 3 compares temperature and CO mass fraction fields from the rocket simulation using monolithic FRC, monolithic FPV, and *a posteriori* data-assisted (DA) simulations. The data-assisted simulations integrate the trained random forest model in the computational fluid dynamics solver to carry out dynamic submodel selection. Note that the FRC simulation results in the highest fidelity, and requires the most computational resources. When compared to monolithic FRC LES, some notable differences are observable from the FPV simulations. In particular, a thicker thermal boundary layer is seen for the FPV simulation, resulting in cooler temperature near the wall. Low near-wall temperature results in greater CO formation.

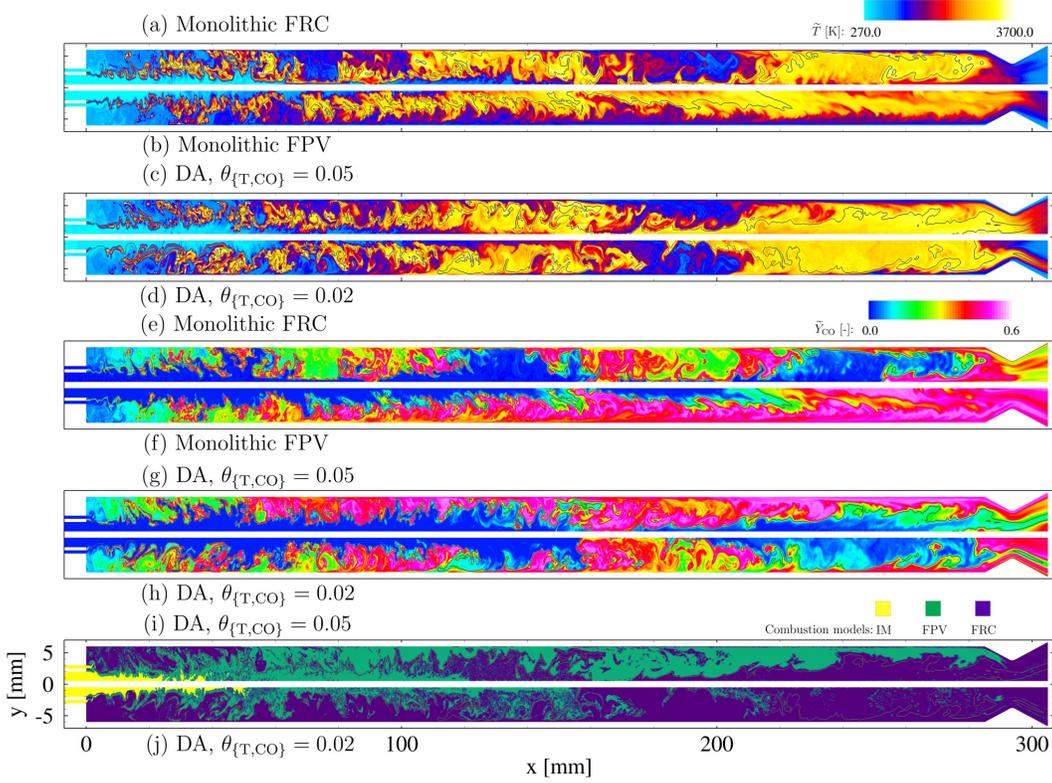


Figure 3: Instantaneous (a,b,c,d) temperature and (e,f,g,h) CO mass fraction for all simulations. The location of the stoichiometric mixture,  $\tilde{Z}_{st} = 0.2$ , is shown by black lines. (i,j) Combustion submodels used in the data-assisted (DA) simulations are also presented.

