JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES

The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Multiobjective optimization of operating conditions in an HCCI model using an artificial neural network trained on CHEMKIN simulations

Otimização multiobjetivo de condições operacionais em um modelo HCCI utilizando uma rede neural artificial treinada com simulações CHEMKIN

Article Info:
Editor handled this article: **Ibrahim Ademola Fetuga**
Article history: Received 2026-10-01 / Accepted 2026-11-20 / Available online 2026-11-24
doi: **10.18540/jcecvl12iss1pp25236**

Estêvão Faria Duarte Freitas
ORCID: https://orcid.org/0009-0003-6007-5119
Universidade Federal de Viçosa, Brasil
Email: estevao.freitas@ufv.br
Júlio César Costa Campos
ORCID: https://orcid.org/ 0000-0002-9488-8164
Universidade Federal de Viçosa, Brasil
Email: julio.campos@ufv.br
Álvaro M. Bigonha Tibiriçá
ORCID: https://orcid.org/ 0000-0002-3300-1988
Universidade Federal de Viçosa, Brasil
Email: alvaro.tibirica@ufv.br
Bianca Lamas Clem de Siqueira
ORCID: https://orcid.org/0009-0002-3465-665X
Universidade Federal de Juiz de Fora, Brazil
Email: 166603593769@estudante.ufjf.br
Pedro Ribeiro de Oliveira
ORCID: https://orcid.org/0009-0002-5993-3350
Universidade Federal de Viçosa, Brasil
Email: pedro.oliveira7@ufv.br
Antônio Marcos de Oliveira Siqueira
ORCID: https://orcid.org/0009-0008-9051-4633
Universidade Federal de Viçosa, Brasil
Email: antonio.siqueira@ufv.br
Rogério Fernandes Brito
ORCID: https://orcid.org/0000-0002-6833-7801
Universidade Federal de Itajuba, Brasil
Email: rogerio6497@gmail.com
Washington Orlando Irrazabal Bohorquez
ORCID: https://orcid.org/0000-0002-9762-0665
Universidade Federal de Juiz de Fora, Brazil
Email: washington.irrazabal@ufjf.br

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES

CC BY OPEN ACCESS

The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Abstract

Chemical-kinetic simulations of engines are accurate but expensive when many operating conditions must be screened. This study integrates CHEMKIN simulations of an HCCI engine model fuelled with four diesel surrogate fuels, an artificial neural network (ANN), multiobjective optimization with NSGA-II and an independent CHEMKIN check of the predictions. A data set of 973 valid runs (four fuels, EGR from 0 to 0.6, equivalence ratio phi from 0.30 to 0.80, seven speeds from 1600 to 3019 rpm) was split by EGR-phi pair into 722 training, 111 validation and 140 interpolation-test records. A feedforward ANN (8-64-32-16-5) predicted power, torque, specific fuel consumption (BSFC), thermal efficiency and NOx. On the test set, R-squared was at least 0.9991 for all outputs; the mean absolute percentage error (MAPE) was 0.10-0.58% for power, torque, BSFC and efficiency and 5.99% for NOx. NSGA-II maximized power and efficiency and minimized NOx over EGR and phi, and two conditions per fuel and speed were selected: a compromise (EGR 0.58-0.60, phi 0.78-0.80) and a low-NOx alternative (phi 0.30-0.38), which reduced the predicted NOx by more than 99% at a power loss of 52-60%. Fifty-six new CHEMKIN runs (eight EGR-phi configurations at seven speeds, absent from the training set) agreed with the ANN, with MAPE of 0.04-0.50% for four outputs and 6.19% for NOx. Combustion indicators (CA50, peak pressure and temperature, CO, HC) were analysed as complementary diagnostics, not as constraints; unburned hydrocarbons increased at low phi and high EGR. A comparison with the manual of a real diesel engine (EGR = 0, estimated phi) gave power and torque deviations of 3.0-5.3% but BSFC and efficiency deviations of 18.5-29.2%, and is not an experimental validation. The conclusions refer to the simulated model, and experimental validation remains necessary.

Keywords: Artificial neural network. CHEMKIN. HCCI. Diesel surrogate fuels. NSGA-II. Multiobjective optimization.

Resumo

Simulações cinéticas de motores são precisas, mas custosas quando muitas condições operacionais precisam ser avaliadas. Este estudo integra simulações CHEMKIN de um modelo de motor HCCI alimentado por quatro combustíveis substitutos do diesel, uma rede neural artificial (RNA), a otimização multiobjetivo NSGA-II e uma verificação independente das previsões com novas simulações CHEMKIN. Uma base de 973 execuções válidas (quatro combustíveis, EGR de 0 a 0,6, razão de equivalência phi de 0,30 a 0,80, sete rotações de 1600 a 3019 rpm) foi dividida por pares EGR-phi em 722 registros de treinamento, 111 de validação e 140 de teste de interpolação. Uma RNA direta (8-64-32-16-5) previu potência, torque, consumo específico (BSFC), eficiência térmica e NOx. No teste, R ao quadrado foi de pelo menos 0,9991 para todas as saídas; o erro percentual absoluto médio (MAPE) foi de 0,10-0,58% para potência, torque, BSFC e eficiência e de 5,99% para o NOx. O NSGA-II maximizou potência e eficiência e minimizou o NOx em função de EGR e phi, e duas condições por combustível e rotação foram selecionadas: uma de compromisso (EGR 0,58-0,60; phi 0,78-0,80) e uma alternativa de baixo NOx (phi 0,30-0,38), que reduziu o NOx previsto em mais de 99% com perda de potência de 52-60%. Cinquenta e seis novas simulações CHEMKIN (oito configurações EGR-phi em sete rotações, ausentes do treinamento) concordaram com a RNA, com MAPE de 0,04-0,50% para quatro saídas e de 6,19% para o NOx. Indicadores de combustão (CA50, pressão e temperatura máximas, CO, HC) foram analisados como diagnóstico complementar, não como restrições; os hidrocarbonetos não queimados aumentaram com phi baixo e EGR alto. A comparação com o manual de um motor diesel real (EGR = 0, phi estimado) resultou em desvios de 3,0-5,3% para potência e torque, mas de 18,5-29,2% para BSFC e eficiência, e não constitui validação experimental. As conclusões referem-se ao modelo simulado, e a validação experimental permanece necessária.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Palavras-chave: Rede neural artificial. CHEMKIN. HCCI. Combustíveis substitutos do diesel. NSGA-II. Otimização multiobjetivo.

Nomenclature

A/F – air–fuel ratio (dimensionless)
A2CH3 – surrogate component, named as in the kinetic mechanism
ANN – artificial neural network
ATDC – after top dead center
BMEP – brake mean effective pressure / bar
BSFC – specific fuel consumption / (g kWh^{-1})
CA50 – crank angle of 50% heat release / ° ATDC
CO – carbon monoxide / ppmvd
EGR – exhaust gas recirculation rate (dimensionless)
HC – unburned hydrocarbons / ppm
HCCI – homogeneous charge compression ignition
HMN – surrogate component, named as in the kinetic mechanism
LHV – lower heating value / (MJ kg^{-1})
MAE – mean absolute error
MAPE – mean absolute percentage error / %
NOx – nitrogen oxides / ppmvd
NSGA-II – non-dominated sorting genetic algorithm II
R^2 – coefficient of determination (dimensionless)
RMSE – root mean square error
rpm – revolutions per minute (engine speed)
D – distance to the ideal point (dimensionless)
fk – normalized objective k (1 = best, 0 = worst; dimensionless)
N – number of records
yi – CHEMKIN value
ŷi – ANN prediction
ȳ – mean of the CHEMKIN values
η – thermal efficiency / % (fraction in Equation 1)
Φ – fuel–air equivalence ratio (dimensionless)

1. Introduction

Homogeneous charge compression ignition (HCCI) combustion combines high fuel economy with very low NOx and particulate emissions, but it has no direct actuator for the start of combustion: ignition timing and the subsequent combustion phasing are governed by the chemical kinetics of the fuel and by the temperature, pressure and composition history of the charge (Duan et al., 2021). Fuel management, charge preparation and exhaust gas recirculation (EGR) are among the strategies used to shape that history (Duan et al., 2021).

Because ignition in HCCI is kinetically controlled, detailed chemical-kinetic modelling is a natural tool for exploring operating conditions. For diesel-range fuels, however, the real fuel contains too many species to be modelled directly, and simulations rely on surrogate mixtures of a few components that emulate selected chemical and physical properties of a target fuel (Pitz & Mueller, 2011; Kukkadapu et al., 2023; Kundu et al., 2021). Mechanisms for multicomponent surrogates can be very large and stiff, which makes their direct use in engine simulations computationally expensive (Kundu et al., 2021), and a substantial literature addresses how to build

The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

and reduce them (Yang & Yang, 2024; Li et al., 2026). The fidelity of a surrogate depends on its composition, on the kinetic mechanism and on the conditions under which it was validated; surrogates are therefore not equivalent to commercial diesel.

