19th Brazilian Congress of Thermal Sciences and Engineering
November 06th-10th, 2022, Bento Gonçalves - RS - Brazil

ENC-2022-0220
PERFORMANCE OF COMPRESSION IGNITION ENGINES FUELED BY DIESEL-BIODIESEL-ETHANOL BLENDS USING MODELS BASED ON ARTIFICIAL NEURAL NETWORKS

Naiara Rinco de Marques e Carmo
Sergio Leal Braga
Florian Pradelle

Department of Mechanical Engineering, Pontifical Catholic University of Rio de Janeiro - (PUC-Rio), 225 St Marquês de São Vicente, Gávea, Rio de Janeiro, RJ, ZIP code:22453-900, BRAZIL
naiararinco14@gmail.com; slbraga@puc-rio.br; pradelle@puc-rio.br

Abstract. *This work aimed to develop different topologies of Artificial Neural Networks (ANNs) to obtain performance parameters of Internal Combustion Engines (ICE) fueled with blends of diesel – biodiesel – ethanol. Eight targets were evaluated: thermal efficiency, brake specific consumption of the blend and of ethanol, exhaust gas temperature, compression and expansion polytropic coefficients, maximum rate of heat released, and maximum pressure. The possible inputs were fuel consumption, ethanol content, lower heating value (LHV), among others. Different settings (type of network, number of neurons in the hidden layer, transfer function, among others) were considered to have ANNs with a high value of coefficient of determination (R^2) for training and test datasets, and smaller errors like Mean Absolute Percentage Error (MAPE). Two scenarios were compared: networks with the significant inputs of the correlation matrix and another with less inputs. For the networks with better performances, response surfaces were plotted for qualitatively analysis. It was possible to obtain models with good representation of almost all the mentioned parameters (obtaining R^2 values in the ranges of 0.606 – 0.995 for training and of 0.685 – 0.992 for test data). In consequence, the models can be used for optimization of the operation of the studied engine with such blends. Only the polytropic coefficients for compression and expansion could not be modelled properly, because the response surfaces did not represent the expected behavior.*

Keywords: *Internal combustion engines; Machine Learning; Diesel-Biodiesel-Ethanol Mixtures; Efficiency; Combustion.*

1. INTRODUCTION

The percentage of renewability of the national energy matrix is about four times higher than the rest of the world, and since 2014 this share of renewable sources has been increasing (EPE 2021). One of the causes is the investment in renewable sources, such as hydraulics, biomass, and more recently, solar and wind (Lima et al. 2020). When it comes to the transport sector, it is known that ethanol has been playing an important role, along with the increasing percentage of biodiesel in diesel oil, in compliance with the law 11.097/2005 (Brasil 2005). In addition, Brazil has a history of public policies for energy efficiency and conscious consumption, such as the Energy Efficiency Law (nº10295/2001), the Air Pollution Control Program by Motor Vehicles (PROCONVE), RenovaBio, among others (Pradelle 2017).

However, the main consumer of fossil fuels is the transport sector (EPE 2021), accounting for 60% of fossil fuel consumption in 2019 (EPE 2020). Since the 1980s, diesel oil has represented the most expressive share, around 50% of total consumption. Additionally, according to the PDE 2030, biodiesel will be among the biofuels that most grow, with a mean rate of 5.8% per year, until 2030. This is basically linked with the increase in its share in diesel oil, expected to reach 15% in 2030 (MME 2021). In addition, consumption of the entire transport sector tends to increase considerably until 2030. A demand increase of 1.9% per year is forecast for the entire sector, and of 2.8% only for diesel oil (MME 2021). Another fuel that has gained prominence is bioethanol, obtained from the fermentation of raw materials such as sugarcane, corn, among others (Hassan and Kalam 2013). This fuel has also contributed, mainly through flex-fuel spark ignition engine vehicles, for Brazil to stand out with a transport sector with relatively low emissions when compared to other economies, even though the increase in acquisition of cars due to increase in the population's incomes (Vieira do Nascimento 2014). It is possible to notice that Brazil has a mature and expanding market for the development of this biofuels market, considering the development of public policies that promote this value chain, as well as its territorial extension, presence of a favorable climate for the cultivation of various materials that can be used in the production of biodiesel and bioethanol (EPE - Empresa de Pesquisa Energética 2016).

An Internal Combustion Engine (ICE) is a machine that converts the chemical energy of a fuel (such as gasoline, natural gas, biodiesel, among others) into mechanical energy (Sonntag, Borgnakke, and Van Wylen 2002). The two most common types are spark ignition (SI) and compression ignition (CI)(Heywood 2018). Basically, emissions from an ICE fueled by carbon-based fuels tend to emit CO,CO_2, particulate matter (MP), HC,NO_x, e SO_x (Heywood 2018). In these substances there are greenhouse gases and acid rain triggers, along with toxic materials to humans. Currently, research in this area aims to improve the quality of fuels or the use of mixtures, for example. More restrictive environmental laws are another incentive, such as the Air Pollution Control Program by Motor Vehicles (PROCONVE) (Pradelle 2017). This shows that a better understanding of the effects of combustion contributes to combining low consumption, high efficiency and low emissions (Heywood 2018).

