Advanced analytics for process analysis of turbine plant and components

Prepared by:

Yashveer Maharajh

MHRYAS018

Department of Mechanical Engineering

University of Cape Town

Supervisor:

Prof. Pieter Rousseau

Co-supervisor:

A/Prof Amit Mishra

July 2019

Submitted to the Department of Mechanical Engineering at the University of Cape Town in partial fulfilment of the academic requirements for a Masters of Science degree in Mechanical Engineering

Key Words: Thermofluid process modelling, FLOWNEX® SE, feed water heater, machine learning, deep learning, artificial neural networks, multi-layer perceptron, ReLU, Adam optimisation, regularization, data augmentation, condition monitoring, predictive analytics

Abstract

This research investigates the use of an alternate means of modelling the performance of a train of feed water heaters in a steam cycle power plant, using machine learning. The goal of this study was to use a simple artificial neural network (ANN) to predict the behaviour of the plant system, specifically the inlet bled steam (BS) mass flow rate and the outlet water temperature of each feedwater heater. The output of the model was validated through the use of a thermofluid engineering model built for the same plant. Another goal was to assess the ability of both the thermofluid model and ANN model to predict plant behaviour under out of normal operating circumstances.

The thermofluid engineering model was built on FLOWNEX® SE using existing custom components for the various heat exchangers. The model was then tuned to current plant conditions by catering for plant degradation and maintenance effects. The artificial neural network was of a multi-layer perceptron (MLP) type, using the rectified linear unit (ReLU) activation function, mean squared error (MSE) loss function and adaptive moments (Adam) optimiser. It was constructed using Python programming language.

The ANN model was trained using the same data as the FLOWNEX® SE model. Multiple architectures were tested resulting in the optimum model having two layers, 200 nodes or neurons in each layer with a batch size of 500, running over 100 epochs. This configuration attained a training accuracy of 0.9975 and validation accuracy of 0.9975. When used on a test set and to predict plant performance, it achieved a MSE of 0.23 and 0.45 respectively. Under normal operating conditions (six cases tested) the ANN model performed better than the FLOWNEX® SE model when compared to actual plant behaviour. Under out of normal conditions (four cases tested), the FLOWNEX SE® model performed better than the ANN.

It is evident that the ANN model was unable to capture the “physics” of a heat exchanger or the feed heating process as a result of its poor performance in the out of normal scenarios. Further tuning by way of alternate activation functions and regularisation techniques had little effect on the ANN model performance. The ANN model was able to accurately predict an out of normal case only when it was trained to do so. This was achieved by augmenting the original training data with the inputs and results from the FLOWNEX SE® model for the same case.

The conclusion drawn from this study is that this type of simple ANN model is able to predict plant performance so long as it is trained for it. The validity of the prediction is highly dependent on the integrity of the training data. Operating outside the range which the model was trained for will result in inaccurate predictions. It is recommended that out of normal scenarios commonly experienced by the plant be synthesised by engineering modelling tools like FLOWNEX® SE to augment the historic plant data. This provides a wider spectrum of training data enabling more generalised and accurate predictions from the ANN model.

Declaration

I, Yashveer Maharajh, hereby declare the work contained in this dissertation to be my own. All information which has been gained from various journal articles, text books or other sources has been referenced accordingly. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as their own work or part thereof.

Signed by candidate

Yashveer Maharajh

Acknowledgements

Above all else I thank God for blessing me with the opportunity, strength and ability to complete this dissertation. I thank my parents Kishore and Reena for instilling in me the value of education and the drive to set and surpass my goals. I thank my wife Ashalia for her strength, support and love through this process. I thank her for keeping me grounded, giving me perspective and helping me stay present in life and through this experience. I thank my daughters Azaya and Maya who teach me on daily basis the meaning of joy and living in the moment. Thank you for helping me embody the words of Mahatma Ghandi “live as if you were to die tomorrow. Learn as if you were to live forever”.

I thank Eskom for the opportunity to further my studies under the EPPEI Program, my manager Jason Hector for the support and time to complete this degree. I thank my industrial mentor Pravin Moodley for his willingness to assist whenever needed and for his advice and perspective, as well as my mentor James L’Etang for his consistent interest and support.

Special thanks go to my academic supervisor Professor Pieter Rousseau for his undying support and encouragement through this process. His ability to provide clear and practical direction has had an enormous impact on this project. I thank my co-supervisor, Associate Professor Amit Mishra, for sharing his expertise and vast knowledge in the field of machine learning and for his guidance through uncharted territory. I also thank Associate Professor Wim Fuls who has never refused to assist and provide insight whenever needed.

Table of Contents

List of Figures

List of Tables

List of Nomenclature

General symbols

$\dot{m}$	Mass flow rate
e	Error function
g	Gradient
h	Enthalpy
i	Neuron layer in current layer
j	Neuron reference in next layer
l	Layer
n	Number of neurons in layer l
Q	Heat Energy
s	Sum of the products of weight and input signals
s^t	Exponential average of squares of gradients
T	Temperature
UA	Overall heat transfer coefficient with area
v^t	Exponential average of gradients
w	Weight variable
x	Input signal to neuron
y	Target value or label

Greek symbols

θ	Output of activation function
η	Learning rate
B_1	Adam optimiser hyperparameter
B_2	Adam optimiser hyperparameter
ε	Mathematical constant

Acronyms and Abbreviations

Adam	Adaptive Moments
AI	Artificial Intelligence
ANN	Artificial Neural Network
DA	De-aerator
DCA	Drains Cooler Approach
FFN	Feed Forward Network
FWH	Feed Water Heater
HP	High Pressure
HPH	High Pressure Heater
IIoT	Industrial Internet of Things
IT	Information Technology
kPa	Kilo Pascal
LP	Low Pressure
LPH	Low Pressure Heater
ML	Machine Learning
MLP	Multi-layer Perceptron
MPa	Mega Pascal
MSE	Mean Squared Error
ORC	Organic Rankine Cycle
OT	Operational Technology
P	Pressure
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SC	Sub-cooling
SVM	Support Vector Machine
TTD	Terminal Temperature Difference

1. Introduction

1.1 Background and motivation for the study

Data science and specifically machine learning has become a widely used concept in many spheres of the business, banking and industrial sector. It adds value as a result of the predictive capabilities it offers. In the current age, we generate vast amounts of data but the human brain is limited in that it is capable of analyzing a relatively small number of variables simultaneously. Advances in computer science, specifically in artificial intelligence (AI), as well as hardware to provide the computational processing capability required by AI, may be the solution to handling and analyzing these large multivariable data sets. This project attempts to leverage the advances in the computer science and technology sphere as a tool in the engineering space, specifically in thermal power plants. It serves to assess the ability, practicality and limitations in implementation of such a tool.

The concept of machine learning, a technique used in AI, was initially attempted by the author and a team of engineers on a power station through a project whereby the condenser back pressure profile was predicted with an input of the hourly ambient temperature forecast for a day. This proved to be a success as the profile was predicted with sufficient accuracy.

The objective of this project is to apply machine learning principles to a more complex plant area namely a train of feed water heaters that are in both series and parallel configurations and are comprised of both open and closed heaters. This will promote the use of machine learning in this highly technical environment, and allow for the validation of the machine learning model output through comparison with a physics-based thermofluid process model of the same plant area.

Such applications are becoming more significant especially in the industrial sector with the advent of Industry 4.0. This is a term used to describe the current shift of industries towards a digitized center of operation which sees Information Technology (IT) and Operational Technology (OT) platforms converging. This is more commonly known as the Industrial Internet of Things (IIoT).

Eskom is currently embarking on such a transformation, which provides a unique opportunity to leverage the available information and raw plant data. Using machine learning techniques with the available data, it is deemed possible to achieve improvements in plant process efficiency, equipment performance, production capacity and reduction in unplanned plant maintenance or downtime. This becomes even more important due to the age of the current asset base. One key differentiator that sets machine learning apart is the ability to apply the models and generate

these predictive outputs in real time. This sets the stage for potentially increasing revenue growth of Eskom through improved operational efficiency.

1.2 Problem description of the study

Machine learning, while a widely used concept throughout the world is only starting to be used in the industrial sector, including power plants, to predict plant process performance. Simplistically, the concept works through providing the model with historic input data which "trains" the model. The model then "learns" the statistical relationships of these parameters. Finally, it predicts an output based on an input of the same structure as the training data set.

Fundamentally, using the training data, which provides hindsight, allows the model to provide some foresight though its prediction. However, without sound knowledge and application of engineering and plant performance principles, that is, without proper insight, the predictions could be incorrect. The risk is that this could be used to make erroneous decisions regarding plant operations and production. This project serves to mitigate this risk through proper validation using sound engineering principles.

In this project the feed water heating train will be modelled using machine learning methods. The various methods and algorithms associated with the construction of such a model will be more defined after a literature study on the topic. The results of the machine learning model will then be validated via a thermofluid process model for the same plant. This would ensure that the insight that is provided is correct and that the output of the machine learning model correlates with that of the thermofluid process model.

The machine learning model itself can be one or a combination of many machine learning methodologies i.e. regression, classification, clustering, or deep learning constituting artificial neural networks. This is an aspect that will be addressed as part of the literature study.

1.3 Goals and objectives

The project primarily serves to answer the question: Can machine learning methods be used in an environment such as a power station, to adequately predict plant performance?

- In the process of answering this, a tuned thermofluid process model of the feed water heating system will be constructed.
- Should the use of machine learning methods prove to be successful, a working machine learning model for a feed water heating train will exist.

- This would then establish a new capability and methodology of predicting plant behaviour which assists where first principle models are not possible or have limitations.
- The capability of both the machine learning and thermofluid process model will be tested in the context of out of normal operating conditions. This allows the investigation of the models' ability to extrapolate beyond normal operation and highlight potential limitations. Some scenarios that will be tested are:
    - placing heaters out of service,
    - simulating tube fouling, and
    - operation at a lower load than the model was trained for.

Further to this, if successful, it will inadvertently demonstrate that free opensource software can be used to model and accurately predict plant performance. This highlights the ability to overcome issues such as

- the use of expensive engineering software,
- modelling complexity and potential inaccuracies using engineering first principles methods,
- limitations of software capability.

To accomplish the above-mentioned objectives, the following activities were executed:

- A literature survey was conducted to provide a better understanding of the breadth of use of machine learning, particularly in the industrial sector, and specifically to predict plant performance. This exercise served to critically analyze the method of building the machine learning architecture, methods of assessing model accuracy and approaches to validation of the model.
- A thermofluid process model was built on FLOWNEX SE to mimic the performance of an integrated train of feed water heaters of a particular power station.
- Real plant data was used to calibrate and validate the thermofluid model such that it is representative of the real plant operation. A manual mass and energy balance calculation was conducted to verify the thermofluid model output.
- A suitable machine learning (ML) architecture was selected and a model created with which to predict the plant performance.
- Both the ML model and the thermofluid model were used to predict the full set of output data and compared using a common training data set.

- A parametric study was conducted to understand the effect of changing certain parameters in order to understand the benefits and limitations of applying the ML model rather than a fundamental thermofluid model in an industrial setting. This included testing both models outside the normal operating parameters of the plant.

2. Literature review

In this section, a critical investigation into the research and applications of machine learning, particularly artificial neural networks (ANN), will be presented. The focus of research and applications was narrowed to heavy industry and to the power generation sector and then specifically to heat exchangers. This study was intended to be broad in the sense that both the type of application and the modelling approach were considered. This was done in order to better understand the methodologies applied and the quality of the subsequent results obtained.

In the engineering world, the behaviour of real systems is described using algebraic and differential equations. Mathematical models are developed to represent such systems. However, this requires specific knowledge of the system dynamics, estimation techniques and numerical calculations to emulate such systems. This can easily introduce uncertainties which make this type of modelling unrealistic or inaccurate. Depending on the type of system or application, like computational fluid dynamics for example, the computational power required to achieve the desired result can be prohibitive. Artificial neural networks are increasingly being used to overcome such limitations.

For these reasons, as stated by Meireles et al. (2003), a review of the industrial applicability of artificial neural networks was conducted. Particular emphasis was placed on highlighting differences in methodologies and architecture for different applications. From a model training perspective, careful consideration is to be given to the selection of input and output variables, the size of the training data set, initialising the weights, learning rates and stopping criteria. There are no rules or methods to accurately determine this, therefore it is largely trial and error. The same applies to the network design, when considering the number of hidden layers and neurons per layer. General practice suggests methods such as increasing the number of hidden layers to increase the model performance. Similarly, maintaining three hidden layers and increasing the number of neurons per layer achieves the same goal. However, increasing the number of neurons increases the model complexity. The effect of this is an increase in computational power and solution time required (Gullil & Pal, 2017).

Training or model learning can be supervised or unsupervised. The most successful applications for predicting plant performance and plant control have used supervised learning. Unsupervised learning is unsuitable due to slow adaptation of the model and the required time it takes for the network to settle into stable conditions. It is used more for pattern recognition type of problems.

The most widely used algorithm for training is the backpropagation algorithm. It has been criticized for being slow, especially if many hidden layers are used, and of "losing its memory" in that it forgets the old when something new is learned. However, it is still the most broadly used.

Neural networks are a preferred modelling method compared to multiple regression methods. It does not require the definition of the most important independent variables since the network automatically adjust the weight (relevance) of each variable. Furthermore, with specific importance to industrial processes, the learning capability of neural networks allows more complex and subtle interactions between the independent (input) variables to be discovered. They are more robust and show more immunity to noisy data allowing for greater precision in its prediction. A model is deemed to have good estimation performance when it achieves more than 95% accuracy in overall data recalls (Chow, et al., 1993).

Tumer et al. (2015) investigated the use of various ANN architectures to predict the performance of a wastewater treatment plant. The plant is made up of five sub-processes or treatment units that occur in sequence in order to treat the water to an acceptable quality. One neural network was constructed for the entire plant using six input variables to predict a single output variable. The input variables selected were largely performance metrics associated with the sub-processes with the exception of the water flow rate. The training data was obtained directly from plant measurements. It was taken over a four-month period so as to cover any seasonal variation of the plant performance. The neural network itself was constructed to test multiple architectures (nine in total) to ascertain the most suitable one. The parameters varied where the number of hidden layers, number of neurons per layer and the activation function used at each layer. Using the minimum mean squared error and maximum correlation coefficient as performance measures, it was found that the architecture with three neurons and one hidden layer provided the best results with an R value of 0.96.

Nasr et al. (2010) also developed an ANN model for a waste water treatment plant in order to optimise control of the plant through prediction of its performance. Like the method employed by Tumer et al. (2015), a feed forward back propagation based training algorithm was used. Here, however, three input variables with three hidden layers were used to predict three output variables. Each hidden layer had 10, 30 and 0 neurons respectively. Their model resulted in an R value of 0.903, which was deemed satisfactory.

What the above research highlights is that water treatment plants are complex processes. Traditional methods to predict performance would require balanced chemical equations together with chemical kinetics or rate equations for each sub process. This is considerably complicated since some reactions are nonlinear and time variable when coupled with environmental interactions. Traditional methods have thus shown some limitations (Nasr, et al., 2010). Artificial

neural networks have been proven as an alternative method to simulate and accurately predict the behaviour of complex, non-linear systems.

As a result of its ability to handle complex, nonlinear problems, artificial neural networks have been used widely in industry. Buratti et al. (2016) used neural networks to predict the energy efficiency of a building due to the lack of data needed for the methodology required by regulatory authorities. A two-layer network was used with a sigmoidal activation function. The number of neurons where determined through a sensitivity study while being cognisant of the issue of overfitting when a large number of neurons are used. The results of the model showed an R value of greater than 0.95. The neural network model was also compared to a 3D simulation model created on TRNSYS. It was found that the difference between the neural network and the simulation was 6%, which is below the acceptable error of 10% as dictated by regulations.

Basile et al. (2015) tested the use of neural networks to model the carbon monoxide (CO) conversion and hydrogen (H_2) production in a membrane water gas shift reactor and compared this to experimental results. Using a feedforward back propagation model with three layers, one hidden layer with 20 nodes and sigmoid activation function, it was found that the error for CO conversion was less than 0.5% and that for H_2 production approximately 10%.

Neural networks are also becoming an essential tool in the renewable energy industry. Rodríguez et al. (2018) was able to predict the output of a solar photovoltaic plant using highly variable weather forecast data as input. The model differed from 0.5 – 9% when compared to the actual production, low enough to be used as predictive tool in integrated energy systems that incorporate solar energy generation.

In addition to the incorporation of variable energy sources into a network, Haque and Kashtiban (2007) investigated the use of neural networks in fault diagnostics, load forecasting, economic dispatch and harmonics analysis in power systems. The study concluded that besides the positive aspects of ANN mentioned previously, the drawback is that it relies on simulation models for the training data before it can be applied to a real system. It should be noted that the use of simulation software for generating training data is a common practice. This is seen more especially in cases where ANNs are built for process control and performance prediction on plant that has not been fully constructed or commissioned.

Mohatram et al. (2011) performed a similar review. While the method was found to be successful, some of the areas of concern where noted. The time taken as a result of the trial and error nature of determining the network architecture and configuration was found to be long and open ended. There is no formal method or modelling theory hence it relies on expert knowledge in the design of the neural network to achieve acceptable results.

In the fossil based power generating industry, ANNs are being used in a variety of ways. Fast and Palme (2010) created an online condition monitoring and diagnostic system for a combined heat and power (CHP) plant using neural networks to improve plant profitability. Due to the large and complex system at hand, the gas turbine, heat recovery steam generator, boiler and steam turbine were modelled separately, that is an ANN for each plant area. Data pre-processing was deemed necessary to remove outliers. Transient operation was removed by using five-minute averaged data i.e steady state operation was modelled. It was concluded that the ANN method for modelling plant components achieves high accuracies. The model itself is plant specific, however, the method is general and applicable to other power plants and plant configurations.

Strušnik et al. (2015) performed a similar exercise with the intention of optimising the steam production between a set of boilers, which differ in capacity. The aim was to optimise the efficiency of a CHP plant. This was achieved by breaking up the problem into two ANNs, one for the steam required for power production and the other for the production of steam needed for heating.

Chokshi et al. (2018) conducted a performance analysis on a coal fired power plant. Empirical data, specifically process related variables, were used as training data to 65 ANNs of differing architecture and loss functions. It was found that a specific architecture would yield an acceptable error level when a particular loss function was used. However, this was not the case when a different loss function was used for the same architecture. It is therefore important to properly establish the method of model performance evaluation.

Looking specifically at the performance of a coal fired boiler, Smrekar et al. (2009) was able to predict the performance using ANNs. It was found that this method of prediction becomes superior to physical simulation models when plant degradation is considered. Simulation models tend to lose their accuracy as plant deteriorates unless such factors are specifically catered for. This limitation is overcome by simply using updated plant data as a training set to retrain the model.

While trying to achieve the same goal, Suresh et al. (2011) attempted to predict performance and optimise operating parameters on a supercritical coal fired plant that was not yet built. Interestingly, a simulation program was used to create the training data needed for the ANN to be able to predict the energy required from coal. The output of the ANN was then used in a genetic algorithm to optimise the individual plant parameters. This proves that training data generated by simulation software can be successfully used to predict plant performance in an ANN. This is especially useful if out of normal operating data from the plant does not exist. This approach enables the out of normal scenario data to be generated using simulation software and augment the training data set to the ANN model.

The performance prediction of an organic Rankine cycle (ORC) plant was carried out using two different machine learning methodologies by Dong et al. (2018). Support Vector Machine (SVM) using both Gauss Radial Bias kernel function and linear function was compared to an ANN using back propagation. Ultimately it was found that ANN and SVM linear function where the most consistent in terms of its predictive capability.

Rashidi et al. (2011) used a novel combination of machine learning methods to optimise an ORC cycle. It took the form of two parts; firstly three ANNs were trained to produce outputs of specific nett work, thermal efficiency and exergy efficiency. The training data to these networks were generated via a thermodynamic cycle analysis using Engineering Equation Solver. The second part, the optimisation, used a concept called Artificial Bees Colony which uses the output of the ANNs as fitness functions to be optimised. Conducting a parametric analysis within a predefined range allowed for the optimum process conditions to be estimated.

ANNs are also being used in the control systems of coal fired power stations as outlined by Mikulandric et al. (2013). The results of the study suggest that the objective was achieved through the use of neural network based control methods to optimise combustion, reduce nitrogen oxide production and increase plant efficiency. Combustion optimisation was found to be the most critical parameter for overall plant improvement.

Luo et al. (2018) applied ANNs in order to predict the product distribution after coal devolatilisation, a pivotal step in the coal combustion process. It was found to predict the distribution well when compared to experimental data and provide a better prediction of yield of each component compared to experimental and theoretical based models. From a bulk materials handling perspective, Bekat et al. (2012) successfully predicted the bottom ash formed in a coal power plant using ANNs by using coal quality related information only. Even though operational variables such as mill performance, mill configuration and combustion efficiency were neglected as input variables, high accuracy predictions were still achieved thus highlighting the power and flexibility of ANNs.

Another typical application in the fossil based power generation industry is the prediction of gaseous emissions. Tunckaya and Koklukaya (2015) performed this study by using post combustion flue gas process parameters, which include temperature, flow rate, velocity, particulate and oxygen content, as input to the model. Interestingly, neither coal, combustion parameters nor boiler load was used as input which one would assume to be critical to the prediction. The predictive performance of the ANN was found to be superior to the autoregressive integrated moving average method and the multiple linear regression method based on a comparison of R values. The ANN's average error was 0.6% (vs actual). Even though the plant experienced some out of normal operating conditions in the boiler on a handful of occasions, the NO_x prediction still

managed to “react” to this abnormal occurrence with an error of 9.5%. This highlights the ability of the ANN to adapt to an out of normal condition that did not exist in the training set, but not the same level of accuracy when compared to normal conditions for which it was trained. The authors used a similar methodology to successfully predict the production rate of a coal fired power plant (Tunckaya & Koklukaya, 2015).

From a boiler design and heat transfer perspective, estimating tube temperatures of the water wall tubes is critical for an economically feasible design and stable operations. Dhanuskodi et al. (2015) found that analytical methods of prediction under the conditions experienced by these tubes are unreliable. Of particular concern, where the empirical correlations of non-dimensional numbers that were used to predict wall temperatures of turbulent flow in the phase transition region of the boiler i.e. the water walls. ANNs were used for the same prediction and was found to have an accuracy of 81.94% when compared to experimental data with a deviation of ±7° C. This was based on the optimum ANN model selected out of 127 models of differing characteristics. The performance of the analytical method is unclear as no evidence of its predictive accuracy was presented. It is therefore difficult to gauge the performance of the ANN against the analytical method.

Heat exchangers are widely used in engineering applications and process plants such as refrigeration systems, air conditioning, chemical and petrochemical plants. In power generation, particularly that of coal fired plants, heat exchangers make up the bulk of the major plant components. With the exception of the main turbine set, the boiler, condenser and feed water heating train, including the deaerator, are all components that transfer heat between two fluid streams. This is a complex process, especially where phase change is involved, due to the geometry of the heat transfer component and the physical phenomena involved. They are typically analysed analytically and experimentally using the first and second law of thermodynamics (Mohanraj, et al., 2015). The theoretical heat exchanger analysis involves complicated equations with major intrinsic assumptions, whereas experimental methods are expensive due to the capital investment required in creating the test rig. Over the past decade ANNs have been increasingly used to simulate, optimise and predict performance of heat exchangers, overcoming the abovementioned constraints. This offers a major benefit in the power generation industry where heat exchange is at the heart of the process.

In addition to some of the literature discussed above, there has been specific work done on heat exchangers, such as estimating fouling and fouling factors using ANNs with Bayesian based training methods (Davoudi & Vaferi, 2018). Heat transfer and pressure drop characteristics of nanofluids in a heat exchanger were modelled using ANNs by Esfe (2017) with mean squared error less than the fifth decimal point. An air and water/ ethylene glycol fin tube compact heat exchanger was

modelled using ANNs to predict its thermal performance. It was found to be superior to non-linear regression models and the spatial variations in air temperature were also well predicted (Tan, et al., 2009). Multi-steam heat exchangers are most commonly found in cryogenic process and in the petrochemical industry. It was modelled using ANNs and first principle methods on a platform called Aspen Hysys, a very strong thermodynamic modelling tool (Khan, et al., 2012). Hysys was used to model the plant as well as generate the training data to feed into the ANN model by applying typical variations experienced by the process. Ultimately, the ANN proved to predict the desired output above the predetermined acceptable mean squared error.

Mohanraj et al (2015) conducted a detailed review of the various applications of ANNs in heat exchanger analysis. It critically analyses the methods employed for twelve different categories of heat exchangers found in industry as well as the various predictive applications which it can be used for. The aim was to compare network architectures used and identify limitations and areas for further research. Of particular importance to this study is that of condensers. Interestingly, only one case of a wet cooled condenser found in typical power generation plant was found. A multi-layer feed forward network was developed to predict the condenser heat transfer rate, heat transfer coefficient and cleanliness factor within a deviation of 1%, 17% and 8% respectively. Prediction of plant and production limiting factors such as condenser vacuum (or backpressure) has not yet been attempted.

Direct contact type heat exchanges were reviewed by the author. However, the studies discussed revolved around cooling towers. It appears that, after an extensive review of literature available, open feed water heaters (or deaerators) have not yet been studied. Shell and tube heat exchangers were also analysed. Predominantly oil/water and water/water systems have been studied to predict outlet temperatures and heat transfer rate. Again, the multi-layer feed forward network was favoured for this type of application, with the optimum architecture found by trial and error. Very good prediction was achieved in all cases studied. Steam/water shell and tube heat exchangers, commonly known has feed water heaters on a power plant, have not been studied. Suffice to say combinations of heat exchangers in parallel and/or in series, such as a feed water heating train, have also not been attempted.

An expansive list of the shortcomings and areas for further research has been identified. Those of particular interest to this study are related to the model and architecture development. One consideration is that genetic algorithms where not used in the training of the ANN models. It could be a beneficial exercise to assess this method against the more commonly used gradient descent backpropagation training method.

From the hundreds of papers that the author's review considered, there is concrete evidence that ANNs can be used successfully in the analysis of heat exchangers. While multi-layer feed forward

networks with back propagation were preferred due to its simplicity, it is not limited to this one method. It was found however that radial basis functional networks and generalised regression neural networks should not be used due to its unfavourable performance prediction, forecasting and estimation capabilities.

More recently, research by LeCun et al. (2015) has found that the most popular activation function used in neural networks is the ReLU (rectified linear unit). It is preferred to the sigmoid function and hyperbolic tangent since ReLU learns much faster, especially in networks with many layers. Goodfellow et al. (2016) states that the method for determining the type of activation function to use is trial and error based on the network's performance. However, the Rectified linear unit is an excellent choice. This is due to their ease of optimisation as a result of their derivative (gradient) being zero across half of it's domain and 1 when the unit or node is active. Large and consistent gradients from ReLU are far more useful for training than activation functions that introduce second order effects. Due to the widespread saturation of the sigmoid function at high and low input values, i.e. only strongly sensitive when input is near zero, it is now discouraged in feedforward networks.

Adapted from the stochastic gradient descent method, Adam (Adaptive moments) optimisation is currently one of the most widely used optimisation functions. It is an adaptive learning rate optimisation algorithm that incorporates momentum directly as an estimate of the first order moment (with exponential weighting) of the gradient which is calculated by the backpropagation algorithm (Goodfellow, et al., 2016).

Based on this study of literature, and given the intended application, artificial neural networks, specifically feed forward networks, will be used to predict plant performance. This network will be a fully connected feed forward neural network (multi-layer perceptron) using ReLU as the activation function at each layer. Adam optimisation with backpropagation will be used to minimise the error function.

3. Theoretical background

3.1 Feed water heaters

The purpose of this section is to present a theoretical foundation for understanding the power station's water/ steam cycle, a description of the major components and the role they play in the cycle. This enables one to understand the complexity of a feed water heater by gaining an appreciation of the technology and understand the variations in terms of type, orientation and performance of the component. This serves to highlight the complexity of modelling such a component, even more so when a combination of various types of heaters are employed in a system of series and parallel process configurations.

3.1.1 Rankine cycle and the role of feed water heaters

The Rankine cycle is the basis for all modern steam power cycles. From the figure below, chemical energy in a fuel source such as coal is converted to thermal energy in a boiler. This energy is absorbed by water, which is pumped into the boiler tubes (1-2), thereby generating steam (3). The thermal energy in the steam is then converted to mechanical energy through expansion through the turbine (3-4). The turbine and generator rotor are connected by a common shaft, which ultimately results in electrical energy being produced. The exhaust steam from the turbine is condensed back to a liquid to be used again in the cycle (4-1). Figure 1 below illustrates a simple closed loop process.

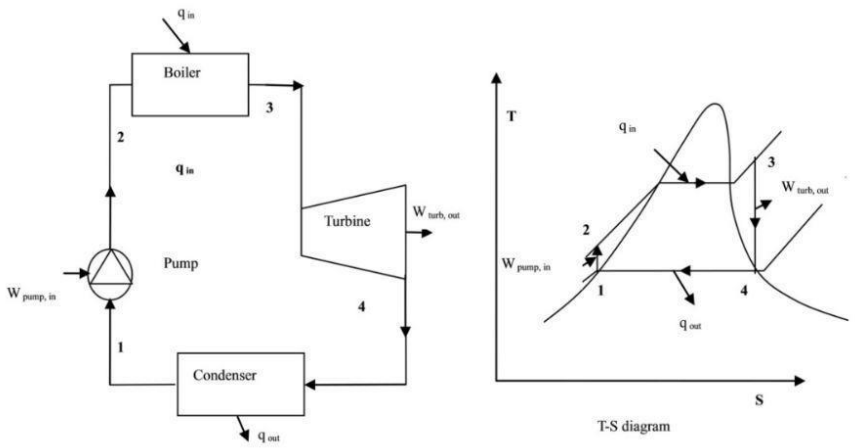


Figure 1- Basic closed loop Rankine cycle

The efficiency of such a cycle can be improved by means such as reducing the condensing temperature, increasing the boiler pressure, using superheated steam, using reheat steam with an additional turbine or turbine set, and regeneration using feed water heaters.

The principle of regeneration is simply to extract steam from the turbine and use it to preheat the water to a higher temperature before it reaches the boiler. This steam could have been used to do work in the turbine. However, by using it for heating, the primary energy input to the boiler is reduced. This increases the overall efficiency of the cycle. A simple regenerative cycle is depicted in Figure 2 where steam is extracted from the turbine (at point 7 in the figure below).

