

Highlights

Validation of an integrated data-driven surrogate model and a thermo-hydraulic network based model to determine thermal response of a 620 MW_e utility scale boiler using a fully connected mixture density network

B.T. Rawlins,Ryno Laubscher,Pieter Rousseau

- Surrogate model development using CFD solution data.
-

Validation of an integrated data-driven surrogate model and a thermo-hydraulic network based model to determine thermal response of a 620 MW_e utility scale boiler using a fully connected mixture density network

B.T. Rawlins^{a,*}, Ryno Laubscher^b and Pieter Rousseau^a

^aDepartment of Mechanical Engineering, Applied Thermal-Fluid Process Modelling Research Unit, University of Cape Town, Library Road, Rondebosch, Cape Town, 7701, South Africa

^bDepartment of Mechanical Engineering, Stellenbosch University, Banghoek Road, Stellenbosch, 7600, South Africa

ARTICLE INFO

Keywords:

Mixture density network
data-driven surrogate modelling
Computational fluid dynamics

ABSTRACT

An integrated data-driven surrogate model and 1-dimensional (1D) process model of a 620 MW_e utility scale boiler was developed. A robust and computationally inexpensive computational fluid dynamic (CFD) model of the utility boiler was utilized to generate the solution dataset for surrogate model training and testing. In the present work, a standard multi-layer perception (MLP) and MLP connected mixture density network (MLP-MDN) machine learning architectures are compared for use as a surrogate model to predict the boilers' heat exchanger heat loads and the flue-gas inlet conditions to the convective pass. A hyperparameter search was performed to find the best MLP and MLP-MDN architecture, with the MLP-MDN model being capable of predicting the uncertainty. The MLP-MDN showed approximately a 20% decrease in the mean absolute error (MAE) and was selected for surrogate model integration. Validation of the integrated model against plant data was performed for a wide range of loads. Sufficient resolution of the thermal-hydraulic response was observed for all load ranges with a 5-8% error discrepancy. The validated model was subsequently used to investigate the effects of using a poor quality fuel for the 100% maximum continuous rating (MCR) load case. The results show that poor quality fuel leads to a lower operational efficiency for the same load case.

1. Introduction

2. Applicable machine learning theory

The present work makes use of various machine learning architectures to develop a surrogate model that can predict the thermal and combustion characteristics of a utility scale boiler using high level inputs. The current section discusses the details of the respective architectures, namely linear regression, ANN and MLP-MDN models.

2.1. Linear regression

The present work makes use of a multiple linear regression model as a base model for comparative purposes. The main assumption of a multiple linear regression model is that the output/s can be calculated from a linear combination of the variable inputs. In other words a linear regression models aims to determine the quantitative relationship between the dependent and independent variables [16]. The representation of the i -th dependent variable (y_i) can be written as follows for m -number of independent variables (x_{mi}):

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi} \quad (1)$$
$$i = 1, 2, 3, \dots, n$$

where β_0 is a constant term, β_m is the m -th coefficient, and n is the total number of observations.

The optimal solution can be estimated by minimizing the cost function (J). A cost function usually calculates the difference between the estimated and the desired values and is reported as a single number. Multiple linear regression problems namely utilize the mean square error (MSE) between the desired and estimated (\hat{y}_i) values [6], which is given in Equation 2.

$$J_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

To minimize the cost function, the gradient descent algorithm is the preferred method [16], which is an iterative procedure used to find the local minimum/maximum. Considering the cost function as a function of the weight the algorithm can be written as shown in Equation 3.

$$\beta_m = \beta_m - \eta \frac{\partial}{\partial \beta_m} J(\beta_m) \quad (3)$$

where η is known as the learning rate.

In most cases the relationship between the dependent and independent variables are not always linear. Special linear basis models, such as polynomial, sinusoidal and radial can be used to optimize the training results [11]. For the current work a multiple linear progression model is utilized as a base model for comparison.

*Corresponding author

✉ rwlbra001@myuct.ac.za (B.T. Rawlins)

ORCID(s):

2.2. Multilayer perception networks (MLPs)

Artificial neural networks (ANN) are machine learning systems inspired by biological animal brains [11]. There are many classifications of ANNs, with multilayer perception networks (MLP) being the standard representation [4]. Typically MLPs are adapted for supervised learning problems where the input variables are mapped to labelled output variables. The relationship between the input and output variables are learned by optimizing the weights (\bar{w}) and biases (\bar{b}) to minimize a selected cost function, which for most cases is the MSE given in equation 2. A standard MLP schematic is given in figure 1, illustrating the common features of a MLP, that being the input, hidden and output layers.

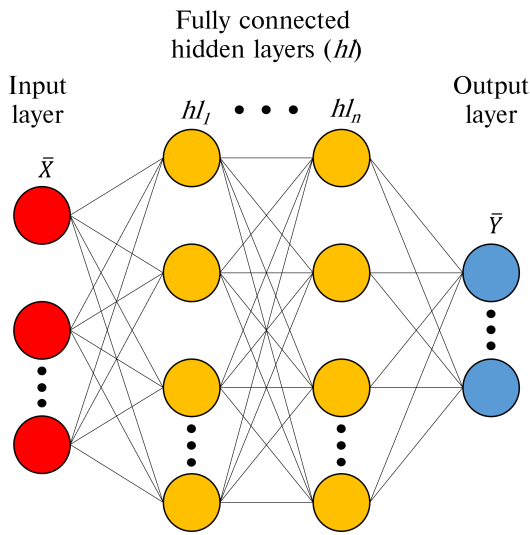


Figure 1: Traditional MLP schematic

To calculate the output values (\hat{y}_i) the forward propagation algorithm is utilized, which calculates the output for each layer and moves sequentially through the network till an output is determined. Each network layer output is calculated using two steps, the first being the calculation of the summed signal \bar{z}_l and secondly the use of an activation function to generate the output signal \bar{h}_l . Equation 4 highlights the first step, where \bar{h}_{l-1} is the output signal from the previous layer.

$$\bar{z}_l = \bar{h}_{l-1} \cdot \bar{w}_l + \bar{b}_l \quad (4)$$

The result of the equation 4 is subsequently passed to an activation function ($\bar{h}_l = \sigma_l(\bar{z}_l)$). There are various activation functions that can be utilized, such as linear, ReLu, Elu, and the hyperbolic-tangent [4], with the final layer activation function usually being linear for regression models. The current work makes use of ReLu and linear activation functions for the hidden layers. The ReLu and

linear activation functions are shown in Equation 5.

$$\bar{h}_l = \sigma_{ReLU}(\bar{z}_l) = \bar{z}_l = \begin{cases} \bar{h}_{l-1} \cdot \bar{w}_l + \bar{b}_l & \text{if } \bar{z}_l > 0 \\ 0 & \text{if } \bar{z}_l < 0 \end{cases} \quad (5)$$

$$\bar{h}_l = \sigma_{linear}(\bar{z}_l) = \bar{z}_l = \bar{h}_{l-1} \cdot \bar{w}_l + \bar{b}_l$$

When the forward propagation step is complete, the network weights and biases can be updated to minimize the cost function (refer to Equation 2) using the backward propagation method [12]. The methodology calculates the gradient of the cost function with respect to the weights and biases for each layer. Once the gradients have been calculated the weights and biases are updated using the gradient descent algorithm. The current work makes use of the Adam [13] alternative to the gradient descent algorithm and is illustrated in Equation 6. Forward- and backward-propagation algorithms would be iteratively conducted until the cost function is reduced to below a desired threshold.

