

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

- Computational Fluid Dynamics capture trend of boiler heat uptake at reduced loads.
- Computational Fluid Dynamics capture trend of boiler heat uptake at reduced loads.
- Computational Fluid Dynamics capture trend of boiler heat uptake at reduced loads.

Validation of an integrated data-driven surrogate model and a thermo-hydraulic network based model to determine thermal response of a 620 *MWe* 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 model
Steady State

ABSTRACT

ss fhvosjv
sfjvnsfklvjkfsfn
skfnvkpsfn
sjkvnsf skfnv
ksfvn;sfv,

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 ANN-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 [15]. 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 [5], 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 [15], 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 [10]. For the current work a multiple linear progression model is utilized as a base model for comparison.

2.2. Multilayer perception networks (MLPs)

Artificial neural networks (ANN) are machine learning systems inspired by biological animal brains [10]. 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.

*Corresponding author

 rw1bra001@myuct.ac.za (B.T. Rawlins)

ORCID(s):

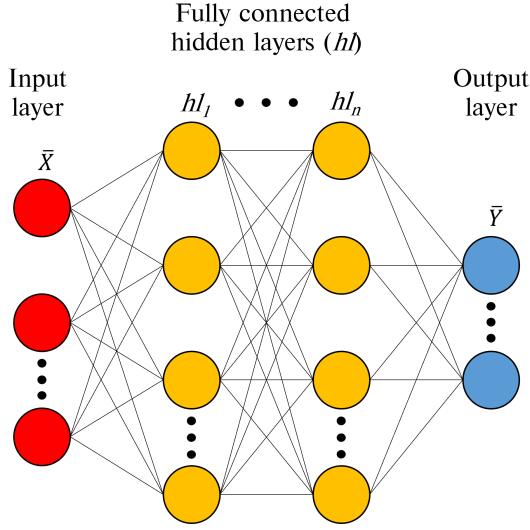


Figure 1: Traditional MLP schematic

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 [4]. 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 [4] 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 [9]. 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 ($\bar{\pi}_k$) and component densities are flexible functions of the input data (\bar{X}). Equation 7 illustrates the conditional probability function.

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

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 ANN-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

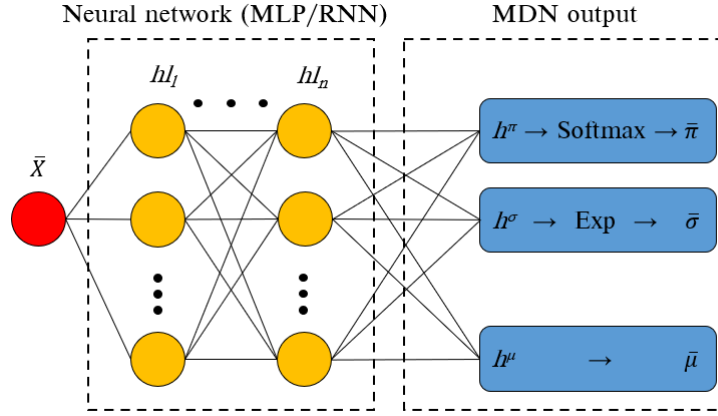


Figure 2: Simple ANN-MDN network

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 . Equations 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).

$$\begin{aligned} \frac{\partial}{\partial x_i}(\rho \bar{u}_i) &= S \\ \frac{\partial}{\partial x_i}(\rho_{eff} u_i u_j) + \frac{\partial \bar{p}}{\partial x_j} &= \\ \frac{\partial}{\partial x_i} \left[\mu \left\{ \frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} - \frac{2}{3} \delta_{ij} \frac{\partial u_i}{\partial x_i} \right\} \right] + \frac{\partial}{\partial x_i}(-\rho \overline{u'_i u'_j}) + S_m \\ \frac{\partial}{\partial x_i}(u_i [\rho E + p]) &= \frac{\partial}{\partial x_j} \left[\lambda \frac{\partial T_g}{\partial x_j} \right] + S_h \\ \frac{\partial}{\partial x_i}(\rho u_j Y_k) &= - \frac{\partial}{\partial x_j}(\bar{J}_k) + \sum_r R_{j,r} + S_k \end{aligned} \quad (12)$$

