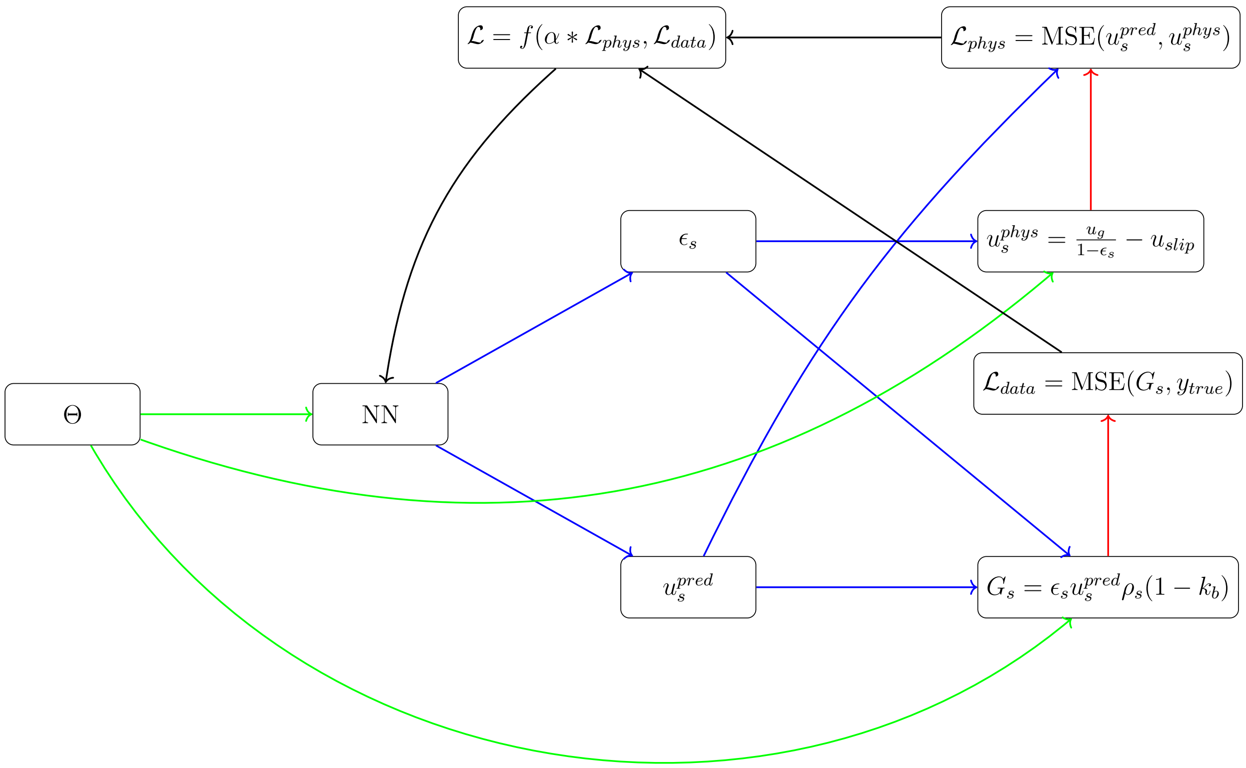
CHALMERS
UNIVERSITY OF TECHNOLOGY



Enhancing Entrainment Modeling of Fluidized Bed Reactors Using Physics Informed Machine Learning

Master's thesis in Innovative and Sustainable Chemical Engineering

Lucas Guné Herman Hellsten

DEPARTMENT OF Chemistry and Chemical Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Enhancing Entrainment Modeling of Fluidized Bed Reactors Using Physics Informed Machine Learning

LUCAS GUNÉ HERMAN HELLSTEN

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Chemistry and Chemical Engineering
Division of Chemical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Enhancing Entrainment Modeling of Fluidized Bed Reactors Using Physics Informed Machine Learning
LUCAS GUNÉ HERMAN HELLSTEN

Supervisor: Ronnie Andersson, Department of Chemistry and Chemical Engineering
Examiner: Jia Wei Chew, Department of Chemistry and Chemical Engineering

Master's Thesis 2026
Department of Chemistry and Chemical Engineering
Division of Chemical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Flowchart for physics-informed neural network structure.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Enhancing Entrainment Modeling of Fluidized Bed Reactors Using Physics Informed Machine Learning
Lucas Guné Herman Hellsten
Department of Chemistry and Chemical Engineering
Chalmers University of Technology

Abstract

Entrainment is a fluidization phenomenon which has been thoroughly studied during the last decades. Despite this, current entrainment models lack a fundamental understanding of the underlying physics and yield poor results. In this project, the recent developments in scientific machine learning, more specifically physics informed machine learning; has been applied to improve modeling performance. The data used in this project was taken from a 12 MW circulating fluidized bed boiler from Chalmers Power Central. Three different physical models were studied and applied to a neural networks training process, resulting in average R^2-scores of 0.376, 0.363 and 0.370. Owen values were calculated for the trained models, which alongside analysis of the data with correlation matrices, principle component analysis and partial least squares regression was used to find important parameters for explaining entrainment. The analysis supports previous hypotheses of relevant parameters like air flows, while also suggesting the importance of new parameters such as temperatures. Results suggest that model performance worsens when both data availability and underlying physics are limited. Improving performance therefore requires compensating for one limitation by strengthening the other. If further studies are to be made within the field, data from multiple boilers with different operating parameters, such as geometry; should be considered.

Keywords: Elutriation, Entrainment, Fluidization, Principle component analysis, Partial least squares regression, Physics informed machine learning, Explainable AI, Circulating fluidized bed, Correction term model.

Acknowledgements

This project has received a lot of help from multiple Chalmers supervisors. First, we would like to thank our examiner Jia Wei Chew and supervisor Ronnie Andersson for suggesting the project and guiding us along the way. We would like to thank Teresa Berdugo Vilches and David Pallarès for providing us data, as well as answering our questions on fluidized beds. Our thanks also goes to Bo Leckner for helping us learn about fluidized bed regimes. Additionally, we would like to thank Tobias Mattisson, Xiaoyun Li and Juan Manuel Rubio Vanegas for helping us expand on the scope and objectives of the project. Outside of Chalmers, we are ever thankful for the help and encouragement of everyone in our families who supported us throughout our studies. Lastly, we would like to thank all of the friends we have made during our five year Chalmers journey. It would not have been the same without you!

Lucas Guné and Herman Hellsten, Gothenburg, May 2026

Disclaimer

This project has used Artificial Intelligence (AI) as part of writing code and reviewing text. The purpose of this was mainly to provide fast debugging, while also ensuring a certain scientific standard while writing.

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ADAM	Adaptive Moment Estimation
ADAMW	Adaptive Moment Estimation With Decoupled Weight Decay
ANN	Artificial Neural Network
BB	Black Box
CFB	Circulating Fluidized Bed
CTM	Correction Term Model
MAE	Mean Absolute Error
MSE	Mean Squared Error
NN	Neural Network
PCA	Principle Component Analysis
PIML	Physics Informed Machine Learning
PINN	Physics Informed Neural Network
PLS	Partial Least Squares Regression
ReLU	Rectified Linear Unit
SHAP	Shapley Additive Explanations
UDE	Universal Differential Equation

Nomenclature

Below is the nomenclature of Fluidization, Data analysis and Machine Learning

Fluidization

a	Decay constant in splash zone (m^{-1})
Air_1	Mass air flow at bottom of boiler (kg/s)
Air_2	Secondary air mass flow (kg/s)
Air_{tot}	Total mass air flow (kg/s)
Ar	Archimedes number
C_D	Drag coefficient of a falling particle
D_e	Effective diameter (m)
d_p	Particle diameter (m)
$d_{p,sauter}$	Sauter mean particle diameter (m)
$Flue_{recirc}$	Flue gas recirculation mass flow (kg/s)
g	Gravitational constant ($\mathrm{m/s^2}$)
g_c	Gravitational unit conversion factor
G_s	External circulation rate ($\mathrm{kg/m^2/s}$)
$G_{s,i}$	Elutration rate for a specific particle size ($\mathrm{kg/m^2/s}$)
G_s^*	Dimensionless entrainment solids flux
G_s^{Corr}	External circulation predicted by correlation ($\mathrm{kg/m^2/s}$)
G_s^{NN}	External circulation predicted by neural network ($\mathrm{kg/m^2/s}$)
G_s^{pred}	Sum of G_s^{Corr} and G_s^{NN} ($\mathrm{kg/m^2/s}$)
h	Evaluated height (m)
h_{mf}	Height of bed at minimum fluidization (m)
H_{Tz}	Height of transport zone (m)
K	Decay constant in transport zone (m^{-1})
k_b	Backflow ratio

L_x, L_y	Side length of boiler cross section (m)
Δp_b	Pressure drop over bed (Pa)
Δp_{fr}	Frictional pressure drop (Pa)
Re_p	Particle reynolds number
$Re_{p,mf}$	Particle reynolds number evaluated at minimum fluidization velocity
T_b	Temperature in top of boiler (^{o}C)
T_t	Temperature in bottom of boiler (^{o}C)
$u_{g,bot}$	Primary gas velocity (m/s)
$u_{g,top}$	Gas velocity at the top of the boiler (m/s)
u_{mf}	Minimum fluidization velocity (m/s)
$u_{mf,i}$	Minimum fluidization velocity for a given particle size (m/s)
u_s	Solids velocity at the top of the boiler (m/s)
u_{slip}	Slip velocity of a particle (m/s)
u_t	Terminal particle velocity (m/s)
u_t, i	Terminal particle velocity for a given particle size (m/s)

ϵ_m	Voidage of packed bed
ϵ_{mf}	Voidage in a fluidized bed
ϵ_{se}	Voidage at boiler exit
μ	Viscocity of gas (Pa*s)
ρ_g	Gas density (kg/m^3)
ρ_s	Solids density (kg/m^3)
$\rho_{s,entr}$	Density of gas and particle mixture entrained from packed bed (kg/m^3)
ρ_{s,H_b}	Density of gas and particle mixture at boiler exit height (kg/m^3)
ϕ_s	Sphericity of a particle

Data analysis

C	Y-weight matrix for PLS
d_i	Difference in rank between two observations
M_k	Principle component k
P	Loading matrix for PCA

$p_{k,i}$	Loading for principle component k of variable i
T	Score matrix for PCA/X-score matrix for PLS
U	Y-score matrix for PLS
W	X-weight matrix for PLS
X	Feature variable matrix
$\bar{x}$	Average value of variables
x_i	Individual occurrence of a feature variable
$\bar{y}$	Average value of targets
y_i	Individual occurrence of a target variable
z	Normalized variable value
μ_x	Average of a variable

Physics Informed Machine learning

a_i	Neuron output
b	Bias vector
b_i	Bias in a neural network
g	Correction function
$\mathcal{L}$	The loss function of a neural network
n_{in}	Number of inputs going into a neuron
p_i	Neuron input
R^2	Coefficient of determination
R^2_{test}	Coefficient of determination for the test set
R^2_{val}	Coefficient of determination for the validation set
v_i	Potential from a neuron
w	Weight vector
w_i	Weight in a neural network
y_{pred}	The predicted output of the neural network
y_{true}	The target value
α	Physical loss factor
γ	Weight decay
η	Learning rate of a neural network

θ	Parameters of a given neural network
Θ	Input variables
λ	Physical loss normalization factor

Shapley and Owen values

j	Feature variable
p	Total number of model features
S	Subset of model features
$val(S)$	Prediction of model with information about variables in subset S
ϕ_j	Shapley value of feature j

Contents

List of Acronyms x

Nomenclature xiii

List of Figures xxi

List of Tables xxiii

1 Introduction 1
- 1.1 Background 1
  - 1.1.1 Power central data 2
  - 1.1.2 Physics informed machine learning 2
- 1.2 Objectives 3
  - 1.2.1 Primary objective 3
  - 1.2.2 Secondary objective 3
- 1.3 Scope 3

2 Theory 5
- 2.1 Fluidization 5
  - 2.1.1 Void fractions 5
  - 2.1.2 Frictional pressure drop 5
  - 2.1.3 Minimum fluidization velocity 5
  - 2.1.4 Terminal velocities of particles 6
- 2.2 Density and viscosity 7
  - 2.2.1 Sutherland's law 7
  - 2.2.2 Ideal gas law 7
- 2.3 Physical models 7
  - 2.3.1 Concentration decay in a CFB 8
  - 2.3.2 Merrick and Highley empirical correlation 10
  - 2.3.3 Particle slip velocity 11
- 2.4 Data pretreatment and analysis 11
  - 2.4.1 Correlation Matrix 11
  - 2.4.2 Z-score normalization 12
  - 2.4.3 Principle component analysis 12
  - 2.4.4 Partial least squares regression 13
- 2.5 Artificial neural networks 14
  - 2.5.1 Neurons 14

2.5.2 Feedforward network 14
2.5.3 Training of a network 15
2.5.4 Hyperparameters 15
2.6 Optimization and generalization of a neural network 16
2.6.1 Stochastic gradient-based optimization 16
2.6.2 Weight decay 16
2.6.3 Dropout regularization 16
2.6.4 Kaiming weight initialization 17
2.7 Grey-box modeling 17
2.7.1 Correction term model 17
2.7.2 Physics informed loss function 18
2.8 Shapley and Owen values 18

3 Method **21**
3.1 Limitations 21
3.2 Data 21
3.2.1 Pretreatment and data curation 23
3.2.2 Data analysis 23
3.3 Physical model testing 24
3.4 Neural network 24
3.5 Implementation of physical models 24
3.5.1 Correction term model 24
3.5.2 Decay model 25
3.6 PINN model 25
3.7 Experimental setup 26
3.7.1 Train-validation-test split 26
3.7.2 Grid search 26
3.7.3 Cross validation 27
3.8 Owen values 27

4 Results **29**
4.1 Data analysis 29
4.1.1 Correlation matrix 29
4.1.2 PCA 30
4.1.3 PLS 30
4.2 Merrick and Highley empirical correlation 31
4.3 Decay model 32
4.4 Model performance 33
4.5 Cross validation of performance 36
4.6 Owen values 38

5 Discussion **45**
5.1 Data analysis 45
5.1.1 Correlation matrices 45
5.1.2 PCA 45
5.1.3 PLS 46
5.2 Merrick and Highley correlation 46

5.3 Decay model 47
5.4 Model performance 48
5.4.1 PINN performance 48
5.4.2 CTM and BB performance 48
5.5 Owen values 49
5.5.1 Temperature 49
5.5.2 Air flows 50
5.6 Sustainability 51

6 Conclusion **53**

Bibliography **55**

A Appendix 1 **I**

List of Figures

2.1 Concentration decay of particles in the CFB boiler. 10
2.2 Neural network structure. 15

3.1 Correction term model. 25
3.2 PINN model. 26

4.1 Correlation matrices for the data. 29
4.2 Score and loading plots for PCA analysis. 30
4.3 Score and weight plots for PLS analysis. 31
4.4 Result plot for Merrick and Highley correlation. 32
4.5 Result plot for decay model. 33
4.6 Best performing PINN network. 34
4.7 Best performing CTM network. 35
4.8 Best performing BB network. 36
4.9 Violin plot of R2 performance values for each network. 37
4.10 Violin plot of MAE of each network. 38
4.11 Beeswarm and bar plots with Owen values for BB models. 40
4.12 Beeswarm and bar plots with Owen values for PINN models. 41
4.13 Beeswarm and bar plots with Owen values for CTM models. 42

List of Tables

4.1 Neural network hyperparameters and performance. 34
4.2 R^2_{test} for seeds evaluated with Owen values. 39

1
Introduction

This chapter presents the history of fluidized beds, Chalmers power central and scientific machine learning. It also presents the scope and objective of this project.

1.1 Background

Fluidized beds have long been used in the chemical industry, and their technology has been steadily improving. Early use of the technology included fluid catalytic cracking, which is still used today. Fluidization is a phenomenon where particles in a bed behave like a fluid when suspended by an upward flowing fluid [1]. The fluid will also suspend some of the particles, which will be carried away from the bed up into the riser. The finer particles will in a circulating fluidized bed (CFB) eventually leave the riser in a process called external circulation.

The interest in external circulation and fluidized bed behavior is driven by its impact on heat transfer and temperature distributions within reactors, as well as its central role in emerging fluidized technologies for fuel conversion, such as chemical looping combustion and hydrogen production. These processes rely heavily on accurate descriptions of particle dynamics to ensure efficiency, scalability, and operational stability [1].

One of the first empirical correlations for such processes was published in 1955 [2]. More recently, CFBs have gained popularity due to the high fuel flexibility they provide [3]. Nevertheless, the numerical and analytical descriptions of the underlying phenomena for CFBs remain largely empirical, lacking a fundamental theoretical understanding. As a result, existing external circulation correlations can sometimes deviate from experimental data by up to two orders of magnitude [2].

Most empirical correlations have been derived from small operational units [2]. The differing geometries lead to difficulty generalizing over different scales, especially larger ones. In addition to this, data measured from larger operating parameters is scarce, often lacking in terms of resolution and range [4].

1.1.1 Power central data

The Chalmers Power Central project began in 1947, with the purpose of researching boilers on a larger scale. Over the years, the technology has gradually been upgraded in order to meet the demand of a carbon dioxide-free power plant [5]. This extensive research project has been documenting process data on a large operational scale in order to enhance the knowledge of the fluidization phenomenon.

