Article

Comparison of Physics-Based, Semi-Empirical and Neural Network-Based Models for Model-Based Combustion Control in a 3.0 L Diesel Engine

Song Hu [1,2], **Stefano d'Ambrosio [1]**, **Roberto Finesso [1,]***, **Andrea Manelli [1]**,
Mario Rocco Marzano [1], **Antonio Mittica [1]**, **Loris Ventura [1]**, **Hechun Wang [2]** and **Yinyan Wang [2]**

[1] Department of Energy, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[2] College of Power and Energy Engineering, Harbin Engineering University, Harbin 150001, China
[*] Correspondence: roberto.finesso@polito.it; Tel.: +39-011-090-4493

check for
updates

Abstract: A comparison of four different control-oriented models has been carried out in this paper for the simulation of the main combustion metrics in diesel engines, i.e., combustion phasing, peak firing pressure, and brake mean effective pressure. The aim of the investigation has been to understand the potential of each approach in view of their implementation in the engine control unit (ECU) for onboard combustion control applications. The four developed control-oriented models, namely the baseline physics-based model, the artificial neural network (ANN) physics-based model, the semi-empirical model, and direct ANN model, have been assessed and compared under steady-state conditions and over the Worldwide Harmonized Heavy-duty Transient Cycle (WHTC) for a Euro VI FPT F1C 3.0 L diesel engine. Moreover, a new procedure has been introduced for the selection of the input parameters. The direct ANN model has shown the best accuracy in the estimation of the combustion metrics under both steady-state/transient operating conditions, since the root mean square errors are of the order of 0.25/1.1 deg, 0.85/9.6 bar, and 0.071/0.7 bar for combustion phasing, peak firing pressure, and brake mean effective pressure, respectively. Moreover, it requires the least computational time, that is, less than 50 μs when the model is run on a rapid prototyping device. Therefore, it can be considered the best candidate for model-based combustion control applications.

Keywords: model-based; control; diesel engine; ANN; physics-based model; semi-empirical model

1. Introduction

Nowadays, emission and fuel consumption reductions are the two main challenges for internal combustion engines, and in particular for diesel technology [1–3]. Many techniques have been proposed to achieve this aim, such as variable geometry turbocharger (VGT), high-pressure common rail systems [4–8], exhaust gas recirculation (EGR) [9], innovative combustion concepts such as HCCI and PCCI [10], and innovative combustion controls [11–14]. As far as these solutions are concerned, model-based combustion control is expected to make a significant contribution in the near future, thanks to the possibility of its integration with the emerging Vehicle-to-Everything (V2X) architectures. This conclusion is also supported by recent international projects, such as IMPERIUM H2020 [14], in which a significant reduction in CO_2 emissions will be achieved (−20%) for heavy-duty trucks, thanks to the use of model-based controllers coupled with V2X systems. The model-based control technology will also be boosted by the development of new sensors and by the increasing computational performance of the new multi-core processors that are now available for mobility applications.

Several advantages may be obtained from adopting a model-based controller instead of a conventional map-based one, such as the possibility of realizing a real-time optimization of the engine

parameters, and the need for less calibration effort. It should, in fact, be considered that modern internal combustion engines (ICEs), especially of the diesel type, require a heavy tuning procedure at the engine test bench [12,15] when a conventional map-based controller is adopted. Therefore, interest in model-based combustion control for diesel ICEs has been increasing more and more over the last few years. It is generally believed that the diesel technology will remain the best solution for light-duty and heavy-duty vehicles over the next decades [14], and research efforts to reduce fuel consumption and pollutants are therefore still required.

As a result of the increasing interest in the development of model-based combustion controllers, there is a need for accurate and low computational time demanding models. However, several types of modeling approaches, which can roughly be divided into physics-based (white box), and mathematical (gray-box or black-box) methods, can be used to this purpose. A detailed comparison of the different approaches is therefore required in order to identify the advantages and drawbacks of each modeling approach in view of their use for control-oriented applications.

As far as physics-based models are concerned, previous studies carried out by the authors [11,12,16] have shown that zero-dimensional approaches are the best solution for the development of model-based controllers. A zero-dimensional combustion model, which is capable of simulating the heat release rate, on the basis of the accumulated fuel mass approach and the in-cylinder pressure, on the basis of a single zone thermodynamic model, was presented in [17]. The accumulated fuel mass approach [18–21] is more physically consistent than the classical Wiebe function methodology [22] for heat release prediction. Highly accurate simulation results have been obtained using this model in many types of single- and multi-injection diesel engines [11–13,17,22].

Apart from physics-based models, mathematical models are also widely used for control-oriented applications. The aim of purely mathematical models is to identify the correlations between the input and output variables, without having knowledge of the physics of the system.

A first category of mathematical models includes empirical or semi-empirical approaches. These methods are easy to calibrate and can provide accurate results, if robust input variables are selected. Moreover, they require a short computational time, which makes them suitable for control applications. An example of a semi-empirical model for the estimation of combustion metrics is reported in [23].

Artificial intelligence systems, which include methods such as support vector machine (SVM) and artificial neural networks (ANNs) [13,24–28], can also be adopted for control-oriented applications. The use of ANNs has increased more and more in the last few years, due to their capacity to accurately predict the behavior of complex systems with short computational times. Many applications based on ANNs can be found in the literature on engine control and diagnostics [13,26–34]. From an analysis of the literature, it may be observed that the predictive performance of ANNs is influenced to a great extent by the selection of their main parameters, such as the number of layers, the training algorithm and the number of neurons [31–37]. The study reported in [37] pointed out that networks that are too small result in underfitting. However, larger networks may result in overfitting. Therefore, a sensitivity analysis is always needed in order to identify what the optimal number of neurons is for a specific application.

Although the previously mentioned modeling approaches have been reported in the literature in detail, to the best of the authors' knowledge there is a lack of studies concerning the comparison of the performances of physics-based, semi-empirical and ANN-based models, for the specific type of application investigated in this paper. The present study therefore addresses this research need and compares the performance of four different modeling approaches for the prediction of the main combustion metrics which are generally considered for combustion control applications, i.e., crank angle at which 50% of fuel mass has burnt (MFB50), peak firing pressure (PFP) and brake mean effective pressure (BMEP):

- Baseline physics-based model: this model was previously presented by the authors in [12] for control-oriented applications. However, in the present study, the model calibration procedure has been refined and the performance of the model has been improved with respect to the previously

reported version. In this approach, the tuning parameters are modeled by means of power-law functions, in which the input parameters are the main engine operating variables.

- ANN physics-based model: this is a new modeling approach, which is based on the use of ANNs to predict the tuning parameters of the aforementioned physics-based model.
- Direct semi-empirical model: in this approach, semi-empirical correlations, which are constituted by power-law functions, are used to directly estimate MFB50, PFP, and BMEP.
- Direct ANN model: this approach exploits feed-forward artificial neural networks to directly estimate MFB50, PFP, and BMEP. Details of the methodology used for the training and optimal selection of the number of neurons are also provided in this study.

The four different models have been assessed considering the same experimental dataset, for a 3.0 L FPT diesel engine for light-duty applications, and their performances have been compared under steady-state conditions and in transient operation over a WHTC.

In addition to the comparison of the performance of the previous models, another innovative contribution introduced in this paper concerns the methodology used for the optimal selection of the input parameters of the models. The proposed method is based on the sequential use of the Pearson correlation and partial correlation coefficients, which are used to identify the least and most robust correlation variables that should be excluded or included as model inputs. The use of the Pearson correlation and partial correlation analysis allows the computational effort required for the identification of the input parameters to be reduced significantly.

The paper is organized as follows. Section 2 summarizes the experimental setup and the engine specifications. Section 3 reports the description of the four investigated models. Section 4 is focused on the presentation of the methodology used for the selection of the model input variables. Finally, Section 5 reports the main results, which include:

- A comparison of the performance of the physics-based model tuned using the methodology introduced in this paper (based on Pearson correlation and partial correlation analysis) with that of the previous version reported in [12].
- A comparison of the accuracy of the four investigated models under steady-state conditions and in transient operation over a WHTC.
- A comparison of the performance of the four investigated models, in terms of required computational time, when they are run on an ETAS ES910 (ETAS, Stuttgart, Germany) rapid prototyping device.

This research is a first step towards the development of a model-based combustion controller, which will be developed and tested on the real engine, through rapid prototyping, in the near future.

2. Experimental Setup and Engine Conditions

The experimental tests were run on a Euro VI FPT F1C 3.0 L diesel engine (FPT Motorenforschung AG, Arbon, Switzerland), within a research project conducted in collaboration with FPT Industrial [11]. The activity was carried out in the dynamic test bench facility at the Politecnico di Torino. The main technical specifications of the engine are reported in Table 1.

Table 1. Main technical specifications of the FPT F1C 3.0 L diesel engine.

Engine Type	Number of Cylinders	Displace Ment	Bore × Stroke	Rod Length	Compres Sion Ratio	Valves per Cylinder	Turbo- Charger	Fuel Injection System
FPT F1C Euro VI diesel engine	4	2998 cm^3	95.8 mm × 104 mm	160 mm	17.5	4	VGT type	High pressure Common Rail

The fuel is characterized by a density of 835 kg/m^3 at 14 °C, a viscosity of 2 mm^2/s at 40 °C and a cetane number equal to 43. The considered injection pattern features two pilot shots and a main shot.

The engine is shown in Figure 1. It is equipped with a VGT turbocharger, a short-route cooled EGR system and a flap valve located on the exhaust side downstream from the turbine. Details about the layout of the engine, the test bench and the used sensors (including the main accuracy) have already been described in detail in [38] and are not repeated in this paper for the sake of brevity.
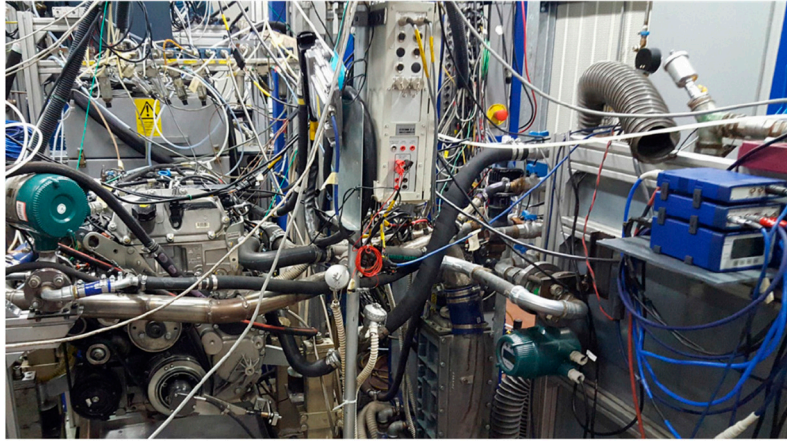

Figure 1. Picture of the FPT F1C 3.0 L Euro VI diesel engine installed at the Politecnico di Torino. The rapid prototyping device is located on the right side.

An ETAS ES910 rapid prototyping device, which is equipped with a main Freescale PowerQUICCTM III MPC8548 processor with an 800 MHz clock, was used to verify the computational time required for the models investigated in this paper.

The experimental tests were conducted under steady-state and transient conditions. The steady-state tests have been divided into three categories (Figure 2):

- a complete engine map (123 tests, indicated with the blue circles in Figure 2).
- EGR-sweep tests (162 tests, carried out on the points indicated with the red diamonds in Figure 2).
- sweep tests of the main injection timing (SOI_{main}) and injection pressure (p_f) (125 tests, carried out on the points indicated with the black circles in Figure 2).
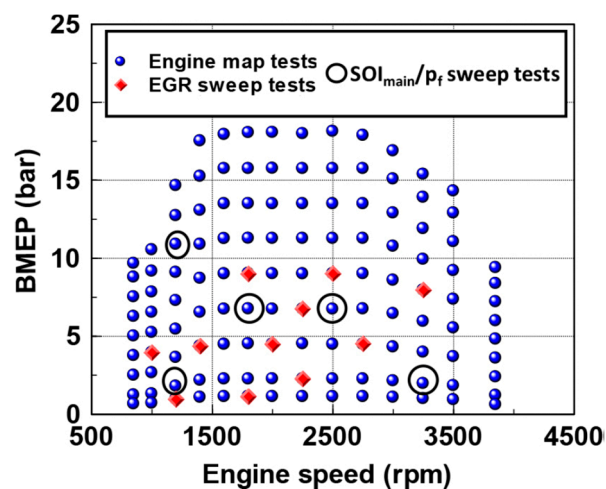

Figure 2. Experimental tests acquired at the test bench.

Validation of the methods was carried out over a WHTC. Details on these validations are given in Section 5.

3. Description of the Models

3.1. Physics-Based Model

Although the physics-based combustion model has already been presented in other studies, a summary of the approach and of the main equations is reported hereafter.

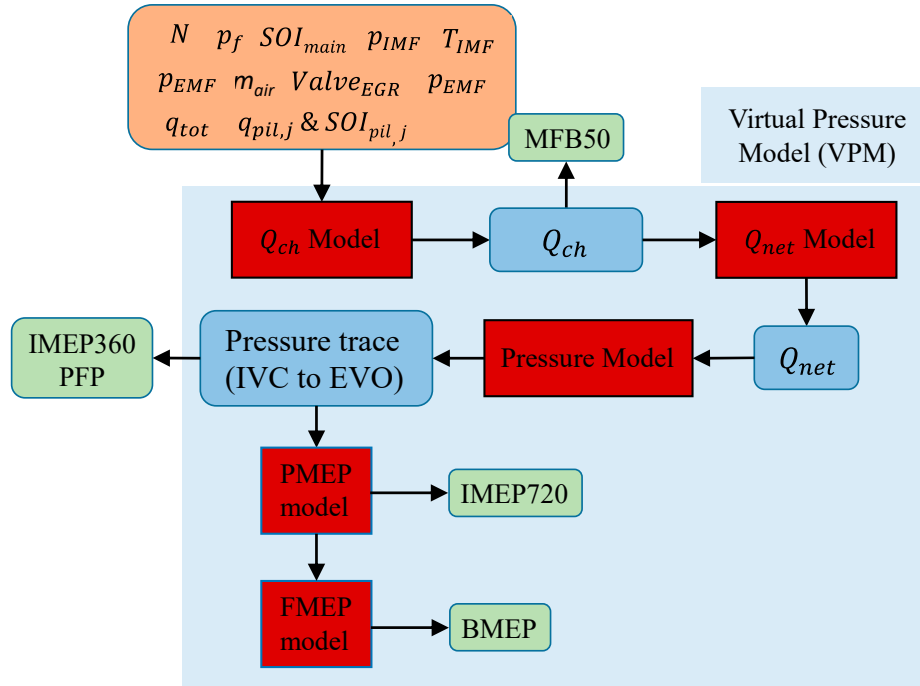A scheme of the model is shown in Figure 3.

Figure 3. Scheme of the physics-based combustion model [12].

N in Figure 3 indicates the engine speed, while p_{IMF} and T_{IMF} the intake manifold pressure and temperature, respectively, p_{EMF} the exhaust manifold pressure, SOI_{main} the start of injection of the main pulse, $SOI_{pil,j}$ the start of injection of the pilot pulse j, $q_{pil,j}$ the injected fuel quantity of pilot injection j, q_{tot} the total injected fuel quantity, $Valve_{EGR}$ the opening position of the high pressure EGR valve and m_{air} the fresh air trapped mass per cycle/cylinder.

The first step consists in the evaluation of the chemical heat release, according to the following equations (AFM approach):

$$\frac{dQ_{ch,pil,j}}{dt}(t) = \underline{K_{pil,j}}[Q_{fuel,pil,j}(t - \underline{\tau_{pil,j}}) - Q_{ch,pil,j}(t)] \tag{1}$$

$$\frac{dQ_{ch,main}}{dt}(t) = \underline{K_{1,main}}[Q_{fuel,main}(t - \underline{\tau_{main}}) - Q_{ch,main}(t)] + \underline{K_{2,main}}\frac{dQ_{fuel,main}(t - \underline{\tau_{main}})}{dt} \tag{2}$$

$$Q_{fuel,j}(t) = \int_{t_{SOI,j}}^{t} \dot{m}_{f,inj}(t)H_L dt \qquad t \leq t_{EOI,j} \tag{3}$$

$$Q_{fuel,j}(t) = \int_{t_{SOI,j}}^{t_{EOI,j}} \dot{m}_{f,inj}(t)H_L dt \qquad t > t_{EOI,j} \tag{4}$$

$$Q_{ch} = \sum_{j=1}^{n} Q_{ch,j} \tag{5}$$

where j indicates the generic injection pulse, H_L indicates the lower heating value of the fuel, and K and τ are the combustion rate coefficient and ignition delay coefficients, respectively.

Subsequently, the net energy of the charge (Q_{net}) is calculated as follows:

$$Q_{net,ht} \cong Q_{ch} \frac{m_{f,inj} H_L - \underline{Q_{ht,glob}}}{m_{f,inj} H_L} \tag{6}$$

$$Q_{net} \cong Q_{net,ht} - \underline{Q_{f,evap}} \tag{7}$$

where $Q_{f,evap}$ and $Q_{ht,glob}$ indicate the fuel evaporation heat from SOI to SOC and the heat exchanged by the charge with the walls over the combustion cycle; $m_{f,inj}$ is the total injected fuel mass per cyc/cyl. Q_{net} is then used to derive the in-cylinder pressure, which is based on the inversion of a single-zone thermodynamic model [39]:

$$p^i = \frac{\Delta Q_{net} - \frac{p^{i-1}}{2}(V^i - V^{i-1}) + \frac{1}{\gamma-1}p^{i-1}V^{i-1}}{\frac{V^i-V^{i-1}}{2} + \frac{V^i}{\gamma-1}} \quad (\gamma = 1.4, \ valid\ from\ SOC\ to\ EOC) \tag{8}$$

The compression and expansion phases are modeled with polytropic processes:

$$pV^{\underline{n}} = const \quad (valid\ from\ IVC\ to\ SOC) \tag{9}$$

$$pV^{\underline{n'}} = const \quad (valid\ from\ EOC\ to\ EVO) \tag{10}$$

while the pressure at IVC (i.e., the starting condition) is correlated with the pressure in the intake manifold:

$$p_{IVC} = p_{IMF} + \underline{\Delta p_{IMF}} \tag{11}$$

The main combustion metrics, that is, MFB50, IMEP360 (gross Indicated Mean Effective Pressure) and PFP (Peak Firing Pressure), can now be evaluated. Finally, friction (FMEP) and pumping (PMEP) models are used, and the net IMEP (IMEP720) and BMEP (Brake Mean Effective Pressure) are estimated:

$$IMEP360 = \frac{\int_{0}^{360} pdV}{V_0} \tag{12}$$

$$IMEP720 = IMEP360 - PMEP \tag{13}$$

$$BMEP = IMEP720 - FMEP \tag{14}$$

The tuning parameters of the physics-based model are underlined in Equations (1)–(14).

The Chen-Flynn approach [40] was adopted to estimate FMEP:

$$FMEP = x_1 + x_2 \cdot N + x_3 \cdot N^2 + x_4 \cdot PFP \tag{15}$$

where x_{1-4} are fitting parameters.

The intake oxygen concentration (O_2) is an operating variable that is closely correlated with the combustion process, and it is therefore used as an input variable in several sub-models. It is measured, on the experimental test bench, by means of a paramagnetic sensor, which is included in the test cell gas analyzer. However, if the developed combustion model is intended to be used for model-based

control, this sensor is not available onboard, and O_2 therefore needs to be estimated by means of a specific sub-model.

The following function has been used to estimate O_2 [11,41]:

$$O_2 = x_1 \frac{X_{r,EGR}}{RAF} + x_2 \tag{16}$$

where x_{1-2} are fitting parameters, RAF is the relative air-to-fuel ratio and $X_{r,EGR}$ is the EGR rate.