The resolution of the Reynolds stress term found in the momentum equation, $-\rho \overline{u'_i u'_j}$, was approximated using the Boussineq equation [14]. 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

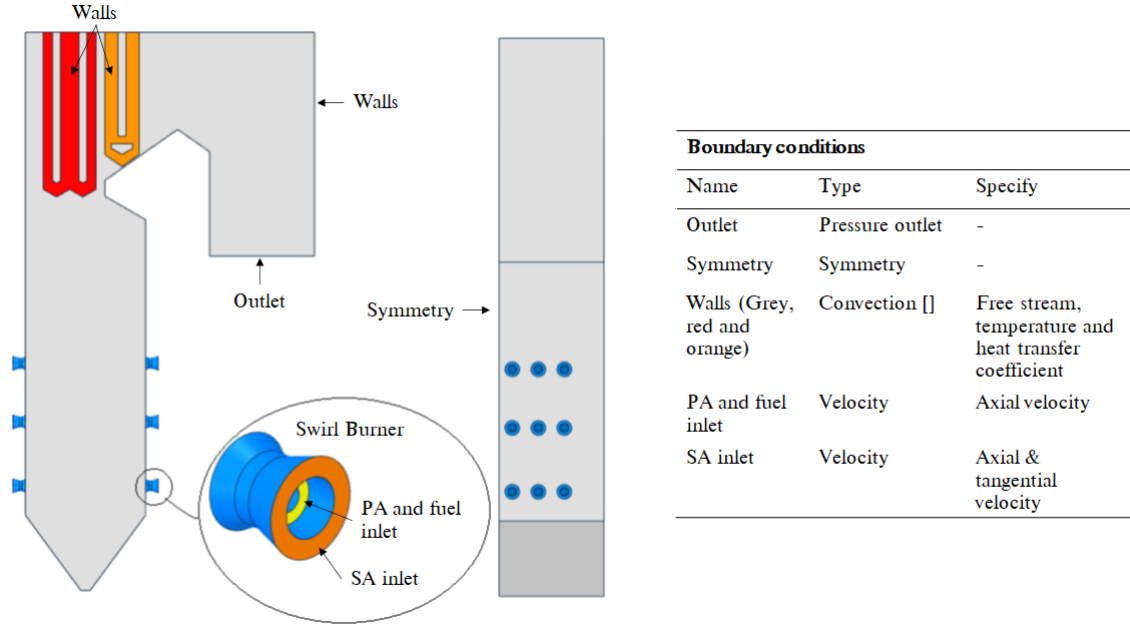


Figure 3: CFD model geometry and boundary condition descriptions.

in modelling the effects of coal-fired swirl burners [7].

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 [11]. 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, 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

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	$\frac{dm_{vol}}{dt} = R_{vol}(m_{0,vol} - m_{vol}),$ $R_{vol} = A_{vol} \exp\left(\frac{E_{a,vol}}{RT_p}\right)$	$A_{vol} = 2 \times 10^5 [s^{-1}],$ $E_{a,vol} = 6.7 \times 10^7 [J/kmol] - [12]$
<i>Char oxidation</i>		
Diffusion/kinetic - [2]	$\frac{dm_{char}}{dt} = -A_p p_{O_2} \frac{R_{diff} R_c}{R_{diff} + R_c},$ $R_c = A_c \exp\left(\frac{E_{a,c}}{RT_p}\right),$ $R_{diff} = \frac{5 \times 10^{-12}}{d_p} \left(\frac{T_g + T_p}{2}\right)^{0.75}$	$A_c = 0.0053 [kg/m^2 s Pa],$ $E_{a,c} = 8.37 \times 10^7 [J/kmol] - [12]$
<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

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 (ANN-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 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 ANN-MDN model. The ANN-MDN has an added parameter namely the number of additional

Table 3

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

Parameter	MLP search space	ANN-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

distributions the ANN-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 & ANN-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 ANN-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 ANN-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 increase in the number of distributions tends to improve the MAE, however it is evident that a threshold of 3 distributions results in the best improvement.

