17th Brazilian Congress of Thermal Sciences and Engineering
November 25th-28th, 2018, Águas de Lindóia, SP, Brazil

ENCIT-2018-0207

STEAM FLOW ESTIMATION WITH ARTIFICIAL NEURAL NETWORK BASED ON POWER PLANT OPERATIONAL DATA

Helena Haas Reichert
Federal University of Rio Grande do Sul, Department of Mechanical Engineering, Porto Alegre, RS, Brazil
helena.reichert@ufrgs.br

João Gari da Silva Fonseca Junior
The University of Tokyo, Institute of Industrial Science, Meguro, Tokyo, Japan
jfonseca@iis.u-tokyo.ac.jp

Paulo Smith Schneider
Federal University of Rio Grande do Sul, Department of Mechanical Engineering, Porto Alegre, RS, Brazil
pss@mecanica.ufrgs.br

Abstract. *Artificial neural networks ANN are gaining space due to their ability on modeling complex heat transfer problems. This paper brings the development of an ANN trained to estimate the steam flow generation of a real thermoelectric power plan, based on past records of its operation. ANN configuration is obtained after a series of tests aiming to reduce the prediction error, which are further on compared to a reference multiple linear regression model. ANN prediction quality is assessed through the calculated mean absolute percentage error MAPE of 3 validation sets, which display results as 3.74%, 2.28% and 3.02%. Regression procedures applied to the same set of data have MAPE values of 7.19%, 3.02% and 4.19%, and show the ANN ability to estimate steam production from actual equipment.*

Keywords: *Artificial neural networks, Steam generator, thermoelectric power plant operation modelling.*

1. INTRODUCTION

Thermoelectric power generation represents 25.8% of the Brazilian energy matrix (ANEEL, 2018), and survey of these plants are important to reduce greenhouse gases emissions, and to keep and to improve their operation efficiency. Thus, the development of control and monitoring techniques is a continuous activity of researchers and stakeholders in thermoelectric power generation. The process of steam generation is a multi-parameter activity and its modeling involves complex phenomena. Actual plants can provide a large number of operation data, continuously acquired by online plant monitoring systems. The complexity of directly modeling the steam generation process in real operation conditions make this problem particularly suitable to the application of machine learning models, whenever a large amount of data is available. Machine learning models such as artificial neural networks (ANN) have the ability to recognize patterns and to infer relations from complex sets of data.

ANNs have already been successfully applied to reproduce and simulate the behavior of heat transfer problems involving gas modeling, optimization of energy efficiency and NO_x emissions, forecasting of energy resources, among others (GHUGARE et al., 2014). ANN modeling of real power plant data were previously investigated (DE et al., 2007; MESROGHLI; JORJANI; CHEHREH CHELGANI, 2009; SMREKAR et al., 2009; STRUŠNIK; GOLOB; AVSEC, 2015), but it is worth noticing that each case deserves the development of a specific network to reproduce the actual equipment or system.

The objective of the present work was to present a step-by-step procedure of the ANN modelling of the estimation of steam generation on a coal-fired power plant. The model was based on a real operational database from the PECEM [1]

[1] www.energiapecem.com.br

power plant (2x360 MW), with records of steam flow rate and pressure, fuel flow rate and water temperature at the steam generator inlet, for several periods and operation conditions.

2. PROBLEM DESCRIPTION

PECEM power plant generates 720 MW of electricity with two similar size power groups, equipped with coal-fired superheated steam generator SSG. Each SSG was designed to produce 360 MW with 1200 t/h of superheated steam at 540ºC and 18MPa. Figure 1 presents the SSG schematic layout and displays the set of input parameters and its output.

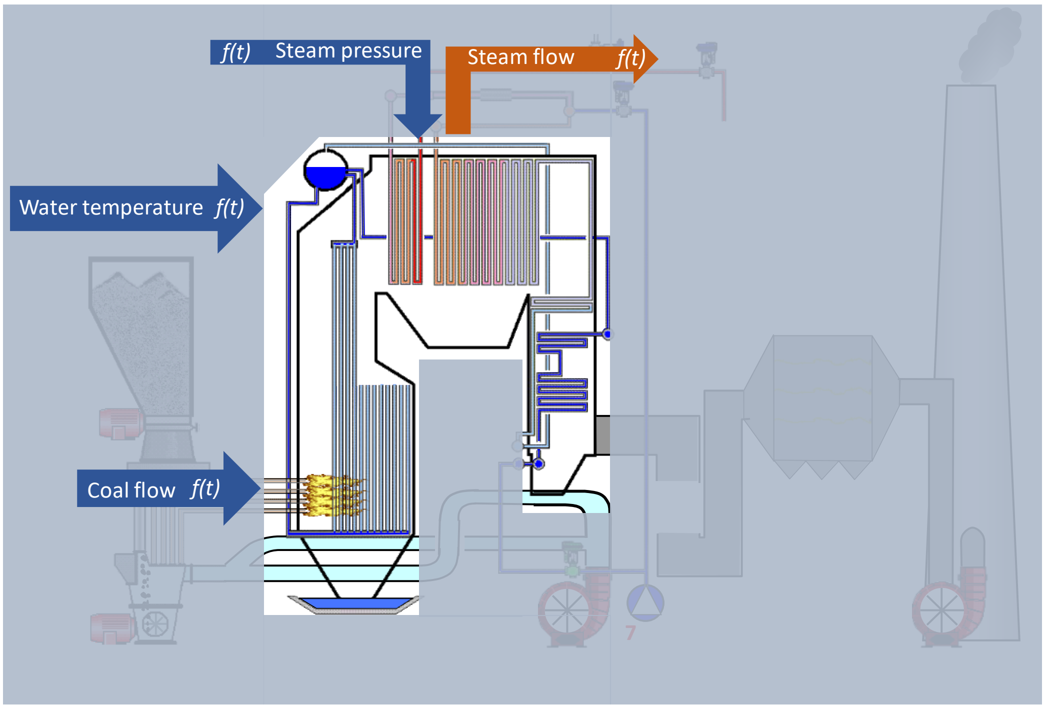

Figure 1 – Superheated Steam Generator SSG with its chosen input and output (blue and orange arrows) parameters

Input parameters were selected from a yearlong data set, allowing to identify two separate power plant electric output regimes, such as 240 MW and 360 MW, imposed by the electric national system operator[2]. Steam generation mass flow rate $\dot{m}_s$ was chosen as the parameter to be estimated by the ANN, based on the SSG steam pressure steam P_s, inlet water temperature T_w, coal mass flow rate $\dot{m}_c$, as a function of time. Reference values for those parameters are displayed in Table 1.