1.1.2 Physics informed machine learning

Recent developments within some areas of science have been heavily reliant on deep learning, using large amounts of data in order to train data driven models. Although this has shown great promise, other areas of science that do not have access to large amounts of data are still bound to mechanistic models, as deep learning techniques do not reach a satisfactory level of precision with smaller data sets [6]. A developing field called scientific machine learning, aims to integrate traditional physical engineering into the learning process of a neural network to improve performance. Two of these methods are known as the universal differential equation (UDE) or physics informed neural network (PINN). Compared to models without mechanistic knowledge, a PINN can perform better even when the dataset is reduced by 80 percent, as shown by Garakani and Chew [7].

In recent years, the application of physics-informed and physics-constrained machine learning in chemical engineering has expanded rapidly. Mukherjee and Zavala [8] describe how physics-constrained machine learning has shown great promise within the field of closed-loop experimental design, real-time dynamics and control and handling of multiscale phenomena; while also highlighting the technical and intellectual challenges that come with it. It has also been applied within fludization engineering by Feng, Deng, Chaffart, Yuan and Zhu in order to investigate the hydrodynamics of the system [9].

Hybrid approaches have also demonstrated more accurate predictive capability in process engineering problems. For example, Mondal and Sharma [10] developed a PINN framework for gas-liquid annular flow prediction, where empirical correlations are combined with machine-learning models to improve model generalization. Similarly, Mohammadghasemi, Soesanto, Singh, Maciszewski and Huang [11] applied scientific machine learning to primary separation vessels by integrating physical constraints within a neural network. This led to models which could extrapolate predictions to industrial-scale systems.

Furthermore, recent advancements within combustion modeling have been reviewed by Deng, Wang, Kim and Koenig [12] and found to be successful for mechanism discovery and more. Lastly, scientific machine-learning has also been found to be applicable within fields outside of chemical engineering, such as enzyme engineering and anomaly detection [13],[14].

These developments illustrate how the integration of physical knowledge and machine learning is becoming increasingly important for modeling, optimization, and discovery in chemical engineering.

1.2 Objectives

Existing elutriation correlation precision vary drastically depending on the operating parameters. Currently, most of the research done for the elutriation phenomenon remain empirical. The lack of a generalizing correlation highlights the need for approaches that can capture the underlying physical mechanisms while keeping a predictive accuracy across a wide range of operating conditions. Physics-informed machine learning is particularly suitable for this purpose because it combines data-driven modeling with physical principles, without relying on having large datasets or highly precise physical models. This motivates the following objectives.

1.2.1 Primary objective

Develop a physics informed neural network with a wide range of inputs which can predict elutriation with greater accuracy, in terms of R^2-score and MAE, than existing correlations.

1.2.2 Secondary objective

Extract meaningful mechanistic understanding between input variables and entrainment using explainable AI methods.

1.3 Scope

The project is limited to data from Chalmers power central, and the error from measurement instruments is considered negligible. The scalability of the model will not be evaluated as there only exists data from the Chalmers power central boiler.

2

Theory

2.1 Fluidization

Fluidization is a phenomenon where a bed of particles will start acting as a fluid once a sufficient force of flowing medium is applied from below [1]. For this project, the medium used is always a gas. The phenomenon begins at a velocity called the minimum fluidization velocity, and as the velocity increases, the emulsion of gas and solids will pass through different regimes. The fluid will also carry some suspended particles with it [1]. This solid flow is called entrainment. The entrained flow rises up through the freeboard, but a part of it back-mixes to the bed [3]. Particles of different sizes do however have different terminal velocities, leading them to have different flow rates causing separation of the different particle sizes. This is called elutriation. The flow rate of one given particle size is called the elutriation rate [2]. The mass fraction weighted sum of all elutriation rates is therefore the same as the entrainment rate. The particles which reach the top of the freeboard can leave the system, and in a circulating fluidized bed (CFB) they are recirculated through a cyclone back to the bottom bed [15]. This rate is called the external circulation.

2.1.1 Void fractions

In a bed of particles, one can not assume a continuous phase. Rather, it is a heterogenous emulsion of gas and solids with voidage spread out across the bed. This can be described by the void fraction ϵ, which is essential in fluidization engineering [1].

2.1.2 Frictional pressure drop

As stated by Kunii and Levenspiel, the frictional pressure drop Δp_{fr} has been found to be correlated to the voidage of a packed bed by Equation 2.1 [1].

$$\frac{\Delta p_{fr}}{h} g_c = 150 \frac{(1-\epsilon_m)^2}{\epsilon_m^3} \frac{\mu u_{g,bot}}{(\phi_s d_p)^2} + 1.75 \frac{1-\epsilon_m}{\epsilon_m^3} \frac{\rho_g u_{g,bot}^2}{\phi_s d_p} \tag{2.1}$$

2.1.3 Minimum fluidization velocity

Fluidization occurs when particles resting on a bed interact with an upward flow of gas from below. Each particle is dragged down by gravity and lifted up by drag from the flowing gas. The minimum fluidization velocity, u_{mf}, is found when the following two forces are equal

$$\begin{pmatrix}\text{drag force by}\\ \text{upward moving gas}\end{pmatrix} = \begin{pmatrix}\text{weight of}\\ \text{particles}\end{pmatrix} \tag{2.2}$$

or

$$\begin{pmatrix}\text{pressure drop}\\ \text{across bed}\end{pmatrix}\begin{pmatrix}\text{cross-sectional}\\ \text{area of tube}\end{pmatrix} = \begin{pmatrix}\text{volume}\\ \text{of bed}\end{pmatrix}\begin{pmatrix}\text{fraction}\\ \text{consisting}\\ \text{of solids}\end{pmatrix}\begin{pmatrix}\text{specific}\\ \text{weight}\\ \text{of solids}\end{pmatrix} \tag{2.3}$$

[16] or, by rearranging Equation 2.3

$$\frac{\Delta p_b}{h_{mf}} = (1-\epsilon_{mf})\left[(\rho_s-\rho_g)\frac{g}{g_c}\right] \tag{2.4}$$

Combining Equation 2.1 and Equation 2.4, we get

$$\frac{1.75}{\epsilon_m^3\phi_s}\left(\frac{d_p u_{mf}\rho_g}{\mu}\right)^2 + \frac{150(1-\epsilon_{mf})}{\epsilon_m^3\phi_s^2}\left(\frac{d_p u_{mf}\rho_g}{\mu}\right) = \frac{d_p^3\rho_g(\rho_s-\rho_g)g}{\mu^2} \tag{2.5}$$

or the dimensionless alternative of Equation 2.5

$$\frac{1.75}{\epsilon_m^3\phi_s}Re_{p,mf}^2 + \frac{150(1-\epsilon_{mf})}{\epsilon_m^3\phi_s^2}Re_p mf = Ar \tag{2.6}$$

where the variables Re_{mf} and Ar can be described by Equation 2.7 and Equation 2.8.

$$Re_{p,mf} = \frac{d_p u_{mf}\rho_g}{\mu} \tag{2.7}$$

$$Ar = \frac{d_p^3\rho_g(\rho_s-\rho_g)g}{\mu^2} \tag{2.8}$$

These equations can then be used to solve for u_{mf}. The voidage ϵ_{mf} can be extracted from table 3 in [1].

2.1.4 Terminal velocities of particles

When particles are fluidized, it means that the drag force from the primary gas is equal to or greater than the gravitational pull on the particle. The terminal velocity of a particle is the maximum velocity it can reach while falling freely. This velocity can be derived from the following correlation

$$u_t = [\frac{4d_p(\rho_s-\rho_g)g}{3\rho_g C_D}]^{1/2} \tag{2.9}$$

where, the drag coefficient C_D can be calculated using

$$C_D = \begin{cases} \frac{24}{Re_p} & \text{if } 0 < Re_p < 0.4 \textit{ (Stoke's law)} \\ \frac{10}{Re_p^{0.5}} & \text{if } 0.4 < Re_p < 500 \textit{ (Intermediate law)} \\ 0.43 & \text{if } 500 < Re_p \textit{ (Newtons law)} \end{cases}$$

and the particle Reynolds number is evaluated at the free stream velocity with which the fluid approaches the particle. This velocity can either be $u_{g,bot}$ or $u_{g,top}$ depending on the phenomenon of interest [16].

$$Re_p = \frac{u d_p \rho_g}{\mu} \tag{2.10}$$

2.2 Density and viscosity

Parameters such as density and viscosity of a gas vary depending on the surrounding environment, enough so that they cannot be assumed to be constant. With this in mind, sufficient formulas need to be applied in order to extrapolate true density and viscosity as a function of temperature.

2.2.1 Sutherland's law

At lower pressures, the viscosity is found to increase with temperature. In order to determine the viscosity change with a two percent error, the following equation can be used

$$\frac{\mu}{\mu_0} = \left(\frac{T}{T_0}\right)^{3/2} \frac{T_0 + S}{T + S} \tag{2.11}$$

where $S = 111.4K$ for a temperature range of $170K - 1900K$, $\mu_0 = 1.716 \times 10^{-5}$ and $T_0 = 273.15K$ [17].

2.2.2 Ideal gas law

In order to calculate the density, dependent on temperature and pressure, the ideal gas law can be utilized;

$$\rho_g = \frac{p}{R_{air} T} \tag{2.12}$$

where

$$R_{air} = \frac{R}{M} = \frac{8.314 J K^{-1} mol^{-1}}{0.02897 kg mol^{-1}} = 287 \frac{J}{K kg} \tag{2.13}$$

2.3 Physical models

There are three physical models being considered for this project for modeling external circulation. The first one considers exponential decay of solids along the height of the boiler, the second one has been empirically derived by Merrick and Highley [18] and the third is based on slip velocity and voidage at the top of the boiler.

2.3.1 Concentration decay in a CFB

In 1995, Johnsson and Leckner introduced a model for the vertical distribution of solids in CFBs, consisting of three distinct zones [19]. This model was then further developed by Djerf, Pallares and Johnsson into the version presented below [3]. The three distinct zones are the dense bed, the splash zone, and the transport zone. In the bottom of the furnace is the dense bed, which has an even distribution of fluidized solids across the height. The dense bed does however have bubbles which explode, leading to clusters of particles being ejected into the freeboard. This is the dominant mechanism for solid transport in the area just above the dense bed, also called the splash zone. The clustered solid concentration strongly decreases exponentially, with a decay constant a which is found through a correlation by Johnsson and Leckner [19]. The expression is shown in Equation 2.14

$$a = 4 * \frac{u_t}{u_{g,bot}} \tag{2.14}$$

The other mechanism for solid transport is entrainment, which happens due to a strong core flow in the boiler, dragging dilute particles upward. Some particles will however escape from the core flow toward the edges, where they will back-flow to the dense bed. This also leads to an exponential decrease in dilute solid concentration over the height, which is described by the decay constant K. This constant is found through a simple equation suggested by Johansson, Johnsson and Leckner, shown as Equation 2.15 [15].

$$K = \frac{1}{H_{tz}} \tag{2.15}$$

This mechanism is dominant in the uppermost region, called the transport zone. The concentration of solids just above the dense bed $\rho_{s,entr}$ is found through a correlation by Djerf, Pallarès and Johansson [20] and shown as Equation 2.16.

$$G_s^* = 0.0054 * (\frac{u_{g,bot}}{u_t} - 1)^2 \tag{2.16}$$

G_s^* is a dimensionless entrained solids flux, which can be used to get $\rho_{s,entr}$ through Equation 2.17.

$$\rho_{s,entr} = \frac{G_s^* \rho_s u_{g,bot}}{u_s} \tag{2.17}$$

Using the two obtained decay constants, the concentration profile along the height can be described by Equation 2.18.

$$\rho(h) = (\rho_{s,H_b} - \rho_{s,entr}) * e^{-a(h-H_b)} + \rho_{s,entr} * e^{-K(h-H_b)} \tag{2.18}$$

The first term of the equation describes the clustered, splashing solids, and the second describes the entrained solids. Given the large decay constant of the clusters, the first term in Equation 2.18 is negligible for large h. This means the concentration in the top can be given by inserting the solid exit height into only the second term, removing the need to calculate ρ_{s,H_b}. This concentration can then be used to describe the external circulation through Equation 2.19

$$G_s = \rho(H_{exit}) * (u_{g,top} - u_t) * (1 - k_b) \tag{2.19}$$

where $u_{g,top}$ is calculated by Equation 2.20, where the constants 1.276 and 1.293 are used with the temperature to describe the air and flue gas density respectively at the given temperature.

$$u_{g,top} = \frac{1}{A} Air_{tot} \frac{T_t}{1.276 * 273} + \frac{1}{A} Flue_{recirc} \frac{T_t}{1.293 * 273} \tag{2.20}$$

In Equation 2.19, $u_{g,top} - u_t$ describes the solid velocity, and $1 - k_b$ describes the proportion of particles reaching the exit height which actually leaves the system. The back-flow ratio k_b is dependent on exit geometry, local particle size and concentration and is calculated by Equation 2.21 which is a fit to experimental data done by Johansson, Johnsson and Leckner [15].

$$k_b = 0.0509 * ln(H_t/D_e) - 0.015 \tag{2.21}$$

D_e is the equivalent bed diameter calculated by Equation 2.22 where L_x and L_y are the side lengths of the boiler cross section.

$$D_e = \frac{2L_x L_y}{L_x + L_y} \tag{2.22}$$

The model in its entirety is shown as Figure 2.1, which is taken from Djerf, Pallarès and Johnsson [21].

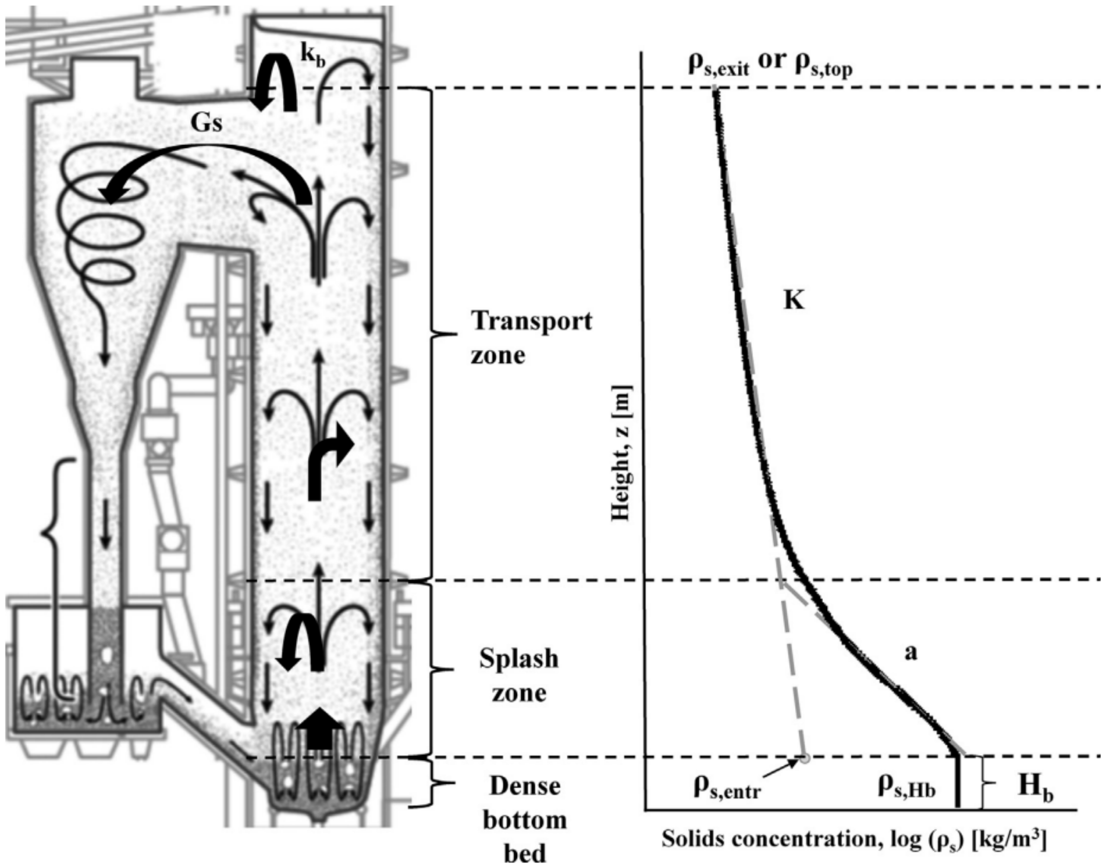


Figure 2.1: Concentration decay of particles in the CFB boiler.

2.3.2 Merrick and Highley empirical correlation

The data used in the Merrick and Highley study was obtained from a fluidized bed combustion plant with a cross section of 0.91×0.91 m^2, with fluidization velocities operating in a range of $0.61 - 2.43$ m/s and temperatures in a range of $704 - 899$ celsius.

The correlation aims to quantify the elutriation of discretized particle sizes $G_{s,i}$. The total external circulation is the sum of each particle contribution, where the contribution of each one is proportional to the total weight fraction of the particle size in the distribution.

Dimensional analysis found $\frac{G_{s,i}}{\rho_g u_{g,bot}}$, $\frac{u_{t,i}}{u_{mf,i}}$ and $\frac{u_{mf,i}-u_{g,bot}}{u_{g,bot}}$ to be the most important groups for determining elutriation. Using these groups and fitting an equation empirically gave Equation 2.23.

$$G_{s,i} = \rho_g u_{g,bot}(A + exp\left[- 10.4\left(\frac{u_{t,i}}{u_{g,bot}}\right)^{0.5}\left(\frac{u_{mf,i}}{u_{g,bot} - u_{mf,i}}\right)^{0.25}\right] \tag{2.23}$$

Merrick and Highley also mentioned bed height might affect the elutriation, although they did not have enough data to discern this fact [18].