Algorithm 1.

$$\begin{aligned} \bar{m} &\leftarrow \beta_1 \bar{m} + (1 - \beta_1) \nabla_{\theta} J_{MSE}(\bar{\theta}) \\ \bar{s} &\leftarrow \beta_2 \bar{s} + (1 - \beta_2) \nabla_{\theta} J_{MSE}(\bar{\theta}) \otimes \nabla_{\theta} J_{MSE}(\bar{\theta}) \\ \bar{m} &\leftarrow \frac{\bar{m}}{1 - \beta_1^t} \\ \bar{s} &\leftarrow \frac{\bar{s}}{1 - \beta_2^t} \\ \bar{\theta} &\leftarrow \bar{\theta} - \eta \bar{m} \otimes (\sqrt{\bar{s}} + \epsilon)^{-1} \end{aligned} \quad (6)$$

The variable $\bar{\theta}$, of Equation 6, represents the model parameters weights (\bar{w}) and biases (\bar{b}) for each layer. Scaling (\bar{s}) and momentum (\bar{m}) matrices are initialized to zero when beginning the training phase, t is the iterative counter, β_1 and β_2 are the momentum and scaling decay hyper-parameters set to values of 0.9 and 0.999 respectively. Lastly ϵ is a smoothing term set to a value of 10^{-8} .

2.3. Mixture density networks (MDNs)

Mixture density networks are fundamentally built from two components, that being an ANN and a mixture model, this allows for multi-modal predictions. The ANN can be a standard feed-forward MLP or a recurrent neural network (RNN), with RNNs traditionally being used in dynamic applications with at least one feedback loop [10]. For this study the traditionally MLP is used.

MDNs are used to predict the parameters of a probability distribution ($P(\bar{X} | \bar{Y})$) allowing for non-Gaussian distributions to be modelled, thus making MDNs a probabilistic machine learning framework. MDNs estimate the conditional probability distribution as a mixture of Gaussian distributions where the mixing coefficients (π_k) and component densities are flexible functions of the input data (\bar{X}). Equation 7 illustrates the conditional probability function.

$$P(\bar{X} | \bar{Y}) = \sum_{k=1}^K \pi_k(\bar{X}) \cdot \mathcal{N}(\bar{Y} | \bar{\mu}_k(\bar{X}), \bar{\sigma}_k^2(\bar{X})) \quad (7)$$

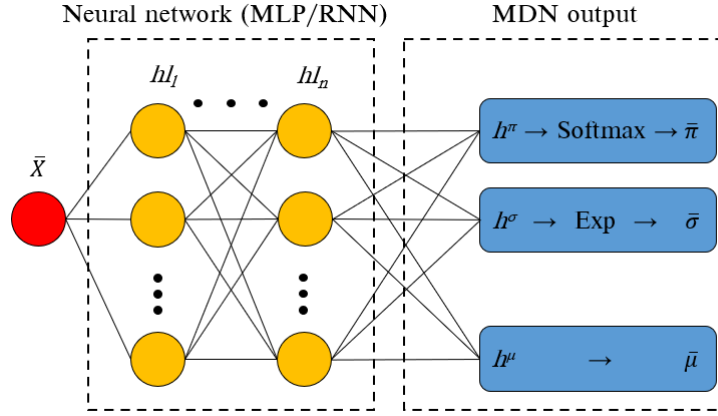


Figure 2: Simple MLP-MDN network

where K represents the number of selected normal distributions, while $\bar{\mu}_k$ and $\bar{\sigma}_k^2$ are the predicted means and variances for each distribution k given the input data \bar{X} respectively.

A schematic of a simple MLP-MDN network is given in Figure 2, highlighting the mixing coefficients, predicted means and deviations. It can be shown that modifications are made to the output layer by splitting the network output into three parts in order to calculate the $\bar{\pi}_k$, $\bar{\mu}_k$ and $\bar{\sigma}_k^2$ for each k distribution. This enables the MDN network to learn the $P(\bar{X} | \bar{Y})$.

Bishop [3] proposed the following restrictions for the mixing coefficients and variance components of the MDN output layer. Since the mixing coefficients contain the discrete probabilities of an output belonging to each K normal distribution for all observations, the mixing coefficients must satisfy the constraints listed in Equation 8.

$$\begin{aligned} \sum_{k=1}^K \bar{\pi}_k^n &= 1 \\ 0 \leq \bar{\pi}_k^n &\leq 1 \end{aligned} \quad (8)$$

These constraints are met by using a Softmax function on the output of the hidden layer h^π . The mixing coefficient is calculated using Equation 9.

$$\bar{\pi}_k^n(\bar{x}^n) = \frac{\exp(h_k^{\pi,n})}{\sum_{k=1}^K \exp(h_k^{\pi,n})} \quad (9)$$

Similarly a constraint is applied to the standard deviation values ensuring a positive value, this is achieved through the use of an exponential function being applied to the standard deviation leg of the MDN output layer, namely h^σ . Equation 10 shows the imposed constraint for the standard deviation leg.

$$\begin{aligned} (\bar{\sigma}_k^n(\bar{x}^n))^2 &\geq 0 \\ \bar{\sigma}_k^n(\bar{x}^n) &= \exp(h_k^{\sigma,n}) \end{aligned} \quad (10)$$

The MDN weights and biases are optimized by minimizing the negative log-likelihood for all observations in a batch, this is shown in Equation 11.

$$\begin{aligned} J_{NLL}(\bar{Y}, \bar{\pi}, \bar{\sigma}, \bar{\mu}) &= \\ - \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \bar{\pi}_k(\bar{X}^n, \bar{\theta}) \cdot \mathbb{N}(\bar{Y}^n | \bar{\mu}_k(\bar{X}^n, \bar{\theta}), \bar{\sigma}_k^2(\bar{X}^n, \bar{\theta})) \right\} \end{aligned} \quad (11)$$

3. Data generation

A steady-state multiphase non-thermal equilibrium CFD model was used to generate the target data, which was subsequently used for training/testing of an appropriate surrogate model.