Table 1 – SSG design and minimum values

Parameter		Minimum	Maximum
Water temperature	T_w	51.07 ºC	279.10 ºC
Steam pressure	P_s	0.225MPa	18.54 MP
Coal mass flow rate	$\dot{m}_c$	0.18 t/h	149.81 t/h
Steam mass flow rate	$\dot{m}_s$	0.27 t/h	1285.63 t/h

These minimum and maximum values were found in the dataset. Although they were not time coincident, data were a reference binderies of the SSG operation.

[2] ://ons.org.br

3. METHODOLOGY

The present section aims to describe a step by step sequence of actions, or methodology, applied to estimate the SSG steam mass flow rate $\dot{m}_s$ and the model Mean Absolute Percentage Error MAPE. The methodology was developed in four blocks due to the problem complexity. Block 1 (Figure 2) refers to the data processing.

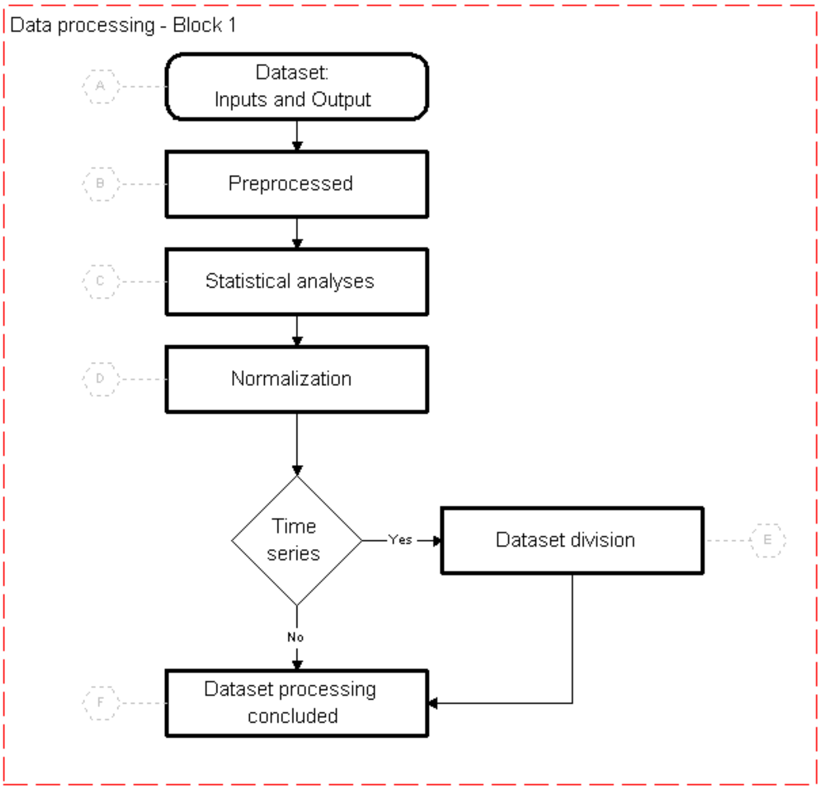

Figure 2 –Block 1- Data processing

Data set (letter A) is presented in Table 1. The preprocessed procedure (letter B) is in fact dedicated to data cleaning, as the original data source presented defective values, with zeros and negatives. The faulty data were replaced by interpolated adjacent values.

Statistical analyses (letter C) performed the correlations analyses were also made to understand the relationship between the parameters. By the correlation study was possible to conclude that the dataset does not present a time series profile, once there is no correlation between the timestep on the dataset. The Pearson correlation index was used to assess the mutual influence between parameters (Table 2). The upper diagonal was not filled due to the matrix symmetry.

Table 2 – Data set Pearson correlation index

	Water temperature	*Steam pressure*	*Coal mass flow rate*	***Steam mass flow rate***
Water temperature	1			
Steam pressure	0.924	1		
Coal mass flow rate	0.931	0.958	1	
Steam mass flow rate	0.953	0.983	0.967	1

It is possible to conclude that steam pressure was the most correlated parameter to steam mass flow rate, followed by coal mass flow rate to water temperature. Input parameters also displayed high correlation coefficients, showing that they also are strongly correlated.

The data normalization (letter D) was performed based on the minimum and the design values showed in Table 1. Time scale was also normalized, by associating the unitary value for the first measured time and 5309 to the last one.

The data set division (letter E) aimed to separate and individualize complete sequence of data. Plant shutdowns caused interruptions on data sequences whose lack of information could cause errors in data assessment. Table 3 details the three sets of data sequences used in the study.

Table 3 – Data set division from letter E in Figure 2

Data set	**Start at**	**End at**	**Number of samples**	**Data set size**
Set_1	2/1/2017 4:51:41 PM	3/7/2017 2:51:41 PM	815	16%
Set_2	3/10/2017 10:51:41 AM	08/12/2017 5:51:41 AM	3716	72%
Set_3	8/16/2017 5:51:41 AM	9/10/2017 10:51:41 PM	617	12%

Data processing conclusion (letter F) is reached at the very end and the data set is prepared be processed along the next blocks of the methodology. Data set 2 was the largest sequence, with 72% of the available information, and it was chosen to be on the base of the two next procedures depicted in Figure 3, blocks 2 to 4, dedicated to linear multiple regression, ANN and conclusions, respectively.

