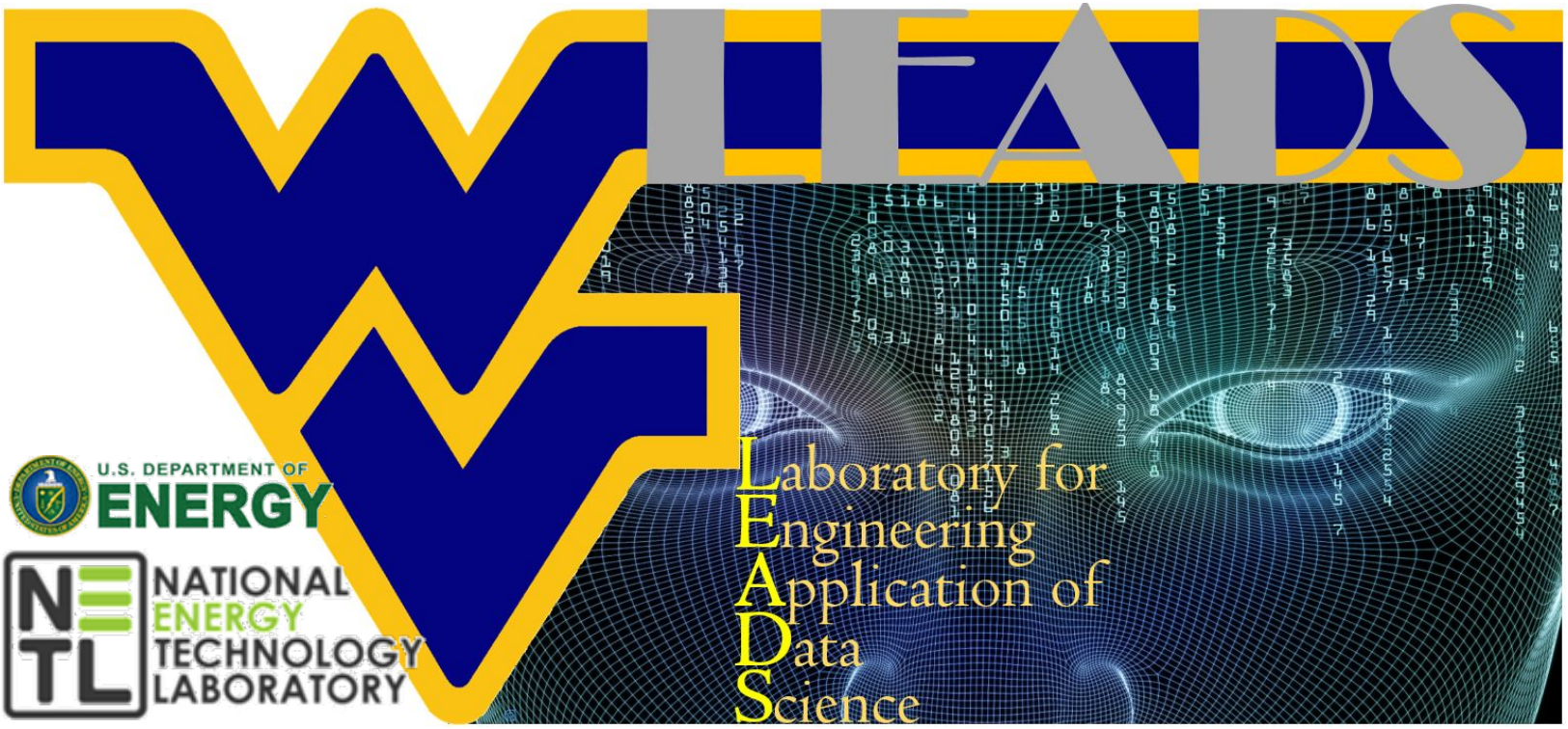


Smart Proxy Modeling

Application of Artificial Intelligence & Machine Learning in Computational Fluid Dynamics

Aboaba[1], A., Martinez[1], Y., Mohaghegh[1], S., Shahnam[2], M., Guenther[2], C., and Liu[3], Y.

[1]WVU-LEADS
[2]DOE-NETL
[3]Leidos Research Support Team

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

Acronyms, Abbreviations, and Symbols

Term	Description
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
CCADS	Combustion Control and Diagnostic Sensor
CFD	Computational Fluid Dynamics
DPM	Discrete Phase Model
GHG	Green House Gases
HPC	High-Performance Computing
MFiX	Multiphase Flow with Interphase eXchanges
RAM	Random Access Memory
RTE	Radiative Transfer Equation
TDM	Top-Down Model
VTK	Visualization Tool Kit
WSGGM	Weighted-sum-of-grey-gases model

Acknowledgements

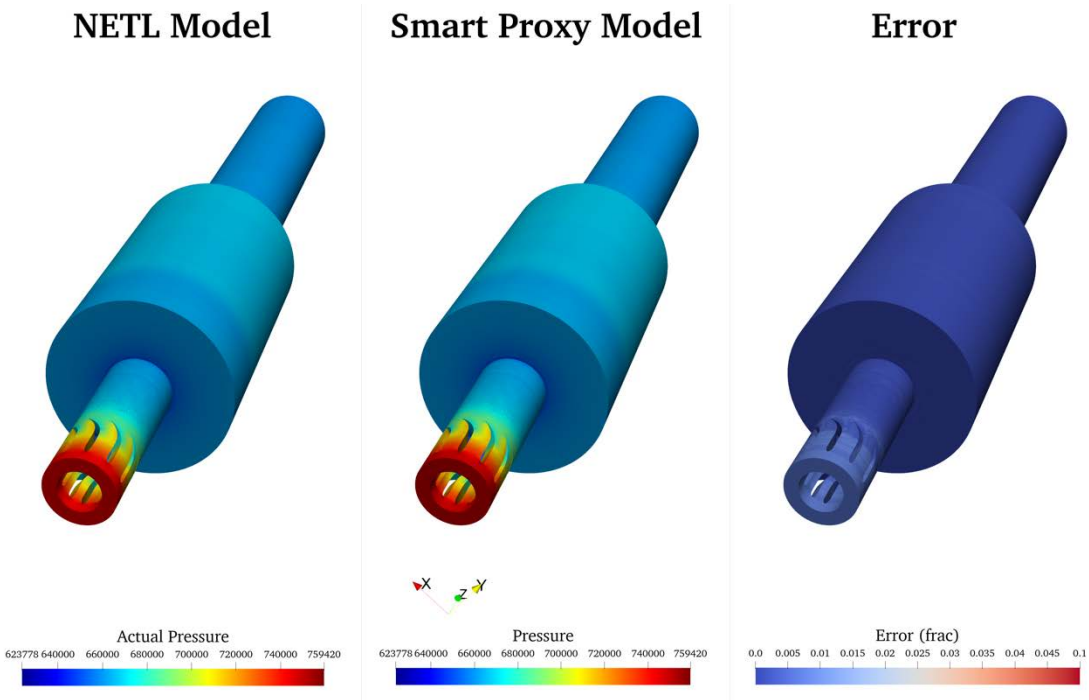
This work was performed in support of the National Energy Technology Laboratory's ongoing research under the RSS contract 89243318CFE000003.

Disclaimer

EXECUTIVE SUMMARY

Smart proxy technology leverages the art of artificial intelligence and machine learning in order to build accurate and very fast proxy models for highly complex numerical simulation models. In this project, smart proxy technology is used to replicate the results of a highly complex single phase CFD simulation with reasonable degree of accuracy while reducing the computational cost associated with such CFD simulations.

The CFD model under study simulates the combustion of natural gas under various conditions such as varying natural gas composition and flow rate, inlet air flow rate and temperature, and outlet pressure in a High-Pressure Combustion Facility (B6 Combustor) with more than 4 million simulation cells. Only eight CFD simulation runs were used to create a smart proxy model that replicates the detail distribution of Pressure, Temperature, Nitrogen, Oxygen and Carbon-dioxide concentration in the CFD simulation model in seconds with less than 10% error. Following figure shows a validation example of the smart proxy results for the distribution of Pressure, comparing the CFD simulation result (left) to the smart proxy result (middle) and the error plot (right) showing the difference between the two simulation runs (Fluent vs. Smart Proxy).



Blind Validation Smart Proxy Results for Pressure – Blend Case 10

1. INTRODUCTION

DOE-NETL is supporting projects to develop technologies that will improve the efficiency, cost, and environmental performance of complex power generation systems such as gas turbines and coal-fired power plants. In-situ monitoring of combustion phenomena is a critical need for optimal operation and control of such systems. CFD is an important tool currently being used to investigate and understand the dynamics of the combustion process in these systems. Gas turbine combustion is a complex process, and it can be a challenge to achieve accurate and reliable CFD simulation results at a reasonable computational cost.

The challenge in CFD simulation of complex reaction flows is to adequately resolve the structures that exist at different spatial and temporal scales in an inherently transient flow. Additionally, in reacting gas-solid flow simulations, small time steps are needed in order to not only resolve the temporal scales of the flow, but also ensure numerical stability of the solution. A rule of thumb for adequate spatial resolution is for the grid spacing to be about 10 times the particle diameter [1]. The grid requirement for maintaining such a ratio of grid size to particle diameter for smaller size particles makes such simulations computationally costly and impractical [2]. Recent work at NETL [2] has shown the number of simulations, which is required for non-intrusive uncertainty quantification, can easily exceed many tens of simulations. The spatial and temporal resolution requirements for multiphase flows make CFD simulations computationally expensive and potentially beyond the reach of many design analysts [3].

The goal of this research work is to develop a data driven predictive model capable of replicating the thermal-flow pattern and species distribution results of CFD simulation of natural gas combustion in a High-Pressure Combustion Facility (B6 Combustor). Achieving this goal will significantly reduce the typical long time-to-solution characteristics of CFD simulations while preserving traditional CFD solver accuracy for the CFD simulation model under study. The developed smart proxy will contribute greatly to the development of technologies that improve the efficiency, cost, and environmental performance of complex power generation systems.

1.1 Structure of the Work

The report presented in this document details the research work performed in building a data driven predictive model that replicates transport variables in the CFD simulation of the combustion of natural gas in a High-Pressure Combustion Facility (B6 Combustor). The data driven modeling framework presented for the B6 Combustor model in this report will ultimately be applied to a more complex system (the Tri-State Coal-fired Boiler) in a separate project.

In chapter one (this chapter), the problem was defined, and the final objective of the research was articulated. In chapter two, a brief background information is provided on key elements of the

research work. A brief description of the reaction flow and governing equations used in the numerical CFD simulation software is provided in order to lay the groundwork for understanding the engineering and scientific details associated with the CFD model being studied. A summary of the different machine learning and computing techniques used on the project is included in chapter two to provide a background to the solution methodology utilized. Also, a brief literature review on the use of AI and Machine learning relating to fluid dynamics is provided.

In chapter three, a general overview of the end-to-end workflow of the smart proxy development process is introduced with a description of the design and implementation of the framework used for the development of the smart proxy model. This chapter provides a description of B6 Combustor CFD simulation model, a detailed description of the simulation data received including the boundary conditions, a brief description of steps taken to develop the smart proxy model and a detailed description of the different machine learning algorithms used in the development process.

Chapter four provides a lot more detail on every step taken to meet the objective of the research. Detail information is provided regarding the input training dataset and the neural network setup at each step of the development process. This chapter includes results and discussions for each step taken towards building the smart proxy model. Conclusions and recommendations for the next phase of the research are presented in chapter five.

2 BACKGROUND

This section of the report provides some basic but necessary background on some key components of this research work.

2.1 NETL's High-Pressure Combustor Facility (B6 Combustor)

Unconventional gas supplies, like shale gas, are expected to grow which will make U.S. natural gas composition more variable and the composition of fuel sources may vary significantly from existing domestic natural gas supplies. The effect of gas composition on combustion behavior is of interest to allow end-use equipment to accommodate the widest possible gas composition. The B6 Combustor is a high-pressure facility used at NETL to study both research and commercial gas turbine fuel injectors. B6 combustor experiments have been conducted at NETL to investigate the effect of varying fuel composition on combustion dynamics.

The CFD simulation model was first validated with available experimental data which had been collected from the pressurized single injector combustion test rig. The tests were conducted at 7.5 atm with a 589K preheated air. A propane blending facility was used to vary the site natural gas composition. The CFD simulation model predicted results were within the experimental error bar. After the CFD simulation model was validated with experimental results, a much wider range of gas composition was simulated to investigate the effect of gas composition on combustion. Another purpose of the high-pressure combustion facility is to develop a Combustion Control and Diagnostic Sensor (CCADS) to in-situ monitor the combustion phenomena which is based on the mechanisms for ion formation and electrical properties of a flame.

2.2 Ansys Fluent

The CFD model is based on the mass, momentum and energy balance equations with some other constitutive equations such as the equation of state to calculate the gas phase density.

$$\rho_g = \frac{PM_{wg}}{RT_g}$$

Equation 2-1

Continuity equation:

$$\frac{\partial}{\partial t}(\rho_g) + \nabla \cdot (\rho_g \overrightarrow{u_g}) = \sum_{i=1}^{N} R_{gi}$$

Equation 2-2

Momentum equation:

$$\frac{\partial}{\partial t}(\rho_g \overrightarrow{u_g}) + \nabla \cdot (\rho_g \overrightarrow{u_g}\overrightarrow{u_g}) = -\nabla P_g + \nabla \cdot \overline{\overline{\tau_g}} + \rho_g \vec{g}$$

Equation 2-3

$$\overline{\overline{\tau_g}} = \mu_e \left[\left(\nabla \overrightarrow{u_g} + \nabla \overrightarrow{u_g}^T \right) - \frac{2}{3} \nabla \cdot \overrightarrow{u_g} I \right]$$

Equation 2-4

As the gas phase is composed of several components such as the O_2, N_2, CO_2 etc., the species transport equation

$$\frac{\partial}{\partial t}(\rho_g Y_i) + \nabla \cdot (\rho_g \overrightarrow{u_g} Y_i) = -\nabla \cdot \vec{J_\iota} + R_{gi}$$

Equation 2-5

$$\vec{J_\iota} = -\left(\rho_g D_{im} + \frac{\mu_t}{Sc_t} \right) \nabla Y_i$$

Equation 2-6

Where:

- $\boldsymbol{\rho_g}$ = gas density (kg/m^3).
- P = operating gas pressure (outlet pressure).
- $\boldsymbol{M_w}$= average molecular weight of gas.
- R = universal gas constant (8.314 J/mol/K)
- T = gas phase temperature.
- $\boldsymbol{\overrightarrow{u_g}}$ = gas phase velocity in x, y and z direction respectively.
- $\boldsymbol{\overline{\overline{\tau_g}}}$ = stress tensor.
- $\boldsymbol{Y_i}$ = fraction of species i in the gas phase.
- $\boldsymbol{R_{gi}}$ = net rate of production of species i by chemical reaction.

- $\vec{g}$ = gravity.
- $\vec{J_i}$ = diffusion flux of species due to the gradients of concentration.
- D_{im} = mass diffusion coefficient for species j in the mixture.
- μ_e = effective viscosity ($\mu_e = \mu_t + \mu$).
- μ_t = turbulent viscosity ($\mu_t = \rho C_\mu \frac{k^2}{\varepsilon}$).

**Subscript "g" means the gas phase, subscript "i" means the species i.

Incompressible ideal gas law is used to calculate the gas density as the temperature changes a lot but the pressure changes little. The operating pressure is the pressure at the coal boiler outlet. For turbulent flows, the molecular viscosity is much smaller than the turbulent viscosity.

2.2.1 Turbulence Model

Realizable k-ε with standard wall functions as realizable k-ε model is more suitable for flow with swirling [4] [5] [6] [7] [8] [9] [10]. The realizable k-ε model differs from the standard k-ε model in two important ways: the realizable model contains an alternative formulation for the turbulent viscosity. A modified transport equation for the dissipation rate, has been derived from an exact equation for the transport of the mean-square vorticity fluctuation. The term "realizable" means that the model satisfies certain mathematical constraints on the Reynolds stresses, consistent with the physics of turbulent flows. Neither the standard k-ε model nor the RNG k-ε model is realizable.

The difference between the realizable k-ε model and the standard and RNG k-ε models is that C_μ is no longer constant but a function of the mean strain and rotation rates, the angular velocity of the system rotation, and the turbulence fields.

k is the turbulence kinetic energy

$$\frac{\partial}{\partial t}(\rho_g k) + \nabla \cdot (\rho_g k \overrightarrow{u_g}) = \nabla \cdot \left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\nabla k\right] + G_k + G_b - \rho\varepsilon - Y_m + S_k$$

Equation 2-7

ε is the dissipation rate of turbulence kinetic energy

$$\frac{\partial}{\partial t}(\rho_g \varepsilon) + \nabla \cdot (\rho_g \varepsilon \overrightarrow{u_g}) = \nabla \cdot \left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\nabla \varepsilon\right] + \rho C_1 S\varepsilon - \rho C_2 \frac{\varepsilon^2}{k + \sqrt{\upsilon\varepsilon}} + C_{1\varepsilon}\frac{\varepsilon}{k} C_{3\varepsilon} G_b + S_\varepsilon$$

Equation 2-8

Where:

- G_k is generation of turbulence kinetic energy due to the mean velocity gradients.

- $\boldsymbol{G_b}$ is the generation of turbulence kinetic energy due to buoyancy.
- $\boldsymbol{Y_m}$ is the contribution of the fluctuating dilatation in compressible turbulence to the overall dissipation rate.
- $\boldsymbol{C_2}$ (1.9) and $\boldsymbol{C_{1\varepsilon}}$ (1.44) are constants. $\boldsymbol{\sigma_k}$ (1.0) and $\boldsymbol{\sigma_\varepsilon}$ (1.2) are the turbulent Prandtl numbers for k and ε respectively.
- $\boldsymbol{S_k}$ and $\boldsymbol{S_\varepsilon}$ are user-defined source terms.

2.2.2 CFD Reaction Eddy-Dissipation Model

Most fuels are fast burning, and the overall rate of reaction is controlled by turbulent mixing. The net rate of production of species due to reaction $R_{i,r}$, is given by the smaller (that is, limiting value) of the two expressions below:

$$R_{i,r} = \upsilon'_{i,r} M_{w,i} 4.0 \rho \frac{\varepsilon}{k} min_R \left(\frac{Y_R}{\upsilon'_{R,r} M_{w,R}} \right)$$

Equation 2-9

$$R_{i,r} = \upsilon'_{i,r} M_{w,i} 2.0 \rho \frac{\varepsilon}{k} \frac{\sum_P Y_P}{\sum_j^N \upsilon''_{j,r} M_{w,j}}$$

Equation 2-10

Where:

- $\boldsymbol{Y_P}$ = mass fraction of any product species, P
- $\boldsymbol{Y_R}$ = mass fraction of a reactant, R
- $\boldsymbol{\upsilon'_{i,r}}$ = stoichiometric coefficient for reactant i in reaction r
- $\boldsymbol{\upsilon''_{i,r}}$ = stoichiometric coefficient for product i in reaction r

2.2.3 CFD Heat Transfer Model

$$\frac{\partial}{\partial t}(\rho_g H) + \nabla \cdot (\rho_g H \overrightarrow{u_g}) = \nabla \cdot \left(\frac{k_t}{C_p} \nabla H \right) + S_h$$

Equation 2-11

$$H_j = \int_{T_{ref}}^{T} C_{p,j} dT + H_j^0(T_{ref,j})$$

Equation 2-12

H is the total enthalpy defined as $H = \sum_j Y_j H_j$ where Y_j is the mass fraction of species j and H_j is the enthalpy of species j. The heat capacity $C_{p,j}$ is defined as a function of temperature for each species. When the radiation model is being used, the source term S_h includes radiation source terms.

Both conduction and convection require matter to transfer heat. Radiation is a method of heat transfer that does not rely upon any contact between the heat source and the heated object. Thermal radiation (often called infrared radiation) is a type of electromagnetic radiation (or light). Radiation is a form of energy transport consisting of electromagnetic waves traveling at the speed of light. No mass is exchanged, and no medium is required for radiation.

Radiation Model

Discrete Ordinates (DO) model is used as DO model needs more computational resource than other radiation model but the DO model is more complete [4] [11] [12] [5] [6] [7] [13] [9] [10].

DO is recommended by Fluent. Emissivity of gas can be calculated from Weighted-sum-of-grey-gases model (WSGGM), which has been widely used in computational fluid dynamics and reached good balance between calculating efficiency and accuracy [14] [15]. WSGGM assumed that the emissivity of flue gas was decided by local temperature and partial pressure of gas species

2.3 Machine Learning

Artificial intelligence and machine learning are widely known technologies that aim to teach machines to learn from input data. Machine learning algorithms can be classified mainly into Supervised and Unsupervised learning algorithms. Supervised learning algorithms learn a function that, given a sample of data and desired outputs, best approximates the relationship between input features and output (also known as ground truth) observable in the data. Unsupervised learning algorithms, on the other hand, do not have labeled outputs; so, the goal is to infer the natural structure or underlying pattern present within a set of data points.

2.3.1 Fuzzy Clustering

Clustering is a form of unsupervised learning technique which involves assigning data points (or objects) to clusters (groups) such that points in the same cluster are as similar as possible. The simplest form of cluster analysis is the hard clustering in which a data point exclusively belongs to a single cluster. Fuzzy clustering is useful in avoiding the arbitrariness of assigning an object or data point to only one cluster when it may be close to several. In fuzzy clustering (also called soft clustering), every object or data point belongs to every cluster with a membership weight that is between 0 (absolutely does not belong) and 1 (absolutely belongs) [16]. Cluster membership weights for any data point must sum up to 1.

In this project, the skfuzzy package from a popular open-source machine learning library called Scikit-learn is used in performing fuzzy clustering tasks.

2.3.2 Artificial Neural Networks (ANN)

One of the most common supervised learning algorithms is the Artificial Neural Network (ANN). An ANN is a simple mathematical computational algorithm that is capable of learning from input data (machine learning) as well as discovering patterns (pattern recognition) [17].

ANN is biologically inspired by the interconnections that take place between neurons in a human brain. Neurons carry and pass information from one neuron to another via synapse. The architecture of artificial neural networks consists of an input layer, one or more hidden layers, and an output layer. The input layer contains the information provided to the neural network in the form of attributes. The hidden layer is responsible for translating the information from the input layer to the output layer by a system of weighted connections and non-linear activation functions [17]. Figure 2-1 shows a typical ANN with four input attributes, three neurons in the hidden layer and a single neuron in the output layer. The strength of information passed from one artificial neuron to another is assigned by its “weight”. Optimization of these weights is crucial in the development of a well-trained neural network.

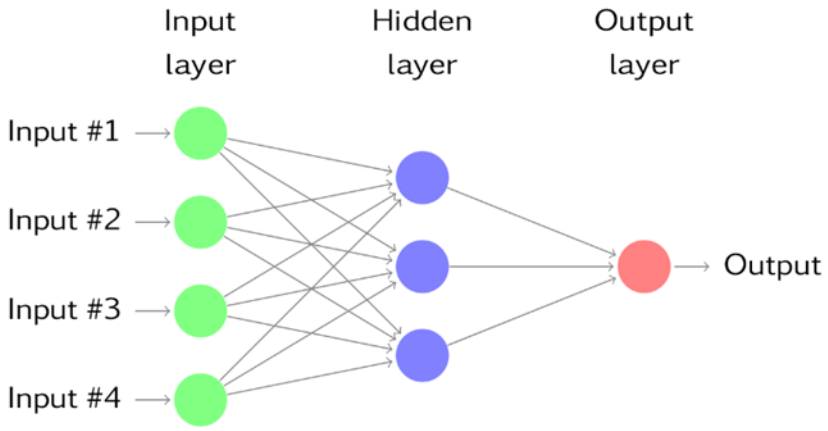


Figure 2-1: Artificial Neural Network Architecture

In this project, a machine learning library in Python called Keras is used in modeling artificial neural networks [18]. Keras is an open source high-level neural networks API written in Python and capable of running on top of TensorFlow, CNTK, or Theano. In this project, the TensorFlow backend is used.

2.3.3 Artificial Neural Network Performance Evaluation Metrics

During the ANN training process, the objective function (also called loss function) is used in updating the weights of the neurons of the neural network. The performance of a neural network is often evaluated in the context of minimizing the total error between the neural network predicted values and the actual output values (ground truth). The most commonly used objective function (especially for regression tasks) is the Mean Square Error (MSE) which is the sum of squared differences between the predicted values and the actual output values, as shown by Equation 2-13.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_{actual} - y_{predicted}\right)^2$$

Equation 2-13

In practice, while a portion of the entire dataset (called test or blind dataset) is set aside prior to training, to test the performance of the neural network on unseen data, a portion of the remaining data is usually set aside for calibration of the neural network as it is being trained. An optimization of the loss function on the calibration set helps to ensure that the neural network is not over-trained

(overfitting) or under-trained. In this project, we have taken a more unique approach to data partitioning for the development of neural networks. This approach is described in more detail in the body of this report.

There are various ways to evaluate the performance of a neural network model. In machine learning, a very common metric is the percentage error which expresses the difference between the predicted value and actual value as a percentage of the actual value as shown in Equation 2-14.

$$\%\, Error = \frac{CFD_{value} - \ Smart\ Proxy_{value}}{CFD_{value}}\ x\ 100$$

Equation 2-14

2.3.4 Data Batching

Artificial neural networks are trained using the gradient descent optimization algorithm in which the difference between the ANN predicted values and the actual values is used in estimating the error gradient. The error gradient is then used to update the model weights and the process is repeated. The estimate of the loss gradient is usually calculated based on all or a subset of the training dataset. The number of training samples used in estimating the error gradient is called batch size and it is an important hyper parameter that influences how the neural network learns. We refer to the concept of controlling the batch size as data batching.

The error gradient is a statistical estimate. The more training examples used in the estimate, the more accurate this estimate will be and the more likely that the weights of the network will be adjusted in a way that will improve the performance of the ANN model in fewer number of iterations [19]. However, computing the gradient over a very large number of examples could be very computationally expensive especially when dealing with a large amount of data. However, using too few examples from the training data could result in less accurate and noisy estimates of the gradient. Nevertheless, these noisy updates can result in faster learning and sometimes a more robust model [19]. Data batching can be used to manage the tradeoff between computational cost and neural network performance.

To manage the large amount of dataset involved (as we will show in this report) during the training of neural networks, we have used a combination of computing techniques in python programming called data generators and memory-mapping. Memory-mapping is a python computing technique used for accessing small segments of large files on disk, without reading the entire file into memory [20]. When it is not practical to load entire training dataset into the machine learning library due to memory limitation, data generators could be used to generate data in batches and continuously feed the data to the machine learning algorithm. In this project, we have memory-mapped very

large files containing training datasets and used custom built-in functions called data generators to feed the data in batches to the artificial neural networks.

2.4 Previous Work

The idea of using Artificial Intelligence in solving petroleum engineering problems was first introduced by Mohaghegh and Ameri [21] . They showed that an ANN can be used to automate the task of determining formation permeability based on geological well logs, thereby eliminating the need to perform the task repeatedly by log analysts. Mohaghegh and Ameri [21] also stated that neural network can handle far more complex tasks. Mohaghegh et al. [22] used ANN for predicting gas storage well performance after hydraulic fracture in their later investigations. This technology has been used to implement a unique approach to petroleum reservoir modeling by constructing top-down reservoir models ("TDM") that use field measurements (i.e. production history, hydraulic fracturing, well logs etc.) to predict oil and gas production from shale reservoirs [23].

Data-driven smart proxy models have been used to take advantage of the "Big Data" solutions (machine learning and pattern recognition) to develop highly accurate replicas of numerical models with very fast response time [24]. Data-driven smart proxies implement machine learning and pattern recognition techniques, using generated numerical simulation data with efforts to significantly reduce the computational footprint and the time spent to conduct large complex numerical simulation runs [25].

Boosari [26] in a study developed a smart proxy model to predict the unsteady state behavior of fluid flow resulting from wall collapse in a 2-dimensonal rectangular water tank, using dataset generated from OpenFOAM CFD simulations. The results from the study showed that a smart proxy model can predict the CFD simulation results with less than 10% error, within a significantly reduced amount of time compared to the large computational footprint of the CFD simulations. Ansari et al. [3] used AI and machine learning to construct a smart proxy model to replicate the flow and particle behavior for a gas-solid multiphase flow in a non-reacting rectangular fluidized bed. The smart proxy was developed using cell-level data generated from CFD simulation runs using MFiX CFD simulation software. The work performed by Ansari et al. [3] showed that this technology can reproduce CFD simulation results with less than 10% error within just a few seconds.

3 B6 COMBUSTOR MODEL

The objective of this chapter of the report is to provide a general overview of the end-to-end workflow involved in developing the smart proxy model for the B6 Combustor simulation models. This chapter provides a detailed description of the simulation data received including the boundary conditions, a brief description of descriptive analytical techniques used to develop the smart proxy model and a detailed description of the different machine learning algorithms used in the development process. A more detailed description of specific steps taken to develop the smart proxy model is provided in the next chapter, together with the results of the smart proxy model at each step of the development process.

3.1 B6 Combustor Problem Definition

It is known that relatively small changes in turbine engine ambient conditions and fuel composition can affect the combustion dynamics of operating engines. Combustion dynamics are a form of oscillating combustion that produces pressure oscillations at hundreds of cycles per second. If left uncontrolled, these oscillations can be very damaging, cracking metal combustion liners, triggering flashback, and producing thermal failure from enhanced heat transfer. A combination of numerical models and experimental testing have been used at NETL to investigate the effects of changing ambient conditions and fuel interchangeability (the volume fraction of propane addition to natural gas) on combustion dynamics (instabilities) [27].

The B6 combustor has been used to study the effect of fuel composition on premix turbine combustion as well as to develop a Combustion Control and Diagnostic Sensor (CCADS) to in-situ monitor the combustion phenomena. The experimental test rig has a propane blending facility which is used to vary the natural gas composition. The resultant pollutant and dynamic response from lean-premixed gas turbine systems relies heavily on adequate mixing of the fuel and air prior to reaching the reacting zones within the combustor [27]. The objective is to develop a smart proxy model capable of replicating the B6 combustor CFD simulation results of Pressure, Temperature, Nitrogen, Carbon-di-oxide and Oxygen distribution in the system, for varying composition and flow rates of air and fuel. Figure 3-1 shows a picture of the combustor rig [27].

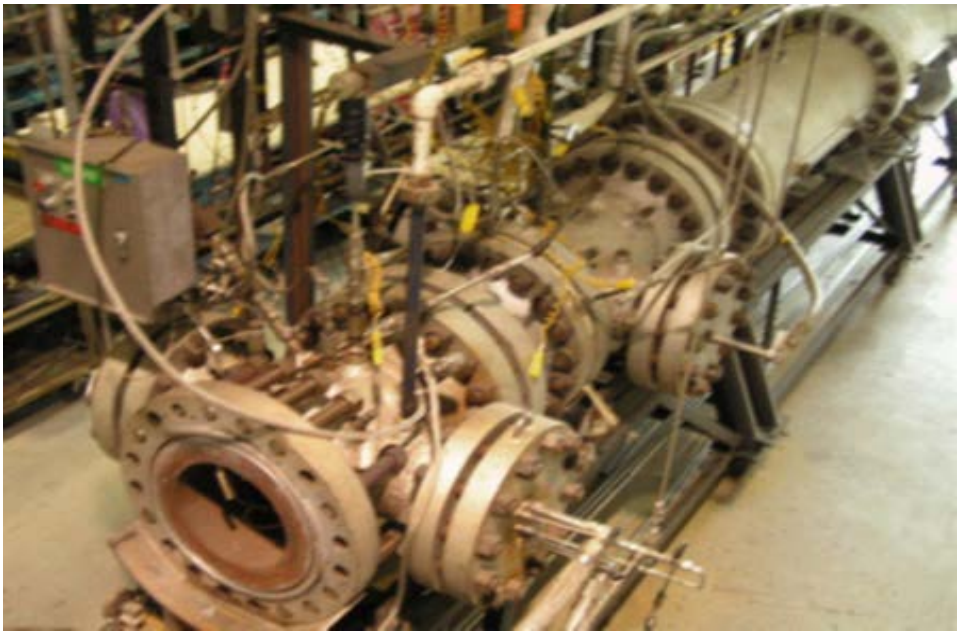


Figure 3-1: NETL's B6 Combustion Rig

3.2 B6 Combustor CFD Simulation Model

NETL's High-Pressure Combustion Facility test rig is roughly 17.8cm (7.0in) diameter combustion chamber and the length of combustion zone is 0.91m (36in). It has been used to investigate the effect of fuel interchangeability on combustor performance (the volume fraction of propane addition to the natural gas) by testing different fuel-air equivalence ratios (0.42~0.48) on combustion performances.