Therefore, it is necessary to estimate RAF and $X_{r,EGR}$ in order to evaluate the intake O_2 concentration. The relative air-to-fuel ratio is defined as follows:

$$RAF = \frac{m_{air}}{m_{f,inj} AF_{st}} \tag{17}$$

where m_{air} is the trapped air mass, $m_{f,inj}$ is the injected fuel mass and AF_{st} is the stoichiometric air-to-fuel ratio (taken as 14.4 in this work).

The accuracy of the estimation of RAF depends to a great extent on the accuracy of the estimation of m_{air}. If m_{air} is obtained from the engine air flow sensor, its accuracy is generally not very high (the error is typically of the order of 5%).

An alternative way of estimating RAF (the method that was adopted in this study) is to develop a semi-empirical correlation, such as:

$$RAF = x_1 \cdot p_{IMF}^{x_2} \cdot T_{IMF}^{x_3} \cdot m_{f,inj}^{x_4} \tag{18}$$

where x_{1-4} are fitted parameters. The correlation is fitted using a calibration dataset, in which the experimental values of RAF are obtained from steady-state measurements carried out using an accurate sensor (Bosch Lambda sensor LSU 4.9 (Bosch, Gerlingen, Germany)).

With reference to the EGR rate $X_{r,EGR}$, it is defined as follows:

$$X_{r,EGR} = \frac{m_{EGR}}{m_{EGR} + m_{air}} \tag{19}$$

where m_{EGR} is the trapped EGR mass.

The EGR rate can in general be derived experimentally on the basis of the measured emissions and of the measured intake CO_2 concentration. However, if the combustion model is intended to be applied for control-oriented applications, a real-time estimation of the EGR rate is needed, through the use of a dedicated submodel. To this aim, several approaches can be used (e.g., see [41]).

The simplest method to estimate the EGR rate (which was adopted in this paper) is to resort to look-up tables that depend on the engine speed and pedal position. These look-up tables were derived on the basis of the analysis of steady-state measurements.

As far as the calibration procedure is concerned, as reported in the previous sections, the physics-based model is characterized by several tuning parameters, which are mentioned in Table 2.

Table 2. Tuning parameters of the physics-based model.

Submodel	Calibration Parameter
Heat release model	$K_{pil,j}$; $K_{1,main}$; $K_{2,main}$; $\tau_{pil,j}$; τ_{main}
Net energy release model	$Q_{f,evap}$; $Q_{ht,glob}$
Pressure model	Δp_{IMF}; n; n'
BMEP model	FMEP, PMEP

The calibration of these parameters is carried out in two steps.

In the first step, the optimal values of all the parameters are identified, test by test, on the basis of the available experimental data. The process used for the identification of these values is reported in detail in [16], and is based on the best matching between the experimental and predicted heat release and in-cylinder pressure curves. At the end of this first step, a dataset of the optimal values of the model calibration parameters is available.

In the second step, on the basis of the dataset identified in step 1, the model calibration parameters are correlated with input variables, which are generally related to the engine operating conditions (e.g., intake manifold pressure/temperature, intake oxygen concentration, engine speed and load ...), using such mathematical methods as look-up tables, empirical or semi-empirical functions or machine learning tools. Semi-empirical correlations, based on power-law functions, are used in the baseline version of the physics-based model (see [12]), since they are capable of capturing the non-linear behavior of the correlations. However, in previous studies, the input parameters for these correlations were identified on the basis of a trial and error approach.

In the next sections, two aspects are investigated in order to improve the physics-based model with respect to previously reported versions:

1. A new procedure for the identification of the optimal set of input parameters is developed. This procedure is based on the joint use of the Pearson correlation analysis, partial correlation analysis, and sensitivity analysis, and it allows the performance of the baseline physics-based model to be improved with respect to the previously developed versions.
2. An alternative mathematical method (i.e., ANNs) is used to estimate the model calibration parameters:

$$K_{pil}, K_{1,main}, \ldots = \text{ANN}(x_0, x_1, \ldots, x_n) \qquad (20)$$

   where $x_0, x_1, \ldots, x_n$ are correlation variables. This approach has never been investigated before, and is denoted as "ANN physics-based model".

3.2. Direct Semi-Empirical Models

In addition to the baseline and ANN physics-based models, semi-empirical correlations were also developed in order to directly predict the MFB50, PFP, and BMEP metrics.

These correlations are based on power-law functions of the type:

$$\text{MFB50, PFP, BMEP} = A \cdot x_0^{a_0} \cdot x_1^{a_1} \cdot \ldots \cdot x_n^{a_n} \qquad (21)$$

The identification of the $x_0, x_1 \ldots x_n$ correlation variables and the fitting of the model are analyzed in the next sections. The semi-empirical models are trained on the basis of the same experimental data used for the physics-based model calibration.

3.3. Direct Artificial Neural Networks

The final method investigated to predict MFB50, PFP, and BMEP is based on feed-forward artificial neural networks with one inner layer:

$$\text{MFB50, PFP, BMEP} = \text{ANN}(x_0, x_1, \ldots, x_n) \qquad (22)$$

where $x_0, x_1, \ldots, x_n$ are the input variables of the neural networks.

The networks are trained on the basis of the same experimental data used for the physics-based model calibration. The methodology used for the calibration of the model, including the selection of the optimal number of neurons, is discussed in detail in Section 5.

4. Selection of the Model Input Variables

Selecting the most appropriate input parameters as independent variables for all the previously mentioned approaches is a key-aspect to obtain accurate and robust models.

In this paper, optimal sets of input parameters have to be defined for:

- The power-law correlations or the ANNs that evaluate the calibration parameters of the baseline physics-based model and of the ANN physics-based model, respectively.
- The direct semi-empirical model that estimates MFB50, PFP, BMEP.
- The direct ANN model that estimates MFB50, PFP, BMEP.

In general, the input parameters are related to the engine operating conditions, and can be either directly measured quantities or derived quantities. The directly measured parameters (such as engine speed N, the start of injection of the pilot/main pulses SOI_{pil}/SOI_{main}, the total injection quantity of the pilot pulses $q_{pil,tot}$, the injection quantity of the main pulse q_{main}, the total injection quantity q_{tot}, the intake/exhaust pressure and temperature $p_{IMF}, T_{IMF}/p_{EMF}, T_{EMF}$, the common rail pressure p_f, the trapped air mass m_{air} and the EGR valve opening position $Valve_{EGR}$) are the most commonly used input variables for control-oriented applications, since they are readily available for the ECU without the need of any additional sub-model. However, derived parameters (such as intake density ρ_{IMF}, in-cylinder gas density and temperature at the start of each injection pulse $\rho_{SOI,pil}, \rho_{SOI,main}, T_{SOI,pil}, T_{SOI,main}$, the start of combustion of the pilot and main pulses SOC_{pil}/SOC_{main}, in-cylinder gas density and temperature at the start of combustion for each injection pulse $\rho_{SOC,pil}, \rho_{SOC,main}, T_{SOC,pil}, T_{SOC,main}$, the pressure difference between the exhaust and intake manifold $\Delta p_{exh\text{-}int}$) or even outcomes of sub-models (such as PFP, MFB50) have also been used as input variables in models reported in the literature, since their correlation with the dependent variables is sometimes greater than that of the directly measured parameters. In this study, all these kinds of parameters were initially considered as candidate input variables to develop the models.

Table 3 reports the list of the directly measured input parameters (1st column), of the generated parameters (from the directly measured ones or from the combustion model) which can be used as input parameters (2nd column), and of the dependent output variables which have been considered in the present investigation. A total of 46 parameters have been included. It should be noted that some of the dependent variables are the calibration parameters of the physics-based model, which needs to be estimated for the model application (3rd column), while MFB50, PFP, and BMEP are the combustion metrics that are adopted for combustion control applications, i.e., the final output variables of all the four modeling approaches that have been investigated in this study (4th column).

The procedure that was identified to select the optimal set of input parameters is based on the joint use of the Pearson correlation and partial correlation coefficients (see [42–45] for details concerning the use of these coefficients), and is reported in detail in Appendix A for the paper readability purposes. However, a synthetic description is reported hereafter.

Basically, the procedure involves 3 steps:

1. First, the Pearson correlation coefficient is calculated for all the possible combinations of the 46 variables reported in Table 3 (the results are reported in Appendix A in Table A1). This coefficient measures the linear dependence between two variables, x_i and x_j, and is defined as follows:

$$\rho\left(x_i, x_j\right) = \frac{cov\left(x_i, x_j\right)}{\sqrt{var(x_i) \cdot var(x_j)}} \tag{23}$$

where $cov(x_i, x_j)$ is the covariance between x_i and x_j, while $var(x_i)$ and $var(x_j)$ are the variances of x_i and x_j, respectively. The Pearson correlation coefficient varies between -1 and $+1$. After the analysis of the Pearson correlation coefficients, an initial candidate set of input variables is identified for each output quantity that has to be estimated (see Table A2 in Appendix A).

2. In the second step, the partial correlation coefficient is evaluated between the dependent variables and each input variable of the initial candidate set, when removing the effect of the remaining input variables of the same set. It should be recalled that the partial correlation coefficient is able to measure the linear dependence between two variables, x_i and x_j, when the effect of the other input variables variable ($z_1, z_2, \ldots, z_i$) is removed:

$$\rho\left(x_i, x_j; z_1, z_2, \ldots, z_i\right) = \frac{\rho\left(x_i, x_j; z_1, z_2, \ldots, z_{i-1}\right) - \rho\left(x_j, z_i; z_1, z_2, \ldots, z_{i-1}\right) \cdot \rho\left(x_i, z_i; z_1, z_2, \ldots, z_{i-1}\right)}{\sqrt{\left[1 - \rho\left(x_j, z_i; z_1, z_2, \ldots, z_{i-1}\right)^2\right] \cdot \left[1 - \rho\left(x_i, z_i; z_1, z_2, \ldots, z_{i-1}\right)^2\right]}} \tag{24}$$

The partial correlation coefficient is more robust than the Pearson coefficient. However, the number of parameters should not be too high, otherwise, the consistency of the method decreases. Also, the partial correlation coefficient varies between −1 and +1. As reported in Appendix A (see Table A3), the simultaneous analysis of the Pearson and of the partial correlation coefficients has allowed the least and most robust correlation variables of the initial candidate set to be identified for each dependent output variable. Therefore, some variables were excluded from the initial candidate set identified in step 1 (since they showed low values of the Pearson and partial correlation coefficients), while other variables were kept since they showed high values of both the Pearson and partial correlation coefficients.

3. The remaining correlation variables of the initial candidate set which showed intermediate values of the Pearson and partial correlation coefficients were analyzed on the basis of a sensitivity analysis, in order to identify those that had to be kept and those that had to be excluded. A power law function, such as that reported in Equation (21), was used to model each output dependent parameter, and all the possible combinations of input variables were identified by evaluating, for each combination, the related model fitting precision, which was quantified by a correlation coefficient R_{adj}:

$$R_{adj} = \sqrt{1 - \frac{n_{\exp} - 1}{n_{\exp} - k}\left(1 - R^2\right)} \tag{25}$$

where n_{exp} is the total number of experimental data and k is the total number of free coefficients. The adjusted correlation coefficient R_{adj} takes into account the number of available experimental tests used for the model fitting, in relation to the number of free parameters of the model. In the end, the combinations which led to the best trade-off between the prediction accuracy and the number of input parameters were selected. Details on the sensitivity analysis have not been reported in this paper for the sake of brevity.

It should be noted that the proposed approach, i.e., the preliminary selection of the input variables through Pearson and partial correlation coefficients, and the subsequent sensitivity analysis carried out on a reduced set of variables, is more computationally efficient than a pure sensitivity analysis approach which includes the entire set of input variables.

In fact, the number of possible combinations that have to be investigated in the sensitivity analysis can be calculated through the following formula:

$$N_c = 2^j - 1 \tag{26}$$

where N_c is the number of possible combinations for the sensitivity analysis and j is the number of independent variables.

Table 3. List of the input and dependent parameters.

Input Parameters (Directly Measured Quantities)	Generated Quantities Which Can Be Used as Input Parameters	Dependent Output Variables (RAF + Calibration Parameters of the Physics-Based Model)	Dependent Output Variables (Combustion Metrics for Use for Model-Based Control)
N	X_{rEGR}	RAF	MFB50
p_f	O_2	τ_{pil}	PFP
SOI_{pil}	ρ_{IMF}	τ_{main}	BMEP
SOI_{main}	$\rho_{SOI,pil}$	K_{pil}	
$q_{tot,pil}$	$\rho_{SOI,main}$	$K_{1,main}$	
q_{main}	$\rho_{SOC,pil}$	$K_{2,main}$	
q_{tot}	$\rho_{SOC,main}$	$Q_{ht,glob}$	
p_{IMF}	$T_{SOI,pil}$	$Q_{f,evap}$	
T_{IMF}	$T_{SOI,main}$	Δp_{IMF}	
p_{EMF}	$T_{SOC,pil}$	n	
T_{EMF}	$T_{SOC,main}$	n'	
m_{air}	$\Delta p_{exh-int}$	PMEP	
$Valve_{EGR}$	SOC_{pil}	FMEP	
	SOC_{main}		
	IMEP360		
	IMEP720		
	Q_{netmax}		

According to Equation (26), N_c increases to a great extent as j increases. Therefore, the combined use of the Pearson correlation and partial correlation analysis is a very useful and efficient way of decreasing the computational effort of the sensitivity analysis, since it allows the poorly correlated variables to be excluded a-priori, or the highly correlated variables to be selected, so that the sensitivity analysis (which is highly time consuming, since it requires model fitting of each possible combination), is only carried out for a reduced set of variables.

5. Results and Discussion

The improvement in the prediction performance of the physics-based model (compared to the previous version reported in [12]), when adopting the new methodology for the input parameter selection described in Section 4, is reported first. Subsequently, the performances of the four investigated models described in Section 3 are compared, in terms of accuracy at steady-state conditions and over WHTC, and in terms of required computational time. The main advantages and drawbacks of each kind of approach are also discussed.

5.1. Improvement to the Baseline Physics-Based Combustion Model

The performance of the baseline physics-based combustion model has been improved, with respect to the previous version reported in other papers [12]. The set of input variables for the model correlations has been revised, on the basis of the Pearson correlation and partial correlation analysis, according to the procedure described in the previous section.

The final list of selected input variables for each dependent variable and the corresponding fitting precision are shown in Table 4. It should be noted that some correlations are improved with respect to pure power-law functions.

Table 4. Final list of the selected input variables, the related correlations, for each tuning parameter of the physics-based model and the corresponding fitting precision in terms of correlation coefficient (R^2), root mean squared error (RMSE) and relative RMSE (RMSE_{rel}).

Dependent Variable	Final List of the Independent Variables	Fitted Correlation	R^2, RMSE, RMSErel
RAF	N, q_{tot}, P_{IMF}, $Valve_{EGR}$	$RAF = 43.172 \times N^{-0.0349} \times qtot^{-0.9530} \times P_{IMF}^{0.8989} \times Valve_{EGR}^{-0.2163}$	0.98, 0.135, 5.8%
τ_{pil}	N, SOI_{pil}, $\rho_{SOI,pil}$, $T_{SOI,pil}$	Power-law function : $\tau_{pil} = 8.177 \times 10^6 \times N^{0.5816} \times SOI_{pil}^{-2.828}$ $\times \rho_{SOI,pil}^{-1.043} \times T_{SOI,pil}^{0.1445}$ Improved correlation : $\tau_{pil} = 0.8688 \times N^{0.4900}$ $\times \left(365 - SOI_{pil}\right)^{0.2647} \times \rho_{SOI,pil}^{-0.9870} \times T_{SOI,pil}^{0.0468}$	0.906, 1.06, 26.9% 0.908, 1.04, 25.6%
τ_{main}	N, p_f, q_{tot}, $\rho_{SOI,pil}$, $\rho_{SOI,main}$	$\tau_{main} = 0.1862 \times N^{2.2085} \times p_f^{-0.7866} \times q_{tot}^{0.2847} \times \rho_{SOI,pil}^{-4.6496} \times \rho_{SOI,main}^{1.5852}$	0.886, 0.339, 47.9%
K_{pil}	N, p_f, SOI_{pil}, $q_{tot,pil}$, $\rho_{SOI,pil}$, $T_{SOI,pil}$	$K_{pil} = 1.2273 \times 10^6 \times N^{-1.0983}$ $\times p_f^{0.4252} \times SOI_{pil}^{3.5487} \times q_{tot,pil}^{-0.4497} \times \rho_{SOI,pil}^{-0.6886} \times T_{SOI,pil}^{-4.2469}$	0.365, 0.0894, 40.4%
$K_{1,main}$	N, SOI_{main}, q_{tot}, p_{IMF}, T_{IMF}, $\rho_{SOI,main}$	$K_{1,main} = 1.0106 \times 10^{-13} \times N^{-0.1030} \times SOI_{main}^{6.7329} \times q_{tot}^{-0.3414} \times p_{IMF}^{1.0216}$ $\times T_{IMF}^{-1.7314} \times \rho_{SOI,main}^{-0.2703}$	0.833, 0.0026, 5.4%
$K_{2,main}$	N, p_f, q_{main}, $\rho_{SOI,main}$, $T_{SOI,main}$	Power $-$ law function : $K_{2,main} = 85.0613 \times N^{-0.2778} \times p_f^{0.3397} \times q_{main}^{0.3948}$ $\times \rho_{SOI,main}^{-1.2203} \times T_{SOI,main}^{-0.5323}$ Improved correlation : $K_{2,main} = 197.5434 \times N^{-3.2244 \times 10^{-4}}$ $\times p_f^{3.6373 \times 10^{-4}} \times q_{main}^{4.793 \times 10^{-4}} \times \rho_{SOI,main}^{-0.0015} \times T_{SOI,main}^{-7.6126 \times 10^{-4}} - 1957.1$	0.634, 0.0311, 61.5% 0.663, 0.0299, 55.2%
$Q_{f,evap}$	N, p_f, SOI_{pil}, q_{tot}, m_{air}, $\rho_{SOI,pil}$	$Q_{f,evap} = 2.016 \times 10^4 \times N^{0.4478} \times p_f^{-0.2891} \times SOI_{pil}^{-3.499} \times q_{tot}^{-0.1624}$ $\times m_{air}^{-0.5281} \times \rho_{SOI,pil}^{0.0807}$	0.532, 0.00145, 15.5%
n	N, SOI_{pil}, p_{IMF}, T_{IMF}, $Valve_{EGR}$, $\rho_{SOI,pil}$, $T_{SOC,main}$	$n = 2.89 \times 10^5 \times N^{-0.013} \times SOI_{pil}^{-2.103} \times p_{IMF}^{-0.154} \times T_{IMF}^{0.347} \times Valve_{EGR}^{0.0042}$ $\times \rho_{SOI,pil}^{0.168} \times T_{SOC,main}^{-0.340}$	0.654, 0.0053, 0.39%
n'	N, p_f, q_{tot}, p_{IMF}, $Valve_{EGR}$	Power $-$ law function : $n' = 2.2173 \times N^{-0.0647} \times p_f^{0.0218} \times q_{tot}^{-0.0613}$ $\times p_{IMF}^{0.0783} \times Valve_{EGR}^{-0.0025}$ Improved correlation : $n' = 5.6312 \times 10^{-4} \times N^{0.6839} \times p_f^{-0.2032}$ $\times q_{tot}^{0.6574} \times p_{IMF}^{-1.0419} \times Valve_{EGR}^{0.0418} + 1.4812$	0.913, 0.01, 0.8% 0.923, 0.00936, 0.709%
$Q_{ht,glob}$	N, p_f, q_{tot}, $\rho_{SOI,pil}$	$Q_{ht,glob} = 0.4596 \times N^{-0.3909} \times p_f^{-0.1981} \times q_{tot}^{1.4489} \times \rho_{SOI,pil}^{-0.5557}$	0.99, 0.0143, 11.5%
Δp_{IMF}	N, q_{tot}, p_{IMF}	Power $-$ law function : $\Delta p_{IMF} = 0.0929 \times N^{0.032} \times q_{tot}^{-0.0319} \times p_{IMF}^{1.15}$ Improved correlation : $\Delta p_{IMF} = 0.12254 \times \left[1 - exp\left(\frac{-N+538.806}{131.94}\right)\right]$ $\times q_{tot}^{-0.044} \times p_{IMF}^{1.1932}$	0.966, 0.00733, 4.6% 0.967, 0.00723, 4.37%
$PMEP$	N, p_{IMF}, $Valve_{EGR}$, $\Delta p_{exh-int}$	$PMEP = 9.5393 \times 10^{-4} \times N^{0.8798} \times p_{IMF}^{0.3571} \times Valve_{EGR}^{0.0010} \times \Delta p_{exh-int}^{0.5164}$	0.994, 0.0204, 14.4%
$FMEP$	N, SOI_{pil}, q_{tot}, p_{IMF}	$FMEP = 2.4730 \times 10^{-5} \times N^{0.5640} \times SOI_{pil}^{1.1170} \times q_{tot}^{-0.0502} \times p_{IMF}^{0.6530}$	0.939, 0.0978, 7.91%