ANNs are a good research tool in the field of Combustion engines (Zhou et al. 2022), either by modeling the phenomenon itself, or by its effects and analysis parameters. It is possible to notice, for example, that percentages of the order of 70% for training, and 15% for validation and testing are often adopted. In addition, ANNs can assist in the performance analysis of an engine operating with only one fuel or mixtures (Singh, Jain, and Mahla 2020). It is also noted that feed-forward networks with backpropagation are applied the most, with a single input layer, and log-sigmoid, tangent sigmoid, and linear transfer functions. As for the training algorithm, the most popular are the Levenberg-Marquardt method (Manjunath, Puttaboregowda, and Chandrashekar 2018), and the gradient descent method. Such studies show the predictive capacity of ANN and its applications in research related to MCI, using simple topologies (Salam et al. 2020). ANNs can deal with non-linear relationships between variables of interest, and it is possible to make different adjustments to inputs and configurations of the network structure, allowing it to learn new features and generalize them to new situations, being used in different situations (Salam et al. 2020).

The purpose of this work is to build models able to characterize the characteristics and performance of internal combustion engines powered by diesel-biodiesel-ethanol mixtures using Artificial Neural Networks (ANN), namely the compression and expansion polytropic coefficients; the specific fuel and ethanol consumptions; the thermal efficiency; the maximum gross heat liberation rate; the maximum pressure; and the exhaust gas temperature. In addition, another parameter used to evaluate the performance of an engine is the thermal efficiency (η_{th}), the ability to convert the chemical energy of a fuel into effective power. Its value is, in general, between 30% and 50% (Heywood 2018). It is important to highlight that many works use expansion and compression coefficients, something that reinforces the importance of studies like this one. (Bao et al. 2020), for example, seeks to characterize this coefficient in the case of engines with direct hydrogen injection, and the work by Lee and Min (Lee and Min 2019) seeks to estimate this coefficient for engines operating in different regimes. In addition, the exhaust gas temperature represents the energy released during combustion, and it is important when it comes to exhaust gas treatment. Moreover, the maximum pressure is important in terms of selecting suitable alloys depending on the combustion process. Other interesting parameters are the maximum heat release rate (or maximum gross heat release rate), and the emissions profile (Seo et al. 2022). For that, the following input data were considered: Lower heating value (LHV); Lambda, Air-Fuel Ratio; Fuel consumption, specific consumption of ethanol (in the case of diesel-biodiesel-ethanol mixtures); Dry air consumption; Engine ambient and water temperatures; Relative humidity; Torque. The specific objectives are: (1) Build and assess the database' characteristics, (2) Identify the variables with significant impact on each target, (3) Compare models with many high impact variables and a reduced number of them, seeking to evaluate the robustness of neural networks considering the physics of the problem. Thus, we sought to obtain a model for each target that had the following characteristics: Be made up of inputs that have strong correlation with the target; Provide good generalization and low errors (for training and test data); Be made up with few inputs as possible.

2. METHODOLOGY

2.1. Building the database

The database was built based on data available in the literature (Pradelle 2017), characterizing the performance of combustion engines operating with diesel-biodiesel-ethanol mixtures with different properties that were also analyzed. The MWM 4.10 TCA (Euro III) model was used. The original database is composed of measured values and standard deviations, with a total of 75 occurrences. In this work, these measured values were considered three repetitions, with 225 occurrences in total: (1) without considering the standard deviation, (2) these values added to the double of the standard deviation and (3) the same values subtracted from the double of the standard deviation. The database was processed. First, for the Maximum gross heat liberation rate, it was stated that occurrences less than 20 J/° and greater than 300 J/° were removed from the database to avoid to create outliers, removing 4% of the total amount of data. These upper and lower limits were based on the rounded histogram values of this parameter. For the specific consumption of ethanol, null occurrences were removed, which represented 48% of the total database. The outputs were analyzed considering two scenarios: one of them without any treatment and the other considered the natural logarithm of the properties, with the network response then converted back to decimal base. These transformations were compared to smooth the distribution of outputs.

Next, a statistical analysis of the database was performed, which consisted of two main steps: first, boxplots were prepared for each variable to show possible outliers to be removed and the frequency of data in each quartile. A median close to the mean value of the data is sought because this indicates that no more data of one type will be presented to the neural networks. Then, qq-plot plots were created to verify that the data distribution is normal. Neural networks receive the normalized data because this treatment removes the effect of high data amplitudes, since it places them between -1 and 1. This also contributes to the patterns being learned evenly. Figure 1 shows the most common behaviors observed for this database.

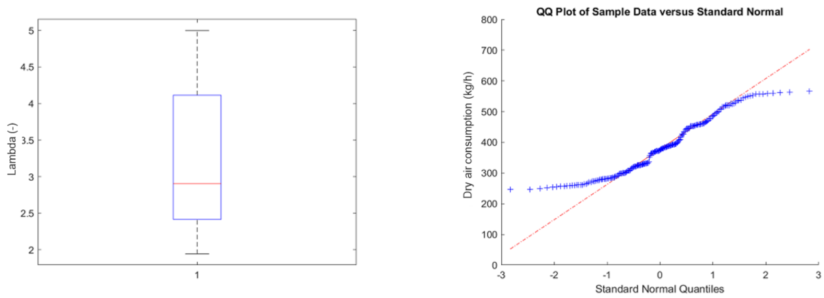

Figure 1: Boxplot of Lambda (left) and qq-plot of dry air consumption (right)
(database used to model the maximum gross heat liberation rate).