The B6 CFD simulation model is 17.8cm in diameter, and 0.91m in length, with a total of 9,291,712 grid cells. Figure 3-2 shows the schematic of the simulation model. The B6 combustor is a steady state single phase (gas only) combustion model. Pre-heated air and pre-mixed fuel (main and pilot fuel at room temperature) is injected through the inlet. A set of nozzles at the inlet provide swirling effect for turbulent mixing of the gas as it enters the combustion chamber. A total of 8 CFD simulation runs were initially generated with which the smart proxy model was developed. An additional 2 simulation runs were provided as a blind test following the development of the smart proxy model. CFD simulation runs were generated using the Ansys FLUENT software. Table 3-1 and Table 3-2 show the percent fuel composition, and other boundary conditions respectively for the initial 8 CFD simulation runs. Four of the eight simulation runs have methane composition ranging from 88.96% to 89.13%, these were tagged as the "Base" composition simulation cases. The other four cases were blended with more propane and have methane composition ranging from 84.15% to 84.425%, and these were tagged as the "Blend" cases. The gas composition of air at inlet and the fuel temperature at inlet remained the same across all cases.

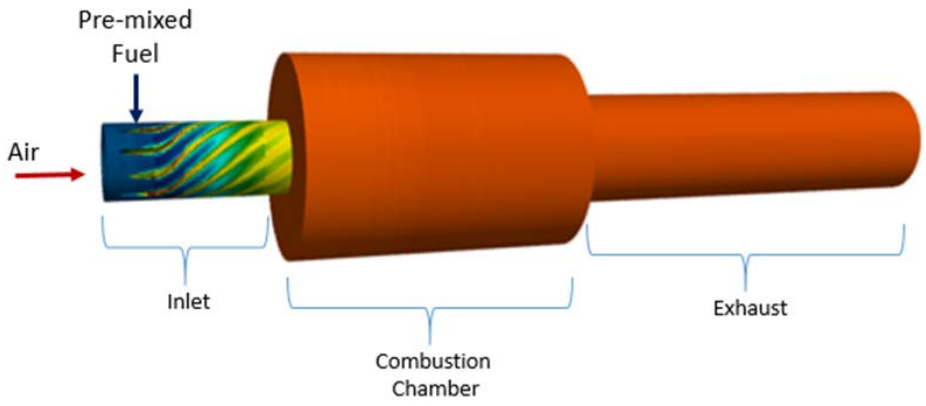


Figure 3-2:Schematic of the B6 Combustion Simulation Model

Table 3-1: Boundary Conditions for Development Cases – Inlet Fuel Composition

Case Number	1	2	3	4	5	6	7	8
	Base vol.%	Blend vol.%	Base vol.%	Base vol.%	Blend vol.%	Blend vol.%	Base vol.%	Blend vol.%
CH_4	89.025	84.425	89.113	89.130	84.283	84.153	88.961	84.235
C_2H_6	7.727	7.488	7.636	7.622	7.450	7.510	7.728	7.500
C_3H_8	1.223	6.415	1.197	1.182	6.273	6.329	1.213	6.264
C_4H_{10}	0.468	0.455	0.471	0.472	0.457	0.462	0.479	0.463
C_5H_{12}	0.130	0.125	0.132	0.132	0.126	0.127	0.135	0.128
N_2	1.061	1.006	1.067	1.075	1.047	1.060	1.106	1.051
CO_2	0.366	0.356	0.384	0.387	0.364	0.358	0.378	0.359

Table 3-2: Other Boundary Conditions for Development Cases

			1	2	3	4	5	6	7	8
Boundary Condition	**Description**	**Unit**	Base	Blend	Base	Base	Blend	Blend	Base	Blend
N2_Air_Inlet	Inlet Air Gas Composition	%	78.08	78.08	78.08	78.08	78.08	78.08	78.08	78.08
O2_Air_Inlet		%	20.95	20.95	20.95	20.95	20.95	20.95	20.95	20.95
Ar_Air_Inlet		%	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
CO2_Air_Inlet		%	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Air_Flow_Rate	Air Flow Rate	scf/hr	66269.58	66222.6	66272.97	66315.72	66241.71	66241.83	66244.22	66253.56
Main_Fuel_Flow	Fuel Flow Rate	scf/hr	2592.92	2783.134	2948.073	2592.51	2783.444	2448.18	2947.512	2448.165
Pilot_Fuel_Flow			137.099	146.068	155.573	136.815	146.725	129.074	155.804	128.781
Air_Inlet_Temp	Air Temp. at Inlet	deg. K	589.046	588.824	588.615	588.543	588.666	588.579	588.579	588.579
Main_Fuel_Inlet_Temp	Fuel Temp. at Inlet	deg. K	293.15	293.15	293.15	293.15	293.15	293.15	293.15	293.15
Pilot_Fuel_Inlet_Temp			293.15	293.15	293.15	293.15	293.15	293.15	293.15	293.15
Pressure_Outlet	Pressure at Outlet	kPa	95.4662	95.4687	95.5355	95.4484	95.5262	95.4461	95.5731	95.5846

Table 3-3 and Table 3-4 show the percent fuel composition, and other boundary conditions respectively for the additional 2 CFD simulation runs. One of the simulation runs is a "Base" case and the other a "Blind" case. These two simulation runs are sometimes referred to as "Extra Base" and "Extra Blind" respectively in this report. The gas composition of air at inlet and the fuel temperature at inlet used are the same as was used in the development simulation case runs.

Table 3-3: Boundary Conditions for Blind Validation Cases – Inlet Fuel Composition

	9	10
Case Number	**Blind Validation Run**	**Blind Validation Run**
	Base vol.%	**Blend vol.%**
CH_4	89.122	84.225
C_2H_6	7.600	7.600
C_3H_8	1.200	6.200
C_4H_{10}	0.471	0.468
C_5H_{12}	0.132	0.130
N_2	1.102	1.009
CO_2	0.373	0.368

Table 3-4: Other Boundary Conditions for Blind Validation Cases

			9	10
Boundary Condition	**Description**	**Unit**	**Blind Base**	**Blind Blend**
N2_Air_Inlet	Inlet Air Gas Composition	%	78.08	78.08
O2_Air_Inlet		%	20.95	20.95
Ar_Air_Inlet		%	0.93	0.93
CO2_Air_Inlet		%	0.04	0.04
Air_Flow_Rate	Air Flow Rate	scf/hr	66280	66240
Main_Fuel_Flow	Fuel Flow Rate	scf/hr	2750	2599
Pilot_Fuel_Flow			142	135
Air_Inlet_Temp	Air Temp. at Inlet	deg. K	588.772	588.66
Main_Fuel_Inlet_Temp	Fuel Temp. at Inlet	deg. K	293.15	293.15
Pilot_Fuel_Inlet_Temp			293.15	293.15
Pressure_Outlet	Pressure at Outlet	kPa	95.5	95.5

The simulation results (solution data) generated from Fluent for each CFD run includes cell-level distribution of five attributes of interest: Pressure, Temperature, Nitrogen, Carbon-dioxide and Oxygen.

3.3 B6 Smart Proxy Development Overview

Development of the smart proxy model for the Simplified B6 Combustor model was carried out in three stages with multiple tasks performed in each stage. The general overview of these stages is shown in Figure 3-3 below. The first stage involves performing data quality check and building a data visualization tool for the project. In the second stage, a detailed exploratory and descriptive analysis of the data was performed, and the predictive model developed in the last stage.

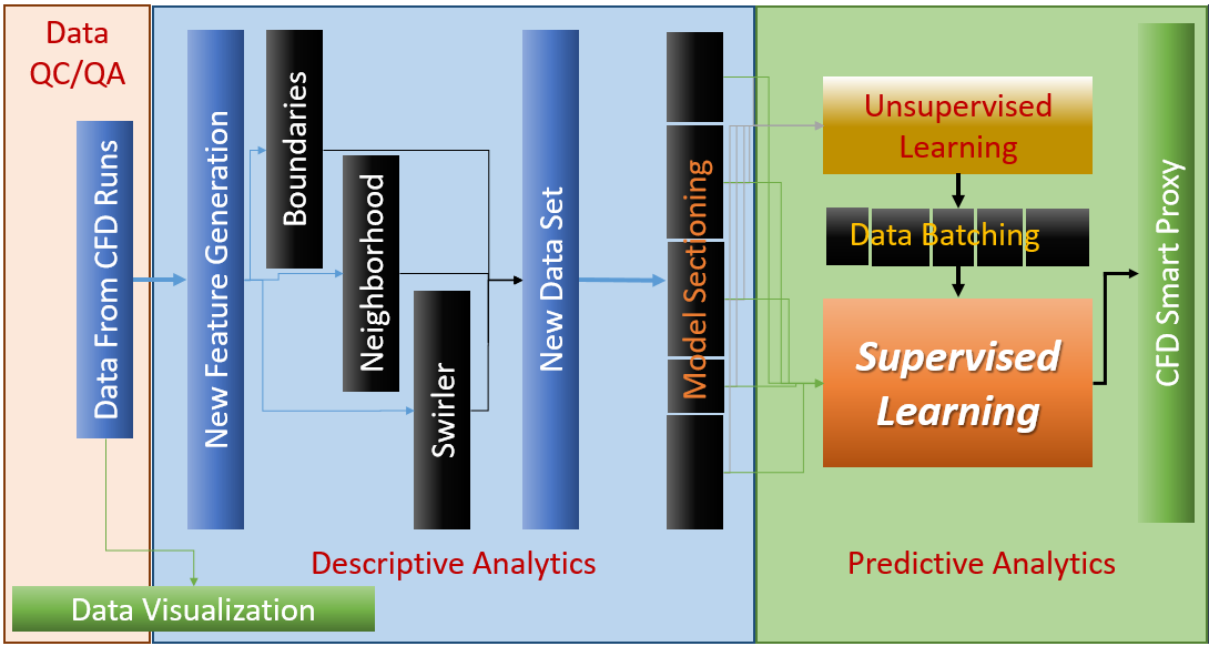


Figure 3-3: Smart Proxy Development Framework for the B6 Combustion Model

It is important to note that the descriptive and predictive modeling tasks of the framework were carried out in multiple steps, such that the modeling approach was continuously refined based on the resulting performance of the smart proxy model at each step. At each step, new features were generated and added to the database. A summary of the features that were added at each modeling step is provided in the Descriptive Analytics section of this chapter (3.3.3). A more detail description of these features with the corresponding smart proxy model results at each step is provided in the next chapter.

3.3.1 Data Received from CFD Simulation Runs

In addition to the B6 model boundary condition information described in Section 3.2, other data received from NETL include the solution data and model geometry data. The solution data contains cell level information on the distribution of each of the five transport variables of interest. The model geometry data contained the cell id, cell center coordinates (x, y, z), adjacent cell ids (i.e. ID of neighboring cells), and node ids of each cell in the simulation model. Data received for each simulation run contained a total of 21 files in **.txt* file format with a total hard disk storage size of about 1.7 GB per simulation run. The boundary condition data was received in a single **.csv* file.

3.3.2 Data Visualization Tool

In order to check the quality of data received and for the purpose of validating the results of the B6 Smart Proxy model to be developed, a python script was developed to transform the data received (originally stored in **.txt* file format) into a file format in which the model results can be visualized in three dimensions. As illustrated in Figure 3-4 below, the data was processed into **.vtk* (Visualization Tool Kit) format files which were then imported into ParaView for visualization. ParaView is an open-source data analysis and visualization application which allows for 3D interactive data exploration.

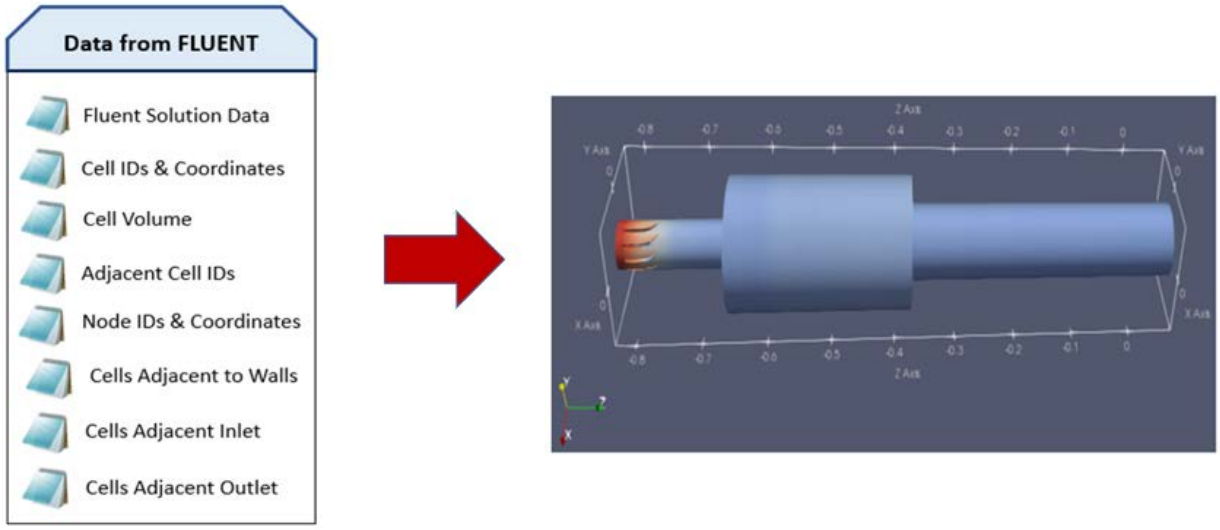


*Figure 3-4: ParaView image generated from *.vtk files built from CFD Simulation Data*

The B6 Combustor model simulation mesh contains a total of 9,291,712 elements (i.e. cells) with about 8 million tetrahedral cells and approximately 950,000 wedge cells. Figure 3-5 provides a brief description of these cell types in terms of the number of faces and points.

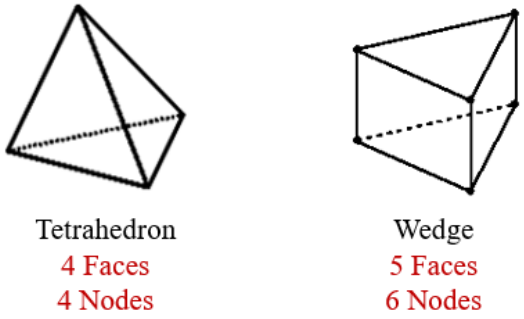


Figure 3-5: B6 Combustor Simulation Mesh Cell Types

Based on the entire simulation dataset collected from Fluent and the model boundary conditions described in Section 3.2, a structured database was generated for each CFD simulation case (for a total of 8 development cases). The following provides a description of the data received for each case run as shown in Figure 3-4.

a. Fluent Solution Data: These set of files contain the output of CFD simulation calculations performed by Fluent for the five different transport variables of interest, at the cell level: Pressure (P), Temperature (T), Oxygen concentration (O_2), Nitrogen concentration (N_2) and Carbon-dioxide concentration (CO_2).

b. Cell IDs and Coordinates: These files contain the ID and cell center coordinates (x, y, z) for each element (i.e. focal cell) in the mesh.

c. Cell Volume: File contains the volume of each focal cell in the mesh.

d. Adjacent Cell IDs: These files contain the ID of face-bounding (adjacent by face) cells for every focal cell in the mesh.

e. Node IDs and Coordinates: These files contain the ID and coordinates of points (nodes) defining each focal cell in the mesh.

f. Cells Adjacent to Walls: These files contain the ID of cells that are adjacent to the external walls of the model.

g. Cells Adjacent to Inlet: These files contain the ID of cells that are adjacent to the inlet wall.

h. Cells Adjacent to Outlet: These files contain ID of cells that are adjacent to the outlet wall.

3.3.3 Descriptive Analytics

Machine learning algorithms cannot work without data. Little to nothing can be achieved if there are too few features to represent the underlying pattern in the data to a machine learning algorithm. Comprehensive descriptive analytics of the simulation model dataset was performed and more features that further represents the underlying physics and dynamics of the combustion phenomena were generated into the project database. The descriptive analytics process was completed in four iterative steps. Below is a brief summary of features added at each step. More detail information and results at each step is provided in chapter four.

In the first step of the descriptive analytics process, features added include information that represent the geometry of the cells and their distances to wall boundaries. Based on the performance of the smart proxy at this step, more features were generated and added to the database in step 2 of the process. In step 2, features representing the location of each focal cell and its neighborhood were generated into the database. This included information on the volumes of neighboring (face-bounding) cells, and information indicating if the focal cell was bounding a wall or not. In a realistic combustor, the overall rate of reaction is controlled by turbulent mixing. Mixing does not only occur as a result of a jet crossflow of the preheated air and fuel at the inlet but typically through some additional mechanism such as swirling flow which increases the turbulence intensity [27]. Features generated in step 3 of the process included a representation of this swirling effect on the combustion process. These include distances of each focal cell in the system to each of the swirler nozzles (see detail in chapter four). In step 4, an unsupervised machine learning technique called fuzzy clustering was applied to the entire system in order to group the over 9 million cells into classes based on the distribution of each transport variable of interest across all development simulation cases. This technique helps to provide more information to the training algorithm regarding the behavior of the B6 combustor simulation models. More detail regarding fuzzy clustering is provided in the Predictive Analytics section (Section 3.3.4) of this chapter.

B6 Combustor Model Sectioning

The computer machine available to the research team at the time of the project had 500GB hard disk storage, and 24GB of memory. Generating dataset required to train the smart proxy from the database required a significant amount of memory larger than the computer memory that was available to the research team. In fact, in steps 1 and 2 of the descriptive analytics process, the smart proxy model was developed for a small section of the B6 Combustor model at each step. This made it possible to complete the training of the model in a shorter length of time and quickly assess the contribution of newly added features at each step of the process. Due to this limitation in compute resource (memory and speed), the B6 Combustor simulation model was divided into seven sections in step 3, based on a detailed analysis of the distribution of the transport variables across all simulation runs. Figure 3-6 shows the model sectioning and naming convention for each section. Table 3-5 shows the number of cells by section for a single CFD simulation run. For each

transport variable of interest, an artificial neural network was trained for each section of the model and these networks were later combined to form the smart proxy model for the B6 Combustor. More information is provided in Section 3.3.4 regarding the setup of neural networks that were developed.

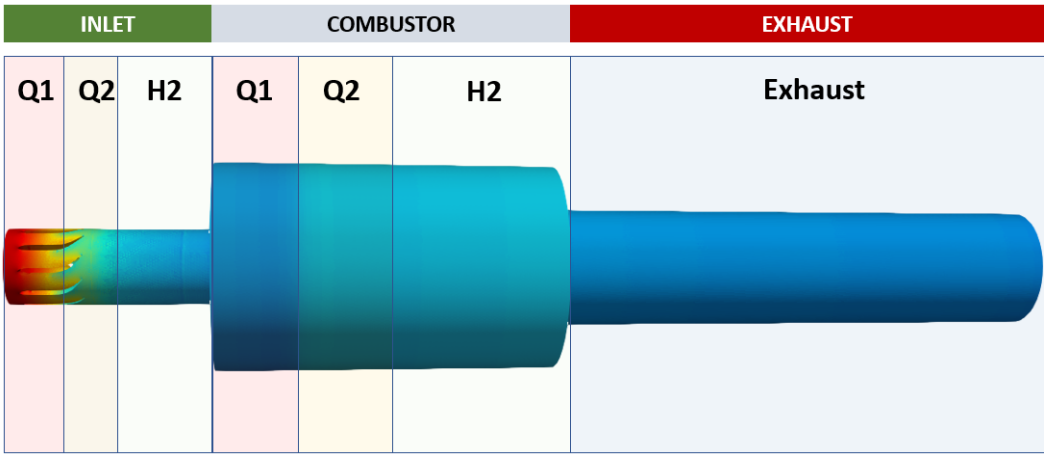


Figure 3-6: B6 Combustor Model Sectioning

Table 3-5: Total Number of Cells by Model Section

Total No. of Cells by Model Section				
Section Name	**Convention Name**	**No. Cells**	**Z-axis BGN**	**Z-axis END**
1st Quarter Inlet	Q1_Inlet	1,623,745	-0.83058	-0.78868
2nd Quarter Inlet	Q2_Inlet	1,662,119	-0.78868	-0.74678
2nd Half Inlet	H2_Inlet	344,757	-0.74678	-0.66298
1st Quarter Combustor	Q1_Combustor	962,001	-0.66298	-0.58981
2nd Quarter Combustor	Q2_Combustor	971,256	-0.58981	-0.51664
2nd Half Combustor	H2_Combustor	1,898,924	-0.516642	-0.37030
Exhaust	Exhaust	1,903,951	-0.37030	0.0762
Entire Model	CFD	9,291,712	-0.83058	0.0762

3.3.4 Predictive Analytics

This section of the report presents the details of the machine learning algorithms used in learning the underlying pattern in the dataset, and the approach taken to validate the performance of the developed smart proxy model.

Both supervised and unsupervised learning methods were used to meet the objective of the project. The unsupervised learning technique involves a cluster analysis of the dataset based on the cell-level distribution of each CFD simulation output of interest across multiple simulation runs, while the supervised learning technique involves the use of artificial neural networks to learn the pattern in the dataset.

Data Partitioning

Before any learning (that is model training) was performed (whether supervised or unsupervised), data partitioning approach was decided and carefully implemented to avoid data leakage and ensure true confidence in the performance of the smart proxy to be developed. The importance of data partitioning has been highlighted in Section 2.3.1 in the background section of this report.

Figure 3-7 illustrates the data partitioning approach used in the development of the smart proxy. We have taken a "double blind" approach. To successfully validate the performance of the proposed smart proxy, the two blind validation runs (referenced as Blind Validation Base - Case 9 and Blind Validation Blend – Case 10 in Section 3.2) were completely kept out for double blind validation of the neural networks to be developed. During the development phase (i.e. neural network training), two (Base Case 1 and Blend Case 5) out of the eight development simulation cases were used as validation to assess the performance of the neural network before adjudging it as being completely developed. So only a total of six CFD simulation runs was used in training the final smart proxy model. The smart proxy was deployed on the blind validation cases 9 and 10 only after completing all of the four development steps identified in the Descriptive Analytics section. The detail smart proxy results for the blind validation cases (9 and 10) is provided in chapter four, as well as the results of performance on the validation cases (1 and 5) at every step of the development process.

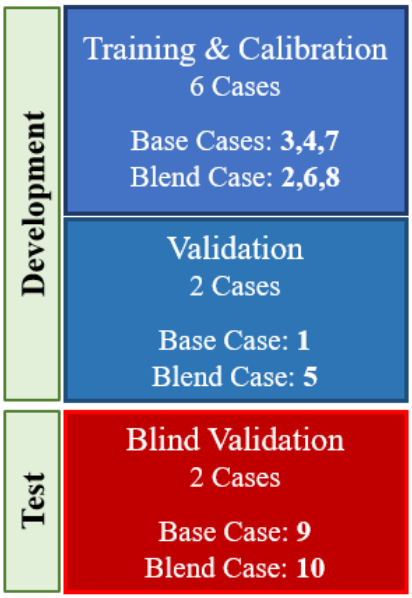


Figure 3-7: Data Partitioning

Table 3-6: Boundary Conditions for B6 Combustor Simulation Model – Fuel Composition

Case Number	1	2	3	4	5	6	7	8	9	10
	Base Vol. (%)	Blend Vol. (%)	Base Vol. (%)	Base Vol. (%)	Blend Vol. (%)	Blend Vol. (%)	Base Vol. (%)	Blend Vol. (%)	Base Vol. (%)	Blend Vol. (%)
CH_4	89.025	84.425	89.113	89.130	84.283	84.153	88.961	84.235	89.122	84.225
C_2H_6	7.727	7.488	7.636	7.622	7.450	7.510	7.728	7.500	7.600	7.6
C_3H_8	1.223	6.415	1.197	1.182	6.273	6.329	1.213	6.264	1.200	6.2
C_4H_{10}	0.468	0.455	0.471	0.472	0.457	0.462	0.479	0.463	0.471	0.468
C_5H_{12}	0.130	0.125	0.132	0.132	0.126	0.127	0.135	0.128	0.132	0.130
N_2	1.061	1.006	1.067	1.075	1.047	1.060	1.106	1.051	1.102	1.009
CO_2	0.366	0.356	0.384	0.387	0.364	0.358	0.378	0.359	0.373	0.368

Validation Case Blind Validation Case

Table 3-7: Additional Boundary Conditions for B6 Combustor Simulation Model

Case Number	1	2	3	4	5	6	7	8	9	10
	Base	Blend	Base	Base	Blend	Blend	Base	Blend	Blind Base	Blind Blend
N2 Air Inlet (%)	78.08	78.08	78.08	78.08	78.08	78.08	78.08	78.08	78.08	78.08
O2 Air Inlet (%)	20.95	20.95	20.95	20.95	20.95	20.95	20.95	20.95	20.95	20.95
Argon Air Inlet (%)	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
CO2 Air Inlet (%)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Air Flow Rate (scf/h)	66269.58	66222.59	66272.97	66315.71	66241.711	66241.8	66244.22	66253.56	66280.00	66240.00
Main Fuel Flow Rate (scf/h)	2592.92	2783.134	2948.073	2592.51	2783.444	2448.18	2947.512	2448.165	2750	2599
Pilot Fuel Flow Rate (scf/h)	137.099	146.068	155.573	136.815	146.725	129.074	155.804	128.781	142	135
Air Inlet Temperature (°K)	589.046	588.824	588.615	588.543	588.666	588.579	588.579	588.579	588.772	588.66
Main Fuel Inlet Temperature (°K)	293.15	293.15	293.15	293.15	293.15	293.15	293.15	293.15	293.15	293.15
Pilot Fuel Inlet Temperature (°K)	293.15	293.15	293.15	293.15	293.15	293.15	293.15	293.15	293.15	293.15
Pressure Outlet (psi)	95.4662	95.4687	95.5355	95.4484	95.5262	95.4461	95.5731	95.5846	95.5	95.5

Validation Case Blind Validation Case

Fuzzy Clustering

A background to fuzzy clustering is already provided in Section 2.3.1(in the background section of this report). Fuzzy clustering helps to discover distribution of patterns in datasets. When meaningful patterns are identified, it can be very valuable input for a neural network to learn. The objective of the fuzzy clustering analysis performed in this project is to find the distribution pattern of each transport variable (i.e. CFD simulation output) of interest (for example Pressure) across the simulation case runs at the cell level in the system. Clustering involves assigning the cells to clusters (groups) such that cells in the same cluster are as similar as possible, in terms of the distribution of each transport variable of interest. Table 3-8 shows an example of the inputs and output information used in the analysis for n number of cells, where n is the total number of cells in the specific model section being developed or trained. The input to the clustering algorithm is the number of clusters to group the cells into, and cell-level values of the attribute of interest across the training and calibration Cases. For example, $P_{3,2}$ refers to the pressure value in cell 3, simulation case 2. For this problem we have used three clusters and only the six training and calibration cases identified in Data Partitioning section above as input. The output is the degree of membership (membership weight) of each cell in each of the clusters (shown in cluster membership columns CM1, CM2, CM3 in Table 3-7). In practice, fuzzy clustering is often converted to an exclusive clustering by assigning each data point to the cluster in which its membership weight is highest. For this problem, we have added the exclusive clustering information (Cluster ID column in Table 3-8) to the fuzzy clustering information (CM1, CM2, CM3 columns) by assigning each cell to the cluster in which it has the highest membership. As mentioned in the Descriptive Analytics section (Section 3.3.3), clustering analysis was applied in

the last step of the development process and so, these two pieces of information (total four attributes) were included as input to every neural network model that was developed for the final B6 Smart Proxy model.

Table 3-8: Input and Output Data to the Fuzzy Clustering Algorithm

	Cell Pressure Values by CFD Case						Cluster Memberships & ID			
Cell ID	**2**	**3**	**4**	**6**	**7**	**8**	**CM1**	**CM2**	**CM3**	**Cluster ID**
1	$P_{1,2}$	$P_{1,3}$	$P_{1,4}$	$P_{1,6}$	$P_{1,7}$	$P_{1,8}$	$CM_{1,1}$	$CM_{1,2}$	$CM_{1,3}$	$f(\text{Max}(CM_{1,1},CM_{1,2},CM_{1,3})$
2	$P_{2,2}$	$P_{2,3}$	$P_{2,4}$	$P_{2,6}$	$P_{2,7}$	$P_{2,8}$	$CM_{2,1}$	$CM_{2,2}$	$CM_{2,3}$	$f(\text{Max}(CM_{2,1},CM_{2,2},CM_{2,3})$
3	$P_{3,2}$	$P_{3,3}$	$P_{3,4}$	$P_{3,6}$	$P_{3,7}$	$P_{3,8}$	$CM_{3,1}$	$CM_{3,2}$	$CM_{3,3}$	$f(\text{Max}(CM_{3,1},CM_{3,2},CM_{3,3})$
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
n	$P_{n,2}$	$P_{n,3}$	$P_{n,4}$	$P_{n,6}$	$P_{n,7}$	$P_{n,8}$	$CM_{n,1}$	$CM_{n,2}$	$CM_{n,3}$	$f(\text{Max}(CM_{n,1},CM_{n,2},CM_{n,3})$

Artificial Neural Network Setup

A neural network was trained, calibrated and validated for each model section and transport variable (a total of 35 neural networks). These neural networks were then coupled together to form the B6 smart proxy model. 80% of the samples (randomly selected) from the six training and calibration cases (base cases 3,4,7 and blend cases 2,6,8) identified in the Data Partitioning section was used to train the neural networks while the remaining 20% was used to continuously calibrate the neural network and check when to stop training in order to avoid overfitting. Neural network hyper-parameters were continuously tuned for better performance on validation cases 1 and 5.

All neural networks trained were single hidden layer, one output networks. The rectilinear activation function was used in all neural network models. The number of hidden neurons used varied from 15 to 2500 and the total number of training epochs completed for any attribute and model section varied from 15 to 100. The smart proxy model was developed for Pressure, Temperature, Carbon-dioxide, Oxygen and Nitrogen.

Data Batching

Though dividing the model into seven smaller sections helped in managing the compute resource limitation by preparing the training dataset section by section, the total number of training samples

by section was still too large to fit into the computer memory to train a neural network. This problem was addressed by using a combination of computing techniques called memory-mapping and data generators. The basis for these techniques has been described in Section 2.3.4 in the background section of this report. For each section of the model, file containing the training dataset was memory-mapped to the virtual memory of the machine and the data fed in batches to the neural network for training. Memory-mapping is a python computing technique used for accessing small segments of large files on disk, without reading the entire file into memory [20]. When it is not practical to load entire training dataset into the machine learning library due to memory limitation, data generators could be used to generate data in batches and continuously feed the data to the machine learning algorithm. Different training data batch sizes (ranging from 5,000 to about 200,000 samples) were tested in other to optimize the performance of the neural networks.

4 SMART PROXY DEVELOPMENT STEPS

In the previous chapter, we presented a general overview of the tasks performed at different stages in the project – Data quality check and visualization, Descriptive analytics, and Predictive analytics. Data received directly from the CFD simulation runs does not have enough information for the neural network to train as-is in the predictive analytics stage. Multiple steps were taken to generate features that represents the flow, reaction and heat transfer phenomenon occurring in the combustion chamber, so as to assist the neural network to learn the underlying pattern in the distribution of the transport and species variables (Pressure, Temperature, Nitrogen, Oxygen and Carbon-dioxide) throughout the combustor system.