3.1. CFD model setup

The current study makes use of the commercial CFD software package ANSYS® Fluent 2019 R3 to resolve the fluid flow, heat transfer and combustion processes for a 620MW_e utility scale coal fired boiler. The computational domain is modelled on a symmetry plane half way through the depth of the boiler. Figure 3 highlights the computational domain and defines the important boundary conditions.

The general conservation equations, which include, continuity, momentum, energy and species, were solved using a Eulerian approach. The subsequent equations are stated in Equation (12).

$$\begin{aligned} \frac{\partial}{\partial x_i} (\rho \bar{u}_i) &= S \\ \frac{\partial}{\partial x_i} (\rho_{eff} u_i u_j) + \frac{\partial \bar{p}}{\partial x_j} &= \\ \frac{\partial}{\partial x_i} \left[\mu \left\{ \frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} - \frac{2}{3} \delta_{ij} \frac{\partial u_i}{\partial x_i} \right\} \right] &+ \frac{\partial}{\partial x_i} (-\rho \overline{u_i' u_j'}) + S_m \end{aligned}$$

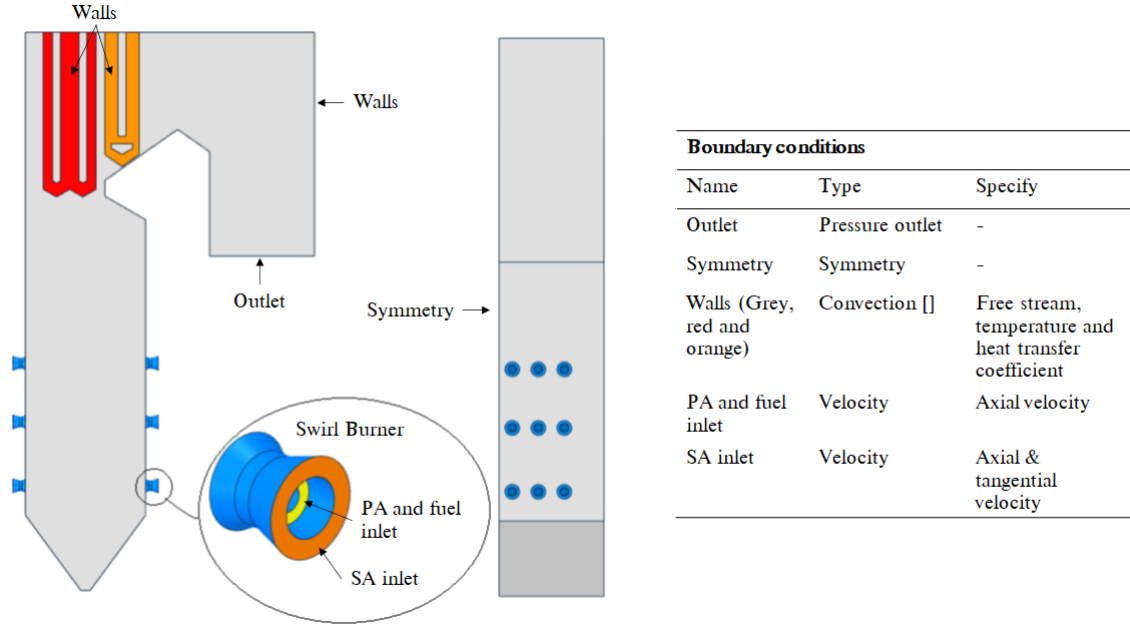


Figure 3: CFD model geometry and boundary condition descriptions.

$$\frac{\partial}{\partial x_i}(u_i[\rho E + p]) = \frac{\partial}{\partial x_j} \left[\lambda \frac{\partial T_g}{\partial x_j} \right] + S_h \quad (12)$$

$$\frac{\partial}{\partial x_i}(\rho u_j Y_k) = -\frac{\partial}{\partial x_j}(\bar{J}_k) + \sum_r R_{j,r} + S_k$$

The resolution of the Reynolds stress term found in the momentum equation, $-\rho \overline{u'_i u'_j}$, was approximated using the Boussineq equation [15]. In the present study the realizable k - ϵ turbulence model was utilized to address the turbulence closure problem, this model was selected for its applicability in modelling the effects of coal-fired swirl burners [8].

The P1 radiation model was used to resolve the radiative field in the domain. Particle transport was modelled using a multiphase approach, further details on the approach are provided in the validation study of Rawlins et al [12]. The combustion follows a four step sequential process, beginning with the heating and evaporation of the moisture present in the fuel, followed by the devolatilization process where the volatiles are liberated from the solid particle, succeeded by the phenomena of char burnout, and finally the gas phase reactions would commence. The char oxidation reaction product species was set to that of carbon monoxide (CO). For the gas-phase reactions the turbulence-chemistry interaction was approximated using the eddy dissipation model (EDM). A summary of the combustion equations and constants are provided in Table 1.

The simulations were solved using the SIMPLE pressure-velocity coupling scheme. The pressure term was discretized using the PRESTO! scheme. Momentum, species

and energy equations were discretized using the second-order upwind scheme and the turbulent kinetic energy and dissipation rate using the first-order upwind scheme. The numerical mesh was generated using quadrilateral elements consisting of 6 million cells. The convergence criteria for the simulation model was set to 1×10^{-3} for the continuity equation, 1×10^{-4} for the velocity equations, 1×10^{-6} for the remaining transport equations, and 1×10^{-4} for monitored key parameters.

3.2. Simulated dataset

As previously mentioned the aim of this study is to illustrate the use of a data-driven surrogate model, integrated with a 1D process model, to predict the heat loads to the various heat exchanging components, the flue-gas composition and exit gas temperatures for a utility scale boiler using various high-level inputs. The inputs include the the following, the excess air ratio per burner, the total mill flowrate for the six mills in operation, the average steam temperatures for the platen and final superheaters, the fouling resistance for the platen and final superheaters, the composition of ash and moisture of the fuel and the gross calorific value of the fuel. Thus the input field has a dimensionality of $d_{inputs} = 14$.