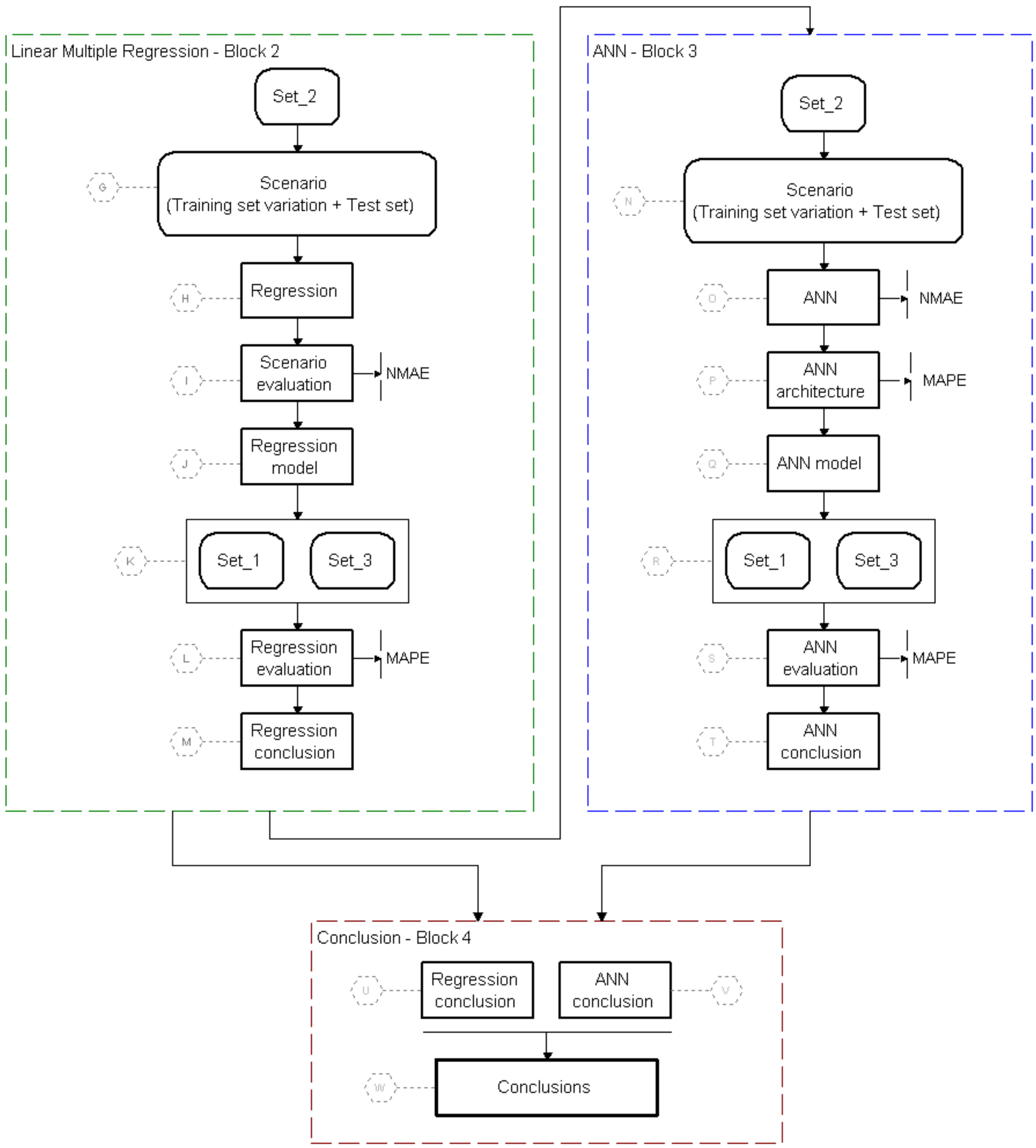

Figure 3 – Methodology scheme for Linear multiple regression (Block 2), ANN (Block 3) and Conclusions (Block 4)

Scenarios were assembled with data from Set_2 (letter G; Block 2), each one built upon a particular training and testing partition. The test group was fixed with 716 samples and the training group range from 3000 to 100 samples. The scenarios were divided into training and test group. Only the size of the training group was changed to compare the training group performance. All the scenarios were used to develop a regression model (letter H). A linear regression was developed to be used as a reference model to the ANN.

Linear regression evaluation was meant to be performed in letter I (Scenario evaluation) by means of the normalized mean absolute error (NMAE). The NMAE was calculated by Equation 1 with data from the test group.

$$NMAE = \left(\frac{1}{n}\sum_{i=1}^{n}|X_{est} - X_{obs}|\right)\frac{1}{(X_{obs,max} - X_{obs,min})} \tag{1}$$

with *X* the steam mass flow rate for both measured (obs) and estimated (est) values, X measured maximum and minimum values found in dataset and *n* the number of data points. After the scenarios evaluation, the regression model (letter J) was tested with Set_1 and Set_3 (letter K) and their Mean Absolute Percentage Error MAPE was calculated by Equation 2, in order to verify the regression model accuracy (letter L).

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{X_{obs} - X_{est}}{X_{obs}}\right| \tag{2}$$

Block 4 presented the methodology for the ANN development. The Set_2 was also divided in training and test group to evaluated the best scenario (letter N; Block 3). All the scenarios were used to develop an ANN and the normalized mean absolute error (NMAE) were calculated to evaluated the best scenario (letter O). The best ANN (letter P) architecture was defined by the MAPE calculation. Finally, Set_1 and Set_3 (letter R) were applied to verify the ANN accuracy. The ANN evaluation (letter S) was made by calculation of MAPE for the sets evaluated and MSE, Equation 3, for the evaluation of each set in relation to the observed values of steam flow.

$$MSE = (X_{est} - X_{obs})^2 \tag{3}$$

ANN Multi-Layer Perceptron MLP type was chosen to simulate the SSG, which is an extension of the perceptron model proposed by Rosenblatt, built with several intermediate or hidden layers of artificial neurons (GEVREY; DIMOPOULOS; LEK, 2003; NETO; BECCENERI, [s.d.]). MLP model comprehends intermediary neurons, responsible for the absorption of progressive knowledge, enabling the execution of more complex tasks. Finally, it is worth emphasizing that the ANN has high connectivity since any modification in the neural network requires a restructuring (HAYKIN, 2001).