The descriptive and predictive modeling stages of the process were carried out in multiple steps, such that the modeling approach was continuously refined based on the resulting performance of the smart proxy model following the inclusion of newly generated features at each step. These steps are summarized into four main steps listed below and are labelled based on the description of what features were generated into the database and included in the neural network training at each step.

- Step 1 – Cell Geometry and Distances to Wall Boundaries
- Step 2 – Cell neighborhood, Location and Euclidian Distances to Wall Boundaries
- Step 3 – Swirler Distances
- Step 4 – Fuzzy Clustering

Detail description of work performed and smart proxy results for each step are presented in this chapter.

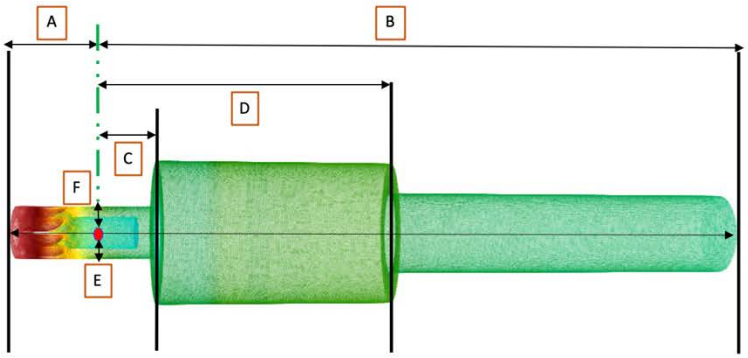
4.1 Model Development Step 1: Cell Geometry and Distances to Wall Boundaries

In gas combustion chambers such as the B6 combustor, a large proportion of the total heat flux to the walls of the combustor is by radiation from the flame [28]. Radiation exchange between surfaces in addition to their radiative properties and temperatures strongly depends on the surface's geometries, orientations and separations distance [29]. The closer a cell is to the source of heat in the combustor, the greater the intensity of radiation energy received.

In step 1 of the development process, features representing the geometry and location of each cell were generated into the database. The cell geometry attribute is represented by the number of nodes bounding each focal cell in the model, which specifies if a cell is a tetrahedron or a wedge. The cell location attribute specifies in which section of the system each focal cell is located (whether in Inlet, Combustor or Exhaust) and additionally specifies how close to the wall the focal cell is.

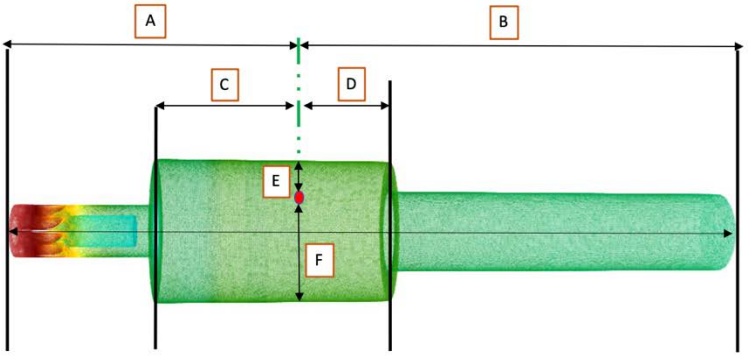
To generate these features, four boundaries were identified as Inlet, Outlet, inlet-to-combustor, combustor-to-exhaust. The inlet and outlet boundaries refer to the walls bounding the inlet and outlet of the system respectively. The inlet-to-combustor refers to the boundary where the system geometry transitions from the inlet into the combustion chamber while the combustor-to-exhaust refers to the boundary where the system geometry transitions from the combustion chamber into the exhaust section. Since the 3 main sections of the system have different diameters, the radial distance of each focal cell to the radial boundary must be accounted for. The distances of each focal cell to the nearest radial boundary and farthest radial boundary was calculated and added to the database. A total of 7 new features were added to the database in step 1.

Figure 4-1, Figure 4-2, and Figure 4-3 illustrate how the distances of cells to the boundaries, were calculated for arbitrary cells located at the inlet, combustion chamber, and exhaust respectively.



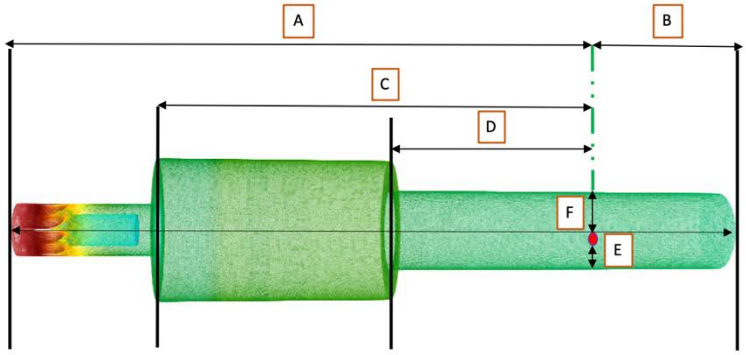
Cells Located Within Inlet	
A	Distance to Inlet
B	Distance to Outlet
C	Distance to Inlet/Combustor
D	Distance to Combustor/Exhaust
E	Closest Radial Boundary
F	Farthest Radial Boundary

Figure 4-1: Calculated distances of cells located at the inlet



Cells Located Within Combustor	
A	Distance to Inlet
B	Distance to Outlet
C	Distance to Inlet/Combustor
D	Distance to Combustor/Exhaust
E	Closest Radial Boundary
F	Farthest Radial Boundary

Figure 4-2: Calculated distances of cells located at the combustion chamber



Cells Located Within Exhaust	
A	Distance to Inlet
B	Distance to Outlet
C	Distance to Inlet/Combustor
D	Distance to Combustor/Exhaust
E	Closest Radial Boundary
F	Farthest Radial Boundary

Figure 4-3: Calculated distances of cells located at the exhaust

A 500 GB storage, 4 cores and 24 GB RAM desktop computer was used in preparing the training dataset and training the neural networks. In order to address the limitations in compute memory and speed considering the large amount of data being processed, only the inlet section of the system (as shown in Figure 4-4) was used in model development in Step 1. The total number of cells at the inlet section is approximately 2,389,668. In order to further address the memory limitation problem and increase the processing and development time, 2 million cells were randomly selected from the inlet section for processing. The same random set of cells were selected for each of the six training and calibration cases already identified in data partitioning (See Section 3.3.4). As a first trial step in the development process, only two attributes (Pressure and Carbon-dioxide) were modeled and analyzed.

Figure 4-4: Step 1 – Target section for training and development of the neural network

4.1.1 Model Development Step 1 – Model Training Information

For each attribute, a total of 23 input features were used in training a neural network for the randomly selected cells at the inlet section of the system. Table 4-1 shows the list of input attributes. 16 out of the 23 features were originally provided as part of data exported from FLUENT software. These includes focal cell location and geometry information (X, Y, and Z coordinates and cell volume), and simulation model boundary conditions (fuel composition, air and fuel flow rates and temperature, and pressure at the outlet). The other seven attributes were generated in Step 1, and these include the number of nodes on each focal cell, and distances to the boundaries described in Section 4.1.

Only the eight development cases (cases 1 through 8) were used in Step 1. The six training and calibration cases (base cases 3, 4, 7 and blended cases 2, 6, 8) identified in Section 3.3.4 were used in training the neural network. A total of 12 million samples (2 million cells from inlet section across all six cases) were used in training and calibration. Specifically, 80% of the samples (selected at random) were used in training while the remaining 20% were used as calibration samples. The base case 1 and blind case 5 were used for validating each neural network after each training attempt. The neural network hyper-parameters were continuously tuned for better performance on the validation cases 1 and 5.

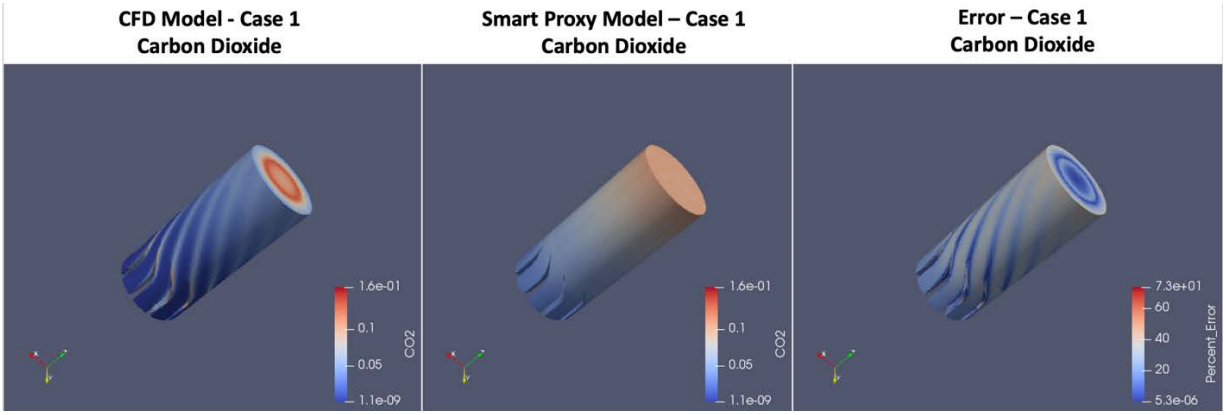
For each attribute, a single hidden-layer neural network with one output was built. The size of the training dataset (approximately 6.39GB) could not fit in memory of the desktop machine being used to train the neural networks and therefore, computing techniques mentioned in Section 2.3.4 (Data Batching) in the background section of this report, were used to address this challenge. File containing the training dataset (for an attribute) was memory-mapped to the virtual memory of the machine and the data fed in batches to the neural network for training. Different training data batch sizes were tested in other to optimize the performance of the neural networks. Batch sizes tested range from about 10,000 samples/batch to 100,000 samples/batch). The rectilinear activation function was used in the hidden and output layers. Different number of hidden neurons were also tested to optimize neural network performance. Number hidden neurons tested ranged from about 50 – 200 neurons.

Table 4-1: Training Input Attributes in Development Step 1

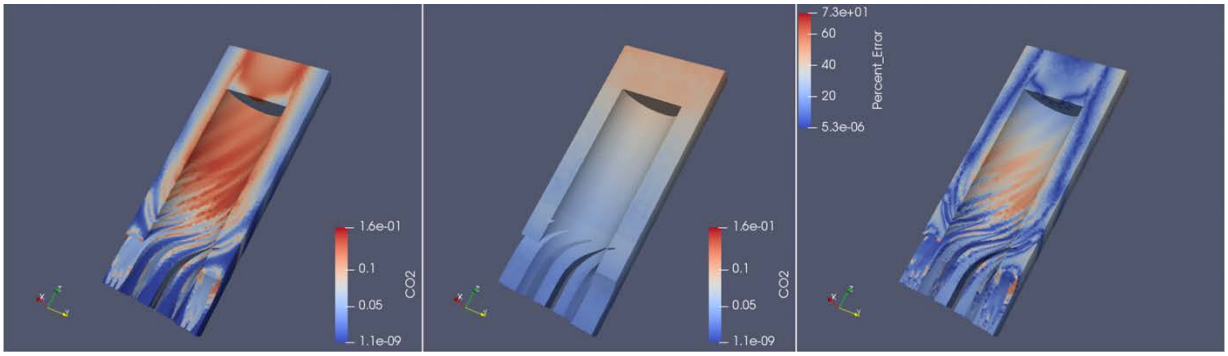
Attributes Provided from FLUENT	Attributes Generated in Step 1
X_Coord	Node_Count
Y_Coord	Dist_To_Inlet
Z_Coord	Dist_To_Outlet
volume	Dist_To_Inlet_Combustor
CH4_Inlet	Dist_To_Combustor_Exhaust
C2H6_Inlet	Closest_Radial_Boundary
C3H8_Inlet	Farthest_Radial_Boundary
C4H10_Inlet	
C5H12_Inlet	
N2_Inlet	
CO2_Inlet	
Air_Flow_Rate	
Main_Fuel_Flow	
Pilot_Fuel_Flow	
Air_Inlet_Temp	
Pressure_Outlet	
Total = 16	**Total = 7**

4.1.2 Model Development Step 1 – Presentation of Results

In this section, the smart proxy results for the distribution of carbon-dioxide in the inlet section for the two validation cases (Base Case 1 and Blend Case 5) are presented. Two figures are displayed in which the first figure shows the distribution of carbon-dioxide in the exterior of the inlet section while the second figure shows a cross-section of the inlet when cut in half. The second figure shows the distribution of carbon-dioxide in the interior parts of the inlet section. In both figures, the first image (on left) shows the result from the CFD simulation model as obtained from FLUENT, followed by the results of the smart proxy model (middle) while the last image (on right) is the percent error plot comparing the CFD simulation result with the smart proxy result. For the carbon-dioxide and other gas species, percent error is calculated by taking the absolute difference between the CFD simulation result and the smart proxy result.

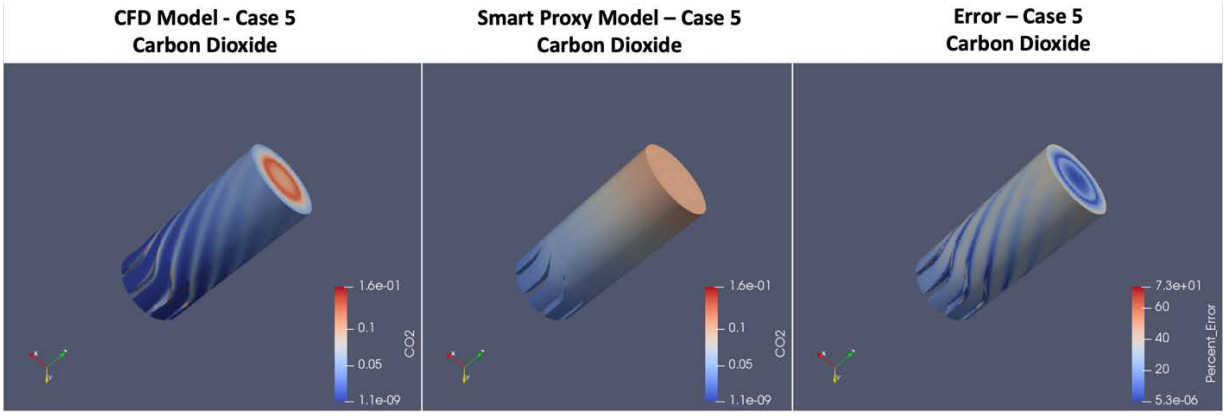


Inlet Exterior

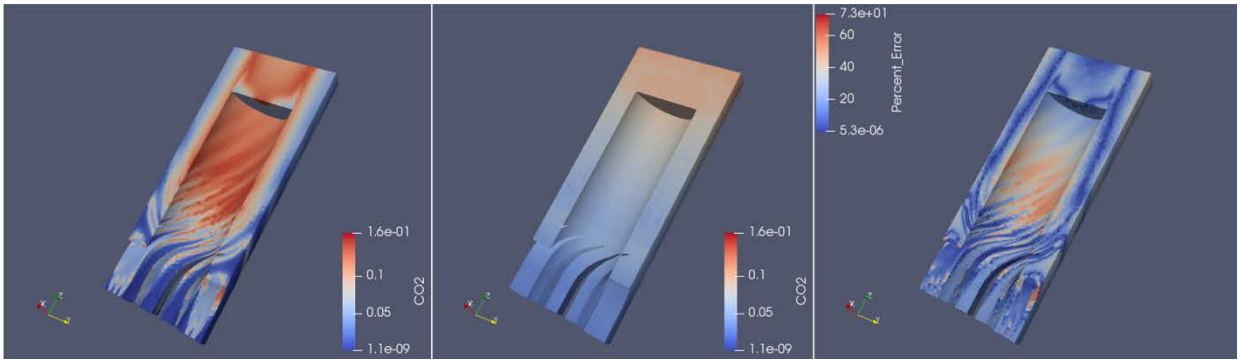


Inlet Interior

Figure 4-5: Step 1 Smart Proxy Results for Validation Case 1 – Carbon Dioxide



Inlet Exterior



Inlet Interior

Figure 4-6: Step 1 Smart Proxy Results for Validation Case 5 – Carbon Dioxide

4.2 Model Development Step 2: Cell neighborhood, Location and Euclidian Distances to Wall Boundaries

Results obtained in Step 1 showed that there was still not enough information for the neural network to learn the underlying pattern in the system. In FLUENT software, as with many other numerical simulation applications, the value of a dynamic variable or properties in any given grid or cell is mostly impacted by the value of the variable in neighboring cells. In addition to already existing features, more features further representing the location of each cell and its neighborhood were therefore generated into the database.

In Step 1, the calculated distances to the inlet, inlet-combustor, combustor-exhaust and outlet boundaries were straight horizontal distances. In order to further communicate focal cell locations to the neural network, the Euclidean distances to a fixed point (center) on these boundaries were calculated and added to the database. Figure 4-7 is an illustration of how the Euclidean distance was calculated for an arbitrary cell at position x,y,z (represented by the yellow circle) to the center (represented by the red circle) of the inlet-combustor boundary. The closest and farthest radial distances were calculated in Step 1, the ratio of these distances was calculated and added to the database in Step 2.

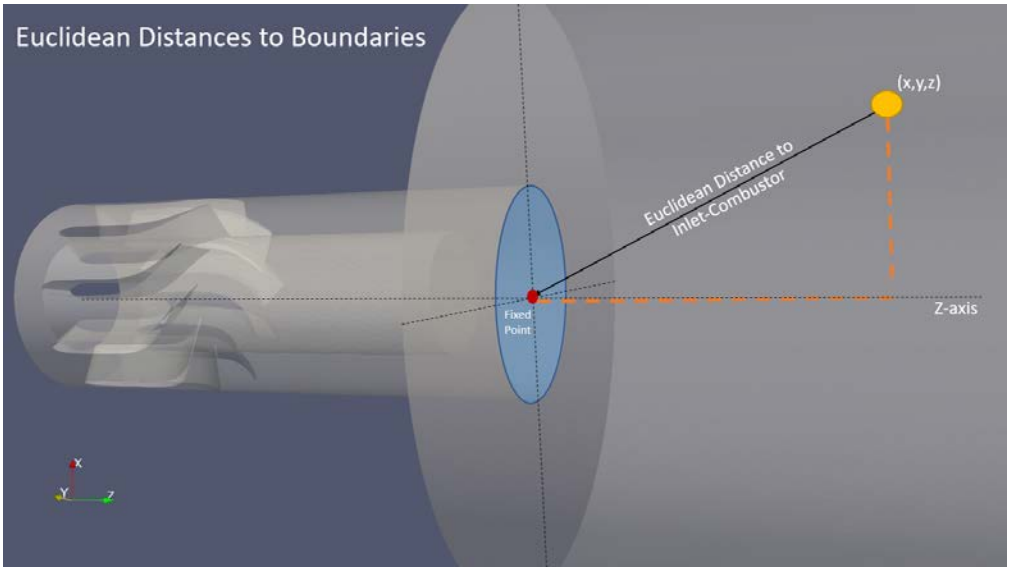


Figure 4-7: Euclidean Distances to a Fixed Point on a given Boundary

In order to represent focal cell neighborhood information, the volume of up to any four face surrounding cells were included for each focal cell. In addition to the volume of surrounding cell attributes, the total number of adjacent (face-bounding) cells were added and features representing whether a focal cell is bounding the inlet, outlet or any surrounding wall were also added to the database. A total of thirteen new features (as shown in Table 4-2) were generated and added to the database in Step 2.

Recall that random sampling of cells in the inlet section was necessary to address the compute resource limitation challenge in order to train the neural networks in Step 1. In order to avoid having to train with a sampled dataset, it was decided to perform the next development analysis with a section of the model where a sizeable portion of the system can be analyzed without exceeding the memory capacity of the machine being used. The first half combustion chamber portion of the system has approximately 2 million cells while it has about the same length along the Z-axis as the inlet section. As shown in Figure 4-8, the first half of the combustion chamber was therefore selected for analysis in Step 2 of the development process.

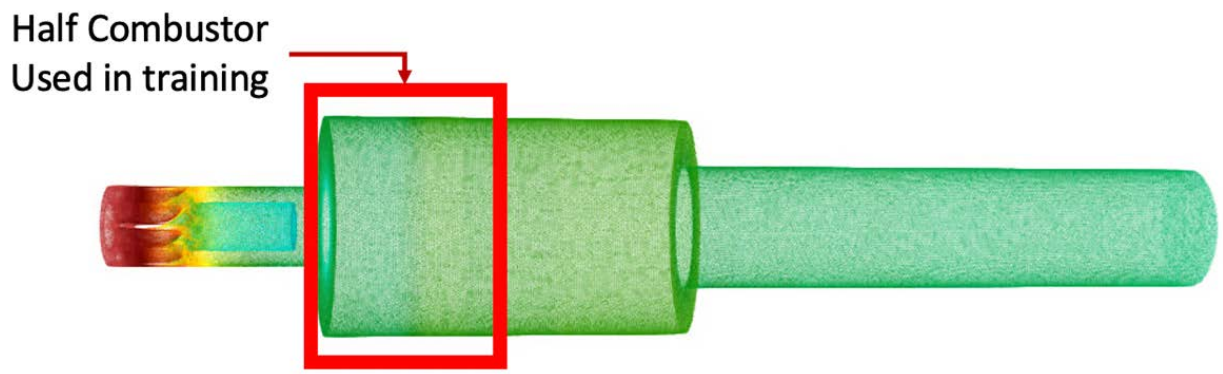


Figure 4-8: Step 2 – Target section for training and development of the neural network

4.2.1 Model Development Step 2 – Model Training Information

In Step 2 of the development process, only the pressure distribution was modeled. The data partitioning approach is as described in the model development Step 1. Six cases (2, 3, 4, 6, 7, 8) were used for training and calibration of the neural network while 2 cases (base case 1 and blend case 5) were used as validation after each training attempt.

The total number of training and calibration samples was 12,034,230 (6 cases x 2,005,705) while 4,011,410 (2 cases x 2,005,705) samples was used to validate the neural network. As mentioned in Step 1, the data batching technique was used at this step to feed data in batches to the training algorithm, and a single hidden-layer neural network with one output was built. A total of 36 input attributes were used in training the neural network as shown in Table 4-2.

Table 4-2: Training Input Attributes in Development Step 2

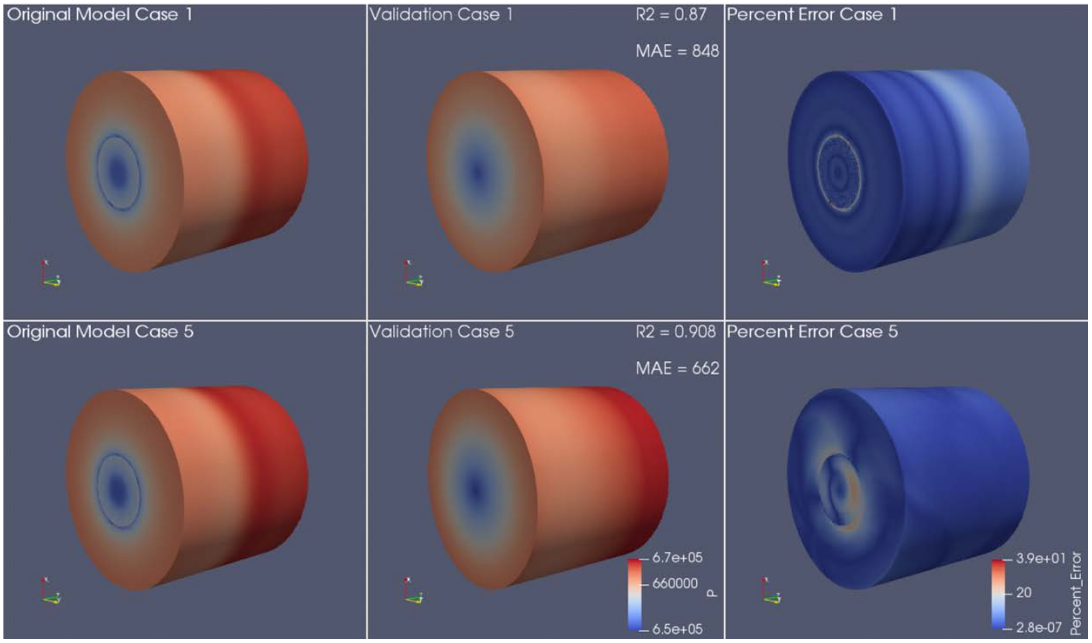
Attributes Provided from FLUENT	Attributes Generated in Step 1	Attributes Generated in Step 2
X_Coord	Node_Count	Euclidean_Dist_Inlet
Y_Coord	Dist_To_Inlet	Euclidean_Dist_Inlet_Comb
Z_Coord	Dist_To_Outlet	Euclidean_Dist_Comb_Exhaust
volume	Dist_To_Inlet_Combustor	Euclidean_Dist_Outlet
CH4_Inlet	Dist_To_Combustor_Exhaust	Radial_Boundary_Ratio
C2H6_Inlet	Closest_Radial_Boundary	Adjacent_Cell_Volume_1
C3H8_Inlet	Farthest_Radial_Boundary	Adjacent_Cell_Volume_2
C4H10_Inlet		Adjacent_Cell_Volume_3
C5H12_Inlet		Adjacent_Cell_Volume_4
N2_Inlet		Total_Adjacent_Cells
CO2_Inlet		Cell_Adjacent_Inlet
Air_Flow_Rate		Cell_Adjacent_Outlet
Main_Fuel_Flow		Cell_Adjacent_Wall
Pilot_Fuel_Flow		
Air_Inlet_Temp		
Pressure_Outlet		
Total = 16	**Total = 7**	**Total = 13**

4.2.2 Model Development Step 2 – Presentation of Results

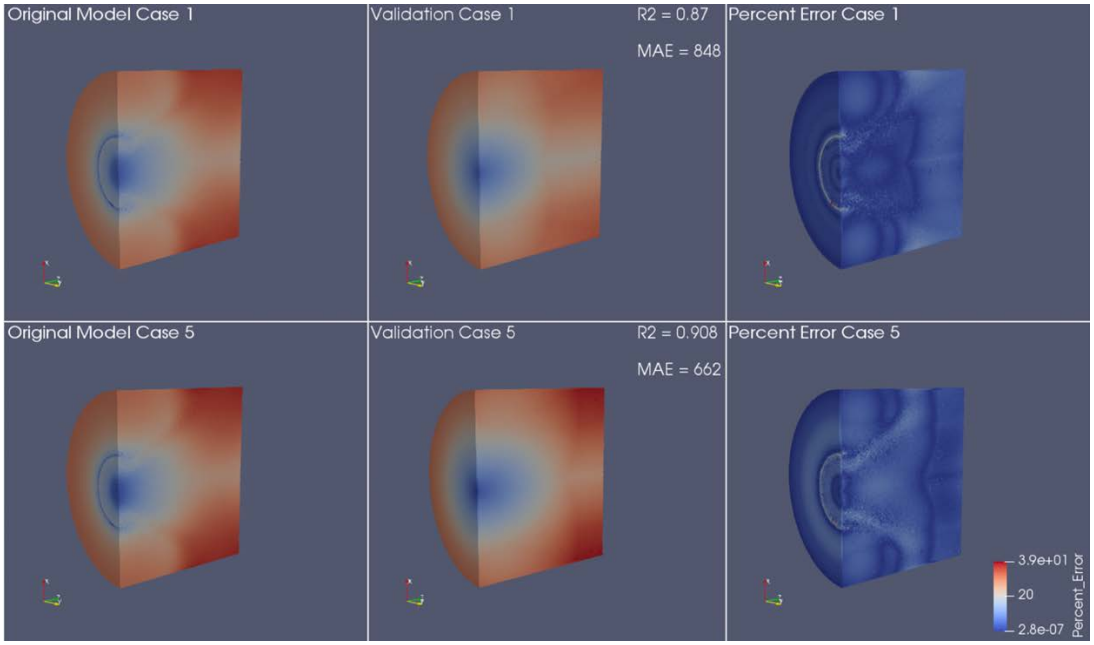
The following figures show the results of the smart proxy model developed in Step 2. The results presented include the pressure distribution for the validation cases (1 and 5) and two of the training cases (2 and 6).

Each figure contains a total of 12 images: the first six images show an exterior view of the model for two cases while the following six images show an interior view of the model for the same two cases. Moreover, the left-hand side images represent the actual CFD model simulation generated from Ansys Fluent, the middle images represent the smart proxy model, and the right-hand side images show the error difference between the actual CFD model and the smart proxy model.

Results for the remaining attributes of interest were not generated as the objective was to evaluate the performance of the developed neural network.

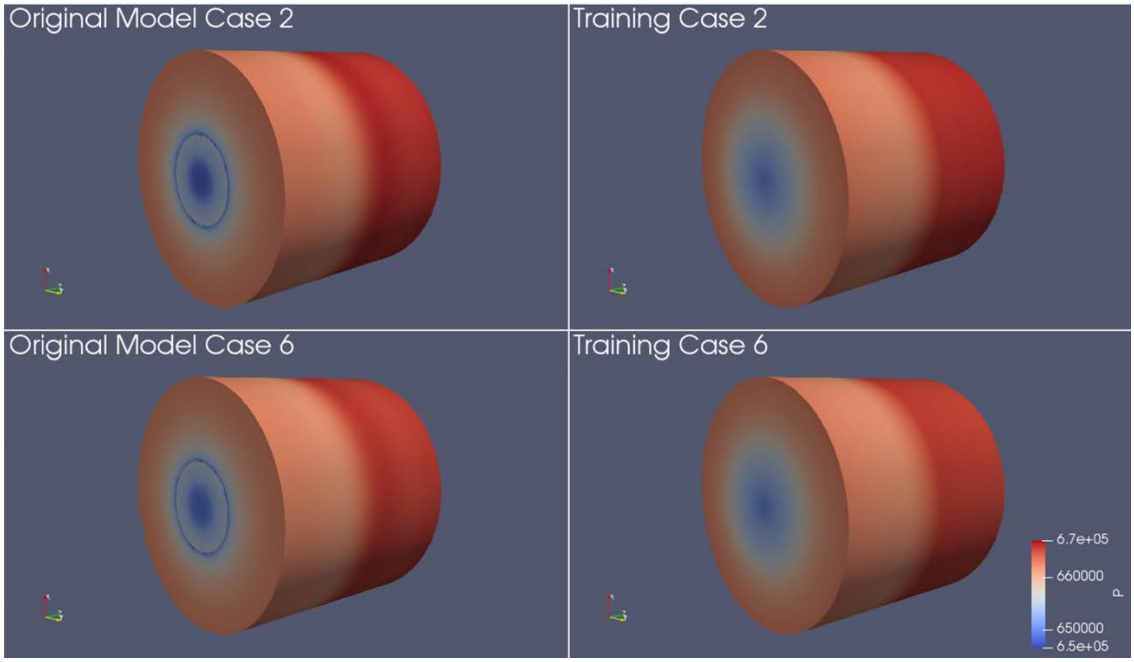


Half Combustor Exterior

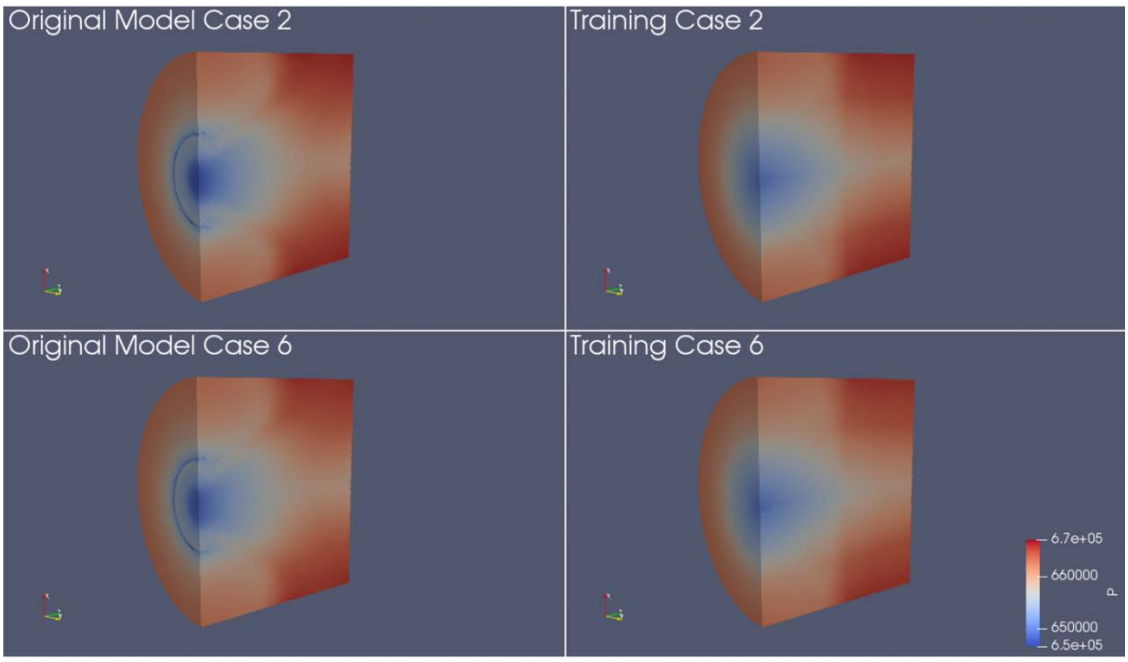


Half Combustor Exterior

Figure 4-9: Step 2 Smart Proxy Results for Training Cases 1&5 – Pressure



Half Combustor Exterior



Half Combustor Interior

Figure 4-10: Step 2 Smart Proxy Results for Training Cases 2&6 – Pressure

4.3 Model Development Step 3: Swirler Distances

In a realistic combustor, the overall rate of reaction is controlled by turbulent mixing. Mixing does not only occur as a result of a jet crossflow of the preheated air and fuel at the inlet but typically through some additional mechanism such as swirling flow which increases the turbulence intensity [27]. The results obtained in Step 2 showed that the neural network was missing the reaction flow pattern (especially at the inlet-combustor boundary) created by the air-fuel mixing effect of the swirlers at the inlet. The newly generated features in Step 3 included a representation of this swirling effect on the combustion process. These includes calculating the distances of each focal cell in the system to each of the thirteen swirler nozzles. Figure 4-11 illustrates how these distances are calculated for an arbitrary cell in the system.