Figures 4 and 5 report the accuracy of the original physics-based model with the baseline calibration (presented in [12]) and the revised model which embeds the improved correlations, concerning the estimation of BMEP, PFP, and MFB50, for the steady-state tests reported in Figure 2.
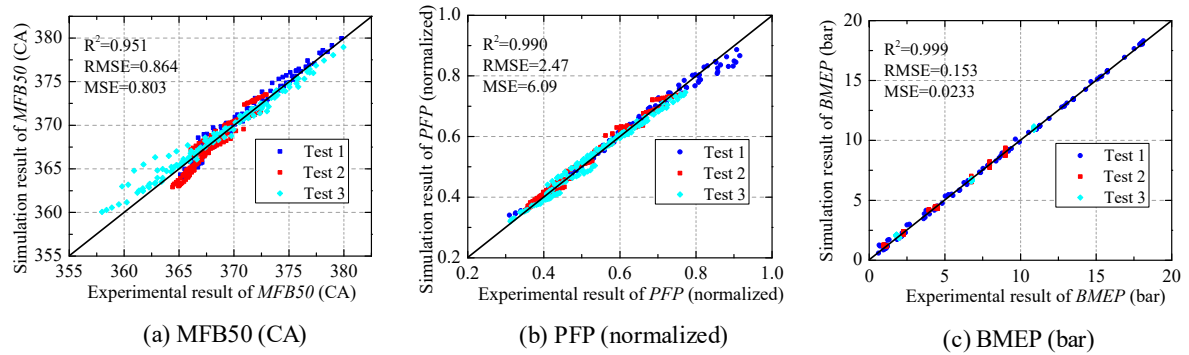

Figure 4. Predicted vs. experimental values of MFB50, PFP and BMEP for the original physics-based model [12], considering the steady-state conditions.
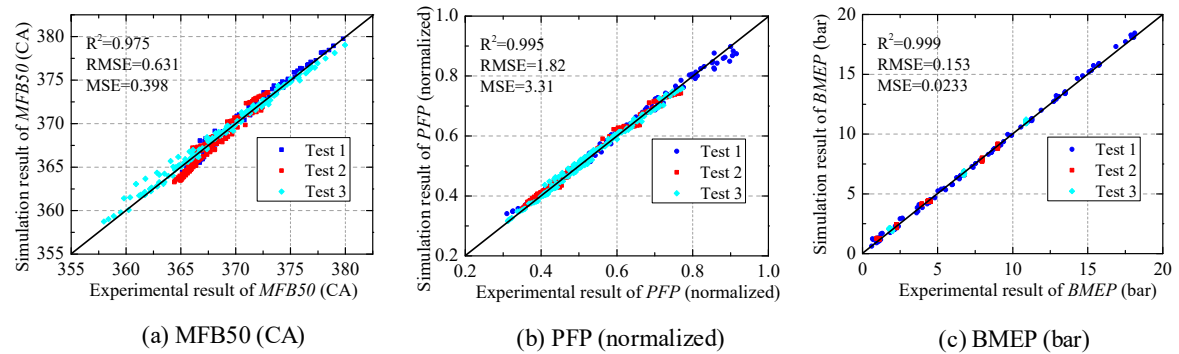

Figure 5. Predicted vs. experimental values of MFB50, PFP and BMEP for the improved physics-based model, considering the steady-state conditions.

It can be seen that, although the accuracy in the estimation of BMEP is somewhat similar, an improvement in the estimation of PFP, and MFB50 can be obtained, since the related RMSE values decrease from 2.47 bar to 1.82 bar and from 0.86 deg to 0.63 deg, respectively.

5.2. Direct Semi-Empirical Models of MFB50, PFP, and BMEP

The procedure used to identify the input parameters of the physics-based model, which was described in Section 4, was also adopted to select the input variables of the direct semi-empirical models. The correlations identified for MFB50, PFP, and BMEP are reported in Table 5.

Table 5. Final list of the selected independent variables, and the related correlations, for MFB50, PFP, and BMEP and the corresponding fitting precision.

Dependent Variable	Final List of the Independent Variables	Fitted Correlation	R^2, RMSE, RMSErel
$MFB50$	$N, p_f, SOI_{main}, q_{tot}, p_{IMF}, T_{IMF}, Valve_{EGR}$	$MFB50 = 0.5125 \times N^{0.0335} \times p_f^{-0.0241} \times SOI_{main}^{1.0036} \times q_{tot}^{0.0102} \times p_{IMF}^{0.0230} \times T_{IMF}^{0.0965} \times Valve_{EGR}^{-0.0038}$	0.974, 0.633, 0.17%
PFP	$N, p_f, SOI_{main}, q_{tot}, p_{IMF}, T_{IMF}, \rho_{SOI,main}$	$PFP = 1.125 \times 10^6 \times N^{-0.386} \times p_f^{0.275} \times (365 - SOI_{main})^{0.147} \times q_{tot}^{0.1625} \times p_{IMF}^{1.116} \times T_{IMF}^{-1.341} \times \rho_{SOI,main}^{-0.610}$	0.983, 3.03, 3.63%
$BMEP$	$N, p_f, SOI_{main}, q_{tot}, p_{IMF}, T_{IMF}, Valve_{EGR}, \rho_{SOI,main}$	$BMEP = 5.7727 \times 10^6 \times N^{-0.2182} \times p_f^{0.2620} \times SOI_{main}^{-3.2588} \times q_{tot}^{1.1765} \times p_{IMF}^{-0.5545} \times T_{IMF}^{0.0131} \times Valve_{EGR}^{0.0304} \times \rho_{SOI,main}^{0.3577}$	0.996, 0.257, 14.9%

5.3. ANN-Based Models

Artificial neural networks have a great capacity to capture complex nonlinear relationships between variables through relatively simple mathematical operations, and require very little computational effort. Therefore, they are good candidates for control-oriented applications. In this section, ANNs have been used to:

- Replace the power law-based correlations of the physics-based models shown in Table 4. The resulting new physics-based model is referred to as "ANN physics-based model".
- Directly simulate the MFB50, PFP, and BMEP combustion metrics, analogously to the semi-empirical models shown in Table 5.

In both cases, the sets of input variables that were used for the ANNs are the same as those shown in Table 4 (physics-based model) and Table 5 (semi-empirical models). The experimental data used for ANN training is constituted by the 410 tests shown in Figure 2.

In general, the implementation of ANNs involves three steps: the selection of the input variables, the training phase, and the testing phase. The available experimental dataset is usually split randomly for training, testing and validation purposes. The main parameters that have to be selected are the net type, the number of inner layers and the number of neurons for each layer, the activation function, and the training algorithm.

As far as the network type is concerned, feed-forward networks are typically used in the literature, and have also been used in this paper. These types of networks contain an input layer, some hidden layers, and an output layer. The back-propagation training algorithm is generally used as a training algorithm for feed-forward neural networks [46]. The Bayesian regularization training function 'trainbr' is one of the best-known BP training algorithms; it can lead to very precise ANNs [46,47] and is able to provide good generalization for difficult, small or noisy datasets [48], even though its number of epochs is usually higher. In this work, the training dataset is not large and the network structure is quite simple, therefore a relatively large epoch number should not lead to an unacceptably long computation time. For these reasons, 'trainbr' was selected as the training algorithm.

In order to speed up the learning process for a high number of networks, the inputs are usually normalized before sending them to the input layer. In this work, the 'mapminmax' process function [49] has been used to normalize the inputs, so that all the values fall into the [−1, 1] interval. Similarly, network outputs can also be associated with processing functions [49].

The activation function that was used in this work is the 'tansig' function. It was found, with reference to the number of hidden layers, that one layer was enough to simulate the parameters of the physics-based model and the combustion metrics (MFB50, PFP, BMEP). The use of one hidden layer for feed-forward neural networks is also quite common in the literature [47,50,51], especially when the number of training data is not very high.

The MSE minimization criterion was used for the training process, as suggested in [52]. The ANN performance is evaluated by comparing the network outputs with the experimental data through regression analysis [53]. The values of the mean-squared-error MSE and the determination coefficient R^2 are used to measure the performance of ANNs. Details of the ANN parameters developed on the MATLAB (R2016b, MathWorks, Natick, MA, USA) platform are shown in Table 6.

The optimal number of neurons in the hidden layer depends on the specific application, and it cannot, therefore, be identified a-priori. The ANN predictive performance is influenced to a great extent by the number of neurons. According to the literature, a too large number of neurons may lead to overfitting, while a too small number of neurons may make the ANN too simple to be able to capture the characteristics of a complex system (that is, underfitting [37]). Therefore, a trade-off must be made between the number of hidden layer neurons and the predictive performance of the ANN, and a sensitivity analysis is always needed in order to identify what the optimal number of neurons is for a specific application. In this section, a suitable number of neurons is selected for each of the dependent parameters that have to be estimated (i.e., those reported in Tables 4 and 5).

Table 6. Details of the artificial neural network (ANN) parameters developed on the MATLAB platform.

Parameter	Selected Option
Network type	Feed-forward back-propagation
Hidden layer	Single layer
Process function	mapminmax
Training algorithm	trainbr
Transfer function	tansig
Loss function	MSE
Performance	MSE, R^2

All the available steady-state experimental data (410 tests) were used to generate the ANNs. Overall, 80% of the data was used as training data, and the other 20% was used to test the trained ANNs. Considering the smallness of the training dataset, the training process features some uncertainties, because the initial value of the weights of the neural networks is generated randomly. Thus, when the training process of a given neural network is repeated, its performance may vary to a great extent, even when the same input training dataset is used. Moreover, the experimental dataset for each training process is randomly divided into training and testing datasets with a certain ratio. This can also influence the performance of the trained neural networks when the training process is repeated. However, it has been shown in the literature that these uncertainties decrease considerably, and may even be ignored, when the size of the training dataset is sufficiently large [54], but this was not the case in this study.

Therefore, 200 training and testing repetitions were carried out for each ANN, with random initial weights and random splitting of the experimental dataset into training and testing datasets. The determination coefficient R^2 and the mean square error MSE were used to measure the performance of the ANNs. The mean values of R^2 and MSE for 200 trials were calculated for the training and testing of each ANN.

The trade-off between the number of hidden layer neurons and the performance of the ANN was explored for each ANN that simulated the parameters of the physics-based model (in the "ANN physics-based model") and for each ANN that simulated the MFB50, PFP, and BMEP combustion metrics (in the "direct ANN model"). The results of the sensitivity analysis are shown in Appendix B.

The sensitivity analysis allowed the optimal number of hidden layer neurons to be identified for each parameter. The most suitable number of neurons for each parameter is listed in Table 7, along with the mean values of R^2 and MSE for the training and testing phases. In general, it can be seen that the most suitable neuron number is no higher than 7 in all cases.

Once the sensitivity analysis had been carried out, the best ANN out of the 200 training trials was selected on the basis of the criteria reported in Appendix B for each parameter.

After this stage, the following investigations were conducted

1. With reference to the physics-based model, the power law-based correlations of the calibration parameters were replaced by the corresponding ANNs, and the performance of the resulting model (i.e., the "ANN physics-based model") was evaluated. Table 8 shows a comparison between the accuracy of the ANN-based correlations and of the power-law based ones in the estimation of the parameters of the physics-based model. It can be observed that all the precisions based on ANNs are higher than those based on the power law correlations. ANNs have in fact been proved to have a powerful ability to catch the non-linear characteristics between input and output parameters

2. The performance of the ANNs that were used to directly simulate MFB50, PFP, and BMEP was evaluated

A detailed comparison of the performance of the four developed approaches, for the estimation of the MFB50, PFP, and BMEP combustion metrics, is reported in the next section.

Table 7. The most suitable number of neurons and the corresponding mean values of R^2 and MSE (over 200 training trials) for the ANNs that simulated the parameters of the physics-based model (a) and the combustion metrics (b).

(a) Physics-Based Model Parameters												
Dependent Parameter	τ_{pil}	τ_{main}	K_{pil}	$K_{1,main}$	$K_{2,main}$	$Q_{ht,glob}$	$Q_{f,evap}$	Δp_{IMF}	n	n'	**PMEP**	**FMEP**
Suitable number of neurons	6	6	6	6	6	4	5	7	5	5	5	4
Mean value of the training and testing R^2	0.977, 0.978	0.981, 0.983	0.889, 0.794	0.975, 0.968	0.939, 0.945	0.998, 0.998	0.914, 0.899	0.992, 0.995	0.977, 0.975	0.989, 0.989	1.00, 1.00	0.988, 0.986
Mean value of the training and testing MSE	0.53, 0.57	0.038, 0.063	0.0026, 0.0034	2×10^{-6}, 2.6×10^{-6}	3.15×10^{-4}, 2.95×10^{-4}	8.1×10^{-5}, 9.6×10^{-5}	7.2×10^{-7}, 9.0×10^{-7}	2.4×10^{-4}, 2.2×10^{-4}	3.6×10^{-6}, 4.3×10^{-6}	2.6×10^{-5}, 2.6×10^{-5}	5.7×10^{-5}, 6.96×10^{-5}	3.9×10^{-3}, 3.9×10^{-3}

(b) Combustion Metrics			
Dependent Parameter	**MFB50**	**PFP**	**BMEP**
Suitable number of neurons	5	6	5
Mean value of the training and testing R^2	0.999, 0.999	1.00, 1.00	1.00, 1.00
Mean value of the training and testing MSE	0.030, 0.043	0.27, 0.40	0.0007, 0.0011

Table 8. Comparison of the correlation coefficient of the ANNs and of the power-law empirical functions that estimate the tuning parameters of the physics-based model.

Dependent Parameter	τ_{pil}	τ_{main}	K_{pil}	$K_{1,main}$	$K_{2,main}$	$Q_{f,evap}$	$Q_{ht,glob}$	n	n'	Δp_{IMF}	$PMEP$	$FMEP$
Most suitable number of neurons	6	6	6	6	6	5	4	5	5	7	5	4
The training and testing R^2 for ANNs	0.977, 0.975	0.977, 0.974	0.878, 0.872	0.981, 0.973	0.941, 0.910	0.919, 0.917	0.998, 0.998	0.978, 0.974	0.989, 0.986	0.992, 0.991	1.00, 0.999	0.975, 0.978
R^2 for the power-law based functions	0.908	0.886	0.365	0.833	0.663	0.532	0.99	0.654	0.923	0.967	0.994	0.939

5.4. Comparison of the Four Different Models under Steady-State and Transient Conditions

In this section, the performance of the four different models has been tested under steady-state and transient operating conditions.

First, the performance of the developed models was compared under steady-state conditions. The results are shown in Figure 6.
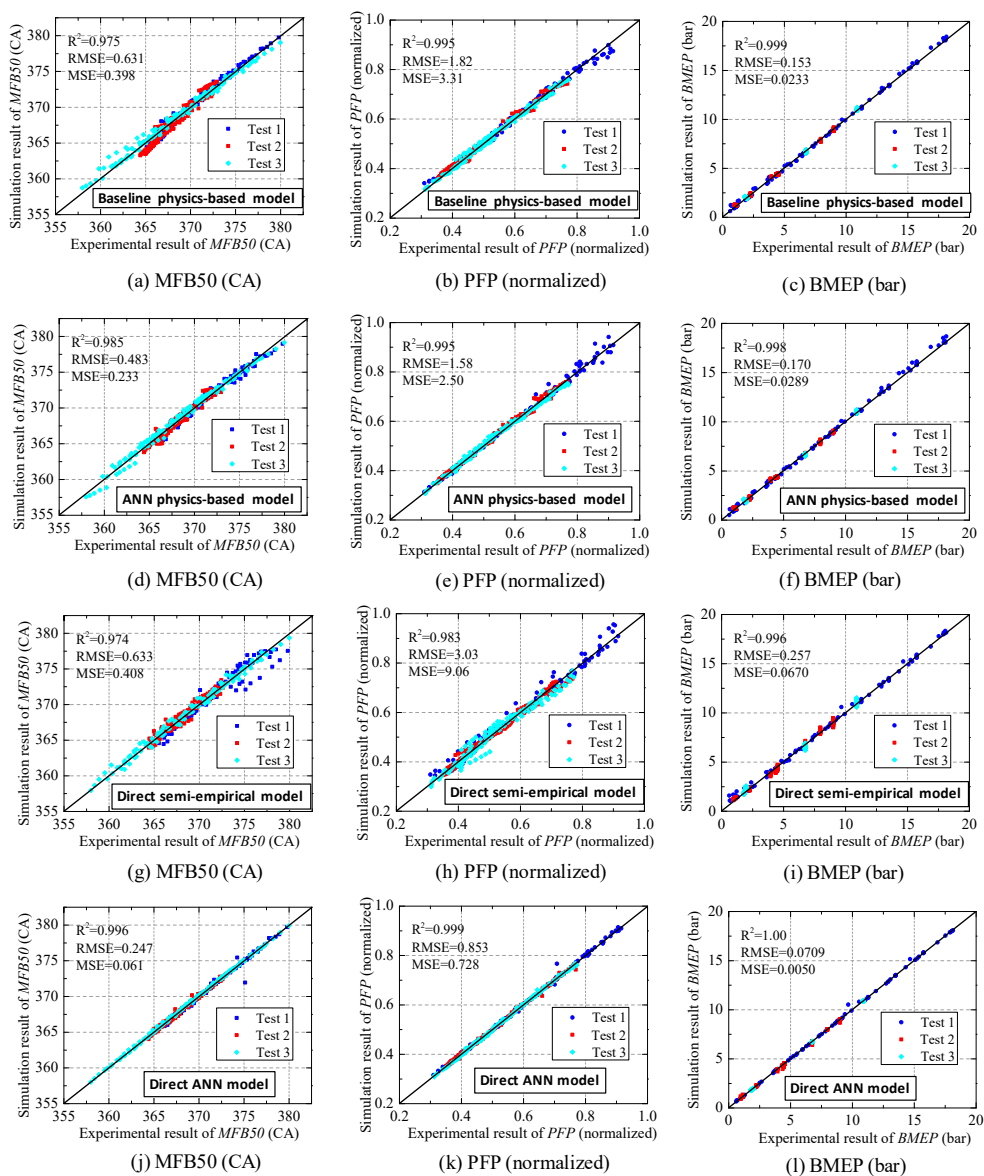

Figure 6. (**a–l**) Comparison of the performance of the developed models under steady-state operating conditions.