The boxplot indicates that networks that use this parameter will tend to learn the behavior linked to the first quartile. The qq-plot shows that the data do not perform a normal distribution.

After processing the data, a symmetric Pearson coefficient matrix (as in Equation 1) was set up between all the possible ones (R) for input data and the desired input. Here, x_i and y_i are the observed and predicted values, respectively, while $\bar{x}$ and $\bar{y}$ are the mean values. N is the number of occurrences while σ_x and σ_y are the standard deviation for both.

$$R = \sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{(N-1)\sigma_x \sigma_y} \quad \textbf{(1)}$$

Those with the strongest relationship, either positive or negative, within the range between -1 and +1, with the target are selected. To select the most suitable inputs for each target, the following criteria were adopted: 1) Inputs obtained through direct measurements and more extreme relationship with the target (close to 1 or -1); 2) Avoid repetition of many correlated quantities, such as dry air consumption, air-fuel ratio and lambda.

2.2. Artificial Neural Networks

Artificial Neural Networks (ANNs) are information processing computational structures. They are highly capable of recognizing non-linear patterns, making them widely applicable to several areas (HAYKIN 2001). It is important to point out that, before entering the network, the database must be normalized. Usually, several configurations of neural networks can be adopted and tested. They are topology, number of neurons in the hidden layer and in the output layer, number of inputs and targets, training algorithm, activation function in the hidden and output layers, number of epochs for training, desired convergence for the model, function objective to be minimized, minimum value of gradient to be met (HAYKIN 2001). In supervised learning, iterations happen from examples of expected inputs and outputs, so the method can detect patterns in a large enough dataset. The method does not represent the physical, chemical and thermodynamic transformations that actually occur (Salam et al. 2020). The best configuration is the one that the errors between the ANN responses and the original outputs are as small as possible, with good generalization. This is the ability of representing the general behavior of the response variable without learning inherent noise data, for example. Another point worth mentioning is that the characterization of the network must be such as to avoid overfitting (when the neural network learns noise from the data unnecessarily, making it rigid and overfitted to the real data) and underfitting (when errors in training and validation are already too high, and the network could not learn the behavior of the desired variable) (HAYKIN 2001).

A random database was set up, which considered the maximum and minimum values of each variable in the training, so that the network can learn the entire range of behaviors. This also allows the defined weights and bias to span any given order of data. This prevents the cyclic behavior inherent to combustion engines from influencing the definition of weights and bias in the neural network. Then, neural networks of different topologies were created, storing training, performance, and output data.

For all networks, 70% of the total data were considered for training, and 30% for testing. A subset dedicated to validation was not considered because they are relatively small databases, and there are studies in the literature along the same lines. Furthermore, there is only one hidden layer, the transfer function in the output layer was linear, and the stopping criterion consisted of training for 1000 epochs or convergence of the Sum of Squared Error (SSE) to 10^{-4} (Equation 2). Here, t^p and s^p are the target and the output from a neuron regarding the pattern p.

$$SSE = \frac{\sum_p (t^p - s^p)^2}{2} \qquad \textbf{(2)}$$

The search for the best performance model for each target was based on the variation of 5 configurations: 6 topologies (Feed Forward, Elman, Layer Recurrent, Cascade Forward, Function fitting, Generate pattern recognition), 17 training algorithms (Levenberg-Marquardt, BFGS Quasi-Newton, Resilient Backpropagation (B), Scaled Conjugate Gradient (CG), CG with Powell/Beale Restarts, Fletcher-Powell CG, Polak-Ribiére CG, One Step Secant, Variable Learning Rate B, Batch training with weight and bias learning rules, Bayesian regularization B, CG B with Fletcher-Reeves updates, Cyclical order weight/bias training, Gradient descent (GD) B, GD with adaptive learning rate B, GD with momentum B, Random order weight/bias training), 15 transfer functions (Competitive, Elliot sigmoid, Positive hard limit, Symmetric hard limit, Logarithmic sigmoid, Inverse, Positive linear, Linear, Radial basis, Radial basis normalized, Positive saturating linear, Symmetric saturating linear, Soft max, Symmetric sigmoid, Triangular basis), number of neurons in the hidden layer between the number of inputs and triple this plus 1 and the quantity and types of inputs. For example, in case of 2 inputs in a neural network, the number of neurons would be between 2 and 7. An experimental approach was chosen, since variables such as thermal efficiency, specific consumption and exhaust temperature have already been previously studied using ANNs (Jahirul, Saidur, and Masjuki 2010; Kurtgoz, Karagoz, and Deniz 2017; Xia et al. 2021), but no studies were found for the other parameters. The best ANNs present high values for during training and testing, along with small errors like SSE, Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (MAPE) (Equation 3). For N occurrences, t_i and o_i are the target and the calculated value for the i pattern.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|t_i - o_i|}{o_i}; \quad RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_i - o_i)^2} \qquad \textbf{(3)}$$

The input parameters varied according to the strength of R between input and target, and the type of arrangement. Models were built with many variables and with a smaller amount to assess how flexible the model would be.