A design of experiments (DOE) was conducted to generate a set of 180 simulation cases to obtain a representative set of results. The various model input ranges used in the DOE are given in Table 2. The ranges were selected to cover a wide range of operational loads with maximum continuous ratings (MCR) between 100% and 30%. A total of 27 output target values ($d_{targets}$) are realised for each CFD simulation case. Due to the quasi-steady nature of the CFD simulations,

Table 1

Summary of combustion models and constants used in the CFD model

Model	Equation/s	Constant/s
<i>Devolatilization</i>		
Single rate kinetic model	$\frac{dm_{vol}}{dt} = R_{vol}(m_{0,vol} - m_{vol}),$ $R_{vol} = A_{vol} \exp\left(\frac{E_{a,vol}}{RT_p}\right)$	$A_{vol} = 2 \times 10^5 [s^{-1}],$ $E_{a,vol} = 6.7 \times 10^7 [J/kmol] - [13]$
<i>Char oxidation</i>		
Diffusion/kinetic - [2]	$\frac{dm_{char}}{dt} = -A_p p_{O_2} \frac{R_{diff} R_c}{R_{diff} + R_c},$ $R_c = A_c \exp\left(\frac{E_{a,c}}{RT_p}\right),$ $R_{diff} = \frac{5 \times 10^{-12}}{d_p} \left(\frac{T_g + T_p}{2}\right)^{0.75}$	$A_c = 0.0053 [kg/m^2 s Pa],$ $E_{a,c} = 8.37 \times 10^7 [J/kmol] - [13]$
<i>Gaseous reactions of volatiles and CO</i>		
Eddy dissipation model - [1]	$R_{k,r,P} = \vartheta_{k,r} M_{w,k} A B \rho_k^\varepsilon \min\left(\frac{\sum_p Y_p}{\sum_j \vartheta_{j,r} M_{w,j}}\right),$ $R_{k,r,R} = \vartheta_{k,r} M_{w,k} A \rho_k^\varepsilon \min\left(\frac{Y_R}{\varepsilon_{R,r} M_{w,R}}\right)$	$A = 4.0, B = 0.5$

Table 2

Design of experiments input ranges for simulations

Input variables	Min	Max	Units
Total fuel flow rate for mills 1 to 6	39.5	120.2	kg/s
Fuel proximate analysis moisture mass fraction, Y_{H_2O}	0.025	0.085	kg/kg
Fuel proximate analysis ash mass fraction, Y_{ash}	0.259	0.559	kg/kg
Platen SH fouling thermal resistance, R_{platen}	0.004	0.007	$m^2 K/W$
Final SH fouling thermal resistance, R_{final}	0.01	0.017	$m^2 K/W$
Dependent variables	Function of		Units
Higher heating value	Fuel constituents Y_{H_2O} and Y_{ash}		J/kg
Excess air	Total fuel flow rate for mills		%
Platen SH internal steam temperature	R_{platen}		K
Final SH internal steam temperature	R_{final}		K

output target values for were taken every 50 iterations for the last 2500 iterations for a converged CFD simulation. This results in each CFD simulation case having a data solution matrix size of ($\tilde{Y} \in 50 \times 27$).

The DOE matrix was populated using the Latin Hypercube Sampling (LHS) method provided in the Python *pyDOE* library. Once the CFD simulations achieved convergence, the target data, comprising of the discretized heat loads to the furnace, platen and final SHs, the exit flue-gas temperatures and combustion characteristics, was stored for each case. The target dataset was split with 80% being used for training purposes and the remainder for testing. To reduce the training time of the neural networks a min-max normalization was utilized scaling all the dataset entries to values between $0 \rightarrow 1$.

4. Model development

The present work makes use of two types of machine learning models, namely a standard artificial neural network (ANN/MLP) and a mixture density designated model

connected to a standard ANN (MLP-MDN). The following section will discuss the hyper parameter tuning and final selected model configuration. The programming language Python 3.7.8 and the Tensorflow machine learning libraries were utilized in the present study.

4.1. Model configuration

The goal of the final machine learning configuration is to be able to predict the heat load distributions to the furnace, platen superheater (SH) and final SH walls, the combustion characteristics and flue-gas temperatures at the exit, using the high level input ranges given in Table 2.

The outputs for the MLP will consist of $y = 50 \times 180 = 9000$ outputs. Thus, for a data batch size, m_b , the output tensor for the MLP model will be $\hat{Y} \in \mathbb{R}^{m_b \times y \times 27}$ since $d_{targets} = 27$. However, the MDN-ANN models output data will consist of three parts, namely; the mixing coefficients tensor of shape $\bar{\pi} \in \mathbb{R}^{m_b \times y \times K}$, the output standard deviation tensor of shape $\bar{\sigma} \in \mathbb{R}^{K \times m_b \times y \times 27}$ and the predicted means of tensor shape $\bar{\mu} \in \mathbb{R}^{K \times m_b \times y \times 27}$. The input data fed into both the MLP and MDN-ANN will have shape $\bar{X} \in \mathbb{R}^{m_b \times d_{inputs}}$. The input features will be varied based on the DOE, to

Table 3

Hyper-parameter search space for fully connected MLP and MLP-MDN models

Parameter	MLP search space	MLP-MDN search space
Number of distributions	-	1,2,3,4
Number of layers	2,3,4	2,3,4
Number of neurons per layer	10, 40, 80, 100	10, 40, 80, 100
Learning rates	1e-3, 1e-4, 1e-5, 1e-6	1e-3, 1e-4, 1e-5, 1e-6
Mini batch sizes	32, 64, 128, 256	32, 64, 128, 256

account for burner mill biasing, fuel quality and SH heating component fouling.

4.2. Hyper-parameter tuning & final model selection

Table 3 highlights the hyper-parameter search spaces for both the MLP and MLP-MDN model. The MLP-MDN has an added parameter namely the number of additional distributions the MLP-MDN would fit to the output data.

The hyper-parameter tuning of both the MLP and MDN-ANN models made use of a single epoch run consisting of 1000 training/testing passes. This was deemed adequate to train/test the models for the various hyper-parameter search spaces, and was utilized to reduce the computational effort/time.

The hyper-parameter search was conducted in a sequential manner with the mean absolute errors (MAE) and root mean square errors (RMSE) being taken as important performance indicators for each sequential step. Firstly both models (MLP & MLP-MDN) hidden layer architectures was varied by considering the amount of hidden layers and neurons per layer. Figure 4 (a) illustrates the MLP model performance for the various hidden layer architectures. It can be seen that the neuron capacity reaches a minimum MAE at 80 neurons per layer for a 4-layer architecture, further neuron capacity results in an increase in the MAE, possibly indicating an over fitting of data. Secondly, the learning rates were varied for the best performing architecture of the first hyper-parameter tuning step. Considering Figure 4 (b), a decrease in the learning rate shows an improvement in the MAE, with a learning rate of 1×10^{-5} being the best for the current epoch size. For comparative purposes, learning rates of 1×10^{-5} and 1×10^{-6} were used in the final step, where the mini-batch sizes are varied. Figure 4 (c) highlights the comparison, for a fixed epoch, with a mini-batch size of 32 and a corresponding learning rate of 1×10^{-5} show the best MAE improvement.