2.3.3 Particle slip velocity

A circulating fluidized bed boiler can be split into different regions, characterized based on the fraction of solids. External circulation occurs at the columns exit, in which the fraction of solids ϵ_{se} is very low compared to the bottom of the bed. Kunii and Levenspiel [1] found that ϵ_{se} is slightly larger than the carrying capacity of the gas, something which can be attributed to the so called slip velocity between the gas and the solids. This slip velocity is defined as $u_{slip} = u_{g,top} - u_s$ where u_s is the mean solid velocity at the exit level of the column.

Knowing that ϵ_{se} is the fraction of solids and that u_s is the velocity of solids, we can now define the elutriation rate with the following equation

$$G_s = \epsilon_{se} u_s \rho_s * (1 - k_b) \tag{2.24}$$

Since Kunii and Lievenspiel aimed to explain the phenomenon in its most general form, $u_{g,top}$ was defined using the solid fraction at the exit $u_{g,top} = \frac{u_{g,bot}}{1-\epsilon_{se}}$, meaning the solids mean velocity can be defined as

$$u_s = \frac{u_{g,bot}}{1 - \epsilon_{se}} - u_{slip} \tag{2.25}$$

At the top of the freeboard of the boiler, the fraction of solids is around the range of $0.02 - 0.06$, leading to the assumption that the particles are almost completely dispersed in the gas stream. Completely dispersed particles implies that the interactions between particles are almost negligible, therefore; it can be assumed that

$$u_{slip} \approxeq u_t \tag{2.26}$$

[1]. With this assumption, the solid velocity in the top of the boiler and elutriation rate can be calculated.

2.4 Data pretreatment and analysis

In order to perform initial data pretreatment and analysis, some mathematical methods were used. These are explained in this section.

2.4.1 Correlation Matrix

A correlation matrix is a matrix visualizing the correlation between all variables. It represents a grid where each row and column is a variable. Where two variables meet, the correlation between the two is shown. There are different correlation metrics that can be used. The primary one is the Pearson correlation coefficient, which is calculated with Equation 2.27

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}} \tag{2.27}$$

where x_i and y_i are the individual occurences of the variables, and $\bar{x}$ and $\bar{y}$ are the average values of the variables [22]. A weakness with this metric is that it only measures linear correlation. To measure non-linear monotonic correlation, one can use the Spearman rank correlation coefficient [22]. It is calculated with Equation 2.28

$$1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{2.28}$$

where d_i is the difference in rank between two observations. Rank is defined as the order in which the variable value comes if all values of that variable are sorted by size.

2.4.2 Z-score normalization

Z-score normalization is a technique used to normalize variables, to prevent the magnitude of different variables from affecting the model. This is done by calculating the mean and variance of a variable, and then subtracting each data point with the mean and dividing with the variance. This makes each variable have zero mean and unit variance. The formula can be found as Equation 2.29.

$$z = \frac{x - \bar{x}}{\sigma_x^2} \tag{2.29}$$

2.4.3 Principle component analysis

Principle component analysis (PCA) is an unsupervised dimensionality reduction method, which is based on finding the straight line which maximizes the variance of the data when projected onto the given straight line [23]. This latent variable is called a principle component, and it is the linear combination of feature variables that has the ability to describe the largest amount of the total data variance. The equation for one principle component is shown in Equation 2.30. The procedure can then be repeated to find new lines, orthogonal to the first, which explain the most of the remaining variance. The dimensionally reduced space then explains a large part of the data variance, and different metrics can be plotted in the principle component space to visualize trends in the dataset, which would otherwise not be possible for a dataset with a lot of variables.

$$M_k = \sum_{i=1}^{n} p_{k,i} x_i \tag{2.30}$$

The sum of squares of the coefficients $p_{k,i}$ (loadings) for each principle component is one to obtain a unique solution and prevent scaling. This constraint is shown in Equation 2.31.

$$\sum_{i=1}^{n} p_{k,i}^2 = 1 \tag{2.31}$$

By projecting a data point $x_i, x_{i+1}, ..., x_n$ onto a principle component by inserting it into Equation 2.30, the data point is transformed into PCA-space. This is called the score t_k of the variable [24]. This can be expressed in matrix form as shown in Equation 2.32 where capital letters represent scores, variables and feature variables across multiple dimensions.

$$T = X * P \tag{2.32}$$

Scores T, alongside the previously mentioned loadings P are the two main metrics that are used in PCA [24]. The loadings are as previously mentioned the coefficents for the variables used when creating the principle components, representing the locations of the variables in principle component space. Scores that are close together are data points that have similar feature values, and loadings that lie in the same direction are variables with high correlation. Loadings with high values also impact the scores more, simply because the value in the linear combination is higher. The scores and loadings are also related. Scores that are located in the same direction as a loading are data points with a high value of that specific variable. This shows the properties of scores in different regions of the plot.

2.4.4 Partial least squares regression

Partial least squares regression (PLS) is also a dimensionality reduction method, which works similarly to PCA [24]. Latent variables are still generated that are a linear combination of all the feature variables, as described in Equation 2.30. However, PLS is not an unsupervised machine learning technique, because the model considers the target variables in order to make the latent variables good predictors of the targets [25]. The targets are therefore also dimensionally reduced to a latent variable through a linear combination. Unlike PCA, the latent variables from the features and targets are not created to explain the most of the variance, but rather to maximize the covariance between the two. There is also a difference in the notation [24]. The linear combination which transforms the data points to scores is in the PLS case called weights and denoted w as opposed to loadings denoted p in PCA. The transformations to x-scores and y-scores for target and feature variables are shown in Equation 2.33 and Equation 2.34 respectively.

$$T = X * W \tag{2.33}$$

$$U = Y * C \tag{2.34}$$

X-scores T and x-weights W for PLS are used and interpreted similarly to scores and loadings for PCA [24]. The change is simply in the way that the latent variables are constructed to maximize covariance instead of variance. Scores that are close together in PLS share variance that is important for explaining the target value, and weights in PLS that lie in the same direction share the same variance used for predicting the target value. Large weights also indicates that a variable is important for predicting the targets, since the variable is important for the latent variable maximizing correlation with the targets. The y-weights C can also be plotted with the feature variables, showing the correlation between features and targets.

2.5 Artificial neural networks

Artificial neural networks is a form of computation inspired by how the human brain stores information. The brain is built on a network of neurons which send signals to each other. Similarly, an ANN learns by changing the connections between their neurons based on data given. These neural networks can be trained to find non-linear patterns and then learn to generalize to other data [26]. A famous theorem found by Hornik [27] states that any *standard multilayer feedforward network with a single hidden layer, arbitrary bounded and nonconstant activation function are universal approximators.*

2.5.1 Neurons

The neurons a neural network is built up out of act as non-linear, parameterized, bounded functions. Each neuron receives n inputs and outputs a value a_i which is a linear combination of the inputs p_i, weighted by the parameters w_i and biased by the parameter b_i. This operation creates the potential v_i which is shown in Equation 2.35.

$$v_i = b_i + \sum_{i=1}^{n-1} w_i p_i \tag{2.35}$$

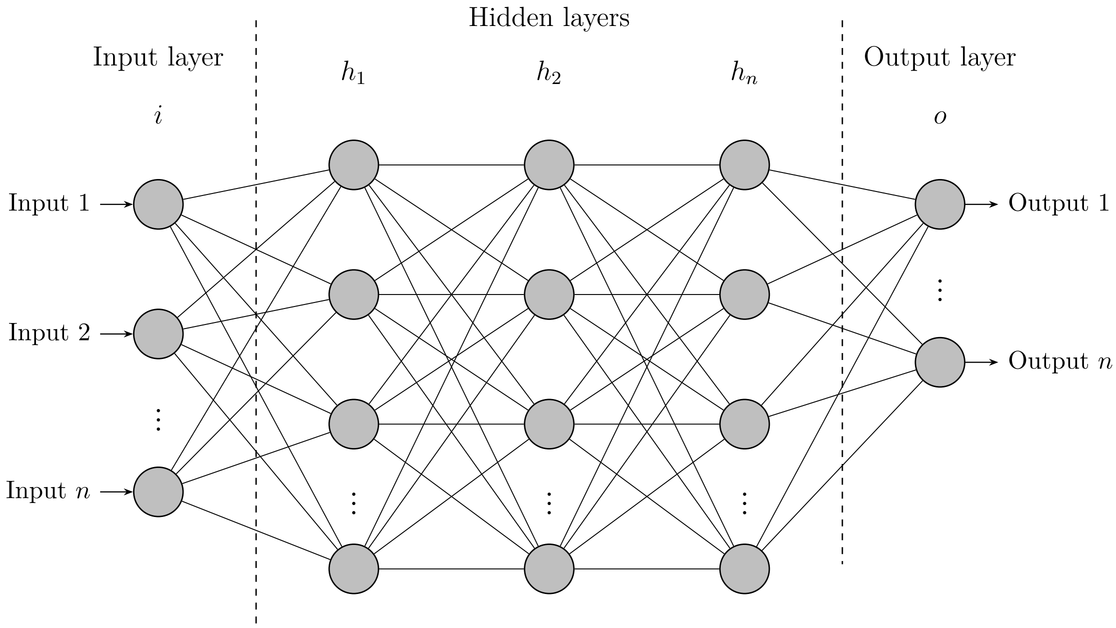
In order to capture non-linear trends, the activation function needs to be non-linear itself. This is often done by using the rectified linear unit (ReLU) activation function. The potential is passed through the ReLU function, leaving the output of the neuron [28].

$$a_i(v_i) = max(0, v_i) \tag{2.36}$$

2.5.2 Feedforward network

A feedforward neural network is built out of layers of neurons. The initial layer of neurons are called input neurons, each of which is connected to each neuron in the next layer. Data is passed through each layer, being affected by biases, weights and activation functions until it reaches the final layer, called the output neurons. The layers in between the input and outer are often described as hidden layers. The whole system is illustrated below.

Figure 2.2: Neural network structure.



2.5.3 Training of a network

In order to train a network, a measurable value needs to be assigned to the difference between predicted output y_{pred} produced by the network and the true target y_{true} given by the dataset. This value is given by a loss function $\mathcal{L}(y_{pred}, y_{true})$. The training process aims to find network parameters $\theta = [w, b]$ that minimize the loss over the training dataset.

With each activation function being known, the minimization of the loss can be done using gradient-based optimization. The gradients of the loss with respect to the network parameters are computed using a backpropagation algorithm, where the gradients are derived from using the chain rule of calculus. Each loss is then weighted by a set learning rate η in order to avoid unstable optimization.

The loss computations are repeated a number of times, a number called epochs, until certain conditions are met. The conditions are unique to the experiment set up. In order to speed up the process of training, multiple inputs can be passed through the network in parallel. This multitude of inputs is denoted by the hyperparameter batch size.

2.5.4 Hyperparameters

Hyperparameters are parameters unique to each network, unaffected by the optimization from the backpropagation algorithm and often used to regulate the training rate. These include the aforementioned number of layers, number of neurons per layer, learning rate, number of epochs and batch size [28].

2.6 Optimization and generalization of a neural network

Even though it is proven that a feed forward neural network with sufficient capacity can approximate any function [Hornik 1989, 1990, 1991], it is not a guarantee that the training process will find this optimal set of parameters. The issues which are the most common are poor optimization, meaning divergence or slow convergence during training; and bad generalization, meaning the model does not perform well outside of the training set. Poor generalization is either called underfitting, a phenomenon where the model performs poorly on both training and test sets, or overfitting; where the model performs well on training data but poorly on unseen data.

2.6.1 Stochastic gradient-based optimization

Stochastic gradient-based optimization methods such as the Adaptive moment estimation (ADAM) optimizer are used to efficiently train neural networks. ADAM improves optimization by calculating moving averages of gradients, which results in adaptive step sizes for each individual parameters. This allows for faster convergence and improved stability, particularly in the presence of noisy gradients.

The Adaptive moment estimation with decoupled weight decay (ADAMW) is an improved version of the ADAM Optimizer, which decouples weight decay from the adaptive gradient update. This leads to a more consistent regularization effect across parameters.
By reducing issues such as unstable updates and slow convergence, the optimizer improves the efficiency of training and helps the model reach a good solution more reliably. This primarily affects optimization and does not have a large effect on generalization [29] [30].

2.6.2 Weight decay

Weight decay is a regularization technique used in machine learning to reduce overfitting. It introduces a weight decay parameter γ that controls the strength of the decay applied to model weights. During training, weights are gradually shrunk toward zero by directly scaling the weights each update (as in ADAMW). This discourages overly large weights and improves generalization.

2.6.3 Dropout regularization

A common issue that often leads to poor generalization in neural networks is co-adaptation, where neurons become overly dependent on specific other neurons. This means that features are only useful when certain other features are also present, which makes the network unstable and prone to overfitting.
Dropout is a regularization technique designed to reduce this co-adaptation. During training, each neuron is randomly omitted with a small probability, meaning the network is effectively trained on different sub-networks of itself each time.

As a result, neurons cannot rely on the presence of specific other neurons and must learn features that are useful in many different network configurations. This reduces co-adaptation and improves generalization [31].

2.6.4 Kaiming weight initialization

Kaiming initialization, introduced by Kaiming [32], is a weight initialization scheme designed to preserve the variance of signals in deep neural networks with ReLU activations. The method is derived under the assumption that inputs and weights are independent and zero-centered, and it seeks to prevent vanishing or exploding activations by maintaining a constant variance across layers. For a layer with n_{in} input units, weights are initialized such that:

$$Var(W) = \frac{2}{n_{in}} \tag{2.37}$$

This scaling compensates for the fact that ReLU activations zero out approximately half of their inputs, thereby reducing the effective variance of the propagated signal. By increasing the initial weight variance accordingly, Kaiming initialization ensures that both forward activations and backward gradients remain well-conditioned during training. The method can be implemented using either a normal distribution $N(0, n_{in})$ or a corresponding uniform distribution. It has become a standard initialization technique for deep networks employing ReLU and its variants. [32].

2.7 Grey-box modeling

In this project, there are two different methods incorporated into the neural network. These methods utilize both machine learning and mechanistic models, often described as grey-box modeling. These are explained below.

2.7.1 Correction term model

The empirically derived formulas for evaluating entrainment have been found to have quite a large margin of error for different operating parameters [2]. In order to compensate for this, a neural network can be trained to correct the fault of the empirical solution. This method is similar to the universal differential equation (UDE), which is a differential equation defined by a universal approximator. In this case, the universal approximator acts as a sort of correction term, meaning it is capable of representing any possible function with a given set of parameters. The general form of this concept is

$$\frac{du}{dt} = f(u, t, \Theta) + g(u, t, \Theta) \tag{2.38}$$

where the function $g(u, t, \Theta)$ acts as the correction term to the physical equation $f(u, t, \Theta)$ [6]. For a system without derivatives, a similar equation can be implemented as seen below.

$$u(\Theta) = f(\Theta) + g(\Theta) \tag{2.39}$$

The goal of this integration into neural networks is to combine the best aspects of machine learning with mechanistic models. As the function $g(\Theta)$ is unknown, it is replaced with a neural network as such $g(\Theta) = NN(\Theta)$ to approximate the correction term.

2.7.2 Physics informed loss function

Another implementation of a physics informed neural network is to inform the network of a physically based loss rather than correction. As mentioned above, neural networks use gradient-based optimization to find the optimal weights. This is called data-based loss and requires a sufficient amount of data to be effective. That data need can be reduced if a physics informed residual, based on the outputs and inputs of the network, is introduced into the loss function in the following way

$$L(y_{pred}, y_{true}, \Theta, \lambda) = L_{data}(y_{pred}, y_{true}) + \lambda L_{physics}(y_{pred}, \Theta) \tag{2.40}$$

This pushes the network to learn not just based on data, but also the physical relations which define the system; reducing the need for data significantly [14],[9]. These loss weights play an important role in improving trainability and are to be used-defined or tuned during automatically [33].

2.8 Shapley and Owen values

Shapley values originate from game theory, where it is a metric used to calculate individual player payout in a coalition based on contribution [34]. It was popularized as an explainable AI metric by Lundberg and Lee in 2017 with the introduction of Shapley additive explanations (SHAP) [35]. Shapley values are calculated using the feature values as players, and the prediction as the payout, splitting the contribution of the model toward the prediction fairly between the feature values. There are multiple SHAP methods to estimate Shapley values, depending on both the data and the model. However, with a small enough dataset and simple enough model, the exact Shapley values can be calculated.

Exact Shapley values ϕ_j are calculated for each prediction by considering coalitions S of feature variables [34]. A coalition is any possible combination of feature variables excluding the feature variable j for which ϕ_j is calculated. For each S, model predictions are made where the variables that are not a part of S or j draws random values from the training set, to represent not having information about variables outside of the coalition. Predictions are also done while pulling random values from the training set for variables that are not a part of S (including j). The Shapley value for S is then the difference between the average over predictions of the model with and without information about j, where lack of information is represented by the previously mentioned random sampling. ϕ_j is then a weighted average over all possible S, where the most weight is put towards coalitions with few or many variables. This is done, since that more clearly shows the isolated effect of j or the interaction effects with the prediction as a whole. This weighting is shown in Equation 2.41

$$\phi_j(val) = \sum_{S \subseteq \{1,\ldots,p\} \setminus \{j\}} \frac{|S|!(p - |S| - 1)!}{p!} (val(S \cup \{j\}) - val(S)) \tag{2.41}$$

where p is the total number of model features and $val(S)$ is the prediction of the model with information about variables in subset S [34].

One weakness with Shapley values is that the method inherently assumes that the variables are independent [34]. This leads to the random sampling creating data points which are highly unprobable if variables are correlated. One example could be taking two highly correlated variables, and maximizing one while minimizing the other. This is a data instance representing a fictional case that will never happen. To circumvent this, one can hierarchically cluster the variables, and calculate the Shapley values for each cluster [36]. The Shapley value calculation can then be done within each cluster to fairly attribute the prediction to correlated variables. The attribution is then called an Owen value.