*A posteriori* data-assisted (DA) simulations using two different model thresholds,  $\theta_{\{T,CO\}} = 0.05$  and  $\theta_{\{T,CO\}} = 0.02$  are performed by employing random forest classifiers in-flight for combustion submodel assignment during simulation runtime. Employing model threshold  $\theta_{\{T,CO\}} = 0.05$  on the DA simulation results in temperature predictions that are in good agreement with the monolithic FRC simulation. However, a greater CO mass fraction is observed compared to monolithic FRC simulations. Tightening the model threshold  $\theta_{\{T,CO\}} = 0.02$  results in temperature and CO mass fraction fields that agree with the monolithic FRC LES. The corresponding combustion submodel assignment is shown in fig. 3. FRC utilization for  $\theta_{\{T,CO\}} = 0.05$  is at 34%, resulting in 70% FRC cost, or – equivalently – a reduction in the computational cost by 30%. Model assignment for  $\theta_{\{T,CO\}} = 0.02$  results in 60% FRC utilization, resulting in 80% FRC cost.

## 6 Conclusions

This work proposes a data-driven modeling approach by which random forest classifiers spatially and dynamically assign three different candidate combustion submodels. This modeling approach is demonstrated in simulations of a complex rocket combustor. Results demonstrated that random forests showed high classification accuracy for this task. *A posteriori* simulations incorporating random forests showed significant improvements from monolithic FPV simulations in all quantities at a 30% lower cost than monolithic FRC calculations. Interesting opportunities for extending this work include (i) the exploration of other classification algorithms, and (ii) the addition of non-local quantities in the feature and label set.

## Impact Statement

Combustion processes are common in engineering applications, such as in power generation, propulsion, and chemical engineering. These processes involve complex physical and chemical phenomena – thermochemistry, chemical kinetics, molecular transport, heat and mass transfer, and laminar and turbulent fluid dynamics. Therefore, accurate combustion simulation techniques are useful for optimizing engineering designs, and can result in reduced design costs, increased engineering performance, and lower emissions. However, straightforward adoption of such high-fidelity simulation techniques are often limited by costly computational resources as a consequence of accurately describing the complex physico-chemical phenomena. By directly addressing this issue, we have developed and demonstrated a method for employing data-driven methods, by integrating random forests into the computational fluid dynamics solver, for reducing computational costs while maintaining high fidelity of combustion simulations.

## Acknowledgments and Disclosure of Funding

The authors gratefully acknowledge financial support from the Air Force Office of Scientific Research under Award No. FA9300-19-P-1502, NASA with award No. 80NSSC18C0207, and from the Stanford University Harold and Marcia Wagner Engineering Fellowship. Resources supporting this work are provided by the High-End Computing (HEC) Program at NASA Ames Research Center.

## References

- [1] S. B. Pope. Small scales, many species and the manifold challenges of turbulent combustion. *Proc. Combust. Inst.*, 34:1–31, 2013.
- [2] F. C. Christo, A. R. Masri, E. M. Nebot, and S. B. Pope. An integrated PDF/neural network approach for simulating turbulent reacting systems. *Proc. Combust. Inst.*, 26:43–48, 1996.
- [3] J. A. Blasco, N. Fueyo, J. C. Larroya, C. Dopazo, and J. Y. Chen. Single-step time-integrator of a methane-air chemical system using artificial neural networks. *Comput. Chem. Eng.*, 23(9): 1127–1133, 1999.
- [4] M. Ihme, C. Schmitt, and H. Pitsch. Optimal artificial neural networks and tabulation methods for chemistry representation in les of a bluff-body swirl-stabilized flame. *Proc. Combust. Inst.*, 32:1527–1535, 2009.
- [5] A. Kempf, F. Flemming, and J. Janicka. Investigation of lengthscales, scalar dissipation, and flame orientation in a piloted diffusion flame by LES. *Proc. Combust. Inst.*, 30:557–565, 2005.
- [6] B. A. Sen and S. Menon. Linear eddy mixing based tabulation and artificial neural networks for large eddy simulations of turbulent flames. *Combust. Flame*, 157(1):62–74, 2010.
- [7] C. J. Lapeyre, A. Misdariis, N. Cazard, D. Veynante, and T. Poinsot. Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates. *Combust. Flame*, 203: 255–264, 5 2019.
- [8] M. T. Henry de Frahan, S. Yellapantula, R. King, M. S. Day, and R. W. Grout. Deep learning for presumed probability density function models. *Combust. Flame*, 208:436–450, 2019.
- [9] J.-L. Wu, H. Xiao, and E. Paterson. Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Phys. Rev. Fluids*, 3:74602, 2018.
- [10] Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, 2001.
- [11] S. Silvestri, M. P. Celano, O. J. Haidn, and O. Knab. Comparison of single element rocket combustion chambers with round and square cross sections. *6th Euro. Conf. Aeronautics Space Sci. (EUCASS)*, 2015.
- [12] S. Silvestri, M. P. Celano, C. Kirchberger, G. Schlieben, O. Haidn, and O. Knab. Investigation on recess variation of a shear coax injector for a single element GOX-GCH4 combustion chamber. *Trans. JSASS Aerospace Tech. Japan*, 14(ists30):101–108, 2016.