The configuration parameters had to be set: activation function; number of hidden layers; number of neurons in each layer; number of inputs to be used; and the amount of data to be used in the training of the ANN. Figure 4 shows the ANN architecture used in this work.

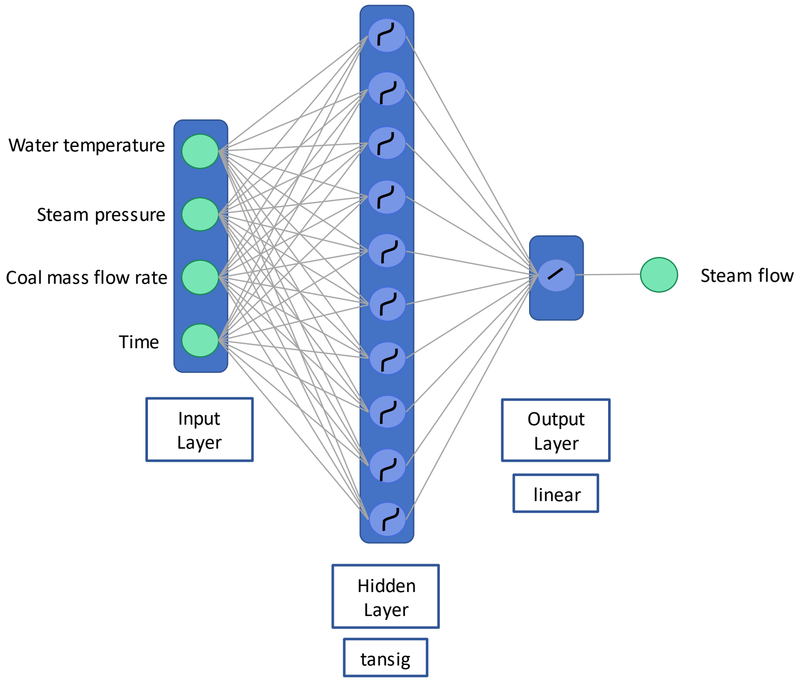

Figure 4 - ANN architecture developed for steam flow estimation

The inputs parameters are presented at the first layer in Figure 4. After that, the hidden layer is showed, in this step the number of neurons needs to be set, this configuration was evaluated varying the number of neurons and analyzing the MSE value of each ANN. The transfer function used was the sigmoidal hyperbolic tangent function (*tansig*). The output layer is the last layer in this architecture with one neuron and the linear function was used as transfer function. The output parameter is the stem flow. The ANN multilayer feed forward perceptron with back propagation mechanism was chosen, together with *trainlm* (Levenberg-Marquardt back propagation) as the training function. Calculations were performed with the aid of Minitab and Excel.

4. RESULTS AND DISCUSSION

3.1 Linear Multiple Regression

The presented methodology was applied to model steam mass flow rate after actual data from the SSG. Data were already assessed using Block 1 suggested steps, which resulted in the assembling of Set_2 as a selected data set. The following steps were used through Block 2 chain of assessments, with the calculation of linear multiple regressions for each scenario (letter G) and the normalized mean absolute error (NMAE), as shown in Figure 5.

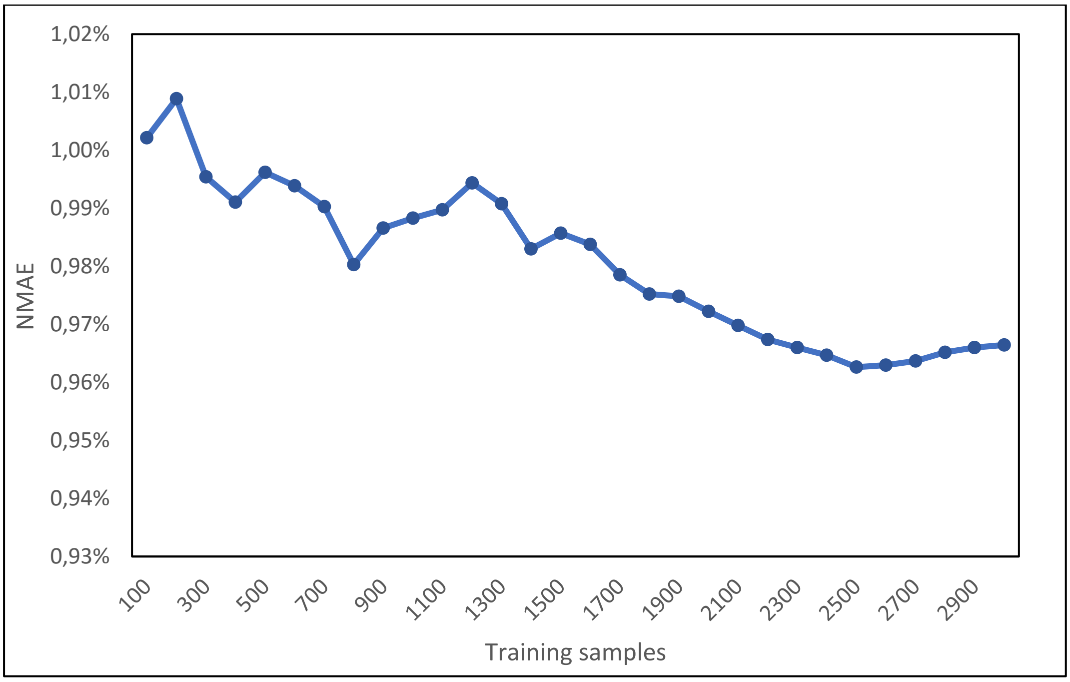

Figure 5 – Regression evaluation for each scenario

Figure 5 presented the regression evaluation for each scenario, it is possible to observed a variation of performance, however this range it 0.05% and this does not allow to conclude with training set has the best performance. Since the scenarios presented NMAE values very similar, the regression scenarios were tested with Set_1 and Set_3 as showed by Figure 6.