For the ANNs with the best performance, response surfaces were built and, on each surface, some points from the original database were plotted. This makes it possible to guarantee the relationship between target and input, helping to validate the model. In addition to the algorithm for creating the surfaces themselves, it was necessary to calculate the fixed variables for each case, based on the median of the training data. The analysis of response surfaces considers the quality of the representation of the target's general behavior in terms of the phenomenon and its relationship with the considered inputs. In addition, the proximity of the original points of the database to the built surface is evaluated. Surfaces whose inputs had the strongest relationship with each target were prioritized. As nomenclature, the network topology data will be mentioned as "topology-training algorithm-transfer function in hidden layer-number of neurons in hidden layer". For example, a Feed-forward network ("newff") with 6 neurons in the hidden layer and using the Levenberg-Marquardt training algorithm ("trainlm") and the log-sigmoid ("logsig") transfer function has the following topology: newff-trainlm-logsig-6.

3. RESULTS AND DISCUSSION

3.1. Thermal efficiency

Table 1 shows the R coefficient between the input candidates in the database and the target. The selected parameters presented extreme values for R and are related to the transformation of chemical energy into effective power.

Table 1: R coefficient between input candidates and thermal efficiency.

Input candidate	R	Input candidate	R	Input candidate	R
Biodiesel content (%)	-0.02	**Fuel consumption (kg/h)**	**0.66**	Absolute humidity (%)	-0.132
Etanol content (%)	-0.159	Dry air consumption (kg/h)	0.351	Water HEX temperature (°C)	0.458
LHV (kJ/kg)	0.159	Air-Fuel ratio	-0.796	**Water temperature (°C)**	**0.754**
Spec. gravity (kg/m3)	0.127	**Lambda**	**-0.803**	Oil temperature (°C)	0.184
Speed (RPM)	-0.098	Exhaust Gas Pressure (bar)	0.362	Room Temperature (°C)	-0.195
Torque (N.m)	**0.794**	BMEP (bar)	0.794	Admission air pressure (bar)	0.285
Acceleration (%)	0.596	IMEP (bar)	0.766	Relative umidity (%)	0.087
Effective power (kW)	0.702	Maximum pressure increase rate (bar/°)	-0.536	Air admission pressure diference	0.34

It is expected that the increase in torque will influence the increase in efficiency, since this tends to increase the effective power delivered by the engine. In addition, the increase in engine water temperature is also linked to the increase in effective power, which indicates a proportional relationship with efficiency (Heywood 2018).

Table 2 compares the performance of the best networks with many and few inputs.

Table 2: Thermal efficiency: Configuration and performance of both networks.

Topology	Many inputs (With NL): newff-trainlm-tansig-8		Few inputs (With NL): newcf-traingdx-logsig-2	
Inputs	Torque, Fuel consumption, Water temperature, Lambda		Torque, Fuel consumption	
SSE	0.675		6.480	
	Training	Testing	Training	Testing
R	0.966	0.893	0.704	0.859
MAPE	0.015	0.033	0.046	0.034
RMSE	0.693	1.275	2.054	1.427

Both networks presented low error values and R very close to 0.9, as evidenced by other studies (Jahirul et al. 2010; Kurtgoz et al. 2017). Furthermore, in both cases the use of logarithm contributed to improve the performance of neural networks. The network with few inputs showed a behavior more consistent with the phenomenon in question, although the R parameters for training and testing and the SSE error were not the highest.

Figure 2 compares the output of the network with the real data for the model with few inputs. The model could provide outputs close to the real data, with a good generalization.

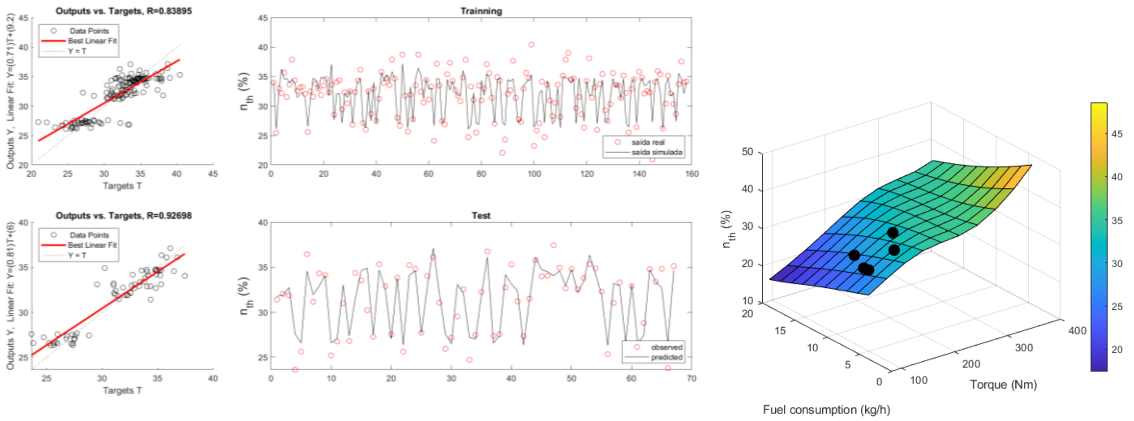

Figure 2: Thermal efficiency: Comparison between real and simulated data for training and testing (left) and main response surface (right) (newcf-traingdx-logsig-2).

An important detail for the network with more inputs is the very high values reached by this model (in some cases exceeding 60%). This would be an efficiency hardly reached by an engine operating in real conditions (Heywood 2018). Therefore, although it presented low errors and better R coefficients, the model does not have a good representation of reality. Another point that reinforces this argument is shown in Figure 3, where it is noted that the region with high torque and low lambda is well populated.