Artificial neural networks (ANNs) have been used as data-driven models that map engine and fuel variables onto performance and emission responses (Bhatt & Shrivastava, 2021). Zandie et al. (2023) developed a multi-input multi-output ANN that predicts 14 performance, emission, combustion and ignition features of diesel–biodiesel–gasoline blends, and Gad and Fawaz (2024) used an ANN to predict the performance and emissions of waste-cooking-oil biodiesel blends as a function of engine speed and blend ratio. Reviews of ANN applications cover diesel engines fuelled with biodiesel-based fuels (Hoang et al., 2021) and engine performance and emissions with alternative fuels, alongside response surface methods (Yusri et al., 2018); these reviews concern engines and fuels different from the HCCI model studied here. For HCCI engines, ANN applications have been reviewed (Veza et al., 2022), and machine-learning models have also been combined with constrained multiobjective optimization (Gharehghani et al., 2021). When the training data come from detailed simulations rather than measurements, an ANN can in principle replace repeated kinetic calculations in screening studies, provided that its accuracy is documented in the region of interest.

This work investigates whether an ANN trained on CHEMKIN simulations of an HCCI engine model fuelled with four diesel surrogate fuels can act as a fast surrogate of the simulator and support the selection of operating conditions that balance power, thermal efficiency and NOx emissions. The workflow integrates five steps: CHEMKIN simulation; ANN training and evaluation on EGR–Φ pairs withheld from training; NSGA-II optimization with three independent objectives; independent CHEMKIN runs at conditions derived from the optimization; and a complementary analysis of combustion indicators. A comparison with the manual of a real diesel engine is included only as a complementary plausibility check. The selected conditions are interpreted as compromise solutions within the investigated domain, and the limitations of the data and of the verification are discussed explicitly.

Section 2 describes the CHEMKIN data, the ANN, the optimization and the verification procedures. Section 3 presents the predictive performance, the Pareto-based selection, the CHEMKIN verification, the combustion indicators, the comparison with the engine manual, the computational cost and the limitations. Section 4 summarizes the conclusions.

2. Materials and Methods

2.1 CHEMKIN Model and Surrogate Fuels

The numerical data were generated with CHEMKIN simulations of an HCCI engine model run in ANSYS Chemkin. The engine geometry and the compression ratio listed in Table 1 were taken from the technical manual of a four-stroke diesel engine from TecQuipment, whose documentation is compatible with the model TD202. This work is based on numerical simulations and on data extracted from that manual: no experimental measurements were carried out. The chemistry was described with the diesel_4comp_MFL2022.cks kinetic mechanism, available in the ANSYS Chemkin library and documented in the ANSYS Chemkin manual (ANSYS, 2022). The initial pressure and temperature are the same in all runs; they are constant metadata of the data set and not inputs of the ANN.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Table 1 – Engine and simulation parameters.

Parameter	Value
Engine type	HCCI
Compression ratio	22:1
Cylinder bore / mm	69
Piston stroke / mm	62
Connecting-rod length / mm	104
Displacement / cm^3	232 (0.232 L)
Initial temperature / K	590
Initial pressure / bar	1.1843
Wall temperature / K	400
Initial crank angle / ° ATDC	−150
Final crank angle / ° ATDC	180

Note: geometry and compression ratio from the reference-engine manual (Section 2.7); the engine speed was varied between 1600 and 3019 rpm (Section 2.1).

Four diesel surrogate fuels were considered, identified in the data set as Substitute 1 to 4. Their compositions, in mole fractions, were read from the columns nc16h34, hmn, a2ch3, c6h5ch3 and nc7h16 of the prepared training file and are given in Table 2. The surrogates are used as representations of selected chemical and physical properties of diesel-range fuels (Kukkadapu et al., 2023; Kundu et al., 2021); no exact equivalence with any commercial diesel is assumed.

Table 2 – Composition of the surrogate fuels (mole fraction).

Fuel	nC16H34	HMN	A2CH3	C6H5CH3	nC7H16
Substitute 1	0.413	0.368	0.219	–	–
Substitute 2	0.450	0.250	0.150	0.150	–
Substitute 3	0.450	0.250	0.200	–	0.100
Substitute 4	0.600	0.250	0.150	–	–

Note: component names as in the mechanism; C6H5CH3 is toluene and nC7H16 is n-heptane; "–" means that the component is absent.

The simulated domain combined seven EGR rates (0 to 0.6 in steps of 0.1), five equivalence ratios (Φ = 0.30, 0.40, 0.50, 0.65 and 0.80) and seven engine speeds (1600, 1766, 2008, 2256, 2507, 2751 and 3019 rpm), i.e. 35 EGR–Φ combinations per fuel and 140 cases. A combustion audit analysed 980 runs and classified 973 as valid; these 973 runs form the data set. In the prepared data set, 133 cases have seven speeds and seven cases have six speeds. Power, torque, BSFC, thermal efficiency and NOx were extracted from each run; thermal efficiency is stored as a fraction (and shown in percent in the figures) and NOx is expressed in ppmvd.

A single lower heating value, LHV = 43.5 MJ kg^{-1} (4.35×10^{11} erg g^{-1}), was adopted for the four fuels. Under this assumption BSFC and the thermal efficiency η are not independent; they are related by Equation 1, with BSFC in g kWh^{-1}, η as a fraction and 3600 the conversion factor between kWh and MJ.

$$\mathrm{LHV_{impl}} = \frac{3600}{\mathrm{BSFC}\cdot\eta} \tag{1}$$

2.2 Data Preparation and Partition

The data were read from the CHEMKIN output tables and checked for missing or non-numeric values, non-positive speed, power, torque or consumption, efficiency outside the interval 0–1, negative NOx and duplicate records; no record was rejected in this step. The data were then partitioned by EGR–Φ pair, so that all fuels and speeds of a pair belong to the same subset and no pair is shared between subsets. Four pairs (EGR, Φ) = (0.1, 0.65), (0.2, 0.40), (0.4, 0.65) and (0.5,

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

0.40) were assigned to validation (111 records), five pairs (0.1, 0.40), (0.2, 0.50), (0.3, 0.65), (0.4, 0.40) and (0.5, 0.50) to the interpolation test (140 records) and the remaining 26 pairs to training (722 records) (Table 3). The test pairs lie in the interior of the domain, so the test measures interpolation between simulated conditions; the domain edges (EGR = 0 and 0.6, Φ = 0.30 and 0.80) are represented only in the training set.

Table 3 – Data set and ANN training settings.

Item	Value
Fuels	4 (Substitute 1–4)
EGR	0–0.6 (steps of 0.1)
Φ	0.30, 0.40, 0.50, 0.65, 0.80
Engine speed / rpm	1600, 1766, 2008, 2256, 2507, 2751, 3019
Runs audited / valid	980 / 973
Records: training / validation / test	722 / 111 / 140
EGR–Φ pairs: training / validation / test	26 / 4 / 5
ANN inputs	speed, EGR, Φ and five mole fractions (8)
ANN outputs	power, torque, BSFC, η, NOx (5)
Architecture; activation	8–64–32–16–5; SiLU
Optimizer; loss	AdamW (learning rate 0.001); Huber (δ = 1)
Batch size; maximum epochs; patience	32; 2000; 120
Best epoch (last epoch run)	322 (442)
Random seed	42

Note: the standardization scalers were fitted to the training set only; NOx was transformed with log(1 + x) before standardization.

2.3 Artificial Neural Network Development and Evaluation

The ANN is a feedforward multilayer perceptron implemented in PyTorch, with eight inputs (engine speed, EGR, Φ and the mole fractions of nC16H34, HMN, A2CH3, toluene and nC7H16) and five outputs (power, torque, BSFC, η and NOx). The hidden layers have 64, 32 and 16 neurons with the SiLU activation function (Table 3). Inputs and outputs were standardized with scalers fitted only to the training set, and NOx was transformed with log(1 + x) before standardization and back-transformed afterwards, which accounts for its range of more than four orders of magnitude (from 1.7 to 21,578 ppmvd in the data set). The network was trained with the AdamW optimizer, the Huber loss on the standardized targets and early stopping on the validation loss; the weights of the best epoch (322) were retained. A single training run (seed 42) was performed, so the variability associated with initialization and with the partition was not assessed.