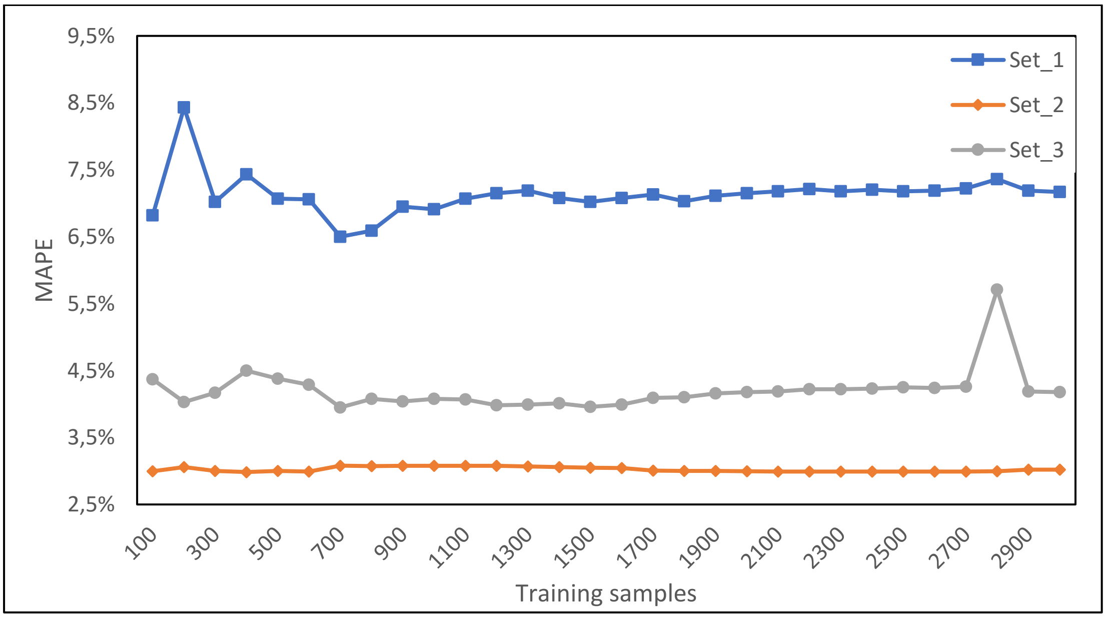

Figure 6 – Regression performance evaluation for Set_1 and Set_3

From Figure 6 it is possible to conclude that the regression has a best performance for Set_2, followed by Set_3 and Set_1. However, it is not possible to define each training set presented a best performance on the regression model.

3.2 Artificial Neural Network

Since the regression evaluation does not allow to conclude which scenario has the best performance, the same analysis was made to the ANN study. Four scenarios were evaluated as showed in Figure 7.

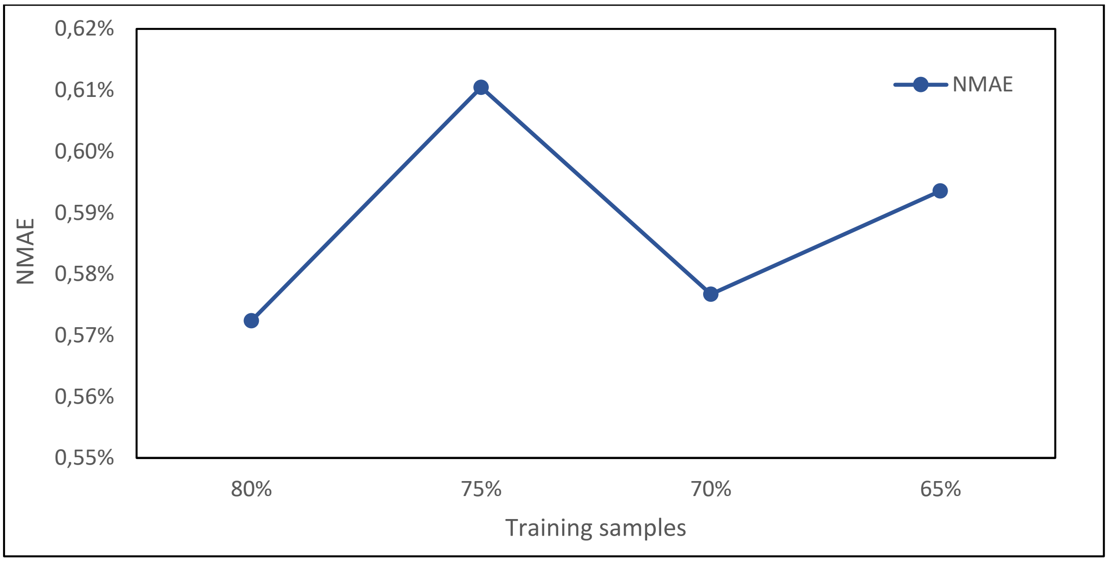

Figure 7 – ANN evaluation for each scenario

The best scenario was composed by 80% of training data and 20% of test data with 0.57% of NMAE, however the others scenarios presented an NMAE values similiter to the first scenario. Figure 8 presents an analysis of the ANN performance for Set_1 and Set_3 to define the best scenario.