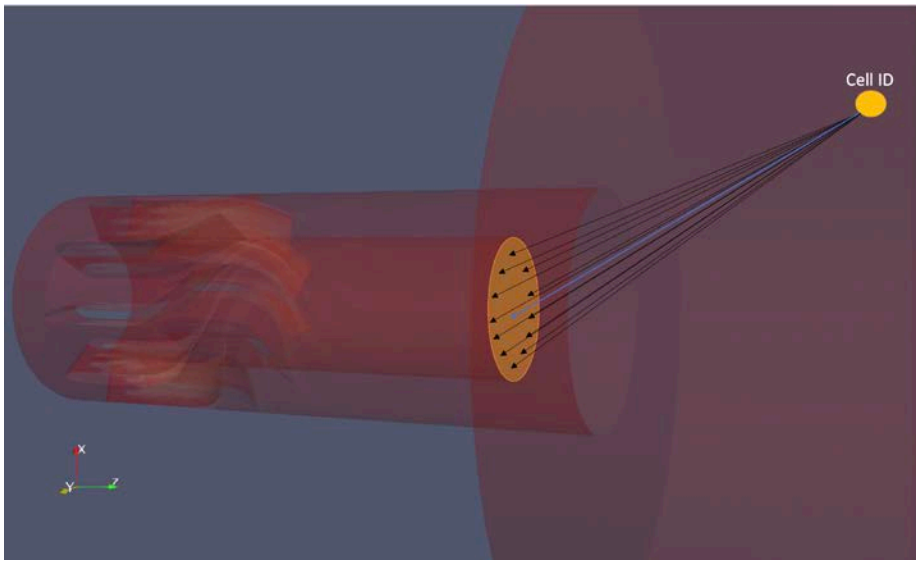


Figure 4-11: Distance to Swirler Nozzles

Based on what was learned in Steps 1 and 2 of the development process, the total number of samples that the available desktop machine could handle in development was already known. This information, coupled with a detail analysis of the distribution of the transport variables of interest (pressure, temperature, oxygen, nitrogen and carbon-dioxide) across all development cases was used in dividing the system into sections for development purpose. The B6 Combustor model was divided into seven different development sections. Figure 4-12 shows the model sectioning and naming convention for each section. Table 4-3 shows the number of cells by section for a single CFD simulation run.

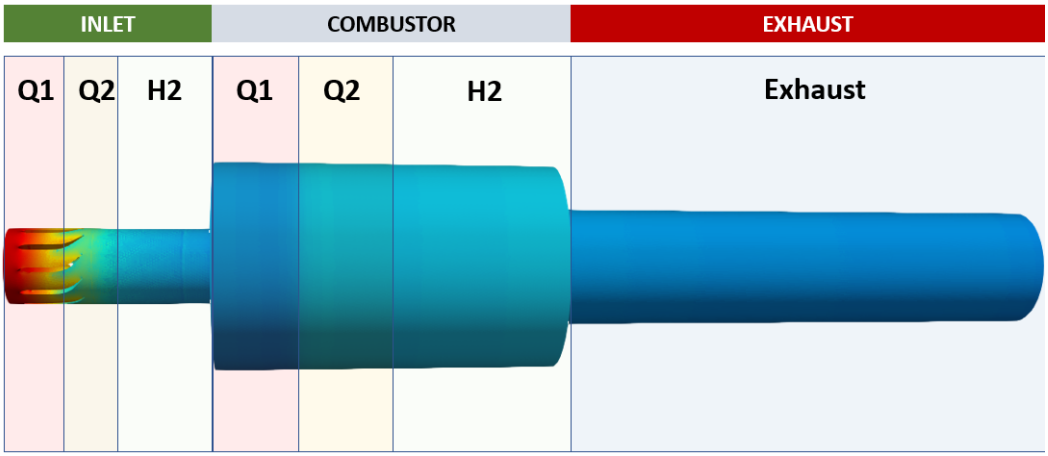


Figure 4-12: Model Sectioning

Table 4-3: Total Number of Cells by Model Section

Total No. of Cells by Model Section				
Section Name	**Convention Name**	**No. Cells**	**Z-axis BGN**	**Z-axis END**
1st Quarter Inlet	Q1_Inlet	1,623,745	-0.83058	-0.78868
2nd Quarter Inlet	Q2_Inlet	1,662,119	-0.78868	-0.74678
2nd Half Inlet	H2_Inlet	344,757	-0.74678	-0.66298
1st Quarter Combustor	Q1_Combustor	962,001	-0.66298	-0.58981
2nd Quarter Combustor	Q2_Combustor	971,256	-0.58981	-0.51664
2nd Half Combustor	H2_Combustor	1,898,924	-0.516642	-0.37030
Exhaust	Exhaust	1,903,951	-0.37030	0.0762
Entire Model	CFD	9,291,712	-0.83058	0.0762

Once the model was properly sectioned, the section with the smallest number of cells, the first Quarter Combustor (Q1-Combustor) as shown in Figure 4-13 was selected as the section to be modeled in the third step of the development process. The number of cells and the combustion activity occurring at the Q1-Combustor section were some of the factors considered to effectively evaluate the effect of the newly generated features on the performance of the neural network in a timely manner.

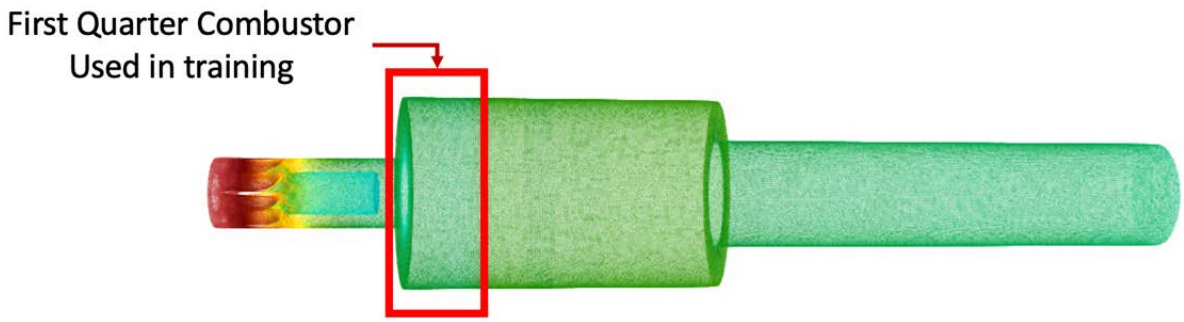


Figure 4-13: Step 2 – Target section for training and development of the neural network

4.3.1 Model Development Step 3 – Model Training Information

The same data partitioning used in earlier steps was implemented in Step 3. Six cases (2, 3, 4, 6, 7, 8) were used for training and calibration while two cases (1 and 5) were used as validation after each training attempt. All five attributes of interest (P, T, N_2, O_2, CO_2) were modeled in this step. Furthermore, a single hidden-layer neural network with one output was built for each attribute. The list of training input attributes in Step 3 is provided in Table 4-4.

Table 4-4: Training Input Attributes in Development Step 3

Attributes Provided from FLUENT	Attributes Generated in Step 1	Attributes Generated in Step 2	Attributes Generated in Step 3
X_Coord	Node_Count	Euclidean_Dist_Inlet	Swirler_1_dist
Y_Coord	Dist_To_Inlet	Euclidean_Dist_Inlet_Comb	Swirler_2_dist
Z_Coord	Dist_To_Outlet	Euclidean_Dist_Comb_Exhaust	Swirler_3_dist
volume	Dist_To_Inlet_Combustor	Euclidean_Dist_Outlet	Swirler_4_dist
CH4_Inlet	Dist_To_Combustor_Exhaust	Radial_Boundary_Ratio	Swirler_5_dist
C2H6_Inlet	Closest_Radial_Boundary	Adjacent_Cell_Volume_1	Swirler_6_dist
C3H8_Inlet	Farthest_Radial_Boundary	Adjacent_Cell_Volume_2	Swirler_7_dist
C4H10_Inlet		Adjacent_Cell_Volume_3	Swirler_8_dist
C5H12_Inlet		Adjacent_Cell_Volume_4	Swirler_9_dist
N2_Inlet		Total_Adjacent_Cells	Swirler_10_dist
CO2_Inlet		Cell_Adjacent_Inlet	Swirler_11_dist
Air_Flow_Rate		Cell_Adjacent_Outlet	Swirler_12_dist
Main_Fuel_Flow		Cell_Adjacent_Wall	Swirler_Center_dist
Pilot_Fuel_Flow			
Air_Inlet_Temp			
Pressure_Outlet			
Total = 16	**Total = 7**	**Total = 13**	**Total = 13**

4.3.2 Model Development Step 3 – Presentation of Results

All attributes were modeled for the Q1 Combustor section, but we only present the results for pressure, temperature, and carbon dioxide distributions for the two validation cases base case 1 and blend case 5. Similar quality of results was obtained for nitrogen and oxygen and these are presented in Appendix 7. The results for each attribute are presented in two figures. The first figure shows the result for the entire Q1 Combustor section while the second figure shows a quarter of the Q1 Combustor. Each figure contains a total of 4 images for a single case. The images on the left represent the CFD model simulation. The images on the right represent the smart proxy model. The top and bottom images show the front and the back of the Q1 combustor section respectively.

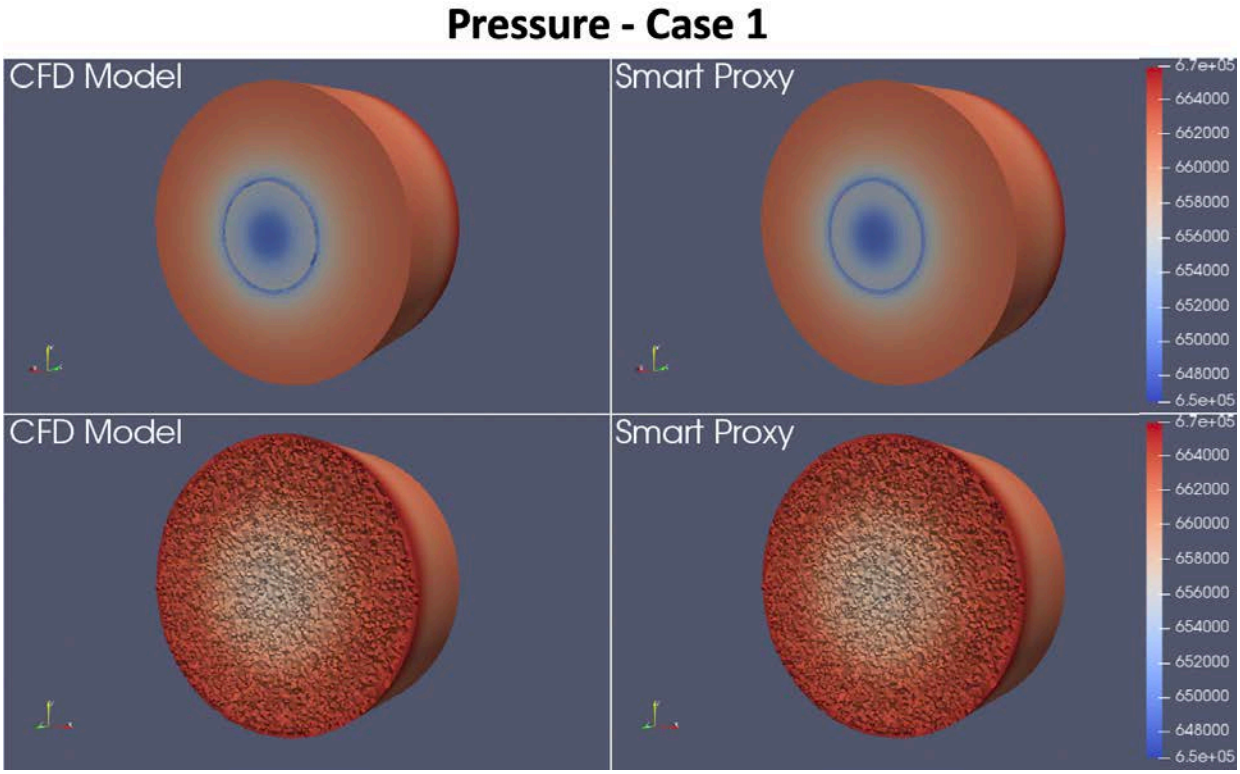


Figure 4-14: Step 3 Results (Q1 Combustor Full View) for Validation Case 1 – Pressure

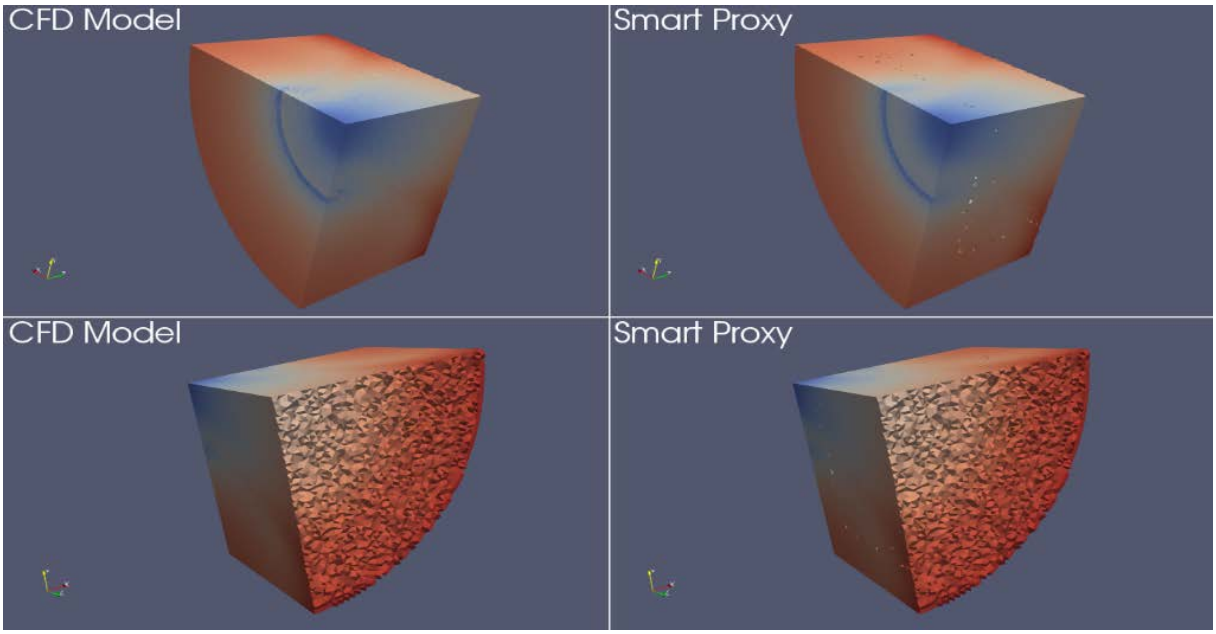


Figure 4-15: Step 3 Results (Q1 Combustor Quarter View) for Validation Case 1 – Pressure

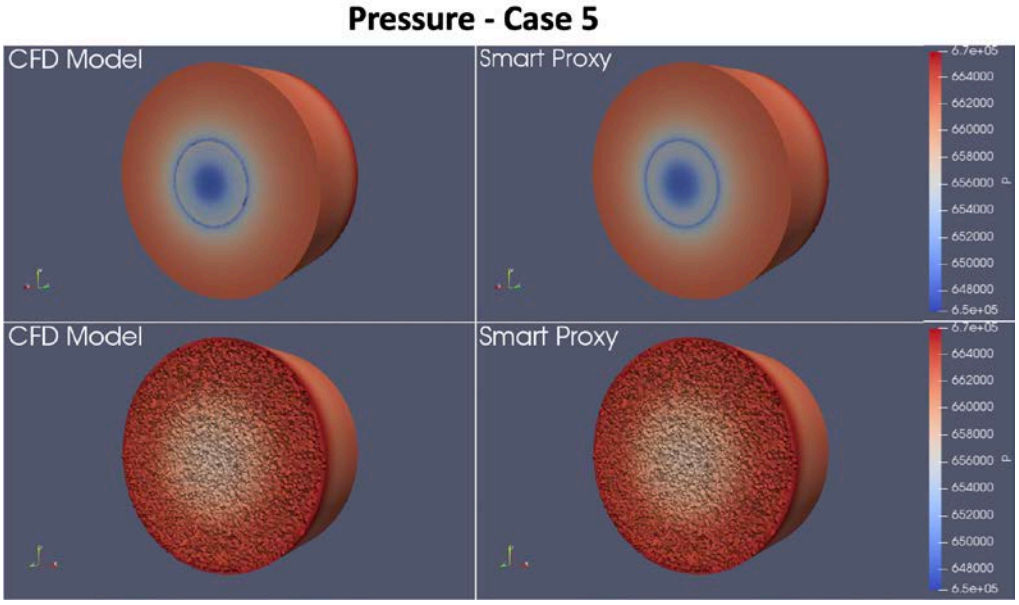


Figure 4-16: Step 3 Results (Q1 Combustor Full View) for Validation Case 5 – Pressure

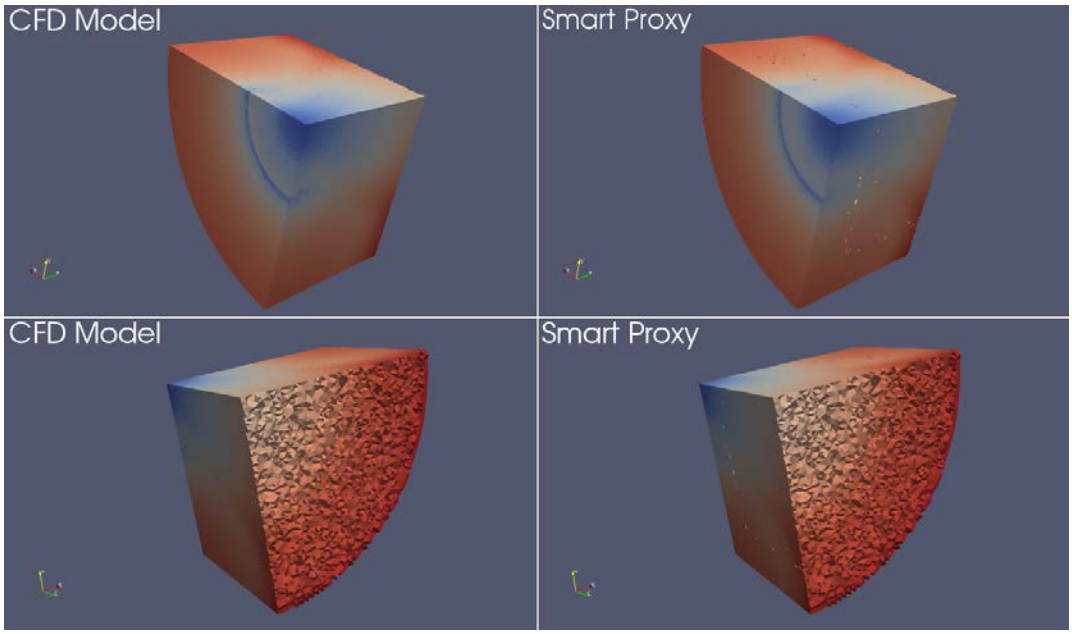


Figure 4-17: Step 3 Results (Q1 Combustor Quarter View) for Validation Case 5 – Pressure

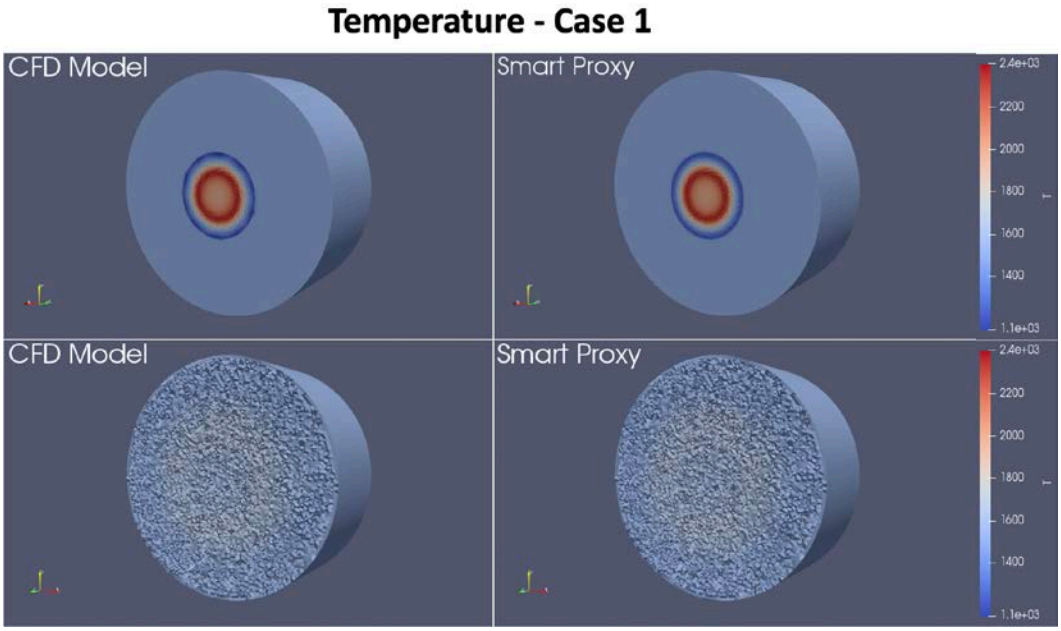


Figure 4-18: Step 3 Results (Q1 Combustor Full View) for Validation Case 1 – Temperature

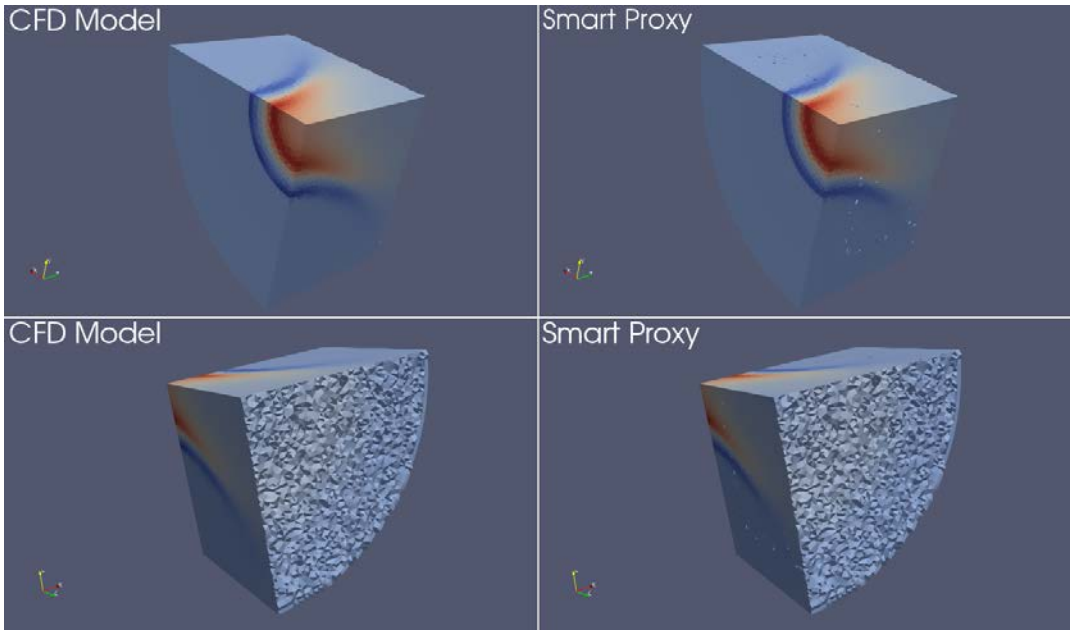


Figure 4-19: Step 3 Results (Q1 Combustor Quarter View) for Validation Case 1 – Temperature

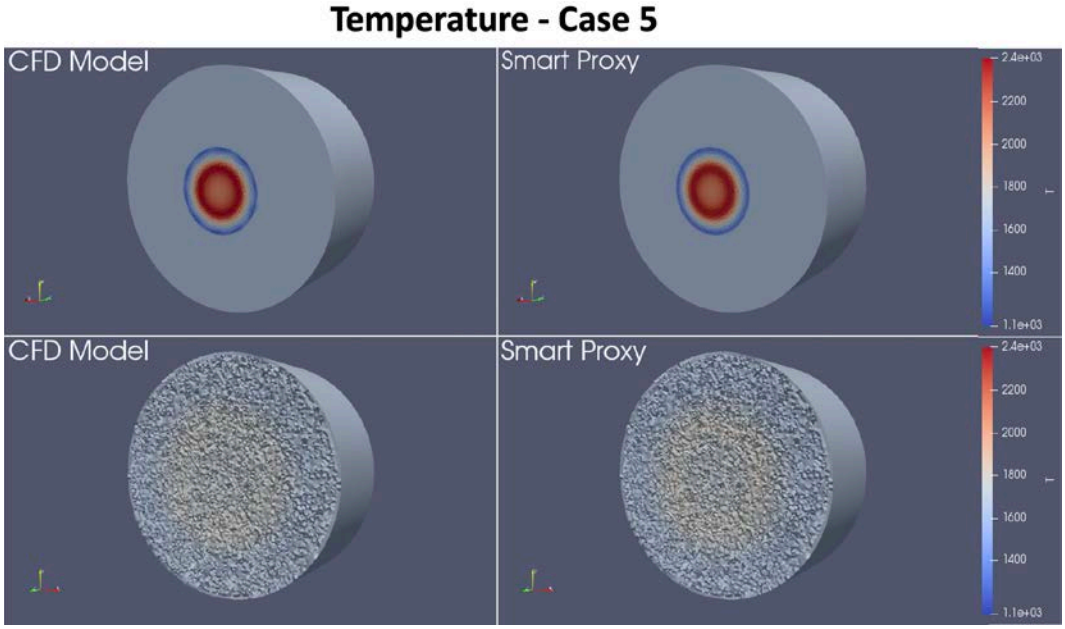


Figure 4-20: Step 3 Results (Q1 Combustor Full View) for Validation Case 5 – Temperature

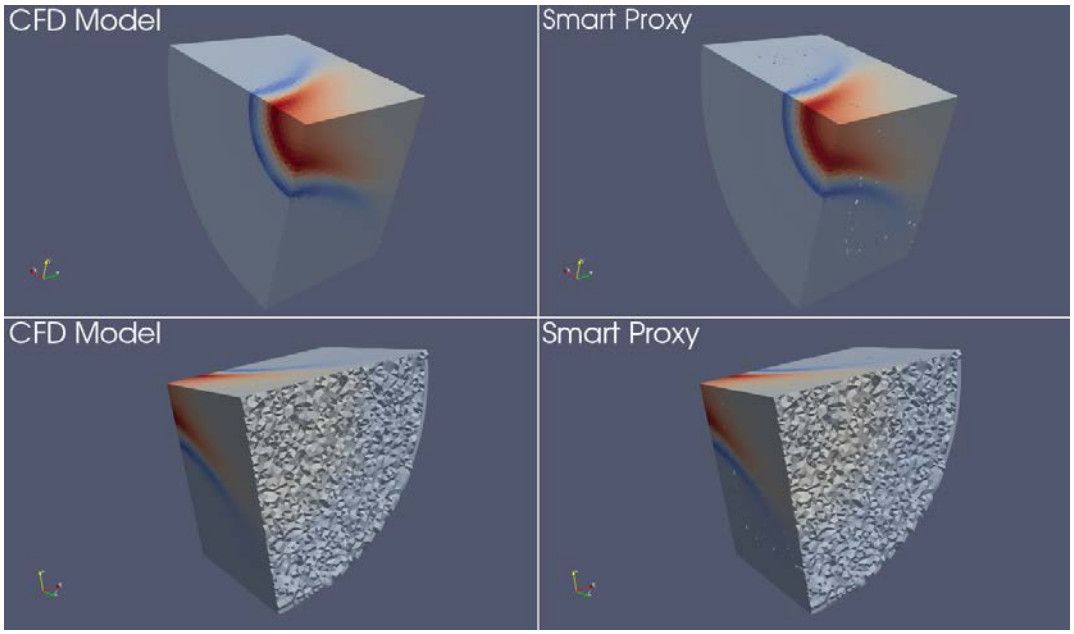


Figure 4-21: Step 3 Results (Q1 Combustor Quarter View) for Validation Case 5 – Temperature

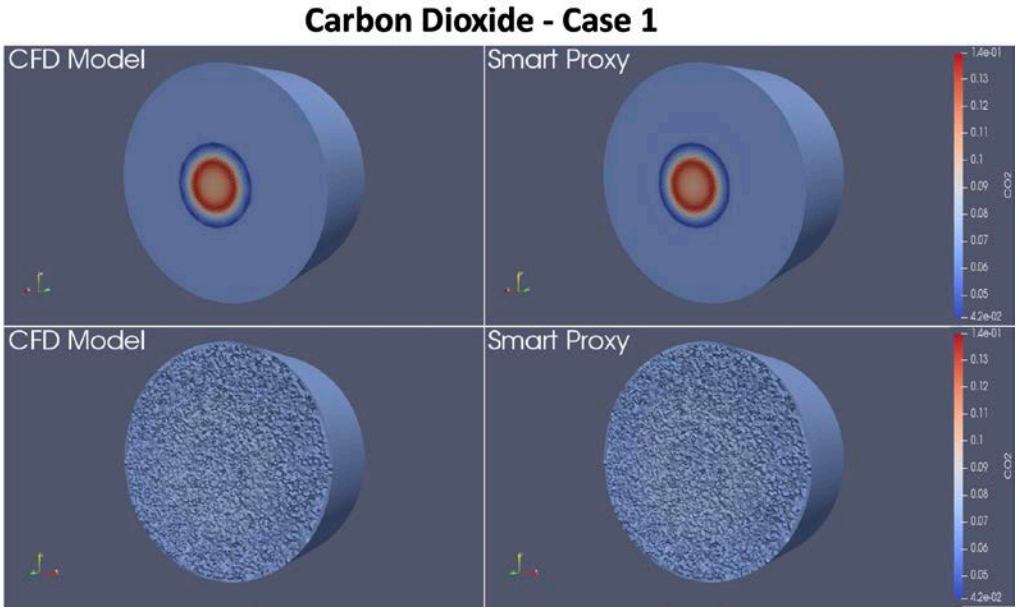


Figure 4-22: Step 3 Results (Q1 Combustor Full View) for Validation Case 1 – Carbon dioxide