The interpolation test set was reserved and used only for the final evaluation. Performance was quantified with the coefficient of determination R^2 (Equation 2), the mean absolute error (MAE), the root mean square error (RMSE), the mean absolute percentage error (Equation 3) and the mean bias (Equation 4), where y_i is the CHEMKIN value, $\hat{y}_i$ the ANN prediction, $\bar{y}$ the mean of the CHEMKIN values and N the number of records. Errors of the efficiency (MAE, RMSE and bias) are given in percentage points. The agreement with CHEMKIN measures the fidelity of the ANN to the simulator, not the accuracy of the simulator for a real engine.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (2)$$

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i|} \quad (3)$$

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \quad (4)$$

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

2.4 Multiobjective Optimization and Selection of Conditions

The ANN was used inside the NSGA-II algorithm (population of 120, 180 generations, seed 42) to explore EGR and Φ for each fuel and engine speed, i.e. for 28 groups. The decision variables are continuous and bounded by the minimum and maximum EGR and Φ observed for each fuel and speed in the data set. Three independent objectives were optimized simultaneously: maximize power, maximize thermal efficiency and minimize NOx. Torque and BSFC were monitored but are not objectives. Candidates with non-physical predictions (non-positive power, torque, BSFC or efficiency, efficiency above 1, or negative NOx) were discarded, and the non-dominated set (Pareto front) of each group was retained; a solution dominates another when it is at least as good in the three objectives and strictly better in at least one.

Two conditions were selected from the front of each group. The compromise condition minimizes the distance D to the ideal point (Equation 5), where f_k is the objective k (power, efficiency and −NOx) min–max normalized within the front (1 is the best and 0 the worst value), so that the three objectives have the same weight after normalization. The low-NOx alternative is the front point with the lowest predicted NOx. Candidates at a normalized EGR–Φ distance of at least 0.025 from the simulated grid were preferred, in order to favor conditions not simulated before, and the two points of a group were required to be separated by at least 0.055 in the scaled EGR–Φ space. The procedure yielded 56 selected points (28 groups × 2). The distance to the grid is not an estimate of the prediction uncertainty, and no constraint on CA50, pressure rise rate, CO or HC was imposed.

$$D = \sqrt{\frac{1}{3}\sum_{k=1}^{3}(1 - f_k)^2} \tag{5}$$

2.5 Independent CHEMKIN Verification

Eight EGR–Φ configurations, two per fuel, were derived from the optimization stage and simulated in CHEMKIN at the seven engine speeds, giving 56 runs (Table 7). They are not the 56 selected points: the same EGR–Φ pair was repeated at the seven speeds, and only nine of the 56 runs coincide exactly with points of the current selection. The audit of the files showed that none of the 56 conditions coincides with a training pair (tolerance of 10^{-5} in EGR and Φ), that all 56 runs are marked as completed in the results file and that the heating value is the same 43.5 MJ kg^{-1} adopted for the data set. The raw solver logs were not audited again, and novelty was checked only against the training set, not against the validation and test sets. The ANN predictions were compared with the 56 CHEMKIN results using the metrics of Section 2.3.

2.6 Combustion Indicators

Complementary indicators were extracted from the CHEMKIN outputs of the 973 valid runs: the crank angle of 50% heat release (CA50), peak pressure, peak temperature, CO, unburned hydrocarbons (HC) and NOx. They characterize the simulated conditions and are neither outputs of the ANN nor constraints of NSGA-II. Negative CA50 values indicate 50% heat release before the reference angle of 0° ATDC, provided that the crank-angle origin of the CHEMKIN output is confirmed. The pressure rise rate was not available in the main audit; the estimates of a later analysis, obtained with a 5° angular window, are reported only as exploratory values because they depend on the angular resolution and on the differentiation method.

2.7 Complementary Comparison with the Engine Manual

The ANN was also compared with the performance data of the reference engine (TecQuipment, model TD202) at six speeds (1766–3019 rpm). BSFC, efficiency and air–fuel ratio were transcribed from the table of the manual, whereas power and torque were read approximately from its curves. The equivalence ratio was estimated as Φ = 14.5/(A/F), where 14.5 is an assumed

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

stoichiometric air–fuel ratio, and EGR was set to zero; the ANN was evaluated at these conditions for the four fuels. This comparison is not a validation: the manual describes a real diesel engine, whereas the model describes an HCCI configuration with surrogate fuels, with different load, mixture preparation, ignition strategy and mean effective pressure.

3. Results and Discussion

3.1 Predictive Performance of the Neural Network

Training stopped at epoch 442, and the weights of epoch 322, with the lowest validation loss, were retained (Figure 1). Table 4 and Figure 2 summarize the performance on the interpolation test set (140 records). R^2 exceeded 0.9991 for all outputs. The MAPE was 0.10% for BSFC, 0.11% for efficiency, 0.47% for torque and 0.58% for power, and NOx had the largest relative error (5.99%) in spite of its R^2 of 0.9991. On the training and validation sets, R^2 ranged from 0.9987 to 0.9999 and from 0.9985 to 0.9998, and the NOx MAPE was 3.60% and 5.80%, respectively.

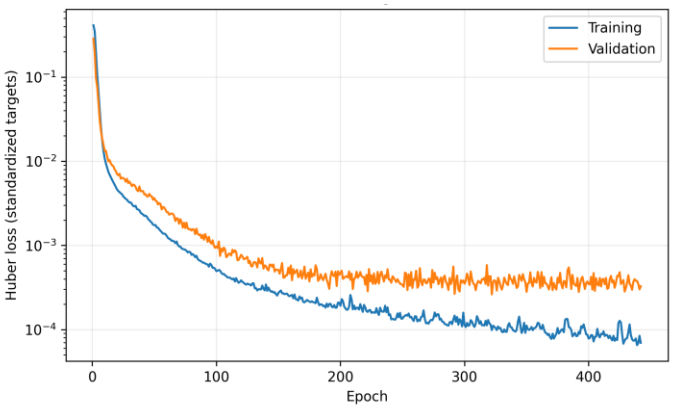


Figure 1 – Learning curves of the ANN (Huber loss on the standardized targets) for the training and validation sets.

Table 4 – ANN performance on the interpolation test set (140 records).

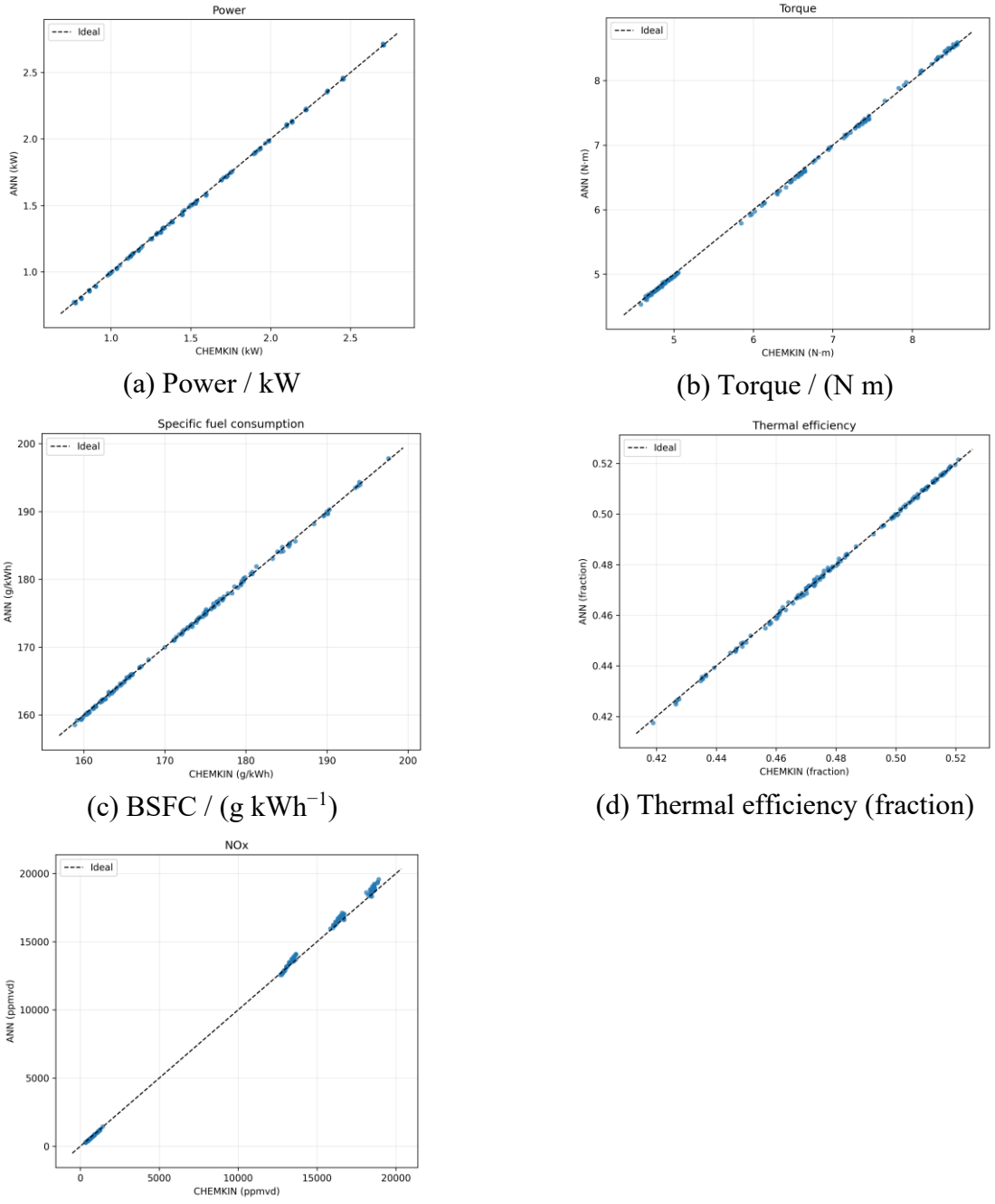
Output	R^2	MAE	RMSE	MAPE / %	Bias	SD of residuals
Power / kW	0.9997	0.0075	0.0090	0.58	−0.0064	0.0063
Torque / (N m)	0.9994	0.0284	0.0323	0.47	−0.0189	0.0261
BSFC / (g kWh^{-1})	0.9995	0.178	0.217	0.10	−0.059	0.209
Thermal efficiency / %	0.9994	0.0513	0.0656	0.11	−0.0117	0.0645
NOx / ppmvd	0.9991	176.7	228.2	5.99	97.1	206.6