The MLP-MDN hyper-parameter tuning was conducted in a similar manner only that an additional step was required. This was to consider the number of distributions the MLP-MDN would use to capture the probabilistic characteristics. Figure 5 shows the MAE for the various distributions, with a distribution of 1 representing the best MLP model. An

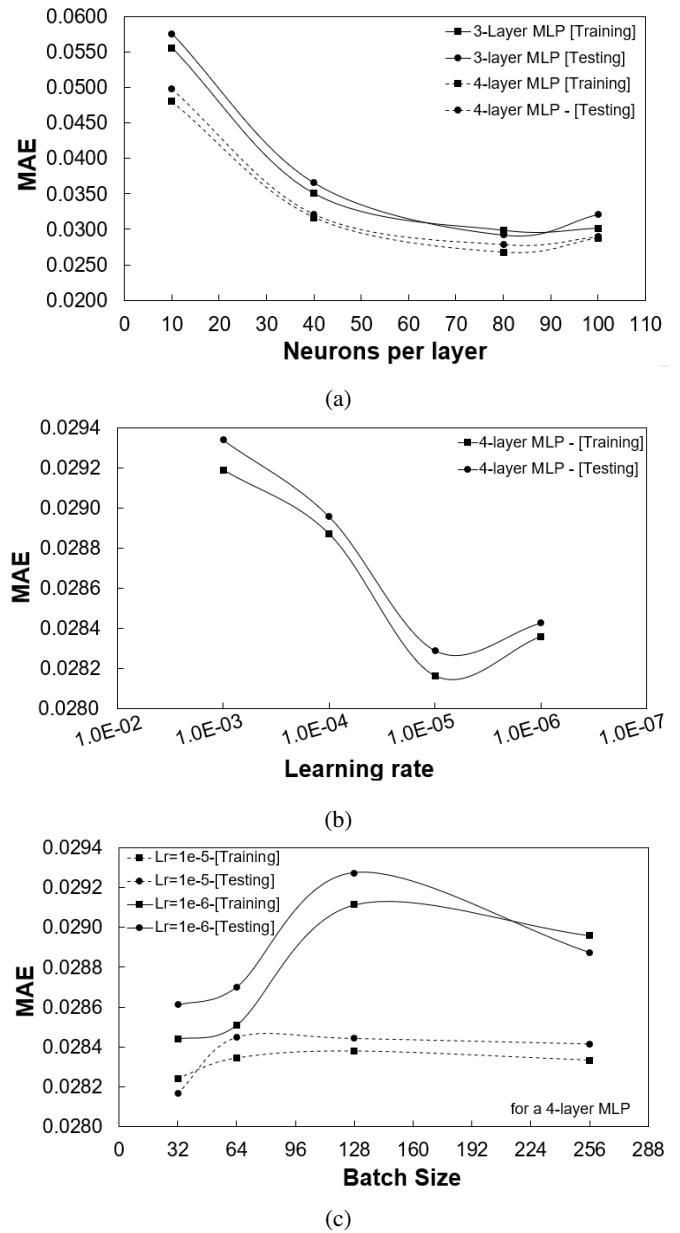


Figure 4: MLP model performance for; (a) hidden layer architecture, (b) varying learning rates, and (c) varying mini-batch sizes

increase in the number of distributions tends to improve the MAE, however it is evident that a threshold of 3 distributions

Table 4
Hyper-parameter search results

Parameter	MLP	MDN-ANN
Number of distributions	-	3
Number of layers	4	4
Number of neurons per layer	80	80
Learning rate	1e-5	1e-5
Mini batch size	32	32
Errors		
RMSE	0.0213	0.0211
MAE	0.0282	0.0263

results in the best improvement.

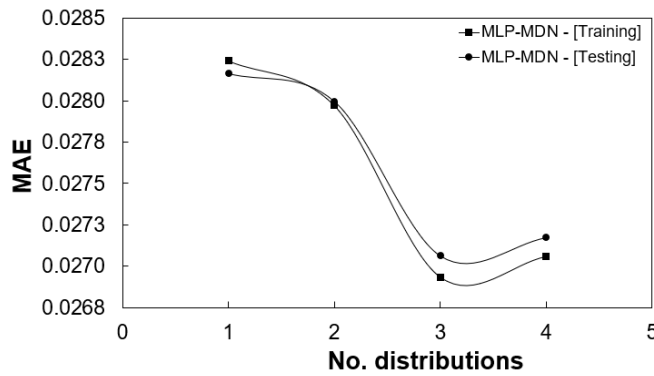
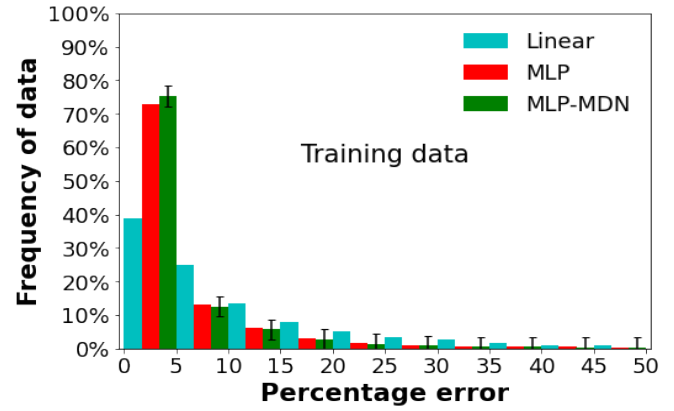


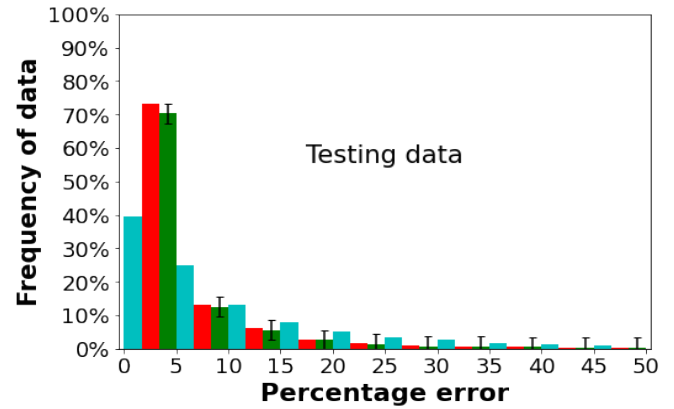
Figure 5: MLP-MDN model performance for number of distributions

The results of hyper-parameter search are given in Table 4. Unlike the MLP, which can only produce a single set of predictions for an input, the MLP-MDN is able to produce a distribution of predictions based on the most probable mixing coefficient. This highlights one of the main advantages of the using the MLP-MDN model is its ability to learn the uncertainty that comes from using a probabilistic model. This is shown in the error distributions graphs of Figures 6 (a) and (b) with addition of uncertainty on the percentage error is introduced. A simple linear regression model was used as a base model for comparative purposes. For the MLP-MDN model it is seen that approximately 80-90% of the training data has mean absolute percentage errors (MAPEs) below 10%, with the MLP model showing a similar trend. Both models show considerable improvement in comparison to the linear model.

Based on the above analysis, the MLP-MDN model was selected as the best model to be used for the surrogate model implementation. As mentioned, the MLP-MDN model is able determine the uncertainty associated with each prediction, an important feature for future investigations incorporating transient/dynamic responses.