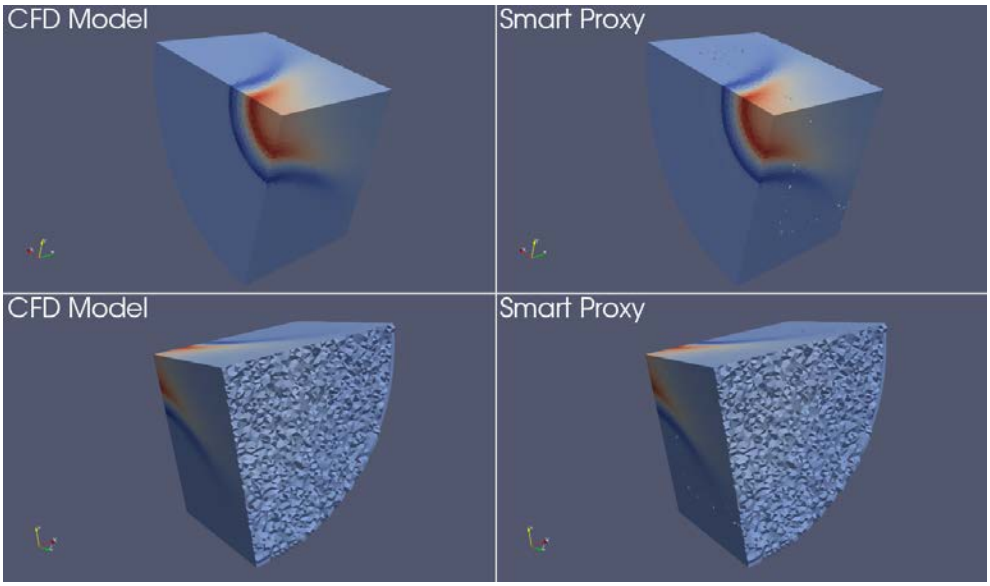


Figure 4-23: Step 3 Results (Q1 Combustor Quarter View) for Validation Case 1 – Carbon Dioxide

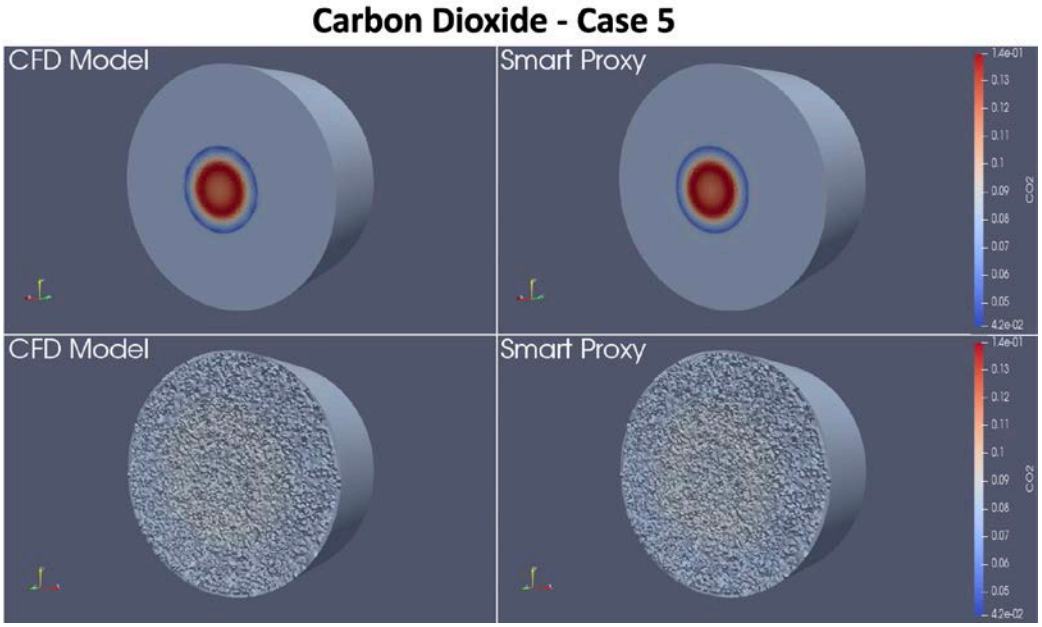


Figure 4-24: Step 3 Results (Q1 Combustor Full View) for Validation Case 5 – Carbon Dioxide

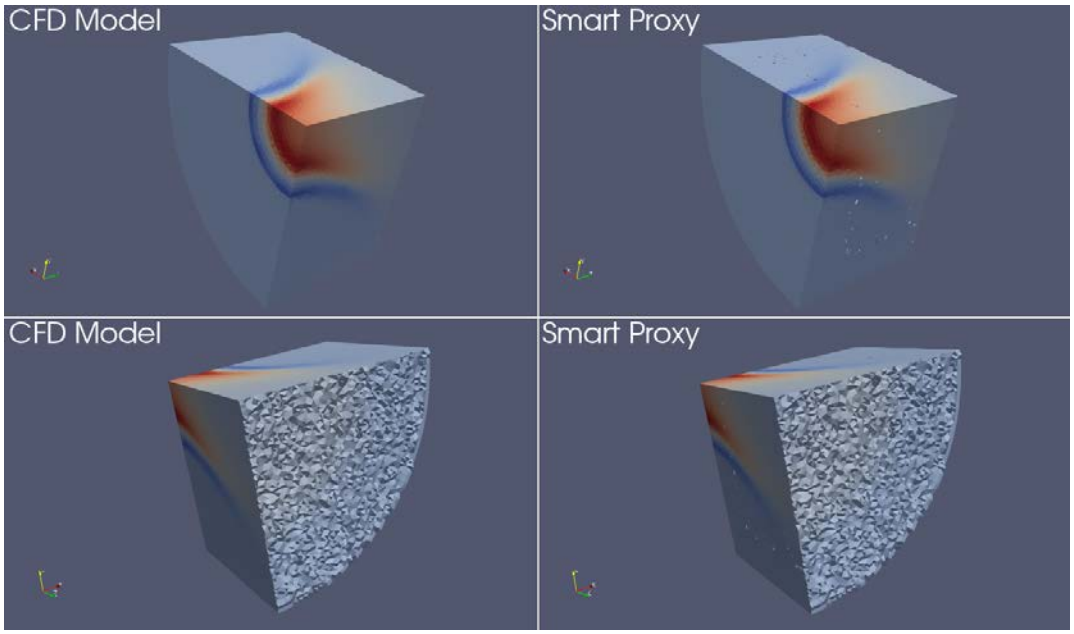


Figure 4-25: Step 3 Results (Q1 Combustor Quarter View) for Validation Case 5 – Carbon Dioxide

4.4 Model Development Step 4: Fuzzy Clustering

While the results obtained in step 3 showed significant improvement in smart proxy performance compared to the results obtained in Step 1 and Step 2, there were still opportunities to be explored for improvement.

In the last step of the model development process, new features were generated to further represent the physics of the reaction flow occurring in the combustion process. This involved clustering cells in each of the seven development sections of the system such that for each attribute, cells with similar distribution of the attribute are grouped together. Detail description of this technique has been provided in Chapter 3 and in the background section of this report.

In addition to the clustering information that was added to the database, a new feature representing the extent of propane blend was added. This was specified by taking the ratio of methane flow rate to propane flow rate in the simulation model. A total of five new features were added in Step 4. Table 4-5 shows the list of added features in this step.

In Step 4a, neural networks were trained, calibrated, and validated for all sections of the B6 combustor. Details of each section are summarized in Table 4-3 and graphically illustrated in Figure 4-12.

4.4.1 Model Development Step 4 – Model Training Information

The same data partitioning approach as used in previous steps was used. Six cases (Cases 2, 3, 4, 6, 7, and 8) were used for training (80% randomly selected samples in the six cases) and calibration (remaining 20% of samples in the six cases) of the neural network. Base case 1 and blend case 5 were used for validation of each neural network at the end of each training attempt. The size of each development dataset is summarized in Table 4-3. As shown in Table 4-5, a total of 54 input attributes were used in training the neural networks.

A neural network was developed for each attribute in each of the seven development sections, therefore a total of 35 neural networks was trained in Step 4. Each fully trained neural network was finally deployed on the Extra Base Case 9 and Extra Blend Case 10, as a blind test of the performance of the neural networks. For each attribute, results for all seven development sections were combined for better visualization of the entire system.

Table 4-5: Training Input Attributes in Development Step 4

Attributes Provided from FLUENT	Attributes Generated in Step 1	Attributes Generated in Step 2	Attributes Generated in Step 3	Attributes Generated in Step 4
X_Coord	Node_Count	Euclidean_Dist_Inlet	Swirler_1_dist	CM_1
Y_Coord	Dist_To_Inlet	Euclidean_Dist_Inlet_Comb	Swirler_2_dist	CM_2
Z_Coord	Dist_To_Outlet	Euclidean_Dist_Comb_Exhaust	Swirler_3_dist	CM_3
volume	Dist_To_Inlet_Combustor	Euclidean_Dist_Outlet	Swirler_4_dist	Cluster_Id
CH4_Inlet	Dist_To_Combustor_Exhaust	Radial_Boundary_Ratio	Swirler_5_dist	Blend Ratio
C2H6_Inlet	Closest_Radial_Boundary	Adjacent_Cell_Volume_1	Swirler_6_dist	
C3H8_Inlet	Farthest_Radial_Boundary	Adjacent_Cell_Volume_2	Swirler_7_dist	
C4H10_Inlet		Adjacent_Cell_Volume_3	Swirler_8_dist	
C5H12_Inlet		Adjacent_Cell_Volume_4	Swirler_9_dist	
N2_Inlet		Total_Adjacent_Cells	Swirler_10_dist	
CO2_Inlet		Cell_Adjacent_Inlet	Swirler_11_dist	
Air_Flow_Rate		Cell_Adjacent_Outlet	Swirler_12_dist	
Main_Fuel_Flow		Cell_Adjacent_Wall	Swirler_Center_dist	
Pilot_Fuel_Flow				
Air_Inlet_Temp				
Pressure_Outlet				
Total = 16	**Total = 7**	**Total = 13**	**Total = 13**	**Total = 5**

4.4.2 Model Development Step 4 – Presentation of Results

This section presents the smart proxy results of the distribution of all attributes in the entire B6 Combustor system. Each attribute result is graphically and numerically presented by (a) a histogram describing the percent error distribution (all attributes shown under a 5% error with the exception of Temperature shown in a 10% error scale), (b) two tables; one table suggesting the number of cells within a 100% error scale and the second table providing a more detailed insight of the error distribution below 10%, and (c) two 3-Dimensional images showing distribution of the attribute in the entire system and a half cross-section view of the system.

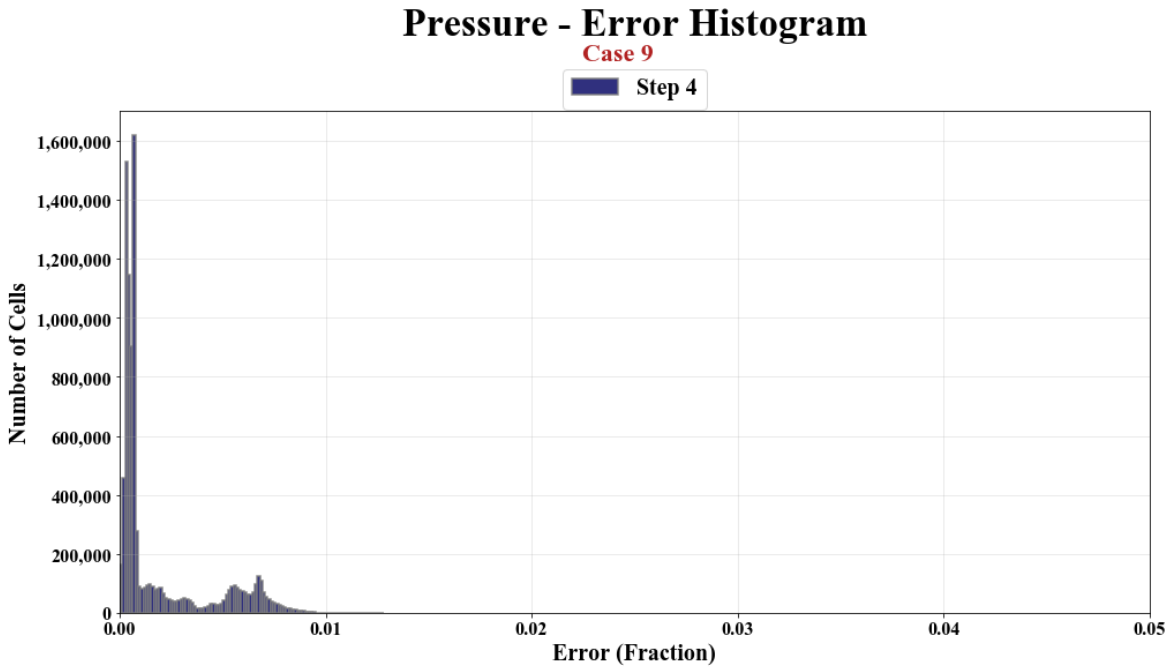


Figure 4-26: Step 4 Error Histogram for Blind Validation Case 9 – Pressure

Table 4-6: No. of Cells Under 100% Error for Blind Validation Case 9 – Pressure

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,291,730	100.0%
10% - 20%	0	0.0%
20% - 30%	0	0.0%
30% - 40%	0	0.0%
40% - 50%	0	0.0%
50% - 60%	0	0.0%
60% - 70%	0	0.0%
70% - 80%	0	0.0%
80% - 90%	0	0.0%
> 90%	0	0.0%

Table 4-7: No. of Cells Under 10% Error for Blind Validation Case 9 – Pressure

Percent Ranges	Number of Cells	Perc. Cells
< 2%	9,291,730	100.0%
2% - 4%	0	0.0%
4% - 6%	0	0.0%
6% - 8%	0	0.0%
8% - 10%	0	0.0%
> 10%	0	0.0%

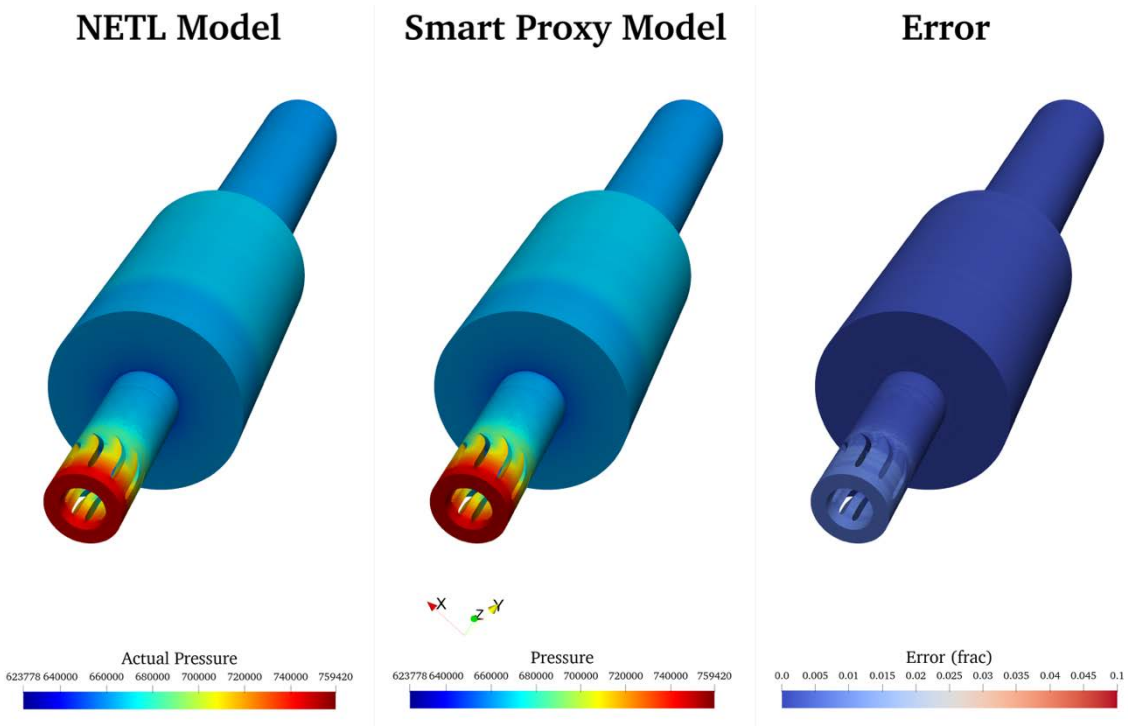


Figure 4-27: Step 4 Results (Entire System) for Blind Validation Case 9 – Pressure

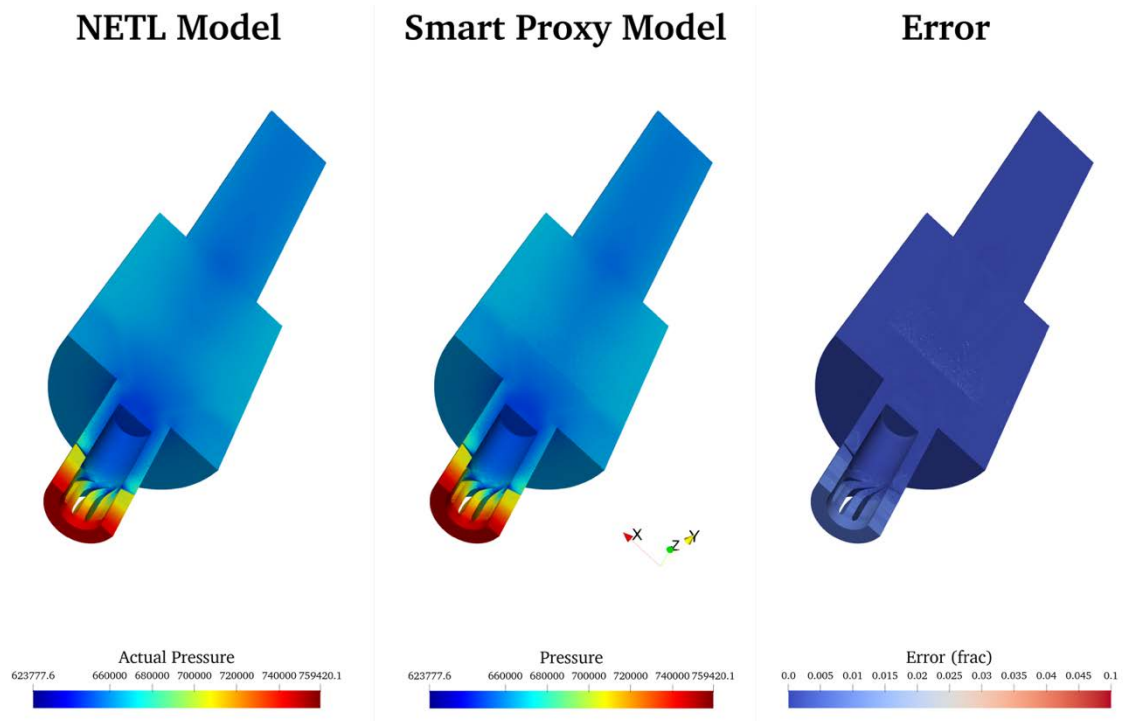


Figure 4-28: Step 4 Results (Half View) for Blind Validation Case 9 – Pressure

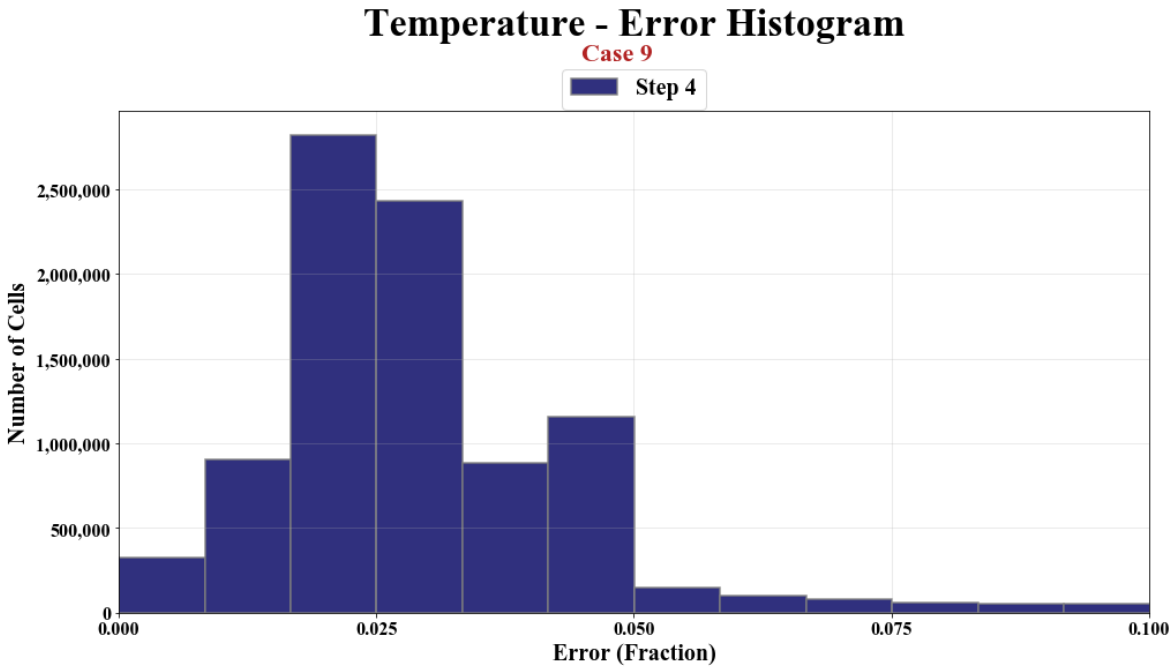


Figure 4-29: Step 4 Error Histogram for Blind Validation Case 9 – Temperature

Table 4-8: No. of Cells Under 100% Error for Blind Validation Case 9 – Temperature

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,021,129	97.088%
10% - 20%	222,291	2.392%
20% - 30%	38,699	0.416%
30% - 40%	8,919	0.096%
40% - 50%	625	0.007%
50% - 60%	51	0.001%
60% - 70%	13	0.0%
70% - 80%	2	0.0%
80% - 90%	1	0.0%
> 90%	0	0.0%

Table 4-9:No. of Cells Under 10% Error for Blind Validation Case 9 – Temperature

Percent Ranges	Number of Cells	Perc. Cells
< 2%	2,579,897	27.766%
2% - 4%	4,465,385	48.058%
4% - 6%	1,647,540	17.731%
6% - 8%	199,023	2.142%
8% - 10%	129,284	1.391%
> 10%	270,601	2.912%

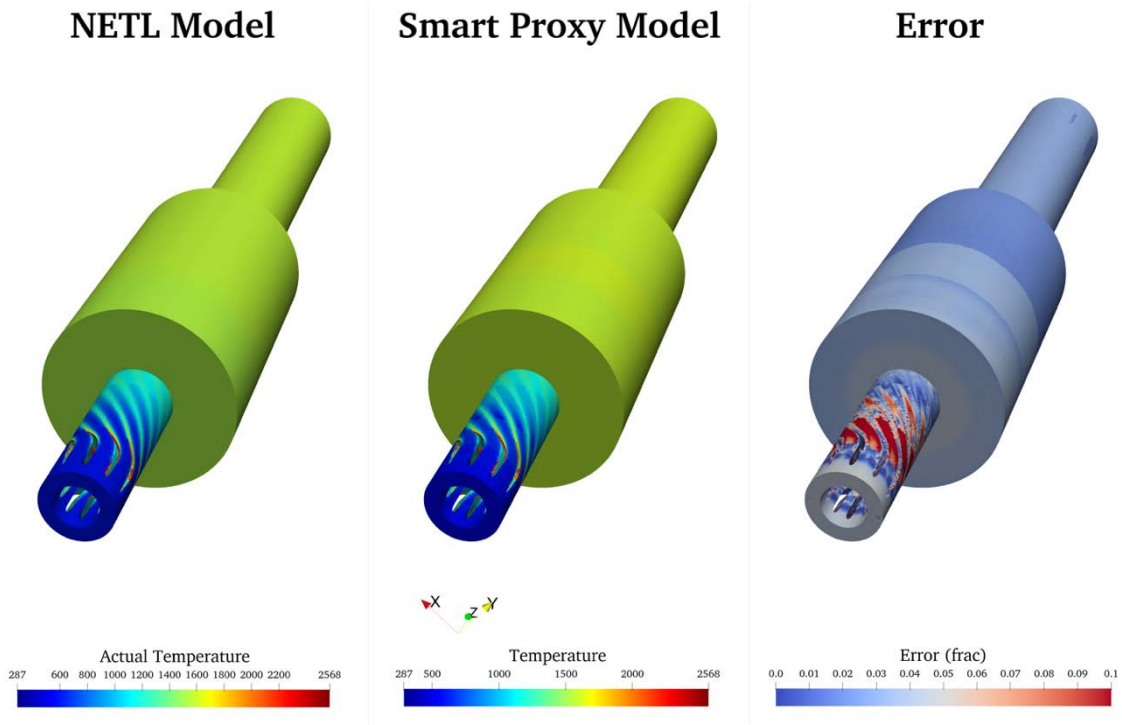


Figure 4-30: Step 4 Results (Entire System) for Blind Validation Case 9 – Temperature

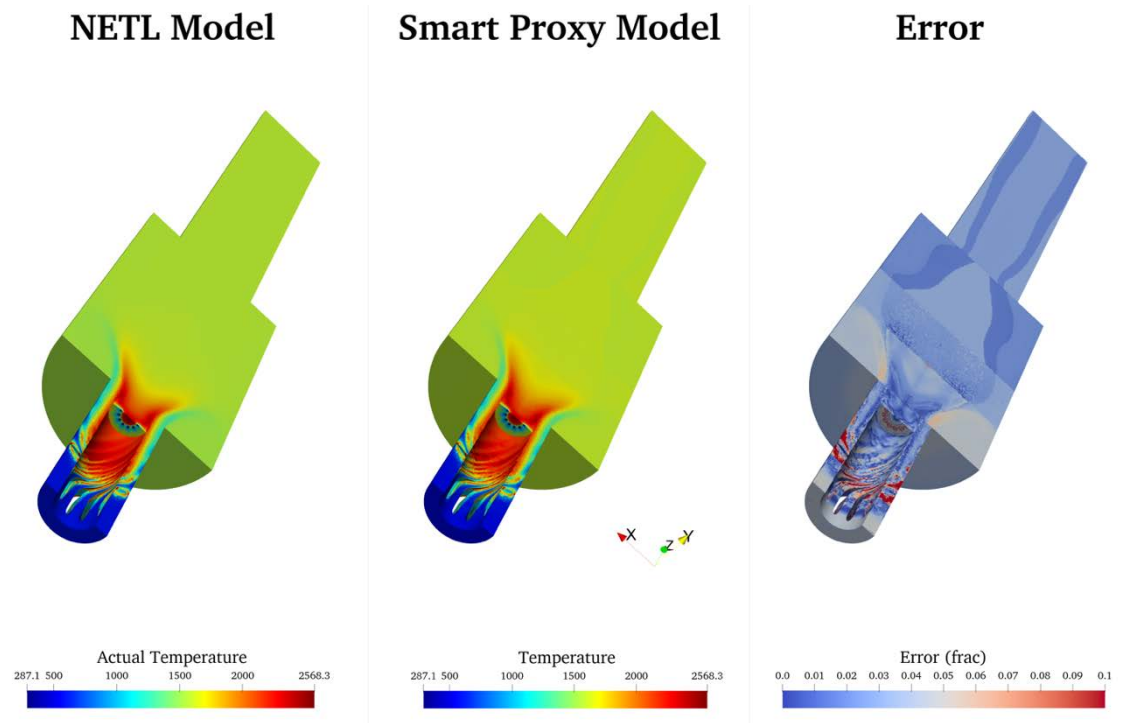


Figure 4-31: Step 4 Results (Half View) for Blind Validation Case 9 – Temperature

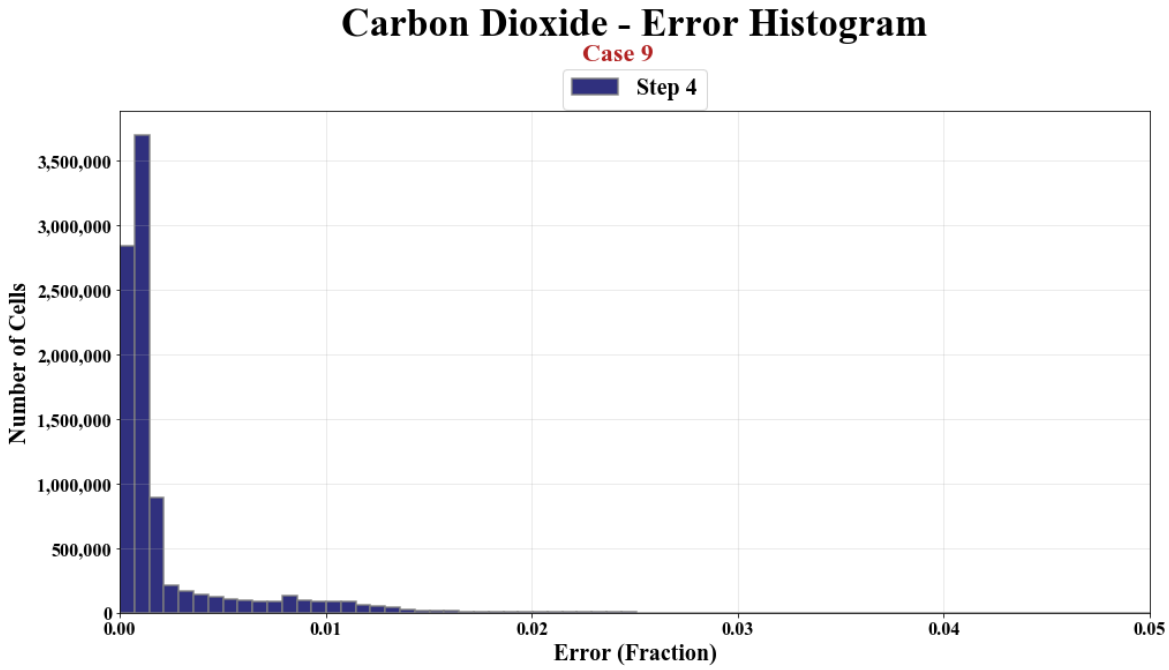


Figure 4-32: Step 4 Error Histogram for Blind Validation Case 9 – Carbon Dioxide

Table 4-10:No. of Cells Under 100% Error for Blind Validation Case 9 – Carbon Dioxide

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,291,730	100.0%
10% - 20%	0	0.0%
20% - 30%	0	0.0%
30% - 40%	0	0.0%
40% - 50%	0	0.0%
50% - 60%	0	0.0%
60% - 70%	0	0.0%
70% - 80%	0	0.0%
80% - 90%	0	0.0%
> 90%	0	0.0%

Table 4-11: No. of Cells Under 10% Error for Blind Validation Case 9 – Carbon Dioxide

Percent Ranges	Number of Cells	Perc. Cells
< 2%	9,222,507	99.255%
2% - 4%	53,275	0.573%
4% - 6%	13,526	0.146%
6% - 8%	2,422	0.026%
8% - 10%	0	0.0%
> 10%	0	0.0%