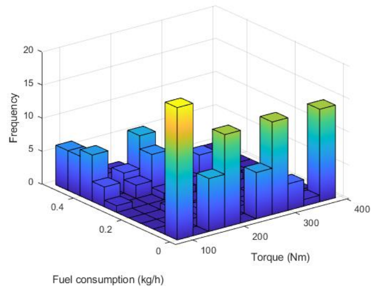

Figure 3: 3D Histogram for torque and fuel consumption.

3.2. Compression polytropic coefficient

During the compression and expansion stages, the relationship between pressure (p) and volume (V) can be considered as polytropic and depends on the value of the exponent (n), given by equation 5.

$$pV^n = constant \quad \textbf{(4)}$$

This coefficient is related to the characteristics of the air-fuel mixture during the thermodynamic cycle. For example, during compression, the temperature of the mixture tends to be lower than that of the chamber walls, so the value of n tends to be greater than the ratio (γ) between the specific heats at pressure (c_p) and at constant volume (c_v). As the compression progresses, the temperature of the mixture increases, to the point that $n < \gamma$ during expansion. This condition is satisfied until the initial stage of the cycle. The compression coefficient is widely used to calculate the ignition delay in compression ignition engines (Pradelle 2017).

Table 3 compares the performance of the best networks with many and few inputs.

Table 3: Compression polytropic coefficient: Configuration and performance of both networks.

Topology	Many inputs (Without NL): newff-trainbr-tansig-12		Few inputs (With NL): newff-trainlm-tansig-5	
Inputs	Ethanol content, Air-fuel ratio, Fuel consumption, Torque, Lambda, Water temperature, Relative humidity		Ethanol content, Water temperature	
SSE	0.320		1.216	
	Training	Testing	Training	Testing
R	0.983	0.823	0.936	0.581
MAPE	0.001	0.003	0.002	0.004
RMSE	0.002	0.005	0.003	0.008

It is noted that, although the relationship between the inputs and this target is considerably weak (generally less than 0.50 for R – data not shown), the performance was generally satisfactory. In addition, the less intense relationship between inputs and targets in this case justifies a high number of neurons in the hidden layer (approximately twice the number of inputs) for the networks to perform well.

Moreover, in the many inputs case, the use of logarithm did not show an improvement in the performance of neural networks. From Figure 4, the network performed well in training, but for the testing it did not adjust well to the Y=T line, something that is also reflected in the low R for the test.

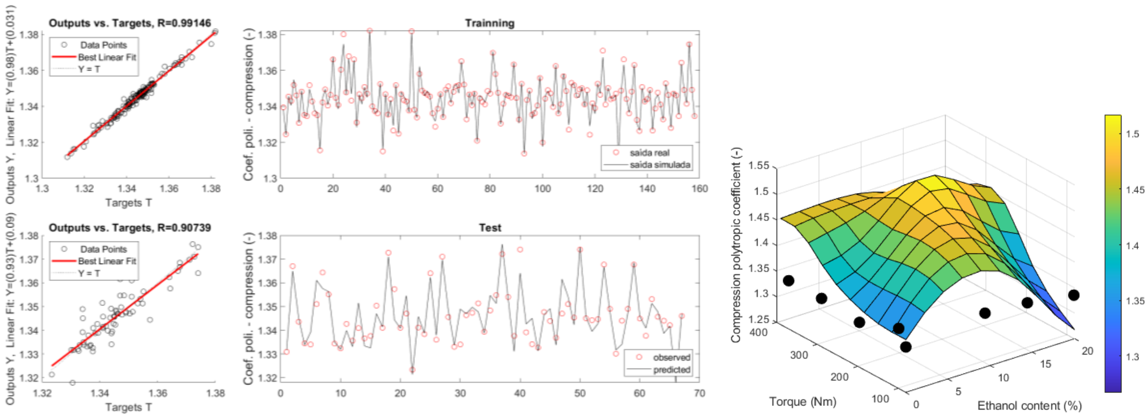

Figure 4: Compression polytropic coefficient: Comparison between real and simulated data for training and testing (left) and main response surface (right) (newff-trainbr-tansig-12).

The main problem detected in both networks is the very high values for the simulated output of the target (up to 1.6) or very low values (reaching 1, which corresponds to an isothermal process, which does not match reality). Values in the range of 1.3 ± 0.05 are expected (Heywood 2018). In Figure 2 it is also possible to see a peak for this coefficient for torque in a rage of 250 N.m, which is not observed in real cases. Therefore, none of the networks provided a model consistent with the phenomenon.

3.3. Maximum pressure

Table 4 compares the performance of the best networks with many and few inputs. It is seen that in both cases the values for the coefficient R were high and the errors very low, and the case with many inputs had better performance. In this case, using the logarithm for the output data offered better results. Figure 5 compares the output data of the neural networks and the real values for the network with many inputs.

Table 4: Maximum pressure: Configuration and performance of both networks.