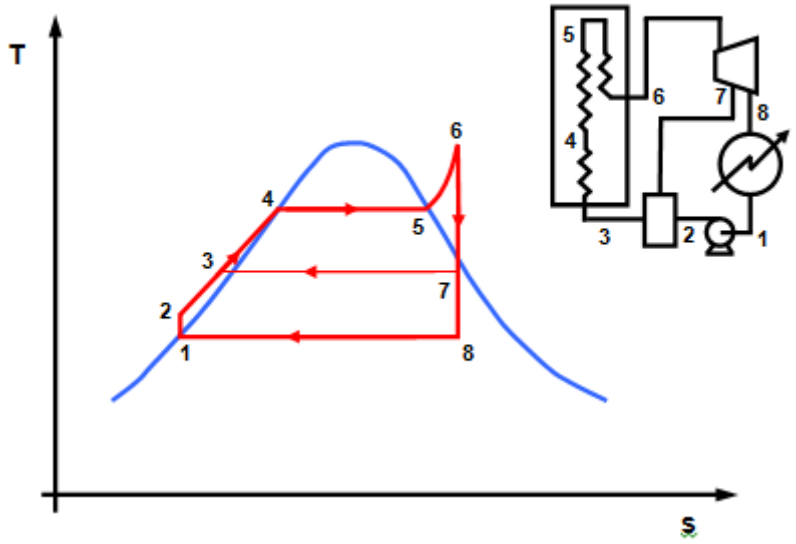


Figure 2- Rankine cycle with feed water heater (Terranova & Gibbard, 2008)

In a utility scale power plant, the design of the process, as seen below in Figure 3, is significantly more complex. It is especially important to select an optimum number and orientation of feed heaters so as to maximise the cycle efficiency. It must be noted that there is a point of diminishing return where the cost of an additional feed water heat does not justify the efficiency gain due to that heater. The feed water heating train is highlighted red in the figure below.

The feed heating train is typically made up of two different types of heaters, open and closed. The open feed water heater is referred to as the deaerator. The closed feed water heater is called either a low pressure heater or a high pressure heater depending on whether it sits on the low pressure side or the high pressure side of the feed water pump.

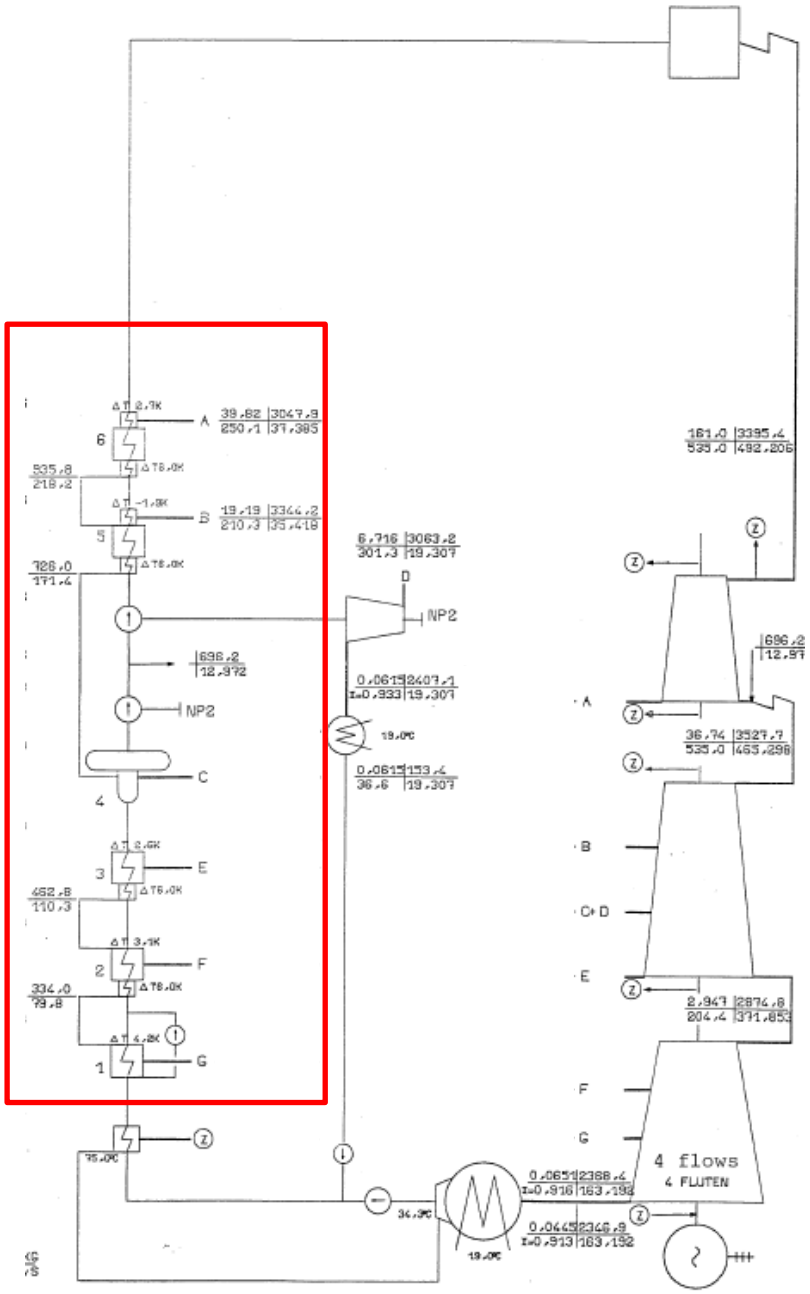


Figure 3- Heat Balance Diagram extract of an Eskom Power Station

3.1.2 Feed water heater types and geometry

Feed water heaters can be open or closed. The simplest form is an open heater where the extracted steam and the water to be heated are in direct contact with each other. Thus the extraction pressure of steam from the turbine must be the same as that of the pressure in the heater. This implies that in a train of heaters, a pump would be required after each heater to ensure circulation of the working fluid, making it uneconomical. For this reason, usually only one

open feed water heater is found in a cycle which also doubles as a deaerator (which removes dissolved gases from the water). A recent study by Banda (2015) found that there are three different types of deaerating feed water heaters. These are tray type, spray-scrubber type or atomiser types, each of which is very different in orientation and construction.

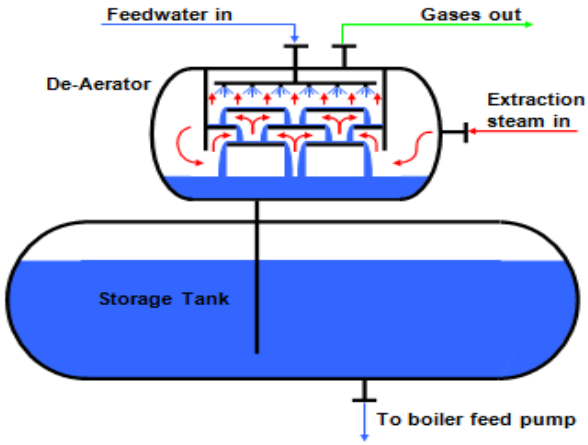


Figure 4- Open feed water heater (deaerator)

Closed feed water heaters are predominantly used to effect the preheating duty before the water enters the boiler. These heaters are generally shell and tube type heat exchangers and are termed closed since the water that is to be preheated and the extraction steam performing the heating are kept separate from each other.

Allie (2016) performed a study in which he found that there are broadly four different criteria which categorises the closed feed water heater. These are tubesheet or header type heaters, vertical or horizontal orientation, one, two or three zoned heat transfer, long or short drain cooling zone (if drains cooler exists).

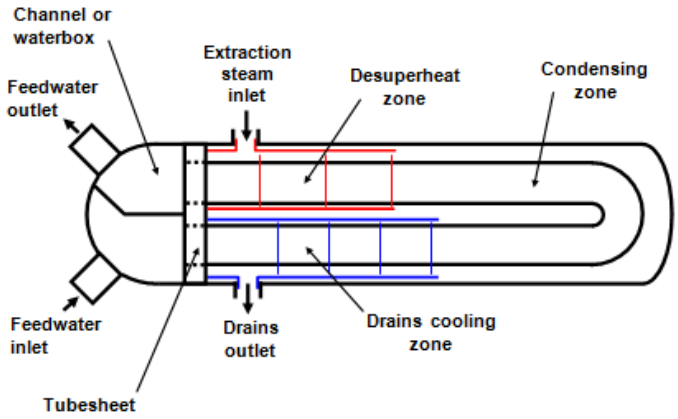
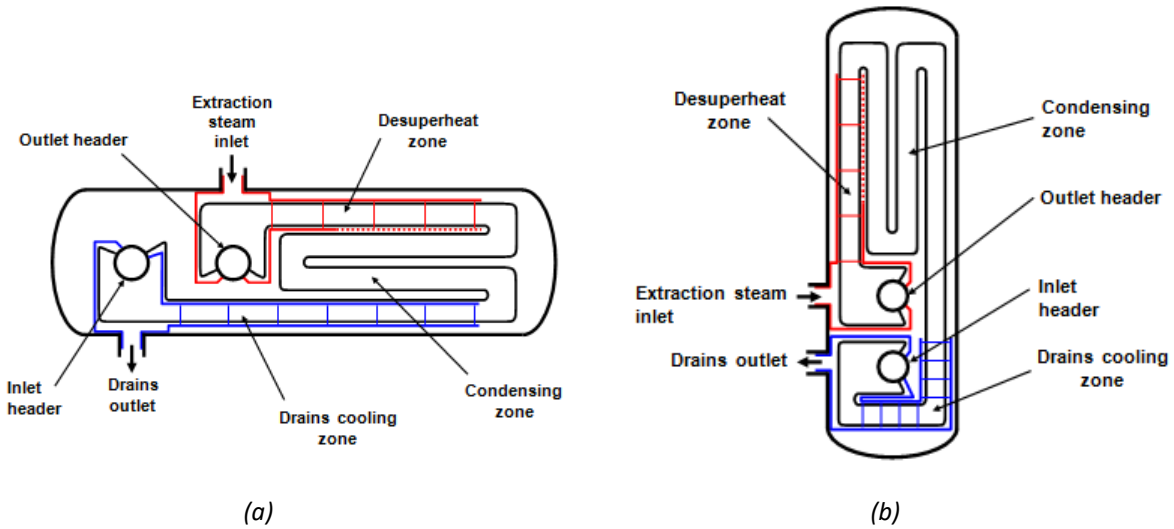


Figure 5- Horizontal three zone tubesheet type heater (Terranova & Gibbard, 2008)



(a) *(b)*

Figure 6- Horizontal (a) and vertical (b) header type heater (Terranova & Gibbard, 2008)

3.1.3 Feed water heater performance parameters

A typical closed feed water heater has one inlet stream and one outlet stream on the tube side which holds the feed water. On the shell side, steam enters from a turbine extraction point and the condensed steam which is formed after the heating duty is performed, termed distillate, exits the heater. The distillate can be sent to a down-stream heater or upstream heater, into either the shell side or the tube side, depending on the position of the heater in the train and the design. Hence on the shell side, there can be two inlet steams and one outlet stream.

Most heaters are designed with three distinct heat transfer zones. They are the desuperheating zone, the condensing zone and the drains cooling zone. The desuperheating zone, as its name suggests, is responsible for removing the sensible heat from the steam to a point where the steam reaches saturation conditions. The saturation temperature is dictated by the shell pressure. The condensing zone, where the bulk of the heat transfer takes place, removes latent heat from the saturated steam thus forming distillate. Finally, the drains cooler further removes sensible heat from the distillate before it is cascaded to a downstream heater or upstream into the feed water side.

There are three critical performance parameters associated with a feed water heater and provides information regarding the effectiveness of the three heat transfer zones. The first is the terminal temperature difference (TTD). This parameter is indicative of the efficiency of heat transfer. Lower TTDs indicate efficient performance whereas higher values indicate that the heater is underperforming.

For heat exchangers in general, the TTD is expressed as the difference between the hot inlet and cold outlet streams. However, this is not the case for feed water heaters. Here it is defined as the difference between the saturation temperature of the extracted steam and the outlet feed water temperature. In addition to this, in a feed water heater with a desuperheating zone, the TTD can also be negative. This is desirable as it would indicate effective heat transfer.

The second parameter is the drains cooler approach (DCA), which is also associated with level control in the heater. The DCA is determined by the difference between the distillate outlet temperature and the inlet feed water temperature. An increase in this value from its target or design value indicates a decrease in level and vice versa.

The last parameter is the sub-cooling temperature (SC) which is defined as the difference between the saturation temperature of the extracted steam and the distillate outlet temperature. These are expressed visually in Figure 7 below.

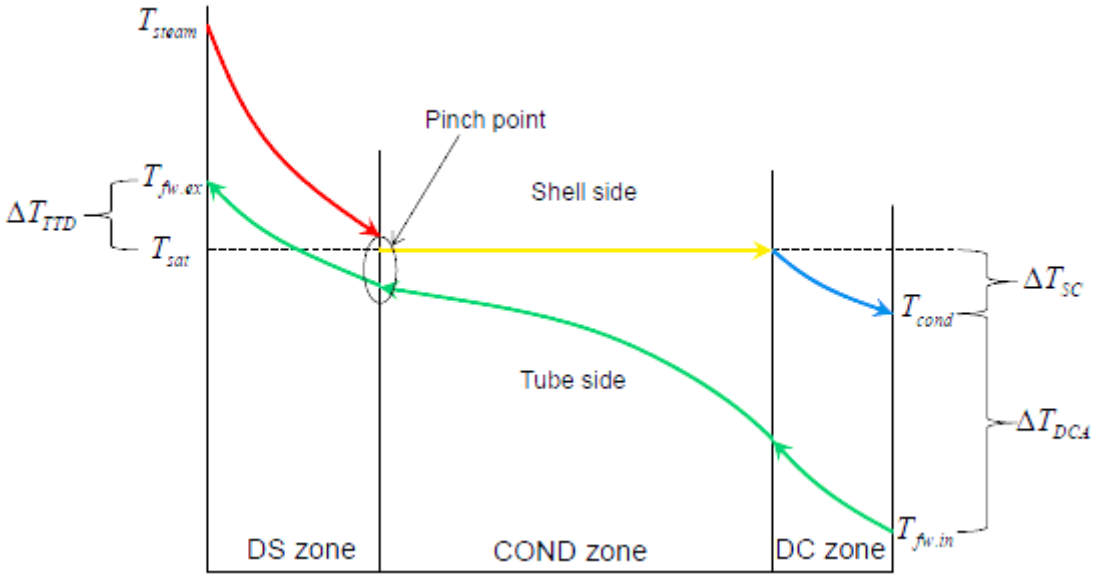


Figure 7 - Temperature profile of FWH streams and performance parameters (Rousseau & Fuls, 208)

The above parameters offer insight to the performance of the heater and potential causes of underperformance. High DCA is an indication of low water level since if the water level decreases, steam from the condensing zone enters the drains cooling causing the distillate temperature to be higher. Conversely a higher level will cause the reduction of the distillate temperature resulting in a lower DCA. This can imply incorrect control philosophy, incorrect valve position requiring stroking to correct or a passing valve altogether. In some cases, it can also point to a tube leak.

TTD can be negatively affected by a number of factors. The most common are changes in extraction steam pressure and temperature, plugged tubes which is a remedy for a tube leak, and

tube fouling which is caused by out of spec water quality causing a build-up of material that impedes heat transfer.

Figure 8 below depicts the typical temperature rise per heater in a train configuration depicted in Figure 3. This is specifically for full load operation. This provides a broad understanding of the duty each heater performs and their significance in attaining the desired final feed water temperature.

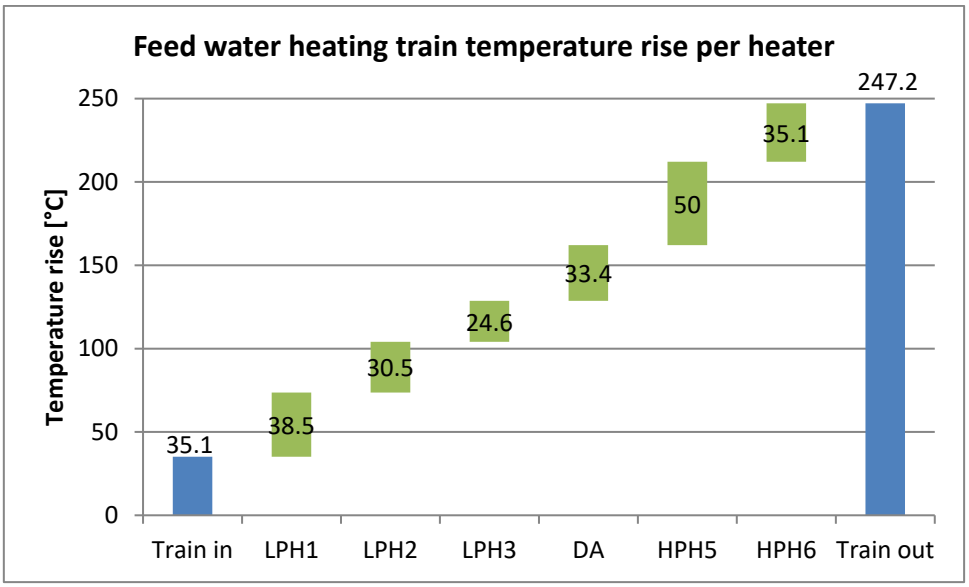


Figure 8 - Temperature rise per heater in feed water train

3.1.4 Thermohydraulic modelling considerations

It is clear that with different variations in design and orientation, accurately modelling a feed water heater can be extremely complicated. Models of this nature are specific to the heater in question since the geometry, both internal and external, plays a significant role in the modelling of key parameters and the performance overall. As an example, accurate knowledge of the tube lengths, tube pitch, baffle spacing, and thickness of baffle rings are all imperative in accurately estimating the free flow area on the shell side of the heat exchanger. The estimation technique is different for different types of baffle arrangements. This is just one parameter out of several that need to be accurately estimated for the overall model to be effective.

An extensive study was conducted by Allie (2016) into modelling feed water heaters taking the above into consideration. He also studied a number of modelling methods relating to the calculation of the heat transfer coefficient. This broadly took the shape of single phase forced convection correlations for both the internal and external tube boundary layer as well as external forced convection correlations for two phase fluids. Ultimately, to adequately model the performance of a feed water heater, the heater was broken down into the three heat transfer zones and further broken down into sub-zones. Each sub-zone therefore was modelled individually

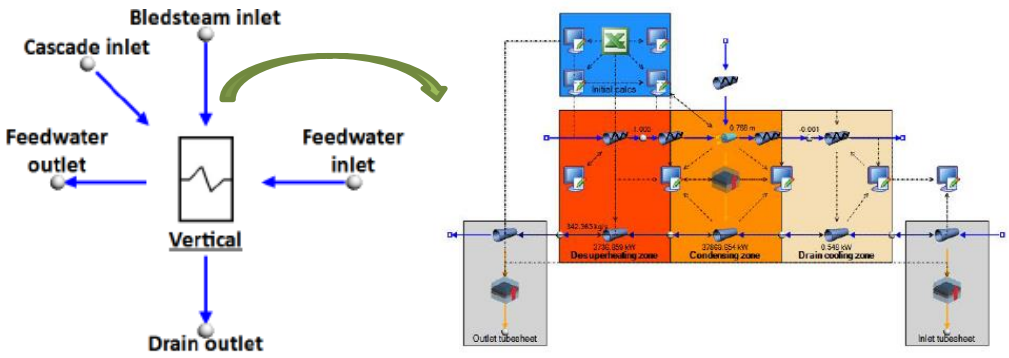
requiring a unique selection of geometry and applicable correlation for heat transfer calculation. This highlights the intricacies and detail required to build models of feed water heaters from first principles.

It must be highlighted that it is natural for additional divergences from design to come about during the operation of a feed water heater in a power station. This is exacerbated over time if the necessary corrective action is not taken to return the plant to its original design state since the plant will continue to degrade. This is usually the case since opportunity to maintain the plant requires that the entire power generating unit be off load for extended periods of time, usually only every two to three years. Hence, the accuracy of engineering-based models can significantly decrease especially if flexibility of this nature is not catered for.

3.1.5 FLOWNEX SE custom components

Research conducted by leGrange (2018) saw the construction of a feed water heater model on a thermofluid process modelling software platform, FLOWNEX® SE. This work saw the construction of the heat exchanger as a single custom component using as little as possible design input data while still adequately capturing the thermodynamic phenomena and realistically accounting for the hydraulic phenomena within the heater. The intent behind this was for the component set up to be as simple as possible while being effective in terms of cycle integration and performance prediction. It also catered for plant degradation in the form of the number of leaking tubes plugged as well as internal and external tube fouling that is generally experienced in a real plant environment. These custom components form the basis of the thermofluid modelling that was conducted as part of this project.

The figure below illustrates the detailed modelling of the desuperheating, condensing and drains cooling zone within the custom component for a vertical feed water heater in FLOWNEX® SE. Horizontal heaters are modelled in a similar manner.



(a) *(b)*

Figure 9 – (a) Custom component and (b) detailed feed water heater model in Flownex

The desuperheating zone was designed to be either active or inactive in the modelling of the heat transfer of the component depending on the steam condition entering the heater. If the steam is superheated, that is a quality greater than 1, then the desuperheating zone will be active. The heat transfer is calculated from equation 1 below. The major assumption is that the quality of steam leaving the desuperheating zone and entering the condensing zone, is at a quality of 1, i.e. saturated vapour.

$$\dot{Q}_{DSH} = \dot{m}(h_1 - h_2) \tag{1}$$

Hence the heat transfer can be calculated by multiplying the extraction steam flow, $\dot{m}$ [kg/s], into the heater by the difference in enthalpy of the extraction steam, h_1 [kJ/kg], and enthalpy of saturated vapour h_2 [kJ/kg] at the extraction pressure and quality of 1. If the steam quality entering the heater is less than 1, no heat transfer takes place in this zone.

The condensing zone, which is responsible for the majority of the heat transfer, is modelled in significantly more detail. The method employed uses design process parameters for multiple load cases which are input by the user. It then calculates the overall heat transfer coefficient, the UA value, using the log mean temperature difference (LMTD) method as per the equation below

$$\dot{Q}_{cond} = UA \frac{\Delta T_2 - \Delta T_1}{\ln(\frac{\Delta T_2}{\Delta T_1})} \tag{2}$$

where ΔT_1 is the temperature difference between the steam entering the condensing zone (quality of 1) and the feed water exiting the condensing zone; ΔT_2 is the temperature difference of the condensed steam (quality of 0) leaving the condensing zone and the feed water entering the condensing zone. Distillate cascading from downstream heaters is also catered for in the calculation of the UA value.

A curve fit (power fit) is applied to the UA values calculated for each load as a function of feed water mass flow rate. This enables the component to calculate the UA value for any load condition. Since FLOWNEX® SE heat transfer components do not cater for the UA value as an input, manipulation of this value was required to allow the program to properly interpret the information. This was done by using an inside to outside heat transfer coefficient ratio to determine the convective heat transfer coefficient to be accepted by FLOWNEX® SE. The tube internal heat transfer coefficient was thus calculated from the UA value, which satisfies the requirement of the heat transfer element of the custom component. The area available for heat transfer is also calculated using the geometry and liquid level as input by the user.

The drains cooling zone heat transfer is calculated using a second order polynomial curve fit based on the design process parameters input by the user. All of these calculations are executed in multiple scripts within FLOWNEX® SE.

Similar to the horizontal and vertical feed water heaters, the de-aerator is also a custom component built in FLOWNEX® SE as seen in figure 10 below.

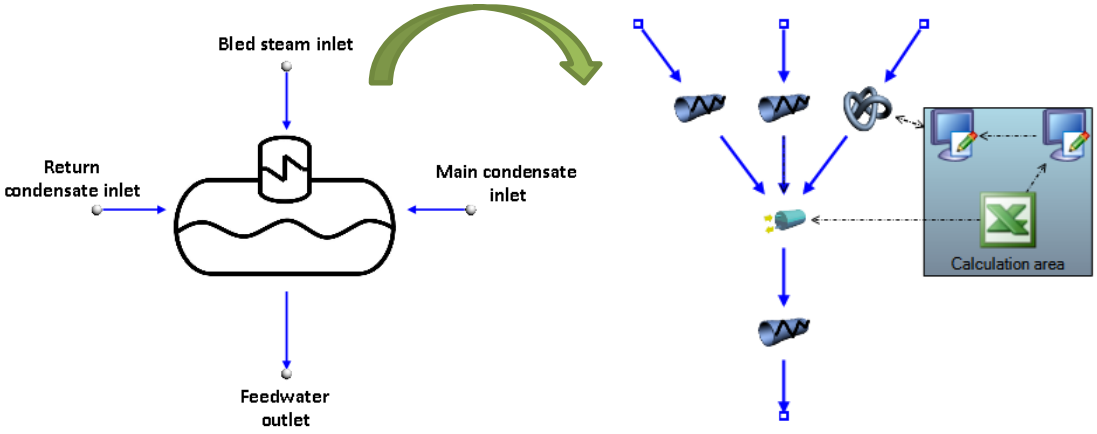


Figure 10 – (a) Custom component and (b) detailed de-aerator model in FLOWNEX® SE

The de-aerator was modelled to allow distillate from downstream heaters, bled steam from the turbine extraction and the feed water to interact in a two phase tank where the heat exchange takes place. The main calculation that takes place in this custom component is the characterisation of the pressure drop over a spray nozzle. This is required on the feed water (or main condensate) inlet line in order to reduce the pressure to that of the de-aerator. The spray nozzle valve was modelled using a variable geometry methodology where a curve is fit to the general empirical relationships loss coefficient (leGrange, 2018). Hence the model allows for variable pressures over the valve with varying load conditions. As in the case of the horizontal and vertical heaters, design process parameters for multiple load cases as well as some geometric data are required to be input by the user. These calculations are performed in scripts within the component.

4. Theoretical background

4.1 Deep learning

Machine learning is a technique that uses algorithms and statistical models to perform a specific task. It forms part of the greater artificial intelligence space. Simplistically it is a technique that enables a computer to behave like a human, by learning and improving the learning autonomously over time. This is done by feeding the model information, or training data, so that it can learn and thereafter make predictions without being explicitly programmed to do so.

The area of machine learning is vast, comprising of various methods to perform a task at hand. These methods are largely categorised as supervised or unsupervised learning. The techniques that can be used are regression, classification, clustering, association rule learning and deep learning, to mention a few. Each of these methods can be further broken down into a number of different algorithms that can be selected based on the application. For instance, if the desired prediction is a binary output (yes or no, true or false), classification would be better suited than regression. However, within the area of classification, one could opt for logistic regression, nearest neighbours, decision tree or random forest algorithms. Each has its unique ability that serves to provide better performance for a particular type of problem and less for others.

Textbooks on machine learning are expansive in their description of the multitude of algorithms that are available in the machine learning space. The intention of this chapter is therefore to focus on the techniques applicable only to this study as defined by the outcome of the literature review.

4.1.1 Artificial neural networks

Artificial neural networks is a subset of deep learning that takes inspiration from biological neural networks of the human brain. The brain consists of approximately 10^{11} connected neurons where each neuron can have a further 10^{4} connections (Hagan, et al., 1996). Each neuron is made up of a cell body, axon and dendrites as can be seen in the figure below.

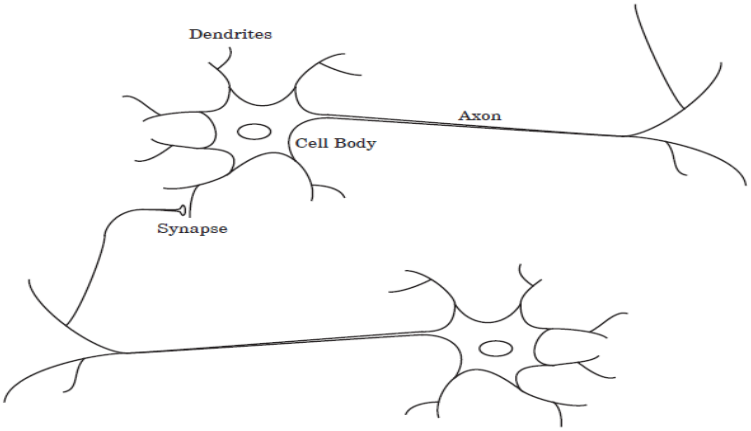


Figure 11 – Schematic drawing of Biological Neurons (Hagan, et al., 1996)

The dendrites are receptors that carry electrical signals to the cell body. The cell body essentially assesses the input signals collectively and produces an output signal when it reaches a specific threshold. The axon carries the output signal from the cell body to other neurons. The point of contact between the axon from one cell and the dendrite of the next is called a synapse. Synapses determine the strength of the incoming signal. As a result, synapses either encourage or inhibit the cell body to generate an output signal depending on the strength of the incoming signal.

Artificial neural networks are mathematical approximations to the biological process described above. Combinations of artificial neurons can thus make for highly complicated mathematical models which are able to identify unique features and approximate extremely complex relationships.

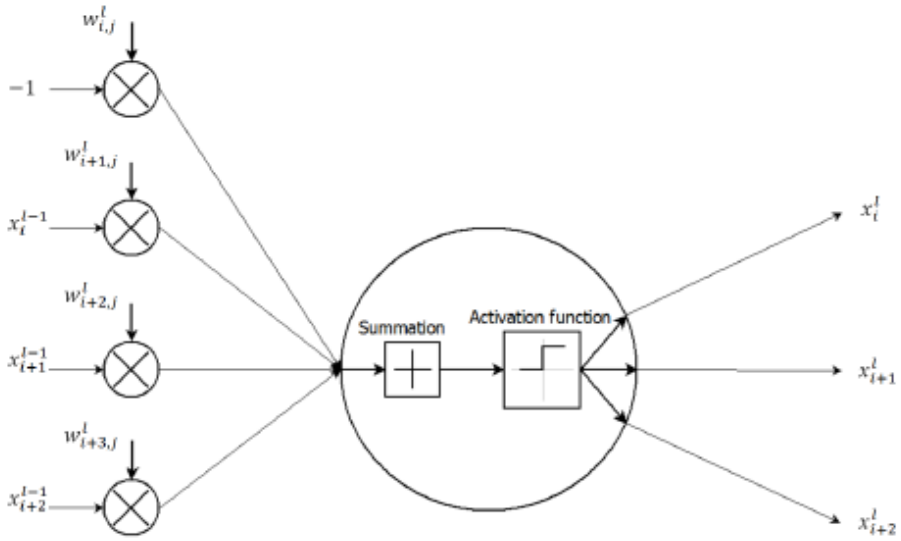


Figure 12 – Schematic of a single artificial neuron (Laubscher, 2017)

The figure above represents a single artificial neuron. x_i^{l-1} represents the input signal to the neuron. It is then multiplied by a weight variable, $w_{i,j}^{l}$. The weight determines the strength of the

incoming signal in that a large positive weight corresponds to a strong excitation and a small negative weight corresponds to a strong inhibition. The weighted signals are then summed together before it is assessed against a threshold. This threshold is determined by the use of an activation function which then generates the output signal. There are a number of different types of activation functions available, all with their strengths and weaknesses depending on the application it is used for. For the purposes of this study, as per the recommendation of current literature, the rectified linear unit (ReLU) activation function will be used. This is due to it being differentiable, which simplifies the implementation of a learning algorithm and due to their ease of optimisation as a result of their derivative (gradient) being zero across half its domain and 1 when the node is active as per Figure 13 below.