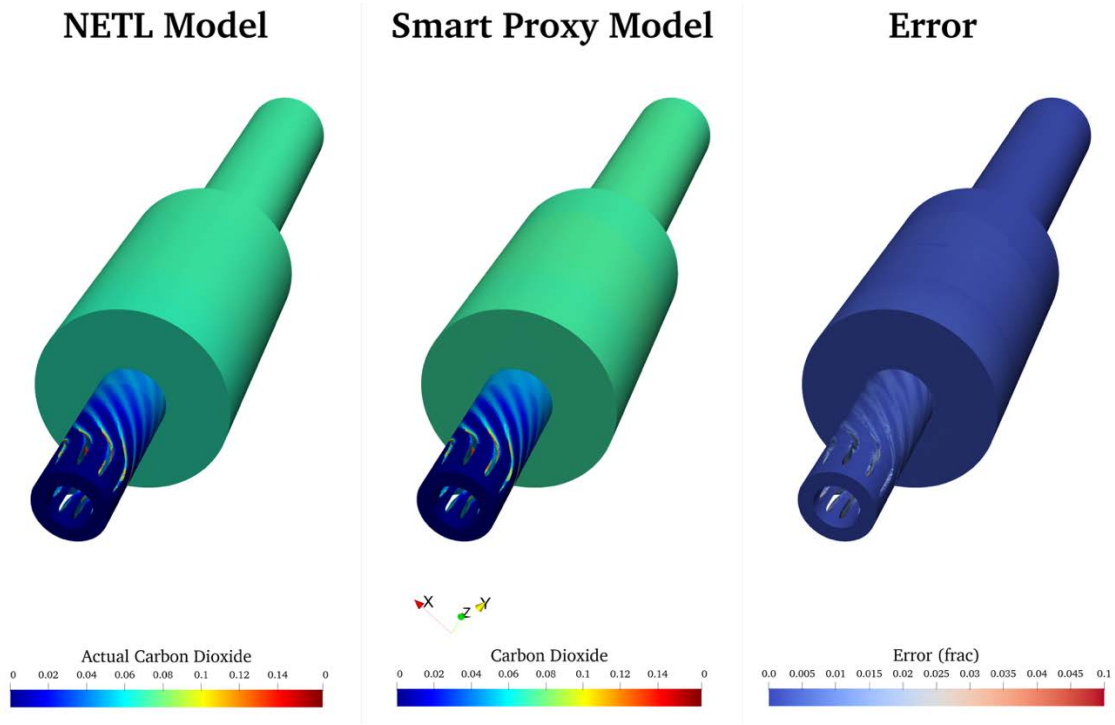


Figure 4-33: Step 4 Results (Entire System) for Blind Validation Case 9 – Carbon Dioxide

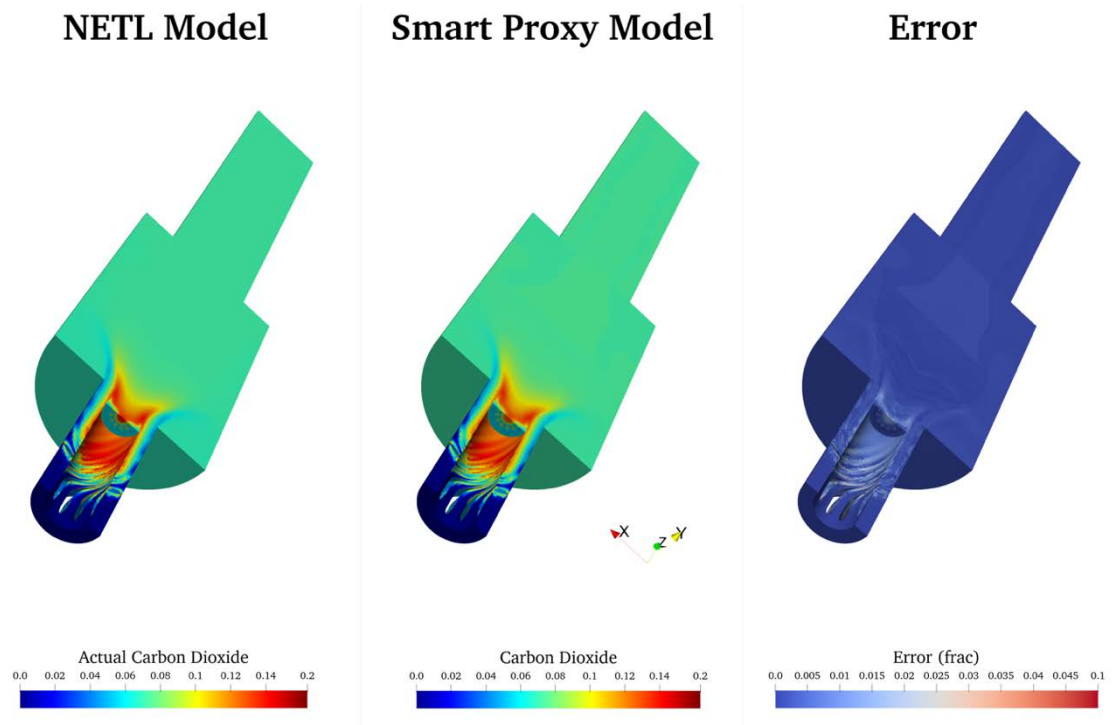


Figure 4-34: Step 4 Results (Half View) for Blind Validation Case 9 – Carbon Dioxide

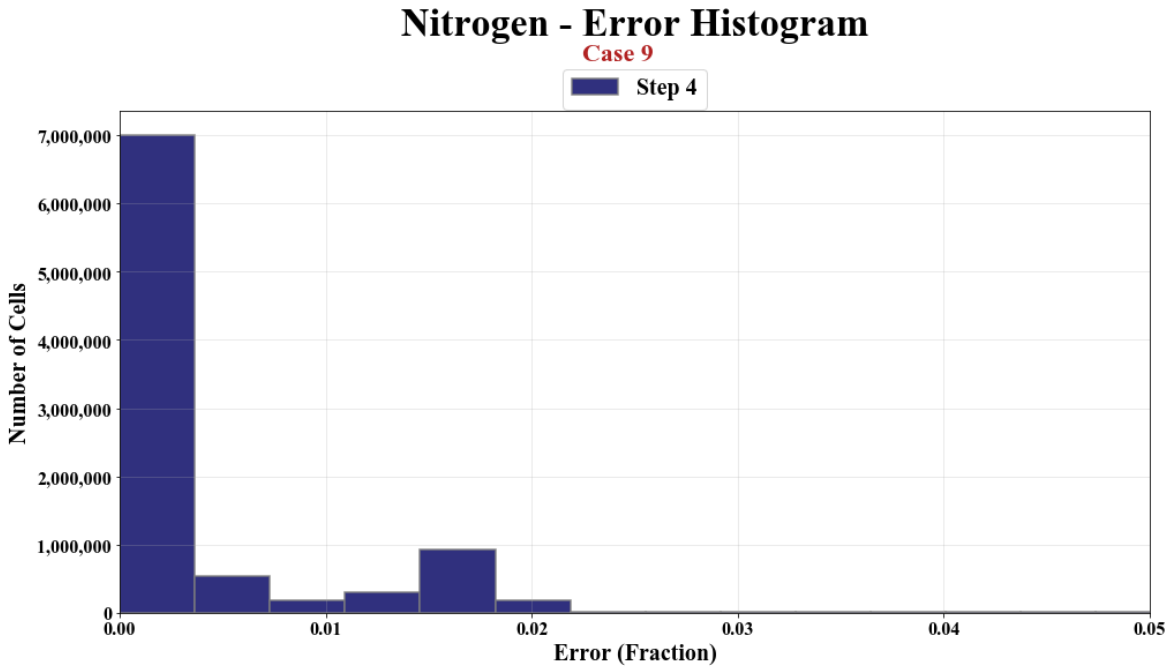


Figure 4-35: Step 4 Error Histogram for Blind Validation Case 9 – Nitrogen

Table 4-12: No. of Cells Under 100% Error for Blind Validation Case 9 – Nitrogen

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,252,113	99.574%
10% - 20%	35,122	0.378%
20% - 30%	4,429	0.048%
30% - 40%	66	0.001%
40% - 50%	0	0.0%
50% - 60%	0	0.0%
60% - 70%	0	0.0%
70% - 80%	0	0.0%
80% - 90%	0	0.0%
> 90%	0	0.0%

Table 4-13: No. of Cells Under 10% Error for Blind Validation Case 9 – Nitrogen

Percent Ranges	Number of Cells	Perc. Cells
< 2%	9,130,895	98.269%
2% - 4%	67,842	0.73%
4% - 6%	28,232	0.304%
6% - 8%	15,280	0.164%
8% - 10%	9,864	0.106%
> 10%	39,617	0.426%

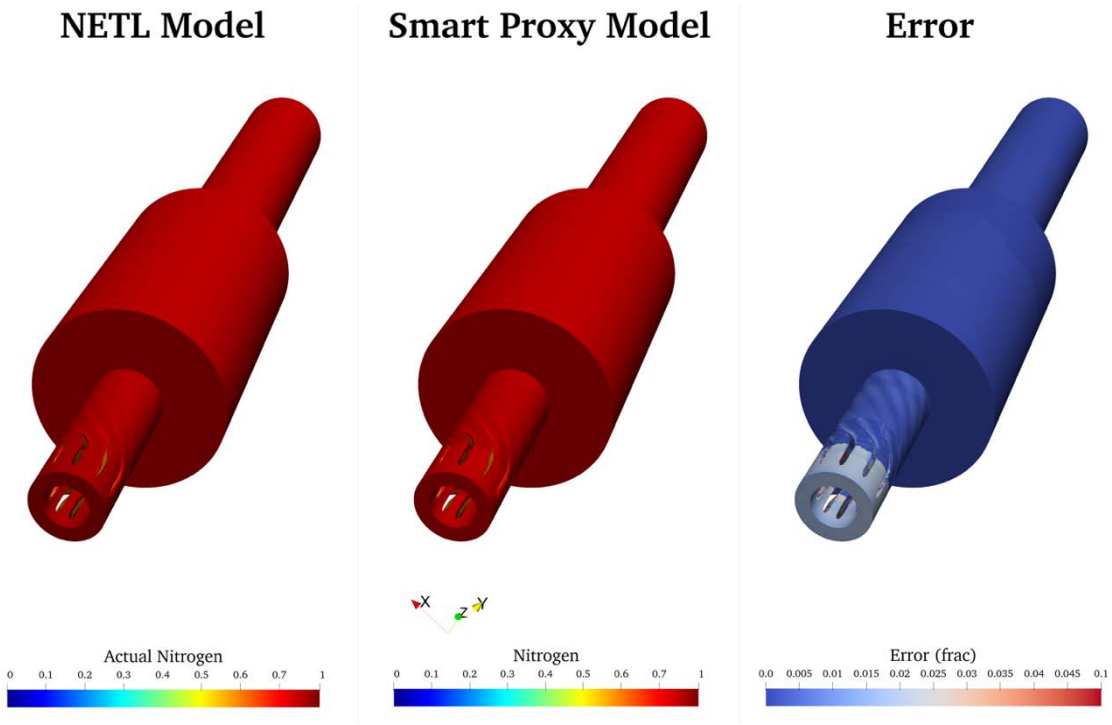


Figure 4-36: Step 4 Results (Entire System) for Blind Validation Case 9 – Nitrogen

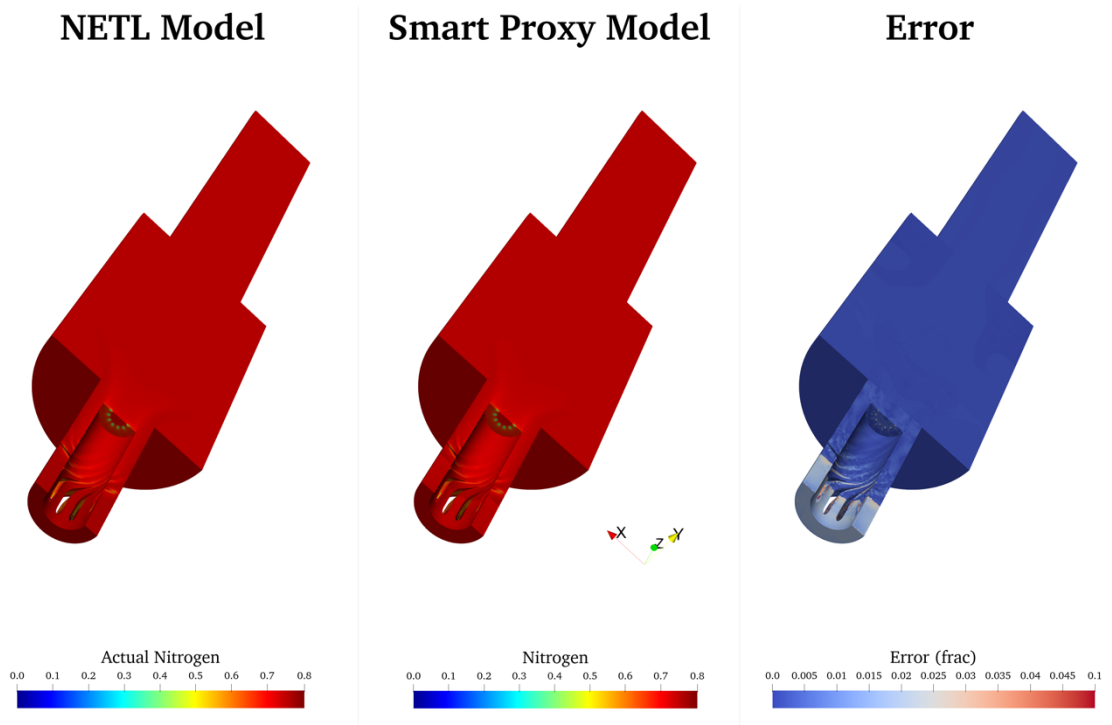


Figure 4-37: Step 4 Results (Half View) for Blind Validation Case 9 – Nitrogen

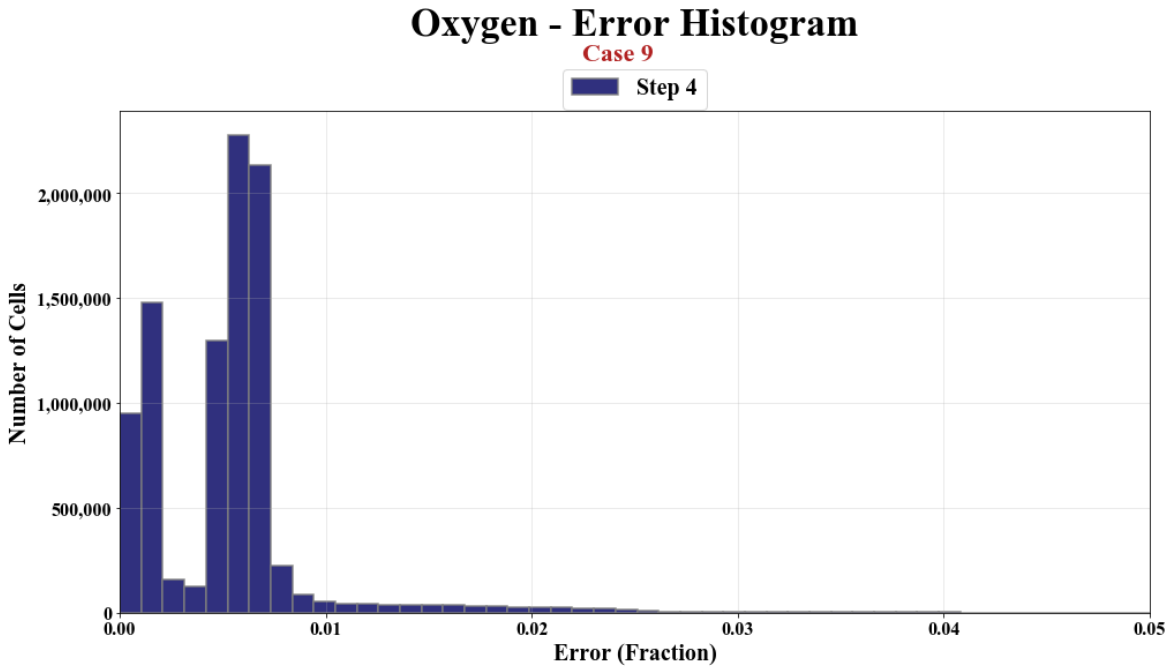


Figure 4-38: Step 4 Error Histogram for Blind Validation Case 9 – Oxygen

Table 4-14: No. of Cells Under 100% Error for Blind Validation Case 9 – Oxygen

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,291,728	100.0%
10% - 20%	2	0.0%
20% - 30%	0	0.0%
30% - 40%	0	0.0%
40% - 50%	0	0.0%
50% - 60%	0	0.0%
60% - 70%	0	0.0%
70% - 80%	0	0.0%
80% - 90%	0	0.0%
> 90%	0	0.0%

Table 4-15: No. of Cells Under 10% Error for Blind Validation Case 9 – Oxygen

Percent Ranges	Number of Cells	Perc. Cells
< 2%	9,113,681	98.084%
2% - 4%	144,830	1.559%
4% - 6%	25,853	0.278%
6% - 8%	7,246	0.078%
8% - 10%	118	0.001%
> 10%	2	0.0%

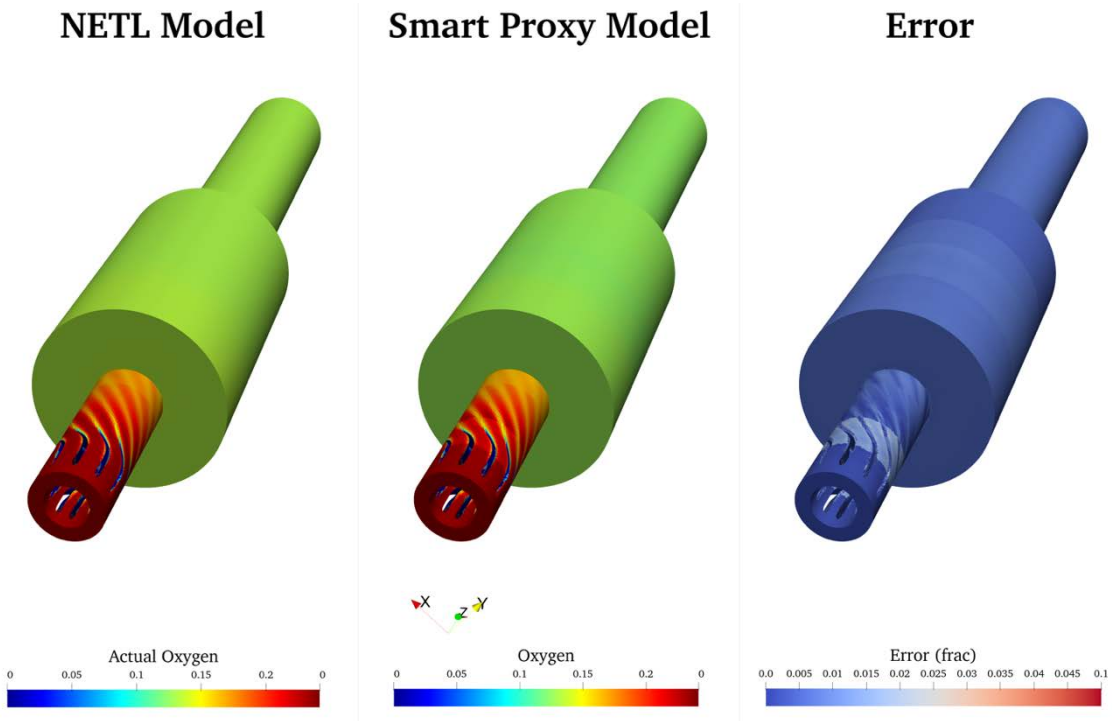


Figure 4-39: Step 4 Results (Entire System) for Blind Validation Case 9 – Oxygen

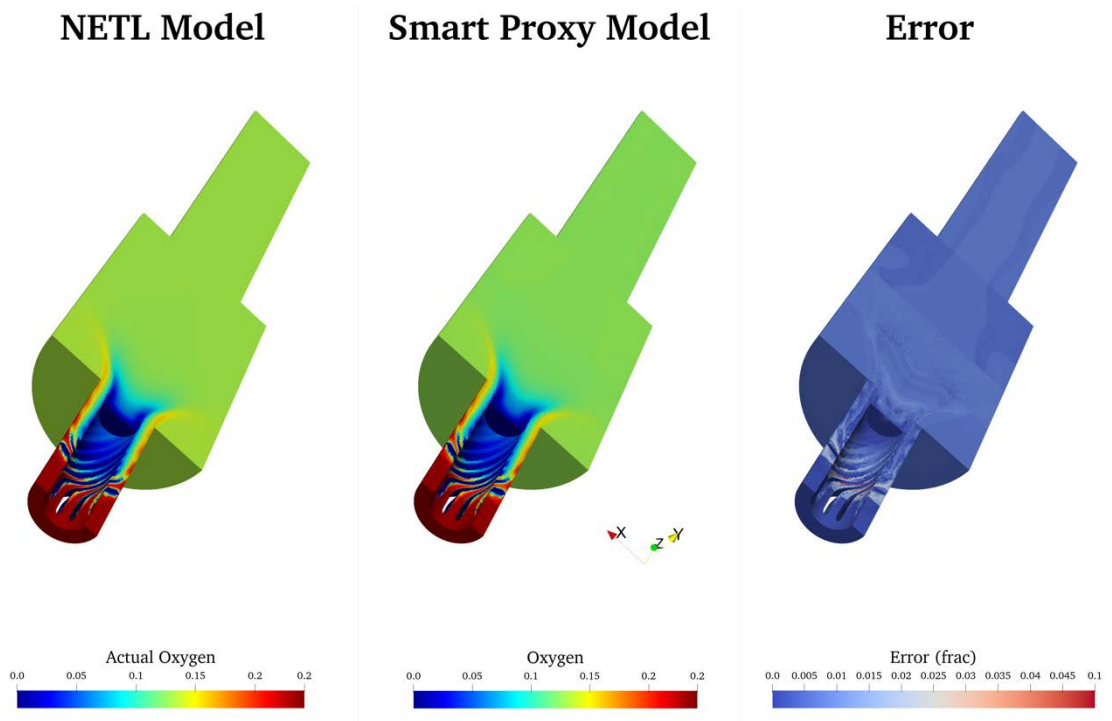


Figure 4-40: Step 4 Results (Half View) for Blind Validation Case 9 – Oxygen

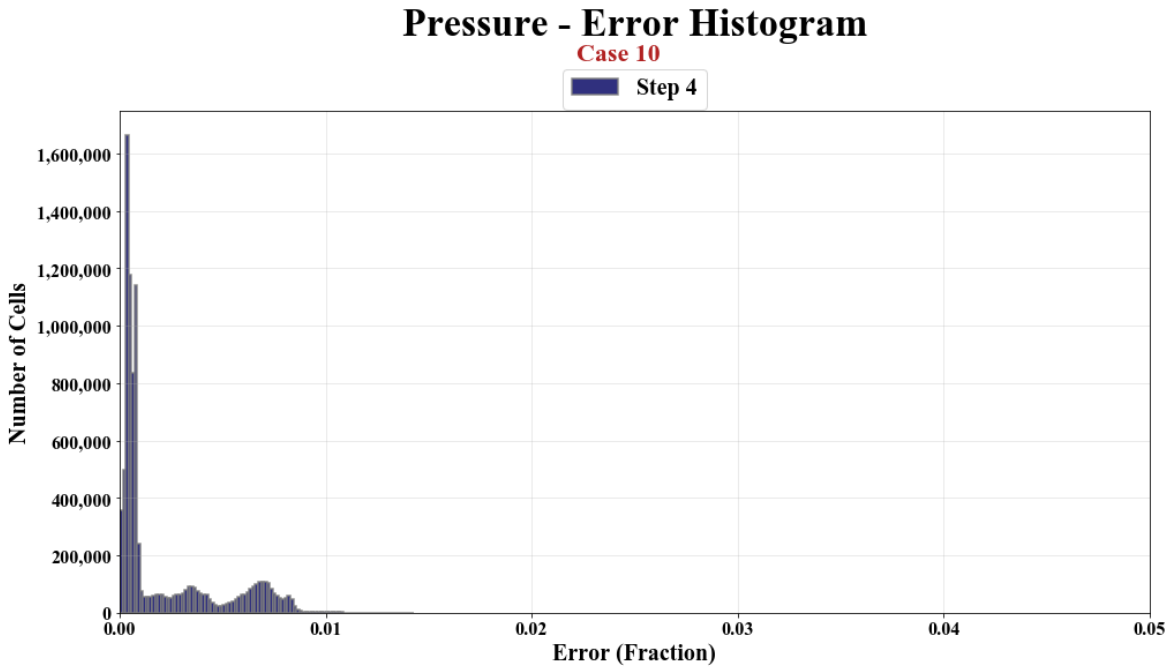


Figure 4-41: Step 4 Error Histogram for Blind Validation Case 10 – Pressure

Table 4-16: No. of Cells Under 100% Error for Blind Validation Case 10 – Pressure

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,291,730	100.0%
10% - 20%	0	0.0%
20% - 30%	0	0.0%
30% - 40%	0	0.0%
40% - 50%	0	0.0%
50% - 60%	0	0.0%
60% - 70%	0	0.0%
70% - 80%	0	0.0%
80% - 90%	0	0.0%
> 90%	0	0.0%

Table 4-17: No. of Cells Under 10% Error for Blind Validation Case 10 – Pressure

Percent Ranges	Number of Cells	Perc. Cells
< 2%	9,291,730	100.0%
2% - 4%	0	0.0%
4% - 6%	0	0.0%
6% - 8%	0	0.0%
8% - 10%	0	0.0%
> 10%	0	0.0%

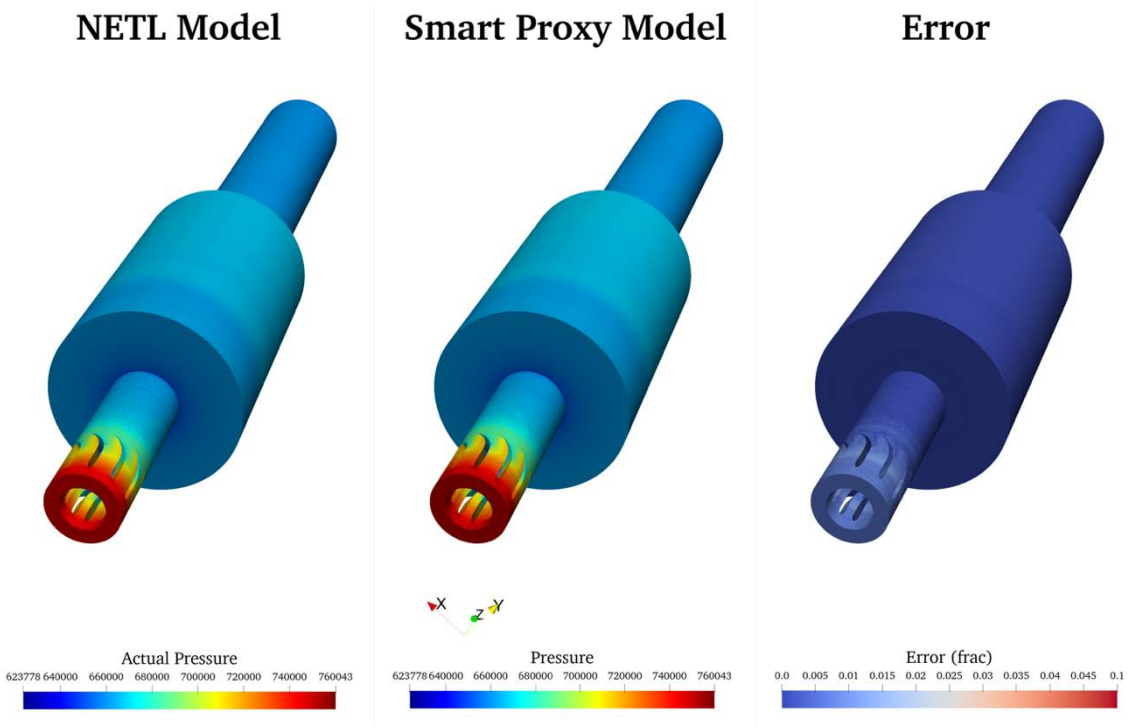


Figure 4-42: Step 4 Results (Entire System) for Blind Validation Case 10 – Pressure

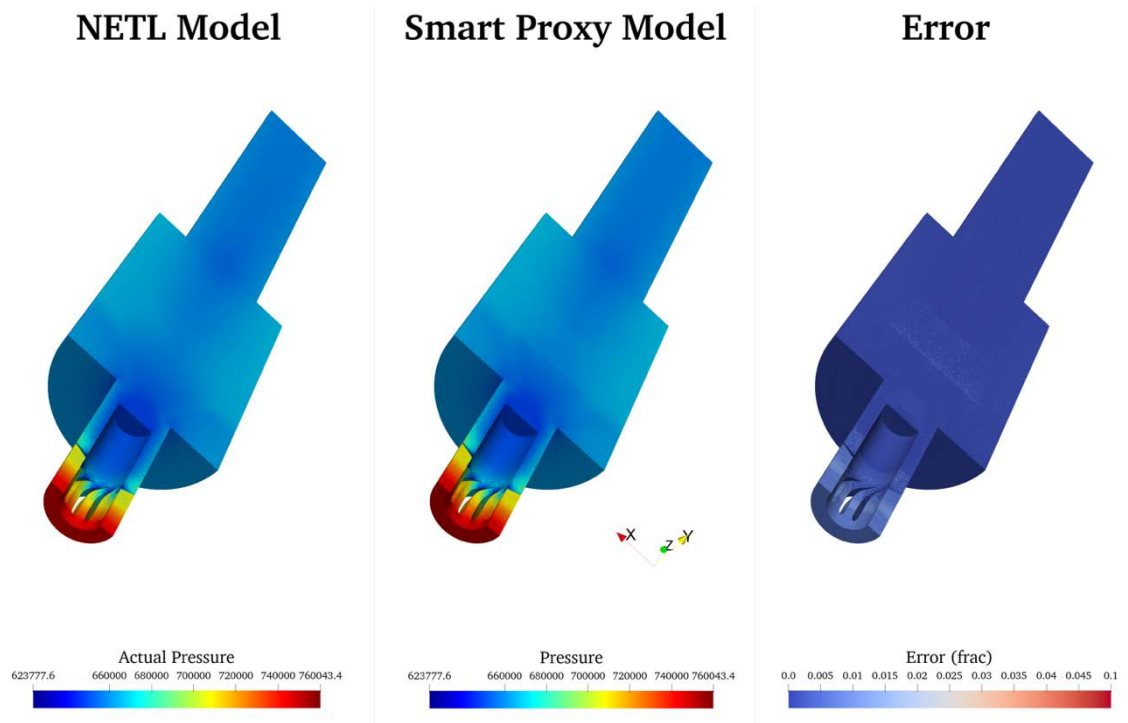


Figure 4-43: Step 4 Results (Half View) for Blind Validation Case 10 – Pressure

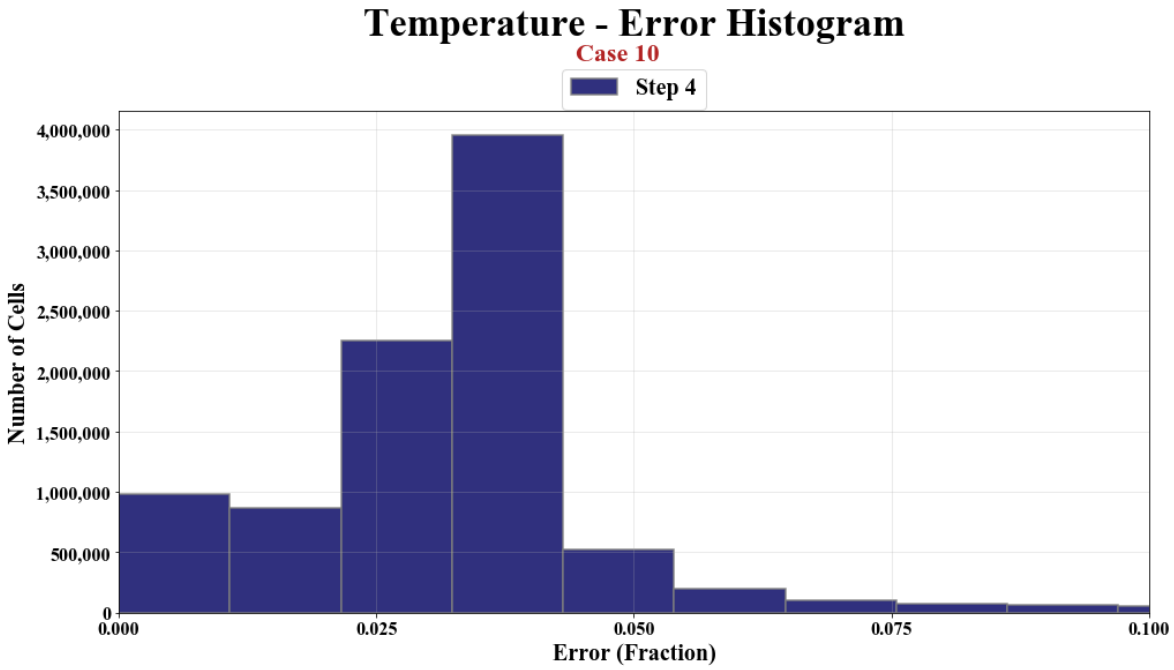


Figure 4-44: Step 4 Error Histogram for Blind Validation Case 10 – Temperature

Table 4-18: No. of Cells Under 100% Error for Blind Validation Case 10 – Temperature