It can be seen, from the previous charts, that the accuracy of the ANN physics-based model is higher than that of the baseline physics-based model concerning MFB50 and PFP. This indicates that the ANNs are more powerful than the power-function-based empirical functions in catching the nonlinear characteristics between the input and dependent parameters of the model for these two metrics under steady-state conditions. Overall, the accuracy of the direct ANN model is the best, while the accuracy of the semi-empirical model is the lowest. However, the absolute values of the error of the semi-empirical model are still acceptable, considering that this kind of model ignores the detailed simulation of the in-cylinder physical process and requires a limited number of free parameters to be tuned (as a consequence, it also requires a limited number of experimental calibration tests). This suggests that the semi-empirical model, based on power functions, is also suitable for control-oriented combustion control applications, since its computational effort is much less than that of the physics-based model.

The performance of the developed models was then evaluated and compared under transient operating conditions over WHTC. The results are shown in Figures 7–9.

Table 9 reports a summary of the accuracy of the models over WHTC and under steady-state operating conditions.
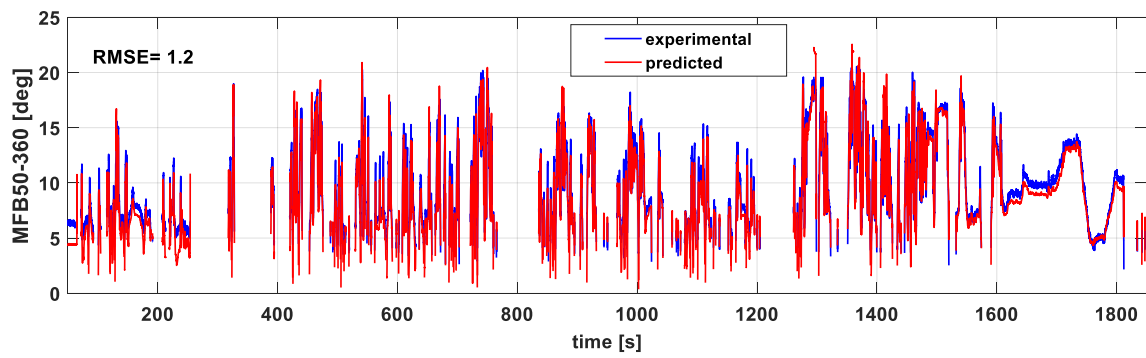
Table 9. Comparison of the performance of the four different models under steady-state and transient operating conditions.

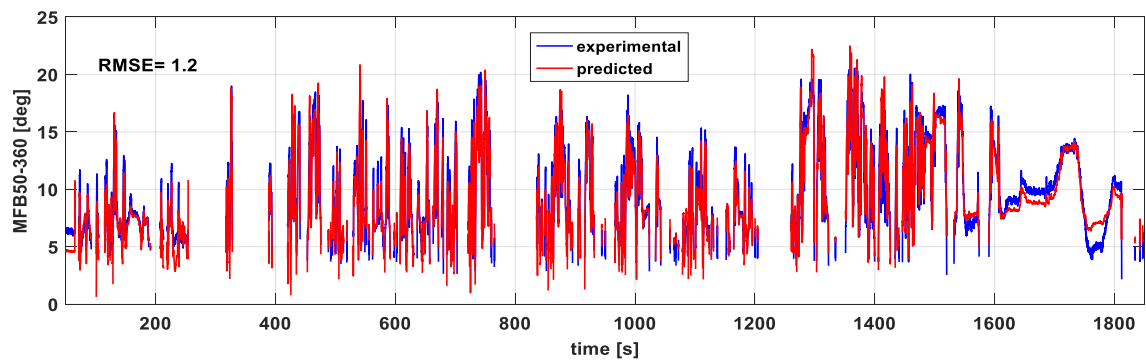| Model Type	Steady-State Condition (R^2 and RMSE)			WHTC Transient Condition (RMSE)		
	MFB50	PFP	BMEP	MFB50	PFP	BMEP
Baseline physics-based model	0.975 0.63 deg	0.995 1.8 bar	0.998 0.15 bar	1.2 deg	10.5 bar	0.7 bar
ANN physics-based model	0.984 0.48 deg	0.995 1.6 bar	0.998 0.17 bar	1.2 deg	13.8 bar	0.8 bar
Direct semi-empirical model	0.974 0.63 deg	0.983 3.0 bar	0.996 0.26 bar	1.4 deg	11.7 bar	0.8 bar
Direct ANN model	0.996 0.25 deg	0.999 0.85 bar	1.0 0.071 bar	1.1 deg	9.6 bar	0.7 bar

It can be seen from the graphs of Figures 6–9 and from Table 9 that the accuracy of the direct ANN model is the best among all the different approaches, not only for steady-state operation (RMSE equal to 0.25 deg, 0.85 bar and 0.071 bar for MFB50, PFP, and BMEP, respectively), but also for transient operation over WHTC (RMSE of 1.1 deg, 9.6 bar and 0.7 bar for MFB50, PFP, and BMEP, respectively).

The accuracy of the baseline physics-based model under transient operating conditions is slightly lower than that of the direct ANN model in terms of RMSE (1.2 deg vs. 1.1 deg for MFB50, 10.5 bar vs. 9.6 bar for PFP, 0.7 bar vs. 0.7 bar for BMEP), but is higher than that of the ANN physics-based model (1.2 deg vs. 1.2 deg for MFB50, 10.5 bar vs. 13.8 bar for PFP, 0.7 bar vs. 0.8 bar for BMEP) and of the semi-empirical model (1.2 deg vs. 1.4 deg for MFB50, 10.5 bar vs. 11.7 bar for PFP, 0.7 bar vs. 0.8 bar for BMEP).
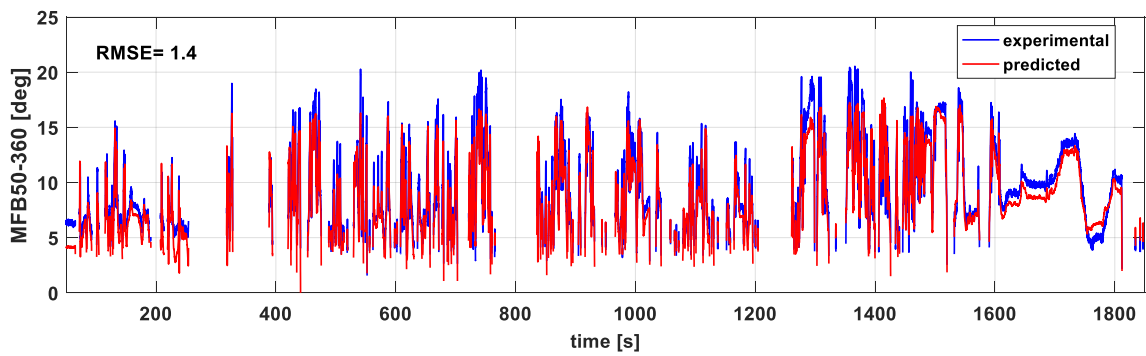
A remarkable result is the deterioration of the accuracy of the ANN physics-based model (compared to the baseline physics-based model) which occurs for transient operation, while this approach leads to more accurate results for steady-state conditions. Therefore, this approach lacks robustness, compared to the baseline physics-based model.
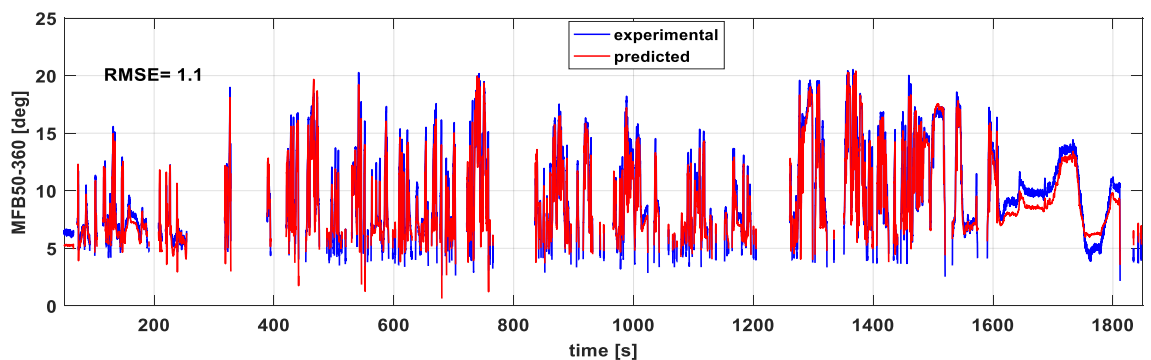
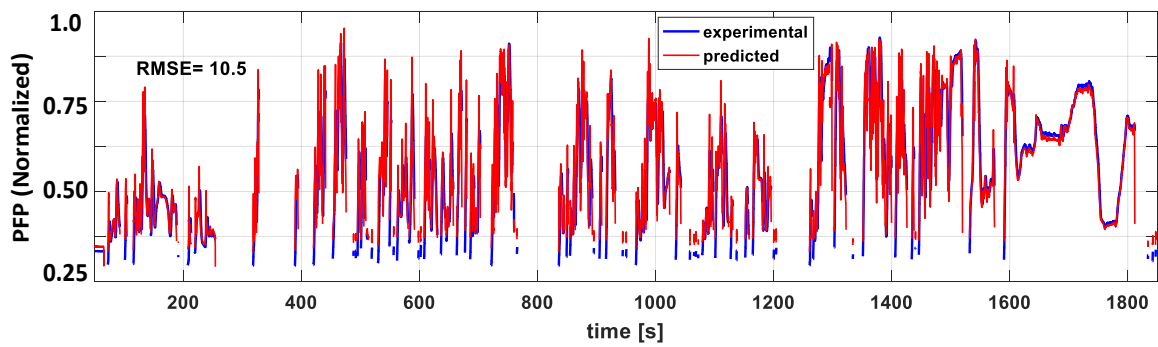(a) MFB50: baseline physics-based model

(b) MFB50: ANN physics-based model

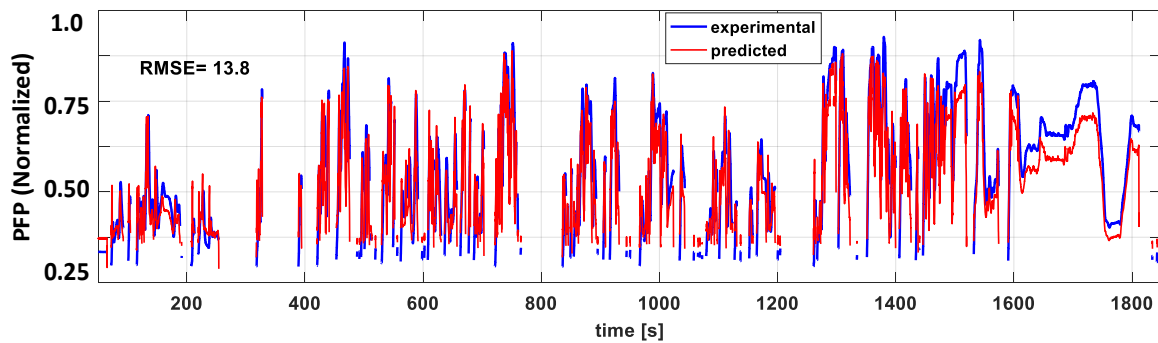

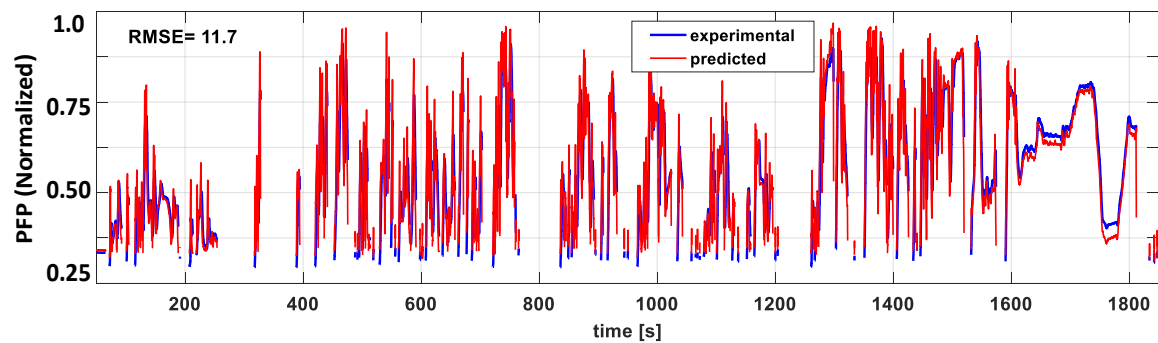(c) MFB50: direct semi-empirical model

(d) MFB50: direct ANN model

Figure 7. (**a–d**) Comparison of the performance of the developed models over WHTC: MFB50.
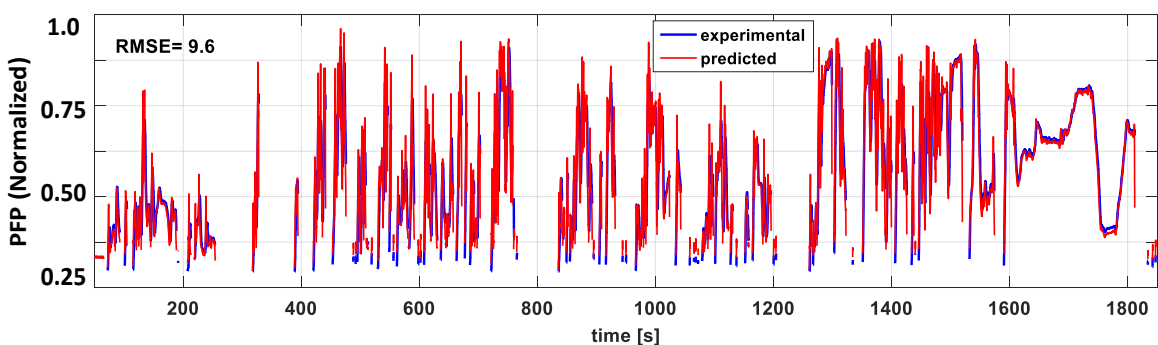
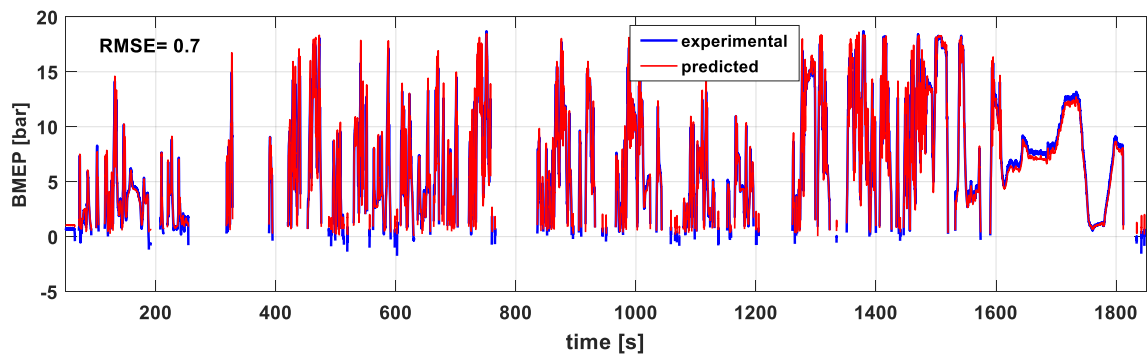(a) PFP: baseline physics-based model

(b) PFP: ANN physics-based model

(c) PFP: direct semi-empirical model

(d) PFP: direct ANN model

Figure 8. (**a**–**d**) Comparison of the performance of the developed models over WHTC: PFP.

(a) BMEP: baseline physics-based model

(b) BMEP: ANN physics-based model
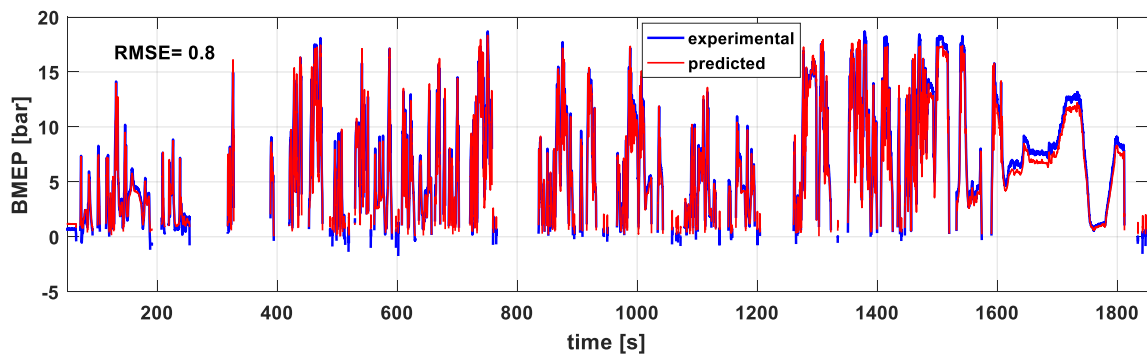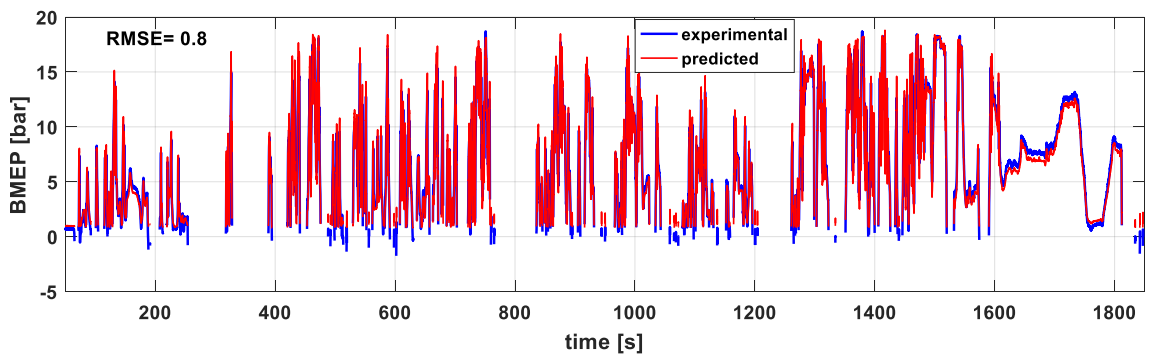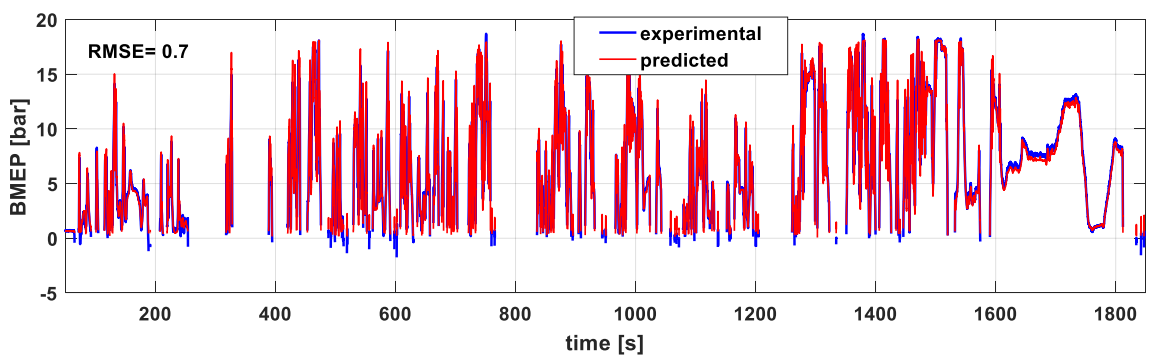

(c) BMEP: direct semi-empirical model

(d) BMEP: direct ANN model

Figure 9. (**a**–**d**) Comparison of the performance of the developed models over WHTC: BMEP.