3
Method

3.1 Limitations

A project of this scale needs to adjust its scope in order to fit within the given time frame. In order to achieve this, certain limitations were set. Many of the methods used for obtaining the data have a margin of error caused by measurement uncertainties. These uncertainties exist for all measurements in the CFB boiler. In order to streamline the scientific process of the project, all data points were considered to be correct.

Within PIML there are a great many different models to consider, even for the case of CFB boilers. Therefore, the second limitation is that only three models were trained and evaluated in light of the time limitation of the project.

3.2 Data

The data for the project comes from the Chalmers power central, where a 12 MW CFB boiler with dimensions 1.5x1.5 meters with a height of 14 meters is used for research [37]. The data was obtained as a time series, where all of the operating parameters of the CFB boiler are recorded over time. The external solid circulation cannot be measured at steady-state and therefore is not recorded over time. It is instead calculated by doing a blow-out test. The blow out test starts with the boiler reaching a steady state, after which the recirculation of solids suddenly is stopped. The solids will then empty out of the boiler, reducing the pressure drop, which is correlated to the external solid circulation by Equation 3.1.

$$G_s = \frac{d\Delta P}{dt} * \frac{A}{g} \tag{3.1}$$

Equation 3.1 is created by assuming that the entire pressure drop is due to the static head, neglecting pressure drops due to acceleration, friction and air density in the boiler. This assumption is justified by Kunii and Levenspiel, who found that 80-90% of the pressure drop in a fast fluidized bed is due to the static head [1]. The pressure drop must therefore be due to solids leaving the boiler creating Equation 3.2

$$\frac{d\rho}{dt} = \frac{d\Delta P}{dt}\frac{1}{gh} \tag{3.2}$$

Each blow out test was repeated 1-4 times, and the true value was taken as the average of all measurements. The test has been done a lot of times at the Chalmers boiler over the last 15 years, operating at different conditions and varying the fluidization particles. 148 data points were able to be recovered for this project, representing five different fluidization particles and a wide range of boiler conditions. The fluidization particles used were olivine, ilmenite, copper smelter slag (järnsand), Baskarp sand and silver sand. Baskarp sand and silver sand is mainly silica oxide, while olivine has both silica oxide and magnesium oxide. The copper smelter slag is mainly iron oxide, and ilmenite is an oxygen carrier mainly comprised of titanium oxide and iron. All particles used are in Geldart group B.

The selection of feature variables were based on Chew, Cahyadi, Hrenya, Karri and Cocco who made a review of entrainment correlations in [2]. All variables that were used for a correlation in that review were considered for this project. The variables are the Archimedes number, the Reynolds number, Temperature, particle size, particle density, air density, air viscosity, gas velocity, terminal velocity of particles, Re number at terminal velocity, minimum fluidization velocity, cross sectional area of boiler, diameter of boiler, particle Froude number, drag coefficient and freeboard height. Some of these values, like the cross sectional area and freeboard height are constant, but the other vary based on the run.

The power central had many different measurements to choose from to use as feature variables, but not all of them were always measured when doing the blow-out tests. In total, there were 12 features that were always measured. Two were discarded, being the oxygen gas percentage in the flue gas and the air mass flow used to transport the fuel. The oxygen gas percentage describes the combustion process, but the main effect of combustion relevant for the external circulation was considered to be the temperature for air density and viscosity calculations. The fuel transport gas flow was included in another metric, total secondary air, meaning that the information was contained in another part of the data. As will be described in the results, the metrics were also not very correlated with the external circulation rate.

The other 10 metrics were used to calculate all other variables and use as inputs for the model. These were pressure drop over column, temperature in the bottom of the column, temperature at the top of the column, gas velocity in bottom of the boiler, primary air mass flow, secondary air mass flow, total air mass flow, recirculated flue gas mass flow, particle density and sauter mean diameter of particles. The pressure drop was kept as an explaining variable since it was used for the measuring technique of the external circulation and describes the total solids mass in the boiler.

The particle size distribution was chosen to be represented by the sauter mean diameter. Due to the data having a mix of materials, voidage at minimum fluidization was assumed to be $\epsilon_m = 0.42$. This is the same as mixed round sand in Table 3, Chapter 3 of Kunii and Levenspiel [1]. As no sphericity is provided, the particles are assumed to be near perfectly spherical, represented by a sphericity of $\phi_s = 0.95$. These assumptions were made to simplify the model.

3.2.1 Pretreatment and data curation

Since the feature data was obtained as a time series of values with one sample/minute, with only one recirculation rate being obtained, the steady-state values had to be obtained from the time series. This was done by averaging the ten samples before the blowout test was started. The sampling time is one minute, making it an average of 10 minutes. The z-values for each explaining variable as well as the external circulation for every data point was plotted as histograms in order to look for outliers. The z-values were calculated as previously described with equation 2.29. The outliers were identified by having an absolute z-value larger than 3. All outliers were manually examined, to see if any part of the calculation was done incorrectly. If the calculation was wrong, it was manually fixed.

The standard deviations of each variable in the ten samples that were averaged before the blowout test was also plotted, in order to look for non steady-state conditions before the blowout test. If the standard deviation at one test was a lot higher than other cases, it was manually examined, to see the cause of the issue. If the values were changing over the sampling time, the plant can be assumed to not be at steady state, making the average a poor representation of the boiler state, and the data point was discarded.

3.2.2 Data analysis

After the initial data curation and pretreatment, the data was analyzed further. First, correlation matrices were made, using both the Pearson correlation as well as the Spearman rank correlation. This was done to show which variables had the largest correlation with the external circulation rate, making them suitable predictors, as well as if any variables were very correlated with each other, and therefore redundant.

PCA was also done, encompassing all of the previous aspects. The scores of the first two principle components were plotted to look for outliers as well as uniformity or clustering of the data. The plot was also color coded according to which particle was used, to see if the behavior of the boiler changed with different particles. The loadings were also plotted to check for correlation between explaining variables, and see how each explaining variable affected the scores. The plot could also be used in combination with the color-coded scores to see how each variable was affected depending on the particle used.

The final data analysis done was PLS. This was done for the same reasons as the principle component analysis, as PLS produces similar results as PCA. However, PLS does connect the feature variables with the target variables. Thus not only showing the inherent structure in the feature data, but instead showing it’s connection to the target data.

3.3 Physical model testing

Two of the physical models proposed, that being Merrick and Highleys empirical correlation, and the decay-model, were tested to see how well they performed without being implemented into a neural network. Both models were applied to the entire data-set, in order to assess their accuracy. This was done to test the models predictive power by itself, with the hopes of later confirming whether the application of machine learning improved upon the physical models.

3.4 Neural network

The neural network was built up on layers of neurons, varying depending on the parameters of the experiment. Each neuron has a ReLU activation function, allowing the network to model non-linear trends. The number of inputs were chosen in accordance to the data analysis explained in section 3.2.

In order to streamline the training process, the ADAM optimizer given by PyTorch was used. Both weight decay, dropout and an adaptive learning rate was passed as hyper-parameters to the optimizer. These hyper-parameters were tested at different values in order to achieve the optimal regression model fit. In addition to implementing physically based models, a pure black box (BB) model was also developed as a benchmark.

3.5 Implementation of physical models

There were three different physical models considered for this project. The first was the Merrick and Highley empirically derived correlation, the second was the exponential decay model and the third was the PINN network.

3.5.1 Correction term model

The empirically derived Merrick and Highley correlation, Equation 2.23, was used in combination with a NN in order to achieve a less data reliant model. The external circulation predicted by the correlation was calculated using the input values of the network. The Merrick and Highley correlation calculates elutriation for a specific particle size based on primary gas velocity, the cross sectional area of the boiler, the terminal velocity of the particles and the minimum fluidization velocity of the particles.

As seen in subsection 2.3.2, the variables required for this computation are then $u_{g,bot}, T_{bottom}, T_{top}, \rho_s$ and d_p. Since the Merrick and Highley correlation assumed a constant temperature throughout the boiler, the average of the bottom and top temperature was used in Equation 2.11 and Equation 2.12. All 10 input variables were passed onto the neural network, which predicted a correction term. The correction term is then added onto the correlation prediction, which lastly gives a final prediction from the network. The final prediction is the value the network then learns to adapt it's loss function to. This is visualized in Figure 3.1

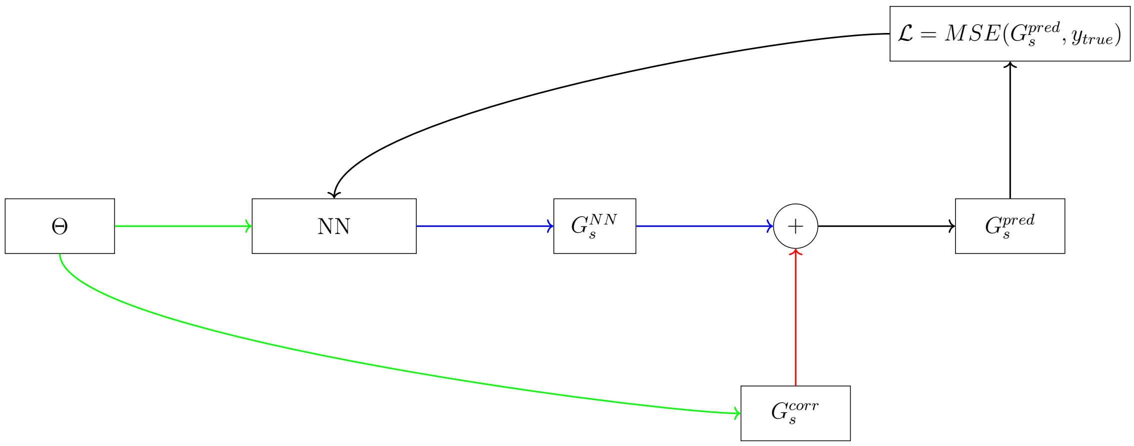


Figure 3.1: Correction term model.

3.5.2 Decay model

The exponential decay model presented in subsection 2.3.1 was also considered for a CTM-setup. However, as presented in the results in section 4.3, the physical model performed quite poorly by itself. Therefore, the choice was made to not incorporate it into the CTM-model, and instead focus on other models. The reasoning for this is explained more clearly in the discussion in section 5.3

3.6 PINN model

Last is the PINN model, which utilized the voidage, solid velocity and slip velocity at the top of the boiler in order to find the flux. The network predicts the solid velocity and voidage, which is then used to calculate the external circulation. The two outputs are also used to compute the residual used in a PINN, mentioned in subsection 2.7.2.

The inputs used for the physical loss were $u_{top}, T_{top}, \rho_s, d_p, Air_{tot}$ and F_{recirc}. As the phenomnenon occurs at the top of the boiler, $u_{g,top}$ was used in order to evaluate C_D and Re_p, and the temperature at the top was used when calculating the viscocity and density of the gas.

The physical loss residual was normalized by dividing the solids velocity predicted by $u_{g,top}$, in order to achieve a dimensionless loss. This is mathematically represented in Equation 2.40 as the variable λ. This was then multiplied with a factor α to make sure the physical loss was around the same scale as the data loss. α was then varied as a hyperparamter during experimentation. The PINN-model is visualized in Figure 3.2

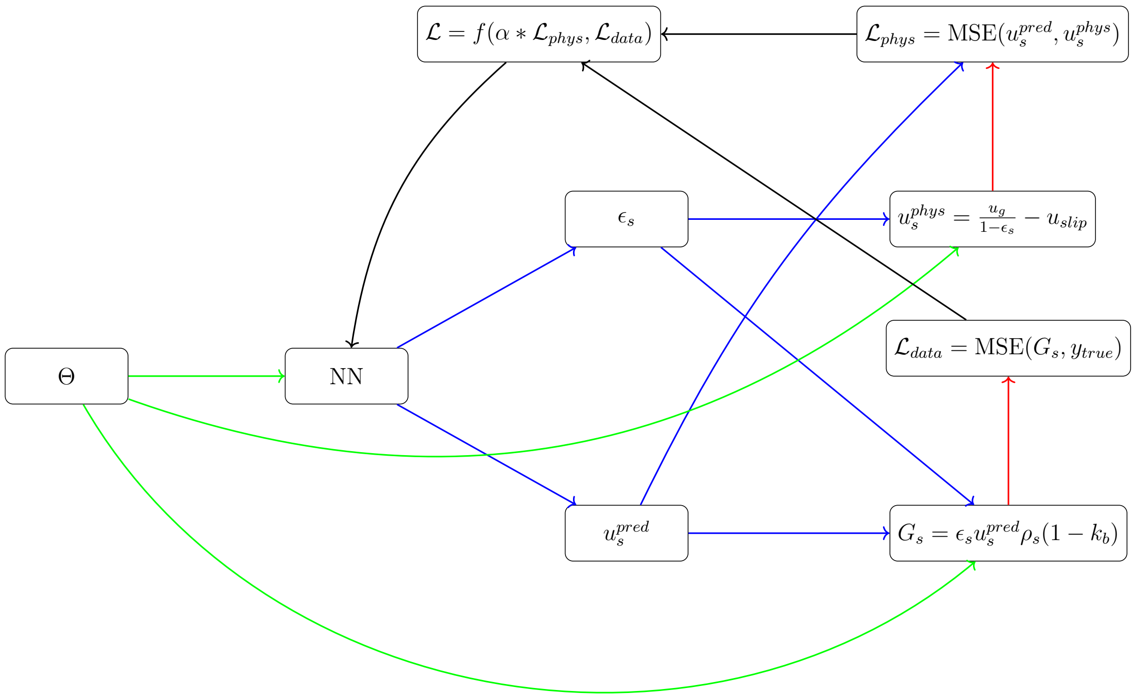


Figure 3.2: PINN model.

3.7 Experimental setup

In order to efficiently experiment with different configurations of the grey-box models, a standard configuration template was implemented, where hyper parameters could be tuned. These parameters included input size, number of layers, size of each layer, dropout rate, batch size, weight decay, learning rate, epochs, patience, which physical model should be used or if one should not be used at all.

3.7.1 Train-validation-test split

To evaluate model performance, data is split into three groups. One is the training data which is used to train the model. The second is the validation data used to validate the performance on unseen data during training. This is an important part of training, as it clearly shows when a model is prone to overfitting. In order to not cherry-pick the model performing best on the validation split, the final split called the test set is used on the best performing model to get a final evaluation of model performance on unseen data. The split used for this project was 64% training, 16% validation and 20% test.

3.7.2 Grid search

A grid search was executed in which different combinations of hyper-parameters were tested. Each hyper-parameters was assigned a finite set of values, hidden layer sizes $\in [[32, 32], [64, 32], [128, 64]]$; dropout rates $\in [0.1, 0.2]$; $\gamma \in [1e^{-3}, 3e^{-4}]$; $\eta \in [1e^{-3}, 3e^{-3}]$; Model type $\in [PINN, CTM, BB]$ and $\alpha \in [0.1, 1, 10]$. The hyper-parameters which were kept static were the Batch Size = 32, $Epochs = 50000$ and $Patience = 15000$. The grid search then evaluated all possible combinations of these values and saved models which had an $R^2_{val} > 0.5$. After the grid search, the models which then had the best average R^2_{val} and R^2_{test} were picked for evaluation.

3.7.3 Cross validation

After the grid search had established the best performing models for each network type, a cross validation was run for each of the configurations for the respective models. For each of the three best performing models, 60 different splits (later denoted as seeds) of the dataset were trained on in order to encapsulate the variance in performance due to data partitioning. This variance is presented in section 4.5 as a violin plot.

3.8 Owen values

As can be seen later on in the results in Figure 4.9, the seed caused large variance in model performance. Therefore, the Owen values were evaluated for the three different kinds of models for four different seeds, resulting in Owen values for 12 different models. The Owen values were evaluated for the test set in order to explain how the model predicts on unseen data. The hierarchical clustering however, was done on the training set; to ensure that the feature correlations identified by the model through training are the same ones that the Owen value calculation considers.

4

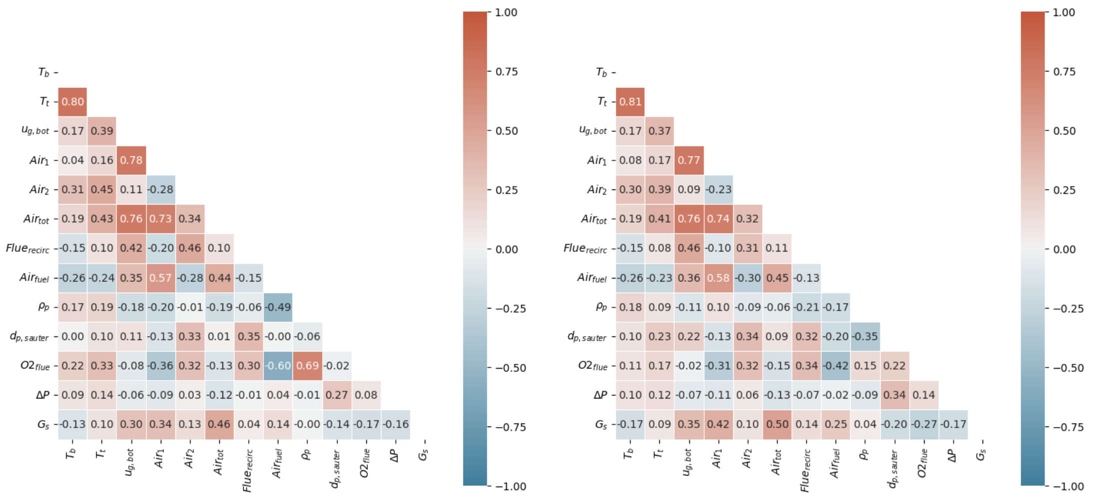
Results

4.1 Data analysis

Here, the results from the data analysis are presented. Each figure is given a brief explanation of what it represents, and interesting results are pointed out.