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,053,161	97.432%
10% - 20%	187,203	2.015%
20% - 30%	40,293	0.434%
30% - 40%	9,600	0.103%
40% - 50%	1,052	0.011%
50% - 60%	308	0.003%
60% - 70%	78	0.001%
70% - 80%	20	0.0%
80% - 90%	9	0.0%
> 90%	6	0.0%

Table 4-19: No. of Cells Under 10% Error for Blind Validation Case 10 – Temperature

Percent Ranges	Number of Cells	Perc. Cells
< 2%	1,748,053	18.813%
2% - 4%	4,973,509	53.526%
4% - 6%	2,002,154	21.548%
6% - 8%	207,036	2.228%
8% - 10%	122,409	1.317%
> 10%	238,569	2.568%

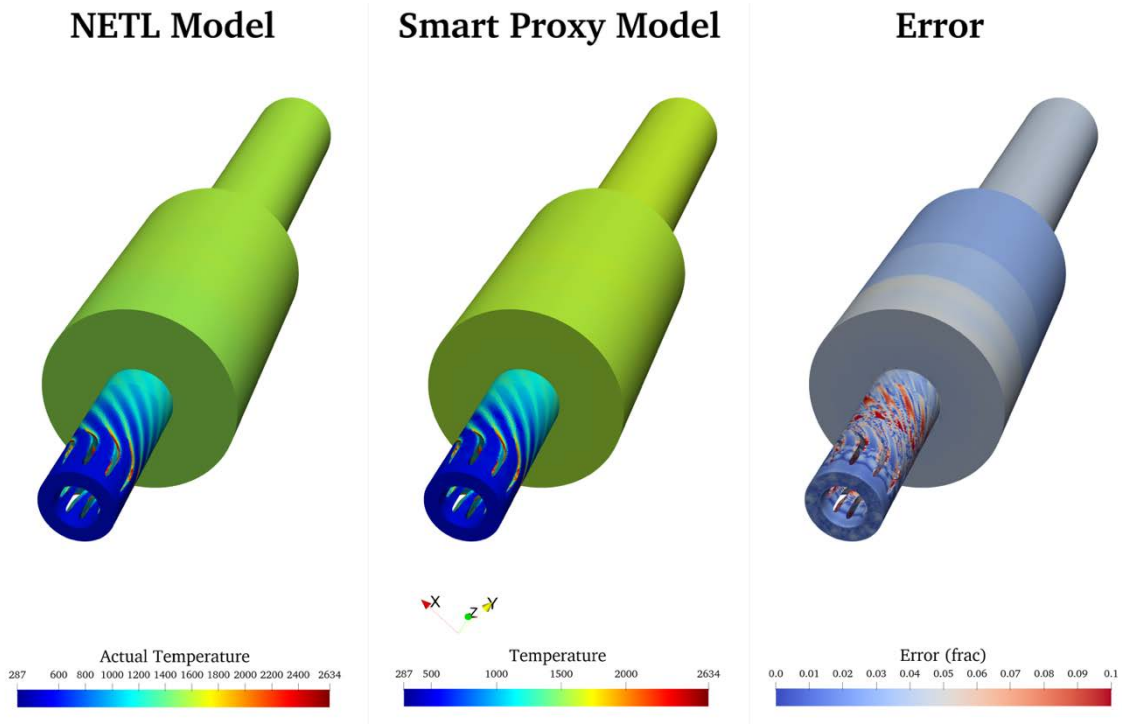


Figure 4-45: : Step 4 Results (Entire System) for Blind Validation Case 10 – Temperature

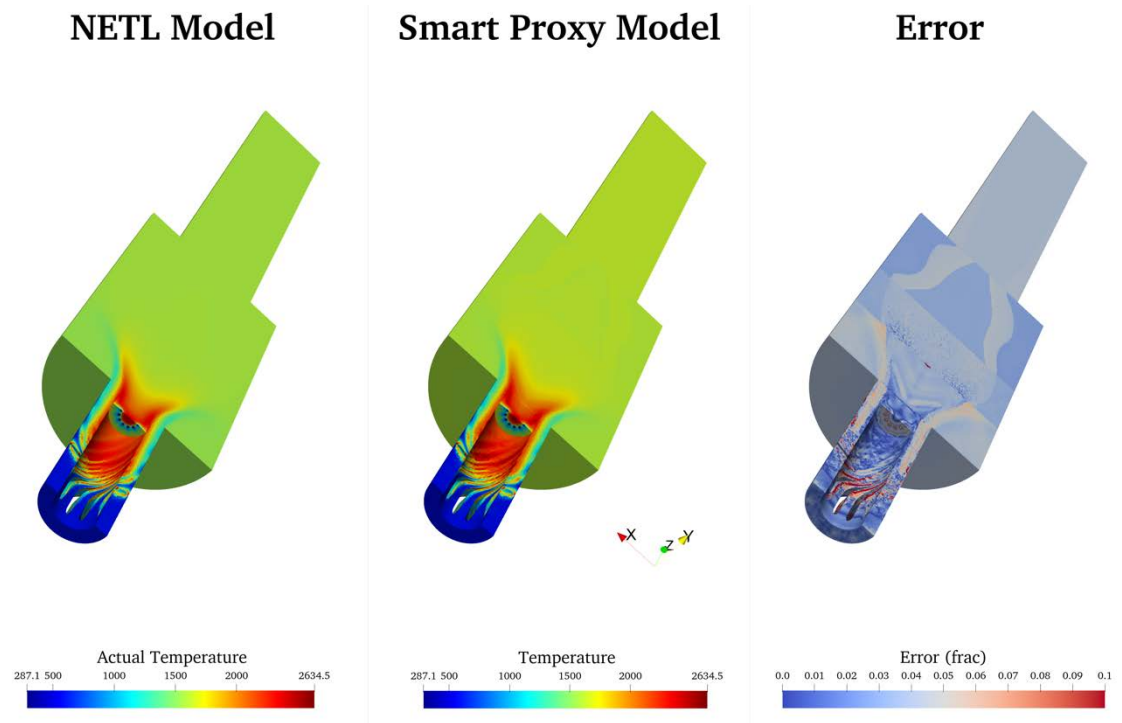


Figure 4-46: Step 4 Results (Half View) for Blind Validation Case 10 – Temperature

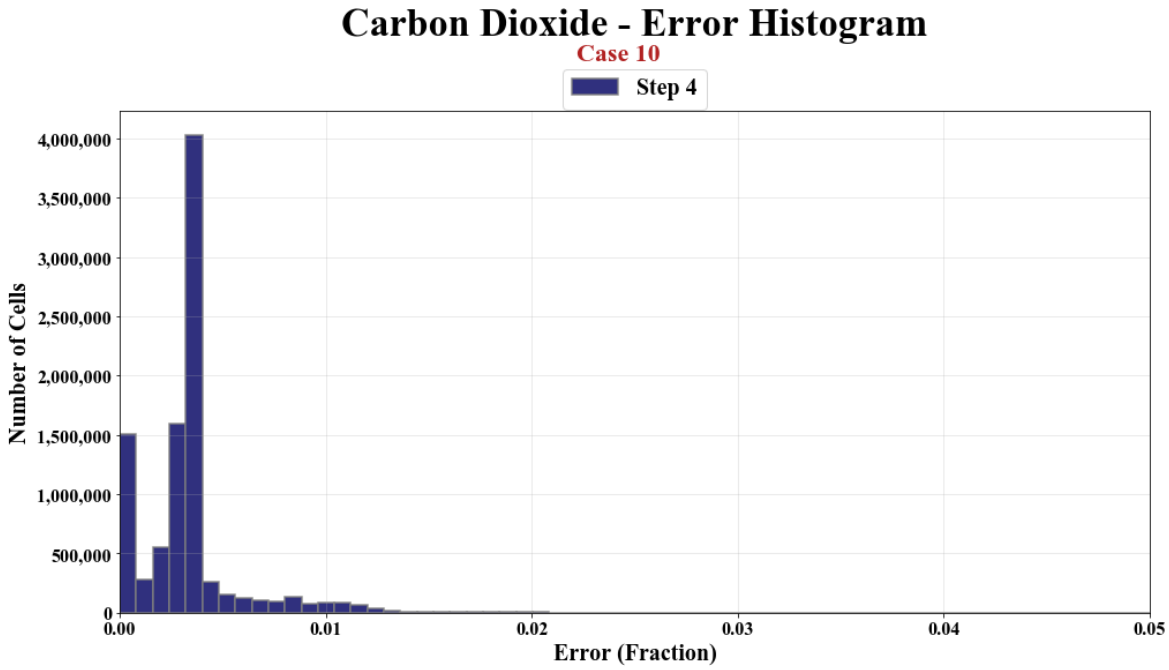


Figure 4-47: Step 4 Error Histogram for Blind Validation Case 10 – Carbon Dioxide

Table 4-20: No. of Cells Under 100% Error for Blind Validation Case 10 – Carbon Dioxide

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,291,730	100.0%
10% - 20%	0	0.0%
20% - 30%	0	0.0%
30% - 40%	0	0.0%
40% - 50%	0	0.0%
50% - 60%	0	0.0%
60% - 70%	0	0.0%
70% - 80%	0	0.0%
80% - 90%	0	0.0%
> 90%	0	0.0%

Table 4-21: No. of Cells Under 10% Error for Blind Validation Case 10 – Carbon Dioxide

Percent Ranges	Number of Cells	Perc. Cells
< 2%	9,252,938	99.583%
2% - 4%	33,072	0.356%
4% - 6%	5,699	0.061%
6% - 8%	20	0.0%
8% - 10%	1	0.0%
> 10%	0	0.0%

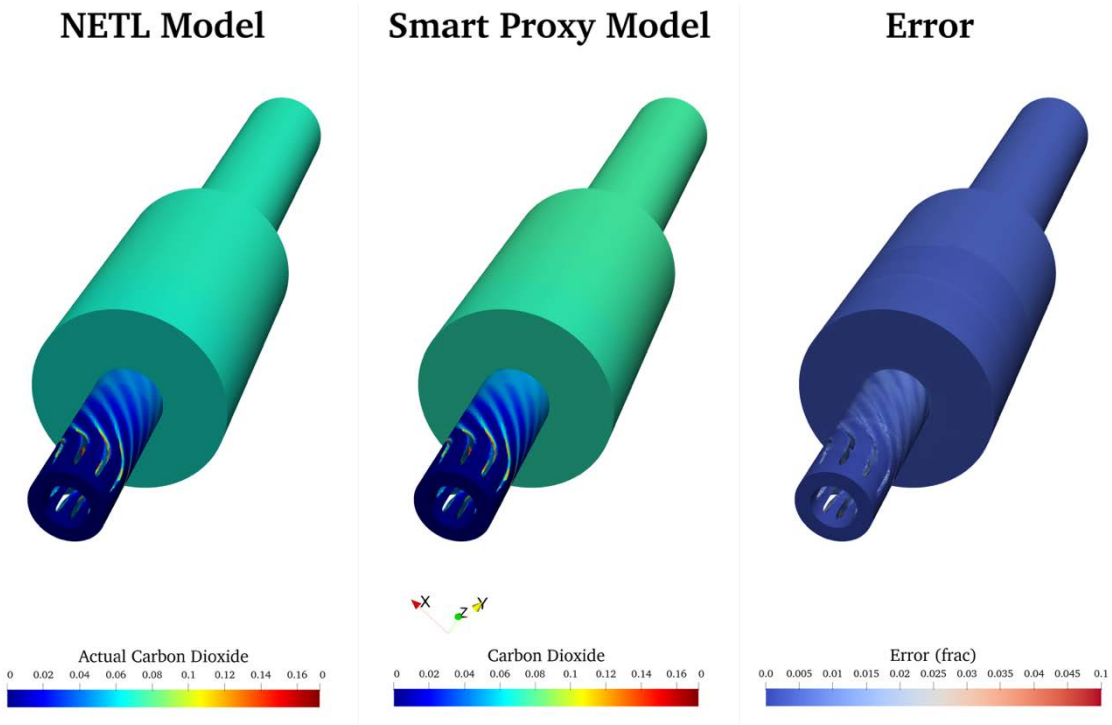


Figure 4-48: Step 4 Results (Entire System) for Blind Validation Case 10 – Carbon Dioxide

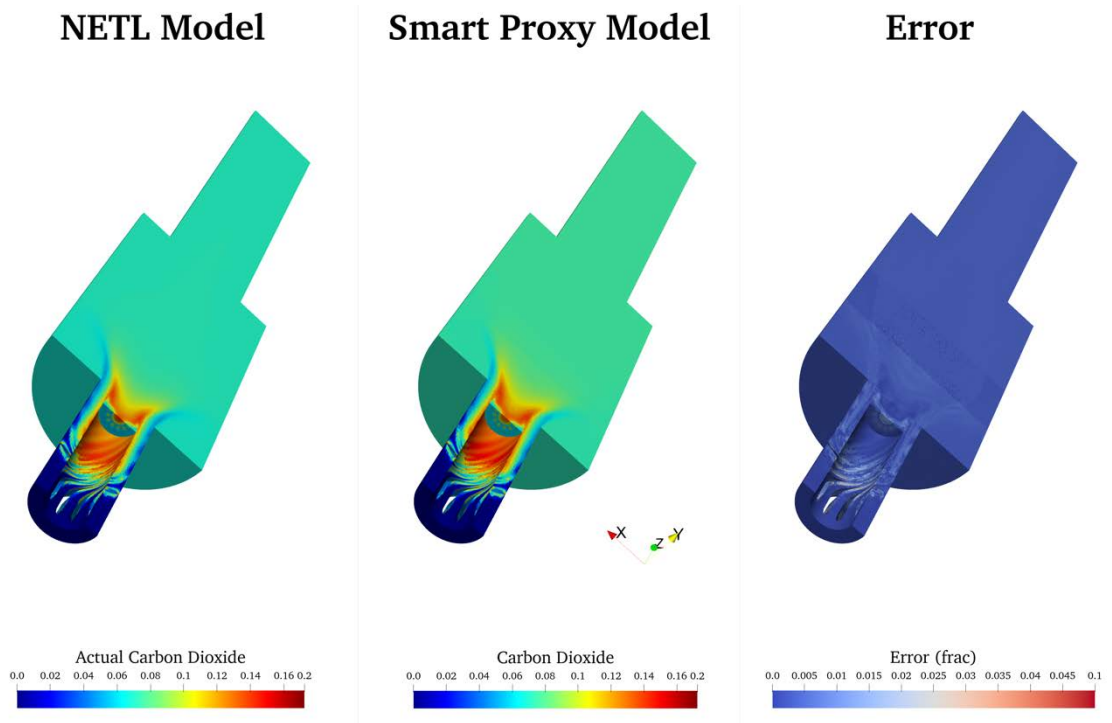


Figure 4-49: Step 4 Results (Half View) for Blind Validation Case 10 – Carbon Dioxide

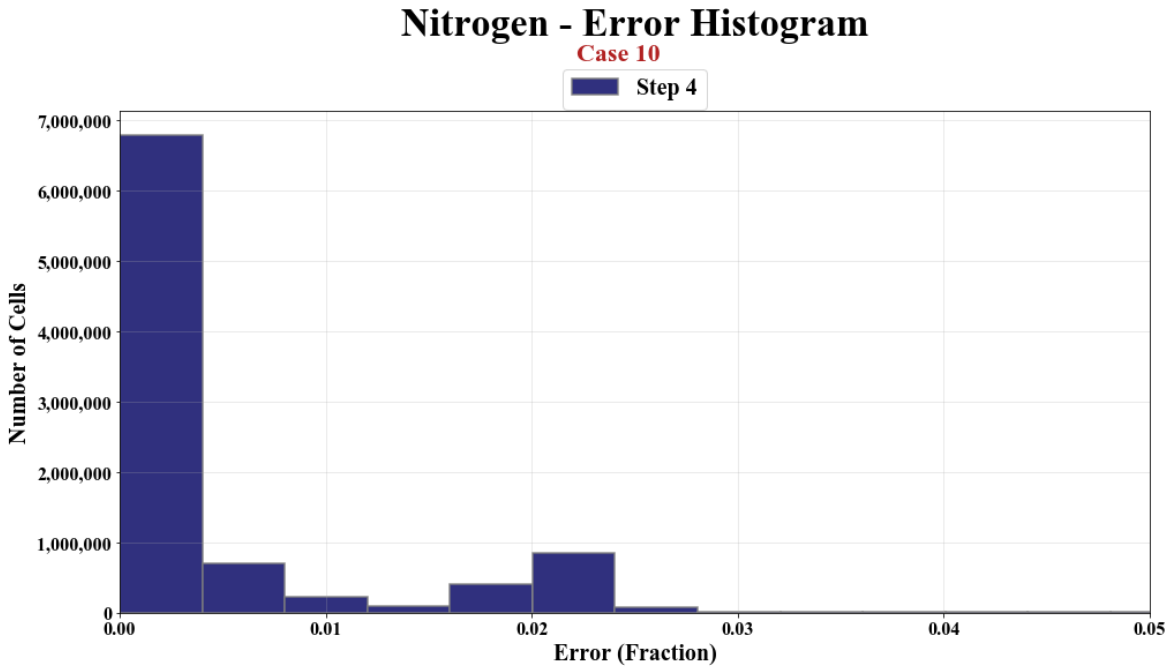


Figure 4-50: Step 4 Error Histogram for Blind Validation Case 10 –Nitrogen

Table 4-22: No. of Cells Under 100% Error for Blind Validation Case 10 – Nitrogen

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,256,463	99.62%
10% - 20%	27,707	0.298%
20% - 30%	7,209	0.078%
30% - 40%	350	0.004%
40% - 50%	1	0.0%
50% - 60%	0	0.0%
60% - 70%	0	0.0%
70% - 80%	0	0.0%
80% - 90%	0	0.0%
> 90%	0	0.0%

Table 4-23: No. of Cells Under 10% Error for Blind Validation Case 10 – Nitrogen

Percent Ranges	Number of Cells	Perc. Cells
< 2%	8,222,525	88.493%
2% - 4%	982,487	10.574%
4% - 6%	25,745	0.277%
6% - 8%	15,571	0.168%
8% - 10%	10,135	0.109%
> 10%	35,267	0.38%

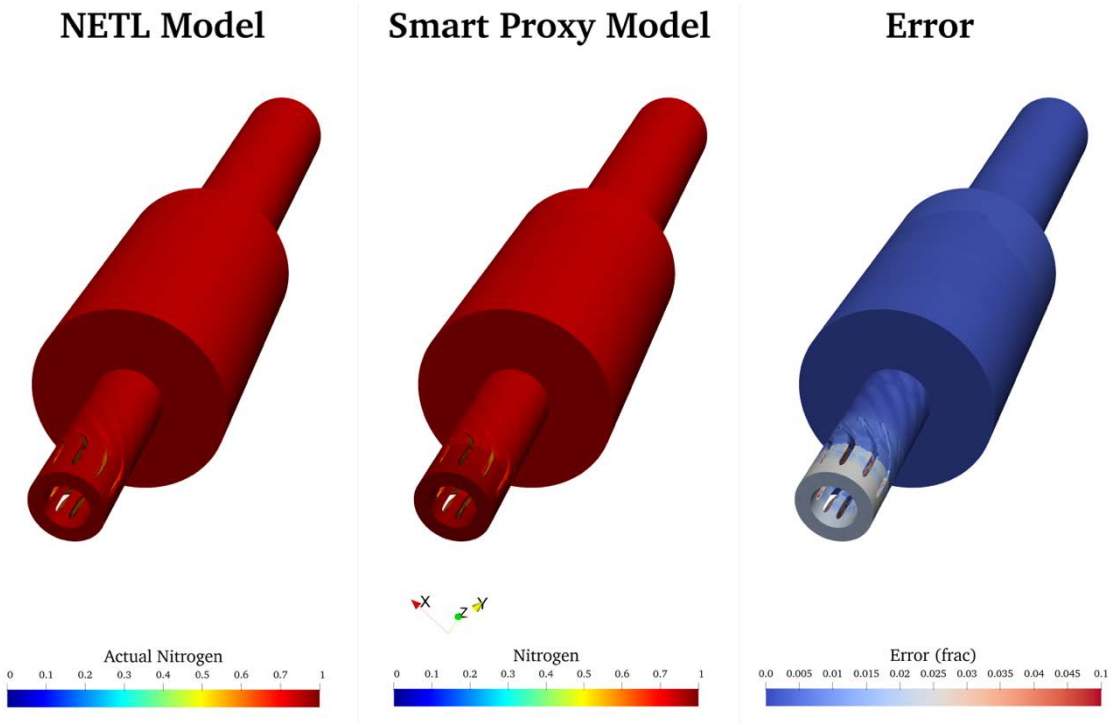


Figure 4-51: Step 4 Results (Entire System) for Blind Validation Case 10 –Nitrogen

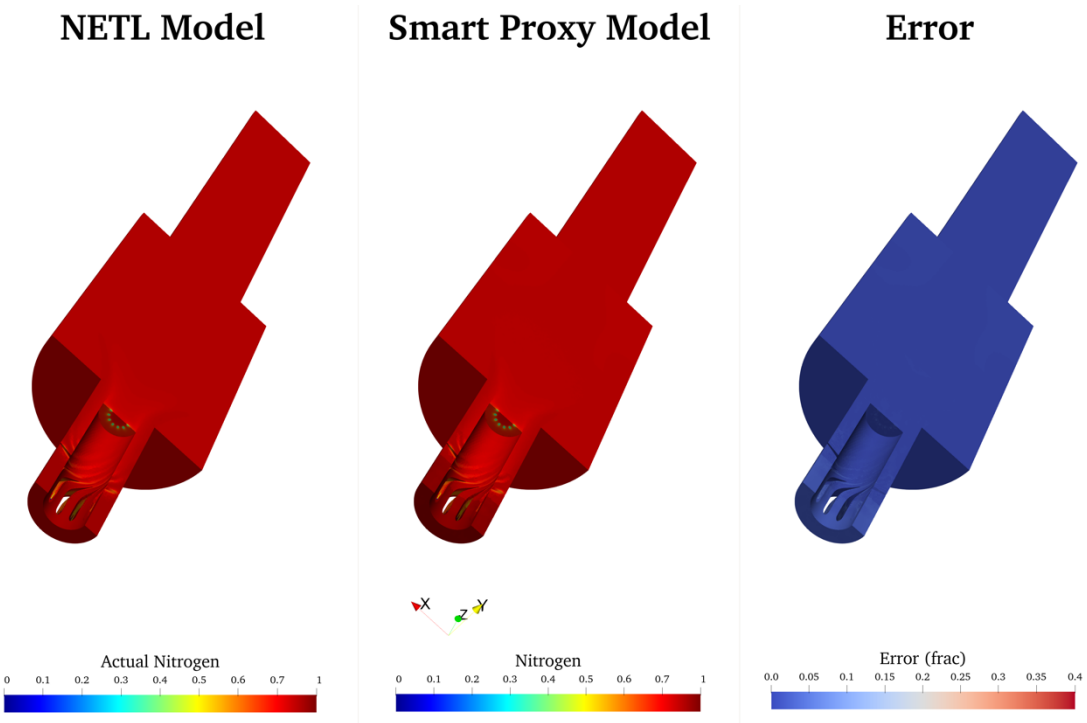


Figure 4-52: Step 4 Results (Half View) for Blind Validation Case 10 –Nitrogen

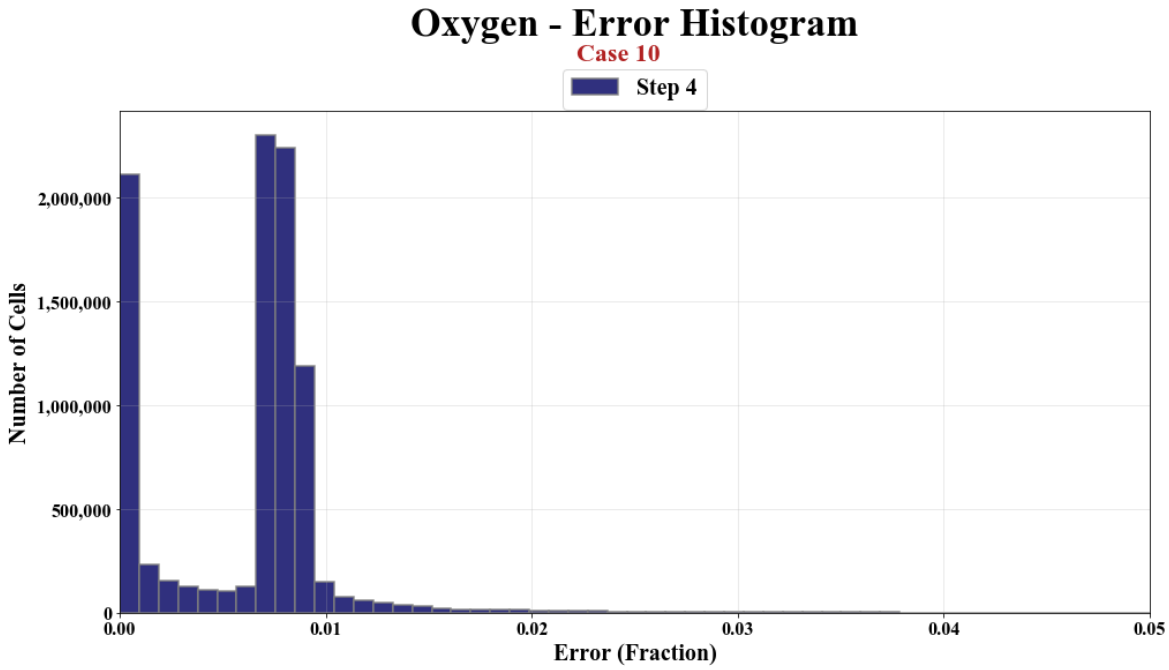


Figure 4-53: Step 4 Error Histogram for Blind Validation Case 10 –Oxygen

Table 4-24: No. of Cells Under 100% Error for Blind Validation Case 10 – Oxygen

Percent Ranges	Number of Cells	Perc. Cells
< 10%	9,291,730	100.0%
10% - 20%	0	0.0%
20% - 30%	0	0.0%
30% - 40%	0	0.0%
40% - 50%	0	0.0%
50% - 60%	0	0.0%
60% - 70%	0	0.0%
70% - 80%	0	0.0%
80% - 90%	0	0.0%
> 90%	0	0.0%

Table 4-25: No. of Cells Under 10% Error for Blind Validation Case 10 – Oxygen

Percent Ranges	Number of Cells	Perc. Cells
< 2%	9,178,770	98.784%
2% - 4%	86,990	0.936%
4% - 6%	23,215	0.25%
6% - 8%	2,726	0.029%
8% - 10%	29	0.0%
> 10%	0	0.0%

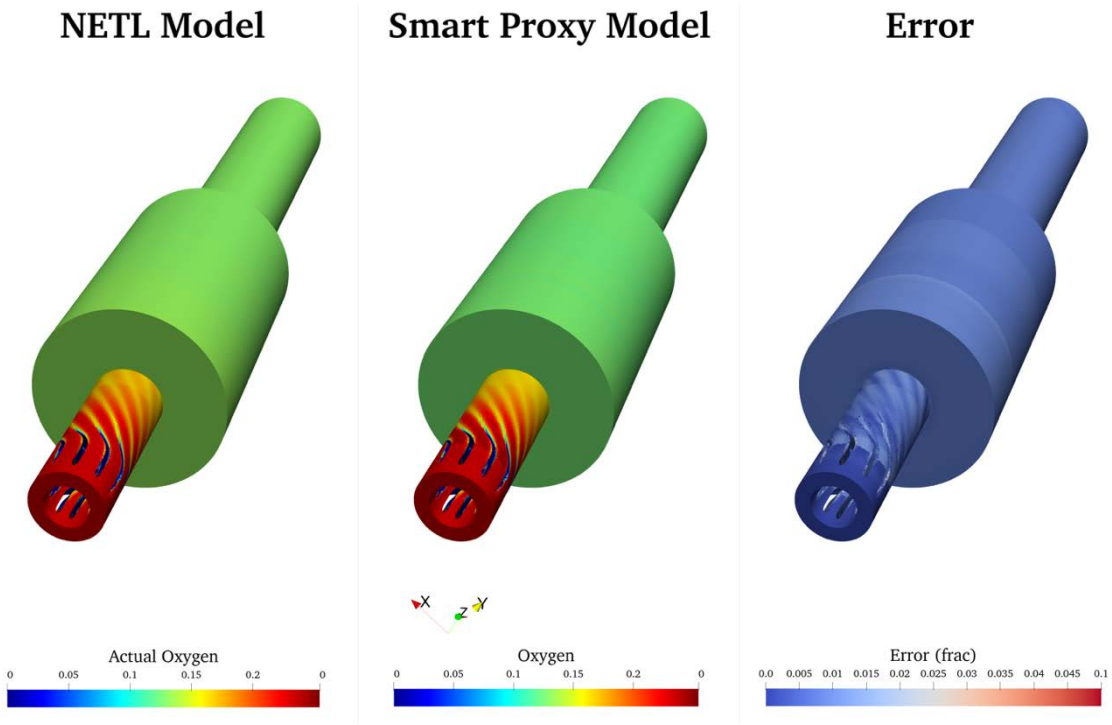


Figure 4-54: Step 4 Results (Entire System) for Blind Validation Case 10 –Oxygen

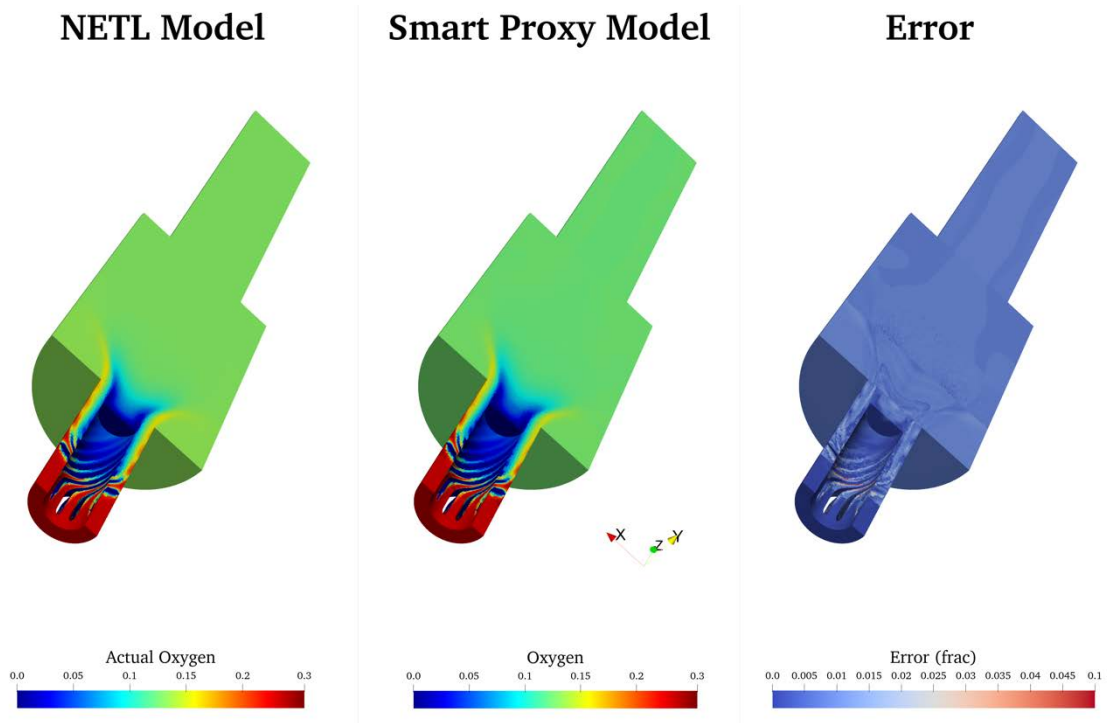


Figure 4-55: Step 4 Results (Half View) for Blind Validation Case 10 –Oxygen

5 CONCLUSIONS

It has been successfully demonstrated that a data driven predictive model can reproduce the results of CFD simulation of natural gas combustion in a high-pressure combustor system. The developed smart proxy replicates the thermal-flow patterns of pressure, temperature, and species concentrations (nitrogen, oxygen and carbon-dioxide) with a percent error of not more than 10%, and a faster execution time compared to the numerical CFD simulation approach. A single CFD simulation run of the B6 Combustor model takes about 24 hours or more to complete on the NETL HPC with single node 196GB RAM, 40 cores while the smart proxy generates results in about 5 - 6 minutes when executed on the same HPC configuration. This time can be further reduced if the compute resource limitations (computer memory and speed) are properly addressed and the system is modeled without having to divide into smaller development sections.