From the Figure 12 above, it is important to note that one of the input signals is set to -1. This is known as a bias. This allows the model to treat the activation function as a weight as opposed to adjusting the threshold function every time the weights are adjusted.

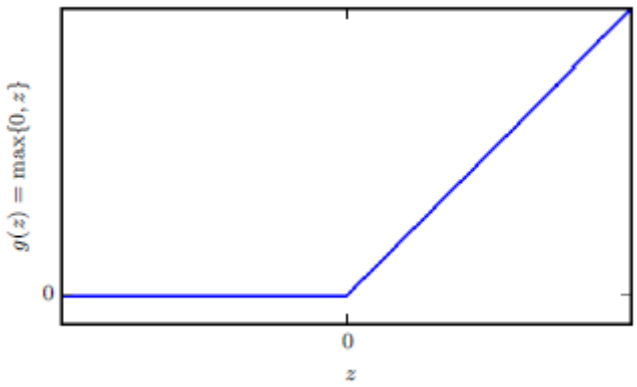


Figure 13 – ReLU activation function used in feed forward artificial neural networks (Goodfellow, et al., 2016)

4.1.2 Feedforward neural networks

Artificial neural networks are therefore made up of multiple neurons interconnected with each other. There are again many types of neural networks, however the specific type used for this study is the feed forward network, also known as the multi layered perceptron (MLP). It is best to think of feedforward networks as function approximation machines that are designed to achieve statistical generalization, occasionally drawing some insights from what we know about the brain, rather than as models of brain function (Goodfellow, et al., 2016). The goal of a feedforward network is to calculate the network's output signal.

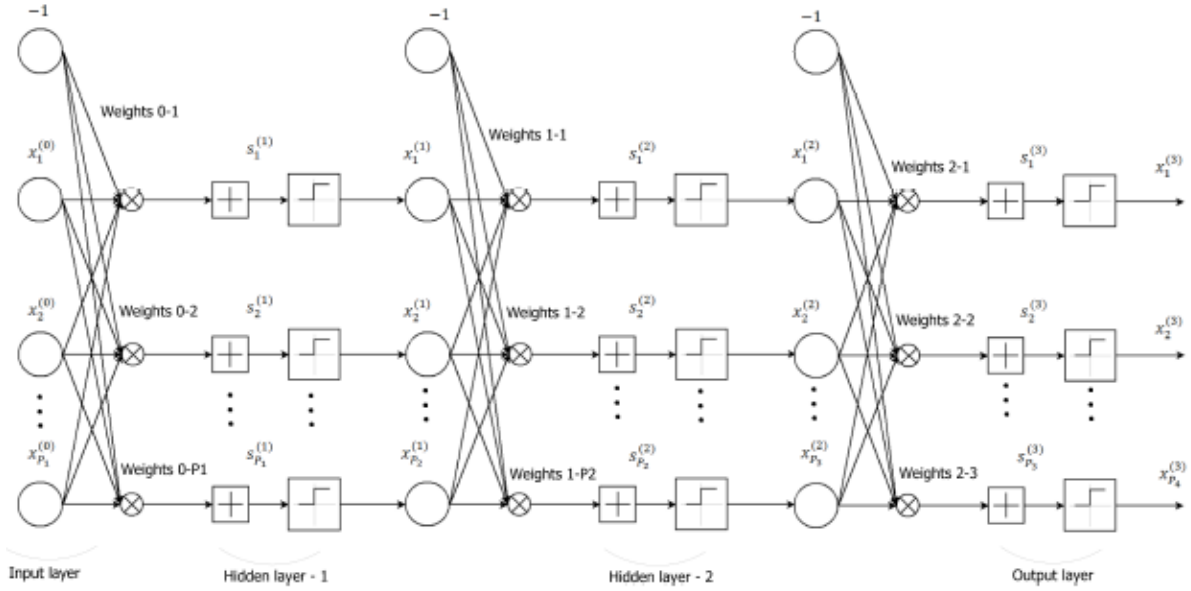


Figure 14 – Schematic of a multi-layer artificial neuron (Laubscher, 2017)

The training data specifies exactly what the output layer must do for each input. The behaviour of the layers in between is not directly specified by the training data. The learning algorithm must decide how to use those layers to produce the desired output, but the training data does not say what each individual layer should do. Because the training data does not show the desired output for each of these layers, these layers are called hidden layers. The number of hidden layers and the number of neurons or nodes per layer forms an integral part of this study.

The term feedforward is because of the flow of information through the function being evaluated from the input, through the intermediate computations, and finally to the output (Laubscher, 2017). There are no feedback connections in which outputs of the model are fed back into itself. In the case of Figure 12 the calculation would propagate through the network as follows:

$$s_1^l = -1 \times w_{i,j}^l + x_i^{l-1} \times w_{i+1,j}^l + x_{i+1}^{l-1} \times w_{i+2,j}^l + x_{i+2}^{l-1} \times w_{i+3,j}^l \tag{3}$$

with:

$$\begin{aligned} 1 \leq l \leq L \\ 0 \leq i \leq n^{(l-1)} \\ 1 \leq j \leq n^l \end{aligned} \tag{4}$$

where n is the number of neurons in layer l. $w_{i,j}^l$ is the weight variable with subscripts that describe the synapse between neuron i in layer $(l-1)$ and neuron j in layer l. x_i^{l-1} refers to the input signal to the neuron. Depending on the position in the network it can either be the actual values for the input layer as per the training data set or an output of a prior layer.

Thereafter this cumulative signal is input into an activation function, in this case the ReLU function described by $\theta(s_j^l) = \max(0, s_j^l)$. The output signal from the activation function, $\theta(s_j^l)$ is the output for neuron j in layer l. This process is repeated for each neuron in layer l before moving on to the next layer. This process then culminates at the output layer where the final result is the approximation for the actual output from the data set.

To assess how well the network performed in its output prediction based on the input dataset, a performance function or error function is used. Literature suggests that with the use of gradient based learning algorithms, the mean-squared error performance function is well suited. The calculation of the error, or in-sample error, for the kth observation is:

$$e(\boldsymbol{w}) = -\frac{1}{2}\sum_{j=1}^{n_{out}} (y_i - x_j^l)^2 \quad (5)$$

Where x_j^l is the approximation from the model output, y_i is the actual output as per the data set.

4.1.3 Backpropagation

It is clear from equation (5) that the error is a function of the weight. Therefore, in order to minimise the error, the weight needs to be optimised such that the mean squared error, equation (5), is minimised. This task is known as model training or learning.

The ability for the model to be trained emanates from the use of a learning algorithm. The purpose of the algorithm is to adjust the weights and biases in an effort to map the input features to the corresponding response features of the training data set within an acceptable level of error. Simply put, it learns the relationship between the input and output variables. Once again, many algorithms are available in the form of supervised, unsupervised and reinforcement learning. Supervised learning using back propagation and adaptive moments (Adam) optimisation was selected for the present research as per the literature study.

Adam optimisation is responsible for updating the model parameters thereby iteratively optimising the model. Adam, being an adaptation of the stochastic gradient descent algorithm, minimises the error by iteratively adjusting the weights in the direction of the steepest descent. However, it also makes use of part of the previous gradient, that being the moving average of the gradient and the root mean square of the exponential average of square of gradients, together with the current gradient. This adaptation is known as momentum. Another feature is that it provides the ability to use an adaptive learning rate (Goodfellow, et al., 2016). Both of these adaptations are designed to accelerate learning and reduce the likelihood of the model converging at local minimum or saddle point. This makes it superior to stochastic gradient descend and RMSProp algorithms.

The back-propagation algorithm is simply the technique used to calculate the gradient of the error function required by the Adam optimisation algorithm. It does this using the chain rule from calculus. The gradient term is determined by:

$$\mathrm{g} = \nabla \mathrm{e}(\mathbf{w}) = \frac{\partial \mathrm{e}(\mathrm{w})}{\partial w_{i,j}^{l}} \; for\ all\ i,j,l \tag{6}$$

The weights are adjusted via Adam optimisation as follows:

$$v^t = \beta_1 \times v^{t-1} - (1-\beta_1) \times g^t \tag{7}$$

$$s^t = \beta_2 \times s^{t-1} - (1-\beta_2) \times g^{t^2} \tag{8}$$

$$w^{t+1} = w^t - \eta \frac{v^t}{\sqrt{s^t + \varepsilon}} \times g^t \tag{9}$$

where η is the initial learning rate, g^t the gradient at iteration t (that is $\nabla \mathrm{e}(\mathrm{w})$), v^t the exponential average of gradients, s^t the exponential average of squares of gradients, β_1 and β_2 the hyperparameters. ε is usually set to a very small positive number to prevent division by zero.

Training of the model thus occurs in two distinct steps at each layer within each iteration. First is the calculation of the derivative of the error function and second is the use of this derivative to adjust the weights. The process starts at the final or output layer and after the weights are adjusted in this layer, the process is then repeated for the previous layer and so on till the first hidden layer is completed. Intuitively, one can see that starting the calculation at the output layer propagates the error moving backwards though the network hence the name backpropagation (Bishop, 2006). The backpropagation algorithm will make a large change to the weight at a node if it results in a large decrease in the observed error.

4.1.4 Model tuning

Ultimately, the point of the model is to be able to perform well when new and previously unseen data is fed into the model. This means that the model must have good generalization i.e. appropriate capacity. In some instances, underfitting can occur, which means that the model is unable to achieve a low enough error on the training set. On the other end of the spectrum, overfitting can occur. This is when the error between the training set and test set is too large (Goodfellow, et al., 2016). This means that the model becomes very good at predicting the output of the training set but cannot do this accurately for data it has not previously seen. Intuitively, one can think of this as fitting an order 9 polynomial equation to a quadratic data set, similar to the image below.

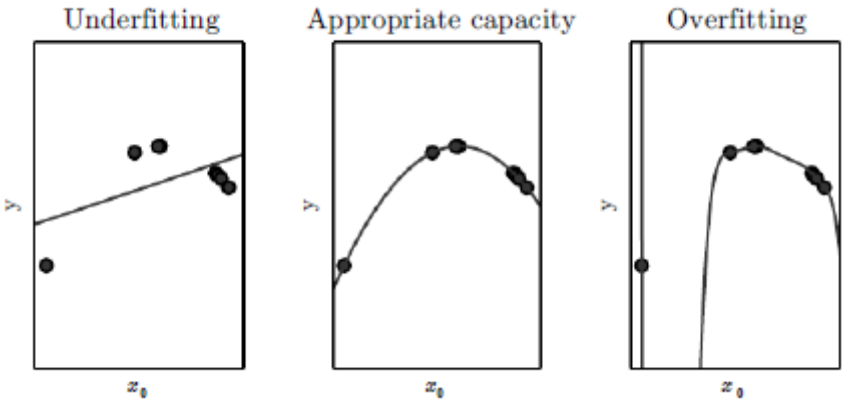


Figure 15 – Schematic of a multi-layer artificial neuron (Goodfellow, et al., 2016)

Strategies that can be considered to overcome the underfitting issue is increasing the data set or using data augmentation techniques if additional data is not available. Further to this, a more complex model can be constructed by increasing the number of layers and nodes per layer to more accurately identify the features of the data set. This strategy is in opposition to reducing overfitting, where model complexity may need to be reduced. Hence careful adjustment is required to find the right balance. Another strategy for reducing overfitting is to relax the regularization parameters. Regularisation is a method used to assist the training of a model when there is noise in the dataset as well as prevent overfitting of the model. It assists in the trade-off of having the model fit the training data well against having it generalise better for new and unseen data. It achieves this by applying a penalty for large weights. Some methods include L2 regularisation and dropout.

These together with batch size (the number of training examples used in a one iteration), epochs (the number of times the network is trained on the entire data set) and decay rate, are major hyperparameters to be considered in the present research in order to construct a network with an optimised architecture, as described in the chapter 5.2. Other parameters that will require tuning are known as learnable parameters that require initialisation to some value at the outset. Once set at the start, the model will automatically tune these parameters as the network is trained. Some of these include the initial weights, biases and learning rate for the Adam optimiser.

5. Methodology

5.1 FLOWNEX® SE modelling

Custom components developed by leGrange (2018) were tailored to construct the feed water heating environment of an Eskom power station. Although the custom component appears to be a simple element requiring minimal design input data, it should be noted that detailed calculations were used to account for level change, thermal inertia and heat transfer when developed as explained in section 3.1.4. It incorporates a level of modelling flexibility in that the heater geometry, heat transfer zones, operational parameters (e.g. tube plugging, fouling) and other physical characteristics such as orientation and may be specified by the user.

Heat balance data are the design data provided by the original equipment manufacturer relating to the operating conditions of the water-steam cycle at various loads. This information was used to characterise the model in terms of the calculation of the overall heat transfer co-efficient. Where information was not readily available from the heat balance diagram, mass and energy balance calculations were performed to estimate conditions.

Geometric information for the component was acquired from design and operational manuals as well as from plant engineers at the power station.

The inlet to the first heater, outlet of the last heater and the steam extraction to each heater was selected as the major boundary points to the model. Heat balance data for the Power Station was used for verification of the model. Four load cases (100%, 80%, 60% and 40% MCR) were applied as input to the model and the modelled performance of each heater was compared to the actual heat balance.

Mass and energy balance calculations on each of the feed heaters for the 100% load case was conducted on Mathcad to validate the results of the FLOWNEX® SE model. Since the custom components were developed outside the scope of this study, the mass and energy balance serve as an additional verification step to ensure the modelled results are consistent on a fundamental level.

Once an acceptable level of accuracy had been achieved, real plant data was obtained from the power station and used to simulate the plant. This data was used for boundary condition inputs and comparison of the FLOWNEX® SE model output. At this juncture, the heaters were tuned in terms of tube plugging and fouling to be more representative of actual conditions on the plant. This assessment using plant data was conducted for six load cases (100%, 90%, 74%, 65%, 56% and 48% MCR) thereby validating the model.

5.2 Artificial neural network model development

Artificial neural network models were constructed to perform the predictive modelling tasks required by this project. It made use of the feed forward or multi-layer perceptron (MLP) architecture with backpropagation and Adam optimisation as the learning algorithm. The rectified linear unit (ReLU) activation function was used for all nodes and mean squared error function was used to assess the model's performance. The training data used for this network was the same set that was used to simulate the real plant conditions in the FLOWNEX® SE model. This data covers an operating load range from 48% to 100%.

With the above already being determined through literature and current industry trends, the first step in developing this network was to define the input and output parameters of the network. The next step was to optimise the pertinent parameters of the network to achieve the most accurate model defined by the model with the lowest error.

Hence the starting point was the definition of an acceptable error level and in turn, the model accuracy. Next required focus on the model itself. From an architecture perspective, this meant adjusting the number of hidden layers and number of nodes per layer to understand the sensitivity of these parameters on performance. From an algorithm perspective, hyperparameters such as initial weights, learning rates and biases, to name a few, also require tuning.

This was done in an iterative manner until the desired result was achieved. At this point, the architecture of the network was completely defined and ready for use. The five load cases used to test the FLOWNEX® SE model was used on the now optimised neural network and the predicted output compared to that of the FLOWNEX® SE model as well as actual plant behaviour. Thereafter the model was used to test its ability to extrapolate under abnormal operating conditions that was not part of the training set.

Figure 16 (a) below illustrates the methodology of attaining an optimised neural network in the form of a flowchart.

5.2.1 Python programming language

Python was chosen as the language for constructing the artificial neural network model. Python is an object orientated programming language that is one of the most popular in the machine learning community. This is because it is easy to understand, information and help is easily available and is well documented. Furthermore, in applications such as this project, for non-computer scientist where the intricacies and fundamental coding of the algorithms are not the

primary focus, Python allows one to use libraries and packages where algorithms are available and can simply be tailored to suit the application at hand. Such is the ease of application with object orientated programming languages. Furthermore, Python is free open source software that can be used on any platform. The libraries pertinent to this project are Tensoflow and Keras, which are specific to the neural network modelling. For data pre-processing and data separation sklearn was be used. Finally, for mathematical functions and visual data representation numpy and matplotlib are the libraries of choice. The programming methodology using these libraries is depicted in Figure 16 (b) below.

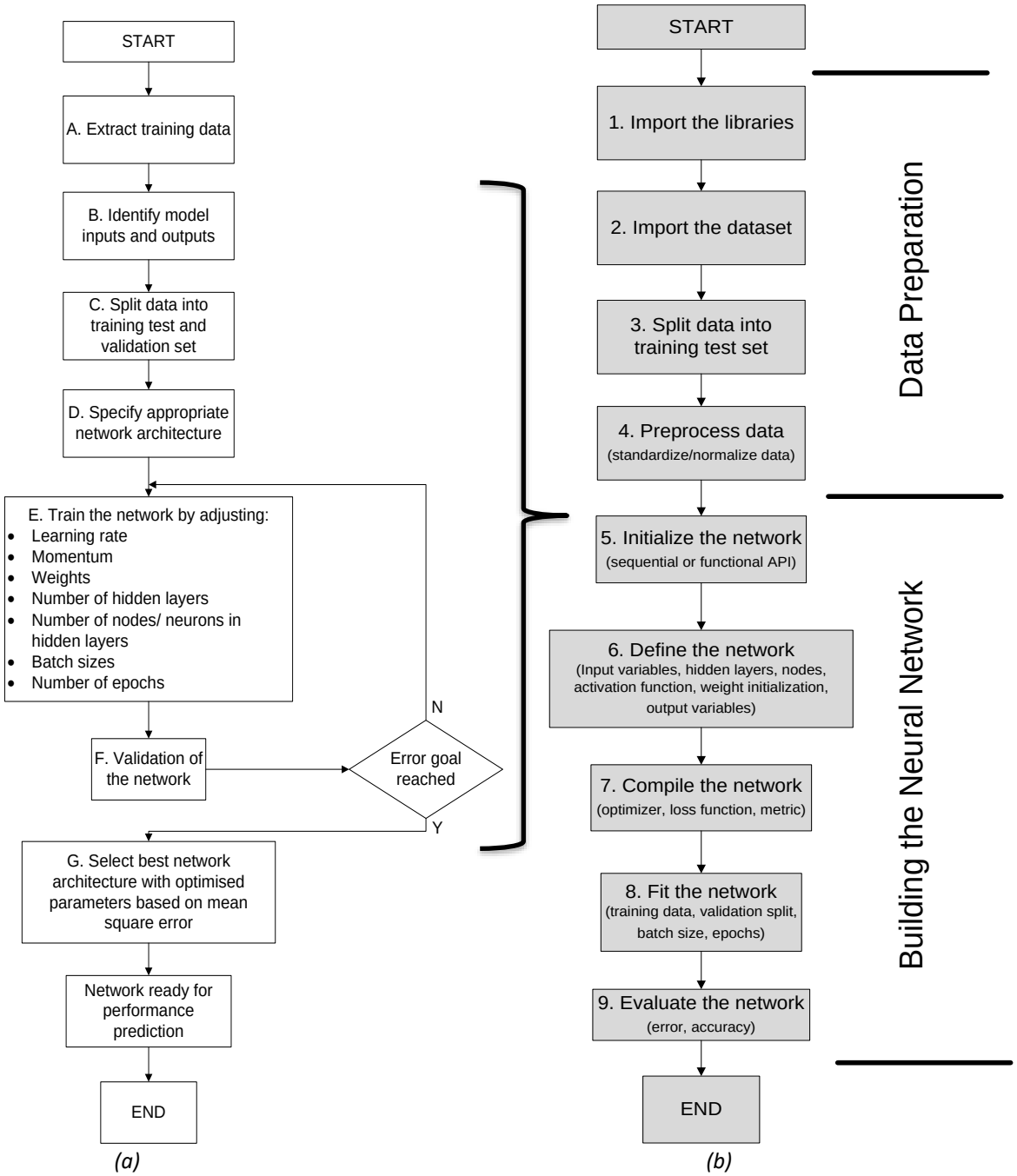


Figure 16 – (a) Flowchart for attaining optimum network (b) Flowchart of building the network using Python and Keras

5.3 Model extrapolation capability

One of the goals of this project is to assess the ability of each model to predict performance and behaviour outside of the normal operating conditions of the plant. This is essentially testing the ability of each model to accurately extrapolate information beyond what it was trained for. As a result, a set of practical scenarios have been designed to simulate out of normal states that could typically be experienced on a power plant. Each model (FLOWNEX® SE and ANN) will experience the same out of normal condition as an input and the predictions will be compared to each other.

The table below illustrates the five scenarios that will be tested as well as the method of achieving the out of normal state in both the FLOWNEX® SE model and the ANN model:

Table 1 – Model extrapolation scenarios

No.	Test	Flownex Method	ANN Method
1	Simulating fouled heaters	Adjust fouling parameters till TTD = 20°C	Set TTD - 20°C
2a	Simulating HPH 5A out of service	Set feed flow to heater to 0 kg/s	Set feed flow to heater to 0 kg/s
2b	Simulating HPH 6A out of service	Set feed flow to heater to 0 kg/s	Set feed flow to heater to 0 kg/s
3	Simulating HPH A bank out of service	Set bypass line to admit 50% of flow and line to A bank to admit 0% flow	Set feed water flow to each heater to 0kg/s
4	Operating at lower load and heaters out of service	Use the downloaded input data for boundary conditions	Use the downloaded input data as input data to the model

6. FLOWNEX® SE model performance

The figure below illustrates the integrated FLOWNEX® SE model of the feed water heating train. Appendix D contains a view of the LP and HP sides independently where the data are more clearly represented.

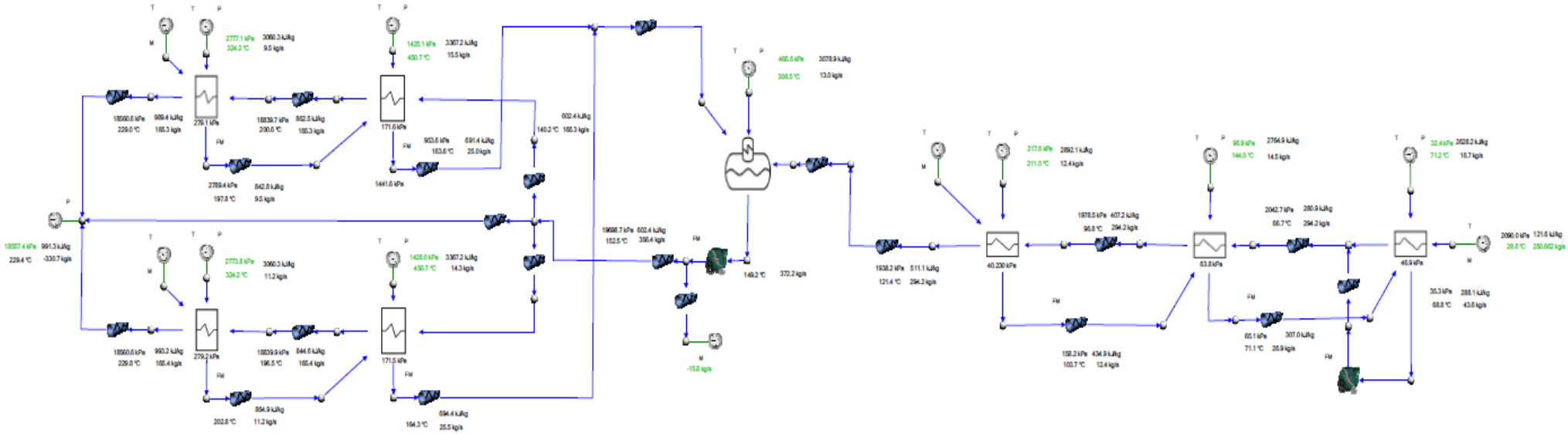


Figure 17– Integrated feed water heating train in FLOWNEX® SE

6.1 Model Performance parameters

There are three important performance parameters assessed for each heater. The bled steam mass flow rate provides a measure of the energy input required to effect the transfer of heat to the feed water. This flow also provides insight into the impact on the turbine cycle since the amount of steam bled to the heater directly affects the amount of steam remaining for power generation. In the current plant setup, if the flowrate of bled steam to the heater was greater than it was designed for, for a particular generating load, the system reacts by making up this flow deficit to the turbine by increasing the feed water and hence steam flow rate of the cycle. It does this to ensure that the power output remains constant. Boiler firing therefore increases to heat this additional fluid to the required steam condition for the turbine. As a result, plant thermal efficiency is reduced since more energy input in the boiler is required for the same power output.

The feed water outlet temperature of each heater tells us if the heater is achieving the outlet temperature desired by the power generation process for a particular load. This parameter, more especially with the high pressure heaters, has a direct and significant impact on the power generation process as a whole. A higher feed water temperature into the boiler implies a higher average temperature at which heat is added in the boiler. From the Carnot principle a higher temperature where heat is added results in a higher efficiency. Upstream heaters also have a negative effect on thermal efficiency if outlet water temperatures are lower than required. This effect is described in greater detail in chapter 9.

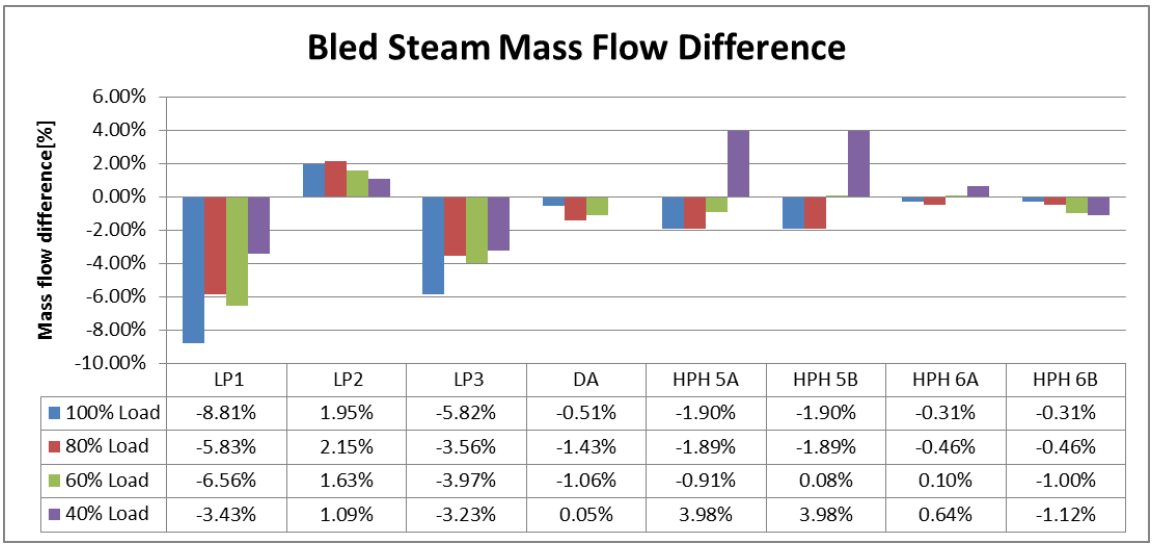
These two parameters are the focus of the results in this and the chapters to follow. The third parameter is the distillate temperature which is used only in the initial assessment of the heater. This parameter allowed the evaluation of the energy balance around the heater to asses if it was consistent with thermodynamic principles. It was therefore not considered after the manual verification step as discussed in section 6.2.

6.2 FLOWNEX® SE model verification

The feed water heating train constructed in FLOWNEX® SE was verified using the heat balance diagram information from the original equipment manufacturer for the Power Station on which this model is based. From Figure 17, the 20 boundary values input to the model are highlighted in green in while those in black are calculated by the model. The boundary values of the model are the inlet mass flow and temperature of the feed water to the train, the outlet pressure of the train, the reheater attemperation spray flow extracted after the pump and the bled steam temperature and pressure to each heater.

The verification of any model is an imperative step since it provides confidence in the model results. The verification was conducted for four load cases, 100%, 80%, 60% and 40%. The graphs below are a depiction of the difference between the FLOWNEX® SE model results and the expected output from the heat balance diagram data.

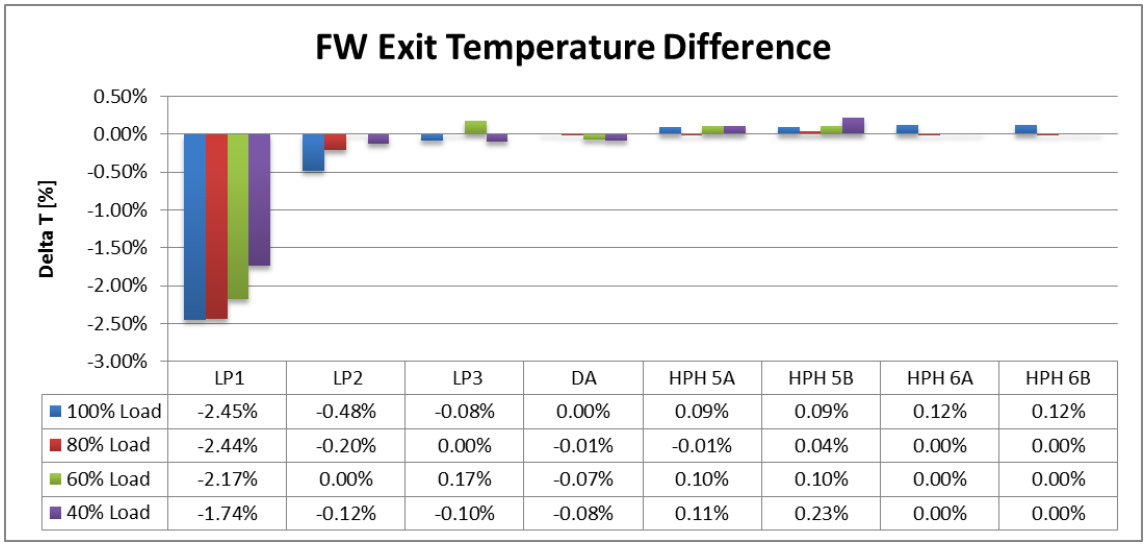
It is evident that the mass flow rate difference of bled steam to the heaters were within 6% with the exception of LP heater 1 at 100% and 60% load cases. It is suspected that this is a result of an inaccurate heat balance steam temperature used to characterise the LP heater 1 in FLOWNEX® SE. OEM heat balance data provided the extraction quality (less than one) and the saturation temperature for this heater while actual temperature and pressure was specified for all the remaining heaters in the train. Thus, the temperature used to characterise LP heater 1 in the FLOWNEX® SE model was estimated to achieve the required extraction enthalpy. The effect of this is seen below were the error around LPH1 is greater than the remaining heaters in terms of flow. It should also be noted that based on the under-prediction of feed water exit temperature (figure 19), LPH2 compensates for this by increasing bled steam mass flow. The remaining heaters and particularly the HP heaters 6 prove to determine the bled steam mass flow rate with greater accuracy.