- [13] P. Moin, K. Squires, W. Cabot, and S. Lee. A dynamic subgrid-scale model for compressible turbulence and scalar transport. *Phys. Fluids A*, 3(11):2746–2757, 11 1991.
- [14] O. Colin, F. Ducros, D. Veynante, and T. Poinso. A thickened flame model for large eddy simulations of turbulent premixed combustion. *Phys. Fluids*, 12(7):1843–1863, 7 2000.
- [15] Y. Khalighi, J. W. Nichols, F. Ham, S. K. Lele, and P. Moin. Unstructured large eddy simulation for prediction of noise issued from turbulent jets in various configurations. *AIAA Paper 2011-2886*, 2011.
- [16] P. C. Ma, Y. Lv, and M. Ihme. An entropy-stable hybrid scheme for simulations of transcritical real-fluid flows. *J. Comput. Phys.*, 340:330–357, 2017.
- [17] H. Wu, P. C. Ma, T. Jaravel, and M. Ihme. Pareto-efficient combustion modeling for improved CO-emission prediction in LES of a piloted turbulent dimethyl ether jet flame. *Proc. Combust. Inst.*, 37(2):2267–2276, 1 2019.
- [18] C. D. Pierce and P. Moin. Progress-variable approach for large-eddy simulation of non-premixed turbulent combustion. *J. Fluid Mech.*, 504:73–97, 4 2004.
- [19] M. Ihme, C. M. Cha, and H. Pitsch. Prediction of local extinction and re-ignition effects in non-premixed turbulent combustion using a flamelet/progress variable approach. *Proc. Combust. Inst.*, 30:793–800, 2005.
- [20] G. P. Smith, D. M. Golden, M. Frenklach, and N. W. Moriarty et. al. GRI-Mech 3.0, 2000. <http://www.me.berkeley.edu/gri-mech/>.
- [21] H. Pitsch. FLAMEMASTER v3.1: A C++ computer program for 0D combustion and 1D laminar flame calculations, 1998.
- [22] H. Wu, Y. C. See, Q. Wang, and M. Ihme. A Pareto-efficient combustion framework with submodel assignment for predicting complex flame configurations. *Combust. Flame*, 162: 4208–4230, 2015.
- [23] D. N. Reshef, Y. A Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. Detecting novel associations in large data sets. *Science*, 334(6062):1518–1524, 2011.
- [24] G. Bradski. The OpenCV library. *Dr. Dobb's J. Softw. Tools*, 2000.
- [25] Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone. *Classification and Regression Trees*. Routledge, 1984.
- [26] N. Perakis and O. J. Haidn. Inverse heat transfer method applied to capacitively cooled rocket thrust chambers. *Int. J. Heat Mass Transf.*, pages 150–166, 3 2019.
- [27] S. Kawai and J. Larsson. Dynamic non-equilibrium wall-modeling for large eddy simulation at high Reynolds numbers. *Phys. Fluids*, 25(1):015105, 1 2013.
- [28] J. Ling and J. Templeton. Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty. *Phys. Fluids*, 27(8):085103, 2015.
- [29] Konrad Schindler. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Trans. Geosci. Remote Sensing*, 50(11 PART 1):4534–4545, 2012.