Note: bias = ANN − CHEMKIN. SD of residuals = population standard deviation (N in the denominator) of the residuals ANN − CHEMKIN, computed from the unrounded residuals of the 140 test records. MAE, RMSE, bias and SD of the thermal efficiency are in percentage points.

The dispersion of prediction errors was further assessed through the standard deviation of the residuals. The values were 0.00633 kW for power, 0.02614 N m for torque, 0.20856 g kWh^{-1} for BSFC, 0.06454 percentage points for thermal efficiency and 206.56 ppmvd for NOx. These values characterize the variability of prediction errors across the 140 test samples and should not be interpreted as variability across independent ANN training runs. Because $RMSE^2 = bias^2 + SD^2$, the difference between RMSE and SD is the systematic part of the error: the bias accounts for about 50% of the mean squared error of power, 34% for torque and 18% for NOx, and for less than 8% for BSFC and efficiency.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

The test NOx values form two groups (Figure 2e): 56 records of the pairs (0.4, 0.40) and (0.5, 0.50), with 338–1,417 ppmvd, and 84 records of the other three pairs, with 12,678–18,913 ppmvd. The MAPE was 12.7% in the low-NOx group and 1.5% in the high-NOx group, and the mean bias was −76 and +213 ppmvd, respectively (values obtained from the test-set prediction file). Hence, the NOx MAPE is dominated by the relative error at low concentrations, and the ANN tends to overpredict NOx at high concentrations (mean bias of +97 ppmvd for the whole test set).



(a) Power / kW

(b) Torque / (N m)

(c) BSFC / (g kWh^{-1})

(d) Thermal efficiency (fraction)

(e) NOx / ppmvd

Figure 2 – CHEMKIN versus ANN-predicted values on the interpolation test set (140 records). Dashed line: ANN = CHEMKIN.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

3.2 Pareto Fronts and Selected Conditions

Figure 3 shows, as an example, the NSGA-II fronts of the four fuels at 2256 rpm in the NOx–power plane, with the thermal efficiency indicated by the colour scale. Along each front, power increases with NOx and the efficiency is highest at the low-power end and lowest at the high-power, high-NOx end. The compromise conditions lie at the beginning of the long, gently inclined section of the front, and the low-NOx alternatives at its low-power end.

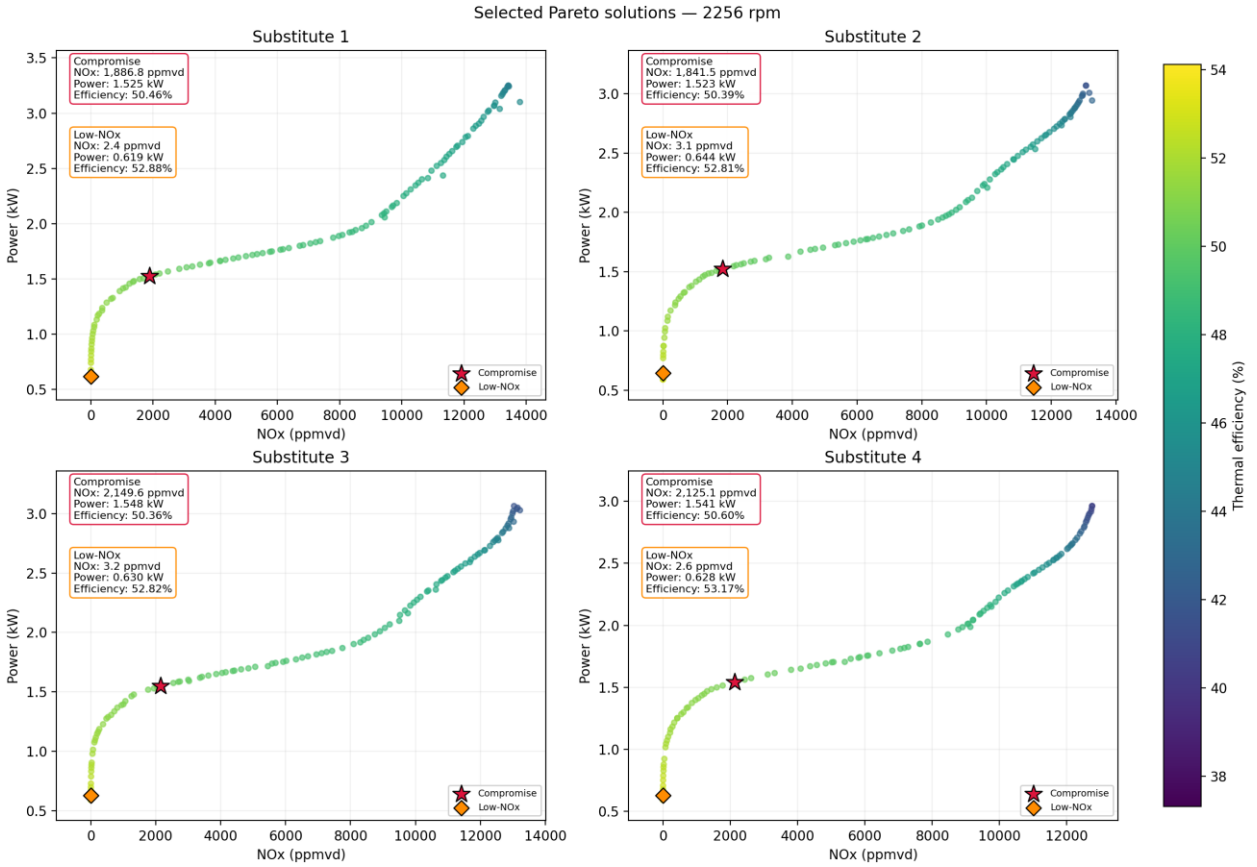


Figure 3 – Selected Pareto solutions at 2256 rpm for the four fuels (ANN predictions). Star: compromise condition; diamond: low-NOx condition; colour: thermal efficiency.

Table 5 summarizes the 28 compromise and the 28 low-NOx conditions. The compromise conditions have EGR of 0.58–0.60 and Φ of 0.78–0.80, and the low-NOx conditions have EGR of 0.56–0.60 and Φ of 0.30–0.38. Relative to the compromise condition of the same fuel and speed, the low-NOx alternative reduced the predicted NOx by 99.6–99.9%, reduced the power by 52–60% and increased the efficiency by 2.1–2.7 percentage points in the 28 groups. Table 6 lists the predictions at 2256 rpm.

Table 5 – Ranges of the selected conditions (28 conditions of each type; ANN predictions).

Condition	EGR	Φ	Power / kW	η / %	NOx / ppmvd
Compromise	0.575–0.599	0.781–0.800	1.01–2.12	49.3–51.5	1,319–2,967
Low-NOx	0.562–0.600	0.300–0.375	0.43–0.88	51.7–54.0	2.2–7.8

Note: ranges over the four fuels and seven speeds; η = thermal efficiency. ANN predictions, not CHEMKIN results.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Table 6 – Selected conditions at 2256 rpm (ANN predictions).