Finally, the semi-empirical model shows a lower level of accuracy than the baseline physics-based model and the direct ANN model. However, the accuracy deterioration can still be considered acceptable, considering the simple structure of this model and the low calibration effort that is required, as discussed above.

Overall, the direct ANN model can be considered as the best candidate for control-oriented applications, since:

- it is much simpler in structure and theory and easier to build than the physics-based model;
- it does not require any detailed knowledge or modeling of the physical and chemical processes;
- it is robust not only under steady-state conditions, but also for transient operation;
- its accuracy is high even when a not so large dataset of experimental tests (410) is used for training;
- it features the best accuracy under steady-state and transient operating conditions;
- the required computational time is less than that of the physics-based model (see the next section).

The main drawback of ANNs is that they are a black-box type models, so they do not provide detailed information about the physical and chemical processes which occur inside the combustion chamber (e.g., heat release trend, in-cylinder pressure, temperatures . . .).

5.5. Analysis of the Computational Time

The computational time of all the models (excluding the ANN physics-based model, which was found to be the least robust approach) was estimated when they were run on the ETAS ES910 device. The results are reported in Table 10.

Table 10. Comparison of the computational time for the different models.

Model	Calculated Quantity	Average Computational Time, per Iteration, on the ETAS ES910 Device. The Reported Computational Time Values are Cumulative.
Baseline physics-based model	MFB50	≈200 μs
	PFP	≈350 μs
	BMEP	≈350 μs
Direct semi-empirical model	MFB50	<50 μs
	PFP	<50 μs
	BMEP	<50 μs
Direct ANN model	MFB50	<50 μs
	PFP	<50 μs
	BMEP	<50 μs

It can be seen that the direct ANN direct model and the semi-empirical model feature the lowest computational times, so that they can be considered the most suitable approaches to be implemented on an ECU for onboard control. The physics-based model requires the highest computational time. However, it is also compatible with real-time combustion control applications, since it is much lower than the typical engine cycle time. On the basis of the results reported in this study, a model-based combustion controller will be developed in the near future, and will be tested on the engine through the ETAS ES910 rapid prototyping device.

6. Conclusions

A comparison of the performance of physics-based, semi-empirical and artificial neural network (ANN)-based models has been carried out in this paper to estimate the main combustion metrics for an FPT Euro VI 3.0 L F1C diesel engine. The models were developed for combustion control-oriented

applications. The combustion metrics that are predicted by the models are MFB50, PFP, and BMEP. These metrics are interesting for control-oriented applications, since they are closely related to the thermal efficiency and to the pollutant formation process of the engine. Four different types of modeling approaches have been investigated. The first approach is based on a previously developed low-throughput mean-value physics-based model, which is capable of simulating the heat release rate, the in-cylinder pressure, and the related metrics on the basis of the accumulated fuel mass principle. In this approach, the tuning parameters of the model are estimated on the basis of physically-consistent correlations, which are modeled using power-law functions. The second investigated approach is based on the joint use of artificial neural networks (ANNs) and physical modeling: instead of using the power-law functions, ANNs have been used to identify the parameters of the physics-based model. The third investigated approach is based on a direct estimation of MFB50, PFP, and BMEP on the basis of semi-empirical correlations. The fourth approach is based on the use of ANNs to directly estimate MFB50, PFP, and BMEP. The performance of the four models was evaluated under steady-state conditions and transient operation over WHTC.

Moreover, a new methodology, which is based on the sequential use of the Pearson correlation and partial correlation analysis, has been developed to identify the optimal set of input model parameters in order to a priori identify the correlation variables that should be excluded and the most robust correlation variables.

The main results can be summarized as follows:

(1) The new methodology set up for the identification of the model input parameters has allowed an improvement to be obtained in the calibration of the baseline physics-based model. Although the accuracy in the estimation of BMEP is quite similar, an improvement in the estimation of PFP, and MFB50 can be obtained, since the related RMSE values decrease from 2.47 bar to 1.82 bar and from 0.86 deg to 0.63 deg, respectively, at steady-state conditions. Moreover, it allows a saving in the model calibration computational time to be achieved.

(2) The direct ANN model has shown the best accuracy in the estimation of MFB50, PFP, and BMEP under both steady-state conditions (RMSE = 0.25 deg, 0.85 bar and 0.071 bar, respectively) and transient operation over the WHTC (RMSE = 1.1 deg, 9.6 bar and 0.7 bar, respectively); the accuracy of the baseline physics-based model is slightly worse than that of the direct ANN model (RMSE = 0.63 deg, 1.8 bar, 0.15 bar under steady-state conditions, RMSE = 1.2 deg, 10.5 bar, 0.8 bar over WHTC).

(3) The accuracy of the ANN physics-based model is higher than that of the baseline physics-based model under steady-state operations (RMSE = 0.48 deg, 1.6 bar, 0.17 bar), but it shows a marked deterioration for transient operation (RMSE = 1.2 deg, 13.8 bar, 0.8 bar), and this approach, therefore, seems to lack robustness.

(4) The accuracy of the semi-empirical model is worse than that of the physics-based model and of the direct ANN model under steady-state conditions (RMSE = 0.63 deg, 3 bar, 0.26 bar) and over the WHTC (RMSE = 1.4 deg, 11.7 bar, 0.8 bar). However, the accuracy can still be considered acceptable, in view of the simple mathematical structure of this kind of model, and the low number of tuning parameters that are necessary (and therefore of experimental data needed for calibration).

(5) The computational time required for the direct ANN model and semi-empirical model is less than 50 μs on an ETAS ES910 rapid prototyping device, while that of the baseline physics-based model is of the order of 350 μs.

Overall, the direct ANN model can be considered the best candidate for control-oriented applications, since:

- it is much simpler in structure and theory and easier to build than the physics-based model;
- it does not require any detailed knowledge or modeling of the physical and chemical processes;

- it is robust not only under steady-state conditions, but also in transient operation;
- its accuracy is high, even when a not so large dataset of experimental tests (410) is used for training;
- it features the best accuracy under steady-state and transient operating conditions
- the required computational time is lower than that of the physics-based model.

The main drawback of this kind of model is that it is of the black-box type. Thus, it does not provide detailed information about the physical and chemical processes that occur inside the combustion chamber, such as the heat release trend, in-cylinder pressure, or temperatures.

Author Contributions: The authors contributed equally to the preparation of the paper. Conceptualization, S.H., R.F., S.d.; Methodology, S.H., R.F., S.d.; Software, S.H.; Formal Analysis, S.H., R.F., S.d.; Data Curation, A.M. (Andrea Manelli), L.V., S.d.; Writing-Original Draft, S.H., R.F.; Writing-Review and Editing, R.F., M.M., A.M. (Antonio Mittica); Supervision, R.F., S.d., A.M. (Antonio Mittica), H.W., Y.W.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ABDC	After Bottom Dead Center
AF_{st}	stoichiometric air-to-fuel ratio
ANN	artificial neural network
BMEP	Brake Mean Effective Pressure (bar)
CFD	Computer Fluid-Dynamics
DoE	Design of Experiment
ECU	Engine Control Unit
EGR	Exhaust Gas Recirculation
EOC	End of Combustion
EOI	End of Injection
EVO	Exhaust Valve Opening
FMEP	Friction Mean Effective Pressure (bar)
H_L	lower heating value of the fuel
HCCI	Homogeneous Charge Compression Ignition
ICE	Internal Combustion Engines
IMEP360	gross Indicated Mean Effective Pressure (bar)
IMEP720	net Indicated Mean Effective Pressure (bar)
IVC	Intake Valve Closing
K	combustion rate coefficient
m	mass
m_{air}	trapped air mass
m_{EGR}	trapped EGR mass
$m_{f,inj}$	injected fuel mass (mg per cycle/cylinder)
$\dot{m}_{f,inj}$	fuel injection rate
MFB50	crank angle at which 50% of the fuel mass fraction has burned (deg)
n	polytropic coefficient for the compression phase
n'	polytropic coefficient for the expansion phase
N	engine rotational speed (1/min)
O_2	intake charge oxygen concentration (%)
p	pressure (bar)
PCCI	Premixed Charge Compression Ignition
p_{EMF}	exhaust manifold pressure (bar abs)
p_f	injection pressure (bar)
PFP	peak firing pressure

p_{IMF}	intake manifold pressure (bar abs)
PMEP	Pumping Mean Effective Pressure (bar)
q	injected fuel volume quantity (mm^3)
Q_{ch}	chemical heat release
$Q_{f,evap}$	fuel evaporation heat
q_{main}	total injected fuel volume quantity of the main pulse per cycle/cylinder
q_{tot}	total injected fuel volume quantity per cycle/cylinder
$q_{tot,pil}$	total injected fuel volume quantity of the pilot pulses per cycle/cylinder
Q_{fuel}	chemical energy associated with the injected fuel
$Q_{ht,glob}$	global heat transfer of the charge with the walls
Q_{net}	net heat release
R^2	squared correlation coefficient
RAF	relative air-to-fuel ratio
RMS	root mean square
RMSE	root mean square error
SOC	start of combustion
SOI	electric start of injection
SVP	support vector machine
t	time
T	temperature (K)
T_{EMF}	exhaust manifold temperature
T_{IMF}	intake manifold temperature
T_{SOC}	charge temperature at start of combustion
T_{SOI}	charge temperature at start of injection
V	volume
V2X	Vehicle-to-Everything
$Valve_{EGR}$	opening position of the high pressure EGR valve
VGT	Variable Geometry Turbine
WHTC	Worldwide Harmonized Heavy-duty Transient Cycle
$X_{r,EGR}$	EGR rate

Greek Symbols

$\Delta p_{exh\text{-}int}$	difference between the exhaust and intake manifold pressure
γ	isentropic coefficient
ρ	density
ρ_{IMF}	density in the inlet manifold
ρ_{SOC}	charge density at the start of combustion
ρ_{SOI}	charge density at the start of injection
τ	ignition delay coefficient

Appendix A Correlation Analysis for the Selection of the Model Input Variables

In order to make the correlation analysis more synthetic, the considered variables have been associated with indexing numbers (IN), as shown in Table A1.

For the sake of clarity, Figure A1 reports the temporal sequence of the calculation of the different variables used in the physics-based model.

The Pearson correlation coefficient was calculated, using Matlab R2016b, in order to identify the correlations between the variables identified in Table A1. The results are reported in Figure A2.

Table A1. Collected variables and corresponding index number (IN).

Variable Name	IN	Variable Name	IN	Variable Name	IN	Variable Name	IN
N	1	$Valve_{EGR}$	13	$T_{SOC,main}$	25	Q_{netmax}	37
p_f	2	X_{rEGR}	14	SOC_{pil}	26	Δp_{IMF}	38
SOI_{pil}	3	RAF	15	SOC_{main}	27	n	39
SOI_{main}	4	O_2	16	$\Delta p_{exh-int}$	28	n'	40
$q_{tot,pil}$	5	ρ_{IMF}	17	τ_{pil}	29	PFP	41
q_{main}	6	$\rho_{SOI,pil}$	18	τ_{main}	30	IMEP360	42
q_{tot}	7	$\rho_{SOI,main}$	19	K_{pil}	31	PMEP	43
p_{IMF}	8	$\rho_{SOC,pil}$	20	$K_{1,main}$	32	FMEP	44
T_{IMF}	9	$\rho_{SOC,main}$	21	$K_{2,main}$	33	IMEP720	45
p_{EMF}	10	$T_{SOI,pil}$	22	MFB50	34	BMEP	46
T_{EMF}	11	$T_{SOI,main}$	23	$Q_{ht,glob}$	35		
m_{air}	12	$T_{SOC,pil}$	24	$Q_{f,evap}$	36		

The numbers in the rows and columns in Figure A2 are the indexing numbers of the variables shown in Table A1. The background color changes from green to white and from white to red when the Pearson coefficient value varies in the $(-1, 0)$ range or in the $(0, 1)$ range, respectively. The linear relationship between two variables is considered strong when their Pearson coefficient value is in the $(-1.0, -0.5)$ or $(0.5, 1)$ ranges, moderate when it is in the $(-0.5, -0.3)$ or $(0.3, 0.5)$ ranges, weak when it is in the $(-0.3, -0.1)$ or $(0.1, 0.3)$ ranges, and non-existent or very weak when from it is in the $(-0.1, 0.1)$ range [43].
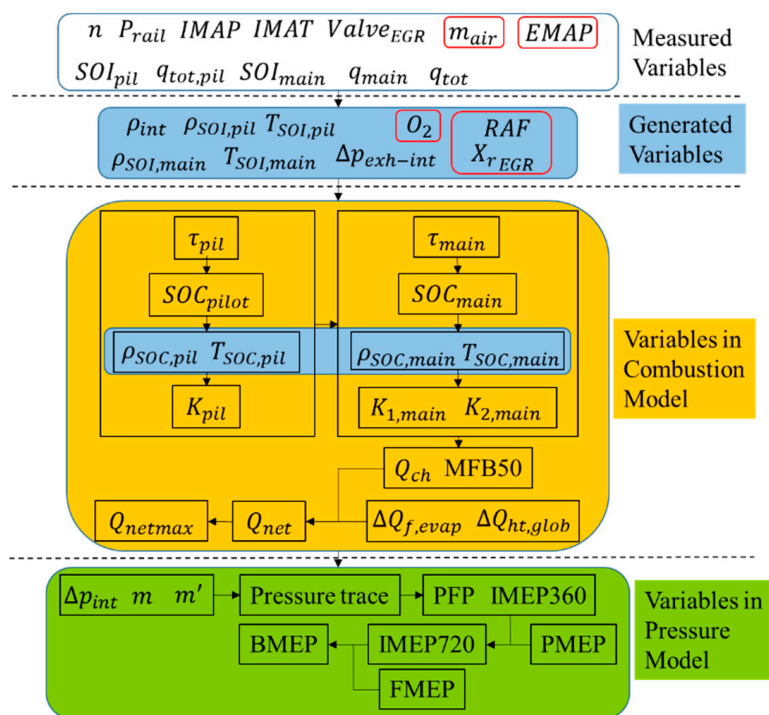

Figure A1. Temporal calculation sequence of the physics-based model variables.

The Pearson coefficient can provide useful information for the selection of:

- the input parameters of the correlations which are used to estimate the calibration parameters of the physics-based models (i.e., those reported in Table 3), or the input parameters of the ANNs that are used to estimate the same calibration parameters;
- the input parameters of the semi-empirical correlations that are used to directly estimate MFB50, PFP, BMEP;
- the input parameters of the ANNs that are used to directly estimate MFB50, PFP, BMEP.
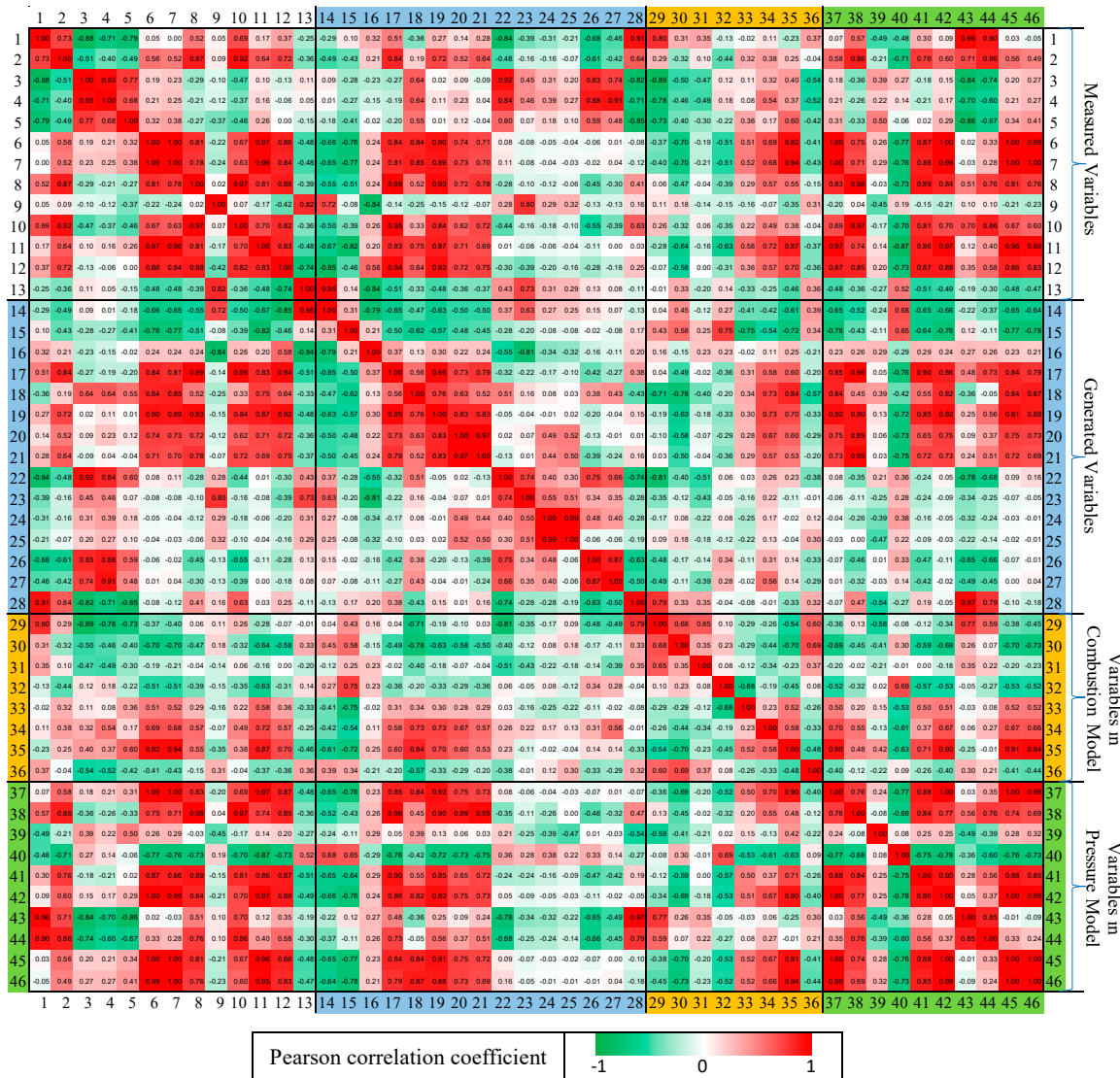

Figure A2. Pearson correlation coefficients for all the possible combinations of the variables reported in Table A1.

As can be seen in Figure A2, the groups of variables [IN 3, IN 4], [IN 6, IN 7], [IN 13, IN 14], [IN 17, IN 19], [IN 20, IN 21], [IN 24, IN 25] and [IN 26, IN 27], namely [SOI_{pil}, SOI_{main}] [q_{main}, q_{tot}], [$Valve_{EGR}$, $X_{r,EGR}$], [ρ_{IMF}, $\rho_{SOImain}$], [$\rho_{SOC,pil}$, $\rho_{SOC,main}$], [$T_{SOC,pil}$, $T_{SOC,main}$] and [SOC_{pil}, SOC_{main}], are very closely correlated to each other, since their crossed Pearson coefficient is almost equal to 1. Therefore, during the identification of the input variables of the models, it is sufficient to select only one of them as representative of each variable group. Moreover, it should be noted that the variables which are calculated at an earlier stage are expected to be closely correlated to the variables that are calculated in a subsequent stage (see Figure A1).