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 ANN-MDN is able to produce a distribution of predictions based on the most probable mixing coefficient. This highlights one of the main

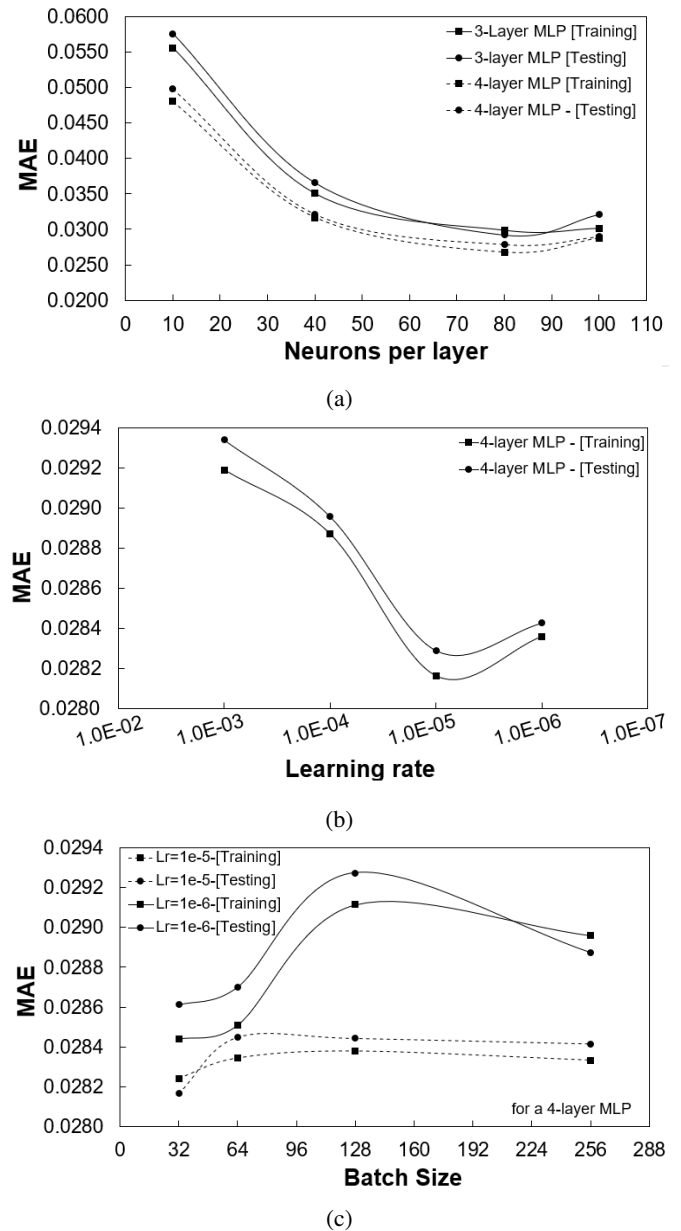


Figure 4: MLP model performance for; (a) hidden layer architecture, (b) varying learning rates, and (c) varying mini-batch sizes

advantages of the using the ANN-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