Topology	Many inputs (With NL): newpr-traingdx-logsig-9		Few inputs (Without NL): newff-trainbr-tansig-5	
Inputs	Torque, Fuel consumption, Water temperature, Dry air consumption		Torque, Fuel consumption	
SSE	0.235		0.464	
	Training	Testing	Training	Testing
R	0.995	0.992	0.988	0.986
MAPE	0.011	0.013	0.017	0.018
RMSE	1.259	1.510	19.000	21.000

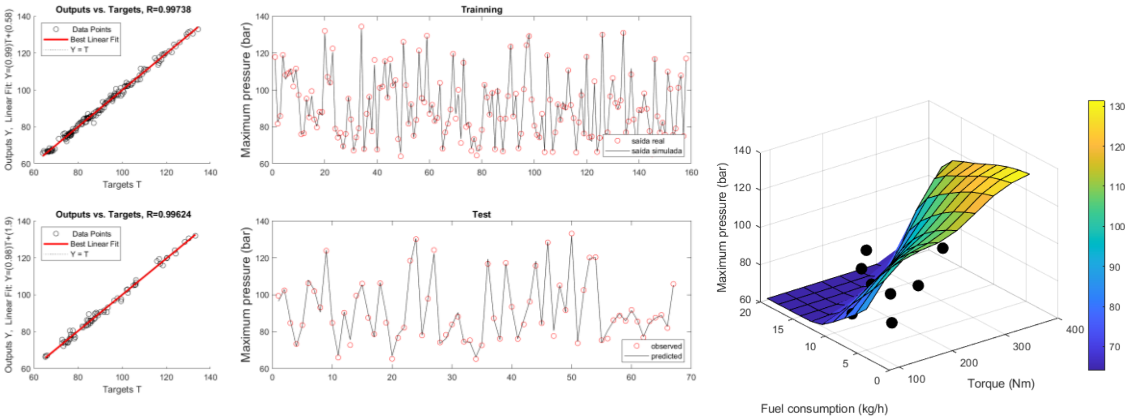

Figure 5: Maximum pressure: Comparison between real and simulated data for training and testing (left) and main response surface (right) (newpr-traingdx-logsig-9).

In both cases the ANN reached ranges for the target between approximately 60 and 140 bar, corresponding to the real values. Furthermore, it is expected that increases in consumption and torque simultaneously correspond to higher pressures, as represented by the network with few inputs. The same does not occur for the network with few inputs, since the model points to regions of high torque and low consumption as the ones with the highest maximum pressure, which is not consistent with the phenomenon. Therefore, the network with many inputs offered better performance.

3.4. Exhaust gas temperature

Table 5 compares the performance of the best networks with many and few inputs.

Table 5: Exhaust gas temperature: Configuration and performance of both networks.

Topology	Many inputs (Without NL): newff-trainbr-tansig-11		Few inputs (Without NL): newff-trainlm-logsig-5	
Inputs	LHV, Torque, Fuel consumption, Lambda, Water temperature		Torque, Fuel consumption	
SSE	0.215		0.949	
	Training	Testing	Training	Testing
R	0.995	0.988	0.976	0.979
MAPE	0.011	0.015	0.024	0.024
RMSE	51.000	71.000	108.000	98.000

Both networks presented R values above 0.95 for training and testing, also including low errors. It is observed here that the use of the logarithm did not contribute to better results. Figure 6 compares the output data of the neural networks and the real values. The output of the neural network was very close to the real values for both training and testing.

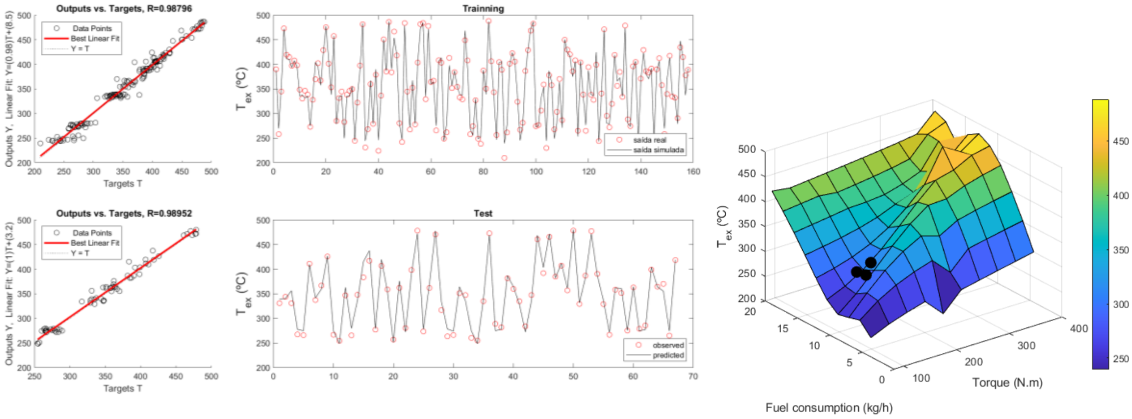

Figure 6: Exhaust gas temperature: Comparison between real and simulated data for training and testing (left) and main response surface (right) (newff-trainlm-logsig-5).

It is expected that the increase in exhaust temperature is linked to the increase in consumption and torque, something that is only represented by the network with few inputs. Therefore, this was the network that best represented the phenomenon.

3.5. Discussion

Table 6 lists the inputs selected to build successful models for each target.

Table 6: Relevant inputs for the ANNs developed for motor MWM 4.10 TCA.