4.1.1 Correlation matrix

Two correlation matrices were calculated. One using the Pearson correlation as calculated by Equation 2.27 and one using the Spearman rank correlation as calculated by Equation 2.28. These are shown as Figure 4.1a and 4.1b respectively. T_b and T_t are the bottom and top temperature in the boiler. Air_{tot}, Air_1 and Air_2 are the total-, primary- and secondary air mass flow into the boiler. $Flue_{recirc}$ is the recirculated flue gas mass flow, u_g is the gas velocity in primary zone (bottom of the boiler), ΔP is the total pressure drop from the bottom to the top of the boiler and $d_{p,sauter}$ and ρ_p is the particle Sauter mean diameter and density respectively. G_s is the target variable, solid external circulation rate.



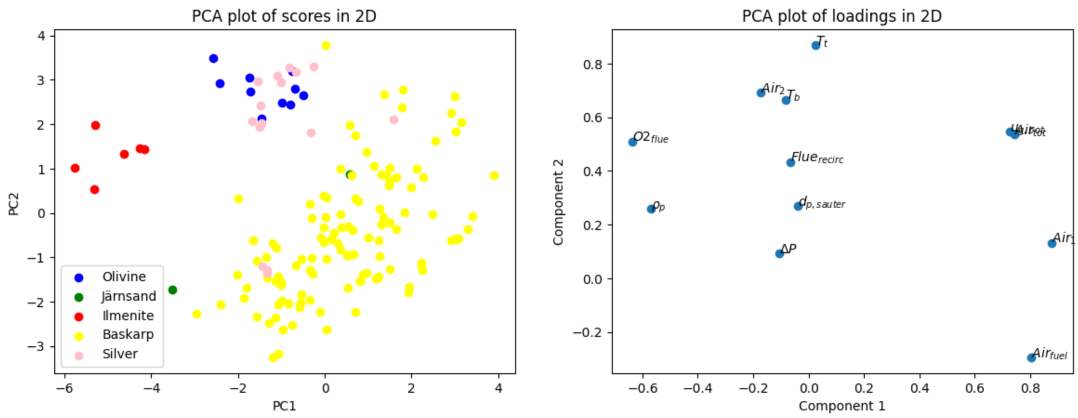
(a) Pearson correlation matrix. **(b)** Spearman rank correlation matrix.

Figure 4.1: Correlation matrices for the data.

The two figures show quite similar results, showing that the correlation between the variables is linear and monotonic. The largest absolute correlation can be seen between top and bottom boiler temperature, followed by the correlations between primary air, total air and gas velocity. All of these are positive correlations.

4.1.2 PCA

PCA analysis is presented through two plots, the score and loading plot. They show the location of each data point and variable in the space of the first two principle components. They are shown as Figure 4.2a and 4.2b. The colorings of the dots refers to the type of particle used in the boiler during each run, with the legend describing which color is which particle.



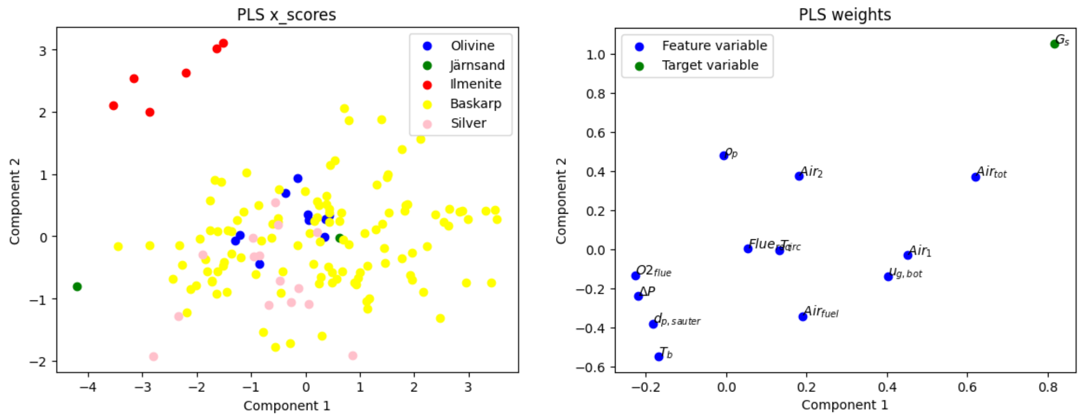
(a) Score plot from PCA. **(b)** Loading plot from PCA.

Figure 4.2: Score and loading plots for PCA analysis.

The score plot shows a clear clustering based on particle type, with ilmenite forming its own cluster and one Baskarp sand data point along with most of the silver sand and all of the olivine grouped together in its own cluster. The other points are grouped together in the main cluster, consisting mainly of Baskarp sand. The loading plot shows that the gas velocity in the bottom of the boiler ends up in the same place as the total air flow. Apart from that, the variables are quite well spread out, with most of the flow related variables on the right side and temperatures alongside secondary air in the top. The fuel spreader and flue gas oxygen also represent unique variance in the dataset, since they are far away from other variables.

4.1.3 PLS

The PLS results are shown in a similar manner to the PCA results, with a score plot and a weight plot. They are shown as Figure 4.3a and 4.3b.



(a) X-scores plot from PLS. **(b)** Weight plot from PLS.

Figure 4.3: Score and weight plots for PLS analysis.

The score plot for PLS also shows clear signs of clustering, but only the Ilmenite now stands as its own cluster. The copper smelter slag has one point outside of the main cluster, but it only has two points and one of them is in the middle of the main cluster. The weight plot shows the air mass flows and gas velocity as the main features describing the external circulation rate. Flue gas oxygen concentration is also close to pressure drop, unlike the PCA loading plot. The fuel spreader flow however shows unique covariance, but with low loading values and low correlation to external circulation . The gas velocity in primary zone has also separated from the total air.

4.2 Merrick and Highley empirical correlation

The results for the empirical correlation are presented in Figure 4.4, where the straight line is $y_{true} = y_{pred}$.

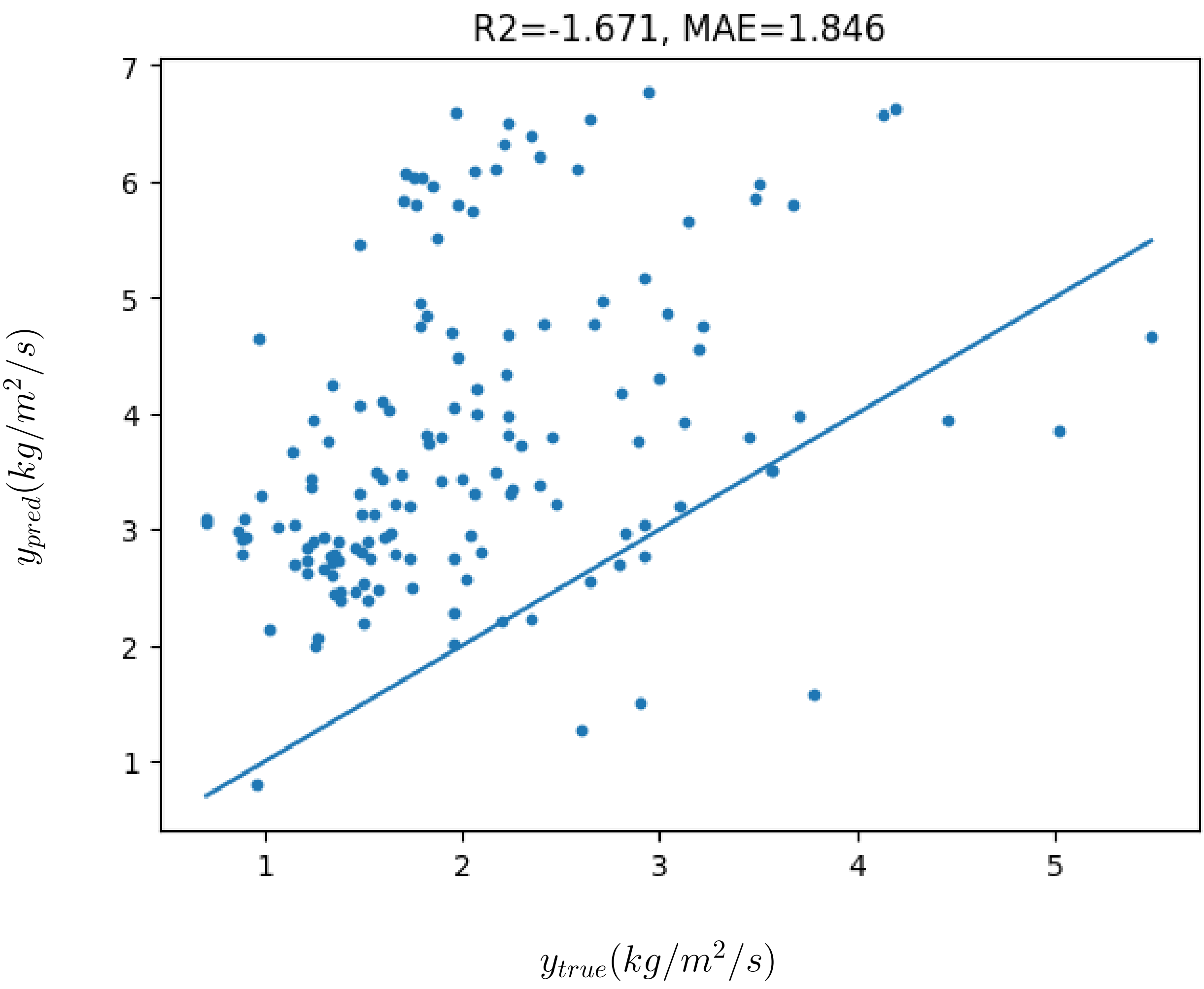


Figure 4.4: Result plot for Merrick and Highley correlation.

The R^2-score of the correlation is -1.671 and the MAE is 1.846. Looking at the figure, it can be clearly seen that Merrick and Highley often overestimates the flux.

4.3 Decay model

The results for the decay model are presented in Figure 4.5. The straight line is $y_{true} = y_{pred}$.

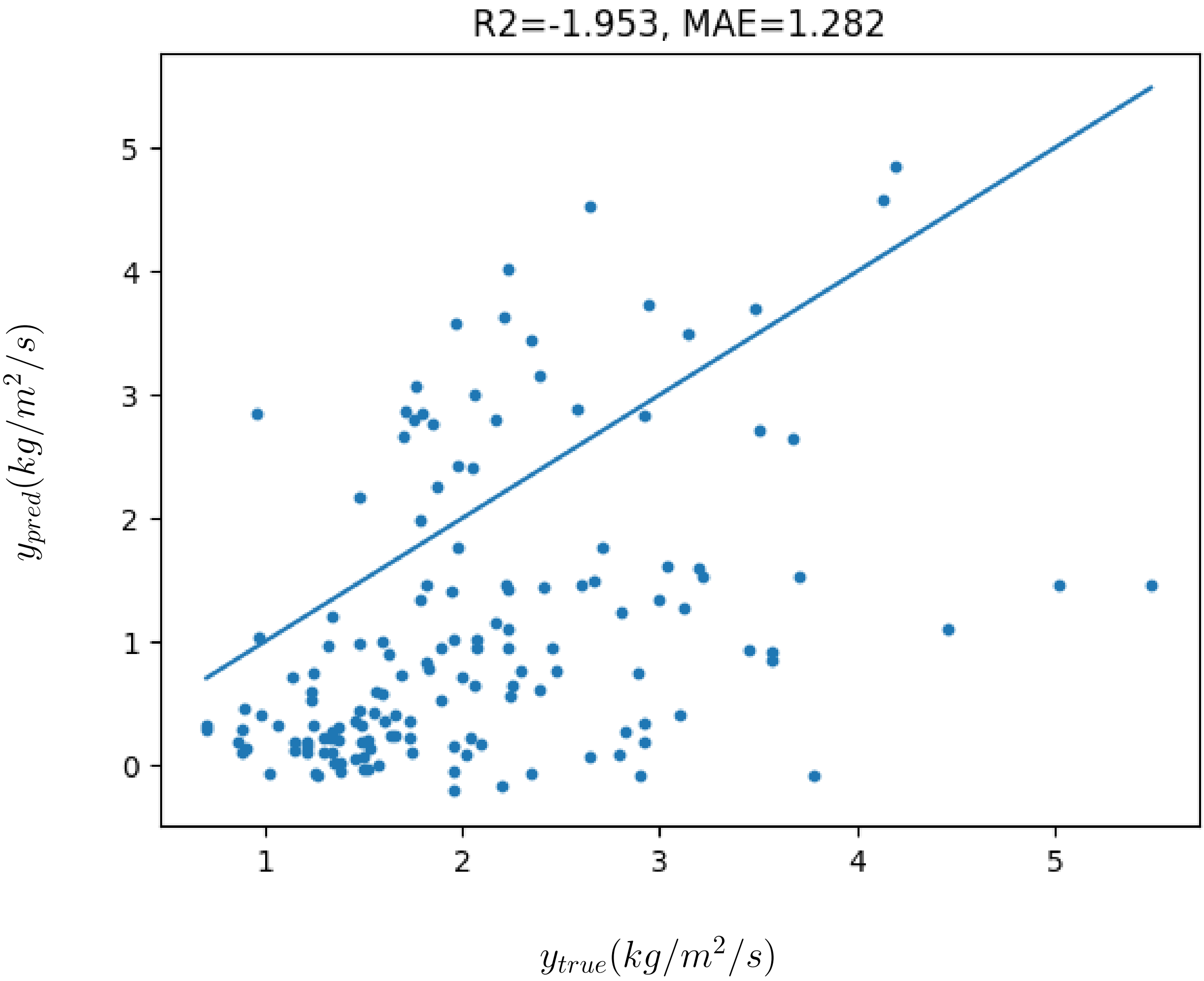


Figure 4.5: Result plot for decay model.

The model has a negative R^2-score of -1.953 and the MAE is 1.282, indicating that it predicts worse than simply taking the average external circulation rate. Very few predictions are close to the straight line, meaning that the predictions are not close to the true value. Most predictions are lower than the true values, predicting near zero external circulation, and some predictions are even negative.

4.4 Model performance

Performance results are shown in Table 4.1, with each of the best performing configurations and their respective R^2_{val},R^2_{test} and MAE. As the black box model did not perform above a $R^2_{val} > 0.5$ more than two times, four PINN models are shown instead.

The performance of the configurations for each type of model is presented in a scatter plot of y_{pred} against y_{true} on the test set, where the R^2_{test} and MAE are displayed in title.

Table 4.1: Neural network hyperparameters and performance.

Network Type	Hidden Sizes	Dropout	α	γ	η	R^2_{val}	R^2_{test}	MAE
PINN	(32,32)	0.1	1	1e-3	3e-3	0.6245	0.2462	0.5900
PINN	(128,64)	0.2	1	3e-3	3e-3	0.6104	0.3306	0.5230
PINN	(128,64)	0.1	1	3e-4	3e-3	0.6034	0.3213	0.5120
PINN	(128,64)	0.2	1	1e-3	1e-3	0.5692	0.4030	0.5190
CTM	(32,32)	0.1	-	3e-4	3e-3	0.5767	0.3798	0.4600
CTM	(128,64)	0.1	-	1e-3	3e-3	0.5457	0.1780	0.5660
CTM	(32,32)	0.2	-	1e-3	3e-3	0.5222	0.4066	0.4860
BB	(32,32)	0.2	-	1e-3	3e-3	0.6351	0.3066	0.4710
BB	(32,32)	0.1	-	3e-4	3e-3	0.5334	0.3341	0.5250

$^{\alpha}$ loss weighting parameter for physics term.
$^{\gamma}$ weight decay parameter.
$^{\eta}$ overall optimizer learning rate.

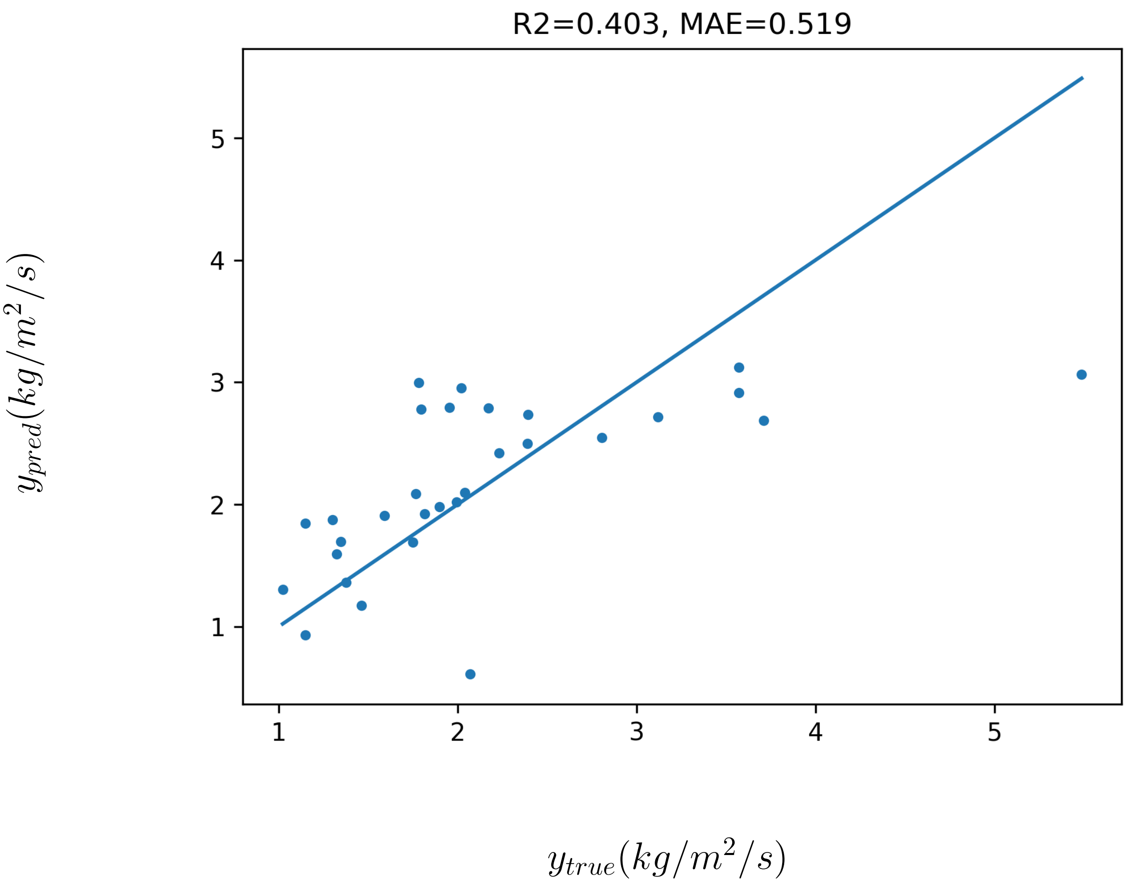


Figure 4.6: Best performing PINN network.