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100% Load	-8.81%	1.95%	-5.82%	-0.51%	-1.90%	-1.90%	-0.31%	-0.31%
80% Load	-5.83%	2.15%	-3.56%	-1.43%	-1.89%	-1.89%	-0.46%	-0.46%
60% Load	-6.56%	1.63%	-3.97%	-1.06%	-0.91%	0.08%	0.10%	-1.00%
40% Load	-3.43%	1.09%	-3.23%	0.05%	3.98%	3.98%	0.64%	-1.12%

Figure 18– Bled steam mass flow rate difference (FLOWNEX® SE vs Heat balance data)

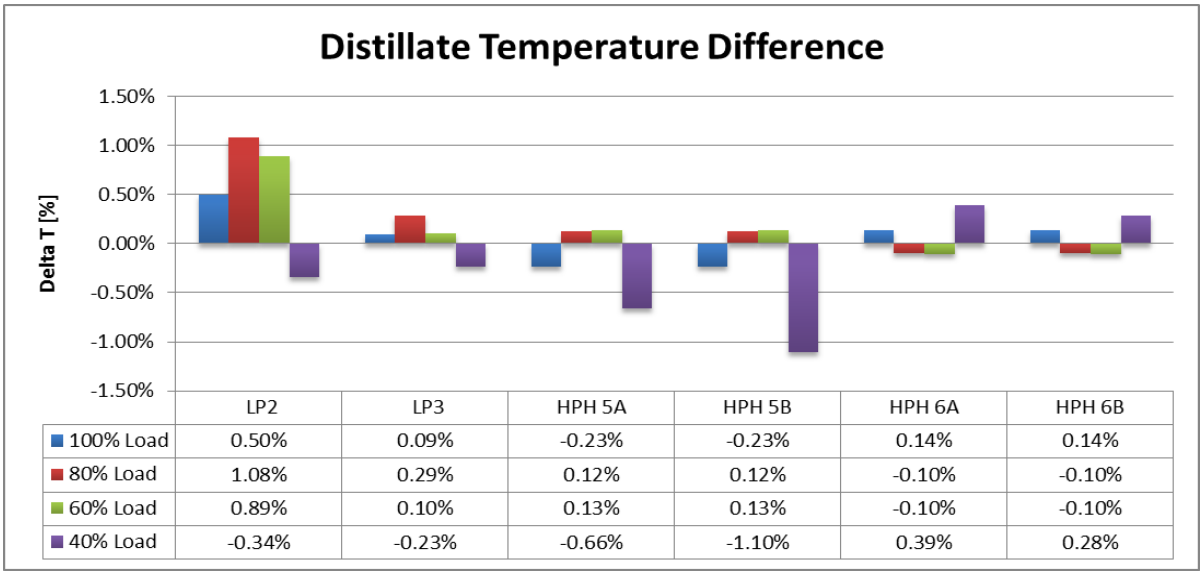
The exit temperatures of the remaining heaters at all load cases, were modelled within 0.5% of heat balance data.



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100% Load	-2.45%	-0.48%	-0.08%	0.00%	0.09%	0.09%	0.12%	0.12%
80% Load	-2.44%	-0.20%	0.00%	-0.01%	-0.01%	0.04%	0.00%	0.00%
60% Load	-2.17%	0.00%	0.17%	-0.07%	0.10%	0.10%	0.00%	0.00%
40% Load	-1.74%	-0.12%	-0.10%	-0.08%	0.11%	0.23%	0.00%	0.00%

Figure 19– Feed water exit temperature difference (FLOWNEX® SE vs Heat Balance data)

The distillate temperature for all heaters in the train was determined within a 1.1% difference of the heat balance data for all load cases.



	LP2	LP3	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100% Load	0.50%	0.09%	-0.23%	-0.23%	0.14%	0.14%
80% Load	1.08%	0.29%	0.12%	0.12%	-0.10%	-0.10%
60% Load	0.89%	0.10%	0.13%	0.13%	-0.10%	-0.10%
40% Load	-0.34%	-0.23%	-0.66%	-1.10%	0.39%	0.28%

Figure 20– Feed heater distillate temperature difference (FLOWNEX® SE vs Heat balance data)

The results depicted above highlight the ability of the FLOWNEX® SE model to accurately determine pertinent performance parameters of the feed water heaters. It stands to reason that the methods implemented within the custom components, and by extension the model of the entire train, are verified to be correct. Further to this, a mass and energy balance calculation was performed in Mathcad as further verification of the heaters on a fundamental level (see Appendix A). In the context of this project, the model was deemed to be sufficiently accurate to enable the objective of the study to be achieved.

6.3 Plant performance deviation

After the verification of the FLOWNEX® SE model using design information, the next step was to extract live plant data from the Power station as input to the model and for comparison for the prediction. This was necessary firstly to validate the model by assessing the modelled output against the actual plant behaviour. Secondly, it served as a point of comparison against the machine learning model outputs using the same plant data as input.

It is important to highlight the difference in actual plant and design (heat balance data) operation. The graphs below serve to highlight these deviations at the 100% load. Interestingly, the mass flow rate of bled steam on the plant is seen to be higher than the heat balance data. Similarly, the exit feed water temperature for LP heaters 2 and 3 are also higher while LP heater 1 and the HP heaters are lower.

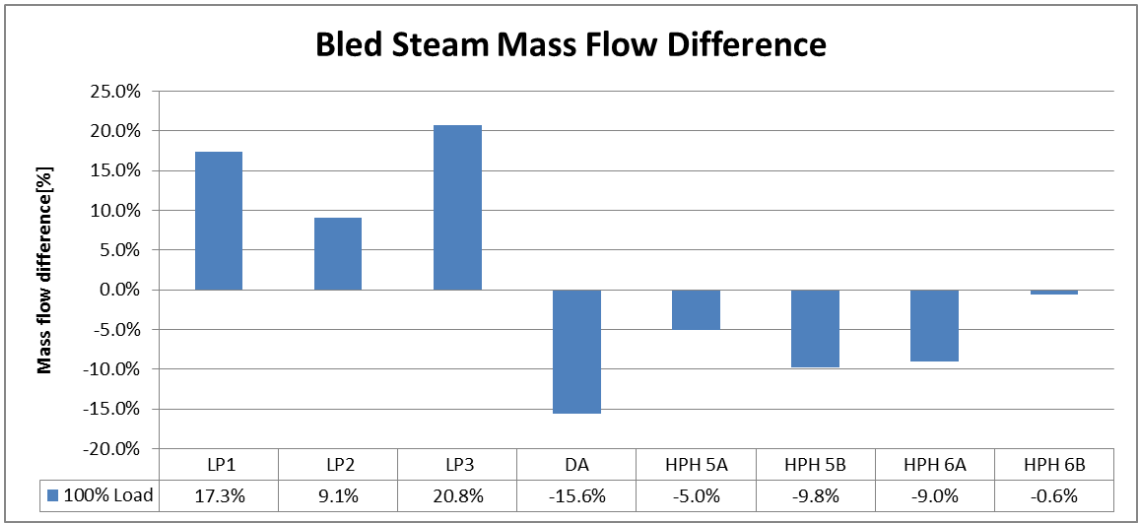


Figure 21– Bled steam mass flow rate difference (Actual plant data vs Heat balance data)

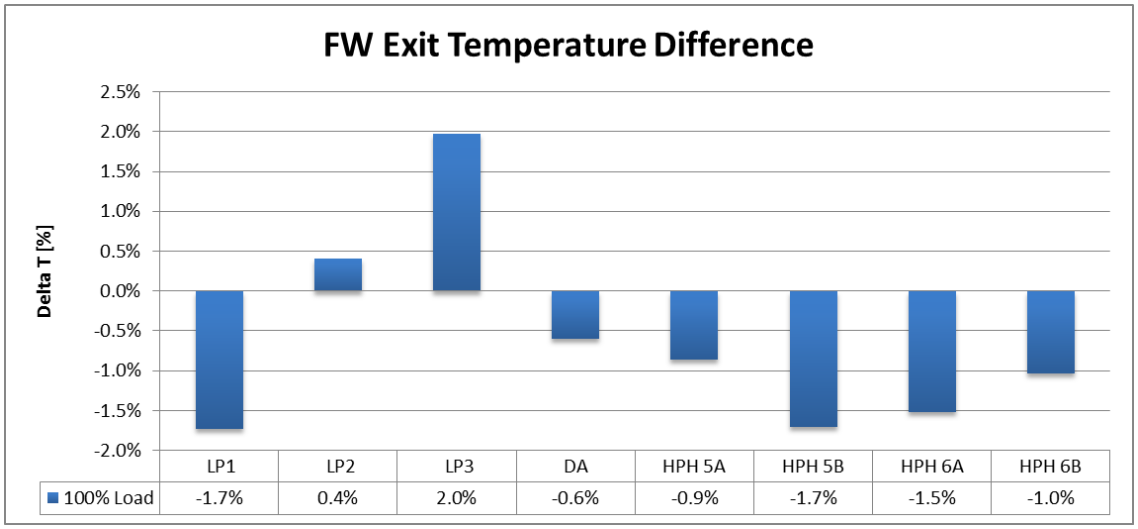


Figure 22– Feed water exit temperature difference (Actual plant data vs Heat balance data)

6.4 FLOWNEX® SE model validation

6.4.1 Model tuning

After verification of the FLOWNEX® SE model, live plant data at various loads was used as input to the model and the pertinent process parameters were predicted. The FLOWNEX® SE model, still characterised by the heat balance data, was initially executed using plant data at 100% load. A difference in predicted and actual performance was expected as a result of physical plant degradation over its lifetime, as illustrated in the previous section.

To account for this deviation to a certain extent, the model was tuned to real conditions by factoring in actual tube plugging as well as internal and external tube fouling. The figures on the following page represent the difference between the un-tuned and tuned model output and actual plant performance. It is evident that the un-tuned model was over predicting both the bled steam mass flow and outlet feed water temperature of the HP heaters. Fouling factors were therefore applied to tune the model to more realistic performance.

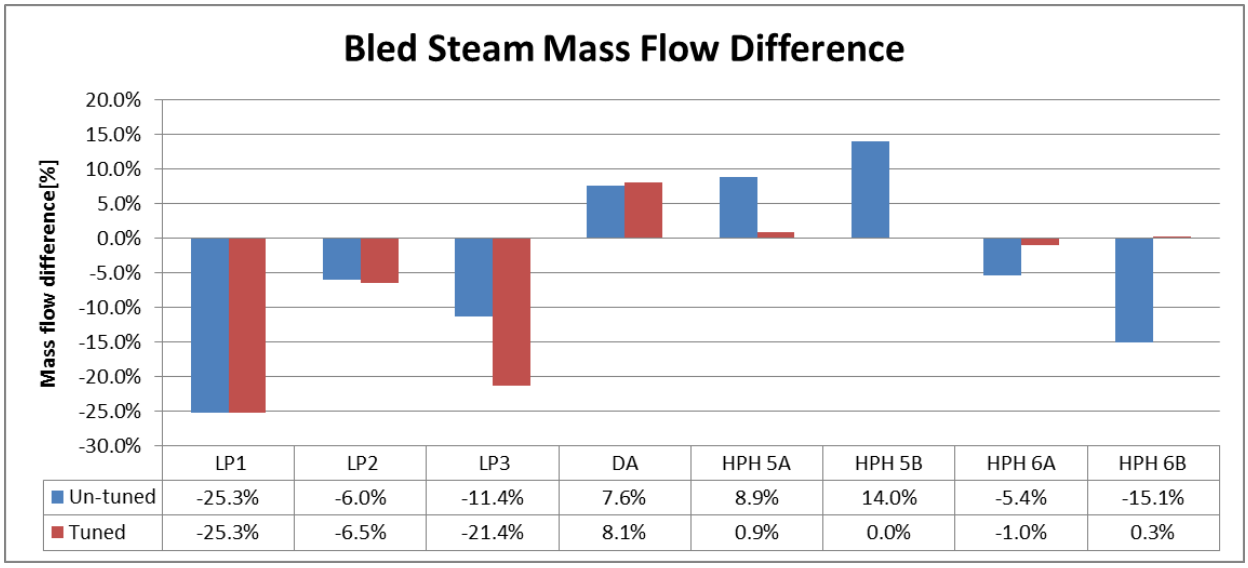
This however was not possible for the LP heaters which seem to be under predicting temperature by up to 1%. More significant is the apparent under prediction of bled steam flow (up to 25.3%). This flow difference is immense especially if one considers the higher flow when plant operation and heat balance data was compared (Figure 21– Bled steam mass flow rate difference (Actual plant data vs Heat balance data)). One way which this flow rate could be a valid is if the bled steam temperature and pressure input was lower than required. This however was not the case.

The FLOWNEX® SE prediction of bled steam flow is 25% lower than the plant data however the input parameters to the model is very close to that of heat balance inputs. Similarly the result from the FLOWNEX® SE model is also very close to that of the heat balance data. This, indicates that the FLOWNEX® SE model is consistent in its determination of the mass flow corroborating the outcome of the manual mass and energy balance verification.

There is therefore a strong possibility that the anomaly in the results is due to the credibility of the point of comparison that is the live plant data for the bled steam mass flow and not the modelling result. The interrogation of the source data entailed gaining access to the IT system for the station and searching for the specific plant parameter code or AKZ. As a result of an AKZ not being available, the conclusion was that this point was not measured and therefore a calculated point. The next step involved accessing the point via its point ID, reverse mapping the data point to its base constituents found in a calculation template and then understanding how the value was calculated.

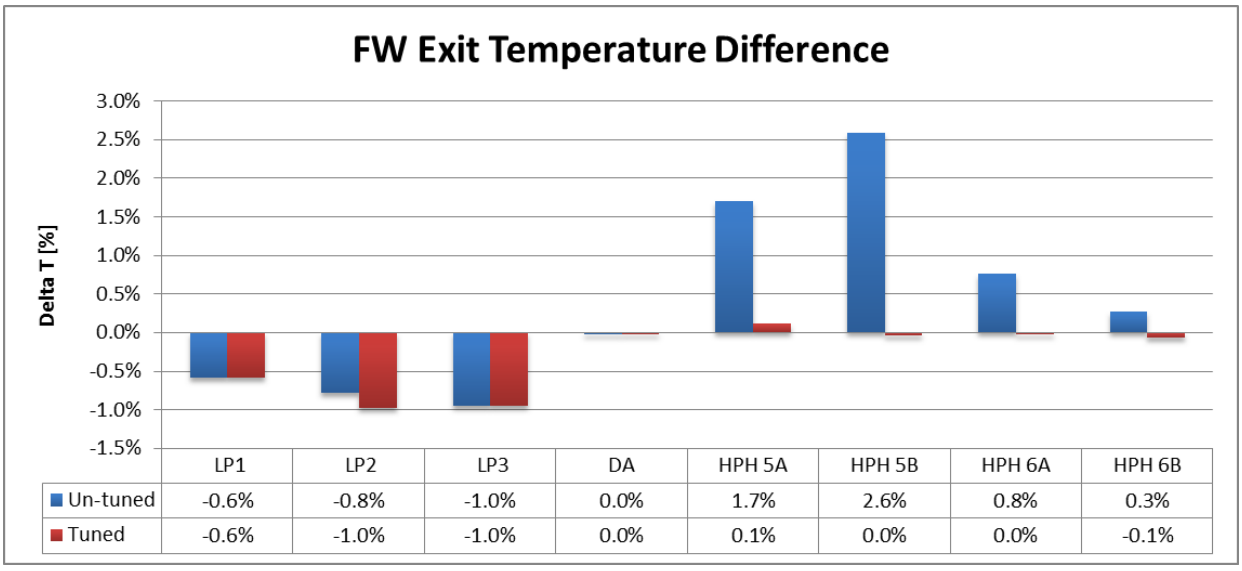
Further analysis into this point revealed that the flow was calculated based on the saturation temperature of steam. This is erroneous since the bled steam to LP heater 1 has a steam quality less than one. This is highlighted again and discussed further in the next section.

For the unit in question, actual tube plugging information from the power station was applied to the heaters. It is interesting to note that of all the heaters, only LP heater 2 had 0.3% of its tubes plugged. None of the remaining heaters had any tubes plugged.



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
Un-tuned	-25.3%	-6.0%	-11.4%	7.6%	8.9%	14.0%	-5.4%	-15.1%
Tuned	-25.3%	-6.5%	-21.4%	8.1%	0.9%	0.0%	-1.0%	0.3%

Figure 23– Feed water heater bled steam mass flow difference (un-tuned vs tuned model using actual plant data)



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
Un-tuned	-0.6%	-0.8%	-1.0%	0.0%	1.7%	2.6%	0.8%	0.3%
Tuned	-0.6%	-1.0%	-1.0%	0.0%	0.1%	0.0%	0.0%	-0.1%

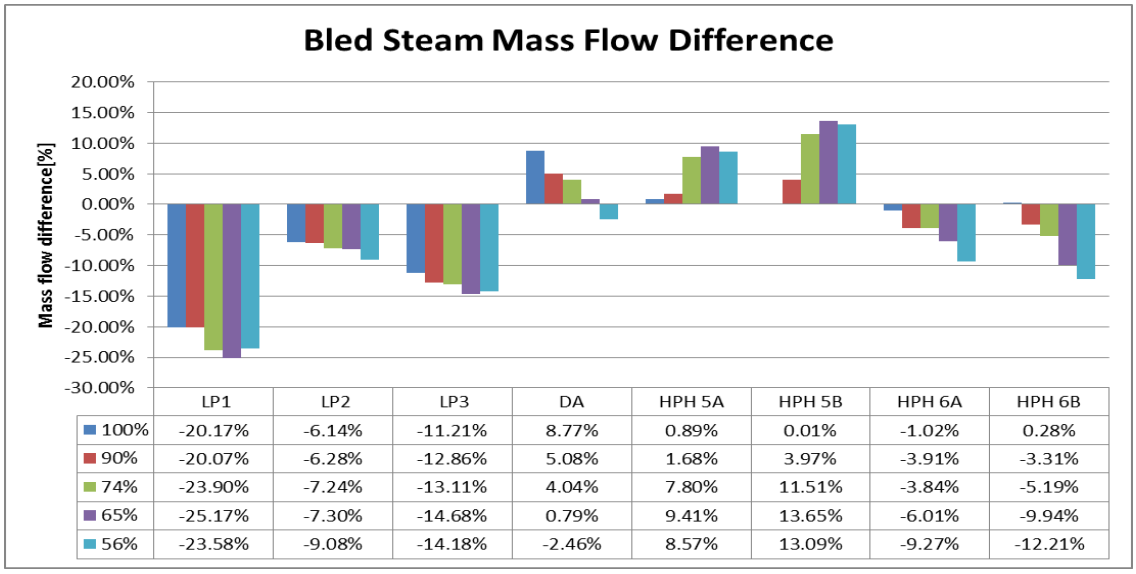
Figure 24– Feed water exit temperature difference (un-tuned vs tuned model using actual plant data)

6.4.2 Actual plant performance prediction

Live plant data of the unit was obtained directly from the power station. From the data, operating loads different to those of the heat balance diagram loads were selected. This rationale applied

was to fully assess the ability of the model to predict plant performance within the actual operating range of the plant, particularly at loads other than those used to characterise it. While the heat balance diagram loads previously tested were 100%, 80%, 60% and 40%, the unit loads selected for live plant data where 100%, 90%, 74%, 65%, 56% and 48%.

In the results below, it should be noted that the 48% load case could not be presented due to the FLOWNEX® SE model being unstable and not converging at this load. The averages of the raw data used as input to the tuned model can be found in Appendix B.



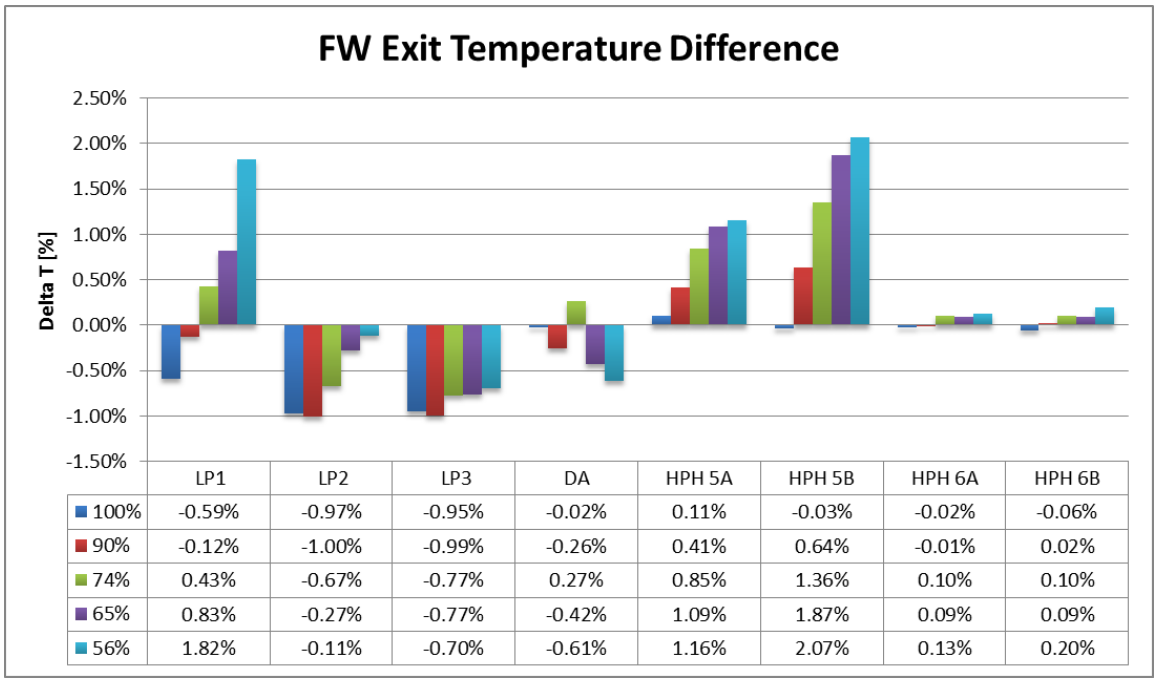
	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100%	-20.17%	-6.14%	-11.21%	8.77%	0.89%	0.01%	-1.02%	0.28%
90%	-20.07%	-6.28%	-12.86%	5.08%	1.68%	3.97%	-3.91%	-3.31%
74%	-23.90%	-7.24%	-13.11%	4.04%	7.80%	11.51%	-3.84%	-5.19%
65%	-25.17%	-7.30%	-14.68%	0.79%	9.41%	13.65%	-6.01%	-9.94%
56%	-23.58%	-9.08%	-14.18%	-2.46%	8.57%	13.09%	-9.27%	-12.21%

Figure 25– Feed water heater bled steam mass flow difference (tuned model vs actual plant performance)

It is evident from the results above that the bled steam mass flows were all within 15% difference with the exception of LP heater 1. As mentioned earlier in this chapter, the source of the bled steam mass flow data to this heater was interrogated and found to be a calculated value based on the saturation temperature of steam. It was not a measured flow nor was the calculation based on the actual bled steam temperature. It is therefore likely that the FLOWNEX® SE model is predicting closer to actual mass flow (if it were to be measured) than the results represent due to this error in the plant source data.

This postulation is further substantiated by the fact that the outlet feed water temperature of LP heater 1 as measured on the plant, and the FLOWNEX® SE modelled output are within a 2% temperature difference across all loads (see Figure 26 below). This is unlikely to be the case if the bled steam mass flow difference of 25% was true since the energy conservation law would not be obeyed. Furthermore, as a result of this flow anomaly, the performance and state of systems and equipment upstream of LP heater 1 was assessed. It was found that the last stage blade of the low pressure turbine on this specific unit was cropped. The impact of this is a reduced extraction pressure and therefore a lower saturation temperature to LP heater 1. As a result, less steam will

be extracted to the heater to raise the feed water to saturated conditions. This again points to the modelled result from FLOWNEX® SE being accurate.



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100%	-0.59%	-0.97%	-0.95%	-0.02%	0.11%	-0.03%	-0.02%	-0.06%
90%	-0.12%	-1.00%	-0.99%	-0.26%	0.41%	0.64%	-0.01%	0.02%
74%	0.43%	-0.67%	-0.77%	0.27%	0.85%	1.36%	0.10%	0.10%
65%	0.83%	-0.27%	-0.77%	-0.42%	1.09%	1.87%	0.09%	0.09%
56%	1.82%	-0.11%	-0.70%	-0.61%	1.16%	2.07%	0.13%	0.20%

Figure 26– Feed water exit temperature difference (tuned model vs actual plant performance)

The feed water outlet temperature prediction for the entire train is within 2.1% of the actual plant performance. The predicted outlet temperature for HP heater 5 however is exacerbated at load conditions 74%, 65% and 56% but still within 2.1% of the actual temperature. Taking a step back to view the FLOWNEX® SE model of the entire system as a black box, even with variability in prediction of temperature and flows of a few heaters, the final feed water temperature to the boiler was predicted with sufficient accuracy. This model was therefore deemed have an acceptable level of performance.

7. Artificial neural network performance

7.1 Training data, input and predicted variables

The block flow diagram of the feed heating train below highlights the parameters used in the training of the ANN model. Parameters in blue are inputs to the model and parameters in orange are the target parameters that will be the result of its prediction. To ensure a fair comparison to the FLOWNEX® SE model, the exact same input variables were used to predict the heater bled steam flow and exit water temperature.

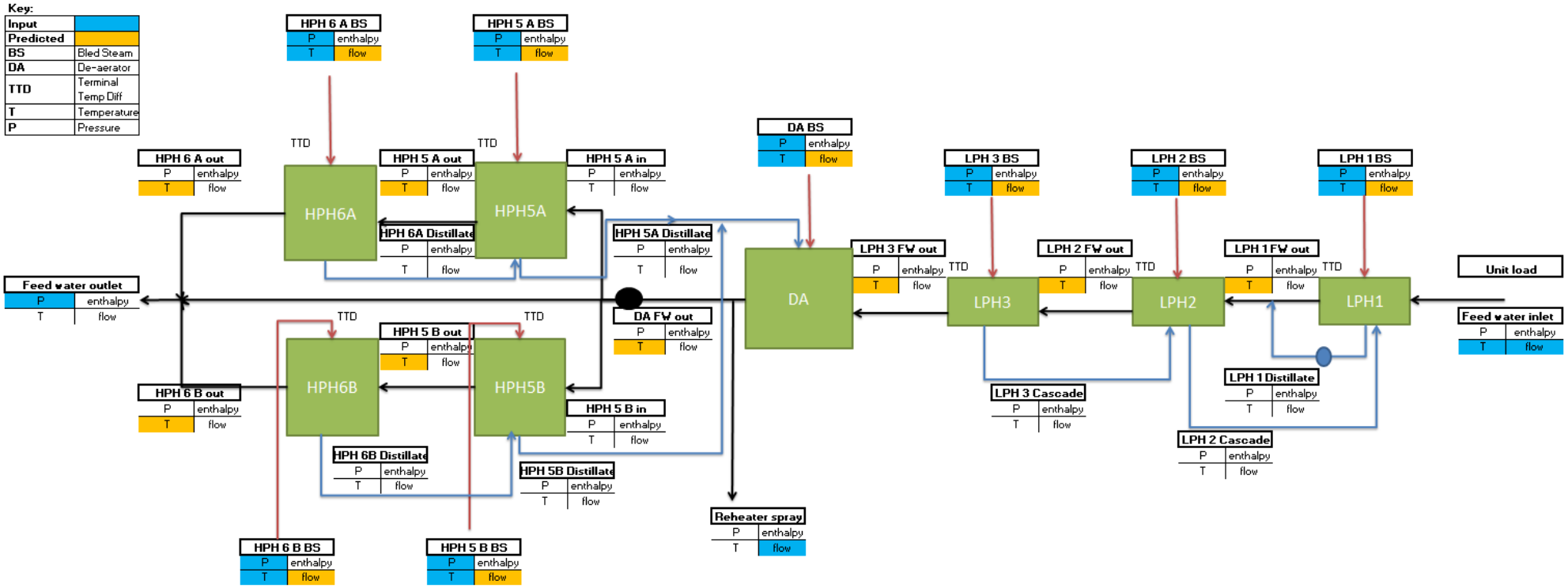


Figure 27 – Block flow diagram illustrating model input and prediction parameters

7.2 Model setup – data handling, parameter settings and architecture permutations

Following on from the method outlined in Figure 16 (Chapter 5.2.1), this section speaks to the details of each of the steps involved in building the ANN model.

The training data set was obtained directly from the power station on which this study was based. A total of 64207 data points were extracted covering a load range from 48% to 100% turbine MCR. In the data set, 20 input parameters and 16 target or prediction parameters were selected as depicted in Figure 27.

The libraries used in this model where then imported as follows:

Table 2 – Libraries imported for the ANN model

Library	Function	Purpose
pandas		reads the data from a csv file to a pandas dataframe required by the model
sklearn.model_selection	train_test_split	Splits the data set into a training set and a test set
sklearn.preprocessing	StandardScalar	Preprocessess the input data set
sklearn.metrics	mean_squared_error	Loss funciton to be employed in assessing the model performance
keras.models	Sequential	type of model used
keras.layers	Dense	type of layer used

The data set was split such that 80% was used for training the model and 20% for testing the model. The input data was preprocessed through a process known as feature scaling. The method employed was standardizing the values of each parameter such that ±68% of the values fall in the range of -1 to 1. This is simply done by:

$$x_{stand} = \frac{x - mean(x)}{Std.\ deviation} \tag{9}$$

This is a critical step since the data from the plant may be in different units. For example for the same pressure, the value in kilo Pascal (kPa) is larger than bar which is larger than Mega Pascal (MPa). The smaller unit (kPa) will consequently have a larger variation in the data set. Therefore, to the ANN model, the value in kilo Pascal is statistically more relevant than that of Mega Pascal due to the Euclidean distance being greater for the data in kilo Pascal versus Mega Pascal.

The ANN itself was then constructed in four steps. First the network was initialized as a sequential model. Next the network was defined in terms of number of layers, nodes per layer, weight and bias initialisation and activation function. The number of layers and nodes per layer were determined through iteration as per Table 4 below. Each layer is defined as a dense layer meaning that it is a fully connected network (every node in a layer is connected to every node in the next layer). The activation function used was ReLU and the weight initialisation method was glorot

normal (Xavier initialisation). This is recommended when the ReLU activation function is used. Weight initialisation is important since it provides a more appropriate starting point for training, which is ultimately the exercise of finding the optimum set of weights in the network. Inappropriate weight initialisation can cause the model not to converge or to converge at low accuracy (model underfitting).

The third step was to compile the model. This was done by specifying the optimisation algorithm to be used, Adam, and the mean squared error as the loss function. This is a loss function typically used with regression type problems. The Adam optimiser has a set of hyper parameters that can be adjusted if the model performance requires improvement. Table 3 below illustrates the parameters and the default values for each.