More importantly, the smart proxy model achieved this level of accuracy using a very minimal amount of data; only six CFD simulation runs of the B6 Combustor model was used in developing the B6 smart proxy model. While CFD simulations require extensive compute resources, the developed smart proxy can be deployed on commodity computers (inexpensive laptop or desktop machines). This proves that this technology can contribute significantly to research studies that are targeted at determining optimal design and operating conditions that would maximize the efficiency of complex power generation systems.

5.1 RECOMMENDATIONS

The final modeling results shown in Section 4.4.2 of this report shows that the smart proxy development framework applied can replicate the results of CFD simulation of a single-phase reaction flow in a high-pressure combustor to a reasonable degree of accuracy. The results however could be improved, especially for a thermal flow field variable such as temperature. Apart from the simulation model boundary conditions, no reaction data from FLUENT software was used in developing the smart proxy model. It might be worth identifying potential contribution of the other successfully modeled transport and specie variables (pressure, oxygen, nitrogen and carbon-dioxide) to the modeling of the thermal flow field. Perhaps, the neural networks may further learn the reaction flow characteristics if transport data such as the turbulent kinetic energy and dissipation rate are also modeled.

Training a single neural network on the entire system for each attribute will increase the amount of information provided to any single neural network in the smart proxy model and improve performance as opposed to a total of seven networks currently built for each attribute. In order to efficiently fulfill the objective of the next phase of the project, which is a more complex multi-

phase reaction flow in a coal-fired boiler, an access to a high-performance-computing (HPC) facility is highly recommended.

6 REFERENCES

[1] W. Fullmer and C. Hrenva, "Quantitative assessment of fine-grid kinetic theory based predictions of mean-slip in unbounded fluidization," *AIChE Journal,* vol. DOI 10.1002/aic. 62, pp. 11-17, 2016.

[2] M. Shahnam, A. Gel, J.-F. Dietiker, A. K. Subramaniyan and J. Musser, " The Effect of Grid Resolution and Reaction Models in Simulation of a Fluidized Bed Gasifier through Nonintrusive Uncertainty Quantification Techniques," *ASME Journal of Verification, Validation and Uncertainty Quantification,* 2016.

[3] A. Ansari, S. Mohaghegh, M. Shahnam, J. F. Dietiker and T. Li, "Data Driven Smart Proxy for CFD Application of Big Data Analytics & Machine Learning in Computational Fluid Dynamics, Part Two: Model Building at the Cell Level," NETL-PUB-21634, NETL Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory, Morgantown, WV, 2017.

[4] ANSYS Inc., *ANSYS Fluent User's Guide, Release 16.1,* Southpointe 2600 ANSYS Drive, Canonsburg, PA 15317, April 2015.

[5] S. Gubba, D. Ingham, K. Larsen, L. Ma and M. Pourkashanian, "Numerical modelling of the co-firing of pulverised coal and straw in a 300MWe tangentially fired boiler," *Fuel processing technology,* vol. 104, pp. 181-188, 2012.

[6] H. Zhou, Y. Yang, H. Liu and Qi. Hang, "Numerical simulation of the combustion characteristics of a low NOx swirl burner: Influence of the primary air pipe," *Fuel,* vol. 130, pp. 168 - 176, 2014.

[7] N. Modliński, P. Madejski, T. Janda, K. Szczepanek and W. Kordylewski, "A validation of computational fluid dynamics temperature distribution prediction in a pulverized coal boiler with acoustic temperature measurement," *Energy,* vol. 92, pp. 77 - 86, 2015.

[8] W. Sun, W. Zhong, A. Yu, L. Liu and Y. Qian, "Numerical investigation on the flow, combustion, and NOX emission characteristics in a 660 MWe tangential firing ultra-supercritical boiler," *Advances in Mechanical Engineering,* vol. 8, 2016.

[9] S. Chen, B. He, D. He, Y. Cao, G. Ding, X. Liu, Z. Duan, X. Zhang, J. Song and X. Li, "Numerical investigations on different tangential arrangements of burners for a 600 MW utility boiler," *Energy*, vol. 122, pp. 287 - 300, 2017.

[10] X. Ge, J. Dong., H. Fan, Z. Zhang, X. Shang, X. Hu and J. Zhang, "Numerical investigation of oxy-fuel combustion in 700 °C-ultra-supercritical boiler," *Fuel*, vol. 207, pp. 602 - 614, 2017.

[11] C. Choi and C. Kim, "Numerical investigation on the flow, combustion and NOx emission characteristics in a 500 MWe tangentially fired pulverized-coal boiler; Fuel," vol. 88, pp. 1720 - 1731, 2009.

[12] P. Edge, P. Heggs, M. Pourkashanian and A. Williams, "An integrated computational fluid dynamics–process model of natural circulation steam generation in a coal-fired power plant," *Computers and Chemical Engineering* , vol. 35, no. 12, pp. 2618 - 2631, 2011.

[13] L. Shi, Z. Fu, X. Duan, C. Cheng, Y. Shen, B. Liu and R. Wang, "Influence of combustion system retrofit on NOx formation characteristics in a 300 MW tangentially fired furnace," *Applied Thermal Engineering*, vol. 98, pp. 766 - 777, 2016.

[14] T.F. Smith, Z.F. Shen and J.N. Friedman, "Evaluation of Coefficients for the Weighted Sum of Gray Gases Model," *Heat Transfer*, vol. 104, no. 4, pp. 602 - 608, 1982.

[15] C. Yin, "Refined Weighted Sum of Gray Gases Model for Air-Fuel Combustion and Its Impacts," *Energy Fuels*, vol. 27, p. 2013, 6287 - 6294.

[16] P.-N. Tan, M. Steinbach, A. Karpatne and V. Kumar, "Cluster Analysis: Basic Concepts and Algorithms," in *Introduction to Data Mining*, Pearson, 2018, pp. 487 - 493.

[17] S. Sharma, "Artificial Neural Network (ANN)," 8 August 2017. [Online]. Available: https://www.datasciencecentral.com/profiles/blogs/artificial-neural-network-ann-in-machine-learning.

[18] F. Chollet, "Keras," https://keras.io, 2015.

[19] J. Brownlee, "How to Control the Stability of Training Neural Networks With the Batch Size," Machine Learning Mastery Pty. Ltd, 21 January 2019. [Online]. Available:

https://machinelearningmastery.com/how-to-control-the-speed-and-stability-of-training-neural-networks-with-gradient-descent-batch-size/. [Accessed 10 March 2020].

[20] T. E. Oliphant, Guide to NumPy, 2006.

[21] S. Mohaghegh and S. Ameri, "Artificial Neural Network As A Valuable Tool For Petroleum Engineers," *SPE 29220,* 1995.

[22] S. Mohaghegh, R. Arefi, S. Ameri and D. Rose, "Design and Development of an Artificial Neural Network for Estimation of Formation Permeability," *SPE Computer Applications, doi:10.2118/28237-PA,* vol. 7, no. 6, pp. 151 - 155, 1995.

[23] S. Mohaghegh, "Top-down, intelligent reservoir modeling of oil and gas producing shale reservoirs: Case studies," *International Journal of Oil Gas and Coal Technology,* p. 12, 2012.

[24] S. Mohaghegh, "Smart Proxy Modeling for Numerical Reservoir Simulations - Big Data Analytics in E&P," 6 October 2015. [Online]. Available: https://webevents.spe.org/products/smart-proxy-modeling-for-numerical-reservoir-simulations-big-data-analytics-in-ep.

[25] F. Alenezi and S. Mohaghegh, "A data-driven smart proxy model for a comprehensive reservoir simulation," IEEE, Riyadh, 2016.

[26] S. Boosari, "Predicting the Dynamic Parameters of Multiphase Flow in CFD (Dam-Break Simulation) Using Artificial Intelligence - (Cascading Deployment)," *Fluids,* 2019.

[27] Geo. Richards, Douglas L. Straub, Donald H. Ferguson and Edward H. Robey, "LNG Interchangeability/Gas Quality: Results of the National Energy Technology Laboratory's Research for the FERC on Natural Gas Quality and Interchangeability," U.S. Department of Energy, National Energy Technology Laboratory (NETL), 2007.

[28] M. G. Carvalho and P. J. Coelho, "Heat Transfer in Gas Turbine Combustors," *J. Thermophysics,* vol. 3, no. 2, 1989.

[29] L. Klobucar, "Thermal Radiation Heat Transfer between Surfaces," University of Ljubljana, 2016.

[30] U.S. Energy Information, "What is the efficiency of different types of power plants?," 8 August 2019. [Online]. Available: https://www.eia.gov/tools/faqs/faq.php?id=107&t=3.

[31] National Association of Clean Air Agencies, "Optimize Power Plant Operations," [Online]. Available: http://www.4cleanair.org/sites/default/files/Documents/Chapter_1.pdf.

[32] Duke Energy, "Electricity from coal," [Online]. Available: https://www.duke-energy.com/energy-education/how-energy-works/electricity-from-coal.

[33] G. Hoffman, "New Mexico's Coal and Electricity Industries," 2014. [Online]. Available: https://geoinfo.nmt.edu/publications/periodicals/litegeology/36/lg_v36.pdf.

[34] H. Gu, H. Zhu, Y. Cui, F. Si, R. Xue, H. Xi and J. Zhang, "Optimized scheme in coal-fired boiler combustion based on information," *Elsevier,* 2018.

[35] T. Sanpasertparnich and A. Aroonwilas, "Simulation and optimization of coal-fired power plants," *Elsevier,* 2009.

[36] R. Saripalli, T. Wang and B. Day, "SIMULATION OF COMBUSTION AND THERMAL FLOW," in *Industrial Energy Technology Conference,* New Orleans, 2005.

[37] U.S. Department of Energy, National Energy Technology Laboratory, "The Advanced Tangentially Fired Combustion Techniques for the Reduction of Nitrogen Oxides (NO) Emissions From Coal-Fired Boilers Demonstration Project: A DOE Assessment," DOE/NETL-2000/1122, 2000.

[38] B. Miller, "Nitrogen oxides formation and control," in *Fossil Fuel Emissions Control Technologies,* Elsevier, 2015, pp. 243 - 280.

[39] J. Mahato, "Boiler in Thermal Power Plant," 2013. [Online]. Available: https://www.coalhandlingplants.com/boiler-in-thermal-power-plant/. [Accessed 15 March 2020].

[40] A.H. Al-Abbas, J. Naser and D. Dodds, "CFD modelling of air-fired and oxy-fuel combustion in a large-scale furnace at Loy Yang A brown coal power station, Fuel," *Fuel,* vol. 102, pp. 646 - 665, 2012.

[41] A.H. Al-Abbas, J. Naser., D. Dodds and A. Blicblau, "Numerical Modelling of Oxy-Fuel Combustion in a Full-Scale Tangentially-Fired Pulverised Coal Boiler," *Procedia Engineering*, vol. 56, pp. 375 - 380, 2013.

[42] D. Tian, L. Zhong, P. Tan, L. Ma, Q. Fang, C. Zhang, D. Zhang and G. Chen, "Influence of vertical burner tilt angle on the gas temperature deviation in a 700 MW low NOx tangentially fired pulverised-coal boiler," *Fuel Processing Technology*, vol. 138, pp. 616 - 628, 2015.

[43] S. Li, Z. Chen, E. He, B. Jiang, Z. Li and Q. Wang, "Combustion characteristics and NOx formation of a retrofitted low-volatile coal-fired 330 MW utility boiler under various loads with deep-air-staging," *Applied Thermal Engineering*, vol. 110, pp. 223 - 233, 2017.

[44] A.A. Bhuiyan and J. Naser, "CFD modelling of co-firing of biomass with coal under oxy-fuel combustion in a large scale power plant," *Fuel*, vol. 159, pp. 150 - 168, 2015.

7 APPENDIX

7.1 Model Development Step 3: Q1 Combustor Results – Nitrogen

Nitrogen – Case 1

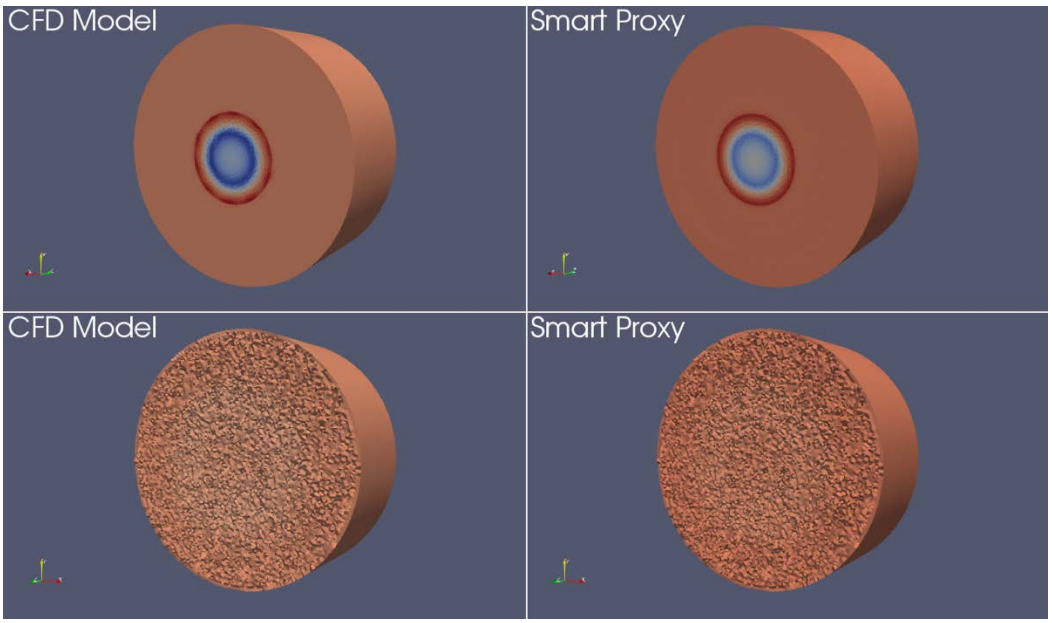


Figure 7-1: Step 3 Results (Q1 Combustor Full View) for Validation Case 1 – Nitrogen

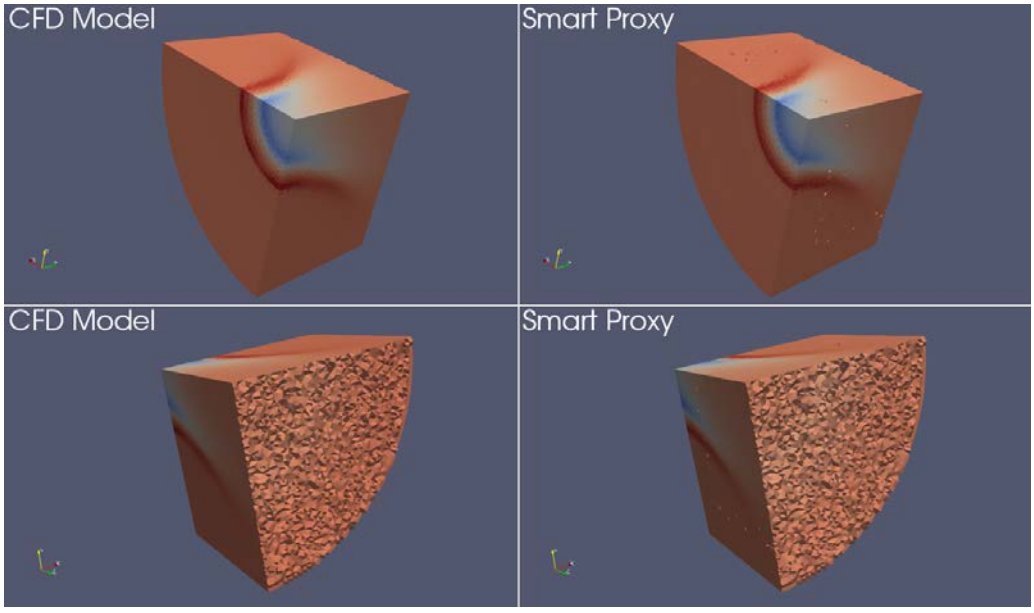


Figure 7-2: Step 3 Results (Q1 Combustor Full View) for Validation Case 1 – Nitrogen

Nitrogen – Case 5

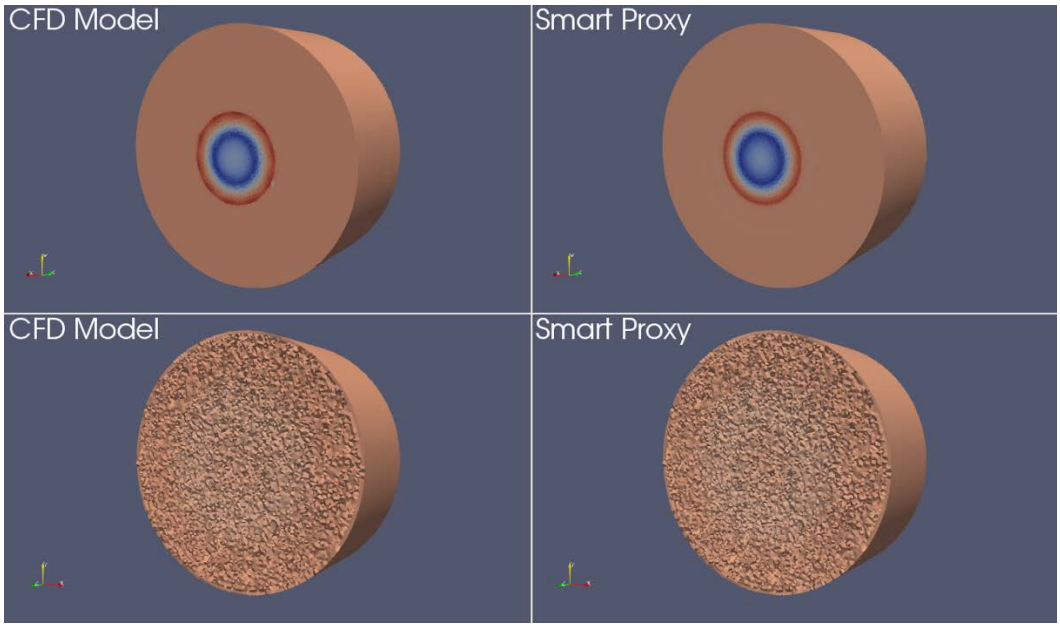


Figure 7-3: Step 3 Results (Q1 Combustor Full View) for Validation Case 5 – Nitrogen

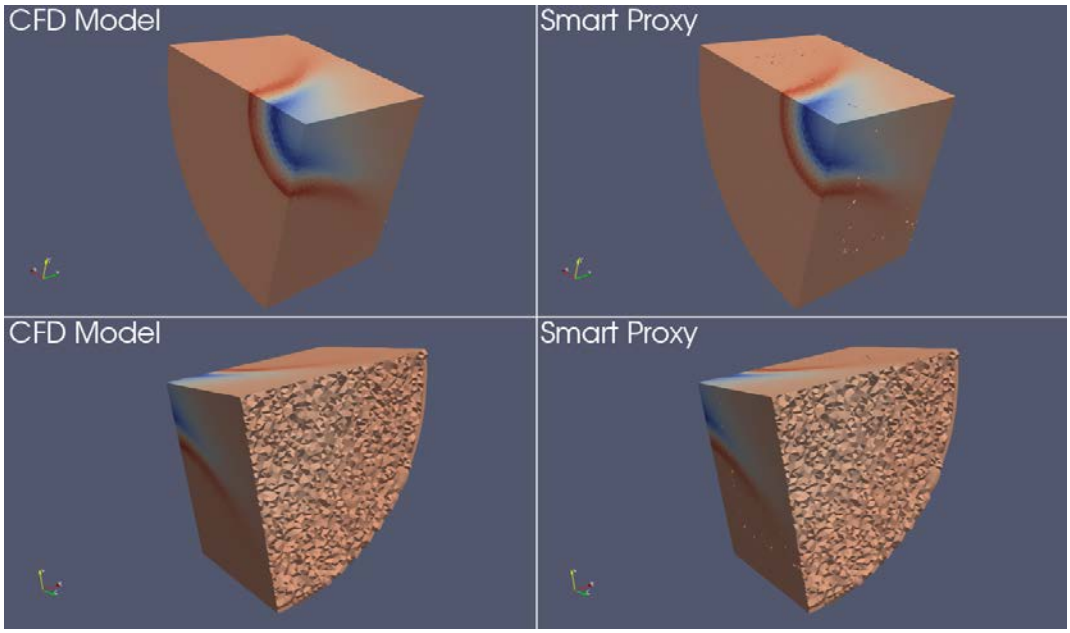


Figure 7-4: Step 3 Results (Q1 Combustor Full View) for Validation Case 5 – Nitrogen

7.2 Model Development Step 3: Q1 Combustor Results – Oxygen

Oxygen – Case 1

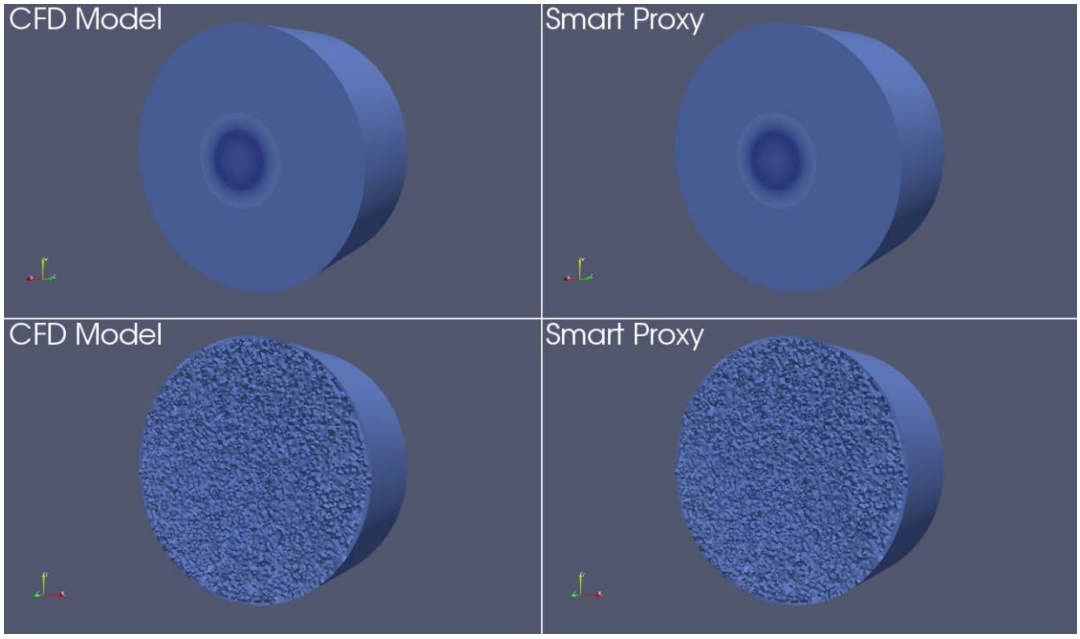


Figure 7-5: Step 3 Results (Q1 Combustor Full View) for Validation Case 1 – Oxygen

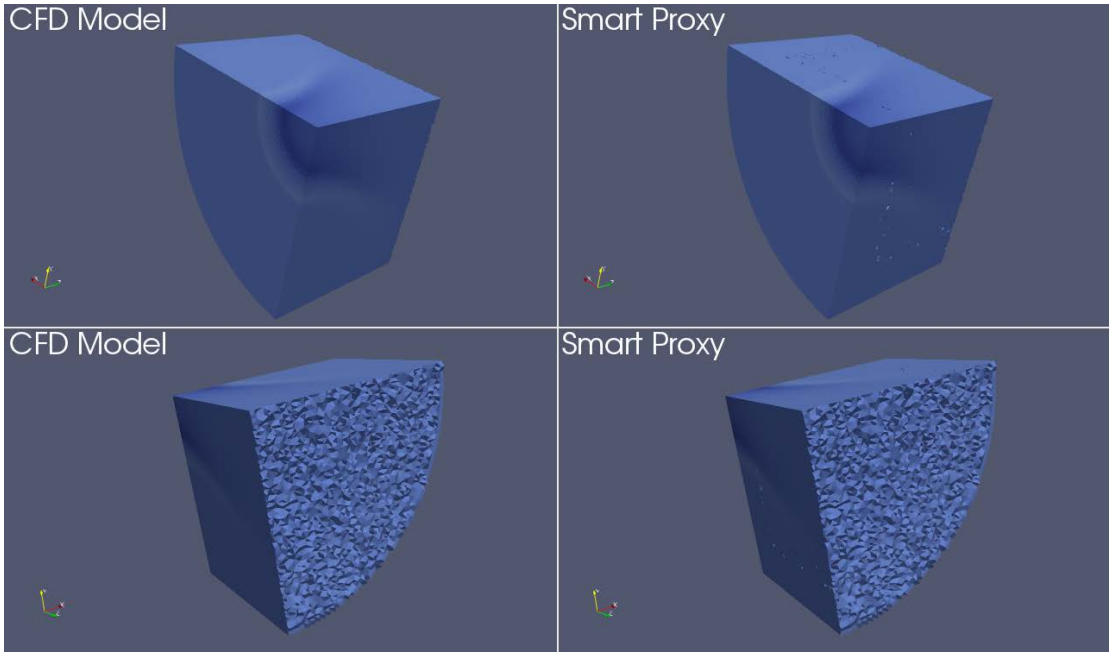


Figure 7-6: Step 3 Results (Q1 Combustor Full View) for Validation Case 1 – Oxygen

Oxygen – Case 5

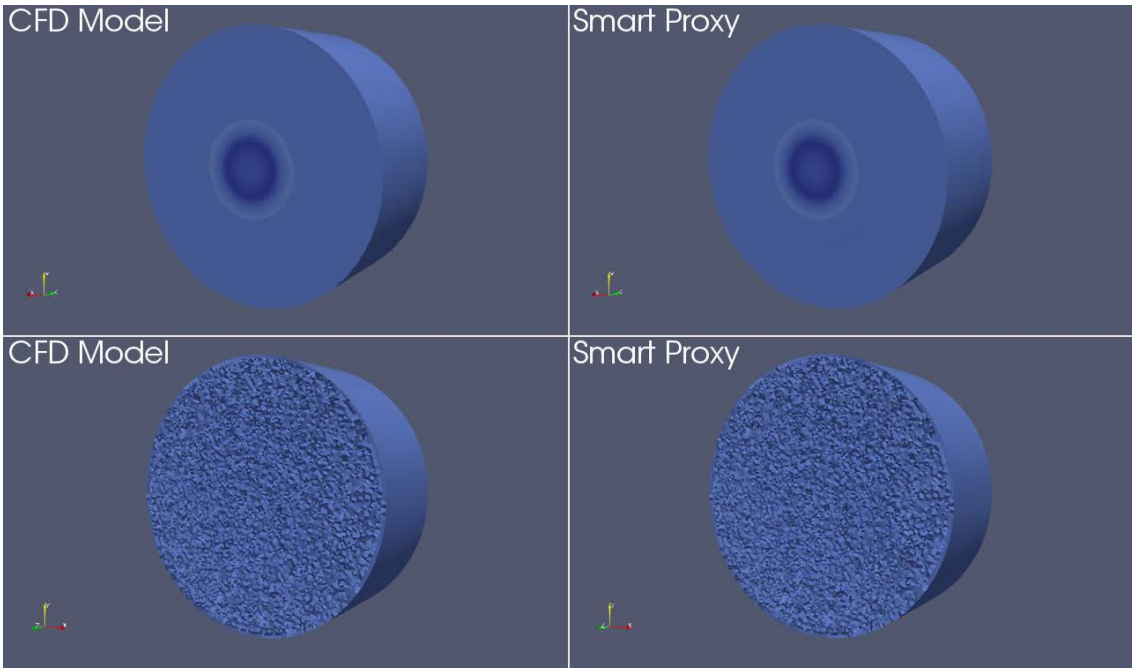


Figure 7-7: Step 3 Results (Q1 Combustor Full View) for Validation Case 5 – Oxygen

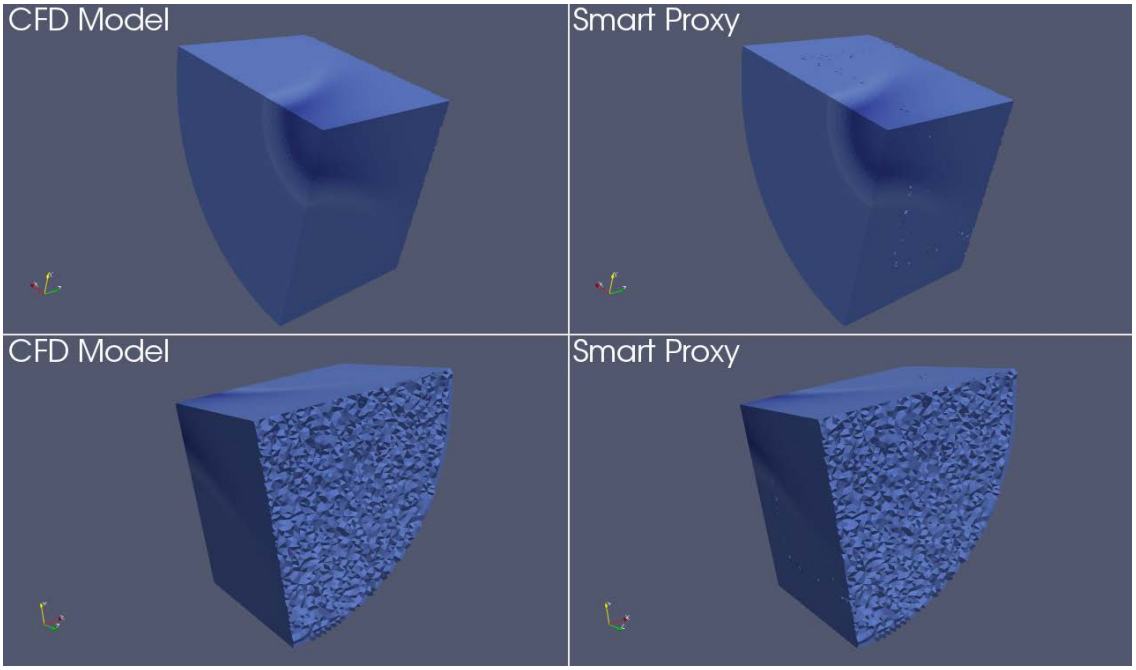


Figure 7-8: Step 3 Results (Q1 Combustor Full View) for Validation Case 5 – Oxygen