Target	Input	R (training)	R (testing)	Quantitative analysis	Qualitativ e analysis
Compression Polytropic Coefficient	Ethanol content, Air-fuel ratio, Torque, Fuel consumption, Lambda, Water temperature, Relative humidity	0.936-0.983	0.581-0.823	No	No
Expansion. Polytropic Coefficient	Ethanol content, Air-fuel ratio, Torque, Fuel consumption, Lambda, Water temperature	0.789-0.966	0.822-0.963	No	No
BSFC	Ethanol content, Torque, Lambda, Water temperature	0.984	0.959	Yes	Yes
BSFC (ethanol)	Ethanol content, Torque, Lambda, Water temperature	0.879	0.770	Yes	Yes
Thermal eff.	Torque, Fuel consumption	0.704	0.859	Yes	Yes
Max. Gross heat lib. Rate	Torque, Fuel consumption, Water temperature, Lambda	0.814	0.663	No	Yes
Maximum pressure	Torque, Fuel consumption, Water temperature, Dry air consumption	0.995	0.992	Yes	Yes
Exhaust gas temp.	Torque, Fuel consumption	0.976	0.979	Yes	Yes

From this round of analysis, satisfactory models were obtained for all targets, except for the two exponents of the polytropic. Note the influence of engine composition, torque, and water temperature in all models. It is worth noting that the parameters did not present a normal distribution, and in some cases, such as lambda, the information tends to be concentrated in certain quartiles, which can make it difficult to generalize neural networks. It is also possible to see that artificial neural networks find good applicability in the analysis of performance parameters in engines, and possibly also in the prediction of these parameters (Kara Togun and Baysec 2010).

4. CONCLUSION AND FUTURE WORK

This work demonstrated the high potential of using neural networks in the modeling of complex phenomena such as combustion. In addition, this tool was able to map many parameters relevant to understanding the performance of internal combustion engines, allowing to represent the influence of aspects related to the air-fuel mixture and thermodynamics of such a complex process. It was possible to find satisfactory models for the specific consumption, thermal efficiency, exhaust temperature and maximum pressure, the first three with results close to the literature. Here, the values of R^2 were between 0.704-0.995 for training and between 0.770-0.992 for testing. On the other hand, the maximum heat release rate presented less satisfactory performances, but still relevant. The R^2 of 0.663 for testing was very low, but from a qualitative perspective the model was satisfactory. On the other way, the compression and expansion coefficients of the polytropic did not present satisfactory results since the response surfaces showed behaviors that do not represent each phenomenon well. One of the possible causes for this is related to the size of the database, which presented some less populated regions.

As suggestions for future works, the importance of employing more robust databases is highlighted, especially for the polytropic exponent in the compression process, a parameter of great interest in many studies, in particular for the determination of ignition delay, and its modeling using neural networks required a high number of neurons to result in a satisfactory model. In addition, the use of tools such as cross-validation applied to a larger dataset can contribute to obtaining models with fewer neurons. In addition, a sensitivity analysis for each variable would be interesting in order to have more idea of which characteristics offered better results. In fact, studies usually focus on algorithms such as "trainbr" or "trainlm" or on functions such as "tansig" and "logsig", and this work has shown that there are other possibilities that can generate good quality models. One example was the "newpr" training algorithm for maximum pressure modeling. When using Neural Networks or some other Machine Learning model, it is common to compare it with other approaches, such as the Random Forest type, or statistical models, such as exponential smoothing. It would be interesting to make a comparative study, aiming to obtain positive and negative points in each method. It is even possible to combine different models, such as neural networks as a regression model, and genetic algorithms for optimization. New analyzes including the time variable would be interesting to assess the impact of engine performance over its lifetime. Models like this could even be used to support the design of engines, since they allow understanding the thermodynamics of the phenomenon.

5. ACKNOWLEDGEMENTS

This paper was done with the support of CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil) – Financing code 001. The authors would like to thank the CNPq/MCTIC for the financial support to the Department of Mechanical Engineering (DEM) at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) and the FAPERJ for the Cientísta do Nosso Estado (CNE) and the Jovem Cientísta do Nosso Estado (JCNE) grants awarded to Sergio Leal Braga and Florian Pradelle, respectively.

6. REFERENCES

Bao, L. Z., Sun, B. G., Luo, Q. H., Wang, X., and Niu, Q. Y. 2020. "Experimental Study of the Polytropic Index of the Compression Stroke for a Direct Injection Hydrogen Engine." *International Journal of Hydrogen Energy*, Vol. 45, No. 52, pp. 28196–203.

Brasil. 2005. "*LEI Nº 11.097, DE 13 DE JANEIRO DE 2005. Dispõe Sobre a Introdução Do Biodiesel Na Matriz Energética Brasileira; Altera as Leis Nºs 9.478, de 6 de Agosto de 1997, 9.847, de 26 de Outubro de 1999 e 10.636, de 30 de Dezembro de 2002; e Dá Outras Providências"*. Diário Oficial da União, seção 1, Brasília, DF, D.O.U. DE 14/01/2005, P. 8.

EPE. 2020. "Combustíveis Renováveis Para Uso Em Motores Do Ciclo Diesel." *Empresa de Pesquisa Energética - Nota Técnica DPG-SDB No. 01/2020* 1–18. < https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-467/NT_Combustiveis_renovaveis_em_%20motores_ciclo_Diesel.pdf>.

EPE. 2021. "BEN 2021 Relatório Síntese." 1–73. < https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2021>.