(a)



(b)

Figure 6: Error distributions for the selected MLP and MLP-MDN model h; (a) training data and (b) testing data

4.3. Process model integration

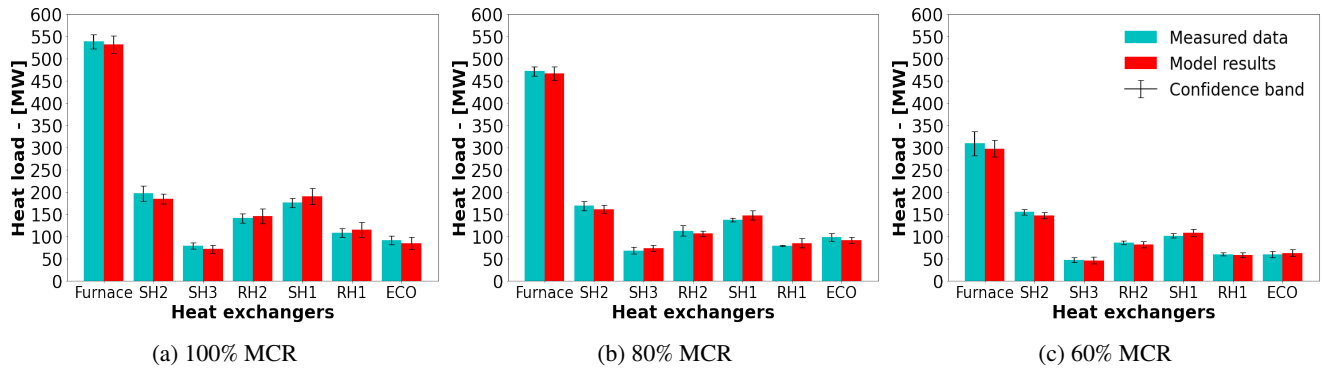
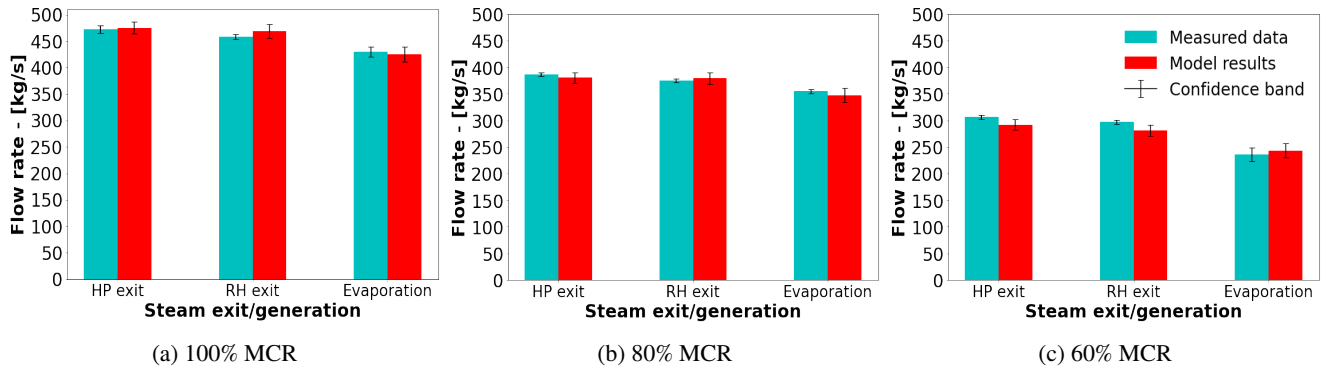
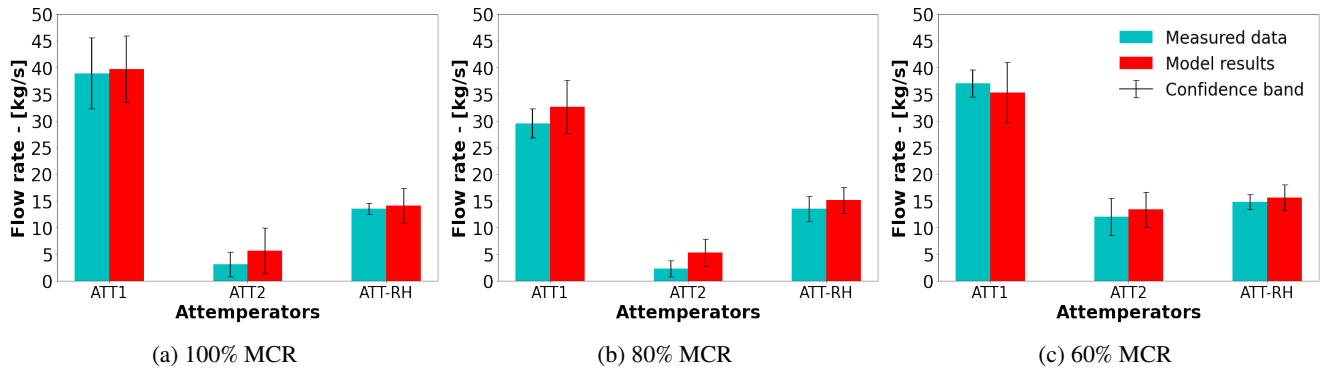
The validation and the subsequent case-studies of Section 5, make use of an integrated surrogate and network based 1D process model. The process model is used to capture the thermodynamic response of the water/steam side of the utility boiler under investigation, with the developed data-driven surrogate model (i.e. MLP-MDN) providing the gas-side thermal predictions and heat-exchanger heat loads.

A C# script is used to access the Python application programming interface (API) available in the process modelling software, Flownex SE® 2021. This allows for predictions to be made using the trained MLP-MDN model. The most probable predictions are retrieved using the script and transferred to the respective process model components as inputs. A schematic of the surrogate and process models integration is provided in Figure 7.