The following rules were identified to select the set of input variables to use in the correlations of the physics-based model or in the semi-empirical and ANN-based models, on the basis of the information derived from the Pearson correlation analysis:

- the temporal sequences of the variables in the model have to be taken into account, i.e., variables cannot be used as the independent variables if they are calculated or occur after the dependent variable, according to the scheme in Figure A1;

- the directly measured engine operation variables (namely N, p_f, SOI, fuel injection quantity, p_{IMF}, T_{IMF}, $Valve_{EGR}$) should be considered as independent variable candidates with top priority, because they are the primary cause of the changes in engine operation. Moreover, these variables are usually very precise because they are measured directly in real-time. The generated parameters (ρ_{SOI}, T_{SOI}), based on SOI, should be considered with higher priority than those based on SOC (ρ_{SOC}, T_{SOC}), because of the uncertainty chain for the estimation of SOC on the basis of SOI;

- the dependent variables (such as τ_{pil}, τ_{main}, SOC, K_{pil}, $K_{1,main}$, $K_{2,main}$, $Q_{f,evap}$, $Q_{ht,glob}$, n, n', Δp_{IMF}, PMEP, FMEP) should be considered with a lower priority as correlation parameters, because they are usually less precise than the directly measured parameters. The $X_{r,EGR}$, RAF, O_2, ρ_{SOC} and T_{SOC} generated parameters should also be considered with a lower priority because of their relatively low precision;

- some variable groups (e.g., [SOI_{pil}, SOI_{main}] [q_{main}, q_{tot}], [$Valve_{EGR}$, $X_{r,EGR}$], [p_{IMF}, m_{air}, ρ_{IMF}], [$\rho_{SOC,pil}$, $\rho_{SOC,main}$], [$T_{SOC,pil}$, $T_{SOC,main}$] and [SOC_{pil}, SOC_{main}]) are constituted by two variables which are closely correlated to each other. Therefore, only one variable should be selected as representative of each group;

- the total number of independent variable candidates should not be too high (in this study a maximum of 15 variables was chosen for the estimation of each dependent one), in order to ensure that the partial coefficient order is not too large and to make the model more robust to input uncertainty.

Some considerations should also been made concerning the m_{air} and p_{EMF} variables. Although m_{air} is measured directly, it is usually not a precise or stable variable, since the accuracy of the engine air mass flow sensor is not very high (~5%), and its use, therefore, needs to be evaluated carefully. Moreover, many medium and heavy-duty engines are not generally equipped with air mass flow sensors and the exhaust manifold pressure is generally not measured in production engines, so its use as an input variable for the model should also be evaluated carefully (in this study it was avoided).

On the basis of the rules explained above, an analysis of the Pearson correlation coefficients was carried out to select independent variable candidates for each dependent variable. A summary is reported in Table A2.

It should be recalled that some of the dependent variables are the calibration parameters of the physics-based model (see Table 3), and they need to be modeled through correlations or ANNs, while other dependent variables (i.e., MFB50, PFP, BMEP) are the direct outcome of the semi-empirical model or of the direct ANN model, i.e., the metrics to be used for combustion control applications.

After the previous step, the partial correlation coefficient between each input variable of the initial candidate set reported in Table A2 and the corresponding dependent variable was calculated. The results are reported in Table A3, where the previously estimated Pearson correlation coefficients are also reported. The simultaneous analysis of the Pearson correlation and of the partial correlation coefficients is very useful, since it allows the least and most robust correlation variables to be identified.

Table A2. Initial candidate set of the input variables for each output parameter, identified on the basis of the Pearson correlation analysis.

Dependent Variable	Initial Candidate Set of the Independent Variables Selected Using Pearson Correlation Analysis
RAF	$N, p_f, SOI_{pil}, q_{tot,pil}, q_{tot}, p_{IMF}, T_{IMF}, X_{rEGR}, \rho_{SOI,pil}, \rho_{SOI,main}$
τ_{pil}	$N, p_f, SOI_{pil}, q_{tot,pil}, p_{IMF}, T_{IMF}, Valve_{EGR}, RAF, \rho_{SOI,pil}, T_{SOI,pil}$
τ_{main}	$N, p_f, SOI_{pil}, q_{tot,pil}, q_{tot}, T_{IMF}, m_{air}, X_{rEGR}, RAF, \rho_{SOI,pil}, \rho_{SOI,main}, \rho_{SOC,pil}, T_{SOI,pil}, \Delta p_{exh-int}, \tau_{pil}$
K_{pil}	$N, p_f, SOI_{pil}, q_{tot,pil}, p_{IMF}, T_{IMF}, Valve_{EGR}, \rho_{SOI,pil}, T_{SOI,pil}, \tau_{pil}, \Delta p_{exh-int}$
$K_{1,main}$	$N, p_f, SOI_{main}, q_{tot,pil}, q_{tot}, p_{IMF}, T_{IMF}, Valve_{EGR}, RAF, \rho_{SOI,main}, \rho_{SOC,pil}$
$K_{2,main}$	$N, p_f, SOI_{pil}, q_{tot,pil}, q_{tot}, T_{IMF}, m_{air}, X_{rEGR}, RAF, \rho_{SOI,pil}, \rho_{SOI,main}, T_{SOI,main}, K_{1,main}$
$Q_{ht,glob}$	$N, p_f, SOI_{pil}, q_{tot,pil}, q_{tot}, T_{IMF}, m_{air}, X_{rEGR}, RAF, \rho_{SOI,pil}, \rho_{SOC,pil}, \tau_{main}, K_{2,main}$
$Q_{f,evap}$	$N, p_f, SOI_{pil}, q_{tot,pil}, q_{tot}, T_{IMF}, m_{air}, \rho_{SOI,pil}, \rho_{SOI,main}, RAF, Valve_{EGR}, T_{SOI,pil}, \Delta p_{exh-int}$
Δp_{IMF}	$N, p_f, SOI_{pil}, q_{tot,pil}, q_{main}, p_{IMF}, X_{rEGR}, \rho_{SOI,pil}, \rho_{SOI,main}, T_{SOI,pil}, \rho_{SOC,main}, \Delta p_{exh-int}$
n	$N, p_f, SOI_{pil}, q_{tot,pil}, q_{tot}, p_{IMF}, T_{IMF}, Valve_{EGR}, \rho_{SOI,pil}, \rho_{SOI,main}, T_{SOI,pil}, T_{SOI,main}, T_{SOC,main}, \Delta p_{exh-int}$
n'	$N, p_f, q_{tot,pil}, q_{main}, SOI_{pil}, T_{IMF}, Valve_{EGR}, RAF, \rho_{IMF}, \rho_{SOI,main}, T_{SOI,pil}, \Delta p_{exh-int}, K_{1,main}, K_{2,main}$
MFB50	$N, p_f, SOI_{main}, q_{tot,pil}, q_{main}, T_{IMF}, X_{rEGR}, RAF, \rho_{IMF}, \rho_{SOI,pil}, \rho_{SOI,main}, T_{SOI,pil}$
PFP	$N, p_f, q_{tot,pil}, q_{main}, SOI_{main}, T_{IMF}, X_{rEGR}, RAF, \rho_{IMF}, \rho_{SOI,pil}, \rho_{SOI,main}, T_{SOI,main}$
PMEP	$N, p_f, q_{tot,pil}, SOI_{pil}, p_{IMF}, T_{IMF}, Valve_{EGR}, \tau_{pil}, T_{SOI,pil}, \Delta p_{exh-int}$
FMEP	$N, p_f, SOI_{pil}, q_{tot,pil}, q_{main}, p_{IMF}, T_{IMF}, Valve_{EGR}, \Delta p_{IMF}, PFP, \rho_{SOI,mian}, T_{SOI,pil}, \Delta p_{exh-int}$
BMEP	$N, p_f, q_{tot,pil}, q_{tot}, SOI_{main}, T_{IMF}, m_{air}, X_{rEGR}, RAF, \rho_{SOI,pil}, \rho_{SOI,main}$

In fact, input variables that are characterized by low Pearson and partial correlation coefficients can be removed from the candidate list, while input variables which are characterized by a high value of the Pearson correlation and the partial correlation coefficient at the same time can be selected as very robust correlation variables. Particular attention should also be paid to the input variables which, although showing a strong correlation with the dependent variables, are characterized by a limited precision (since they are derived parameters or measured parameters with low accuracy sensors). On the basis of the previous considerations, the least robust correlation variables are highlighted in Table A3 with a gray background, while the highly correlated variables are marked with an orange background, and finally the highly correlated variables characterized by a low precision are marked with a blue background.

It was found that some input variables (i.e., those marked with a white background in Table A3) are characterized by an intermediate value of the correlation coefficients. In order to understand whether they should have been included or excluded from the input candidate list, a sensitivity analysis was carried out, considering all the possible combinations of these variables and estimating the accuracy of the related models, as reported in Section 4.

As reported in the same section, it should be noted that the proposed approach, i.e., the preliminary selection of the input variables through Pearson and partial correlation coefficients, and a subsequent sensitivity analysis carried out on a reduced set of variables, is more computationally efficient than a pure sensitivity analysis approach which includes the entire set of input variables.

Table A3. Pearson and partial correlation coefficients for the input variables of the initial candidate set.

Dependent Variable		Initial Candidate of the Independent Variables													
RAF	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{tot}	p_{IMF}	T_{IMF}	X_{rEGR}	$\rho_{SOI,pil}$	$\rho_{SOI,main}$				
	Pearson coefficient	0.10	−0.43	−0.28	−0.41	−0.77	−0.51	−0.08	0.31	−0.62	−0.57				
	Partial coefficient	−0.46	−0.56	−0.47	−0.48	−0.29	−0.24	0.19	−0.50	0.42	0.21				
τ_{pil}	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	p_{IMF}	T_{IMF}	$Valve_{EGR}$	RAF	$\rho_{SOI,pil}$	$T_{SOI,pil}$				
	Pearson coefficient	0.80	0.29	−0.89	−0.73	0.06	0.11	−0.01	0.43	−0.71	−0.81				
	Partial coefficient	0.44	−0.37	−0.13	−0.06	−0.15	0.13	−0.16	0.12	−0.05	0.44				
τ_{main}	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{tot}	T_{IMF}	m_{air}	X_{rEGR}	RAF	$\rho_{SOI,pil}$	$\rho_{SOI,main}$	$\rho_{SOC,pil}$	$T_{SOI,pil}$	$\Delta p_{exh-int}$ / τ_{pil}
	Pearson coefficient	0.31	−0.32	−0.50	−0.40	−0.70	0.18	−0.58	0.45	0.58	−0.78	−0.63	−0.58	−0.40	0.33 / 0.68
	Partial coefficient	0.26	−0.47	−0.27	0.03	−0.18	−0.01	−0.05	−0.01	−0.13	0.34	−0.29	0.06	0.00	0.08 / 0.26
K_{pil}	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	p_{IMF}	T_{IMF}	$Valve_{EGR}$	$\rho_{SOI,pil}$	$T_{SOI,pil}$	τ_{pil}	$\Delta p_{exh-int}$			
	Pearson coefficient	0.35	0.10	−0.47	−0.30	−0.04	−0.14	0.25	−0.40	−0.51	0.65	0.35			
	Partial coefficient	−0.45	0.36	0.02	−0.01	0.13	−0.15	−0.07	−0.13	0.10	0.76	−0.16			
$K_{1,main}$	Independent variable	N	p_f	$q_{tot,pil}$	q_{tot}	SOI_{main}	p_{IMF}	T_{IMF}	$Valve_{EGR}$	RAF	$\rho_{SOI,main}$	$\rho_{SOC,main}$			
	Pearson coefficient	−0.13	−0.44	−0.22	−0.51	0.18	−0.39	−0.15	0.14	0.75	−0.33	−0.36			
	Partial coefficient	−0.34	0.13	−0.20	0.04	−0.44	−0.19	0.15	−0.26	−0.12	−0.49	0.69			
$K_{2,main}$	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{tot}	T_{IMF}	m_{air}	X_{rEGR}	RAF	$\rho_{SOI,pil}$	$\rho_{SOI,main}$	$T_{SOI,main}$	$K_{1,main}$	
	Pearson coefficient	−0.02	0.32	0.11	0.36	0.52	−0.16	0.36	−0.41	−0.75	0.34	0.30	−0.16	−0.68	
	Partial coefficient	0.01	−0.06	−0.09	−0.17	−0.20	0.13	−0.06	−0.50	0.21	−0.27	−0.10	−0.27	0.01	
$Q_{ht,glob}$	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{tot}	T_{IMF}	m_{air}	RAF	X_{rEGR}	$\rho_{SOI,pil}$	$\rho_{SOC,pil}$	τ_{main}	$K_{2,main}$	
	Pearson coefficient	−0.23	0.25	0.40	0.60	0.94	−0.35	0.70	−0.72	−0.61	0.84	0.60	−0.70	0.52	
	Partial coefficient	0.33	−0.48	0.40	−0.04	0.93	−0.44	−0.08	0.34	0.14	−0.43	−0.24	−0.42	0.09	
$Q_{f,evap}$	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{tot}	T_{IMF}	m_{air}	RAF	$Valve_{EGR}$	$\rho_{SOI,pil}$	$\rho_{SOI,main}$	$T_{SOI,pil}$	$\Delta p_{exh-int}$	
	Pearson coefficient	0.37	−0.04	−0.54	−0.42	−0.43	0.31	−0.36	0.34	0.36	−0.57	−0.33	−0.38	0.32	
	Partial coefficient	0.07	−0.14	−0.31	0.04	−0.01	−0.07	−0.27	0.03	−0.03	0.26	0.09	−0.06	0.07	
Δp_{IMF}	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{main}	p_{IMF}	X_{rEGR}	$\rho_{SOI,pil}$	$\rho_{SOI,main}$	$T_{SOI,pil}$	$\rho_{SOC,main}$	$\Delta p_{exh-int}$		
	Pearson coefficient	0.57	0.86	−0.36	−0.33	0.75	0.98	−0.52	0.45	0.90	−0.35	0.95	0.47		
	Partial coefficient	−0.09	−0.13	0.06	−0.03	−0.02	0.42	−0.02	−0.09	0.16	−0.12	−0.15	0.18		
n	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{tot}	p_{IMF}	T_{IMF}	$Valve_{EGR}$	$\rho_{SOI,pil}$	$\rho_{SOI,main}$	$T_{SOI,pil}$	$T_{SOI,main}$	$T_{SOC,main}$	$\Delta p_{exh-int}$
	Pearson coefficient	−0.49	−0.21	0.39	0.35	0.29	−0.03	−0.45	−0.27	0.39	0.13	0.21	−0.25	−0.47	−0.54
	Partial coefficient	−0.45	0.09	−0.15	−0.28	0.24	−0.19	0.19	−0.01	−0.01	0.13	−0.05	0.03	−0.42	−0.16
n'	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{main}	T_{IMF}	X_{rEGR}	RAF	p_{IMF}	$\rho_{SOI,main}$	$T_{SOI,pil}$	$\Delta p_{exh-int}$	$K_{1,main}$	$K_{2,main}$
	Pearson coefficient	−0.48	−0.71	0.27	−0.06	−0.77	0.19	0.68	0.65	−0.76	−0.72	0.36	−0.27	0.69	−0.53
	Partial coefficient	−0.69	−0.13	−0.04	−0.24	−0.36	0.05	0.04	−0.04	0.07	0.05	−0.03	0.25	0.52	0.20
$MFB50$	Independent variable	N	p_f	SOI_{main}	$q_{tot,pil}$	q_{main}	T_{IMF}	X_{rEGR}	RAF	p_{IMF}	$\rho_{SOI,pil}$	$\rho_{SOI,main}$	$T_{SOI,pil}$		
	Pearson coefficient	0.11	0.38	0.54	0.17	0.69	0.57	−0.42	−0.54	0.58	0.73	0.73	0.26		
	Partial coefficient	0.76	−0.64	0.84	0.07	0.35	0.13	−0.10	−0.10	0.07	−0.10	0.16	0.13		
PFP	Independent variable	N	p_f	SOI_{main}	$q_{tot,pil}$	q_{main}	T_{IMF}	X_{rEGR}	RAF	p_{IMF}	$\rho_{SOI,pil}$	$\rho_{SOI,main}$	$T_{SOI,main}$		
	Pearson coefficient	0.30	0.76	−0.21	0.02	0.87	0.89	−0.65	−0.64	0.90	0.55	0.85	−0.24		
	Partial coefficient	−0.68	0.59	−0.69	0.21	0.36	−0.32	−0.10	−0.13	0.35	−0.20	−0.25	0.30		

Table A3. *Cont.*

Dependent Variable		Initial Candidate of the Independent Variables											
PMEP	Independent variable	N	p_f	$q_{tot,pil}$	SOI_{pil}	p_{IMF}	T_{IMF}	$Valve_{EGR}$	$T_{SOI,pil}$	τ_{pil}	$\Delta p_{exh-int}$		
	Pearson coefficient	0.96	0.71	−0.86	−0.84	0.51	0.10	−0.19	−0.78	0.77	0.97		
	Partial coefficient	0.57	−0.05	−0.58	0.25	0.32	0.09	0.04	−0.21	−0.05	0.84		
FMEP	Independent variable	N	p_f	SOI_{pil}	$q_{tot,pil}$	q_{main}	p_{IMF}	T_{IMF}	$Valve_{EGR}$	Δp_{IMF}	PFP	$\rho_{SOI,main}$	$T_{SOI,pil}$	$\Delta p_{exh-int}$
	Pearson coefficient	0.90	0.86	−0.74	−0.67	0.33	0.76	0.10	−0.30	0.56	0.56	0.56	−0.68	0.79
	Partial coefficient	0.28	0.08	−0.03	0.18	−0.25	0.15	0.15	−0.18	−0.28	−0.00	0.25	−0.04	−0.07
BMEP	Independent variable	N	p_f	SOI_{main}	$q_{tot,pil}$	q_{tot}	T_{IMF}	m_{air}	X_{rEGR}	RAF		$\rho_{SOI,main}$		
	Pearson coefficient	−0.05	0.49	0.27	0.41	1.00	−0.23	0.83	−0.64	−0.78	0.87	0.88		
	Partial coefficient	−0.78	0.51	−0.64	0.43	0.97	−0.01	−0.07	−0.11	−0.43	0.20	0.61		