ID	Fuel	Condition	EGR	Φ	Power / kW	Torque / (N m)	BSFC / (g kWh^{-1})	η / %	NOx / ppmvd
P007	Substitute 1	Compromise	0.584	0.800	1.525	6.51	164.1	50.46	1,887
P008	Substitute 1	Low-NOx	0.600	0.317	0.619	2.61	156.5	52.88	2.4
P021	Substitute 2	Compromise	0.585	0.800	1.523	6.48	164.2	50.39	1,841
P022	Substitute 2	Low-NOx	0.593	0.326	0.644	2.72	156.7	52.81	3.1
P035	Substitute 3	Compromise	0.577	0.800	1.548	6.59	164.3	50.36	2,150
P036	Substitute 3	Low-NOx	0.572	0.303	0.630	2.66	156.7	52.82	3.2
P049	Substitute 4	Compromise	0.579	0.800	1.541	6.58	163.6	50.60	2,125
P050	Substitute 4	Low-NOx	0.596	0.318	0.628	2.66	155.6	53.17	2.6

Note: ANN predictions, not CHEMKIN results.

These conditions lie at the boundary of the investigated domain: Φ ≈ 0.80 and Φ ≈ 0.30 are the upper and lower limits of the Φ range, and EGR ≈ 0.6 is the upper limit of the EGR range. A selection at the boundary shows only that, within the domain and under the criterion used, no interior candidate offered a better value of the criterion; it does not show where the real optimum lies. In addition, the test pairs are interior (Section 2.2), so the accuracy of the ANN near the corners where the selected conditions are located was not evaluated with withheld pairs, which makes the independent CHEMKIN runs of Section 3.3 necessary. The low-NOx alternative is the extreme of the NOx objective; its large loss of power (Table 6) shows that it is not a practical compromise by itself.

3.3 Independent CHEMKIN Verification

Table 7 lists the eight configurations. Cases 1, 3, 5 and 7 have EGR of 0.579–0.597 and Φ of 0.782–0.800, similar to the compromise conditions, and cases 2, 4, 6 and 8 have EGR of 0.591–0.600 and Φ of 0.318–0.331, similar to the low-NOx conditions. Table 8 and Figure 4 compare the ANN with the 56 CHEMKIN runs. R^2 ranged from 0.9991 to 0.9999, and the MAPE was 0.04–0.05% for BSFC and efficiency, 0.35% for torque and 0.50% for power. NOx again had the largest relative error (MAPE of 6.19%; 4.72–7.28% among fuels).

Table 7 – Verification configurations and MAPE of the ANN per case (7 speeds each).

Case	Fuel	EGR	Φ	Power / %	Torque / %	BSFC / %	η / %	NOx / %
1	Substitute 1	0.597	0.782	0.67	0.10	0.06	0.07	3.47
2	Substitute 1	0.600	0.318	0.33	0.39	0.08	0.08	8.65
3	Substitute 2	0.582	0.800	0.79	0.35	0.02	0.02	1.88
4	Substitute 2	0.600	0.324	0.45	0.60	0.04	0.02	7.57
5	Substitute 3	0.581	0.800	0.48	0.15	0.02	0.02	1.47
6	Substitute 3	0.591	0.318	0.33	0.50	0.05	0.04	11.88
7	Substitute 4	0.579	0.800	0.65	0.26	0.03	0.07	2.09
8	Substitute 4	0.600	0.331	0.33	0.46	0.05	0.06	12.48

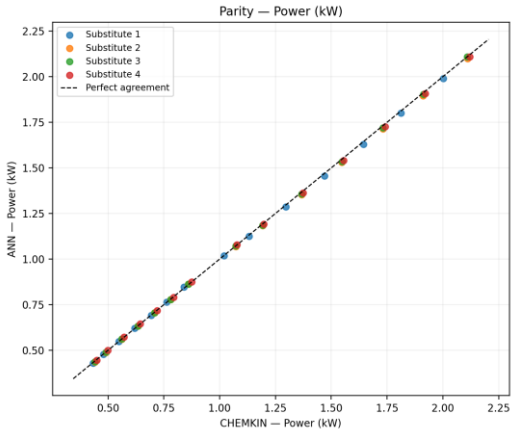
Note: MAPE of the ANN relative to CHEMKIN in each case; EGR and Φ as simulated.

Table 8 – Agreement between the ANN and the 56 CHEMKIN verification runs.

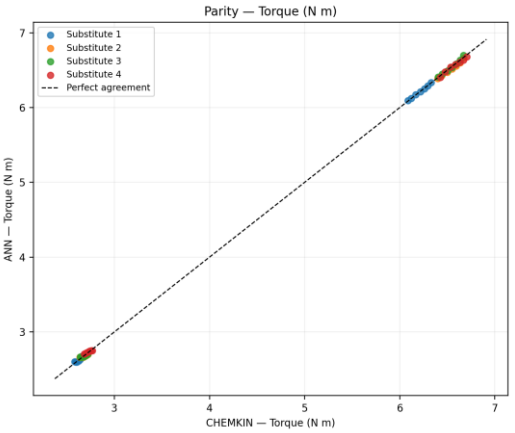
Output	R^2	MAE	RMSE	MAPE / %	Bias
Power / kW	0.9998	0.0062	0.0083	0.50	−0.0054
Torque / (N m)	0.9999	0.0137	0.0164	0.35	−0.0052
BSFC / (g kWh^{-1})	0.9995	0.069	0.093	0.04	−0.001
Thermal efficiency / %	0.9994	0.0241	0.0333	0.05	0.0056
NOx / ppmvd	0.9991	20.28	31.63	6.19	−0.70

Note: bias = ANN − CHEMKIN. MAE, RMSE and bias of the thermal efficiency are in percentage points.

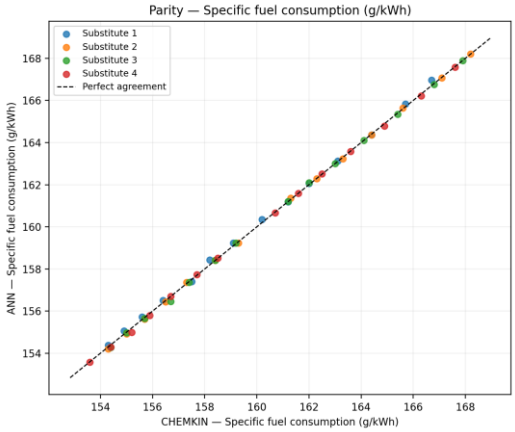
JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416



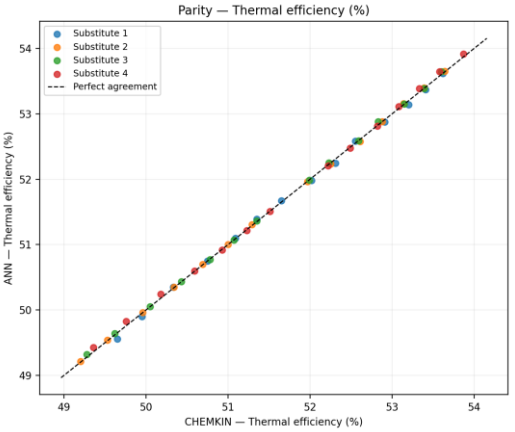
(a) Power / kW



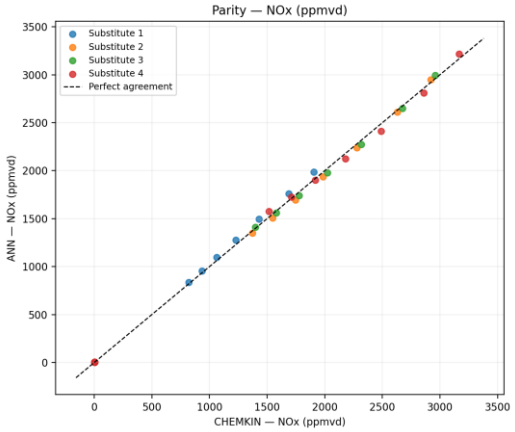
(b) Torque / (N m)



(c) BSFC / (g kWh^{-1})



(d) Thermal efficiency / %



(e) NOx / ppmvd

Figure 4 – ANN versus CHEMKIN for the 56 verification runs (eight EGR–Φ configurations at seven speeds). Dashed line: ANN = CHEMKIN.

Two points require care in the interpretation. First, the high overall R^2 partly reflects the separation between the two groups of configurations, which differ strongly in Φ and therefore in torque (about 2.6–2.8 N m at $\Phi \approx 0.32$ and 6.1–6.7 N m at $\Phi \approx 0.80$; Figure 4b); within the low-Φ cases, only seven speeds are available and the variation of torque and NOx with speed is small, so the case-level R^2 is low (for NOx, from −0.03 to 0.40) and is not informative. The MAPE and the absolute errors are more informative: in the four low-NOx cases the NOx MAPE was 7.6–12.5%,

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

corresponding to a mean absolute error of only 0.21–0.38 ppmvd at NOx values of a few ppmvd, with the ANN overpredicting. Second, the runs verify the ANN at eight configurations in the region of the selected conditions, not at each of the 56 selected points, and novelty was established only with respect to the training set.