The boiler can be split into three sections namely, the furnace, the radiative pass consisting of the platen (SH2) and final (SH3) SHs, and the convective pass consisting of the secondary re-heater (RH2), primary SH (SH1), primary re-heater (RH1) and the economizer (ECO). The furnace heat

Table 5MLP-MDN input vectors (\bar{X}) for the validation and varied fuel case studies

Inputs	Validation load cases			Varied fuel load cases	
	100%	80%	60%	High-ash	High-moisture
Excess air ratio - [-]	1.155	1.209	1.263	1.155	1.155
Y_{ASH} - [kg/kg]	0.409	0.409	0.409	0.5	0.409
Y_{H_2O} - [kg/kg]	0.055	0.055	0.055	0.055	0.146
HHV = $f(Y_{ASH}, Y_{H_2O})$ - [MJ/kg]	15.07	15.07	15.07	12.68	12.68
Platen SH fouling - [K/W]	0.012	0.012	0.012	0.012	0.012
Final SH fouling - [K/W]	0.0067	0.0067	0.0067	0.0067	0.0067
Total fuel flowrate - [kg/s]	108.6	90.72	60.12	108.6	108.6
Platen SH mean steam temp. - [K]	698	697	781	698	698
Final SH mean steam temp. - [K]	787	793	781	787	787

**Figure 8:** Load validation result comparison of the measured data and MLP-MDN model results for the various heat exchangers**Figure 9:** Load validation result comparison of the measured data and MLP-MDN model results for the steam exit and generation flowrates**Figure 10:** Load validation result comparison of the measured data and MLP-MDN model results for the required attemperators flowrates to maintain operational integrity

cases. It can be seen for a range of MCR loads the integrated model is able capture the heat loads to the various heat exchangers (Figures 8 (a-c)) with sufficient accuracy. In general the model under-predicts the mean of the heat loads in the furnace and radiative pass, with slight over-predictions seen in the convective pass. However, the associated uncertainty of the integrated model overlaps with the measured data uncertainty with minimal outliers.

Considering the steam generation and exit flowrates of Figure 9 (a-c), it can be seen that the integrated model is able to sufficiently capture the hydraulic response of the case study boiler for a wide range of loads. A higher degree of uncertainty is associated with the integrated model but this is deemed acceptable since the response overlaps with the measured data. The attenuators of a utility scale boiler, are an integral part of the control systems put in place to ensure safe operating conditions and minimal wear on the downstream components, namely the steam HP and Low-pressure steam turbines [5]. Figure 10 (a-c) highlight the attenuator flowrates across a wide range of loads. The integrated model is shown to adequately resolve the flowrates with a slightly higher mean being predicted for the 80% load case.

It has been shown that integrated model is able to resolve the thermal response based on the predictions of the MLP-MDN model for a wide range of loads with sufficient accuracy. The heat loads are within a 5% tolerance of the measured data's mean value, while the steam generation rates exhibit a maximum difference of 7%, which seen in the 60% MCR load case (refer to Figure 9 (c)). The attenuators results feature the largest uncertainty/confidence band of all the results (Figure 10), however the confidence bands do overlap within 5%.

5.2. Utility boiler response for a fuel with high ash & moisture contents

The following section makes use of the validated model of Section 5.1. The fuel constituents that effect the energy content of the fuel (HHV) the most are primarily the ash (Y_{ASH}) and moisture (Y_{H_2O}) content. Thus, two case studies were made to illustrate the effects of using poor quality fuel.

The input vectors for the two fuel cases are provided in Table 5, it can be seen that the all the inputs are identical to the 100% load case except for the ash, moisture and higher heating value of the respective fuels. This was done to compare the effects of the poor quality coal to that of the 100% load case, in terms of the boiler efficiency, heat pick-up and thermal response. It is important to note that both the high ash and moisture case studies have the same HHV value, thus allowing for a productive correlation of the results.

As with validation case studies of Section 5.1, the Monte Carlo method was employed to ascertain the integrated models uncertainty for all responses. The base model heading of Figure 11 (a-c) refers to the integrated model results for the 100% load case of Figures 8-10. Figure 11 (a) highlights the effects that poor quality fuel have on the heat load to the various heat exchangers, with a 20% drop being observed. This is an expected result since the energy content of the fuel is lower than that of the base case. To maintain 100% load capabilities using the poor quality fuel an increase in the fuel flowrate would be suggested, this provides more energy content, however the mill operational outputs are limited. The operational protocol for the 100% load case makes use of 5 mills operating 30 of the 36 burners, at full capacity with a standby mill and burner arrangement placed in reserve to help mitigate operational risks, such as maintenance schedules. Using the the standby mill would allow the fuel flowrate to increase by maximum 17% , yet this would decrease the operational integrity and increase the associated risks. Since when burning the high-ash content fuel the likelihood of ash deposition/fouling of the heat exchanging components increases.

Figure 11 (b-c) similarly illustrates a decrease in the water/steam flowrates for the steam exit and attenuator conditions. The generation rates exhibit similar characteristics to the the 80% load case of Figure 9 (b), illustrating a significant decrease in the operational capabilities when using a poor quality fuel. The predicted flue-gas temperatures entering the the convective pass are shown in Table 6. It is shown that for an increase in the fuels ash content the inlet temperature increases, which results in a higher radiative heat transfer percentage for the convective pass components (table 6). While the presence of more moisture in the fuel results in a lower inlet temperature, primarily due to the fact that the extra moisture requires a higher heat rate of evaporation, leading to lower radiative heat transfer percentages. Where the radiative heat transfer percentage per component i is defined as follows:

$$\dot{Q}_{rad,i}\% = \frac{\dot{Q}_{rad,i}}{\dot{Q}_{rad,i} + \dot{Q}_{conv,i}} \times 100 \quad (13)$$

The boiler efficiency is the measure used to convey how well the combustion heat is transferred to the working fluid, and is defined as the ratio of total amount of heat absorbed by the heat exchangers (i.e. sum of the furnace, radiative and convective pass heat loads) and the combustion energy released ($m_{fuel}HHV$). A significant decrease in the boiler efficiency is highlighted in Table 6, for both fuel cases with the high ash fuel showing a slightly better result. The higher efficiency of the high-ash fuel is attributed to the fact that in general the heat loads to various heat exchangers are higher than that of the high-moisture fuel, this can be seen in Figure 11 (a).

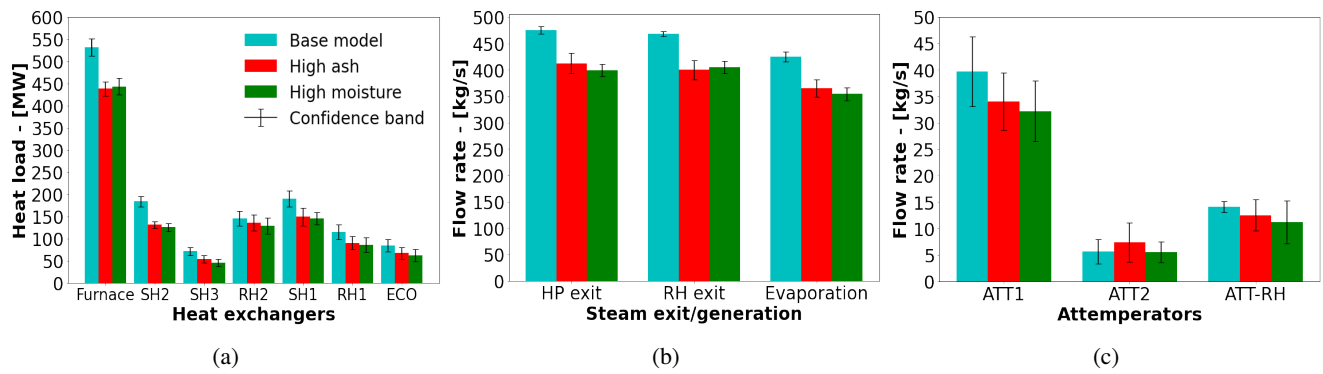


Figure 11: Integrated base, high-ash and high-moisture case study comparison for; (a) the various heat exchangers, (b) the steam exit and generation flowrates, and (c) the required attenuator flowrates to maintain operational integrity