Table 3 – Adam optimiser hyperparameters

Parameter	**Value**	**Purpose**
ε	1 x 10^-8	Set to be very close to zero to prevent divide by 0 error
β_1	0.9	Hyperparameter used in the calcualting the momentum
β2	0.999	Hyperparameter used in the calcualting the momentum
α	0.001	Learning rate, the step size taken with every update of each weight

The final step was to fit the model. Here the training data was split further into 80% for training and 20% for validation. Validation is the assessment of the model's predictive capability which is executed at the same time as the model is being trained. In this step the batch size and number of epochs was also determined through iteration. Table 4 below lists the permutations of layers, nodes per layer, batch size and number of epochs tested. From the first nine combinations of layers and nodes, the best performing model was selected. In this case combination 6.1.0 was the best performing model (discussed in the following section). This model was then tested with varying batch size and epochs.

Table 4 – Model permutations

Combination	Layers	Nodes	Batch size	Epochs
1.1.0	1	100	1000	100
2.1.0	1	200	1000	100
3.1.0	2	50,50	1000	100
4.1.0	2	100,100	1000	100
5.1.0	2	200,200	1000	100
6.1.0	3	100,100,100	1000	100
7.1.0	3	100,100,8	1000	100
8.1.0	4	100,100,100,8	1000	100
9.1.0	4	100,100,100,100	1000	100
6.1.1	3	100,100,100	32	100
6.1.2	3	100,100,100	500	100
6.1.3	3	100,100,100	2000	100
6.1.4	3	100,100,100	500	10
6.1.5	3	100,100,100	500	50
6.1.6	3	100,100,100	500	200

7.3 Model architecture selection

Each model was run 10 times and the minimum, maximum and average accuracy for both the training and validation set was recorded as well as the MSE loss. Accuracy is a predefined metric within Keras and is defined as the mean accuracy rate across all predictions. As per Figure 28 and Table 5 below, model combination 6.1.0 was the best performing model since this architecture achieved the highest average accuracy and lowest mean squared error on the training and validation set. It also produced the lowest error on the test and prediction set (live plant data) making it clear that the prediction capability and generalisation ability of combination 6.1.0 was the best among all architectures tested.

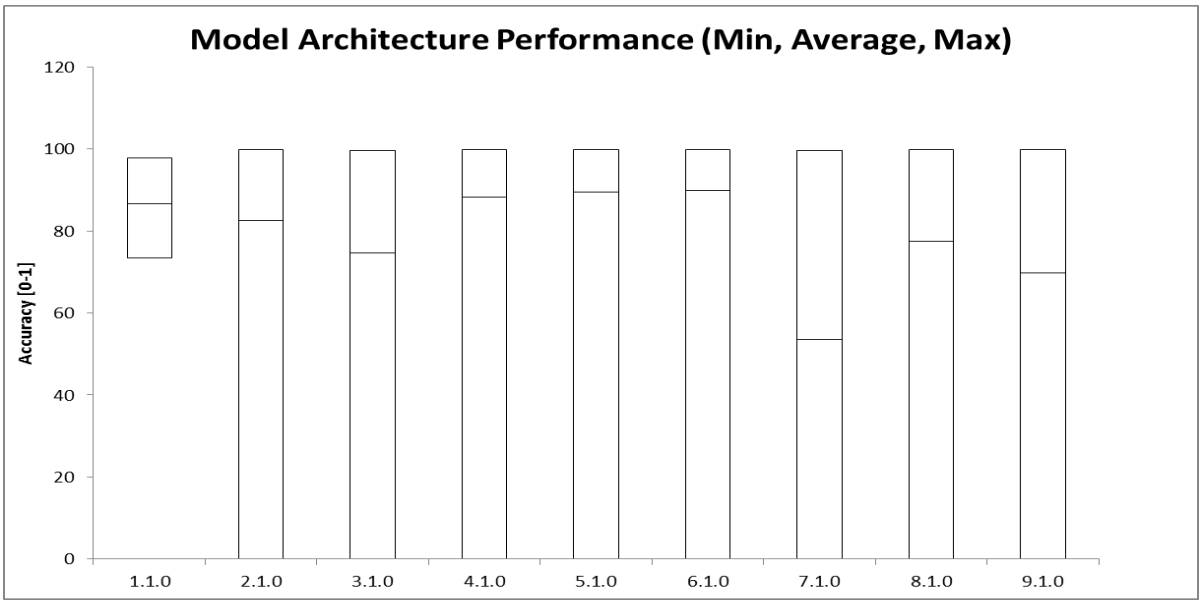


Figure 28 – Minimum, average and maximum model accuracy

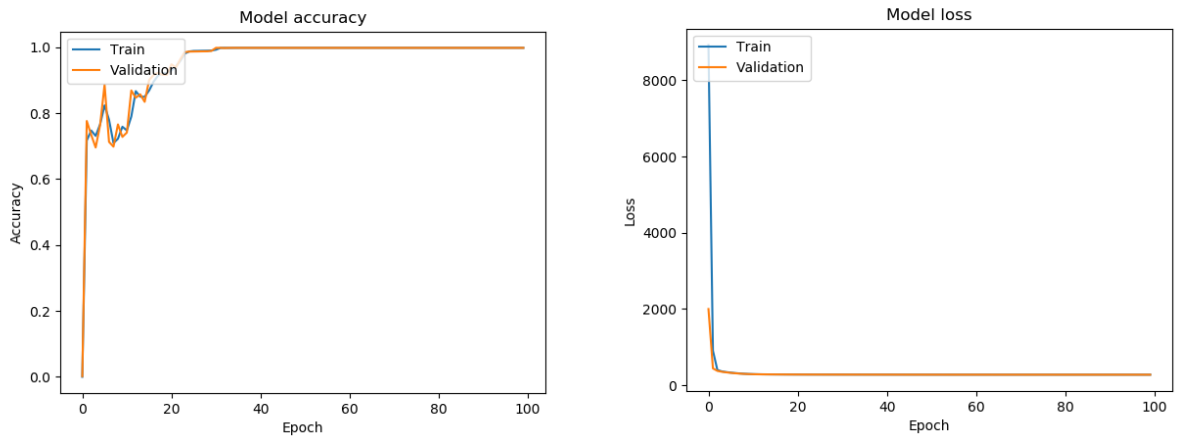
It is important to note that all architectures, with the exception of 1.1.0 have minimum values of 0% accuracy. This is because in at least one of the 10 runs, the model produced an accuracy of 0%. This highlights a level of instability in in each of these architectures. Therefore, as a matter of practice when using the model, one should run the selected model multiple times and use the outputs from the best performing run.

As per Table 5 below, Combination 6.1.0 was then adjusted for different batch sizes per epoch. It is clear that the lower the batch size, the better the model is trained and the better the accuracy and mean squared error loss on the test and prediction sets. The architecture with a batch size of 500 produced a higher accuracy with a comparably low mean squared error to the architecture with a batch size of 1000. The architecture with a batch size of 32 proved to be quite unstable and did not perform as well as expected.

Table 5 – Model architecture performance results

Combination	Layers	Nodes	Batch size	Epochs	Train acc	Val acc	Train MSE	Val MSE	Test MSE	Pred MSE
1.1.0	1	100	1000	100	86.69	86.59	16.40	15.90	16.20	1.86
2.1.0	1	200	1000	100	82.63	82.66	19.04	18.70	18.78	9.77
3.1.0	2	50,50	1000	100	74.57	74.51	7.08	6.97	7.22	0.96
4.1.0	2	100,100	1000	100	88.25	88.22	13.60	13.60	13.53	10.59
5.1.0	2	200,200	1000	100	89.54	89.51	2.80	2.75	2.80	0.76
6.1.0	3	100,100,100	1000	100	89.78	89.82	0.72	0.71	0.73	0.27
7.1.0	3	100,100,8	1000	100	70.24	70.52	1269.40	1267.60	1263.80	1245.10
8.1.0	4	100,100,100,8	1000	100	77.54	77.29	1228.90	1228.50	1225.60	1206.90
9.1.0	4	100,100,100,100	1000	100	69.85	69.84	309.28	309.13	308.08	302.91
6.1.1	3	100,100,100	32	100	97.63	97.82	17.24	17.21	17.03	16.45
6.1.2	3	100,100,100	500	100	99.36	99.74	0.31	0.31	0.32	0.47
6.1.3	3	100,100,100	2000	100	83.40	83.50	33.88	33.80	33.60	29.57
6.1.4	3	100,100,100	500	10	88.60	88.50	37.67	34.90	34.49	12.20
6.1.5	3	100,100,100	500	50	97.25	97.81	48.35	48.30	47.90	44.40
6.1.6	3	100,100,100	500	200	99.80	99.80	613.98	614.06	612.30	605.22

Using a batch size of 500, which is combination 6.1.2, the number of epochs was assessed. From Table 5 it is evident that the lower the number of epochs, the lower the model accuracy due to insufficient training. Having twice the number of epochs compared to combination 6.1.2 marginally improved the model's performance but substantially increased the model running time (from an average of 30 seconds to 85 seconds) and computational intensity. Thus the optimum model and architecture combination was 6.1.2. From Figure 29 below, it is clear that the model has a high accuracy for both the training and validation set therefore it is not underfitting. Furthermore, there no split in loss or accuracy between the training and validation set meaning there is no overfitting. The model is therefore expected to have good generalisation when used for prediction.



(a) *(b)*

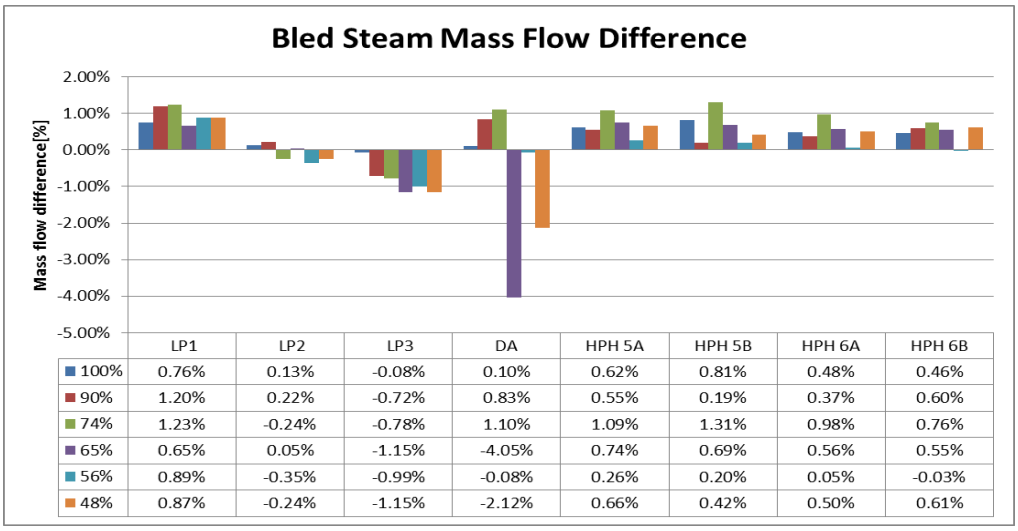
Figure 29 – (a) Model accuracy vs epochs, (b) model loss vs epochs

One of the ideas tested was the combination 7.1.0 and 8.1.0 which was deliberately structured to have the number of nodes in the last hidden layer correlate to the number of heaters in the feed

water heating train. Theory suggests that the initial hidden layers capture the abstract features of the input information while the latter hidden layers capture more holistic information. The hypothesis was that if the last hidden layer consisted of the same number of nodes as there are heaters in the train, each node would capture the features of each individual heater. From the results however, this does not appear to be the case.

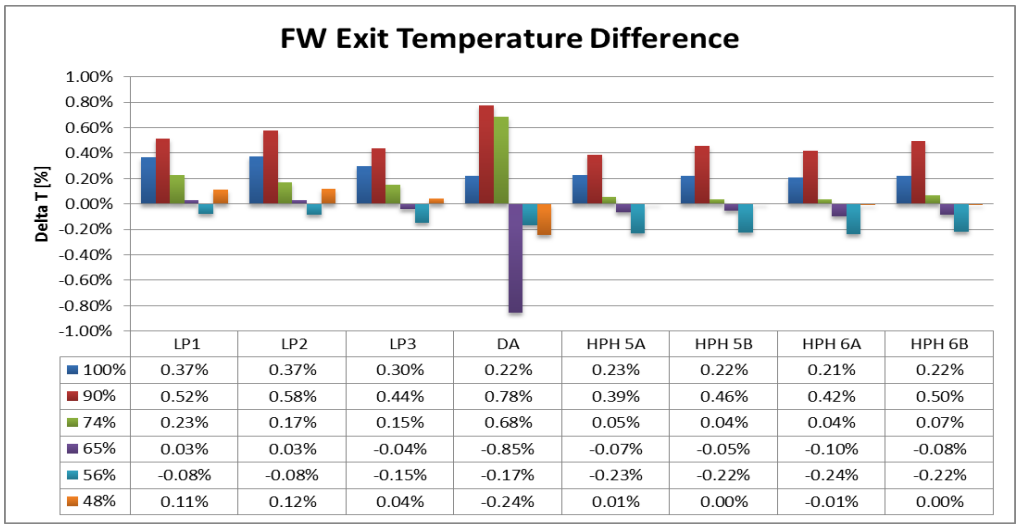
7.4 Model prediction performance

The model was run with the prediction data set which essentially contained the same input cases as that for the FLOWNEX® SE model. It is evident from the graphs below that the artificial neural network was able to predict the bled steam mass flow rate within 4% of the actual value. With respect to the feed water heater outlet temperature, it predicted within 1% of the actual value.



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100%	0.76%	0.13%	-0.08%	0.10%	0.62%	0.81%	0.48%	0.46%
90%	1.20%	0.22%	-0.72%	0.83%	0.55%	0.19%	0.37%	0.60%
74%	1.23%	-0.24%	-0.78%	1.10%	1.09%	1.31%	0.98%	0.76%
65%	0.65%	0.05%	-1.15%	-4.05%	0.74%	0.69%	0.56%	0.55%
56%	0.89%	-0.35%	-0.99%	-0.08%	0.26%	0.20%	0.05%	-0.03%
48%	0.87%	-0.24%	-1.15%	-2.12%	0.66%	0.42%	0.50%	0.61%

Figure 30 - Feed water heater bled steam mass flow difference (ANN model vs actual plant performance)



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100%	0.37%	0.37%	0.30%	0.22%	0.23%	0.22%	0.21%	0.22%
90%	0.52%	0.58%	0.44%	0.78%	0.39%	0.46%	0.42%	0.50%
74%	0.23%	0.17%	0.15%	0.68%	0.05%	0.04%	0.04%	0.07%
65%	0.03%	0.03%	-0.04%	-0.85%	-0.07%	-0.05%	-0.10%	-0.08%
56%	-0.08%	-0.08%	-0.15%	-0.17%	-0.23%	-0.22%	-0.24%	-0.22%
48%	0.11%	0.12%	0.04%	-0.24%	0.01%	0.00%	-0.01%	0.00%

Figure 31 - Feed water heater outlet temperature difference (ANN model vs actual plant performance)

7.5 ANN and FLOWNEX® SE model comparison

The graphs below compare the actual predicted values of the FLOWNEX® SE model and the ANN model to the actual plant behaviour for both bled steam mass flow and feed heater outlet temperature. It is evident that the ANN model more accurately predicts the bled steam flow to each heater than the FLOWNEX® SE model. Both the FLOWNEX® SE model and the ANN are able to accurately predict the outlet water temperature of each heater. However, the FLOWNEX® SE mode is able to predict the outlet temperature of the final heater stage (6A and 6B) with greater accuracy. The FLOWNEX® SE model was not able to converge for the 48% load case hence the absence of results. This highlights the ability of the ANN model to operate well at any load range or at least any load range for which was trained.

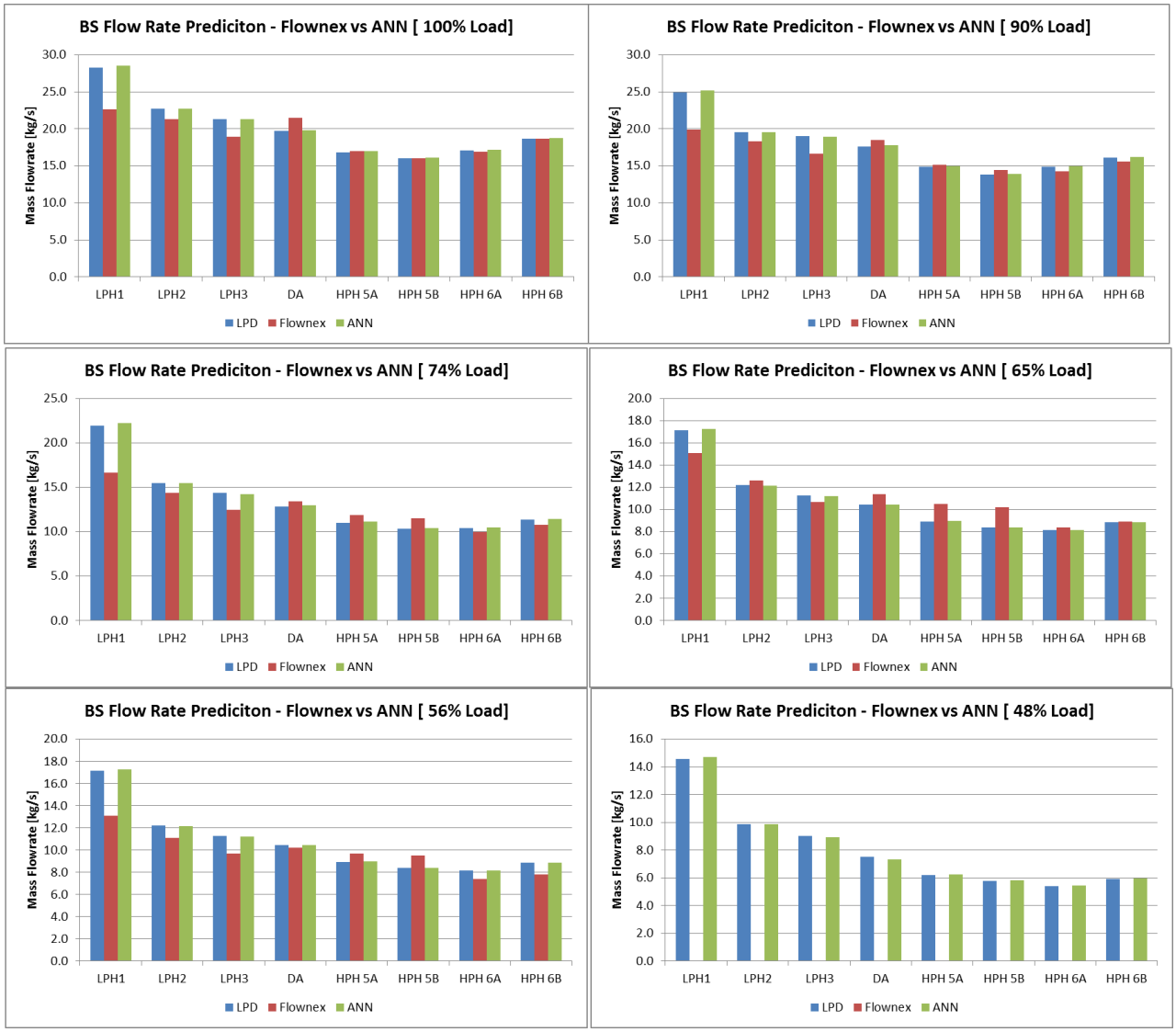


Figure 32 – Bled steam mass flow prediction comparison of FLOWNEX® SE and ANN to actual plant for six load cases

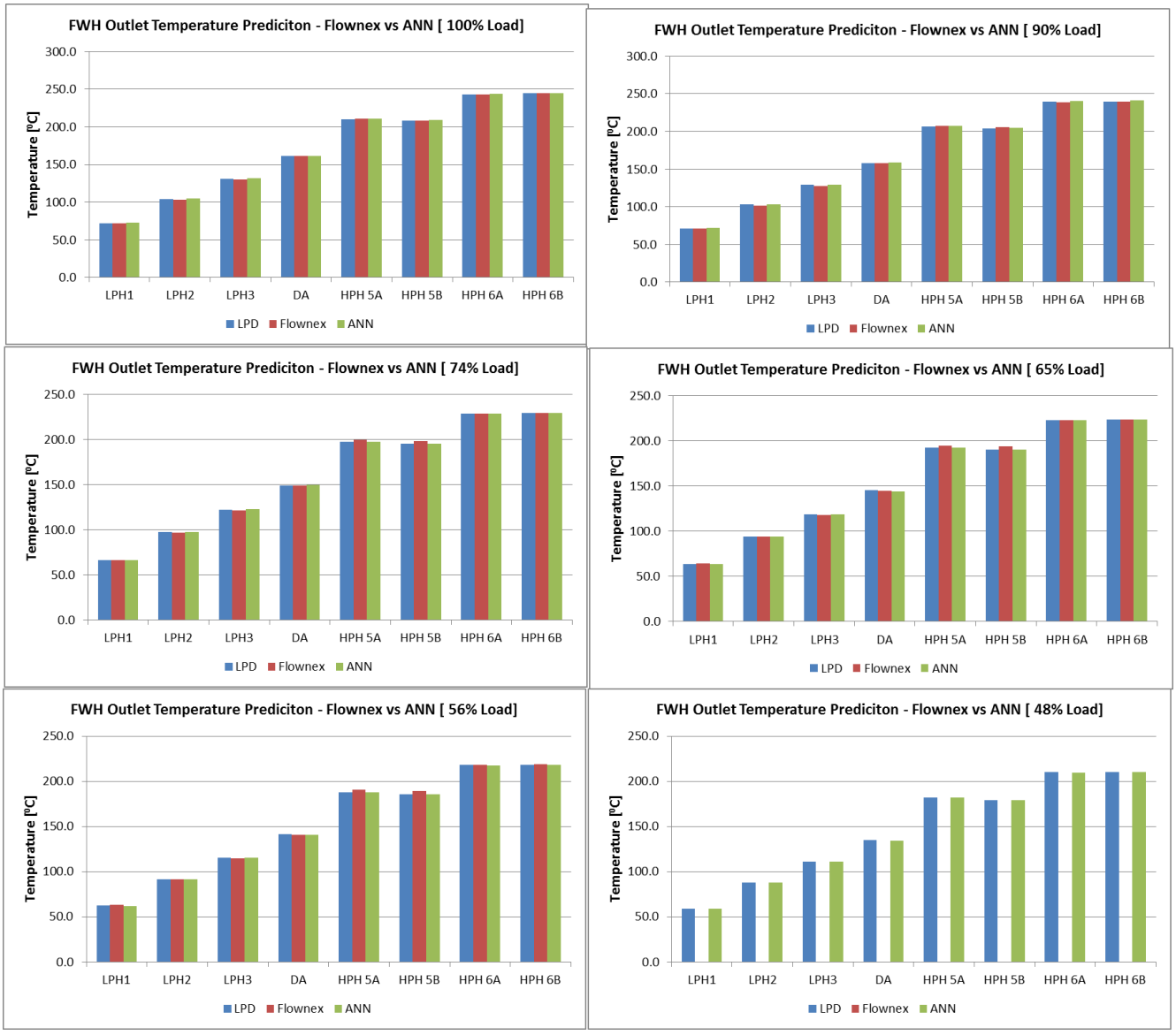


Figure 33 – Feed heater outlet temperature prediction comparison of FLOWNEX® SE and ANN to actual plant for six load cases

Table 6 – Model architecture 6.1.2 bled steam mass flow comparison – low pressure heaters

	LP1			LP2			LP3			DA		
Load	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN
100%	28.3	22.6	28.5	22.7	21.3	22.7	21.3	18.9	21.3	19.8	21.5	19.8
90%	24.9	19.9	25.2	19.5	18.3	19.6	19.0	16.6	18.9	17.6	18.5	17.8
74%	21.9	16.7	22.2	15.5	14.4	15.5	14.4	12.5	14.3	12.9	13.4	13.0
65%	20.2	15.1	20.3	13.6	12.6	13.6	12.5	10.7	12.4	11.3	11.4	10.9
56%	17.1	13.1	17.3	12.2	11.1	12.2	11.3	9.7	11.2	10.5	10.2	10.4
48%	14.6	0.0	14.7	9.9	0.0	9.9	9.0	0.0	8.9	7.5	0.0	7.3

Table 7 – Model architecture 6.1.2 bled steam mass flow comparison – high pressure heaters

	HPH 5A			HPH 5B			HPH 6A			HPH 6B		
Load	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN
100%	16.9	17.0	17.0	16.0	16.0	16.1	17.1	16.9	17.2	18.6	18.7	18.7
90%	14.9	15.1	14.9	13.9	14.4	13.9	14.9	14.3	14.9	16.1	15.6	16.2
74%	11.0	11.9	11.2	10.3	11.5	10.4	10.4	10.0	10.5	11.4	10.8	11.5
65%	9.6	10.5	9.7	9.0	10.2	9.0	8.9	8.4	9.0	9.9	8.9	9.9
56%	8.9	9.7	9.0	8.4	9.5	8.4	8.2	7.4	8.2	8.9	7.8	8.9
48%	6.2	0.0	6.2	5.8	0.0	5.8	5.4	0.0	5.4	5.9	0.0	6.0

Table 8 – Model architecture 6.1.2 feed water outlet temperature comparison – low pressure heaters

	LP1			LP2			LP3			DA		
Load	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN
100%	72.3	71.9	72.6	104.5	103.5	104.9	131.2	130.0	131.6	161.1	161.1	161.5
90%	71.5	71.4	71.9	102.8	101.8	103.4	129.0	127.7	129.5	158.0	157.6	159.2
74%	66.4	66.7	66.6	97.5	96.8	97.6	122.4	121.5	122.6	148.8	149.2	149.8
65%	63.9	64.4	63.9	94.3	94.0	94.3	118.8	117.9	118.8	145.1	144.5	143.9
56%	62.6	63.7	62.5	91.9	91.8	91.8	115.9	115.1	115.7	141.6	140.7	141.3
48%	59.2	0.0	59.3	88.0	0.0	88.1	111.2	0.0	111.2	135.1	0.0	134.8

Table 9 – Model architecture 6.1.2 feed water outlet temperature comparison – high pressure heaters

	HPH 5A			HPH 5B			HPH 6A			HPH 6B		
Load	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN
100%	210.3	210.5	210.8	208.5	208.4	208.9	243.4	243.4	244.0	244.7	244.5	245.2
90%	206.4	207.3	207.2	204.2	205.5	205.1	239.1	239.1	240.1	239.9	240.0	241.1
74%	197.6	199.3	197.7	195.3	197.9	195.3	228.7	228.9	228.8	229.4	229.6	229.5
65%	192.6	194.7	192.5	190.0	193.6	189.9	222.8	223.0	222.6	223.4	223.6	223.2
56%	188.4	190.6	188.0	186.0	189.8	185.5	218.2	218.5	217.7	218.6	219.0	218.1
48%	182.2	0.0	182.2	179.5	0.0	179.5	210.1	0.0	210.1	210.5	0.0	210.5

It is important to note that while the credibility of the raw plant data around LP heater 1 bled steam is questionable, as highlighted with the FLOWNEX® SE model, the ANN model predicts the bled steam mass flow within 0.7% of the plant data. This brings to light the fact that the physics of the plant is not necessarily captured by the ANN model in the way that the FLOWNEX® SE model did, thereby providing the insight into the data quality issue. The ANN model simply finds patterns in the data, thus enabling it to make a prediction whether the data is physically realistic or not.

8. Artificial neural network performance with updated training data

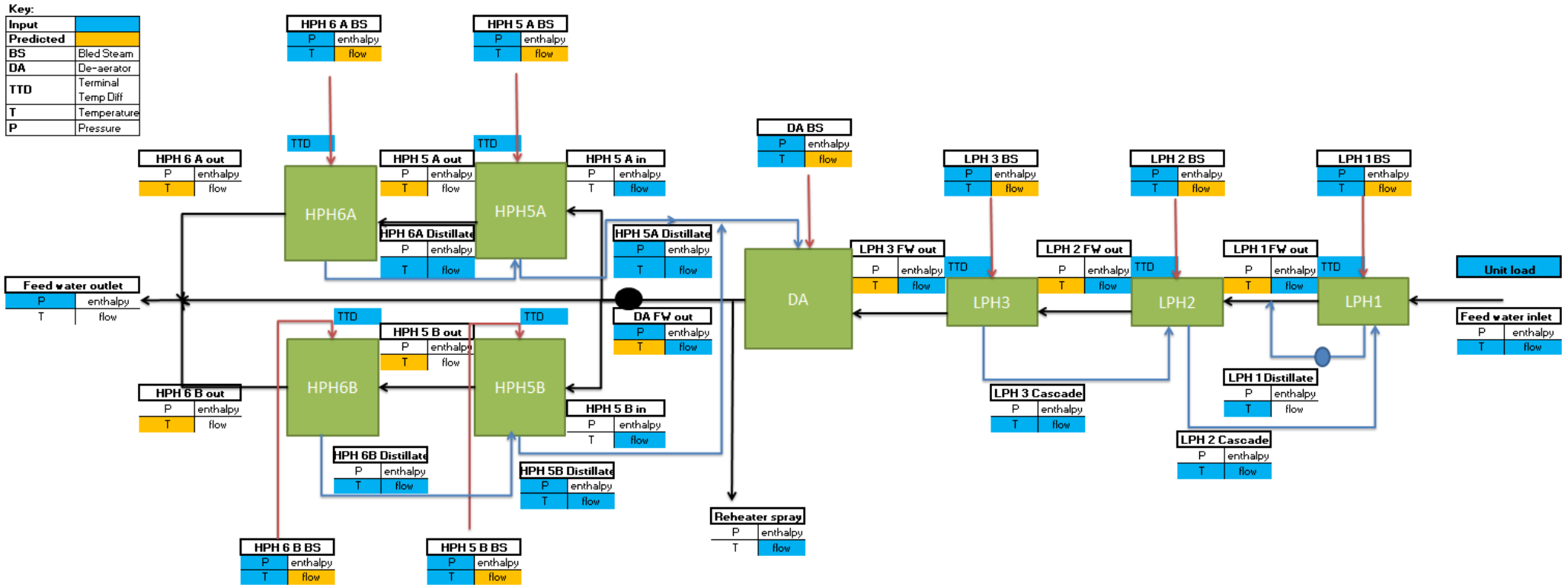


Figure 34 - Block flow diagram illustrating new model input and prediction parameters

8.1 Model input data expansion

As a result of the previous ANN models' limitation in capturing the physics of the plant, an attempt was made to improve its performance by increasing the number of input variables to the model. This is a reasonable approach since in the previous chapter, the comparison of the FLOWNEX® SE model and the ANN model was not strictly a fair one. The FLOWNEX® SE model intrinsically makes use of fundamental mass and energy balance equations that produce intermediate values such as feed water exit flow rate, distillate temperature, cascade temperature and terminal temperature differences of each heater. This is information that the previous ANN model did not have and thus could be the cause of its limitation of not capturing the physics of the plant.