3.4 Combustion Indicators

Table 9 summarizes the indicators of the 973 valid runs. The median CA50 was −25.18° ATDC, and the medians of peak pressure and temperature were 106.8 bar and 2567 K. The ranges are wide, which shows that the domain covers very different combustion regimes. The estimates of the pressure rise rate, which have a median of 8.19 bar $^{\circ -1}$ for the 5° window, are exploratory and sensitive to the angular resolution. These values are descriptive statistics of the simulated conditions and do not define limits of safe or stable operation.

Table 9 – Combustion indicators of the 973 valid CHEMKIN runs.

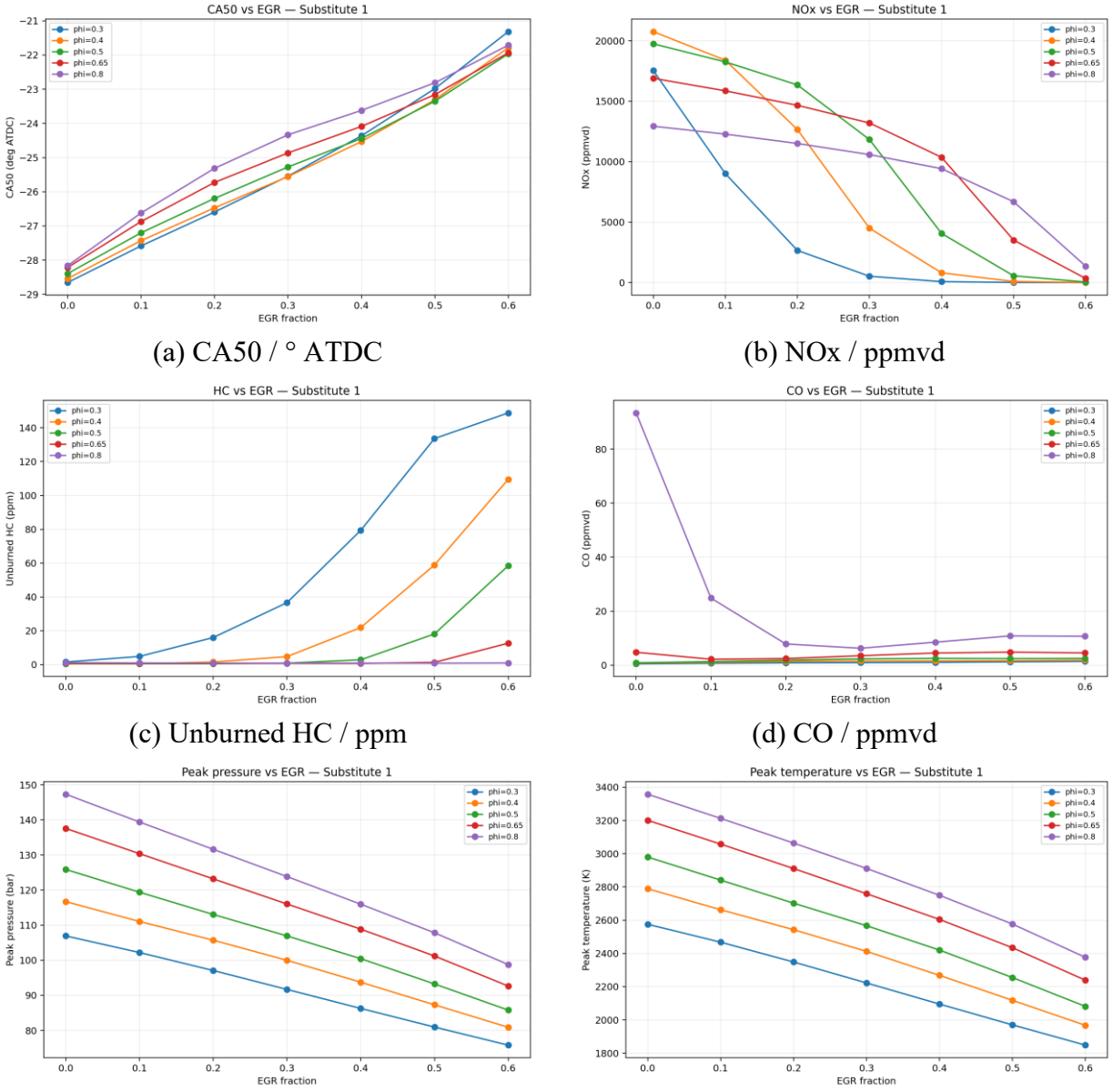
Indicator	Median	Minimum	Maximum
CA50 / ° ATDC	−25.18	−38.94	−19.80
Peak pressure / bar	106.8	75.55	154.0
Peak temperature / K	2,567	1,842	3,480
CO / ppmvd	2.076	0.342	301.1
HC / ppm	1.175	0.294	177.6
NOx / ppmvd	9,663	1.721	21,578
Pressure rise rate / (bar $^{\circ -1}$)[a]	8.19	2.57	20.37

Note: (a) exploratory estimate from a later analysis with a 5° window, not recalculated here; it depends on the angular resolution and on the differentiation method. CA50 convention as in Section 2.6.

Figure 5 shows the trends with EGR for Substitute 1 (means over the available speeds); the other fuels show the same qualitative behaviour in the summary table of the indicators. Increasing EGR retarded the combustion: at Φ = 0.30 the mean CA50 moved from −28.65° (EGR = 0) to −21.32° ATDC (EGR = 0.6). Peak pressure and temperature decreased with EGR and increased with Φ: at Φ = 0.80 the mean peak pressure was 147.3 bar without EGR and 98.8 bar at EGR = 0.6, and the mean peak temperature was 3358 K and 2376 K, respectively. NOx decreased strongly with EGR, from 17,537 to 1.89 ppmvd at Φ = 0.30 and from 12,933 to 1,358 ppmvd at Φ = 0.80. The costs of this reduction appear in the other indicators: at Φ = 0.30, unburned HC increased from 1.57 ppm (EGR = 0) to 148.8 ppm (EGR = 0.6), whereas at Φ = 0.80 it remained close to 1 ppm; CO was highest at Φ = 0.80 without EGR (mean of 93.4 ppmvd for Substitute 1; maximum of 301.1 ppmvd in the data set) and fell to about 10.7 ppmvd at EGR = 0.6.

These trends help to interpret the selected conditions, with the caveat that CO and HC were not simulated at the selected points. At the simulated condition nearest to the compromise points (EGR = 0.6, Φ = 0.80), the four fuels have NOx of 1,358–1,405 ppmvd, CO of 10.7–10.9 ppmvd and HC of 0.73–1.27 ppm, whereas at the condition nearest to the low-NOx points (EGR = 0.6, Φ = 0.30) NOx is only 1.9 ppmvd but HC is 148.7–165.5 ppm. The low-NOx alternative is therefore probably accompanied by high unburned hydrocarbons, a trade-off that was not part of the optimization because CA50, CO, HC and the pressure rise rate were neither ANN outputs nor constraints. The combustion indicators strengthen the physical interpretation of the results but do not fulfil the incorporation of these criteria in the optimization.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416



(a) CA50 / ° ATDC (b) NOx / ppmvd

(c) Unburned HC / ppm (d) CO / ppmvd

(e) Peak pressure / bar (f) Peak temperature / K

Figure 5 – Combustion indicators of Substitute 1 as a function of EGR for the five equivalence ratios (CHEMKIN results; mean over the available speeds).

3.5 Complementary Comparison with the Engine Manual

Table 10 gives the manual data used, and Table 11 and Figure 6 compare the ANN predictions at EGR = 0 and the estimated Φ (0.51–0.67). The MAPE relative to the manual was 3.0–5.3% for power and 3.7–5.1% for torque, but 24.4–29.2% for BSFC and 18.5–26.4% for efficiency. The ANN predicted a thermal efficiency of 41.0–47.0% against 34.8–38.0% in the manual (mean absolute errors of 6.7–9.5 percentage points) and a BSFC of 176–202 g kWh^{-1} against 240–270 g kWh^{-1}.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Table 10 – Data of the reference-engine manual used in the comparison.