Table 6
Model results for poor fuel characteristics

Variable	Base	High-ash	High-moisture
Mean boiler efficiency [%]	88.2	77.6	75.9
Convective pass inlet flue-gas temp. [K]	1352	1368	1311
<i>Radiative heat transfer percentage (Convective pass)</i>			
RH2	52.3	56.2	50.3
SH1	46.3	49.8	47.2
RH1	21.3	26.7	18.2
ECO	6.7	8.3	5.4

6. Conclusion

The present work establishes the basis for utilizing an integrated data-driven surrogate and 1D process model to investigate the operational response for a 620 MW_e utility scale boiler.

A data-driven surrogate modelling approach for a utility scale boiler using CFD simulations was presented that can predict the various parameters needed to capture the combustion heat transfer characteristics. These include the heat loads to the furnace and radiative pass components, and the flue-gas characteristics entering the convective pass, which include predictions of the temperature, species mass fractions and the radiation intensity. Training and testing datasets were generated using a 3-dimensional (3D) CFD model of the 620 MW_e utility scale boiler under investigation. A total of 180 different CFD simulation cases were performed, where the input variables of Table 2 were varied.

It was found that the best surrogate model architecture was the use of a MLP-MDN configuration, allowing for uncertainty to be predicted. The configuration comprised of a MLP-MDN network consisting of 4 hidden layers each containing 80 neurons and 3 normal output distributions while using a learning rate of 1×10^{-5} and batch size of 32. The MLP-MDN network illustrated that approximately

80-85% of the model predictions had MAPEs below 10%. The added benefit was the prediction of the uncertainty.

A validation case study was performed using the integrated model for a wide range of operational loads. The model results were compared to the measured thermal response for all load cases. The resolution of the heat loads, steam flowrates and attenuation flowrates resulted in a maximum error discrepancy of 7%, highlighting the models sufficient accuracy of the key parameters. Using the Monet Carlo method the uncertainty predictions could be diffused through the integrated model, adding valuable insight into the operational limits of the model.

The subsequent analysis results investigated the using of poor quality fuels at 100% MCR operation. The results highlight the drop in performance due to the fuel quality resulting in a substantial decrease in boiler efficiency. Using a high-ash fuel can result in a higher exit flue-gas temperature resulting in a marginal increase in performance of the convective pass heat exchangers when compared to the high moisture fuel. However the increase in particle loading increases the likelihood of ash deposition occurring.

The current work has shown that it is possible to adequately resolve the thermal-hydraulic response of a utility scale boiler using a data-driven surrogate model based off of CFD simulation result data. The integration of the surrogate and 1D process model is paramount in modelling the thermal response and provides valuable operational insight for various load cases. The use of the model also provides a fast turn around and significant reduction in computational time when compared to co-simulation models using a combination of CFD and process modelling.

CRediT authorship contribution statement

B.T. Rawlins: Methodology, Software, Validation, Formal analysis, Investigation, Writing original draft, Visualization.. **Ryno Laubscher:** Writing review & editing, Methodology, Resources, Conceptualization.. **Pieter Rousseau:** Writing review & editing, Resources, Conceptualization.

References

- [1] ANSYS, 2021. ANSYS Fluent Theory Guide. 20 ed., Ansys Inc.
- [2] Baum, M., Street, P., 1971. Predicting the combustion behaviour of coal particles. *Combustion science and Technology* 3, 231.
- [3] Bishop, C., 1994. Mixture Density Networks. Technical Report. Aston University. doi:10.1007/978-3-322-81570-5_8.
- [4] Goodfellow, I., Bengio, Y., Courville, A., 2017. Deep Learning. First edit ed., MIT Press, Chennai.
- [5] Kakac, S., 1991. Boiler Evaporator and Condenser. Wiley Inter-science.
- [6] Kochenderfer, M., Wheeler, T., 2019. Algorithms for optimization. 1 ed., The MIT Press, Cambridge, Massachusetts.
- [7] Laubscher, R., Rousseau, P., 2019. CFD study of pulverized coal-fired boiler evaporator and radiant superheaters at varying loads. *Applied Thermal Engineering* 160. doi:10.1016/j.applthermaleng.2019.114057.
- [8] Modlinski, N., 2010. Computational modeling of a utility boiler tangentially-fired furnace retrofitted with swirl burners. *Fuel Processing Technology* 91, 1601–1608. doi:10.1016/j.fuproc.2010.06.008.
- [9] MTech, 2021. Flownex Theory Guide. MTech inc.
- [10] Oko, E., Wang, M., Zhang, J., 2015. Neural network approach for predicting drum pressure and level in coal-fired subcritical power plant. *Fuel* 151, 139–145. URL: <http://dx.doi.org/10.1016/j.fuel.2015.01.091>, doi:10.1016/j.fuel.2015.01.091.
- [11] Rasmussen, C., Williams, C., 2006. Gaussian Processes for Machine Learning. 1 ed., The MIT Press, Cambridge, Massachusetts.
- [12] Rawlins, B.T., Laubscher, R., Rousseau, P., 2021. Validation of a thermal non-equilibrium Eulerian-Eulerian multiphase model of a 620 MWe pulverized fuel power boiler., in: Skatulla, S. (Ed.), 12th South African Conference on Computational and Applied Mechanics (SACAM2020), MATEC Web of Conferences, Cape Town. doi:<https://doi.org/10.1051/mateconf/202134700004>.
- [13] Sheng, C., Moghtaderi, B., Gupta, R., Wall, T.F., 2004. A computational fluid dynamics based study of the combustion characteristics of coal blends in pulverised coal-fired furnace. *Fuel* 83, 1543–1552. doi:10.1016/j.fuel.2004.02.011.
- [14] Thomopoulos, N.T., 2013. Essentials of Monte Carlo simulation: Statistical methods for building simulation models. volume 9781461460220. doi:10.1007/978-1-4614-6022-0.
- [15] Versteeg, H., Malalasekera, W., 2007. Introduction to Computational Fluid Dynamics, The finite volume method. Second ed., Pearson Prentice Hall. doi:10.1002/9781119369189.
- [16] Wen, D., Pan, Y., Chen, X., Aziz, M., Zhou, Q., Li, N., 2022. Analysis and prediction of thermal stress distribution on the membrane wall in the arch-fired boiler based on machine learning technology. *Thermal Science and Engineering Progress* 28, 101137. doi:10.1016/j.tsep.2021.101137.