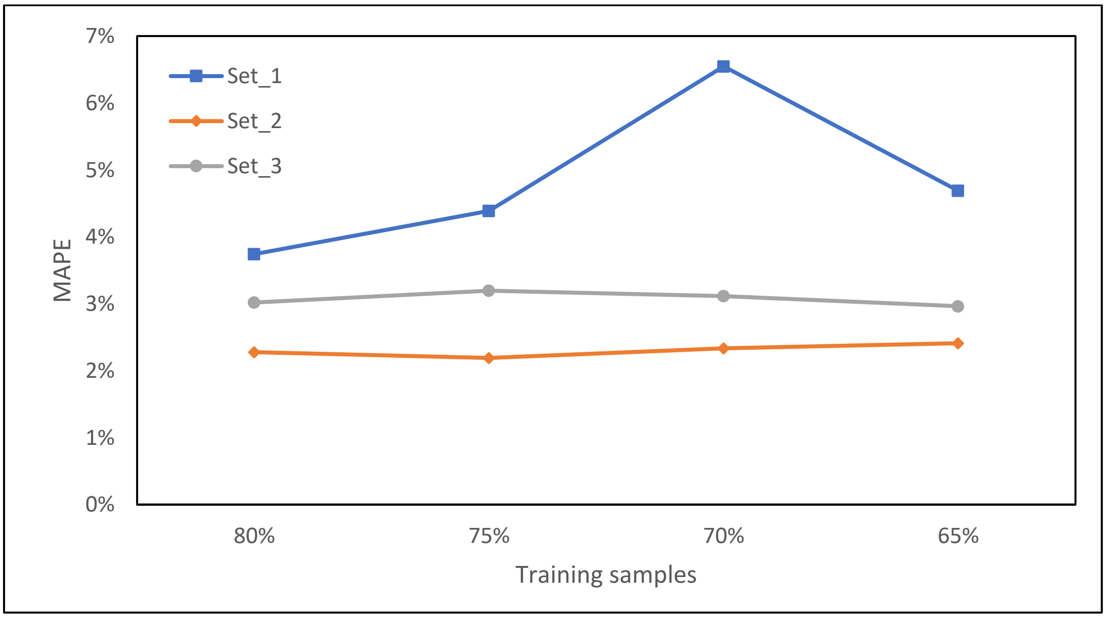

Figure 8 – ANN performance evaluation for Set_1 and Set_3

The analysis of the ANN performance for the sets 1 and 3 allow to conclude that the best scenario is de first one, with 80% of data for training and 20% for test. After the scenarios evaluation, the ANN architecture was study by the neuron numbers variation. Six ANN were compared and evaluated by the MAPE value for each set as showed in Figure 9.

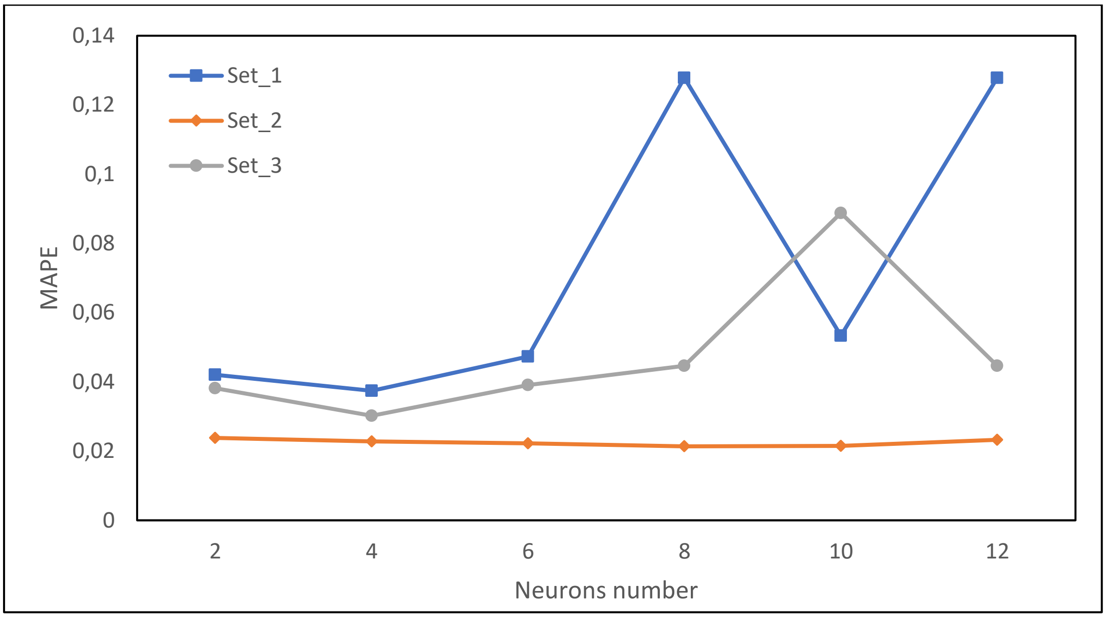

Figure 9 – ANN architecture performance

According to Figure 9, the ANN with four neurons presented the lower MAPE with 3.74% for Set_1, 2.28% for Set_2 and 3.02% for Set_3. The ANN outputs were the values of steam flow. These estimated values were compared to the observed values of the real plant. The ANN outputs presented an MSE value lower than 0.01% in 55% of the samples in Set_2. The others 44% of the data set are between 0.011 and 1% of an MSE value.

Figure 10 presents the ANN evaluation in respect to Set_1 and Set_3 is presented in figures 5 and 6 (letter R).