It is important to note that the additional input variables used in this data set specifies the necessary information for the energy conservation laws to be satisfied for each heater. In this light, it can be perceived that the ANN model was training itself on the pattern of the energy conservation equation. Therefore providing the ANN model with this additional intermediate data actually makes the comparison to the FLOWNEX® SE model more even. In the new data set, 50 input parameters were used, incorporating the available intermediate raw plant data. The number of prediction parameters as well as the total number of data points remains unchanged.

8.2 Model architecture selection

The same method, as conducted in chapter 7.2 was carried out with the new, more comprehensive data set. Table 10 below lists the permutations of layers, nodes per layer, batch size and number of epochs tested. From the first nine combinations of layers and nodes, the best performing model was selected. In this case combination 5.1.0 was the best performing model (discussed in the following section). This model was then tested with varying batch size and epochs.

Table 10 – Model permutations

Combination	Layers	Nodes	Batch size	Epochs
1.1.0	1	100	1000	100
2.1.0	1	200	1000	100
3.1.0	2	50,50	1000	100
4.1.0	2	100,100	1000	100
5.1.0	2	200,200	1000	100
6.1.0	3	100,100,100	1000	100
7.1.0	3	100,100,8	1000	100
8.1.0	4	100,100,100,8	1000	100
9.1.0	4	100,100,100,100	1000	100
5.1.1	2	200,200	32	100
5.1.2	2	200,200	500	100
5.1.3	2	200,200	2000	100
5.1.4	2	200,200	1000	10
5.1.5	2	200,200	1000	50
5.1.6	2	200,200	1000	200

Each model was run 10 times and the minimum, maximum and average accuracy for both the training and validation set was recorded as well as the MSE loss. As per Figure 35 and Table 11 below, model combination 5.1.0 and 9.1.0 where the best performing models with 9.1.0 marginally outperforming 5.1.0. Models 5.1.0 and 9.1.0 where then used to make a prediction on the test set and the live plant date set. The mean squared error values of the two where compared to ascertain the best performing model. As per Table 5 below, it is clear that the prediction capability and generalisation ability of combination 5.1.0 was superior to 9.1.0 based on a lower error from the test and prediction set.

Compared to the models in chapter 7.3, these models with the additional information are deemed more stable since only 3 of the 9 models gave a 0% accuracy reading (vs 8 of 9 in chapter 7.3).

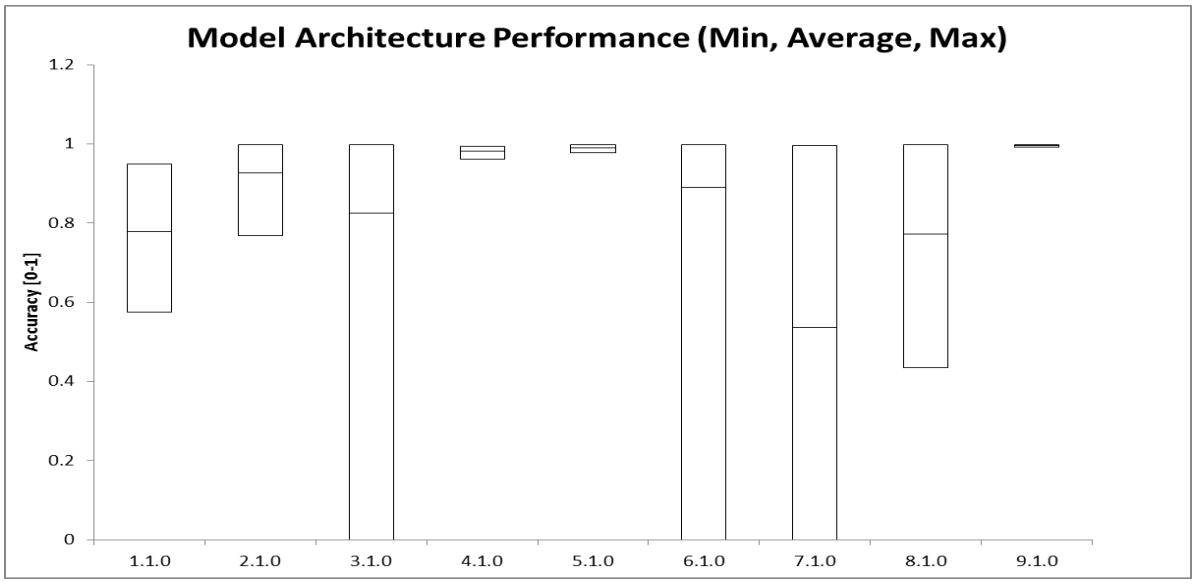


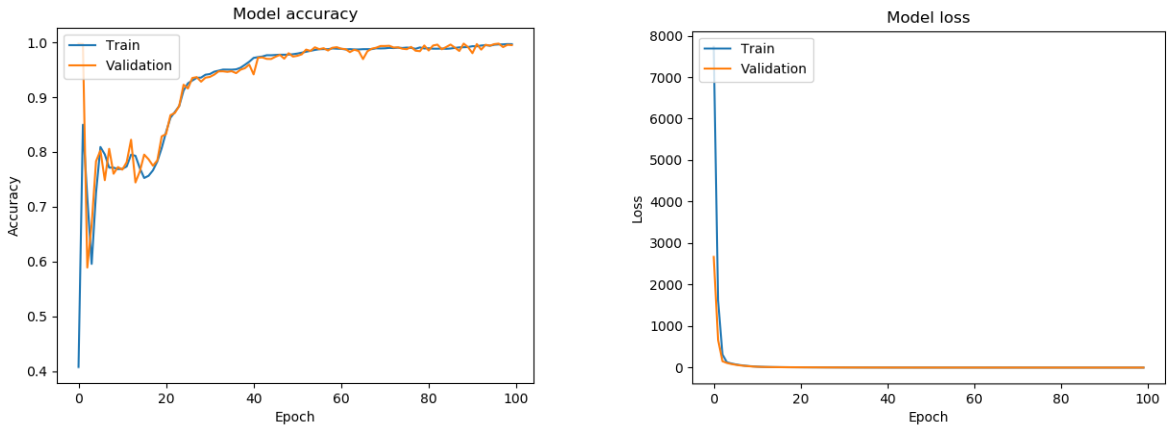
Figure 35 – Minimum, average and maximum model accuracy

Table 11 – Model architecture performance results

Combination	Layers	Nodes	Batch size	Epochs	Train acc	Val acc	Train MSE	Val MSE	Test MSE	Pred MSE
1.1.0	1	100	1000	100	0.7786	0.7772	22.33	22.09	22.22	1.47
2.1.0	1	200	1000	100	0.9259	0.9252	13.13	12.87	12.97	1.97
3.1.0	2	50,50	1000	100	0.8244	0.8229	3.69	3.59	3.63	2.20
4.1.0	2	100,100	1000	100	0.9820	0.9819	1.71	1.67	1.68	1.02
5.1.0	2	200,200	1000	100	0.9907	0.9909	0.91	0.72	0.91	0.71
6.1.0	3	100,100,100	1000	100	0.8894	0.8895	10.06	10.19	10.11	9.82
7.1.0	3	100,100,8	1000	100	0.5368	0.5308	3344.67	3343.81	3336.48	3299.21
8.1.0	4	100,100,100,8	1000	100	0.7723	0.7730	4719.58	4718.35	4706.09	4684.17
9.1.0	4	100,100,100,100	1000	100	0.9958	0.9932	5719.71	5658.32	5467.00	5466.00
5.1.1	2	200,200	32	100	0.9980	0.9980	0.06	0.03	0.07	0.23
5.1.2	2	200,200	500	100	0.9975	0.9975	0.41	0.37	0.23	0.45
5.1.3	2	200,200	2000	100	0.9974	0.9955	1373.21	1372.81	1367.00	1382.00
5.1.4	2	200,200	500	10	0.8860	0.8820	31.49	28.39	40.84	17.14
5.1.5	2	200,200	500	50	0.9816	0.9783	1.84	1.95	1.93	1.65
5.1.6	2	200,200	500	200	0.9976	0.9975	0.09	0.71	0.08	0.46

Combination 5.1.0 was then adjusted for different batch sizes per epoch. From Table 11, it is clear that the lower the batch size, the better the model is trained and the better the accuracy and mean squared error loss on the test and prediction sets. Ultimately, a batch size of 500 was chosen since the batch size of 32 was too computationally intensive for the hardware used on this project, taking three times longer to complete a model run.

Using a batch size of 500, which is combination 5.1.2, the number of epochs was assessed. From Table 11Table 5 it is evident that lower the number of epochs, the lower the model performance due to insufficient training. Having twice the number of epochs compared to combination 5.1.2 marginally improved the models performance but substantially increased the model running time and computational intensity. Thus the optimum model and architecture combination was 5.1.2. From Figure 36 below, it is clear that the model has a high accuracy for both the training and validation set therefore it is no underfitting. Furthermore there no split in loss or accuracy between the training and validation set meaning there is no overfitting. The model is therefore expected to have good generalisation when used for prediction. The ANN program code for this optimised architecture can be found in Appendix C.



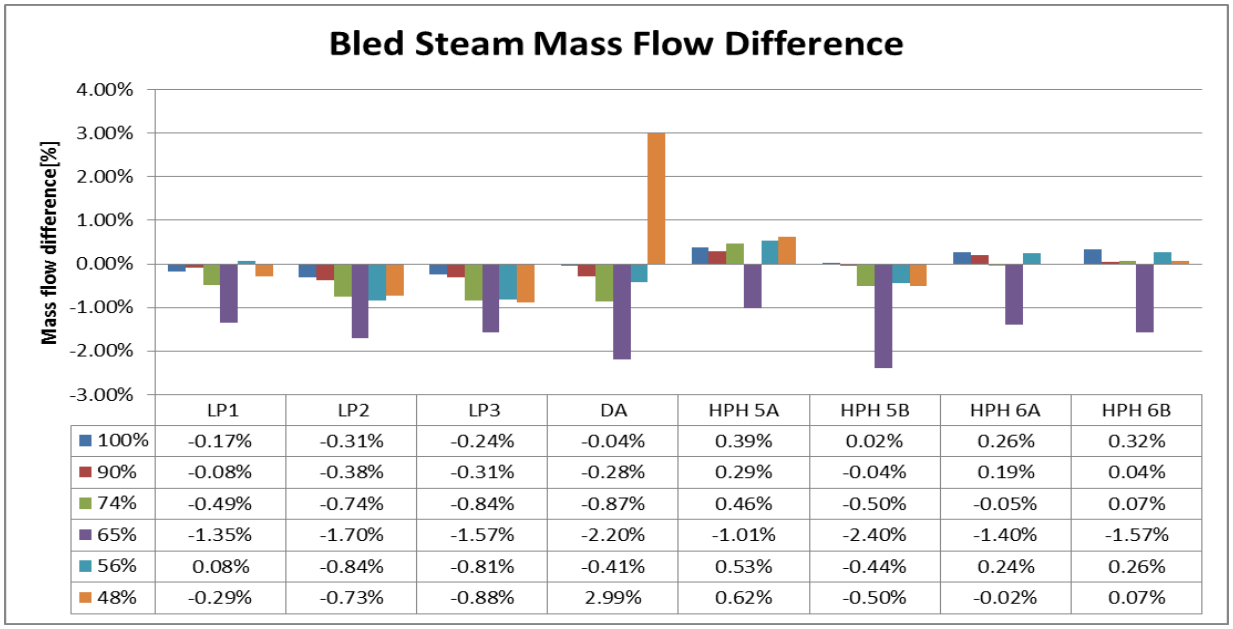
(a) *(b)*

Figure 36 – (a) Model accuracy vs epochs, (b) model loss vs epochs

It is important to note that the optimum model with this data set, comprising a larger number of input variables, has two hidden layers and 200 nodes per layer. The optimum model for the previous data set (used in chapter 7) with fewer input variables had three hidden layers and 100 nodes per layer. It is evident that, due to having less information for training, a model of greater complexity was required to extract the features and patterns in the data. This made it possible to perform comparably to the simpler model produced when more information was available.

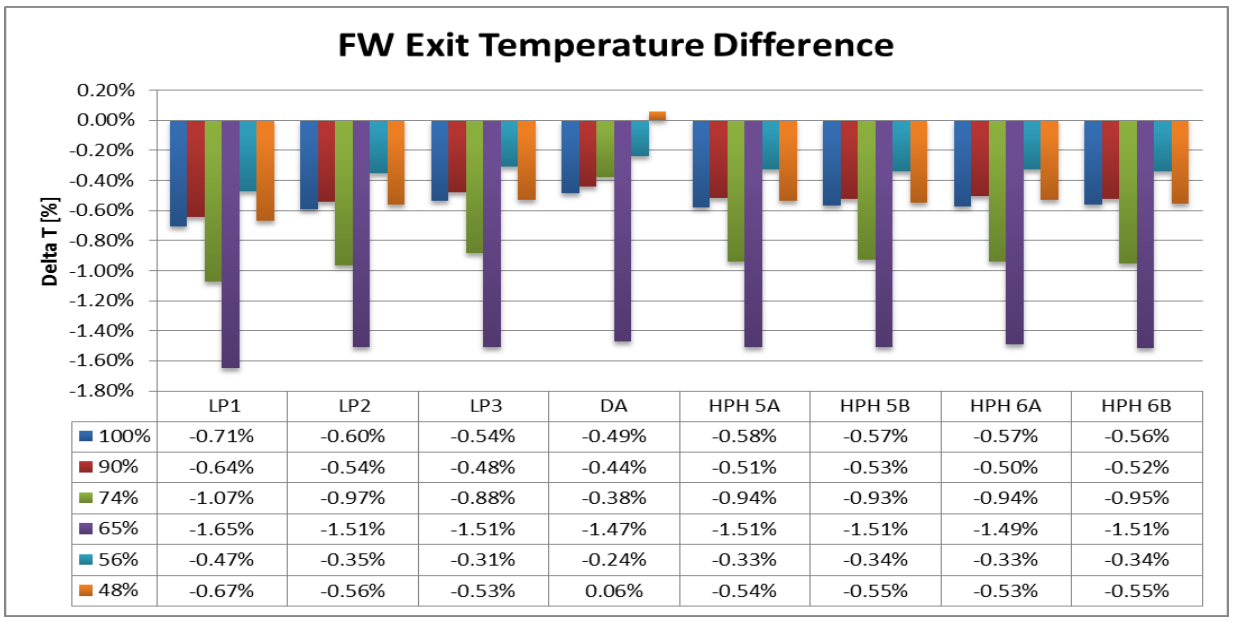
8.3 Model prediction performance

The model was run with the new prediction data set containing the additional intermediate input variables. It is evident from the graphs below that the artificial neural network predicted the bled steam mass flow rate within 3% of the actual values. With respect to the feed water heater outlet temperature, it was under predicting by up to 1.65% depending on the load.



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100%	-0.17%	-0.31%	-0.24%	-0.04%	0.39%	0.02%	0.26%	0.32%
90%	-0.08%	-0.38%	-0.31%	-0.28%	0.29%	-0.04%	0.19%	0.04%
74%	-0.49%	-0.74%	-0.84%	-0.87%	0.46%	-0.50%	-0.05%	0.07%
65%	-1.35%	-1.70%	-1.57%	-2.20%	-1.01%	-2.40%	-1.40%	-1.57%
56%	0.08%	-0.84%	-0.81%	-0.41%	0.53%	-0.44%	0.24%	0.26%
48%	-0.29%	-0.73%	-0.88%	2.99%	0.62%	-0.50%	-0.02%	0.07%

Figure 37 - Feed water heater bled steam mass flow difference (ANN model vs actual plant performance)



	LP1	LP2	LP3	DA	HPH 5A	HPH 5B	HPH 6A	HPH 6B
100%	-0.71%	-0.60%	-0.54%	-0.49%	-0.58%	-0.57%	-0.57%	-0.56%
90%	-0.64%	-0.54%	-0.48%	-0.44%	-0.51%	-0.53%	-0.50%	-0.52%
74%	-1.07%	-0.97%	-0.88%	-0.38%	-0.94%	-0.93%	-0.94%	-0.95%
65%	-1.65%	-1.51%	-1.51%	-1.47%	-1.51%	-1.51%	-1.49%	-1.51%
56%	-0.47%	-0.35%	-0.31%	-0.24%	-0.33%	-0.34%	-0.33%	-0.34%
48%	-0.67%	-0.56%	-0.53%	0.06%	-0.54%	-0.55%	-0.53%	-0.55%

Figure 38 - Feed water heater outlet temperature difference (ANN model vs actual plant performance)

The results above are based on an optimised data set where the input and output parameters where carefully and deliberately selected as stated in section 8.1. However, a separate and an

unrefined data set was employed in an earlier iteration of this model. This made use of inter-stage feed water temperatures as inputs (even though it is a prediction parameter) and feed heater distillate temperatures as predicted parameter. This essentially meant that there were more training and prediction parameters in total. Interestingly, the model performance was significantly poorer in that the best test and prediction set mean squared error was 11.72 and 2.34 respectively (vs 0.23 and 0.45 for the optimised model 5.1.2).

Further to this, the two best performing model architectures with the non-optimised data set happened to be a model with one hidden layer and 100 nodes and the other with two hidden layers and 100 nodes per layer. This again highlights that with more input information, model complexity tends to decrease.

Furthermore, it became evident that an anomaly around the open feed water heater (the de-aerator) existed in that the prediction of both the bled steam mass flow and temperature was off by a substantial amount (33.1%), as can be seen in Figure 39 below.

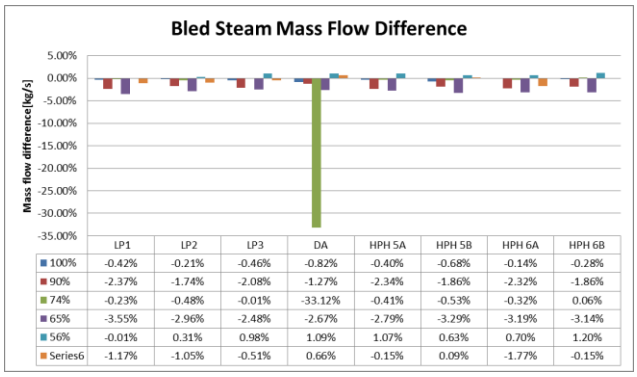


Figure 39 – Model prediction comparison using model with 1 layer and 200 nodes per layer (non-optimised data set)

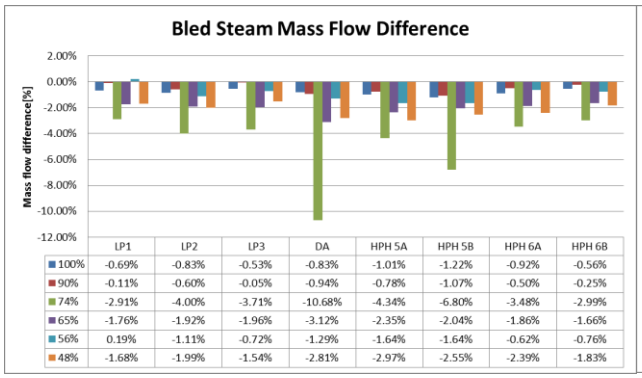


Figure 40 - Model prediction comparison using model with 2 layers and 100 nodes per layer (non-optimised data set)

This arbitrary scenario serves to highlight that when the model complexity was increased, that is using a deeper (more layers) and wider (more nodes per layer) network the anomaly of the de-aerator was significantly reduced (Figure 40). The bled steam mass flow difference reduced from -33.12% to -10.12% whereas the feed water outlet temperature difference reduced from -6.3% to -

4.5%. This is as a result of a model of greater complexity being able to more accurately extract the features from the training data than a model with lower complexity. This further substantiates the strategy of increasing the number of features used in the model as method of trying to capture the physics of the process.

8.4 ANN and FLOWNEX® SE model comparison

The graphs below compare the prediction of the FLOWNEX® SE model and the ANN model to the actual plant behaviour for both bled steam mass flow and feed heater outlet temperature. It is evident that the ANN model behaves in an identical manner as the previous ANN model which had less input variables.

Figure 41 – Bled steam mass flow prediction comparison of FLOWNEX® SE and ANN to actual plant for six load cases

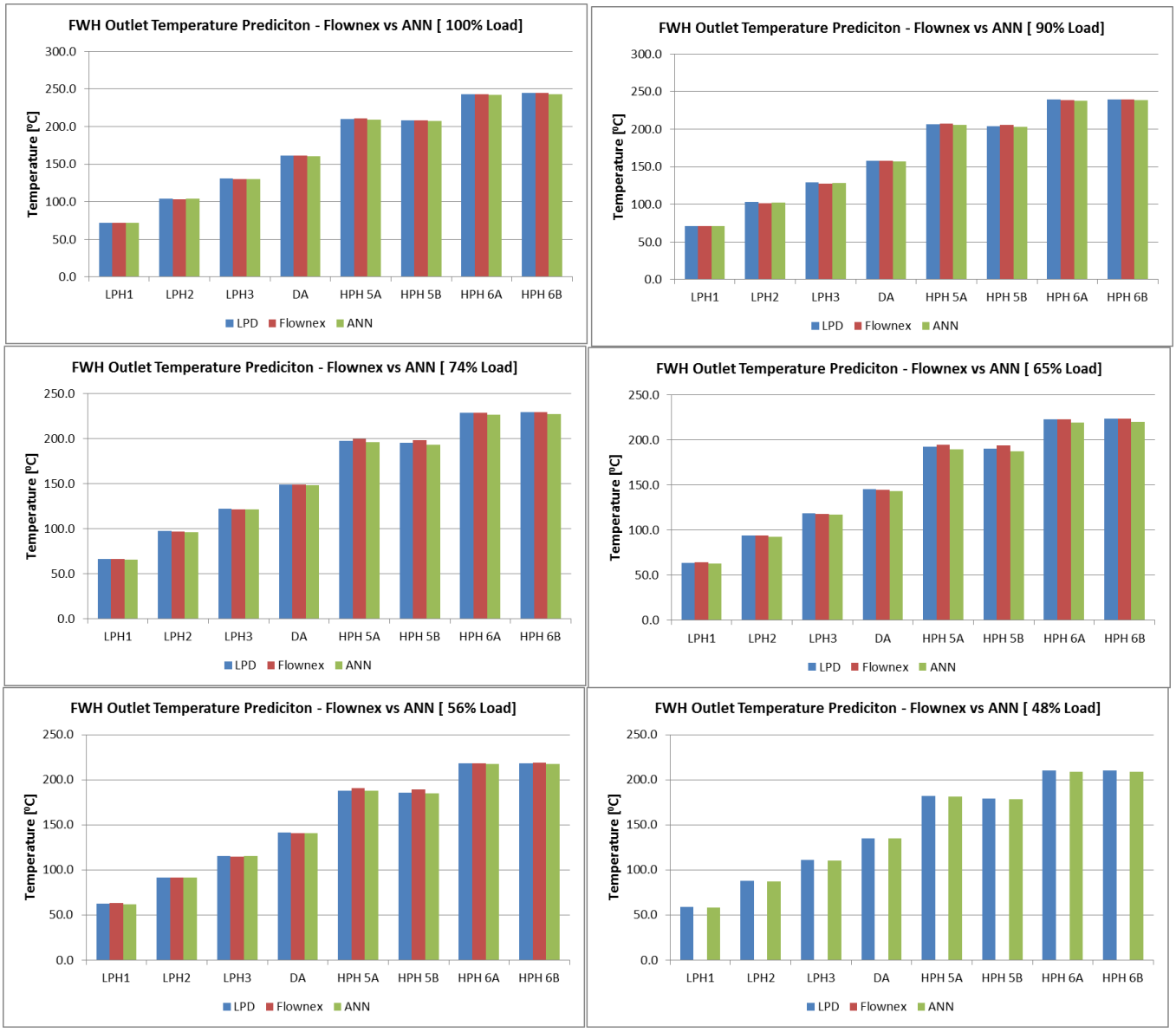


Figure 42 – Feed heater outlet temperature prediction comparison of FLOWNEX® SE and ANN to actual plant for six load cases

Table 12 - Model architecture 5.1.2 bled steam mass flow comparison – low pressure heaters

	LP1			LP2			LP3			DA		
Load	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN
100%	28.3	22.6	28.3	22.7	21.3	22.6	21.3	18.9	21.2	19.8	21.5	19.8
90%	24.9	19.9	24.9	19.5	18.3	19.5	19.0	16.6	19.0	17.6	18.5	17.6
74%	21.9	16.7	21.8	15.5	14.4	15.4	14.4	12.5	14.3	12.9	13.4	12.8
65%	20.2	15.1	19.9	13.6	12.6	13.4	12.5	10.7	12.3	11.3	11.4	11.1
56%	17.1	13.1	17.2	12.2	11.1	12.1	11.3	9.7	11.2	10.5	10.2	10.4
48%	14.6	0.0	14.5	9.9	0.0	9.8	9.0	0.0	8.9	7.5	0.0	7.7

Table 13 - Model architecture 5.1.2 bled steam mass flow comparison – high pressure heaters

	HPH 5A			HPH 5B			HPH 6A			HPH 6B		
Load	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN
100%	16.9	17.0	16.9	16.0	16.0	16.0	17.1	16.9	17.1	18.6	18.7	18.7
90%	14.9	15.1	14.9	13.9	14.4	13.8	14.9	14.3	14.9	16.1	15.6	16.1
74%	11.0	11.9	11.1	10.3	11.5	10.3	10.4	10.0	10.4	11.4	10.8	11.4
65%	9.6	10.5	9.5	9.0	10.2	8.8	8.9	8.4	8.8	9.9	8.9	9.7
56%	8.9	9.7	9.0	8.4	9.5	8.4	8.2	7.4	8.2	8.9	7.8	8.9
48%	6.2	0.0	6.2	5.8	0.0	5.8	5.4	0.0	5.4	5.9	0.0	5.9

Table 14 – Model architecture 5.1.2 feed water outlet temperature comparison – low pressure heaters

	LP1			LP2			LP3			DA		
Load	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN
100%	72.3	71.9	71.8	104.5	103.5	103.9	131.2	130.0	130.5	161.1	161.1	160.3
90%	71.5	71.4	71.0	102.8	101.8	102.3	129.0	127.7	128.4	158.0	157.6	157.3
74%	66.4	66.7	65.7	97.5	96.8	96.5	122.4	121.5	121.4	148.8	149.2	148.2
65%	63.9	64.4	62.8	94.3	94.0	92.8	118.8	117.9	117.0	145.1	144.5	143.0
56%	62.6	63.7	62.3	91.9	91.8	91.6	115.9	115.1	115.6	141.6	140.7	141.2
48%	59.2	0.0	58.8	88.0	0.0	87.5	111.2	0.0	110.6	135.1	0.0	135.2

Table 15 – Model architecture 5.1.2 feed water outlet temperature comparison – low pressure heaters

	HPH 5A			HPH 5B			HPH 6A			HPH 6B		
Load	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN	LPD	Flownex	ANN
100%	210.3	210.5	209.1	208.5	208.4	207.3	243.4	243.4	242.1	244.7	244.5	243.3
90%	206.4	207.3	205.4	204.2	205.5	203.1	239.1	239.1	237.9	239.9	240.0	238.7
74%	197.6	199.3	195.8	195.3	197.9	193.4	228.7	228.9	226.5	229.4	229.6	227.2
65%	192.6	194.7	189.7	190.0	193.6	187.2	222.8	223.0	219.5	223.4	223.6	220.0
56%	188.4	190.6	187.8	186.0	189.8	185.3	218.2	218.5	217.5	218.6	219.0	217.8
48%	182.2	0.0	181.2	179.5	0.0	178.5	210.1	0.0	209.0	210.5	0.0	209.3

With the addition of more input information, the ANN still does not capture the physics of the plant and predicts the bled steam flow of LP heater 1 exactly. It was postulated that the additional input variables used in this data set, specified the necessary information for the energy conservation laws to be satisfied for each heater and that it could be perceived that the ANN model was training itself on the pattern of the energy conservation equation. However, given the ANN prediction for LP heater 1, this cannot be the case. Furthermore, from a data science perspective, theory suggests that with more training data and features (variables), the model generalisation and hence predication capability should improve however it was still unable to highlight the anomaly around LP heater 1.

It stands to reason that the ANN model prediction is only as valid as the input training data it receives. If the training data are erroneous, the prediction, while seemingly accurate, will actually be incorrect in real terms as highlighted by the insight provided by the FLOWNEX® SE model. The integrity of the training data set is therefore critical to the validity of the ANN prediction.

9. Model extrapolation capability

Chapters 6 and 8 demonstrated the ability of the FLOWNEX® SE model and the ANN model respectively, to accurately predict plant behaviour under expected operating conditions as experienced on a day to day basis at a power station. In some instances however, stations experiences out of normal conditions that significantly affect the performance of the feed water heating system and by extension the power generation cycle. It is therefore useful to predict such deviations in performance to have the foresight in terms of its impact on plant capacity and overall thermal efficiency. This chapter highlights the results of five specifically selected out of normal cases (described in section 5.3) when applied to both the FLOWNEX® SE model and the ANN model.

The decision was taken to implement these cases to the high pressure heaters since deviation in its performance has a more pronounced effect on the system at large. This is because when a low pressure heater underperforms, given its position in the train, there are many subsequent heaters that will "make-up" the deficit in performance by increasing the heat transferred in that heater. With there being only two high pressure heaters in parallel at this particular station, there is less opportunity, or none if it is the last heater, to make up the deficit before it feeds into the boiler.

The decision was also taken to perform these simulations at 100% MCR (with the exception of case 4) since it at this load that all major components of the power station operate at the upper end of its capacity. Therefore any deficit in the final temperature produced by the feed heating train will cause the downstream systems and components to be pushed closer to its limits by doing more work. In some cases, these limits are reached which results in the loss of generating capability, commonly known as a partial load loss (PLL) or an unplanned capability loss factor (UCLF). Among other plant areas, the systems that most regularly reach their maximum allowable operating capacity are the milling plant and the draught group (primary and secondary input air fans and/ or induced draught fans).

9.1 Model extrapolation performance

Case 1 was designed to simulate a dirty (fouled) heat exchanger. In this scenario, all four high pressure heaters were fouled. The expected effect of fouling in the heater is the reduction of heat transferred from the bled steam to the water. Hence the outlet temperature of the water is expected to decrease. Fouling on a feed water heater is typically indicated by higher than normal TTD values. Hence, this condition was simulated in both cases by manipulating the TTD of the heater to 20°C.

From Figure 43 below, it is evident that the FLOWNEX® SE predicted that the temperature of the water exiting each high pressure heater decreases. Thus the FLOWNEX® model behaves as expected.