The best performing PINN network generally predicts external circulation well, although it can be seen that predictions worsen as the true external circulation value gets higher. The largest outlier appears at the highest external circulation rate, in which the network predicts a much lower value.

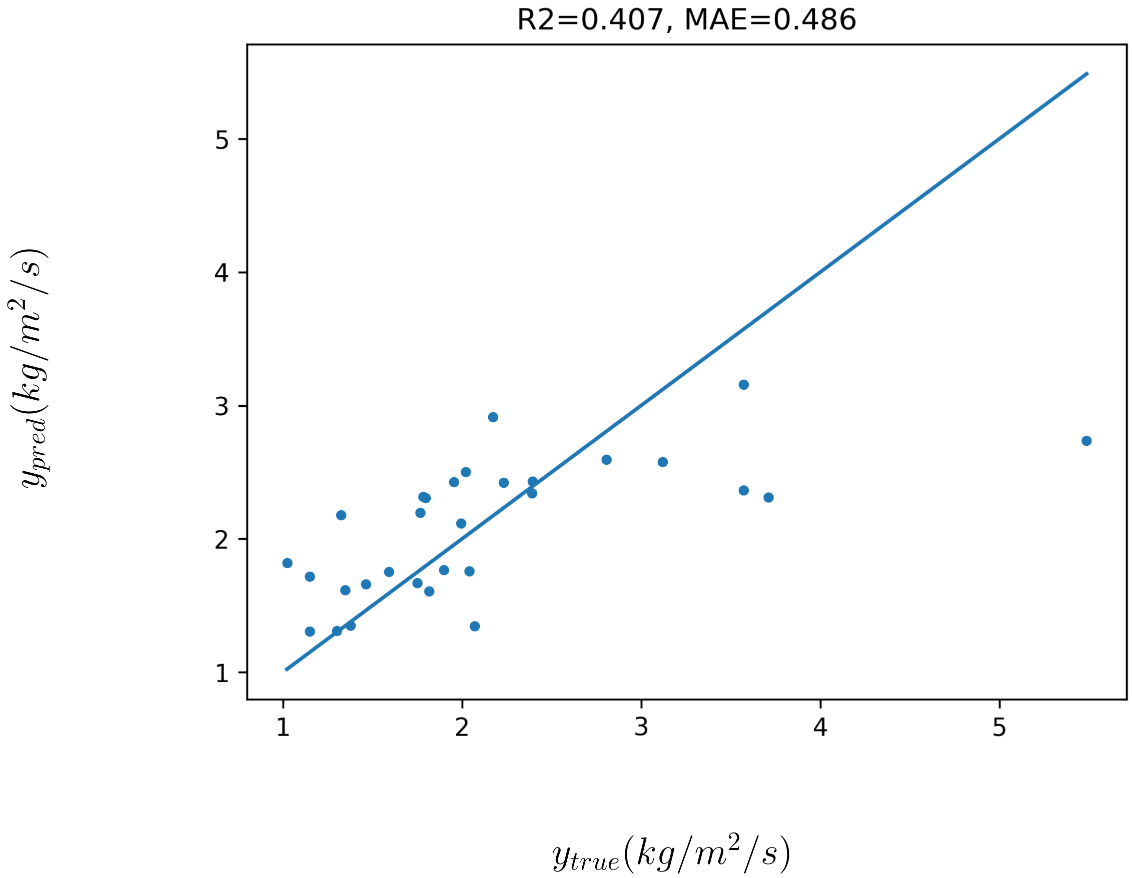


Figure 4.7: Best performing CTM network.

Similarly to the PINN network, the CTM network also predicts lower external circulation rates fairly well while underperforming as the true values increase. The same outlier can be found for the CTM model.

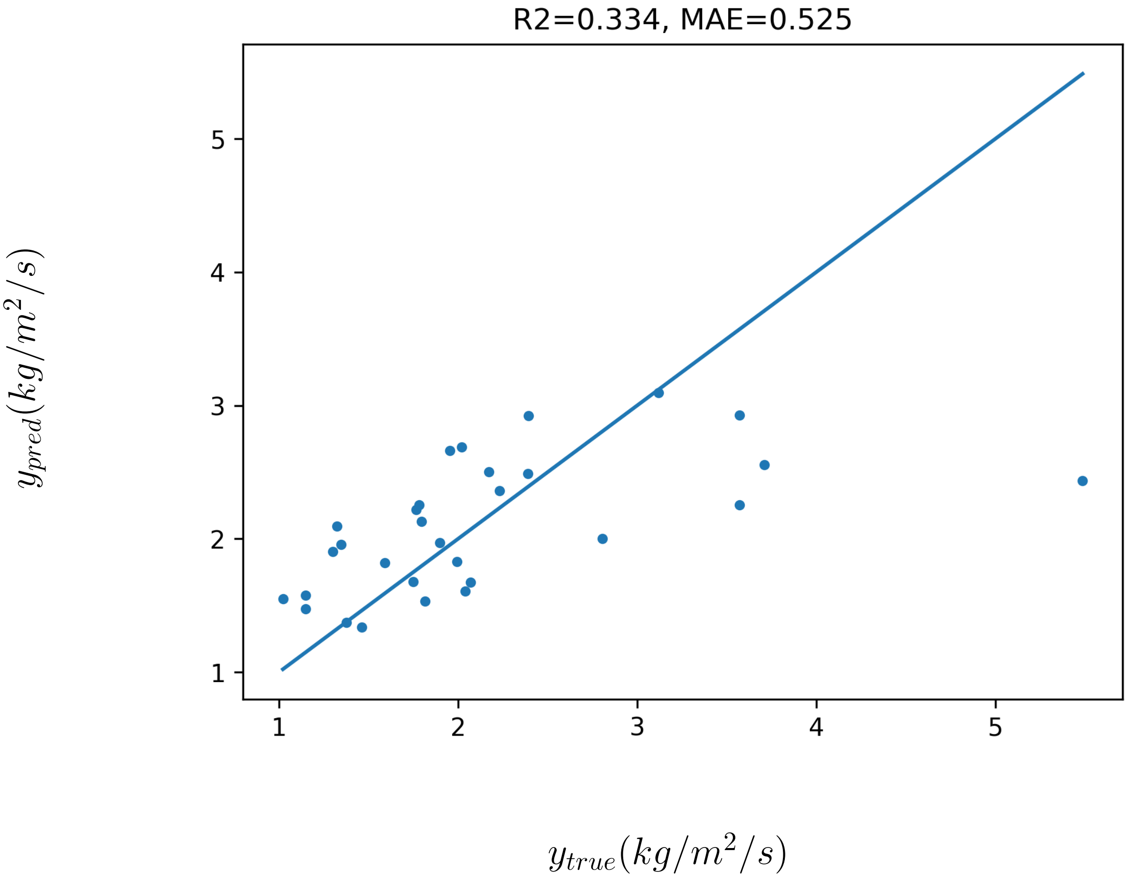


Figure 4.8: Best performing BB network.

Lastly, the BB model also performs better for lower external circulation values, however, the performance is slightly worse than the two previous models. The same outlier can also be found for the BB model.

4.5 Cross validation of performance

The results of the 60 different data-set splits are represented in a violin plot. Each data point is shown as a gray marker and the width of the violin shows an increase in the concentration of similar points. The average test R^2-score is 0.376 for PINN, 0.362 for BB and 0.370 for CTM.

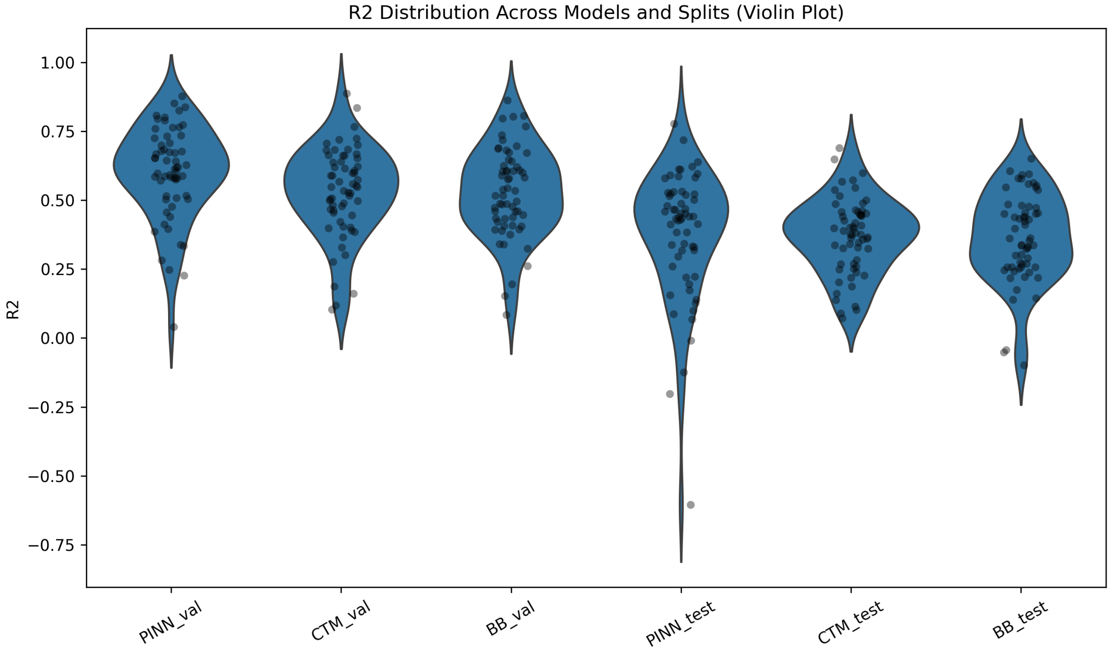


Figure 4.9: Violin plot of R2 performance values for each network.

It can be seen across all models that they generally perform better on the validation set rather than the test set. The key difference between the models is that the PINN models assumes an asymmetric distribution for the validation set, having more points overachieve the mean while only having outliers underperform. The CTM and BB model tend to have a more symmetric shape, where overachieving and underachieving outliers have the same spread on the validation set. As for the test set, the asymmetric trend is even more apparent for the PINN model, while the CTM model distribution is a little more distorted but generally keeps it's shape and the BB model assumes an asymmetric distribution as well.

A violin plot of the test set MAE of each respective model is shown in Figure 4.10.

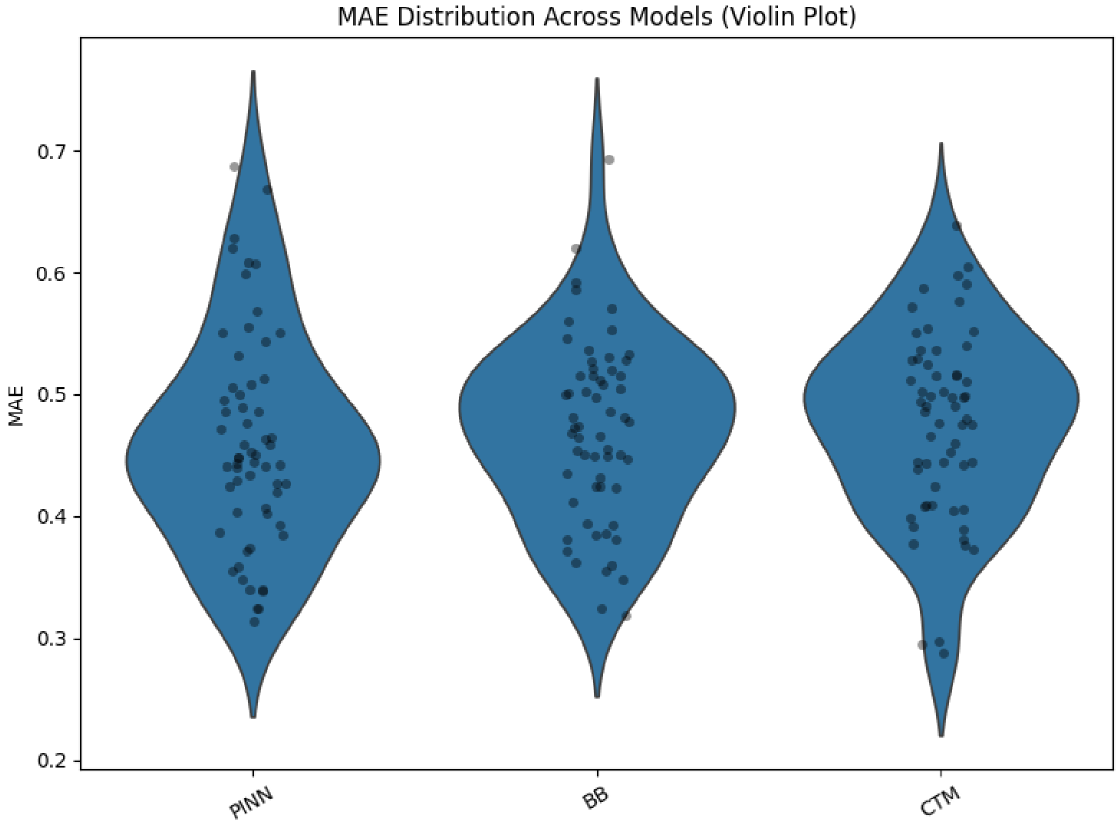


Figure 4.10: Violin plot of MAE of each network.

The violin plots of the MAE for each model, although different, tend to be more similar to each other than the R^2 violin plots are. In Figure 4.10 it can be seen that for both the CTM and BB model, points tend to cluster closer to the mean than the PINN model.

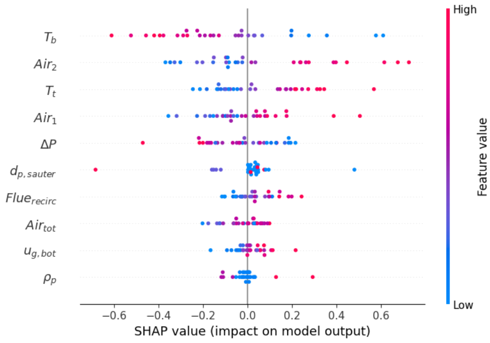
4.6 Owen values

As previously mentioned, the Owen values were evaluated for the PINN, BB and CTM models trained with random seed 0-3. The Owen values are shown in Figure 4.11, Figure 4.12 and Figure 4.13, represented with both a swarm and a bar plot. The swarm plot shows the Owen values for each individual variable and prediction in the test set with dots, colored depending on if the feature value is high or low relative to the rest of the dataset. This way, one can see how the feature value affected the final prediction. The bar plot sums up the mean absolute Owen value for each variable, showing which variables affected the prediction the most. It also shows the hierarchical clusters where the feature variables share more than 50% explaining power.

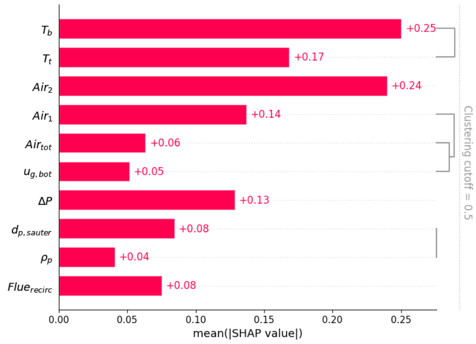
Note that the clustering varies between the seeds since it is based on the training data, which varies with the seed due to the random train-validation-test split. When interpreting the result, consider the variance in model performance as well. The performance for all networks considered is presented in Table 4.2.

Table 4.2: R^2_{test} for seeds evaluated with Owen values.

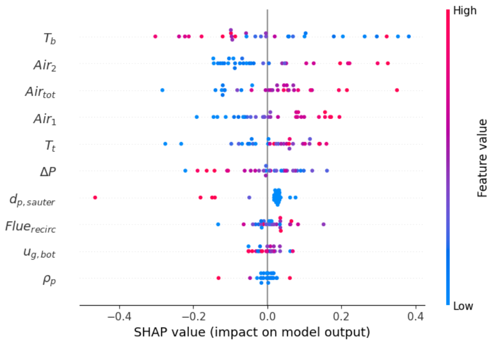
Network Type	Seed=0	Seed=1	Seed=2	Seed=3
BB	0.270	-0.052	0.218	0.175
PINN	0.467	0.126	0.639	0.319
CTM	0.374	0.072	0.443	0.465



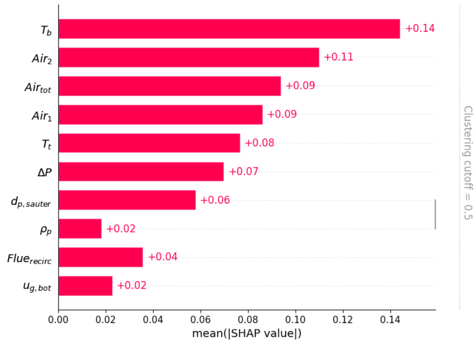
(a) Swarm plot for seed 0 BB model.