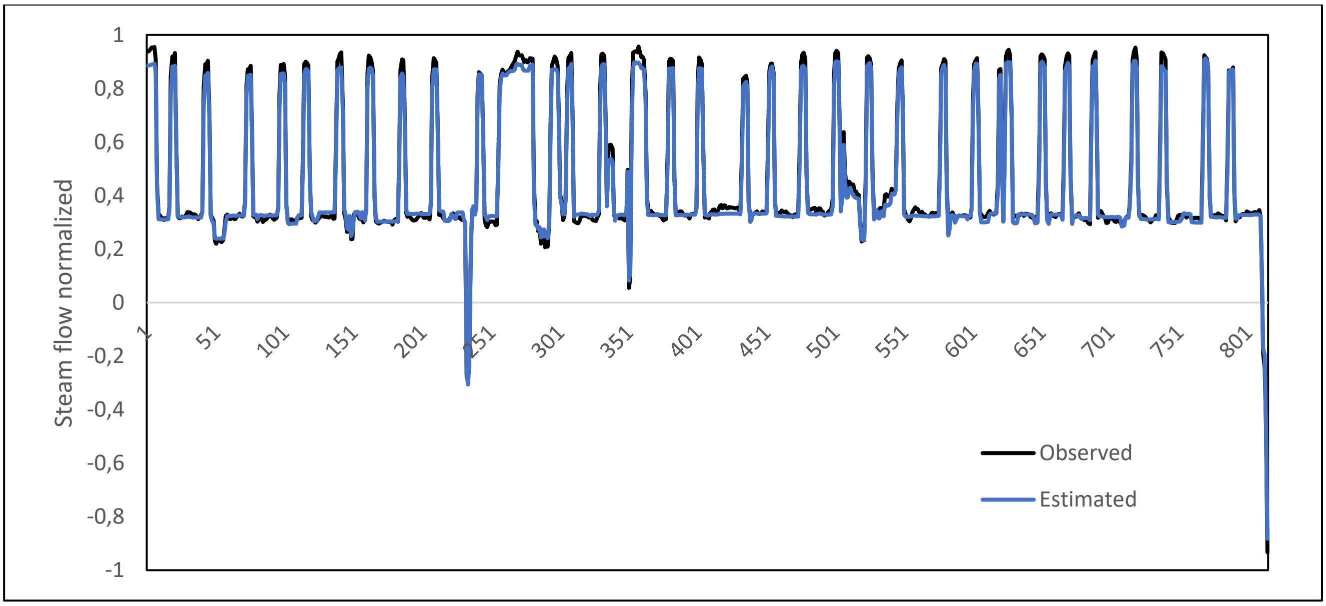

Figure 10 – ANN evaluation: Set_1

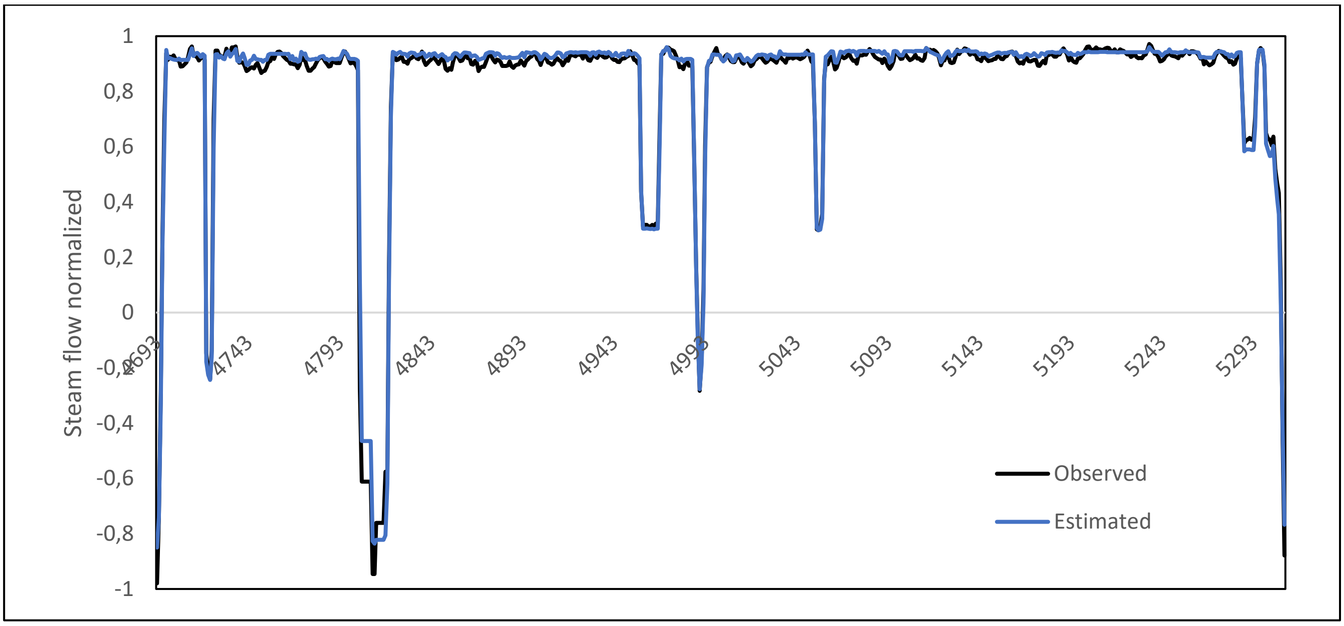

Figure 11 – ANN evaluation: Set_3

Figure 10 shows a satisfactory representation of the real problem, as approximately 83% of Set_1 with a MSE value lower than 0.1%, and 17% of the data set with a MSE value between 0.11 and 0.5%. Figure 11 shows that the ANN was able to estimate with high reliability the phenomena, since the curve of estimated values was very similar to the one of observed values. Set_3 presents 88% of the data set with an MSE value lower than 0.1%, the remaining 12% of the data set displayed an MSE value between 0.11 and 2.5%. However, some discrepancies can be observed when some extreme values were estimated. Finally, MAPE was also calculated to Set_2 in order to quantify and compare the ANN performance to Set_1 and Set_3. Results are presented in Table 4 (letter R).

Table 4 – ANN performance for the three sets of data

Set	MAPE
Set_1	3.74%
Set_2	2.28%
Set_3	3.02%

From Table 4 it is possible to conclude that the ANN has the better performance when estimate the steam generation values with input data from Set_2. It was expected since the ANN was developed based on data from this data set. The Set_1 and Set_3 also presented low values of MAPE, which indicates the quality of the ANN on estimate the steam generator value. The Set_2 presented data from the two-operation profile, 360 MW and 240 MW. Set_1 presented data from the 240 MW regime and Set_3 for the 360 MW. The reason for the Set_3 presented a lower error than Set_1 is due to the Set_2 has more data from 360 MW operation regime. The profile of each set can explain the reason for an lower error for Set_3 than Set_1.

The results in Table 5 and Figure 12 illustrate the ANN results compared with the linear multiple regression ones on the same data sets. Thus, it is possible to verify which the model has the best capacity to estimate the steam generation. Once the regression presented a similar performance for the different scenarios, it was chosen a similar scenario to compare with the best scenario of ANN.