With the ANN model, the temperature is firstly seen to increase which is the contrary to how a heat exchanger is expected to behave. Secondly, it increases past the point of any achievable temperature in this heat exchanger. Furthermore, by making a change to the high pressure heaters, the ANN model predicts that the low pressure heaters, including the de-aerator, will also achieve incredibly high feed water exit temperatures. In the actual process, upstream low pressure heaters should not be affected by fouling on the high pressure heaters. This again highlights the inability of this ANN model to capture the physics of the feed heating process and is a consistent behaviour encountered in the subsequent test cases.

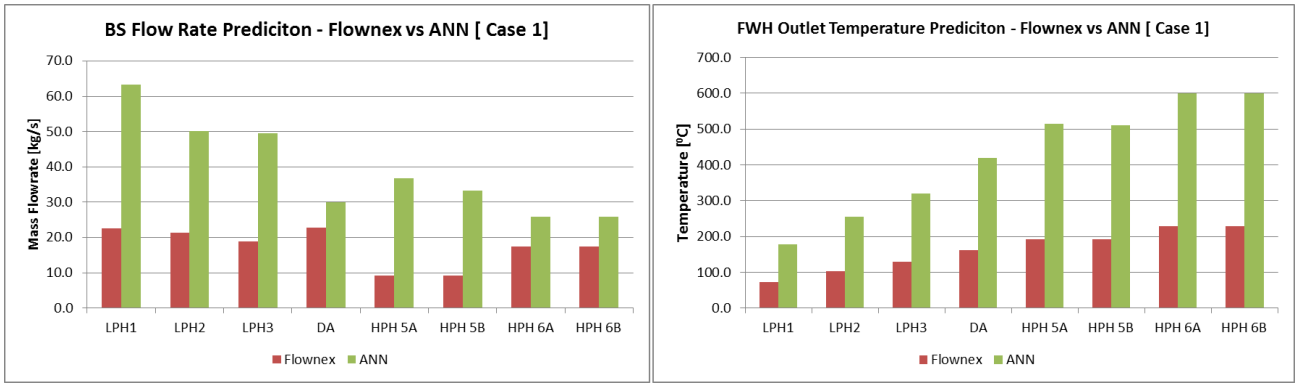


Figure 43 – Extrapolation case 1: Predicted bled steam flow and feed heater exit temperature

Case 2 was designed to simulate a single high pressure heater out of service. This was done first for heater 5A and then for 6A independently while the heaters on the B bank remained in service.

With heater 5A out of service, the temperature of the water exiting the heater should increase only marginally as a result of the heat transferred from the distillate cascading from heater 6A. One would also expect the mass flow of bled steam to increase in heater 6A to make up the deficit in temperature. This effect was seen by the FLOWNEX® SE model as depicted in Figure 44.

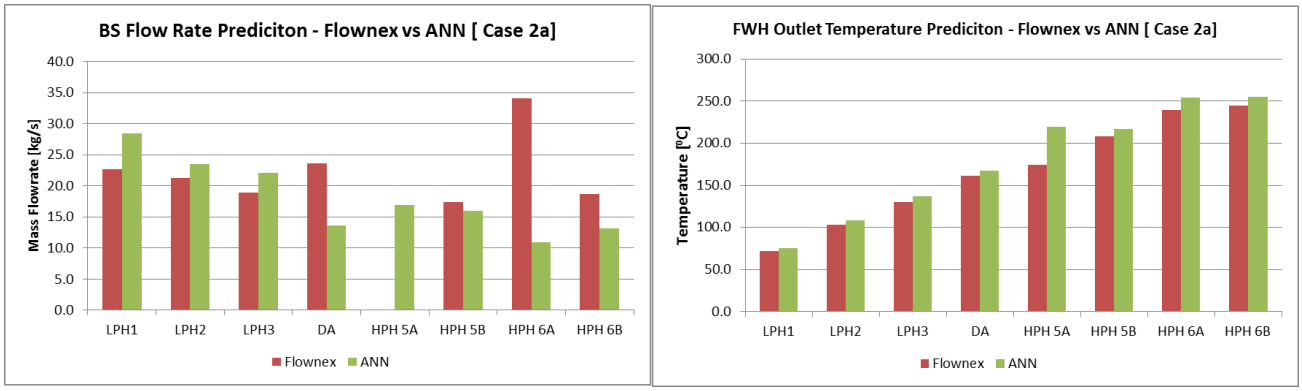


Figure 44 - Extrapolation case 2a: Predicted bled steam flow and feed heater exit temperature

The ANN model however does not see the heater as out of service when given a feed water mass flow rate of 0 kg/s. Once again, predicted temperatures and bled steam mass flows upstream and downstream of the heater, as well as for the heater itself, do not correspond to the expected plant behaviour.

Similar trends can be seen when the last heater in the train, 6A, is taken out of service as per Figure 45 below. Where the heater outlet temperature should be roughly equal to the inlet temperature (as correctly predicted by FLOWNEX® SE), the ANN model firstly predicts a temperature increase. Secondly it predicts an outlet temperature of heater 6A above that which heater is capable of at full load conditions.

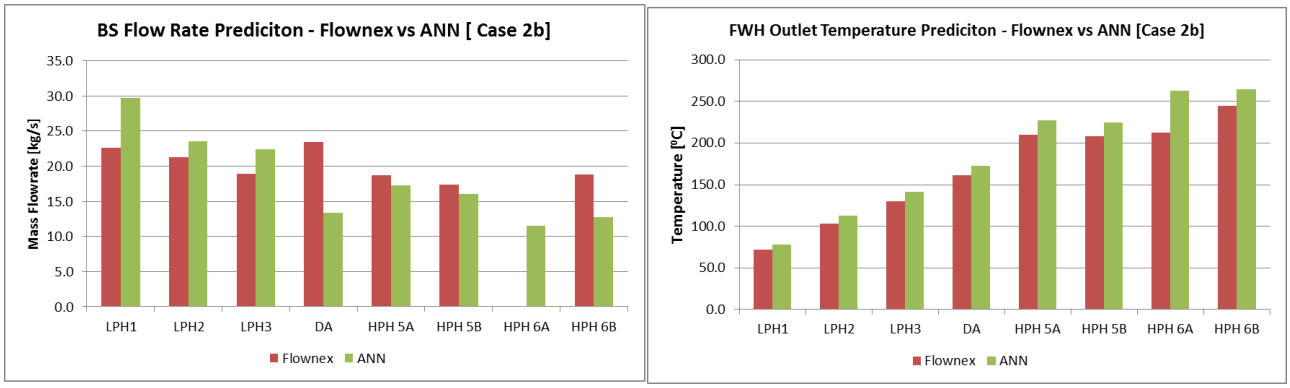


Figure 45 - Extrapolation case 2b: Predicted bled steam flow and feed heater exit temperature

Case 3 was a combination of case 2a and 2b where both HP heaters 5A and 6A were taken out of service at the same time by bypassing the heater bank. In such a case, it is expected that the bank in service (B bank) will operate as normal while the outlet temperature of the bank out of service will be close to the deaerator outlet temperature with the exception of some losses.

It was difficult to simulate this scenario in FLOWNEX® SE due to the pressure calculation convergence being very sensitive. This simulation was ultimately achieved by removing the heaters entirely from the model. The result was an accurate prediction for both mass flow and

temperature, in line with expected plant behaviour. Comparing this to the ANN model, one can see from the figure below that the ANN model once again performed inconsistently to the expected plant behaviour.

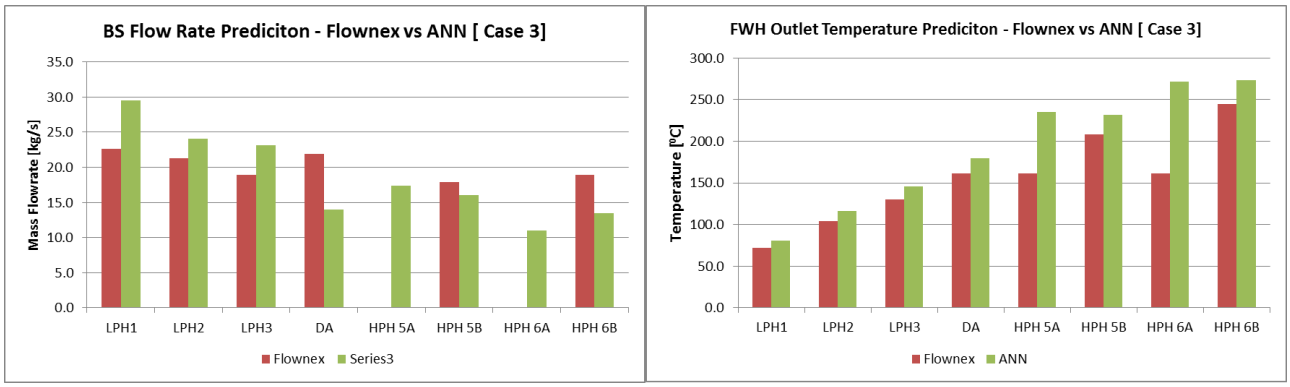


Figure 46 - Extrapolation case 3: Predicted bled steam flow and feed heater exit temperature

Case 4 served to assess the behaviour of both models at a load of 38% with heater 6A and 6B out of service. In this case, the FLOWNEX® SE model was unable to converge and therefore did not produce any results. The ANN was able to produce results albeit inaccurate.

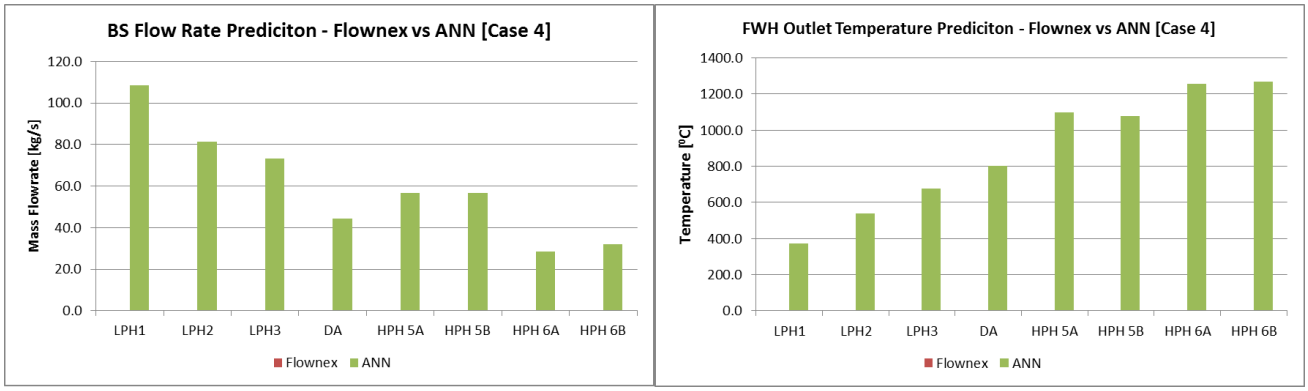


Figure 47 - Extrapolation case 4: Predicted bled steam flow and feed heater exit temperature

9.2 ANN model fine tuning for improved performance

9.2.1 Alternate activation function

In an attempt to improve the ANN model performance with the out of normal cases, three concepts where tested. First was the use of the leaky ReLU activation function in place of ReLU. The reason for this is that ReLU is sometimes capable of creating dead neurons. This means that through the training process, some of the nodes or neurons in the hidden layers become inactive as a result of their output being consistently 0. Due to the ReLU activation function producing a zero value for any non-positive input, these nodes cannot become active again. Leaky ReLU overcomes this since the output for non-positive inputs is a small negative output as per Figure 48

below. This allows the node to remain active (Geron, 2017) compared to ReLU (Figure 13). The hypothesis tested is that if the model with leaky ReLU performs better than with ReLU, the dead neurons where most likely the cause of its inability to extrapolate under the out of normal conditions.

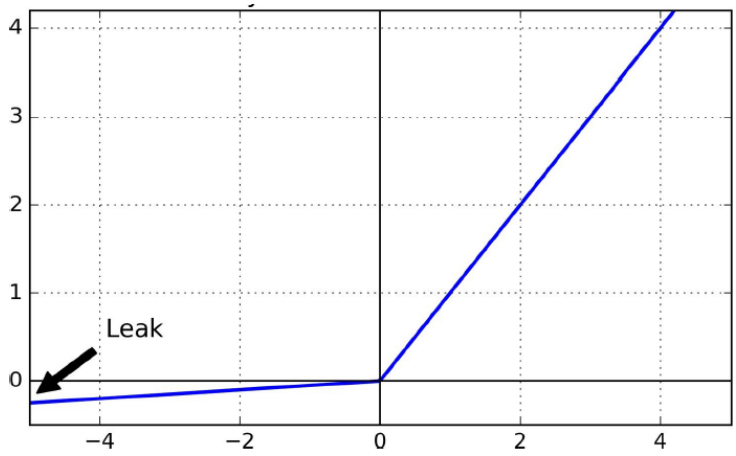


Figure 48 – Leaky ReLU Activation function used in feed forward artificial neural networks (Geron, 2017)

The ANN model using leaky ReLU activation function showed similar performance on the training and validation set to the original model. On the data set with the out of normal scenarios, the model did not show any improvement in predicted values nor the ability to capture the physics of the process. It is can be deduced that dead neurons is not the cause of the model's inability to extrapolate.

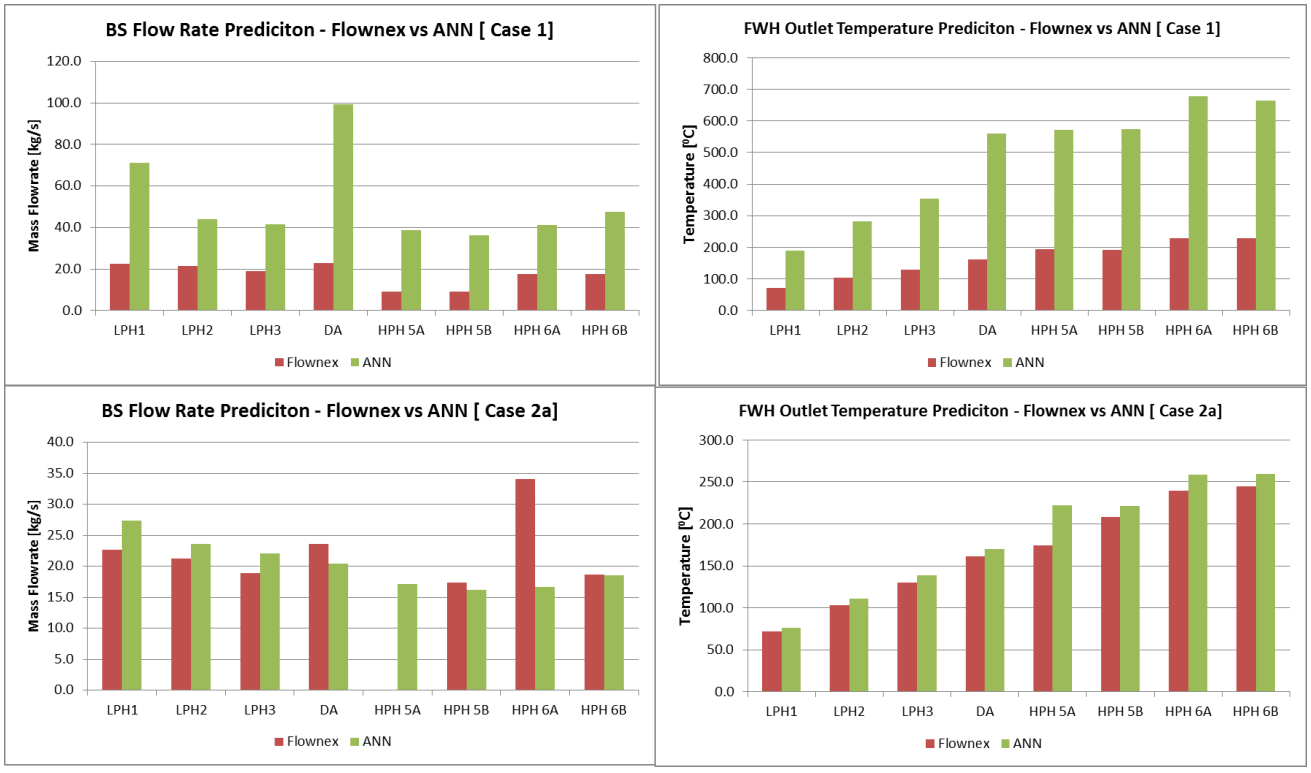
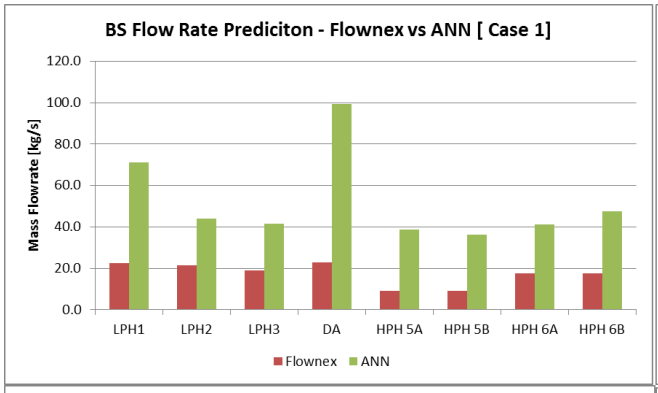
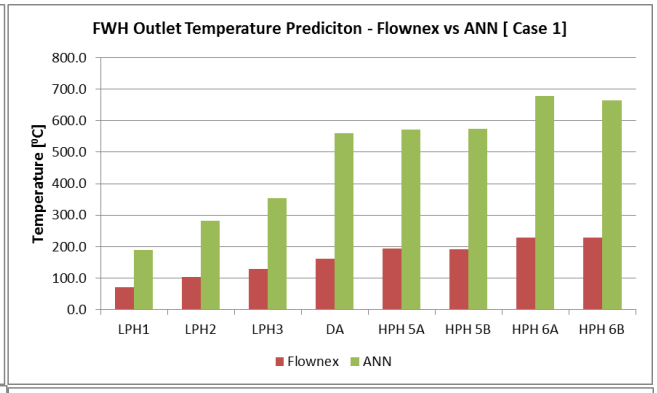


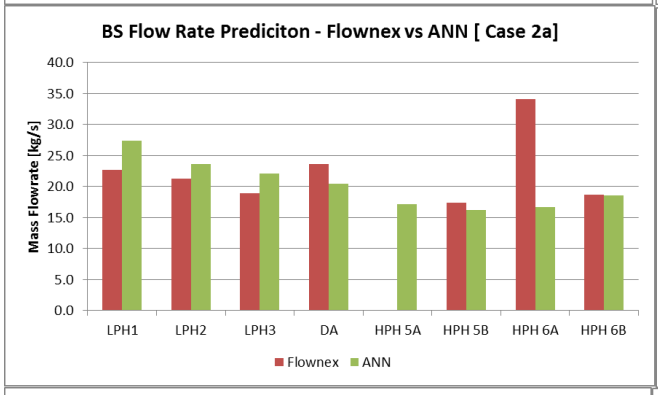
Figure 49 - Extrapolation case results using Leaky ReLU activation function in ANN model

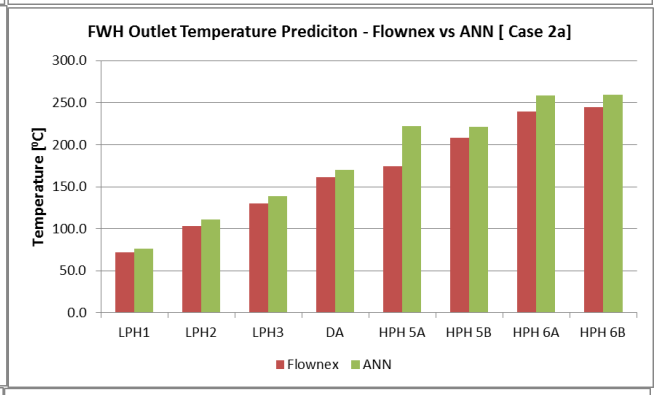
9.2.2 Regularisation

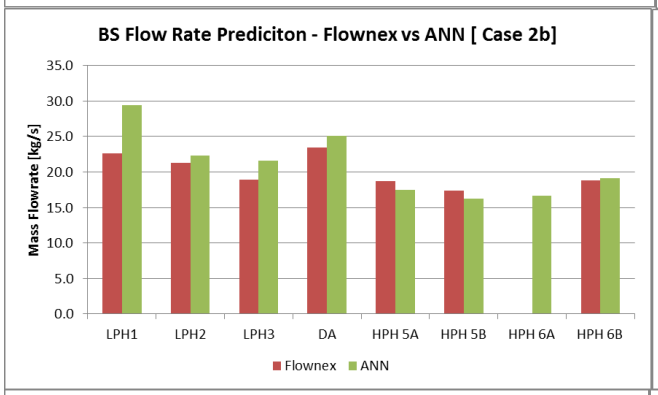
The second concept involves adding dropout as a form of model regularisation. While typically used to reduce overfitting, regularisation is a technique that can be used to improve a model's generalisation. It does this by randomly ignoring neurons in the hidden layers during the training of the model. This means the neuron temporarily has no influence on downstream neurons in the forward pass and therefore will not have a weight update calculated for it in the backward pass. In this model, a dropout of 20% of neurons was selected. As can be seen in the results below, there is no improvement in the ANN model performance under the out of normal conditions.

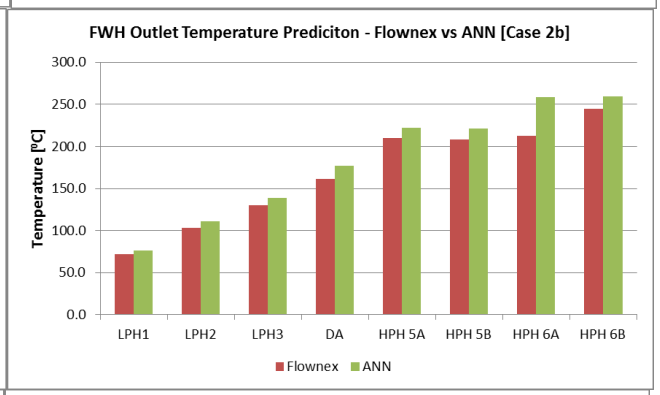

BS Flow Rate Prediciton - Flownex vs ANN [Case 1]
Mass Flowrate [kg/s]
120.0
100.0
80.0
60.0
40.0
20.0
0.0
LPH1 LPH2 LPH3 DA HPH 5A HPH 5B HPH 6A HPH 6B
Flownex ANN

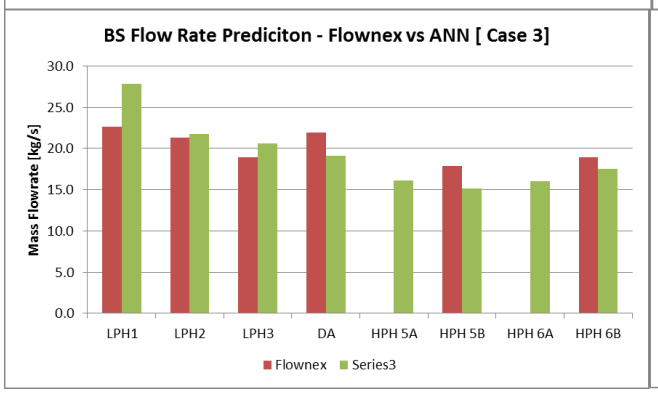

FWH Outlet Temperature Prediciton - Flownex vs ANN [Case 1]
Temperature [°C]
800.0
700.0
600.0
500.0
400.0
300.0
200.0
100.0
0.0
LPH1 LPH2 LPH3 DA HPH 5A HPH 5B HPH 6A HPH 6B
Flownex ANN

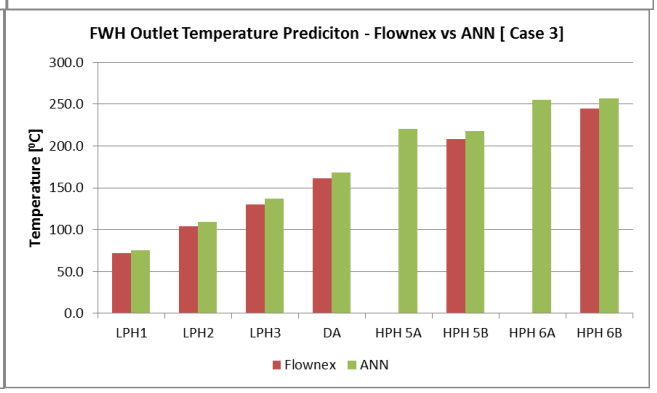

BS Flow Rate Prediciton - Flownex vs ANN [Case 2a]
Mass Flowrate [kg/s]
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
LPH1 LPH2 LPH3 DA HPH 5A HPH 5B HPH 6A HPH 6B
Flownex ANN


FWH Outlet Temperature Prediciton - Flownex vs ANN [Case 2a]
Temperature [°C]
300.0
250.0
200.0
150.0
100.0
50.0
0.0
LPH1 LPH2 LPH3 DA HPH 5A HPH 5B HPH 6A HPH 6B
Flownex ANN


BS Flow Rate Prediciton - Flownex vs ANN [Case 2b]
Mass Flowrate [kg/s]
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
LPH1 LPH2 LPH3 DA HPH 5A HPH 5B HPH 6A HPH 6B
Flownex ANN


FWH Outlet Temperature Prediciton - Flownex vs ANN [Case 2b]
Temperature [°C]
300.0
250.0
200.0
150.0
100.0
50.0
0.0
LPH1 LPH2 LPH3 DA HPH 5A HPH 5B HPH 6A HPH 6B
Flownex ANN


BS Flow Rate Prediciton - Flownex vs ANN [Case 3]
Mass Flowrate [kg/s]
30.0
25.0
20.0
15.0
10.0
5.0
0.0
LPH1 LPH2 LPH3 DA HPH 5A HPH 5B HPH 6A HPH 6B
Flownex Series3


FWH Outlet Temperature Prediciton - Flownex vs ANN [Case 3]
Temperature [°C]
300.0
250.0
200.0
150.0
100.0
50.0
0.0
LPH1 LPH2 LPH3 DA HPH 5A HPH 5B HPH 6A HPH 6B
Flownex ANN

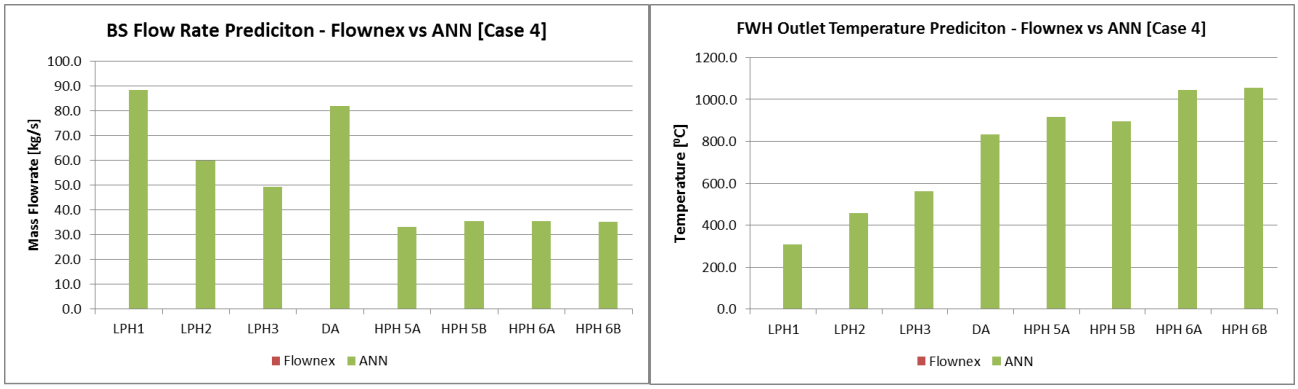


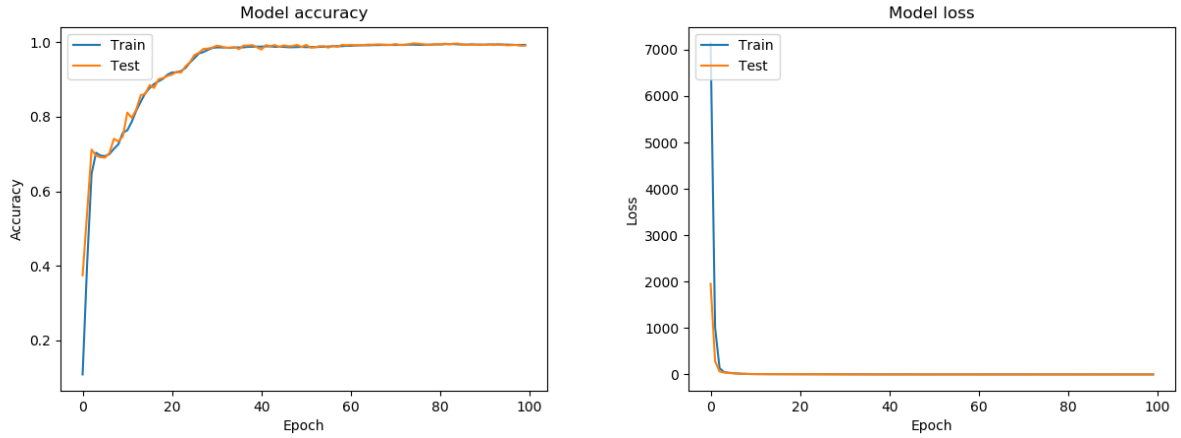
Figure 50 - Extrapolation case results using Dropout regularisation in ANN model

9.2.3 Data synthesis and augmentation

The third concept tested was the method of data augmentation. This is a method typically used when only a small data set is available to train a neural network. As an example, for image recognition models, an image can be replicated and then adjusted to form a new image and hence a new data point (instance). Typical examples would be creating a mirror image of the original, rotating the image a few degrees successively, changing the contrast, colour and brightness of the image all forming new instances.

In this particular case, since the FLOWNEX® SE model was generally successful at predicting the out of normal scenarios, the input and result data was captured and used as training data for the ANN. This was attempted only for case 3 with all the variables relating to the heaters out of service set to a 0 value in the input data. Given that only a single set of input and predictions where generated from FLOWNEX® SE (i.e. one instance), the data had to be manipulated to create more instances required for training the ANN model. Here the data was adjusted by adding 0.00002 to each input parameter successively till 6000 instances where created. This number of instances was arbitrarily chosen. However, it is important to be aware that if too few instances of this scenario where used, the neural net may not see this as significant and not train the network as intended. Adding this very small number to each successive data point meant the difference between the first point and the 6000th was 0.11996.

These 6000 data points where added to the original ANN model training set of 64207. The ANN model was retrained with the new set and executed for the test set. This resulted in a training accuracy of 0.9982, validation accuracy of 0.9981 and a test mean squared error of 0.09135.



(a) *(b)*

Figure 51 - (a) Model accuracy vs epochs, (b) model loss vs epochs

This proves that the model performance was as accurate as the one selected in Chapter 8.3. The out of normal condition, case 3, was then executed. With the new augmented training data set, the ANN model was able to predict the plant behaviour with a great deal more accuracy as seen in Figure 52 below.