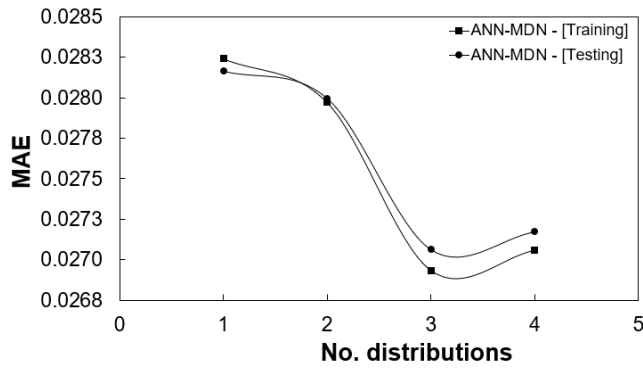


Figure 5: ANN-MDN model performance for number of distributions

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

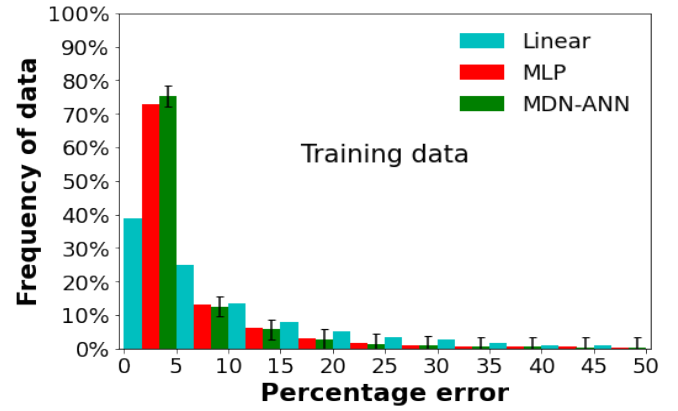
Figures 6 (a) and (b) with addition of uncertainty on the percentage error is introduced. A simple linear regression model was used as a base model for comparative purposes. For the ANN-MDN model it is seen that approximately 80-90% of the training data has mean absolute percentage errors (MAPEs) below 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 ANN-MDN model was selected as the best model to be used for the surrogate model implementation. As mentioned, the ANN-MDN model is able determine the uncertainty associated with each prediction, an important feature for future investigations incorporating transient/dynamic responses.

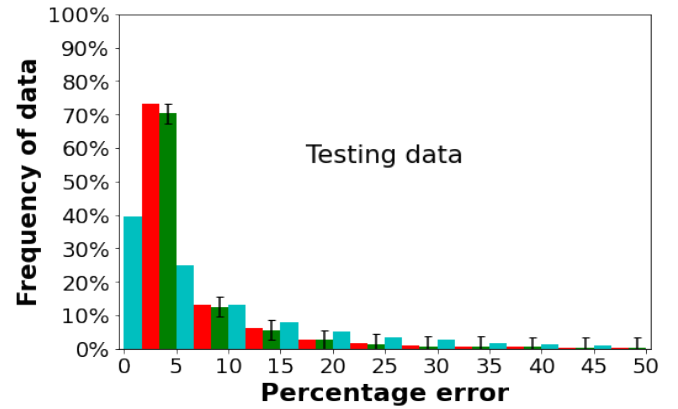
4.3. Process model integration

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. ANN-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



(a)



(b)