Appendix B Training and Selection of the ANNs

Figures A3–A6 report the mean values of R^2 and MSE for 200 training trials of the ANNs that simulate the parameters of the physics-based model (Figures A3–A5) and the combustion metrics (Figure A6).
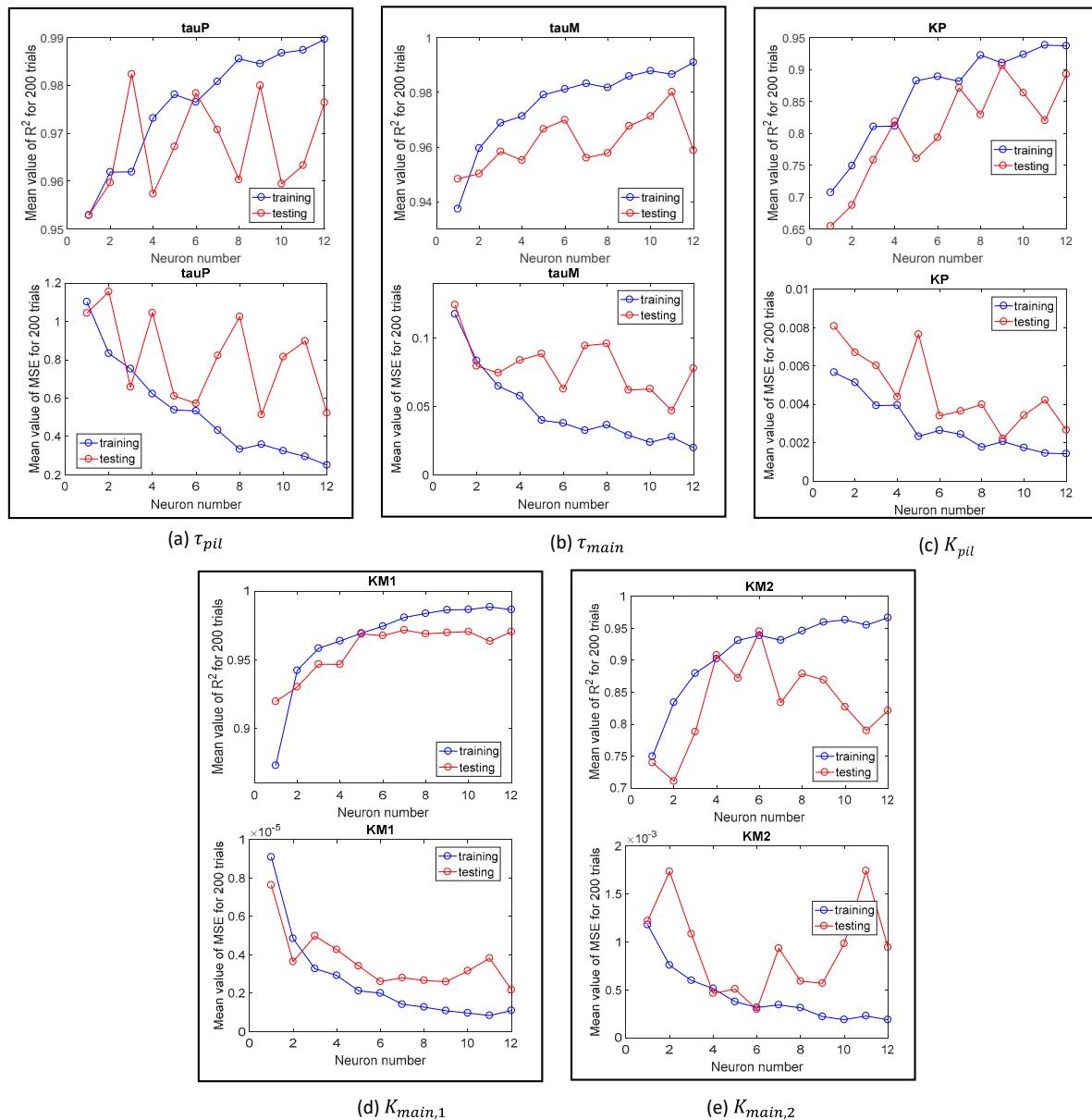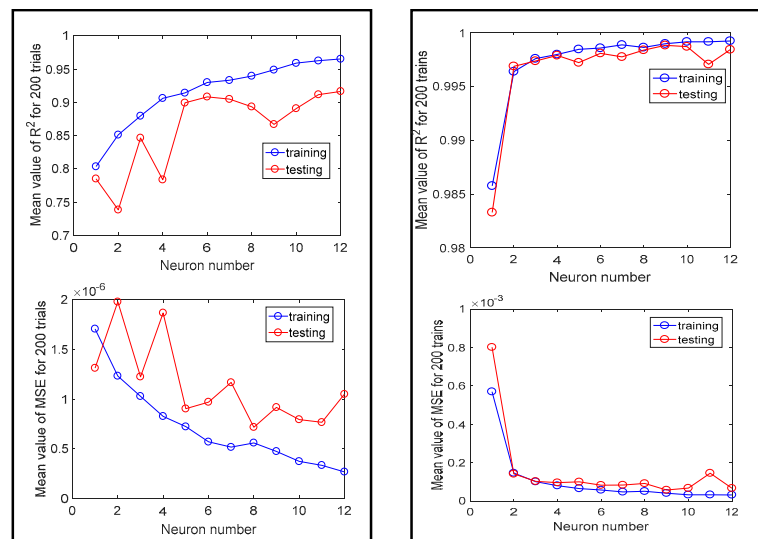

(a) τ_{pil}

(b) τ_{main}

(c) K_{pil}

(d) $K_{main,1}$

(e) $K_{main,2}$

Figure A3. (**a**–**e**) Mean value of R^2 and MSE over 200 training trials for the ANNs that simulate the different parameters of the physics-based model (heat release simulation), as a function of the number of neurons.
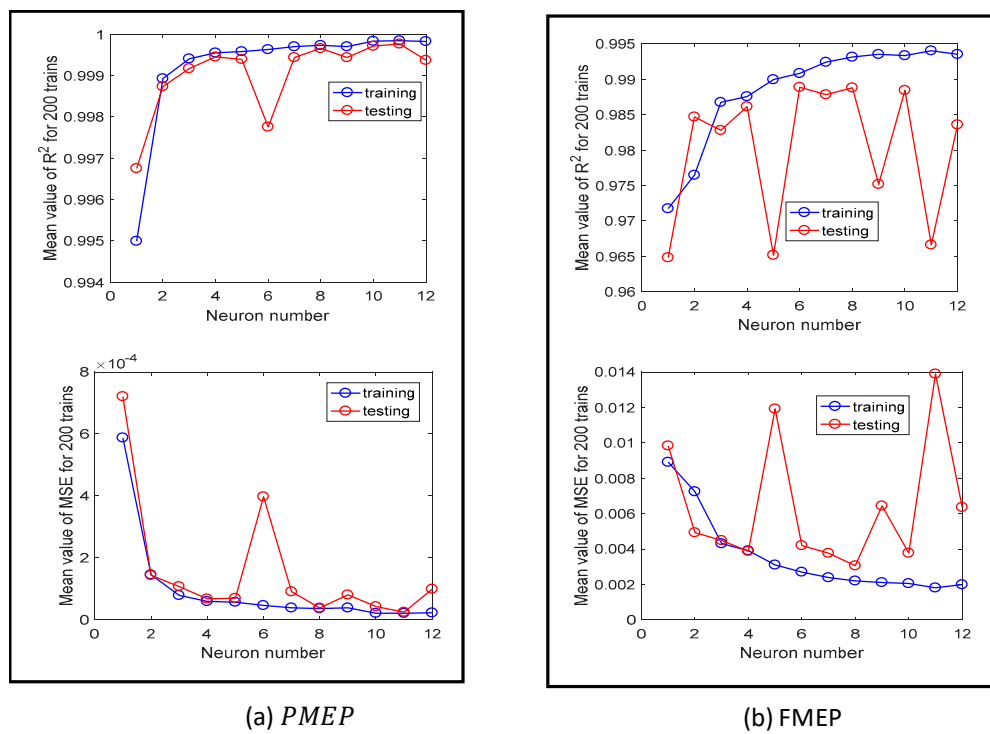
It can be noted, from the charts, that the training precision is generally higher than the testing precision, for both R^2 and MSE, in all cases. It is obvious that the training precision will always increase when the number of neurons increases, since the neural network is tested over the same dataset as that used for training. Moreover, it can be seen that the training precision increases slightly as the number of neurons increases when the number of neurons is above a certain value. As far as the testing precision is concerned, although some fluctuations occur, a rising trend of R^2 and a decreasing trend of MSE are also in general observed at the beginning when the number of neurons increases, and the trends

then tend to become stable or even reverse. Reversing of the trend indicates overfitting. The precision of both the training and testing is quite low when the number of neurons is low, and this means that a too small number of neurons cannot capture the data characteristics accurately (underfitting). An interesting phenomenon can be seen with reference to the trends of the training precision of MFB50, PFP, and BMEP (Figure A6), since they fluctuate less than the other parameters of the physics-based model (Figures A3–A5). This can be ascribed to the fact that the calibration precision of the ANNs that simulate MFB50, PFP, and BMEP is higher than that of the other parameters.

(a) $Q_{f,evap}$ (b) $Q_{ht,glob}$

(c) n (d) n' (e) ΔP_{IMF}

Figure A4. (**a**–**e**) Mean value of R^2 and MSE over 200 training trials for the ANNs that simulate the different parameters of the physics-based model (pressure simulation), as a function of the number of neurons.

(a) *PMEP* (b) FMEP

Figure A5. (**a**,**b**) Mean value of R^2 and MSE over 200 training trials for the ANNs that simulate the different parameters of the physics-based model (BMEP simulation), as a function of the number of neurons.
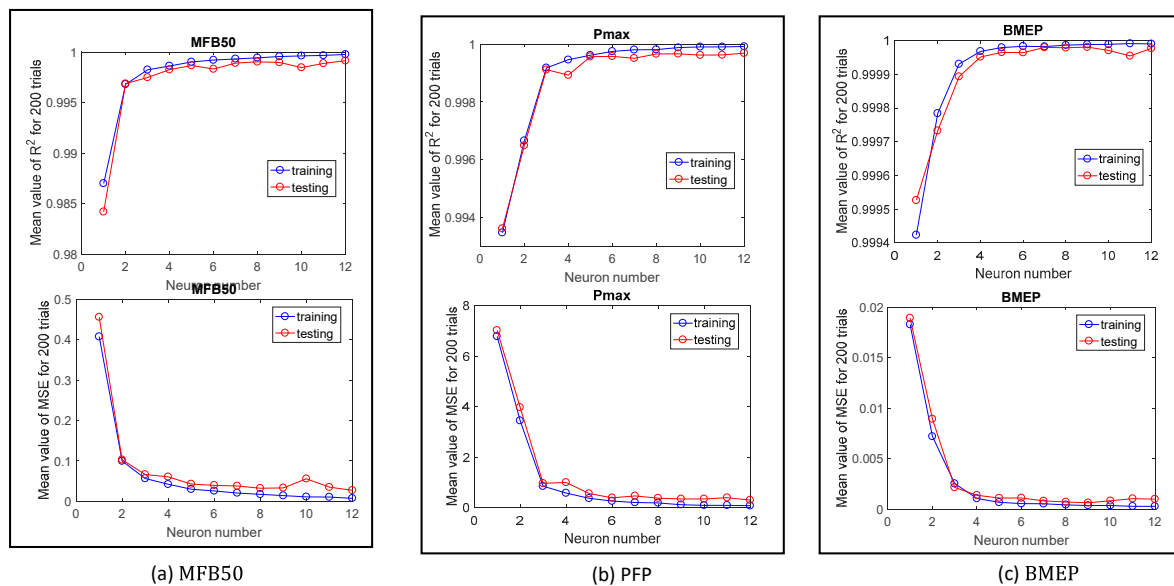

(a) MFB50 (b) PFP (c) BMEP

Figure A6. (**a**–**c**) Mean value of R^2 and MSE over 200 training trials for the ANNs that simulate MFB50 (**a**), PFP (**b**) and BMEP (**c**) combustion metrics, as a function of the number of neurons.

Once the sensitivity analysis had been carried out, the best ANN out of the 200 trials was selected for each parameter. The automatic selection algorithm is shown hereafter (Algorithm A1):

Algorithm A1. Automatic selection

---

    for i = 1:200
if i = 1
ANN_best = ANN_i;
R2_best = [R2(i,1), R2 (i,2), R2 (i,3)];
MSE_best = [MSE (i,1), MSE (i,2), MSE (i,3)];
else
if R2(i,3) > R2_best(3) && (R2 (i,2) > R2_best(2) || R2 (i,1) > R2_best(1)) && abs(R2 (i,1)−R2(i,2)) < 0.10*
R2_best(3)
ANN_best = ANN_i;
R2_best = [R2(i,1), R2 (i,2), R2 (i,3)];
MSE_best = [MSE (i,1), MSE (i,2), MSE (i,3)];
end
end
end

Where, i is the trial number; ANN_best is the best ANN selected by this algorithm; ANN_i is the trained ANN for ith trial; R2(i,1), R2 (i,2) and R2 (i,3) are the R^2 precisions for training, testing and total data validation, respectively; MSE (i,1), MSE (i,2) and MSE (i,3) are the MSE precisions for training, testing and total data validation, respectively; abs is the absolute value function.

Three conditions are used to decide whether to update the ANN_best variable for this algorithm, namely R2(i,3) > R2_best(3), (R2 (i,2) > R2_best(2) || R2 (i,1) > R2_best(1)) and abs(R2 (i,1) − R2(i,2)) < 0.10* R2_best(3), respectively. ANN_best is updated only when all the three conditions are true. The first condition, i.e., R2(i,3) > R2_best(3), means that the precision of the total data validation becomes better; the second one, i.e., (R2 (i,2) > R2_best(2) || R2 (i,1) > R2_best(1)), means that the precision of the training data or testing data becomes better; the third one, i.e., abs(R2 (i,1)−R2(i,2)) < 0.10* R2_best(3), means that the difference between the training and testing precisions is smaller than a given level. The latter condition is used to ensure that the testing and training precisions are not so different, and this can prevent the need to select a trained ANN which may present overfitting or underfitting.

Finally the best ANN out of the 200 trials is selected for each parameter, and the fitting results are compared with experimental data, as shown in Figures A7–A10. It can be observed that both the training and testing precisions are quite high (all the R^2 values are higher than 0.85) and there is not much difference between them. This indicates that all the selected ANNs present quite good fitting and prediction performances.
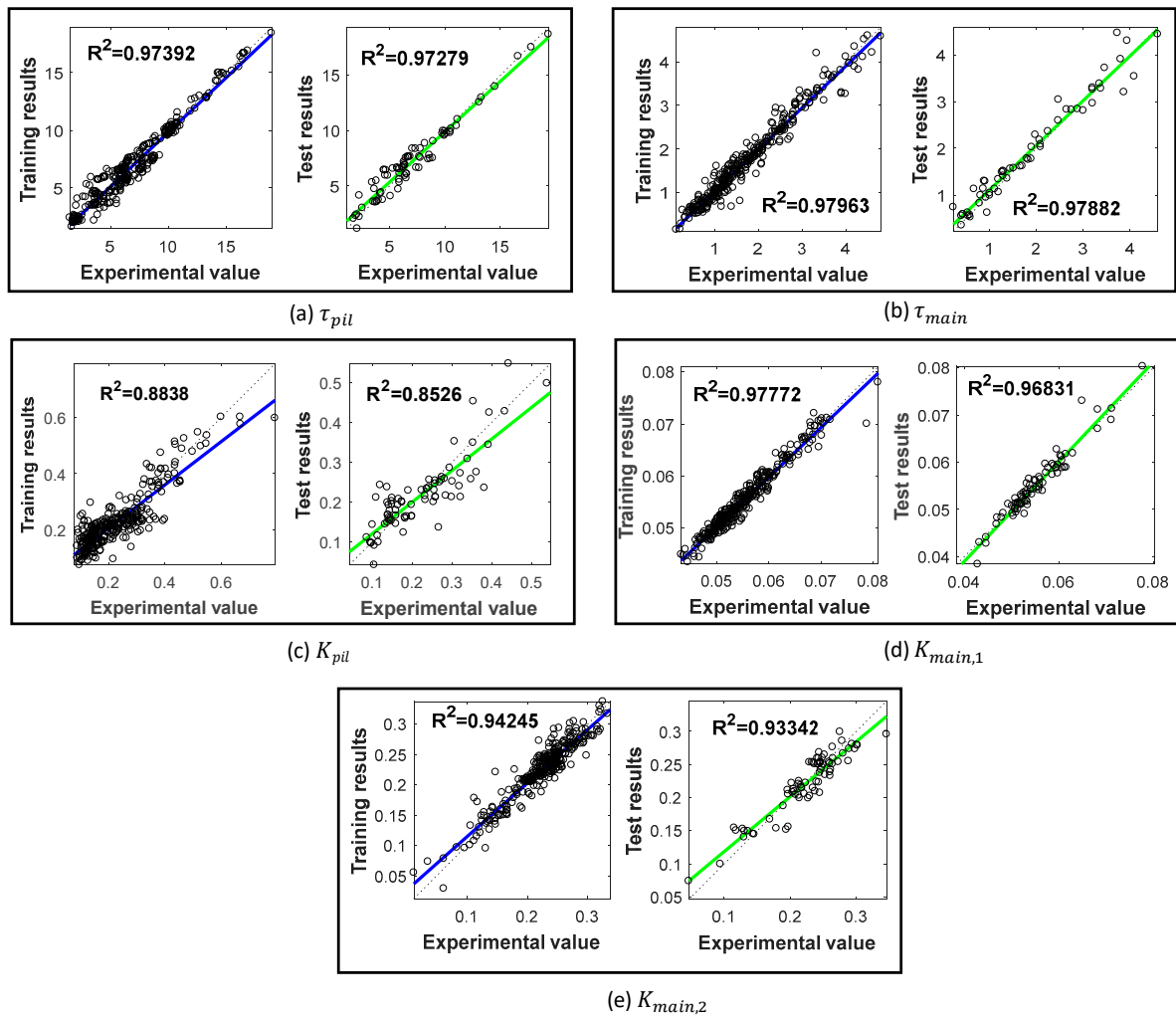
(a) τ_{pil}

(b) τ_{main}

(c) K_{pil}

(d) $K_{main,1}$

(e) $K_{main,2}$

Figure A7. (**a**–**e**) Comparison of the experimental and ANN fitting results for the physics-based model parameters (heat release simulation).
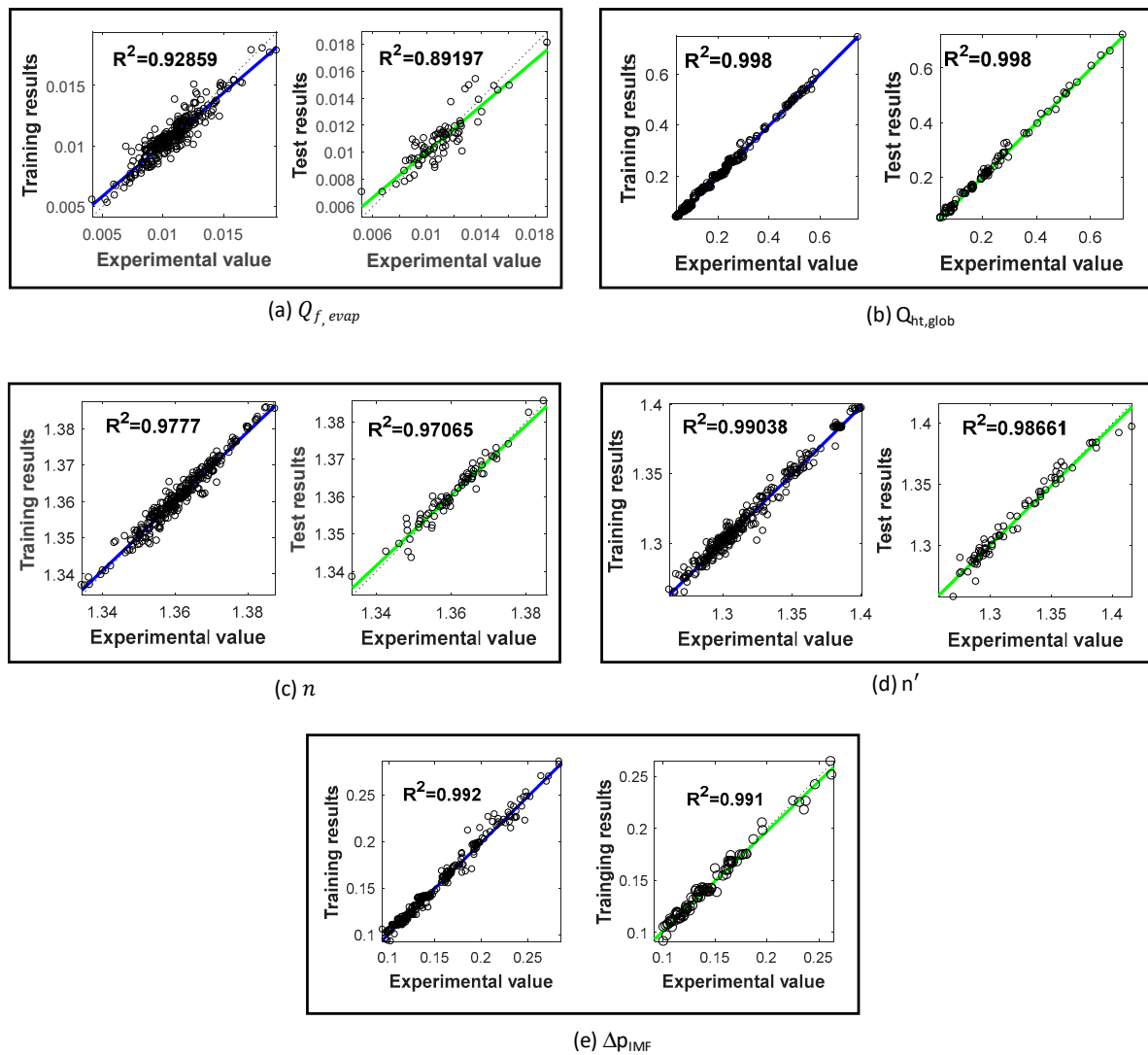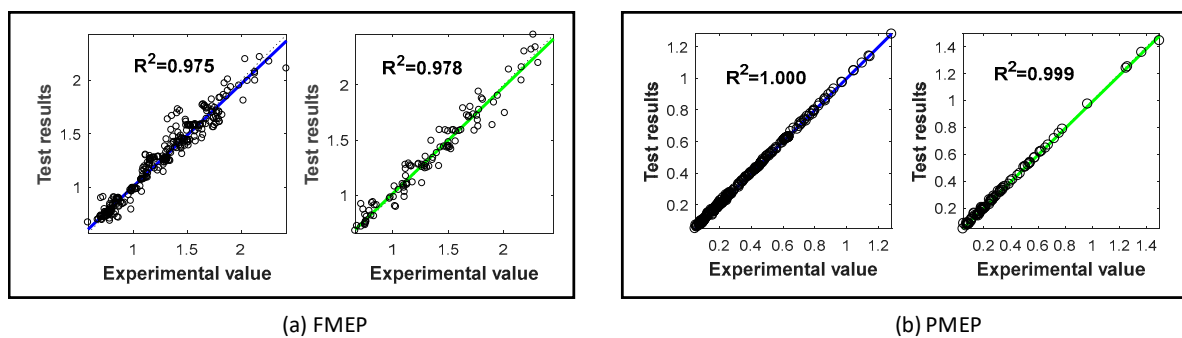
(a) $Q_{f,evap}$

(b) $Q_{ht,glob}$

(c) n

(d) n'

(e) Δp_{IMF}

Figure A8. (**a**–**e**) Comparison of the experimental and ANN fitting results for the physics-based model parameters (pressure simulation).