(b) Bar plot for seed 0 BB model.

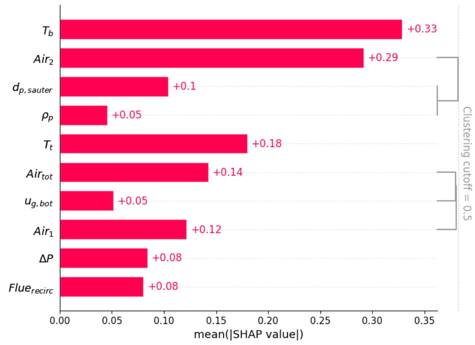


(c) Swarm plot for seed 1 BB model.

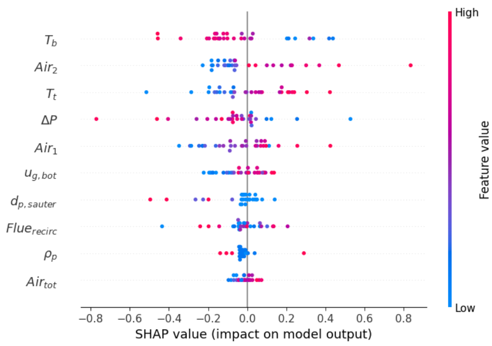


(d) Bar plot for seed 1 BB model.

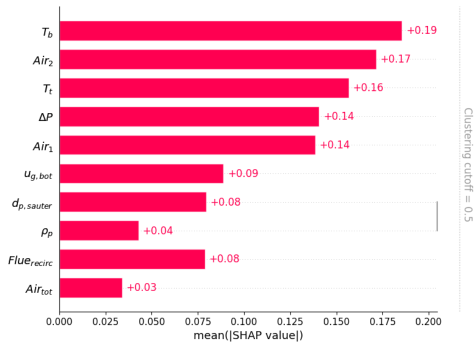
(e) Swarm plot for seed 2 BB model.



(f) Bar plot for seed 2 BB model.

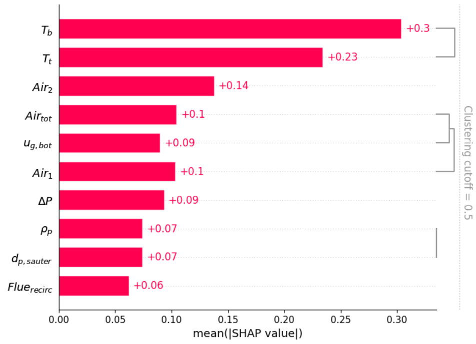


(g) Swarm plot for seed 3 BB model.

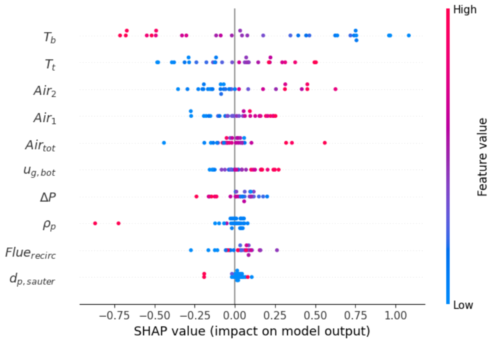


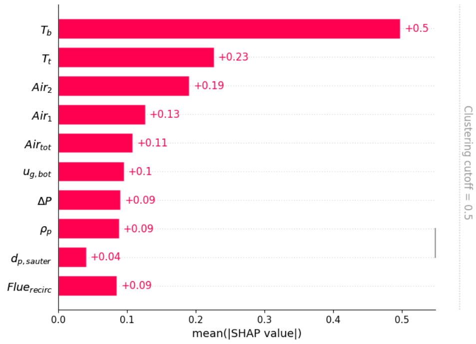
(h) Bar plot for seed 3 BB model.

Figure 4.11: Beeswarm and bar plots with Owen values for BB models.

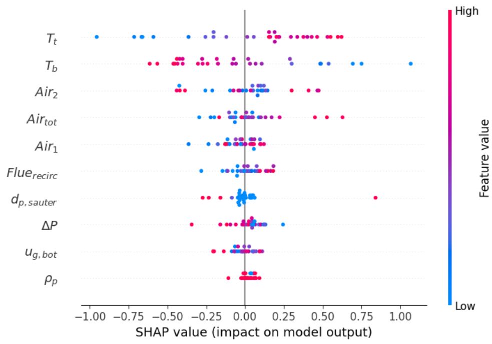


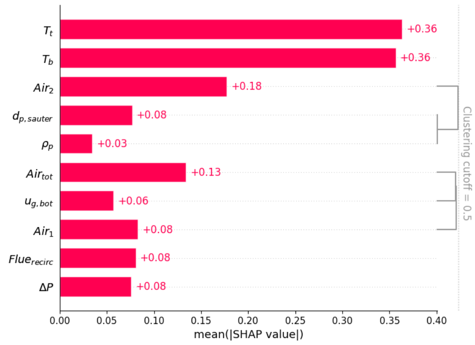
(a) Swarm plot for seed 0 PINN model. **(b)** Bar plot for seed 0 PINN model.



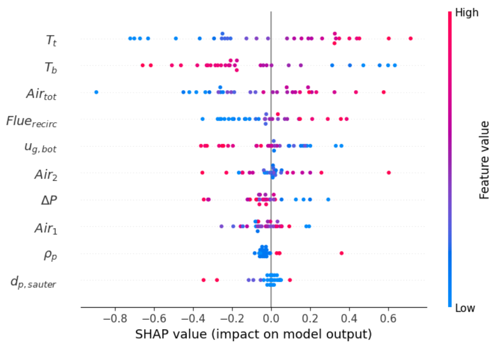


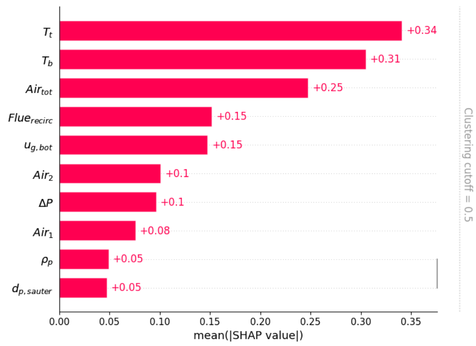
(c) Swarm plot for seed 1 PINN model. **(d)** Bar plot for seed 1 PINN model.





(e) Swarm plot for seed 2 PINN model. **(f)** Bar plot for seed 2 PINN model.

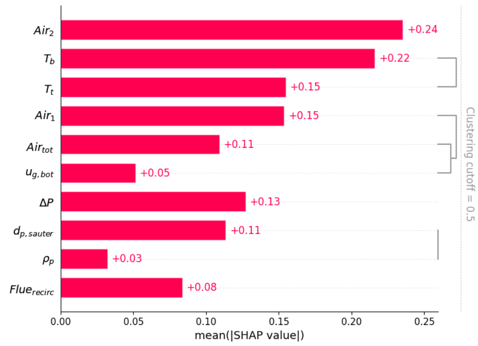




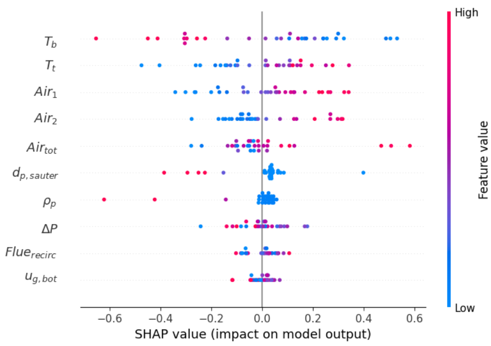
(g) Swarm plot for seed 3 PINN model. **(h)** Bar plot for seed 3 PINN model.

Figure 4.12: Beeswarm and bar plots with Owen values for PINN models.

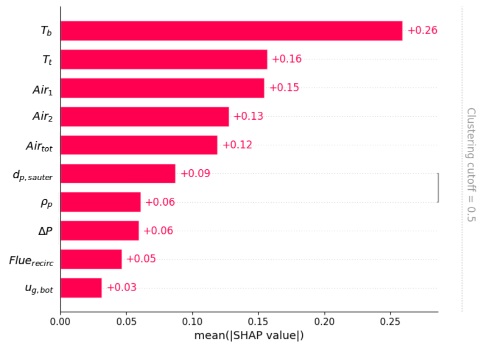
(a) Swarm plot for seed 0 CTM model.



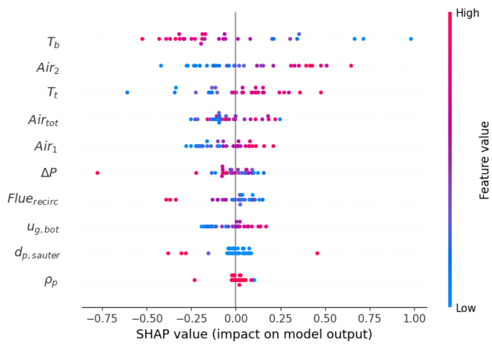
(b) Bar plot for seed 0 CTM model.



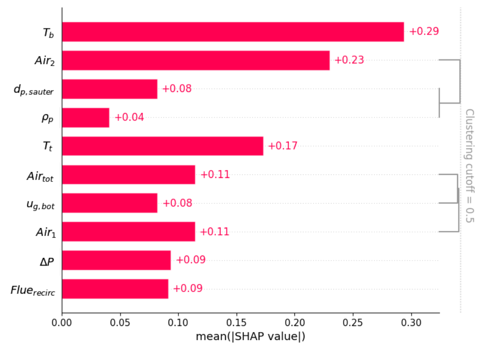
(c) Swarm plot for seed 1 CTM model.



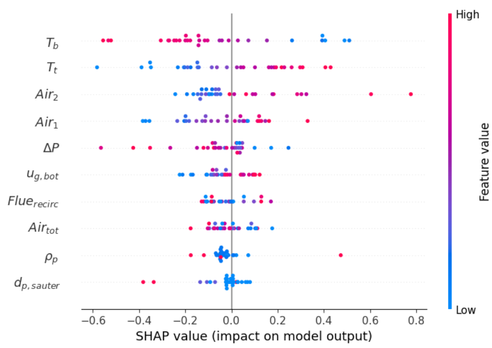
(d) Bar plot for seed 1 CTM model.



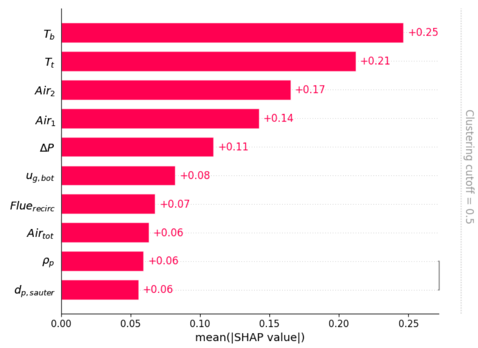
(e) Swarm plot for seed 2 CTM model.



(f) Bar plot for seed 2 CTM model.



(g) Swarm plot for seed 3 CTM. model



(h) Bar plot for seed 3 CTM model.

Figure 4.13: Beeswarm and bar plots with Owen values for CTM models.

One can note that in general, the temperature seems to be an important parameter for the BB models, with a high temperature in the top, and low temperature in the bottom leading to high external circulation. Secondary air does also seem to be an important parameter, with high secondary air leading to a high external circulation prediction. The primary air, total air, and gas velocity in the lower region, which seemed to be good predictors in the correlation matrix, do not seem to be the most important parameters. The other parameters, such as flue gas recirculation, pressure drop, sauter mean diameter and particle density are generally not important.

The PINN models show the same general trends as the black-box models. Temperatures are once again the most important parameters, with high top temperature and low bottom temperature maximizing external circulation. Secondary air is once again very important, alongside the primary air, total air and gas velocity in the lower region. Seed 3 is however interesting, since the flue recirculation is an important parameter and the air velocity in the lower region has a negative effect on external circulation, but it is only the case for one seed.

For the CTM models, the same patterns shown in the BB and PINN models are valid. Temperatures are still the most important parameters, alongside secondary air.

5

Discussion

5.1 Data analysis

5.1.1 Correlation matrices

The correlation between the gas velocity, primary air and total air as seen in Figure 4.1 is very logical. The primary air is the main constituent of the total air, which in turn is the largest gas flow making it the main contributing factor to the gas velocity. The largest correlations with the external circulation rate are also those variables, which makes sense, since it is the air which drags the particles out of the boiler. The higher the velocity of the air, the higher the solid velocity ends up since it is dependent on the gas velocity through Equation 2.25. The temperature correlation also makes sense, since both measurements are from inside the boiler, and the fluidized bed particles increase the heat transfer in the bed. In general, the matrices also generate very similar results, showing that the correlation between the variables is usually monotonic and linear.

5.1.2 PCA

The clustering of particle types in the PCA score plot in Figure 4.2a can be explained by the density and Sauter mean diameter being included as variables in the dataset. It can also in some cases be explained by the boiler being operated differently, or behaving differently. One example of this is the ilmenite cluster, which is in the direction of flue gas oxygen concentration. This indicates that the ilmenite data points has a high oxygen concentration, which could be explained by the ilmenite being an oxygen carrier. The operation of the boiler was perhaps controlled to have a higher oxygen concentration to ensure a correct oxidation state, or the oxygen carriers meant that less oxygen from the air reacted in the combustion.

The PCA loading plot in Figure 4.2b is also interesting. The fuel spreader flow and oxygen percentage in the flue gases show unique variance in the PCA-space which is problematic since they were removed as variables. Losing unique variance in the dataset could be problematic if that variance can be used to describe the external circulation. The primary gas velocity and total air also end up in the same place, showing that they in the first two principle components provide the same information.

5.1.3 PLS

In the PLS score plot in Figure 4.3a, the cluster with olivine and silver sand has been incorporated into the main cluster, and only the ilmenite cluster remains. This is probably due to the previously mentioned oxygen carrier property of the ilmenite, as well as the ilmenite being the heaviest particle. The dots in the main cluster also covers the latent space quite nicely, not leaving holes in the data which would make predictions in those zones difficult due to the lack of data. There is a hole however between the ilmenite and main cluster, which could lead to difficulties.

In the PLS rotation plot in Figure 4.3b, the flue gas oxygen concentration is now shown in the same space as the pressure drop. This indicates that despite the oxygen concentration in the flue gases having unique variance as shown by the PCA loading plot, its covariance with the external circulation rate is represented quite well by the pressure drop. The fuel spreader flow does however show unique covariance with the external circulation rate, which could be problematic. However, the loading values are low, showing a small impact on the scores. The fuel spreader flow is however a part of the total secondary air flow, meaning that it is still represented in the model. This does show a weakness in only having access to the total secondary air, since secondary air inlets exist at multiple different heights in the boiler, and the position of the flows could affect external circulation.

5.2 Merrick and Highley correlation

As seen in Figure 4.4, the empirical correlation did not perform well on the CFB boiler. This issue most likely stems from the two systems being too different. Merrick and Highely fitted their correlation to a system with a much smaller scale, assuming a uniform temperature distribution, not introducing secondary air into the system and using different particles. As all of these factors greatly impact how fluidization behaves, it is no surprise that the correlation itself did not perform well on the CFB dataset.

Merrick and Highley's correlation is also made to be used on a distribution of particle sizes rather than one particle diameter. Simplifying the system to a single particle diameter means the model will fail to capture the difference in terminal velocities of particles. Smaller particles have a lower terminal velocity as shown in Equation 2.9 and therefore get entrained and elutriate more easily. The accurate terminal velocity at the exit height is therefore smaller since the particles there are on average smaller than the total in the system.

5.3 Decay model

The decay model similarly to the Merrick and Highley correlation unfortunately has very poor performance. This can be attributed to multiple factors. Firstly, the decay model was not made to model external circulation to begin with, but rather external circulation. It was suggested to model the vertical distribution of solids inside a CFB boiler, but the phenomenon of interest in this case is the solids that are leaving the boiler. Therefore, the mass concentration at the exit height has to be multiplied with the approximated solid velocity $u_{g,bot} - u_t$ and the back-flow ratio $1 - k_b$ to account for the rate at which the particles leave the system and exit geometry effects respectively.

For the approximated solid velocity, the terminal velocity in this case is based on the sauter mean diameter, which does not consider the fact that there is a distribution of diameters for the particles used. This leaves similar problems to using the sauter mean diameter in the Merrick and Highley correlation and could explain why the decay model so often predicts low or sometimes negative external circulation. A potential solution to this would be to combine the exponential decay with a population balance, not only showing the decrease in particles over the boiler height but also the decrease in particle size over the boiler height.

The model also features two correlations based on experimental fits rather than physical equations. The value of $\rho_{s,entr}$ and k_b are both from experimental fits. Although data from the power central boiler has been used for both of the fits, there is also a lot more data used from other CFB units, meaning that the values used are not best fits for the power central boiler. Since both values are obtained from correlation, the values also lack a physical background and are sensitive to data errors in the data used to calculate the correlation as well as data outside the valid range. This is particularly important since the back-flow ratio is a very important aspect in order to predict external circulation. The back-flow ratio has also as previously mentioned been shown to depend on local particle size and concentration, which are not used in the correlation. This makes it difficult for the correlation to predict the correct back-flow ratio.