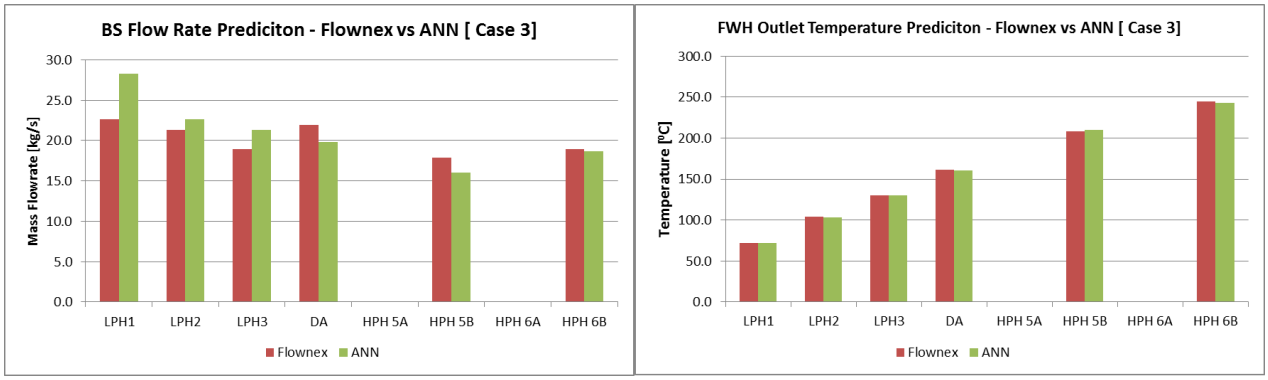


Figure 52- Extrapolation Case 3 results using Data Augmentation and retraining ANN model for OON scenario

This proves that for supervised learning models such as this ANN, the ability of the model to predict plant performance is dependent on the quality and extent of the training data used to train the model. Therefore for this model to be able to predict a scenario, it will have had to be trained to do so. This data can be attained by either mining data from a plant where the condition was previously experienced. Alternatively, the data can be synthesised from simulation tools or modelling software, as tested above, for a broader spectrum of scenarios.

10. Conclusions

From the results obtained the following conclusions can be drawn:

- The FLOWNEX® SE model for the feed heating train was constructed, verified and validated. It proved to accurately predict plant performance under normal operating conditions with the exception for load cases at 48% MCR and lower, where it was unable to converge. The model was found to be unstable at low loads.
- The FLOWNEX® SE model provided insight into the questionable quality of plant data regarding LP heater 1, which would not have been identified if an ANN alone was used.
- The ANN model was built, trained, validated and tested in Python using exactly the same number of input parameters as FLOWNEX® SE and compared. While both models were able to accurately predict plant performance, the ANN was unable to capture the physics of a heat exchanger based on its prediction of the LP heater 1 bled steam mass flow. The comparison was also biased since FLOWNEX® SE has more intermediate information as a result of its fundamental mass and energy balance equations.
- These intermediate variables from plant data, in conjunction with the previous data set, were used to train a new ANN. The ANN, while being able to accurately predict plant performance, was still unable to capture the physics of a heat exchanger.
- The validity of predictions emanating from any of these models hinges on the quality of input data from the plant. It is therefore imperative to assess the integrity and source of the data before creating a model.
- Regarding the ANN model, selecting the pertinent input and prediction parameters, eliminating duplication of parameters and ordering of the dataset (in terms of the columns containing training and predicted variables) is a critical step and must be given a great deal of attention. Further to this, pre-processing of the input data to the neural network is crucial for constructing a robust model.
- The required model complexity (depth and width) was found to decrease with increasing the number of training variables.
- When tested under out of normal conditions, the FLOWNEX® SE model outperformed the ANN model, although getting the simulation to work on FLOWNEX® SE proved to be challenging. It was easier to apply the simulation to the ANN model. However, it was unable to capture the physical operation of the plant processes and produced extremely inaccurate predictions.
- An alternate activation function, leaky ReLU, was used in attempt to improve model generalisation, specifically relating to the out of normal conditions. This proved to make no

improvement to the model's performance, implying that there were no dead neurons created by the ANN model using ReLU activation.

- An attempt was made to use a regularisation technique known as dropout to improve the model's generalisation relating to the out of normal conditions. This did not result in any improvement to the model's performance.
- Data augmentation techniques using the FLOWNEX® SE model results for out of normal case 3 was employed to retrain the ANN model. This concept was proved successful in that the ANN model was able to accurately predict the plant behaviour for case 3. The conclusion drawn is that ANN models can accurately predict plant behaviour only when they are trained to do so.

11. Recommendations

From the results obtained and conclusions made the following are recommended:

- The FLOWNEX® SE model was found to be unstable when trying to simulate out of normal conditions and did not converge at loads less than 56% MCR. The flexibility of this model therefore requires some attention to allow for stable operation and allow for transient modelling. This allows for a wider range of plant operation to be tested as well as for generating additional ANN training data.
- Mass flow of bled steam to LP heater 1 should be physically measured and compared to the plant data.
- In this project, a single artificial neural network was able to predict the behaviour of the plant feed heating plant. However it could not capture the physics of the process as seen by the model's poor performance in the out of normal scenarios. It would be a worthwhile study to assess if one could achieve better prediction if the feed heating plant were discretized and multiple neural networks employed. The first level of discretization could be splitting the heater train into the low pressure and high pressure side each having a neural network to predict their performance. The next level lower could be to discretize it further by building a neural network for each heater in the train.
- For the model to better predict out of normal conditions, training data for these scenarios should be generated using plant simulation models like FLOWNEX® SE. This fills in the gaps potentially created by plant data as a result of that plant not having experienced such an out of normal incident. This synthesised plant behaviour from the simulation of process deviation will provide a wider spectrum of training data. This will improve the neural network generalisation capability.

- A generic loss function from the KERAS toolbox was used in this project. We have seen that the loss function forms the basis of the optimisation of the artificial neural network (see Chapter 4.2). It would therefore be a worthwhile exercise to perform research in understanding if a more suitable loss function can be developed and compare the models' performance on this basis.
- The work conducted in this project focussed on assessing the use of artificial neural networks to predict plant performance. To achieve this, specific periods of operation where considered from which plant data was obtained. For this model to perform well as a plant condition monitoring tool, a recurrent type neural network (RNN) should be considered, particularly the LSTM (Long-Short Term Memory) Network. This type of network is used traditionally for time series data sets. This is applicable in such a scenario due to component degradation and seasonal effects that are seen as a function of time.

12. List of references

Akpan, P. & Fuls, W., 2018. Generic approach for estimating final feed water temperature and extraction pressures in pulverised coal power plants. *Applied Thermal Engineering,* 141(2018), pp. 257-268.

Allie, M. N., 2016. *Thermal modelling of feed water heaters,* s.l.: University of Cape Town.

Banda, R. B., 2015. *Modelling of the deaerator system in flownex,* s.l.: University of Cape Town.

Basile, A. et al., 2015. Water gas shift reaction in membrane reactors: Theoretical investigation by artificial neural networks model and experimental validation. *International journal of Hydrogen Energy,* 40(2015), pp. 5897-5906.

Bekat, T., Erdogan, M., Inal, F. & Genc, A., 2012. Prediction of the bottom ash formed in a coal-fired power plant using artificial neural networks. *Energy,* 45(2012), pp. 882-887.

Bishop, C. M., 2006. *Pattern Recognition and Machine Learning.* s.l.:Springer.

Buratti, C., Orestano, F. C. & Palladino, D., 2016. *Comparison of the energy performance of existing buildings by means of dynamic simulations and Artificial Neural Networks.* Turin, Italy, Energy Procedia.

Chen, R., Rubanova, Y., Bettencourt, J. & Duvenaud, D., 2018. *Neural Ordinary Differential Equations.* Montreal, Neural Information Processing Systems (NeurIPS 2018).

Chokshi, R. B., Chavda, N. K. & Patel, D. A. D., 2018. Prediction of Performance of Coal-Based KWU Designed Thermal Power Plants using an Artificial Neural Network. *International Journal of Applied Engineering Research,* 13(5), pp. 3093-3110.

Chow, M. Y., Sharpe, R. N. & Hung, J. C., 1993. On the application and design of artificial neural networks for motor fault detection—Part II. *IEEE Transactions on Industrial Electronics,* Volume 40, pp. 189-196.

Davoudi, E. & Vaferi, B., 2018. Applying artificial neural networks for systematic systematicestimation of degree of fouling in heat exchangers. *Chemical Engineering Research and Design,* 130(2018), pp. 138-153.

Dhanuskodi, R. et al., 2015. Artificial Neural Networks model for predicting wall temperature of supercritical boilers. *Applied Thermal Engineering,* 90(2015), pp. 749-753.

Diederik P. Kingmal, J. L. B., 2015. *Adam: A Method for Stochastic Optimization.* s.l., s.n.
Dong, S. et al., 2018. Investigation of Support Vector Machine and Back Propagation Artificial Neural Network for performance prediction of the organic Rankine cycle system. *Energy,* 144(2018), pp. 851-864.

Esfe, M. H., 2017. Designing a neural network for predicting the heat transfer and pressure drop characteristics of Ag/water nanofluids in a heat exchanger. *Applied Thermal Engineering,* 126(2017), pp. 559-565.

Fast, M. & Palme, T., 2010. Application of artificial neural networks to the condition monitoring and diagnosis of a combined heat and power plant. *Energy,* 35(2010), pp. 1114-1120.

Geron, A., 2017. *Hands-On Machine Learning with Scikit-Learn and TensorFlow.* s.l.:O'Reilly Media.

Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning.* s.l.:MIT Press.
Gullil, A. & Pal, S., 2017. *Deep lEarning with Keras.* s.l.:Packt Publishing.

Guo, S., Liu, P. & Li, Z., 2018. Enhancement of performance monitoring of a coal-fired power plant via dynamic data reconciliation. *Energy,* 151(2018), pp. 203-210.

Hagan, M. T., Demuth, H. B., Beale, M. H. & Jesus, O. D., 1996. *Neural Network Design.* 2nd ed. s.l.:s.n.

Haque, M. T. & Kashtiban, A., 2007. Application of Neural Networks in PowerSystems; A Review. *International Journal of Energy and Power Engineering,* 1(6), pp. 897-901.

Jestin, L. & Piketh, S., 2018. *Overview of the Power Plant Industry Volume 1.* s.l.:University of Cape Town.

Khan, M. S., Husnil, Y. A., Getu, M. & Lee, M., 2012. *Modeling and Simulation of Multi-stream Heat Exchanger Using Artificial Neural Network.* Singapore, Elsevier.

Laubscher, R., 2017. *Utilization of artificial neural networks to resolve chemical kinetics in turbulent fine structures of an advanced CFD combustion model,* s.l.: University of Stellenbosch.

LeCun, Y., Benigo, Y. & Hinton, G., 2015. Deep Learnong. *Nature,* Volume 521, p. 436.

leGrange, W., 2018. *Component development for a high fidelity transient simulation of a coal-fired power plant using Flownex SE,* s.l.: University of Cape Town.

Luo, K., Xing, J., Ba, Y. & J. F., 2018. Prediction of product distributions in coal devolatilization by an artificial neural network model. *Combustion and Flame,* 193(2018), pp. 283-294.

Massimiani, A., Palagi, L., Sciubba, E. & Tocci, L., 2017. *Neural Networks for small scale ORC optimasiton.* Milano, Italy, Energy Procedia.

Meireles, M. R. G., Almeida, P. E. M. & Simões, M. G., 2003. A Comprehensive Review for Industrial Applicability of Artificial Neural Networks. *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS,* 50(3), pp. 585-601.

Mikulandric, R., Loncar, D., Cvetinovi, D. & Spiridon, G., 2013. Improvement of existing coal fired thermal power plants performance by control systems modifications. *Energy,* 57(2013), pp. 55-65.

Mohanraj, M., Jayaraj, S. & Muraleedharan, C., 2015. Applications of artificial neural networks for thermal analysis of heat exchangers - A review. *International Journal of Thermal Sciences,* 90(2015), pp. 150-172.

Mohatram, M., Tewari, P. & Shahjahan, 2011. Applications of Artificial Neural Networks in Electric Power Industry: A Review. *International Journal of Electrical Engineering,* 4(2), pp. 161-171.

Nasr, M. S., Moustafa, M. A., Seif, H. A. & Kobrosy, G. E., 2010. Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT. *Alexandria Engineering Journal.*

Rashidi, M. et al., 2011. Parametric analysis and optimization of regenerative Clausius and organic Rankine cycles with two feedwater heaters using artificial bees colony and artificial neural network. *Energy,* 36(2011), pp. 5728-5740.

Rodríguez, F., Fleetwood, A., Galarza, A. & Fontan, L., 2018. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renewable Energy,* 126(2018), pp. 855-864.

Rousseau, P. & Fuls, W., 208. *Power Plant System Analysis.* s.l.:University of Cape Town.

Smrekar, J. et al., 2009. Development of artificial neural network model for a coal-firedboiler using real plant data. *Energy,* 34(2009), pp. 144-152.

Strušnik, D., Golob, M. & Avsec, J., 2015. Artificial neural networking model for the prediction of high efficiency boiler steam generation and distribution. *Simulation Modelling Practice and Theory,* 57(2015), pp. 58-70.

Suresh, M., Reddy, K. & Kolar, A. K., 2011. ANN-GA based optimization of a high ash coal-fired supercritical power plant. *Applied Energy,* 88(2011), pp. 4867-4873.

Tan, C., Ward, J., Wilcox, S. & Payne, R., 2009. Artificial neural network modelling of the thermal performance of a compact heat exchanger. *Applied Thermal Engineering,* 29(2009), pp. 3609-3617.

Tan, C., Ward, J., Wilcox, S. & Payne, R., 2019. Artificial Neural Network Modelling of the Thermal Performance of a Compact Heat Exchanger. *Applied Thermal Engineering.*

Tang, Y., 2017. *Deep Learning using Support Vector Machines.* Toronto, s.n.

Terranova, T. & Gibbard, I., 2008. *Power Plant Feed Water Heaters.* s.l.:Power Plant Engineering.

Tumer, A. E., 2015. An Artificial Neural Network Model for Wastewater Treatment Plant of Konya. *International Journal of Intelligent Systems and Applications in Engineering,* pp. 131-135.

Tunckaya, Y. & Koklukaya, E., 2015. Comparative analysis and prediction study for effluent gas emissions in a coal-fired thermal power plant using artificial intelligence and statistical tools. *Journal of the Energy Institute,* 88(2015), pp. 118-125.

Tunckaya, Y. & Koklukaya, E., 2015. Comparative analysis and prediction study for effluent gas emissions in a coal-fired thermal power plant using artificial intelligence and statistical tools. *Journal of the Energy Institute,* 88(2015), pp. 118-125.

Tunckaya, Y. & Koklukaya, E., 2015. Comparative prediction analysis of 600 MWe coal-fired power plant production rate using statistical and neural-based models. *Journal of the Energy Institute,* 88(2015), pp. 11-18.

Tunckaya, Y. & Koklukaya, E., 2015. Comparative prediction analysis of 600 MWe coal-fired power plant production rate using statistical and neural-based models. *Journal of the Energy Institute,* 88(2015), pp. 11-18.

Wiering, M. et al., n.d. *The Neural Support Vector Machine,* Groningen, Netherlands: Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen.
Yann LeCun, L. B. G. B. O. K.-R. M., 1998. Efficient BackProp. *Springer.*

Appendix A. FLOWNEX® SE model Mathcad verification

Each heater was assessed individually. The naming convention for each heater is as follows: Feed water shall be described as *fw*, inlet by subscript *in*, feed water exit by subscript *out*, cascade inlet by subscript *cas*, bled steam inlet will by subscript *bs*, and distillate outlet by subscript *dist*.

LP Heater 1

INPUTS FROM FLOWNEX:

$m_{fw\_in} := 345.54\,\frac{kg}{s}$ $m_{fw\_out} := 345.54\,\frac{kg}{s}$ $m_{bs} := 22.0\,\frac{kg}{s}$ $m_{cas} := 38.23\,\frac{kg}{s}$ $m_{dist} := 60.23\,\frac{kg}{s}$

$h_{fw\_in} := 149.3\,\frac{kJ}{kg}$ $h_{fw\_out} := 299.38\,\frac{kJ}{kg}$ $h_{bs} := 2638.3\,\frac{kJ}{kg}$ $h_{cas} := 332.6\,\frac{kJ}{kg}$ $h_{dist} := 313.58\,\frac{kJ}{kg}$

CALCULATION:

Mass balance:

$$m_{cas} + m_{bs} - m_{dist} = 0\,\frac{kg}{s}$$

$$m_{fw\_out} - m_{fw\_in} = 0\,\frac{kg}{s}$$

Energy balance:

Shell side

$$Q_s := m_{cas}\cdot h_{cas} + m_{bs}\cdot h_{bs} - m_{dist}\cdot h_{dist} = 51.871\,MW$$

Tube side

$$Q_t := m_{fw\_out}\cdot h_{fw\_out} - m_{fw\_in}\cdot h_{fw\_in} = 51.859\,MW$$

Overall

$$Q_s - Q_t = 0.012\,MW$$

LP Heater 2

INPUTS FROM FLOWNEX:

$m_{fw\_in} := 407\,\frac{kg}{s}$ $m_{fw\_out} := 407\,\frac{kg}{s}$ $m_{bs} := 21.7\,\frac{kg}{s}$ $m_{cas} := 17.8\,\frac{kg}{s}$ $m_{dist} := 39.5\,\frac{kg}{s}$

$h_{fw\_in} := 302\,\frac{kJ}{kg}$ $h_{fw\_out} := 435.9\,\frac{kJ}{kg}$ $h_{bs} := 2742.5\,\frac{kJ}{kg}$ $h_{cas} := 462.6\,\frac{kJ}{kg}$ $h_{dist} := 335.3\,\frac{kJ}{kg}$

CALCULATION:

Mass balance:

$$m_{cas} + m_{bs} - m_{dist} = 0\,\frac{kg}{s}$$

$$m_{fw\_out} - m_{fw\_in} = 0\,\frac{kg}{s}$$

Energy balance:

Shell side

$$Q_s := m_{cas}\cdot h_{cas} + m_{bs}\cdot h_{bs} - m_{dist}\cdot h_{dist} = 54.502\,MW$$

Tube side

$$Q_t := m_{fw\_out}\cdot h_{fw\_out} - m_{fw\_in}\cdot h_{fw\_in} = 54.497\,MW$$

Overall

$$Q_s - Q_t = 0.005\,MW$$

LP Heater 3

INPUTS FROM FLOWNEX:

$m_{fw\_in} := 407\,\frac{kg}{s}$ $m_{fw\_out} := 407\,\frac{kg}{s}$ $m_{bs} := 17.8\,\frac{kg}{s}$ $m_{cas} := 0\,\frac{kg}{s}$ $m_{dist} := 17.8\,\frac{kg}{s}$

$h_{fw\_in} := 435.9\,\frac{kJ}{kg}$ $h_{fw\_out} := 541.5\,\frac{kJ}{kg}$ $h_{bs} := 2874.2\,\frac{kJ}{kg}$ $h_{cas} := 0\,\frac{kJ}{kg}$ $h_{dist} := 462.6\,\frac{kJ}{kg}$

CALCULATION:

Mass balance:

$$m_{cas} + m_{bs} - m_{dist} = 0\,\frac{kg}{s}$$

$$m_{fw\_out} - m_{fw\_in} = 0\,\frac{kg}{s}$$

Energy balance:

Shell side

$$Q_s := m_{cas} \cdot h_{cas} + m_{bs} \cdot h_{bs} - m_{dist} \cdot h_{dist} = 42.926\,MW$$

Tube side

$$Q_t := m_{fw\_out} \cdot h_{fw\_out} - m_{fw\_in} \cdot h_{fw\_in} = 42.979\,MW$$

Overall

$$Q_s - Q_t = -0.053\,MW$$

De-Aerator:

INPUTS FROM FLOWNEX:

$$m_{fw\_in} := 407\,\frac{kg}{s} \quad m_{fw\_out} := 502.4\,\frac{kg}{s} \quad m_{bs} := 23.3\,\frac{kg}{s} \quad m_{cas} := 72.06\,\frac{kg}{s}$$

$$h_{fw\_in} := 541.5\,\frac{kJ}{kg} \quad h_{fw\_out} := 684.7\,\frac{kJ}{kg} \quad h_{bs} := 3062\,\frac{kJ}{kg} \quad h_{cas} := 723.7\,\frac{kJ}{kg}$$

CALCULATION:

Mass balance:

$$m_{cas} + m_{bs} + m_{fw\_in} - m_{fw\_out} = -0.04\,\frac{kg}{s}$$

HP Heater 5A

INPUTS FROM FLOWNEX:

$$m_{fw\_in} := 246.5\,\frac{kg}{s} \quad m_{fw\_out} := 246.5\,\frac{kg}{s} \quad m_{bs} := 17.4\,\frac{kg}{s} \quad m_{cas} := 18.6\,\frac{kg}{s} \quad m_{dist} := 36\,\frac{kg}{s}$$

$$h_{fw\_in} := 714.6\,\frac{kJ}{kg} \quad h_{fw\_out} := 915.2\,\frac{kJ}{kg} \quad h_{bs} := 3345.3\,\frac{kJ}{kg} \quad h_{cas} := 936.2\,\frac{kJ}{kg} \quad h_{dist} := 723.7\,\frac{kJ}{kg}$$

CALCULATION:

Mass balance:

$$m_{cas} + m_{bs} - m_{dist} = 0\,\frac{kg}{s}$$

$$m_{fw\_out} - m_{fw\_in} = 0\,\frac{kg}{s}$$

Energy balance:

Shell side

$$Q_s := m_{cas} \cdot h_{cas} + m_{bs} \cdot h_{bs} - m_{dist} \cdot h_{dist} = 49.568\,MW$$

Tube side

$$Q_t := m_{fw\_out} \cdot h_{fw\_out} - m_{fw\_in} \cdot h_{fw\_in} = 49.448\,MW$$

Overall

$$Q_s - Q_t = 0.12\,MW$$

HP Heater 6A

INPUTS FROM FLOWNEX:

$m_{fw\_in} := 246.5\,\frac{kg}{s}$ $m_{fw\_out} := 246.5\,\frac{kg}{s}$ $m_{bs} := 18.7\,\frac{kg}{s}$ $m_{cas} := 0\,\frac{kg}{s}$ $m_{dist} := 18.7\,\frac{kg}{s}$

$h_{fw\_in} := 915.2\,\frac{kJ}{kg}$ $h_{fw\_out} := 1075.2\,\frac{kJ}{kg}$ $h_{bs} := 3047.3\,\frac{kJ}{kg}$ $h_{cas} := 0\,\frac{kJ}{kg}$ $h_{dist} := 936.2\,\frac{kJ}{kg}$

CALCULATION:

Mass balance:

$$m_{cas} + m_{bs} - m_{dist} = 0\,\frac{kg}{s}$$

$$m_{fw\_out} - m_{fw\_in} = 0\,\frac{kg}{s}$$

Energy balance:

Shell side

$$Q_s := m_{cas}\cdot h_{cas} + m_{bs}\cdot h_{bs} - m_{dist}\cdot h_{dist} = 39.478\,MW$$

Tube side

$$Q_t := m_{fw\_out}\cdot h_{fw\_out} - m_{fw\_in}\cdot h_{fw\_in} = 39.44\,MW$$

Overall

$$Q_s - Q_t = 0.038\,MW$$

Appendix B. FLOWNEX® SE live plant data

	Feed water inlet		BS - LP1		BS - LP2		BS - LP3		BS - DA	
Load	**T**	**m**	**T**	**P**	**T**	**P**	**T**	**P**	**T**	**P**
100	34.39	346.18	77.71	42.85	143.00	129.32	214.00	292.23	303.83	635.42
90	35.14	313.46	76.24	40.42	144.00	119.16	211.00	269.89	307.74	582.25
74	28.55	250.66	71.16	32.44	144.00	96.92	211.00	217.61	306.53	465.64
65	25.93	224.40	68.49	28.96	144.00	86.02	211.00	192.32	301.27	410.24
56	28.12	210.57	67.02	27.28	144.00	78.42	211.00	174.56	291.03	368.85
48	26.25	174.44	63.35	22.95	144.00	66.76	211.00	147.97	301.41	311.09

	BS - HPH 5A		BS - HPH 5B		BS - HPH 6A		BS - HPH 6B		Feed water outlet	RH ATTEMP
Load	**T**	**P**	**T**	**P**	**T**	**P**	**T**	**P**	**P**	**m**
100	450.21	1949.44	450.21	1953.40	328.19	3811.11	328.19	3812.55	19640.82	15.84
90	452.36	1773.01	452.36	1776.68	328.69	3461.91	328.69	3449.31	19364.10	16.72
74	450.66	1425.07	450.66	1428.01	324.24	2777.07	324.24	2773.80	18557.37	15.83
65	445.61	1265.21	445.61	1266.30	317.29	2463.23	317.29	2459.33	18557.37	15.86
56	433.18	1147.64	433.18	1151.76	318.71	2236.73	318.71	2229.88	18143.58	16.84
48	447.14	971.24	447.14	974.07	317.87	1873.40	317.87	1867.00	18065.20	16.63

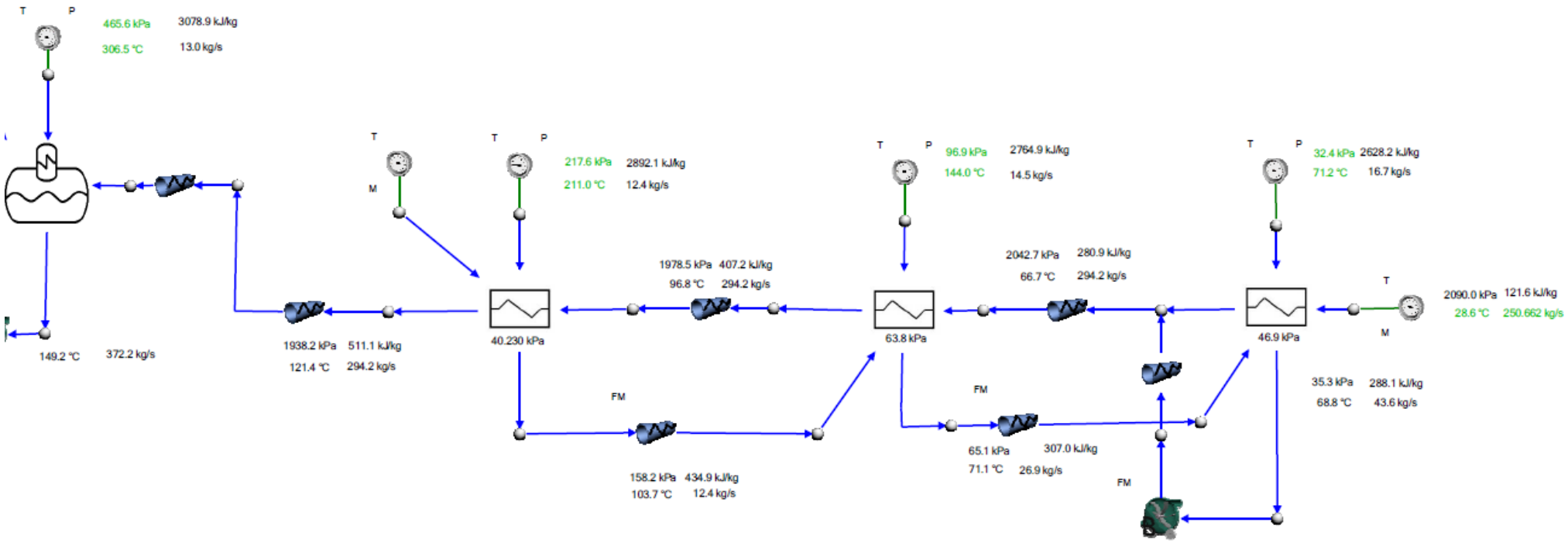
Appendix C. Machine learning program code

```
# FF ANN model to predict the performance of a feed water heating train
# Network will use ReLU activation function and Adam optimisation with Xavier weight
initialization
# OPTIMUM ARCHITECTURE = Case 5.1.2: layers = 2, nodes1 = 200, nodes2 = 200, batch size =
500, epochs = 100

#_____________________________________________________________________________


# 1. IMPORT LIBRARIES

# Import the required libraries for data preprocessing and plotting
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt

# Importing the Keras libraries and packages for ANN model building
import keras
from keras.models import Sequential
from keras.layers import Dense

#_____________________________________________________________________________


# 2. DATA PREPROCESSING

# Import the training data set
dataset = pd.read_csv('plantdata_new.csv')
X = dataset.iloc[:, :-16].values
y = dataset.iloc[:, 55:71].values

# Import the live plant data set
pdataset = pd.read_csv('PredictionData_new.csv')
X_pd = pdataset.iloc[:, :-16].values
y_target = pdataset.iloc[:, 55:71].values

# Splitting the training dataset into the Training set (80%) and Test set (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

# Feature Scaling (standardize data between -1 and 1)
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
X_pd = sc.transform(X_pd)

#_____________________________________________________________________________


# 3. BUILD THE NEURAL NETWORK

# Initialise the network
model = Sequential()

# Define the network
# Add the input layer and the first hidden layer
model.add(Dense(200, kernel_initializer='glorot_normal', activation='relu',
input_dim=50))
```

```
#Add the second hidden layer
model.add(Dense(200, kernel_initializer='glorot_normal', activation='relu'))

# Add the output layer
model.add(Dense(16, kernel_initializer='glorot_normal', activation='relu'))

# Complile the network
model.compile(optimizer='adam', loss='mean_squared_error', metrics=['accuracy'])

# Fit the network (split the training set further into a training and validation set)
history = model.fit(X_train, y_train, validation_split=0.2, batch_size=500, epochs=100,
verbose=2)


#_____________________________________________________________________________


# 4. PLOT TRAINING AND VALIDATION PERFORMANCE

# #Plot training & validation accuracy values
plt.plot(history.history['acc'])
plt.plot(history.history['val_acc'])
plt.title('Model accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(['Train', ' Validation '], loc='upper left')
plt.show()

# Plot training & validation loss values
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('Model loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Train', 'Validation'], loc='upper left')
plt.show()
#_____________________________________________________________________________


# 5. MAKE PREDICTIONS USING THE MODEL

# Predicting the Test set results
y_predt = model.predict(X_test, batch_size=None, verbose=2, steps=None)
print(mean_squared_error(y_test,y_predt))

# Predicting the plant data set results
y_predp = model.predict(X_pd, batch_size=None, verbose=2, steps=None)
print(mean_squared_error(y_target,y_predp))
```

Appendix D. FLOWNEX® SE model

The figure below depicts series of LP heaters 1, 2, 3 and the de-aerator.



The figure below depicts feed pump and the parallel HP heater banks A (above) and B (below).