(a) FMEP

(b) PMEP

Figure A9. (**a**,**b**) Comparison of the experimental and ANN fitting results for the physics-based model parameters (BMEP simulation).
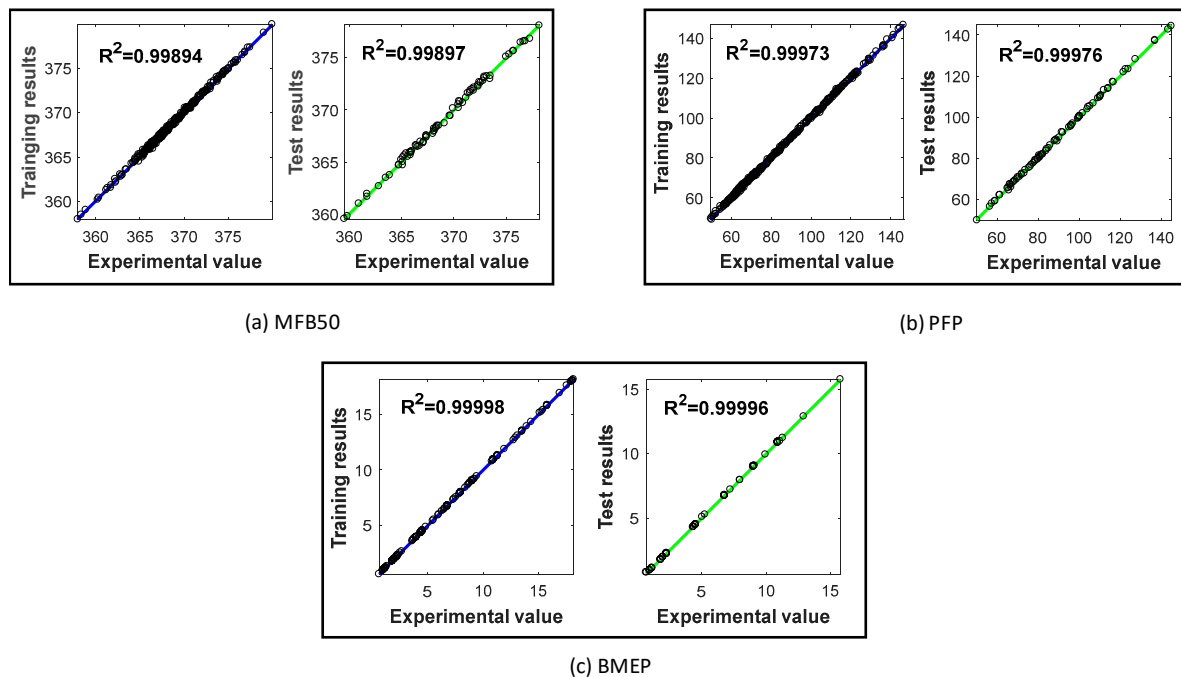
(a) MFB50

(b) PFP

(c) BMEP

Figure A10. Comparison of the experimental and ANN fitting results for the direct estimation of MFB50 (**a**), PFP (**b**) and BMEP (**c**).

References

1. Payri, F.; Olmeda, P.; Martín, J.; García, A. A complete 0D thermodynamic predictive model for direct injection diesel engines. *Appl. Energy* **2011**, *88*, 4632–4641. [CrossRef]

2. Maroteaux, F.; Saad, C.; Aubertin, F. Development and validation of double and single Wiebe function for multi-injection mode Diesel engine combustion modelling for hardware-in-the-loop applications. *Energy Convers. Manag.* **2015**, *105*, 630–641. [CrossRef]

3. Hu, S.; Wang, H.; Niu, X.; Li, X.; Wang, Y. Automatic calibration algorithm of 0-D combustion model applied to DICI diesel engine. *Appl. Therm. Eng.* **2018**, *130*, 331–342. [CrossRef]

4. Ferrari, A.; Mittica, A.; Pizzo, P.; Jin, Z. PID Controller Modelling and Optimization in Cr Systems with Standard and Reduced Accumulators. *Int. J. Automot. Technol.* **2018**, *19*, 771–781. [CrossRef]

5. Ferrari, A.; Mittica, A.; Pizzo, P.; Wu, X.; Zhou, H. New methodology for the identification of the leakage paths and guidelines for the design of common rail injectors with reduced leakage. *J. Eng. Gas Turbines Power* **2018**, *140*, 022801. [CrossRef]

6. Ferrari, A.; Mittica, A.; Paolicelli, F.; Pizzo, P. Hydraulic Characterization of Solenoid-actuated Injectors for Diesel Engine Common Rail Systems. *Energy Procedia* **2016**, *101*, 878–885. [CrossRef]

7. Catania, A.E.; Ferrari, A.; Mittica, A.; Spessa, E. Common Rail without Accumulator: Development, Theoretical-Experimental Analysis and Performance Enhancement at DI-HCCI Level of a New Generation FIS. *SAE Tech. Paper Ser.* **2007**. [CrossRef]

8. Catania, A.; Ferrari, A.; Mittica, A. High-pressure rotary pump performance in multi-jet common rail systems. In Proceedings of the 8th Biennial ASME Conference on Engineering Systems Design and Analysis; ESDA2006, Engineering Systems Design and Analysis, Fatigue and Fracture, Heat Transfer, Internal Combustion Engines, Manufacturing, and Technology and Society, Torino, Italy, 4–7 July 2006; Volume 4, pp. 557–565. [CrossRef]

9. Baratta, M.; Finesso, R.; Misul, D.; Spessa, E. Comparison between Internal and External EGR Performance on a heavy Duty Diesel Engine by Means of a Refined 1D Fluid-Dynamic Engine Model. *SAE Int. J. Engines* **2015**, *8*, 1977–1992. [CrossRef]

10. D'Ambrosio, S.; Gaia, F.; Iemmolo, D.; Mancarella, A.; Salamone, N.; Vitolo, R.; Hardy, G. Performance and Emission Comparison between a Conventional Euro VI Diesel Engine and an Optimized PCCI Version and Effect of EGR Cooler Fouling on PCCI Combustion. *SAE Tech. Paper Ser.* **2018**. [CrossRef]

11. Finesso, R.; Marello, O.; Misul, D.; Spessa, E.; Violante, M.; Yang, Y.; Hardy, G.; Maier, C. Development and Assessment of Pressure-Based and Model-Based Techniques for the MFB50 Control of a Euro VI 3.0L Diesel Engine. *SAE Int. J. Engines* **2017**, *10*, 1538–1555. [CrossRef]

12. Finesso, R.; Marello, O.; Spessa, E.; Yang, Y.; Hardy, G. Model-Based Control of BMEP and NOx Emissions in a Euro VI 3.0L Diesel Engine. *SAE Int. J. Engines* **2017**, *10*, 2288–2304. [CrossRef]

13. Finesso, R.; Spessa, E.; Yang, Y.; Conte, G.; Merlino, G. *Neural-Network Based Approach for Real-Time Control of BMEP and MFB50 in A Euro 6 Diesel Engine*; SAE Technical Paper 2017-24-0068; SAE International: Warrendale, PA, USA, 2018. [CrossRef]

14. Finesso, R.; Hardy, G.; Mancarella, A.; Marello, O.; Mittica, A.; Spessa, E. Real-Time Simulation of Torque and Nitrogen Oxide Emissions in an 11.0 L Heavy-Duty Diesel Engine for Model-Based Combustion Control. *Energies* **2019**, *12*, 460. [CrossRef]

15. Hu, S.; Wang, H.; Sun, Y.; Wang, Y. Zero-Dimensional Prediction Combustion Modelling of a Turbocharging Diesel Engine. *Trans. CSICE* **2016**, *34*, 311–318. [CrossRef]

16. Catania, A.E.; Finesso, R.; Spessa, E. Predictive zero-dimensional combustion model for DI diesel engine feed-forward control. *Energy Convers. Manag.* **2011**, *52*, 3159–3175. [CrossRef]

17. Finesso, R.; Spessa, E.; Yang, Y. Development and Validation of a Real-Time Model for the Simulation of the Heat Release Rate, In-Cylinder Pressure and Pollutant Emissions in Diesel Engines. *SAE Int. J. Engines* **2016**, *9*, 322–341. [CrossRef]

18. Orthaber, G.C.; Chmela, F.G. *Rate of Heat Release Prediction for Direct Injection Diesel Engines Based on Purely Mixing Controlled Combustion*; SAE Technical Paper 1999-01-0186; SAE International: Warrendale, PA, USA, 2018. [CrossRef]

19. Egnell, R. *A Simple Approach to Studying the Relation between Fuel Rate Heat Release Rate and NO Formation in Diesel Engines*; SAE Technical Paper 1999-01-3548; SAE International: Warrendale, PA, USA, 2018. [CrossRef]

20. Ericson, C.; Westerberg, B. *Modelling Diesel Engine Combustion and NOx Formation for Model Based Control and Simulation of Engine and Exhaust Aftertreatment Systems*; SAE Technical Paper 2006-01-0687; SAE International: Warrendale, PA, USA, 2018. [CrossRef]

21. Finesso, R.; Spessa, E.; Yang, Y.; Alfieri, V.; Conte, G. HRR and MFB50 Estimation in a Euro 6 Diesel Engine by Means of Control-Oriented Predictive Models. *SAE Int. J. Engines* **2015**, *8*, 1055–1068. [CrossRef]

22. Hu, S.; Wang, H.; Yang, C.; Wang, Y. Burnt fraction sensitivity analysis and 0-D modelling of common rail diesel engine using Wiebe function. *Appl. Therm. Eng.* **2017**, *115*, 170–177. [CrossRef]

23. Finesso, R.; Spessa, E.; Yang, Y. Fast estimation of combustion metrics in DI diesel engines for control-oriented applications. *Energy Convers. Manag.* **2016**, *112*, 254–273. [CrossRef]

24. Roy, S.; Ghosh, A.; Das, A.K.; Banerjee, R. A comparative study of GEP and an ANN strategy to model engine performance and emission characteristics of a CRDI assisted single cylinder diesel engine under CNG dual-fuel operation. *J. Nat. Gas Sci. Eng.* **2014**, *21*, 814–828. [CrossRef]

25. Brusca, S.; Lanzafame, R.; Messina, M. *A Combustion Model for ICE by Means of Neural Network*; SAE Technology Paper 2005-01-2110; SAE International: Warrendale, PA, USA, 2005. [CrossRef]

26. Yusri, I.; Majeed, A.A.; Mamat, R.; Ghazali, M.; Awad, O.I.; Azmi, W. A review on the application of response surface method and artificial neural network in engine performance and exhaust emissions characteristics in alternative fuel. *Renew. Sustain. Energy Rev.* **2018**, *90*, 665–686. [CrossRef]

27. Niu, X.; Yang, C.; Wang, H.; Wang, Y. Investigation of ANN and SVM based on limited samples for performance and emissions prediction of a CRDI-assisted marine diesel engine. *Appl. Therm. Eng.* **2017**, *111*, 1353–1364. [CrossRef]

28. Turkson, R.F.; Yan, F.; Ali, M.K.A.; Hu, J. Artificial neural network applications in the calibration of spark-ignition engines: An overview. *Eng. Sci. Technol. Int. J.* **2016**, *19*, 1346–1359. [CrossRef]

29. Yap, W.K.; Karri, V. ANN virtual sensors for emissions prediction and control. *Appl. Energy* **2011**, *88*, 4505–4516. [CrossRef]

30. Shi, Y.; Yu, D.-L.; Tian, Y.; Shi, Y. Air–fuel ratio prediction and NMPC for SI engines with modified Volterra model and RBF network. *Eng. Appl. Artif. Intell.* **2015**, *45*, 313–324. [CrossRef]

31. Kshirsagar, C.M.; Anand, R. Artificial neural network applied forecast on a parametric study of Calophyllum inophyllum methyl ester-diesel engine out responses. *Appl. Energy* **2017**, *189*, 555–567. [CrossRef]

32. Xu, K.; Xie, M.; Tang, L.; Ho, S. Application of neural networks in forecasting engine systems reliability. *Appl. Soft Comput.* **2003**, *2*, 255–268. [CrossRef]

33. Pai, P.S.; Rao, B.S. Artificial Neural Network based prediction of performance and emission characteristics of a variable compression ratio CI engine using WCO as a biodiesel at different injection timings. *Appl. Energy* **2011**, *88*, 2344–2354.

34. Kiani, M.K.D.; Ghobadian, B.; Tavakoli, T.; Nikbakht, A.; Najafi, G. Application of artificial neural networks for the prediction of performance and exhaust emissions in SI engine using ethanol- gasoline blends. *Energy* **2010**, *35*, 65–69. [CrossRef]

35. Javed, S.; Murthy, Y.S.; Baig, R.U.; Rao, D.P. Development of ANN model for prediction of performance and emission characteristics of hydrogen dual fueled diesel engine with Jatropha Methyl Ester biodiesel blends. *J. Nat. Gas Sci. Eng.* **2015**, *26*, 549–557. [CrossRef]

36. Bahri, B.; Shahbakhti, M.; Kannan, K.; Aziz, A.A. Identification of ringing operation for low temperature combustion engines. *Appl. Energy* **2016**, *171*, 142–152. [CrossRef]

37. Lawrence, S.; Giles, C. Overfitting and neural networks: Conjugate gradient and backpropagation. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, Como, Italy, 27 June 2000; Volume 1, pp. 114–119.

38. Finesso, R.; Hardy, G.; Maino, C.; Marello, O.; Spessa, E. A New Control-Oriented Semi-Empirical Approach to Predict Engine-Out NOx Emissions in a Euro VI 3.0 L Diesel Engine. *Energies* **2017**, *10*, 1978. [CrossRef]

39. Heywood, J. *Internal Combustion Engine Fundamentals*; McGraw-Hill Intern: Columbus, OH, USA, 1988.

40. Chen, S.K.; Flynn, P.F. *Development of a Single Cylinder Compression Ignition Research Engine*; SAE Technical SAE Technical Paper 650733; SAE International: Warrendale, PA, USA, 2018. [CrossRef]

41. Catania, A.; Finesso, R.; Spessa, E. *Real-Time Calculation of EGR Rate and Intake Charge Oxygen Concentration for Misfire Detection in Diesel Engines*; SAE Technical SAE Technical Pape r2011-24-0149; SAE International: Warrendale, PA, USA, 2018. [CrossRef]

42. Pearson, K. Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **1894**, *185*, 71–110. [CrossRef]

43. Wilson, L.T. Pearson Product-Moment Correlation. Available online: https://explorable.com/pearson-product-moment-correlation?gid=1586 (accessed on 3 June 2018).

44. Melissa, A.S.; Raghuraj, K.R.; Lakshminarayanan, S. Partial correlation metric based classifier for food product characterization. *J. Food Eng.* **2009**, *90*, 146–152. [CrossRef]

45. Marrelec, G.; Krainik, A.; Duffau, H.; Pélégrini-Issac, M.; Lehéricy, S.; Doyon, J.; Benali, H. Partial correlation for functional brain interactivity investigation in functional MRI. *NeuroImage* **2006**, *32*, 228–237. [CrossRef]

46. Rezaei, J.; Shahbakhti, M.; Bahri, B.; Aziz, A.A. Performance prediction of HCCI engines with oxygenated fuels using artificial neural networks. *Appl. Energy* **2015**, *138*, 460–473. [CrossRef]

47. Najafi, G.; Ghobadian, B.; Tavakoli, T.; Buttsworth, D.; Yusaf, T.; Faizollahnejad, M. Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network. *Appl. Energy* **2009**, *86*, 630–639. [CrossRef]

48. MATLAB Documentation. *Matlab User Guide*; The MathWorks, Inc.: Natick, MA, USA, 2016.

49. Beale, M.H.; Hagan, M.T.; Demuth, H.B. *Neural Network Toolbox™ User's Guide*; The MathWorks, Inc.: Natick, MA, USA, 2018.

50. Ismail, H.M.; Ng, H.K.; Queck, C.W.; Gan, S. Artificial neural networks modelling of engine-out responses for a light-duty diesel engine fuelled with biodiesel blends. *Appl. Energy* **2012**, *92*, 769–777. [CrossRef]

51. Yusaf, T.F.; Buttsworth, D.; Saleh, K.H.; Yousif, B.; Yousif, B. CNG-diesel engine performance and exhaust emission analysis with the aid of artificial neural network. *Appl. Energy* **2010**, *87*, 1661–1669. [CrossRef]

52. Roy, S.; Banerjee, R.; Bose, P.K. Performance and exhaust emissions prediction of a CRDI assisted single cylinder diesel engine coupled with EGR using artificial neural network. *Appl. Energy* **2014**, *119*, 330–340. [CrossRef]

53. Sayin, C.; Ertunc, H.M.; Hosoz, M.; Kilicaslan, I.; Canakci, M. Performance and exhaust emissions of a gasoline engine using artificial neural network. *Appl. Therm. Eng.* **2007**, *27*, 46–54. [CrossRef]

54. Baum, E.B.; Haussler, D. What Size Net Gives Valid Generalization? *Adv. Neural Inf. Process. Syst.* **1989**, *1*, 81–90. [CrossRef]