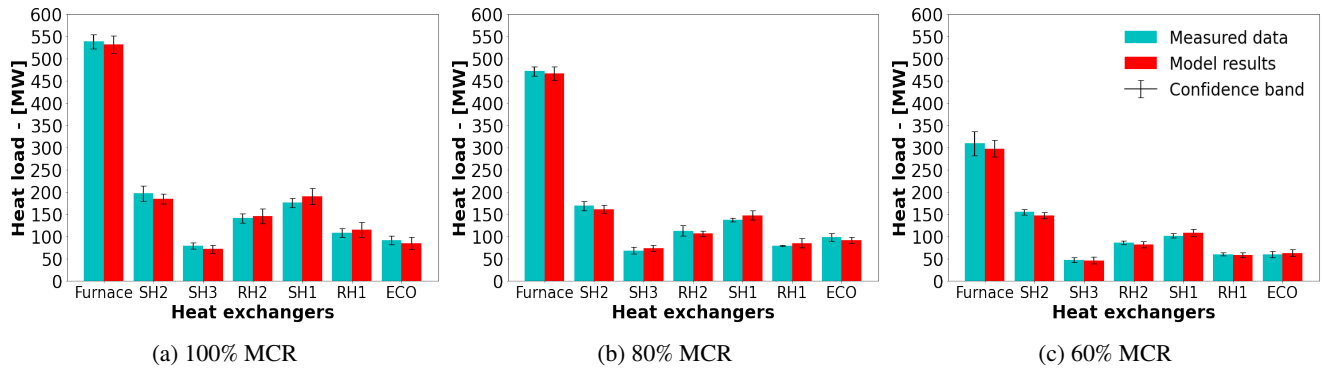
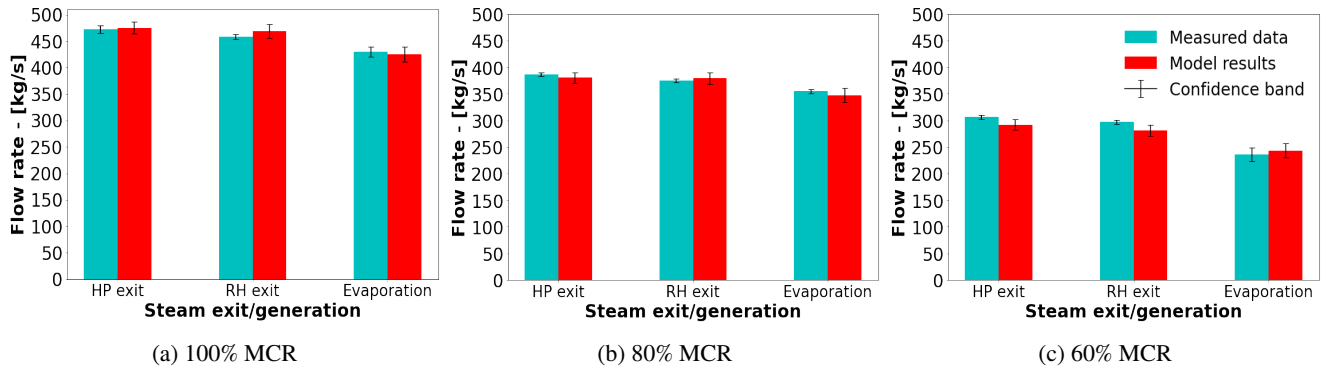
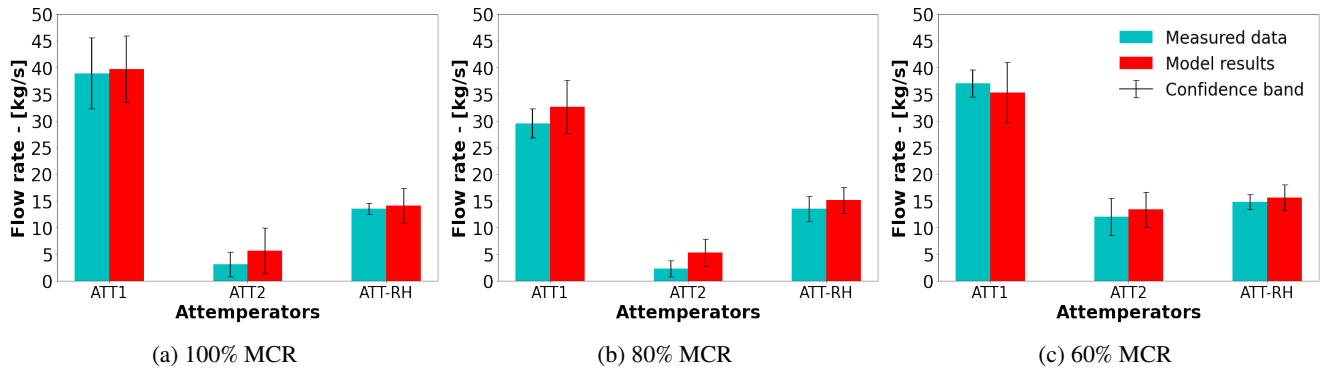
Figure 6: Error distributions for the selected MLP and ANN-MDN model h; (a) training data and (b) testing data