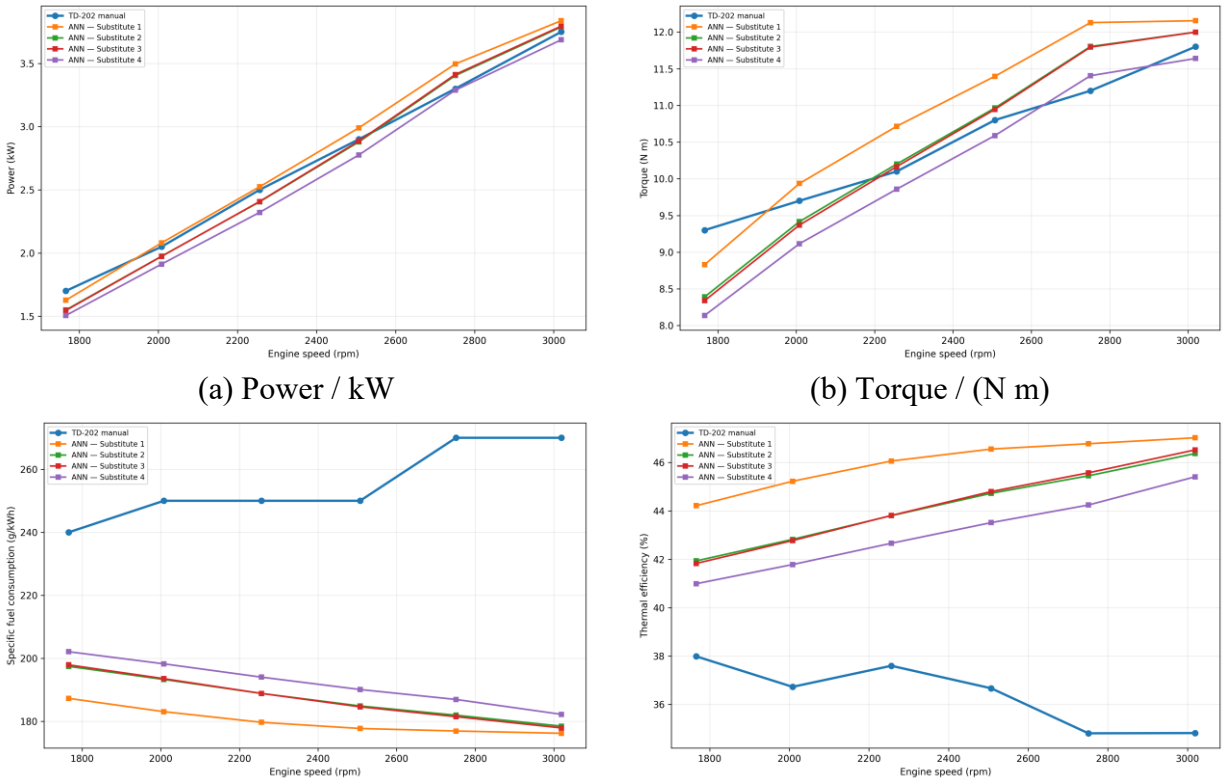
Engine speed / rpm	Power / kW[a]	Torque / (N m)[a]	BSFC / (g kWh^{-1})	η / %	A/F	Estimated Φ
1766	1.70	9.3	240	37.99	28.28	0.513
2008	2.05	9.7	250	36.73	25.68	0.565
2256	2.50	10.1	250	37.60	24.24	0.598
2507	2.90	10.8	250	36.67	23.03	0.630
2751	3.30	11.2	270	34.81	21.75	0.667
3019	3.75	11.8	270	34.82	21.87	0.663

Note: (a) approximate values read from the curves of the manual; BSFC, η and A/F transcribed from its table. Φ = 14.5/(A/F) with an assumed stoichiometric A/F of 14.5.

Table 11 – MAPE of the ANN relative to the manual (EGR = 0, six speeds).

Fuel	Power / %	Torque / %	BSFC / %	η / %
Substitute 1	3.0	5.1	29.2	26.4
Substitute 2	3.5	3.7	26.2	21.5
Substitute 3	3.6	3.8	26.3	21.6
Substitute 4	5.3	4.3	24.4	18.5

Note: deviations between different systems (HCCI model with surrogate fuels versus real diesel engine), not prediction errors of the ANN.



(a) Power / kW (b) Torque / (N m)

(c) BSFC / (g kWh^{-1}) (d) Thermal efficiency / %

Figure 6 – ANN predictions for the four fuels (EGR = 0, estimated Φ) and data of the TD-202 manual as a function of engine speed.

The proximity in power and torque must not be read as validation, because the manual values were read approximately from curves and the ANN has no load variable. The large deviations in efficiency and BSFC are consistent with differences of basis between the simulated model, whose power and torque are identified in the audit files as indicated quantities, and the real engine, and with the assumed heating value and stoichiometric air–fuel ratio; these contributions were not

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

isolated. BSFC and efficiency are linked by the common LHV (Equation 1), so the BSFC deviation should not be attributed to the ANN alone. The comparison therefore indicates only that power and torque are of the order of magnitude of the real engine under the assumptions made.

3.6 Energy Consistency of BSFC and Efficiency

The consistency between BSFC and efficiency in the data set was audited with Equation 1 for the 973 records. The implicit LHV was close to 43.5 MJ kg^{-1} in all records, ranging from 43.48343 to 43.51648 MJ kg^{-1}; the largest deviation from 43.5 MJ kg^{-1} was about 0.017 MJ kg^{-1} (0.038%), compatible with the rounding of the stored indicators. The mean values of the four fuels are given in Table 12, and the 56 verification runs are also consistent with 43.5 MJ kg^{-1}. This is an internal check of two derived quantities obtained with a common heating value; it is not an independent calorimetric or thermochemical determination, and the specific LHV of each surrogate was not determined in this study.

Table 12 – Implicit lower heating value of the data set (Equation 1).

Fuel	N	Mean implicit LHV / (MJ kg^{-1})
Substitute 1	244	43.49955
Substitute 2	244	43.49999
Substitute 3	243	43.50072
Substitute 4	242	43.49986

Note: values derived from stored BSFC and efficiency; they are not measured properties of the mixtures.

3.7 Computational Performance Assessment

The computational cost of the trained ANN was compared with that of the CHEMKIN simulations used to build the data set. The mean CPU time of a CHEMKIN simulation, taken from the Total CPUtime field of the output files, was approximately 3.44 s per case. The ANN inference time was measured with the 140 cases of the test set, using the trained network (8–64–32–16–5) in PyTorch 2.7.0+cu128 on Windows 11, with the processor identifier Intel64 Family 6 Model 165 Stepping 5 (GenuineIntel). The times include the normalization of the inputs and the post-processing of the outputs, and exclude the initial loading of the model and the training.

Table 13 summarizes the comparison and Table 14 the inference benchmark. An individual inference (one case per execution, 1,000 repetitions) required 0.8974 ms on average, with a standard deviation of 0.0703 ms. Processing the 140 test cases in a single batch (300 repetitions) required 1.0136 ms (standard deviation of 0.1091 ms), which corresponds to 0.00724 ms per case, about two orders of magnitude less than in individual inference. The mean time of an individual ANN inference is therefore approximately 3,833 times shorter than the mean CPU time of a CHEMKIN case, which is the main computational advantage of the surrogate model after training.

Table 13 – Computational time comparison between CHEMKIN and ANN.

Metric	CHEMKIN	ANN
Mean time per individual case / ms	≈ 3,440	0.8974
Timing measurement	CPU time	Elapsed time
Indicative computational speedup	1×	≈ 3,833×

Note: the CHEMKIN time is CPU time (Total CPUtime field of the output files) and the ANN time is elapsed (wall-clock) time; the speedup is indicative.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Table 14 – ANN inference benchmark results.

Metric	Individual inference	Batch inference
Number of cases per execution	1	140
Number of repetitions	1,000	300
Mean execution time / ms	0.8974	1.0136
Standard deviation / ms	0.0703	0.1091
Mean time per case / ms	0.8974	0.00724

Note: times include input normalization and output post-processing and exclude model loading and training.

The speedup must be read as indicative, not as a strictly equivalent comparison. The CHEMKIN time is CPU time reported by the solver, whereas the ANN time is elapsed (wall-clock) time, and the comparison does not assume that both were obtained on identical hardware; the number of cores, the memory and the use of accelerators are not reported here. In addition, the times refer only to inference after training. The training of the ANN, the generation of the 973 CHEMKIN runs of the data set and the CHEMKIN verification runs are one-off costs not included in Tables 13 and 14, so the reduction in prediction time should not be extrapolated to the total time spent in developing the methodology. Finally, the ANN is valid only within the domain and for the fuels on which it was trained (Sections 2.2 and 3.2), and the verification of the selected conditions still relied on new CHEMKIN simulations (Section 3.3); the surrogate model complements the physicochemical modelling and does not replace it.

4. Limitations

Simulated data: all results refer to CHEMKIN simulations of an HCCI model and to an ANN trained on them; no experimental measurements were performed, so neither the ANN nor the selected conditions are experimentally validated.

Domain and partition: the data set has 35 EGR–Φ pairs per fuel and seven speeds, the selected conditions lie at the boundary of the domain, and the test pairs are interior. A single training run and a single partition were used, without evaluation of the variability with the seed or with the pairs withheld.

NOx: the relative error of NOx is larger than that of the other outputs, especially at low concentrations, where the selected low-NOx conditions lie (NOx of a few ppmvd), whereas the lowest NOx in the test set is 338 ppmvd.

Verification: the 56 runs derive from eight configurations; only nine coincide exactly with the current selection, novelty was verified only against the training set, and the solver logs were not audited again.