Table 5 – ANN and regression performance

Model	Set	MAPE
Regression	*Set_1*	7.19%
	Set_2	3.02%
	Set_3	*4.19%*
ANN	*Set_1*	3.74%
	Set_2	2.28%
	Set_3	3.02%

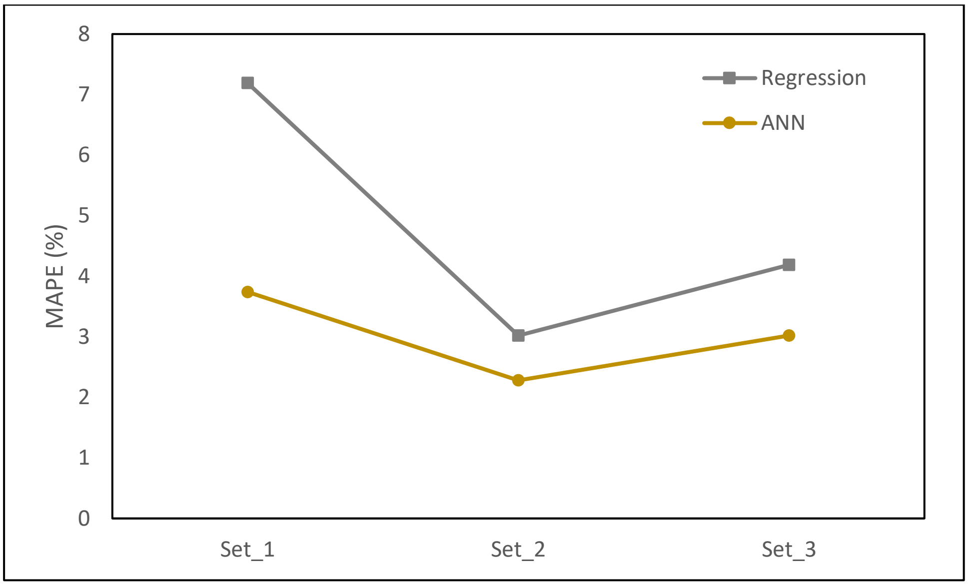

Figure 12 - ANN and regression performance

Table 5 and Figure 12 allow comparing ANN to regression performance, and show that the latest displayed higher MAPE values for Set_1 and Set_3. That model reached its best performance with data from Set_2, as expected, because it was developed based on that data set. The ANN model followed the same trend or general behaviour than the regression model, but with lower deviations, with Set_2 with the best performance followed by Set_3 and Set_1. Nevertheless, comparing to the regression model results, the ANN was clearly able to better estimate steam flow with Set_1 and Set_3. This indicates that the ANN model had better generalization ability than the regression model as it could estimate with higher accuracy steam flow from examples not used in its training.

5. CONCLUSIONS

The aim of this study was to develop an ANN capable to estimate the steam flow of a thermoelectric power plant. Real data of a thermoelectric power plant was obtained and used to develop the model. A reference model was created to evaluate the ANN performance. The regression model, used as reference, was made with different scenarios, where the size of the training group was changed. The same analysis was made for the ANN study. In both cases, the scenarios presented a similar NMAE values. To better evaluated the scenario performance in both cases, Set_1 and Set_3 were used to measured MAPE values.

The training group from scenario 1, from Set_2, was divided in training (80%) and test (20%) groups. The ANN architecture was evaluated regarding the number of neurons on the hidden layer and the low MSE value was for ANN with 4 neurons.

The results presented by the ANN model were compared with the reference model. The best performance of the regression was when this model was tested with the Set_2, with 3.02% of MAPE. When the model was tested with Set_1 .and Set_3 the MAPE values were higher, 7.19% and 4.19%. These values indicate the real performance of the model. The ANN model similar performance to the regression model in the case of Set_2, with 2.28% of MAPE. However, the MAPE for Set_1 and Set_3 were 3.74% and 3.02%. As such, the real performance of the ANN model was considerably better than the regression model, reducing the highest MAPE found (for Set_1) by approximately 50%. Thus, we found that in this problem the use of ANN can be justified by its superior performance.

Futures studies will be evaluated the performance of ANN and regression model regarding in other performance parameters and different operation conditions. Furthermore, fine tuning procedures regarding, the ANN inputs, training structure will be done aiming decreasing further the estimation error.

6. ACKNOWLEDGEMENTS

We acknowledge Energy of Portugal EDP for the financial and technical support to this project; Reichert acknowledges CAPES for his MSc grant; Smith Schneider acknowledges CNPq for his research grant (PQ 305357/2013-1).

7. REFERENCES

ANEEL. BIG - Banco de Informações de Geração. [online] Available at: http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm [Accessed 27 Jul. 2018].2018

DE, S. et al. Development of an artificial neural network model for the steam process of a coal biomass cofired combined heat and power (CHP) plant in Sweden. **Energy**, [s. l.], v. 32, n. 11, p. 2099–2109, 2007.

GEVREY, Muriel; DIMOPOULOS, Ioannis; LEK, Sovan. Review and comparison of methods to study the contribution of variables in artificial neural network models. **Ecological Modelling**, [s. l.], v. 160, n. 3, p. 249–264, 2003.

GHUGARE, S. B. et al. Prediction of Higher Heating Value of Solid Biomass Fuels Using Artificial Intelligence Formalisms. **Bioenergy Research**, [s. l.], v. 7, n. 2, p. 681–692, 2014.

HAYKIN, Simon. **Redes neurais: princípios e prática**. [s.l: s.n.].

MESROGHLI, Sh; JORJANI, E.; CHEHREH CHELGANI, S. Estimation of gross calorific value based on coal analysis using regression and artificial neural networks. **International Journal of Coal Geology**, [s. l.], v. 79, n. 1–2, p. 49–54, 2009.

NETO, Silva; BECCENERI, Carlos. TÉCNICAS DE INTELIGÊNCIA COMPUTACIONAL INSPIRADAS NA NATUREZA: APLICAÇÃO EM PROBLEMAS INVERSOS EM TRANSFERÊNCIA RADIATIVA. [s. l.], [s.d.].

SMREKAR, J. et al. Development of artificial neural network model for a coal-fired boiler using real plant data. **Energy**, [s. l.], v. 34, n. 2, p. 144–152, 2009.

STRUŠNIK, Dušan; GOLOB, Marjan; AVSEC, Jurij. Artificial neural networking model for the prediction of high efficiency boiler steam generation and distribution. **Simulation Modelling Practice and Theory**, [s. l.], v. 57, p. 58–70, 2015.

8. RESPONSIBILITY NOTICE