EPE - Empresa de Pesquisa Energética. 2016. *Energia Renovável: Hidráulica, Biomassa, Eólica, Solar, Oceânica*. edited by M. T. T. (coord.). Rio de Janeiro. < https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/livro-sobre-energia-renovavel-hidraulica-biomassa-eolica-solar-oceanica>.

Hassan, M. H., and Kalam, M. A. 2013. "An Overview of Biofuel as a Renewable Energy Source: Development and Challenges." *Procedia Engineering*, Vol. 56, pp. 39–53.

Haykin, Simon. 2001. *Redes Neurais: Princípios e Prática*. 2ª ed. trad. Paulo Martins Engel. Porto Alegre: Bookman, 2001.

Heywood, J. B. 2018. *Internal Combustion Engine Fundamentals*. 2ª ed. editado por I. McGraw-Hill Editions.

Jahirul, M. I., Saidur, R., and Masjuki, H. H. 2010. “Predictability of Artificial Neural Network (ANN) in Performance Prediction of a Retrofitted CNG Engine.” *International Journal of Mechanical and Materials Engineering*. Vol. 5, No. 2, pp. 268–75.

Kara Togun, N., and Baysec, S. 2010. “Prediction of Torque and Specific Fuel Consumption of a Gasoline Engine by Using Artificial Neural Networks.” *Applied Energy*, Vol. 87, No. 1, pp. 349–55. doi: https://doi.org/10.1016/j.apenergy.2009.08.016.

Kurtgoz, Y., Karagoz, M., and Deniz, E. 2017. “Biogas Engine Performance Estimation Using ANN.” *Engineering Science and Technology, an International Journal*, Vol. 20, No. 6, pp. 1563–70. doi: 10.1016/j.jestch.2017.12.010.

Lee, Youngbok, and Kyoungdoug Min. 2019. “Estimation of the Polytropic Index for In-Cylinder Pressure Prediction in Engines.” *Applied Thermal Engineering*, Vol. 158. doi: https://doi.org/10.1016/j.applthermaleng.2019.04.113.

Lima, M. A., Mendes, L. F. R., Mothé, G. A. Linhares, F. G. de Castro, M. P. P., da Silva, M. G., and Sthel, M. S. 2020. “Renewable Energy in Reducing Greenhouse Gas Emissions: Reaching the Goals of the Paris Agreement in Brazil.” *Environmental Development*, Vol. 33, November 2018. doi: 10.1016/j.envdev.2020.100504.

Manjunath, B. B., B. Puttaboregowda, and R. Chandrashekar. 2018. “Prediction of Performace Parameters Using Artificial Neural Network for 4-Stroke C i Engine Fueled with Esterified Neem Oil and Biogas.” *IOP Conference Series: Materials Science and Engineering*, Vol. 376, No. 1. doi: 10.1088/1757-899X/376/1/012038.

MME, EPE. 2021. “Plano Decenal de Expansão de Energia 2030.” *Ministério de Minas e Energia. Empresa de Pesquisa Energética*. < https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/plano-decenal-de-expansao-de-energia-2030>.

Pradelle, Florian Alain Yannick. 2017. “Use of Biofuels in Compression Ignition Engines – Potential of Diesel-Biodiesel-Ethanol Blends.”. Dissertation (Master's in Mechanical Engineering). - Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, 2017.

Salam, Satishchandra, Tushar Choudhary, Arivalagan Pugazhendhi, Tikendra Nath Verma, and Abhishek Sharma. 2020. “A Review on Recent Progress in Computational and Empirical Studies of Compression Ignition Internal Combustion Engine.” *Fuel*, Vol. 279, May. doi: 10.1016/j.fuel.2020.118469.

Seo, Jigu, Boseop Yun, Juwon Kim, Myunghwan Shin, and Sungwook Park. 2022. “Development of a Cold-Start Emission Model for Diesel Vehicles Using an Artificial Neural Network Trained with Real-World Driving Data.” *Science of The Total Environment*, Vol. 806. doi: https://doi.org/10.1016/j.scitotenv.2021.151347.

Singh, Someet, Anuj Jain, and Sunil Kumar Mahla. 2020. “Sampled-Data Model Validation: An Algorithm and Experimental Setup of Dual Fuel IC Engine.” *Fuel*, Vol. 279, July. doi: 10.1016/j.fuel.2020.118517.

Sonntag, R. E., C. Borgnakke, and G. J. Van Wylen. 2002. *Fundamentals of Thermodynamics*. 8a ed. United States: Wiley, 2013.

Vieira do Nascimento, D. M. 2014. “The Brazilian Experience of Flex-Fuel Vehicles Technology: Towards Low Carbon Mobility.” *WIT Transactions on the Built Environment*, Vol. 138, pp. 545–53. doi: 10.2495/UT140451.

Xia, Y., Li, Y., Liao, C., Liu, J., Wang, S., Qiao, J., and Zhang, S. 2021. “On the Quantitative Relationship of the In-Cylinder Heat to Work Conversion Process of Natural Gas Spark Ignited Engine under Steady State and Transient Operation Conditions.” *Energy*, Vol. 221. doi: 10.1016/j.energy.2020.119607.

Zhou, L., Song, Y.,Weiqi Ji, and Wei, H. 2022. “Machine Learning for Combustion.” *Energy and AI*, Vol. 7. doi: https://doi.org/10.1016/j.egyai.2021.100128.