Physical criteria: CA50, pressure rise rate, CO and HC were not ANN outputs or constraints, and the safe or stable operation of the selected conditions is not demonstrated; the high HC expected at low Φ and high EGR is a relevant trade-off.

Fuels and energy basis: the surrogates are not equivalent to commercial diesel, a common LHV of 43.5 MJ kg^{-1} was adopted for the four mixtures, and no specific LHV was determined for each of them.

Manual comparison: it uses EGR = 0, an estimated Φ and an assumed stoichiometric air–fuel ratio, with power and torque read approximately from curves, and compares different systems.

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

5. Conclusions

An ANN with eight inputs and five outputs reproduced the CHEMKIN results of an HCCI engine model fuelled with four diesel surrogate fuels on EGR–Φ pairs withheld from training ($R^2 \geq 0.9991$ for all outputs), with MAPE of 0.10–0.58% for power, torque, BSFC and efficiency and 5.99% for NOx, whose relative error is dominated by low concentrations.

NSGA-II with three independent objectives (power, efficiency and NOx) provided, for each fuel and speed, a compromise condition (EGR 0.58–0.60, Φ 0.78–0.80) and a low-NOx alternative (Φ 0.30–0.38) that reduced the predicted NOx by more than 99% at the cost of 52–60% of the power. Both lie at the boundary of the investigated domain and are ANN predictions; the low-NOx alternative is probably associated with high unburned hydrocarbons.

Fifty-six independent CHEMKIN runs at eight configurations absent from the training set agreed with the ANN (MAPE of 0.04–0.50% for four outputs and 6.19% for NOx). This supports the ANN as a fast surrogate of the simulator in the region of the selected conditions, but does not verify each selected point. After training, an individual ANN prediction took 0.8974 ms on average, against approximately 3.44 s of CPU time per CHEMKIN case (indicative speedup of about 3,800 times, with non-equivalent time measures); this gain does not include training or data generation and does not replace the physicochemical modelling. The comparison with the engine manual is complementary, showed deviations of 3.0–5.3% in power and torque and of 18.5–29.2% in BSFC and efficiency, and is not a validation. Experimental validation, a wider domain, explicit physical constraints (CA50, pressure rise rate, CO and HC) and the determination of the heating value of each surrogate are the next steps.

Acknowledgements

The authors thank the Federal University of Viçosa (UFV)- Federal University of Juiz de Fora (UFJF) and the Federal University of Itajubá (UNIFEI). This work was carried out with the support of the Minas Gerais State Research Support Foundation (FAPEMIG) and the Arthur Bernardes Foundation (FUNARBE) - grant PPE-00023-21, State Secretariat for Infrastructure and Mobility (Seinfra-MG), Center for Railway Technological Development of Minas Gerais (NDF), and CNPq.

References

ANSYS, Inc. (2022). *Ansys Chemkin-Pro Getting Started Guide*. ANSYS, Inc.

Bhatt, A. N., & Shrivastava, N. (2021). Application of artificial neural network for internal combustion engines: A state of the art review. *Archives of Computat ional Methods in Engineering*, *29*(2), 897–919. https://doi.org/10.1007/s11831-021-09596-5

Duan, X., Lai, M.-C., Jansons, M., Guo, G., & Liu, J. (2021). A review of controlling strategies of the ignition timing and combustion phase in homogeneous charge compression ignition (HCCI) engine. *Fuel*, *285*, 119142. https://doi.org/10.1016/j.fuel.2020.119142

Gad, M. S., & Fawaz, H. E. (2024). Artificial neural network based forecasting of diesel engine performance and emissions utilizing waste cooking biodiesel. *Scientific Reports*, *14*, 21980. https://doi.org/10.1038/s41598-024-71675-x

Gharehghani, A., Abbasi, H. R., & Alizadeh, P. (2021). Application of machine learning tools for constrained multi-objective optimization of an HCCI engine. *Energy*, *233*, 121106. https://doi.org/10.1016/j.energy.2021.121106

CC BY OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Hoang, A. T., Nižetić, S., Ong, H. C., Tarelko, W., Pham, V. V., Le, T. H., Nguyen, X. P., & Chau, M. Q. (2021). A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels. *Sustainable Energy Technologies and Assessments*, *47*, 101416. https://doi.org/10.1016/j.seta.2021.101416

Kukkadapu, G., Whitesides, R., Wang, M., Wagnon, S. W., Mehl, M., Westbrook, C. K., McCormick, R., Sung, C.-J., et al. (2023). Development of a diesel surrogate for improved autoignition prediction: Methodology and detailed chemical kinetic modeling. *Applications in Energy and Combustion Science*, *16*, 100216. https://doi.org/10.1016/j.jaecs.2023.100216

Kundu, P., et al. (2021). Implementation of multi-component diesel fuel surrogates and chemical kinetic mechanisms for engine combustion simulations. *Transportation Engineering*, *3*, 100042. https://doi.org/10.1016/j.treng.2020.100042

Li, C., et al. (2026). Chemical kinetic mechanism reduction and construction strategies of multicomponent fuels for internal combustion engines: A review. *Energies*, *19*(14), 3452. https://doi.org/10.3390/en19143452

Mueller, C. J., Cannella, W. J., Bays, J. T., Bruno, T. J., DeFabio, K., Dettman, H., Gieleciak, R. M., Huber, M. L., Kweon, C.-B., McConnell, S. S., Pitz, W. J., & Ratcliff, M. A. (2016). Diesel surrogate fuels for engine testing and chemical-kinetic modeling: Compositions and properties. *Energy & Fuels*, *30*(2), 1445–1461. https://doi.org/10.1021/acs.energyfuels.5b02879

Pitz, W. J., & Mueller, C. J. (2011). Recent progress in the development of diesel surrogate fuels. *Progress in Energy and Combustion Science*, *37*(3), 330–350. https://doi.org/10.1016/j.pecs.2010.06.004

Veza, I., Afzal, A., Mujtaba, M. A., Hoang, A. T., Balasubramanian, D., Sekar, M., Fattah, I. M. R., Soudagar, M. E. M., EL-Seesy, A. I., Djamari, D. W., Hananto, A. L., Putra, N. R., & Tamaldin, N. (2022). Review of artificial neural networks for gasoline, diesel and homogeneous charge compression ignition engine. *Alexandria Engineering Journal*, *61*(11), 8363–8391. https://doi.org/10.1016/j.aej.2022.01.072

Yang, S., & Yang, R. (2024). Development of a 5-component diesel surrogate chemical kinetic mechanism coupled with a semi-detailed soot model with application to engine combustion and emissions modeling. *SAE International Journal of Advances and Current Practices in Mobility*, *6*(4), 1895–1911. https://doi.org/10.4271/2023-24-0030

Yusri, I. M., Abdul Majeed, A. P. P., Mamat, R., Ghazali, M. F., Awad, O. I., & Azmi, W. H. (2018). A review on the application of response surface method and artificial neural network in engine performance and exhaust emissions characteristics in alternative fuel. *Renewable and Sustainable Energy Reviews*, *90*, 665–686. https://doi.org/10.1016/j.rser.2018.03.095

Zandie, M., Ng, H. K., Gan, S., & Muhamad Said, M. F. (2023). Multi-input multi-output machine learning predictive model for engine performance and stability, emissions, combustion and ignition characteristics of diesel-biodiesel-gasoline blends. *Energy*, *262*(Part A), 125425. https://doi.org/10.1016/j.energy.2022.125425

The Journal of Engineering and Exact Sciences – jCEC
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075 - ISSN: 2446-9416

Declarations

Authors' contributions

Estêvão Faria Duarte Freitas was responsible for the conceptualization, methodology, software development, simulations, formal analysis, data curation, visualization, and writing of the original manuscript. Júlio César Costa Campos, Washington Orlando Irrazabal Bohorquez, Rogério Fernandes Brito and Antônio Marcos de Oliveira Siqueira supervised the research and contributed to the development of the research project. Pedro Ribeiro de Oliveira contributed to specific aspects of the programming. Álvaro M. Bigonha Tibiriçá provided access to laboratory facilities. Bianca Lamas Clem de Siqueira contributed to the review and editing of the manuscript.

Funding Information

The authors received financial support from FAPEMIG - NDF - Project PPE-00023-21 to carry out the aforementioned research of the NDF group. This study was supported by the Brazilian National Council for Scientific and Technological Development (CNPq) through an undergraduate research scholarship.

Conflict of Interests

The authors declare that they have no competing interests.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Artificial Intelligence Use

During the preparation of this manuscript, the authors used AI-assisted tools to help identify and correct errors in Python code, translate the manuscript into English, and identify potentially relevant and recent bibliographic references. The authors reviewed the generated suggestions, verified the references used, and take full responsibility for the accuracy, integrity, and final content of the manuscript.

Handling editor

Editor handled this article: **Ibrahim Ademola Fetuga.**