The performance of the decay model is worse than the Merrick and Highley correlation. The R^2-score is also negative, indicating that the model performs worse than just predicting the mean external circulation. The interpretation from this is that the decay model does not help understand the system, and a CTM-model with the exponential decay equation would be approximately the same as a black box. The theory was supported by results from the CTM-models that were trained with the Merrick and Highley correlation, a correlation which also had a negative R^2-score. This is shown in Figure 4.9, where those CTM-models showed very similar results to the BB-models. Therefore, with the time frame of the project in mind, the decision was made to not apply the decay model in a CTM setup, since it was assumed that it would give similar performance to a BB-model.

5.4 Model performance

The model performance from the grid search, presented in Table 4.1 show that no model ever exceeds a R^2_{val} score of 0.6245. This however, is not the case once cross validation is introduced, as the test and validation performance vary drastically depending on how the data-set split is done. The common denominator for each model is that performance on high external circulation rates is in general worse than for low external circulation rates. The data-set contains two high external circulation data points, meaning the model will have a hard time adapting to predictions of high external circulation. The drastic spread of performance over the cross validation can then be explained by the fact that the high value external circulation points either ended up in the training set or the test set. If they were in the training set, the model did not have to predict those values during the validation process and therefore performed better. If the opposite happened, it had no frame of reference for those high values and therefore performed much worse.

5.4.1 PINN performance

It can be seen in Figure 4.9 that the PINN model has an asymmetric distribution for the R^2_{val}-scores, and even more so for the R^2_{test}-scores. Unlike the CTM and BB model, the PINN model trained under the assumption that the physics incorporated into the training process were the absolute truth. These physics included such assumptions as the terminal velocity of a particle being equal to the slip velocity. The asymmetric behavior of the model reflects the fact that these assumptions can only be applied to specific cases. Therefore, the model performs a lot better when the assumption is correct, and a lot worse when it simply is not correct; leading to a very asymmetric distribution of performance points. This might also explain why the model performed worse for high external circulation values. High external circulation implies more interactions between particles and therefore terminal velocity is not the only force acting on every particle, meaning it won't be equal to the slip velocity.

This conclusion is also supported by the results from other studies where PINN's have been implemented. In Wu, Sicard and Gadsden [14] and Feng, Deng, Chaffart, Yuan and Zhu [9], generalization and performance improves drastically as a consequence of PINN being implemented. However, in their papers, the physics used are well established within the field it is applied and therefore the network benefits in a much more significant way. Fluidization physics is famously unpredictable, meaning a PINN network can benefit from it in certain cases, but also perform worse if the physics are not in line with reality [2].

5.4.2 CTM and BB performance

Given a correlation which performed worse than taking the mean of all values, the CTM model trained as if there was no physics incorporated into the model. As Merrick and Highley provided no relevant information to the neural network, it could only extract information from the data set, which means that it trained under almost the exact same conditions as the BB model. This also explains why the performance of the CTM model mirrors that of the BB model, as their performance distributions are very similar. The slight deviation between distributions comes from the fact that the CTM model had to correct the empirical correlation based on the data, while the BB model had to predict the external circulation based on the data.

5.5 Owen values

Since all three models with all four seeds give approximately the same feature importances, the results can be considered reliable. In order to get more reliable results, testing should be done with more seeds. One should also consider that the models do not have the best performance, ranging from -0.052 to 0.639. Therefore, the feature importances described by the Owen values for the different models might not be the feature importances that best describe the actual physical system, since the models generally do not describe the system very well.

5.5.1 Temperature

The big surprise from the Owen values are the large temperature values, both for the bottom and top of the boiler. This is surprising since the temperature does not have a direct effect on the particle external circulation, unlike the total air flow which is what in fact blows the particles out of the boiler. This was also reflected in the correlation matrices, where both temperatures had very low correlation with the external circulation rate for both correlation metrics, and the total air flow was the feature with the highest correlation to external circulation rate. A physical explanation to the high Owen values could be the temperature dependence on air density and viscosity. Since the air flows are given as mass flows, and the gas velocity is dependent on the volumetric flow and area, the density is needed to convert from mass flow to gas velocity. Therefore, the gas velocity is dependent on the temperature and mass flows (except for in the bottom, where the gas velocity is in the data). The density and viscosity additionally affects the Reynolds number and Archimedes number, which in turn affects the minimum fluidization velocity as well as terminal velocity and therefore solid velocity.

However, what is strange about the temperature Owen values is that the top temperature and bottom temperature affects the prediction in different directions. A high top temperature and low bottom temperature gives a high prediction. However, the top and bottom temperatures are extremely correlated, which also explains itself physically due to both temperatures being dependent on the heat generation by the furnace combustion. This leads to the effect of the temperatures in the prediction often netting a lower effect than the plots visualize, since the effects tend to cancel each other out. The temperatures could also be a description of the hydrodynamic conditions in the boiler, since a high primary air flow would lead to cold air flowing into the bottom of the boiler, lowering the bottom temperature. A high primary air flow would also lead to a higher oxygen concentration, which could increase combustion and therefore temperature if the air-fuel ratio is low. Large heat transport to the top could also be a sign of the fluidized particles being entrained into the top of the boiler, carrying heat with them, and increasing external circulation.

5.5.2 Air flows

The secondary air mass flow also has very high Owen values, with high secondary air leading to a high prediction. This is strange, since the primary air coming through the bottom of the boiler is what fluidizes and entrains the particles. The secondary air still increases the air velocity in the upper part of the boiler, but it is usually not considered when modeling entrainment.

Another interesting result are the low Owen values for primary air mass flow, total air mass flow and gas velocity in the lower region, despite these variables being the most correlated with external circulation rate by both correlation metrics. The values are also as previously mentioned the most important when representing the hydrodynamics of the boiler, since the primary air flow is what enters through the bottom, causing the particle fluidization.

One explanation for the small impact of these three variables could be that these variables are the ones that are the most correlated with each other except for the temperatures. This also makes sense when considering the variables physical meaning. The primary air is the main constituent of the total air flow, and the gas velocity in the lower region is a function of the primary air flow, recirculated flue gas and bottom temperature, with the primary air flow being the main factor. Therefore, all three variables roughly describe the same behavior, and the behavior of the boiler could potentially be described by either one of the variables. Since the variables also affect the prediction in the same direction (higher feature variable value -> higher prediction), the sum of the mean absolute Owen values for all three variables could represent the fluidization effect better. To evaluate it properly, the model would need to be retrained with only one variable, supposedly the primary air mass flow or gas velocity in primary zone. The Owen values for the primary air or gas velocity would then be an accurate description of the importance of the phenomenon.

If the sum of the mean absolute Owen values for all three variables are used, it becomes comparable in size to the bottom boiler temperature and larger than the secondary air. However, if correlated features are to be grouped, the temperatures should also be grouped which would still make temperature the most impactful feature. As previously mentioned, the models do not have the best performance, making them not the best description of the system in question. However, the clear trend of high absolute temperature Owen values shows that temperature is a potentially underestimated feature of consideration when modeling external circulation in fluidized beds.

5.6 Sustainability

PIML implementation for fluidized beds shows promise in multiple aspects of sustainability. Improved modeling makes better process control and optimization possible, which would lead to better energy efficiency and economic feasibility. In turn, this could cause fluidized beds to become a more viable alternative for both energy and material production, while having lower emissions.

The lower need for data compared to traditional machine learning is also important, since this allows a reduction in experiments required to obtain data. Experiments are usually quite expensive for large scale process units such as fluidized beds, and reducing the data need therefore improves the economic feasibility.

The implementation of physical equations into machine learning, as well as the use of explainable AI also leads to a more responsible use of data. It gives strong insight into how a model operates, compared to a BB model which as its name suggests is very hard or impossible to interpret. The insight into the model enables critical analysis use of data, leading to a more ethical decision making.

6
Conclusion

Fluidized beds are still hard to model, both due to lack of data, uncertainty of data and an absence of physical governing equations. However, the project shows that PIML is a feasible modeling method for fluidized beds with promising improvement in performance compared to traditional machine learning. The pre and post data analysis supports previous hypotheses regarding which factors are relevant for external circulation such as primary air and secondary air, while also showing that temperature might be a more important factor than previously thought. This should however be researched further, since the Owen values were uncertain and disagreed with the correlation matrix, PCA and PLS.

The grey-box models trained in this project performs slightly better than current mechanistic models and show promise for further studies within the field. Future projects should aim to improve upon the previously mentioned lack of data. In section 5.2, it is shown that data from different cross-sectional areas and exit configurations do not translate well over systems. In order to improve upon this, data should be obtained from systems with different operating parameters, specifically geometric differences. These include, for example, position of secondary air flow and exit configuration. Investments could also be made into better measurement units to improve the quality of the data by reducing noise. Lastly, efforts should be put towards steady state measurements of external circulation. This allows for the use of the many methodologies used in scientific machine learning which utilize time derivatives, as well as the continuous collection of data which would make the data set a lot larger.

Implementation of more advanced physical models could also reduce the need for data, as a better physical representation could describe a large part of the system by itself. For example, taking into consideration the particle size distribution and the separation of particle sizes along the height through a population balance.

For future research projects, the population balance should be considered, as well as other methods which consider the entire particle size distribution and the difference in terminal velocities. The previously mentioned issues with data uncertainty and scarcity also warrants a more extensive search for data in future projects, as well as research into how physics in CFB boilers can be better measured to obtain more reliable data. Additionally, further investigation into variable importance should be done in order to confirm or deny the strong external circulation dependence on boiler temperature, as well as the relatively weak dependence on gas flows.

Bibliography

[1] D. Kunii and O. Levenspiel, *Fluidization Engineering*. Elsevier Inc, 1991, pp. 64–81, 200–201, 206.

[2] J. W. Chew, A. Cahyadi, C. M. Hrenya, R. Karri, and R. A. Cocco, "Review of entrainment correlations in gas–solid fluidization," *Elsevier*, vol. 260, pp. 152–170, 2015.

[3] T. Djerf, D. Pallarès, and F. Johnsson, "Bottom-bed fluid dynamics – influence on solids entrainment," *Fuel Processing Technology*, vol. 173, pp. 112–118, 2018.

[4] F. J. David Pallarès Tove Djerf, "Solids flow patterns in large-scale circulating fluidised bed boilers: Experimental evaluation under fluid-dynamically downscaled conditions," *Elsevier*, vol. 231, 2021.

[5] *A historical perspective*, Accessed: 2026-03-04 15:07, 2022. [Online]. Available: `https://www.chalmers.se/en/infrastructure/cpc/a-historical-perspective/`.

[6] C. Rackauckas et al., *Universal differential equations for scientific machine learning*, 2021. arXiv: `2001.04385 [cs.LG]`. [Online]. Available: `https://arxiv.org/abs/2001.04385`.

[7] S. S. Garakani and J. W. Chew, "Development of physics-informed machine-learning models to enhance understanding and prediction of membrane fouling," *Journal of Membrane Science*, vol. 728, p. 124 133, 2025, ISSN: 0376-7388. DOI: `https://doi.org/10.1016/j.memsci.2025.124133`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0376738825004466`.

[8] A. Mukherjee and V. M. Zavala, "Physics-constrained machine learning for chemical engineering," *Current Opinion in Chemical Engineering*, vol. 51, p. 101 228, 2026, ISSN: 2211-3398. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S2211339826000031`.

[9] L. Feng, Z. Deng, D. Chaffart, Y. Yuan, and J. Zhu, "A physics-informed neural network (pinn)-based model investigation for hydrodynamics and conversion in a gas-solid fluidized bed reactor," *Powder Technology*, vol. 471, p. 122 037, 2026, ISSN: 0032-5910. DOI: `https://doi.org/10.1016/j.powtec.2025.122037`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0032591025014329`.

[10] A. Mondal and S. L. Sharma, "Physics-informed machine learning-aided framework for entrainment fraction in gas-liquid two-phase annular flow," *Applied Thermal Engineering*, vol. 296, p. 130 778, 2026, ISSN: 1359-4311. DOI: `https://doi.org/10.1016/j.applthermaleng.2026.130778`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S1359431126010860`.

[11] H. Mohammadghasemi, J. F. Soesanto, A. Singh, B. Maciszewski, and B. Huang, "Scientific machine learning for modeling industrial-scale primary separation vessel," *Computers Chemical Engineering*, vol. 205, p. 109 427, 2026, ISSN: 0098-1354. DOI: `https://doi.org/10.1016/j.compchemeng.2025.109427`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0098135425004302`.

[12] S. Deng, L. Wang, S. Kim, and B. C. Koenig, "Scientific machine learning in combustion for discovery, simulation, and control," *Proceedings of the Combustion Institute*, vol. 41, p. 105 796, 2025, ISSN: 1540-7489. DOI: `https://doi.org/10.1016/j.proci.2025.105796`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S1540748925000100`.

[13] J.-P. Li et al., "Machine learning-driven lipase engineering for highly enantioselective synthesis of chiral florfenicol intermediate via enzymatic resolution-chemical racemization coupling system," *Chemical Engineering Journal*, vol. 536, p. 176 120, 2026, ISSN: 1385-8947. DOI: `https://doi.org/10.1016/j.cej.2026.176120`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S1385894726035813`.

[14] Y. Wu, B. Sicard, and S. A. Gadsden, "Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring," *Expert Systems with Applications*, vol. 255, p. 124 678, 2024, ISSN: 0957-4174. DOI: `https://doi.org/10.1016/j.eswa.2024.124678`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0957417424015458`.

[15] A. Johansson, F. Johnsson, and B. Leckner, "Solids back-mixing in cfb boilers," *Chemical Engineering Science*, vol. 62, pp. 561–573, 2007.

[16] P. Basu and S. A. Fraser, *Circulating Fluidized Bed Boiler*. Elsevier Inc, 1991, pp. 17–48, 344–346.

[17] F. M. White, *Viscous fluid flow*. McGraw-Hill Higher Education, 2006, p. 16.

[18] D. Merrick and J. Highley, "Particle size reduction and elutration in a fluidized bed process," *AIChe Symp. Ser.*, pp. 366–378, 1974.

[19] F. Johnsson and B. Leckner, "Vertical distribution of solids in a cfb-furnace," in *Proceedings of the 13th international conference on fluidized bed combustion*, Orlando, 1995.

[20] T. Djerf, D. Pallarès, and F. Johnsson, "Solids backmixing and entrainment in the splash zone of large-scale fluidized bed boilers," *Powder Technology*, vol. 404, 2022.

[21] T. Djerf, D. Pallarès, and F. Johnsson, "Solids flow patterns in large-scale circulating fluidised bed boilers: Experimental evaluation under fluid-dynamically down-scaled conditions," *Chemical Engineering Science*, 2021. DOI: https://doi.org/10.1016/j.powtec.2022.117471.

[22] Z. Luca, *Extending Power BI with Python and R (2nd Edition). 2024.* Packt Publishing, 2024, ISBN: 978-1-83763-953-3.

[23] A. Subasi, *Practical Machine Learning for Data Analysis Using Python.* Elsevier, 2020. DOI: 10.1016/C2019-0-03019-1.

[24] L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold, *Multi- and Megavariate Data Analysis: Principles and Applications.* Umeå: Umetrics Academy, 2001, ISBN: 91-973730-1-X.

[25] S. Wold, M. Sjöström, and L. Eriksson, "Pls-regression: A basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, 2001. DOI: 10.1016/S0169-7439(01)00155-1.

[26] B. Mehlig, *Machine Learning with Neural Networks: An Introduction for Scientists and Engineers.* Cambridge University Press, 2021, pp. 1–12, ISBN: 9781108494939. DOI: 10.1017/9781108860604. [Online]. Available: http://dx.doi.org/10.1017/9781108860604.

[27] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991, ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(91)90009-T. [Online]. Available: https://www.sciencedirect.com/science/article/pii/089360809190009T.

[28] G. Dreyfus, *Neural Networks: An Overview.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1–83. DOI: 10.1007/3-540-28847-3_1. [Online]. Available: https://doi.org/10.1007/3-540-28847-3_1.

[29] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: 1412.6980 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1412.6980.

[30] I. Loshchilov and F. Hutter, *Decoupled weight decay regularization*, 2019. arXiv: 1711.05101 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1711.05101.

[31] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, *Improving neural networks by preventing co-adaptation of feature detectors*, 2012. arXiv: 1207.0580 [cs.NE]. [Online]. Available: https://arxiv.org/abs/1207.0580.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, vol. abs/1502.01852, 2015. arXiv: 1502.01852. [Online]. Available: http://arxiv.org/abs/1502.01852.

[33] K. G. K. I.G, "Physics-informed machine learning," *Nat Rev Phys*, pp. 366–378, 2021.

[34] C. Molnar, *Interpretable Machine Learning, A Guide for Making Black Box Models Explainable*, 3rd ed. 2025, ISBN: 978-3-911578-03-5. [Online]. Available: https://christophm.github.io/interpretable-ml-book.

[35] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[36] J. M. Giménez and M. A. Puente, "The owen and the owen-banzhaf values applied to the study of the madrid assembly and the andalusian parliament in legislature 2015-2019," in *In Proceedings of the 8th International Conference on Operations Research and Enterprise Systems-Volume 1*, Prague, Czech Republic: Scitepress, 2019, pp. 45–52.

[37] B. Leckner, "Regimes of large-scale fluidized beds for solid fuel conversion," *Elsevier*, vol. 308, pp. 362–367, 2017.

A
Appendix 1

Code is made available on request.

DEPARTMENT OF CHEMISTRY AND CHEMICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se