to be made using the trained ANN-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 load predictions to the furnace walls determine the evaporation rate. The high pressure (HP) steam outlet steam flow rate is calculated as the addition of the attenuation flow rates (ATT1 and ATT2) and the evaporation flow rate. The attenuators are used to control the steam temperature to ensure an exit temperature of 808K at both the HP outlet and RH outlet. In addition to the furnace heat loads the surrogate model predicts the platen and final SH heat loads. Lastly the surrogate model is able to predict the flue-gas conditions at the inlet to convective pass (i.e. RH2). The furnace walls, platen and final SHs are represented by a single lumped parameter pipe flow component, interconnected with nodes that introduce the attenuation flows to the water/steam

Table 5ANN-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 ANN-MDN model results for the various heat exchangers**Figure 9:** Load validation result comparison of the measured data and ANN-MDN model results for the steam exit and generation flowrates**Figure 10:** Load validation result comparison of the measured data and ANN-MDN model results for the required attemperators flowrates to maintain operational integrity

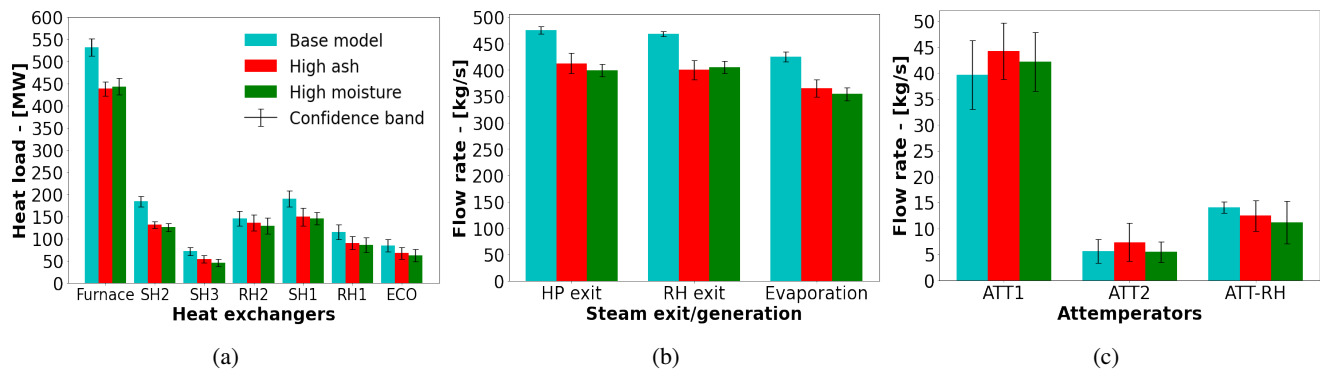


Figure 11: caption here

Table 6
Model results for poor fuel characteristics

Variable	Base	High-ash	High-moisture
Mean boiler efficiency [%]	88.2	76.6	75.9
Inlet flue-gas temperature [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

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

Mention the importance of fuel energy, case The fuel quality is an important factor when determining the Fuels with total moisture contents exceeding The following results illustrate the effects poor quality fuel has on the case studies boiler operational. The study made use of the developed surrogate model to investigate the steady state operation of the boiler burning poor quality coal. In this section the results of a study conducted

6. Conclusion

The present work has shown it is possible

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] Kochenderfer, M., Wheeler, T., 2019. Algorithms for optimization. 1 ed., The MIT Press, Cambridge, Massachusetts.
- [6] 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.
- [7] 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.
- [8] MTech, 2021. Flownex Theory Guide. MTech inc.
- [9] 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.
- [10] Rasmussen, C., Williams, C., 2006. Gaussian Processes for Machine Learning. 1 ed., The MIT Press, Cambridge, Massachusetts.
- [11] 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>.
- [12] 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.
- [13] 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.
- [14] Versteeg, H., Malalasekera, W., 2007. Introduction to Computational Fluid Dynamics, The finite volume method. Second ed., Pearson Prentice Hall. doi:10.1002/9781119369189.

- [15] 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.