This article has been accepted for inclusion in a future issue.

Combining Neural Networks and Genetic Algorithms to Predict and Reduce Diesel Engine Emissions

José M. Alonso, Fernando Alvarruiz, José M. Desantes, Leonor Hernández, Vicente Hernández, and Germán Moltó

Abstract—**Diesel engines are fuel efficient which benefits the reduction of CO_2 released to the atmosphere compared with gasoline engines, but still result in negative environmental impact related to their emissions. As new degrees of freedom are created, due to advances in technology, the complicated processes of emission formation are difficult to assess. This paper studies the feasibility of using artificial neural networks (ANNs) in combination with genetic algorithms (GAs) to optimize the diesel engine settings. The objective of the optimization was to find settings that complied with the increasingly stringent emission regulations while also maintaining, or even reducing the fuel consumption. A large database of stationary engine tests, covering a wide range of experimental conditions was used for this analysis. The ANNs were used as a simulation tool, receiving as inputs the engine operating parameters, and producing as outputs the resulting emission levels and fuel consumption. The ANN outputs were then used to evaluate the objective function of the optimization process, which was performed with a GA approach. The combination of ANN and GA for the optimization of two different engine operating conditions was analyzed and important reductions in emissions and fuel consumption were reached, while also keeping the computational times low.**

Index Terms—**Diesel engines, emission regulations, genetic algorithms (GAs), neural networks.**

I. Introduction

Over the last few decades, the industrial and economic expansion of developed countries has involved a sharp increase in vehicle production and transport volume. Environmental concerns such as global warming, greenhouse effects, acid rain, and air pollution problems related to the emissions of carbon dioxide (CO_2), nitrogen oxides (NOx), particulate matter (PM), carbon monoxide (CO), and unburned hydrocarbons (HC), together with the consumption of fossil fuels, combine to create serious problems at a global level [1].

In that sense, the exhaust emission regulatory bodies around the world have been simultaneously reducing the admissible limits on exhaust emissions over the last decades, while the market has also been striving to maintain or even reduce fuel consumption [2].

In order to comply with these regulations, the diesel engine industry has undergone a great technological development in the last few years, creating a high number of new strategies such as electronic control, new injection systems allowing higher pressures, different injection events, etc., [3]–[6]. As a result, the problem of optimizing the engine management in order to simultaneously comply with emission regulations and fuel economy requirements has become a difficult task, especially due to the increased number of degrees of freedom in the engine operating parameters. This optimization process is carried out during the development of a new engine, and is usually known in the mechanical engineering field as *engine calibration*. Although calibrations were in the past based entirely on empirical results, the development in technology has incorporated new model-based techniques [7]–[12].

The formation of the different kinds of pollutants in diesel engines as a result of combustion is a complex process that depends on local variables, and is also highly dependent on the engine settings, influences, and interconnections. Physical and chemical models have been proposed but a general solution has not yet been found [13]–[16]. Computational fluid dynamic codes require detailed unknown local data and also imply large calculation times [17].

Artificial neural networks (ANNs) are an emerging tool of artificial intelligence, which have been shown to be effective in solving a wide range of problems, including many applications to engine modeling [18]. The structure of ANNs enables them to model complex nonlinear multiple problems, which makes them a well-suited method for pollutant modeling. In addition, an ANN can produce fast prediction responses, which represent an important advantage in comparison with alternative modeling techniques, such as physical and chemical models.

As a first objective in this paper, ANN modeling was used to predict diesel NOx, PM, CO, HC exhaust emissions, and brake specific fuel consumption (BSFC) in terms of engine operating parameters. The working operation of the engine is defined by the combination of the engine operating parameters values, which in turn determine the engine exhaust emission levels and fuel consumption. The number of operating parameters has increased in the last few years due to the engine development, making it difficult to directly combine them to comply with different constraints.

The second main objective of the study was to define an optimization process that, using the ANN predictions, could find combinations of operating parameters that simultaneously minimize fuel consumption, while also keeping the overall emissions

Manuscript received August 6, 2004; revised July 19, 2005 and November 23, 2005.This work was supported in part by the Plan Valenciano de Investigación Científica, Desarrollo Tecnológico e Innovación from the Conselleria de Empresa, Universidad y Ciencia of the Generalitat Valenciana, Spain, under Project CITDIB/2002/34.

J. M. Alonso, F. Alvarruiz, V. Hernández, and G. Moltó are with the High Performance Networking and Computing Group, Universidad Politécnica de Valencia, Valencia 42022, Spain (e-mail: jmalonso@dsic.upv.es; fbermejo@dsic.upv.es; vhernand@dsic.upv.es; gmolto@dsic.upv.es).

J. M. Desantes is with the CMT-Thermal Engines Group, Universidad Politécnica de Valencia, Valencia 46022, Spain (e-mail: jmdesant@mot.upv.es).

L. Hernández is with the Departamento de Ingeniería Mecánica y Construcción, Universitat Jaume I, Castellón de la Plana 12071, Spain (e-mail: lhernand@emc.uji.es).

Digital Object Identifier 10.1109/TEVC.2006.876364

under certain limits. The correct definition of the objective function, exactly reproducing the requirements, was considered an important step in the optimization process.

Classical methods for optimization, based on numerical techniques, have been applied to the optimization of diesel engines in different publications. In [19] and [20], a simple gradient method is used, where only one parameter is changed in each iteration. In [9], [21], and [22], a steepest descent method is followed, while [23] employs a sequential quadratic programming approach. However, numerical methods for optimization present some limitations, such as the difficulty to escape from local minima, and therefore the dependence of the performance on the initial value chosen. Moreover, these methods are only applicable to continuous differentiable functions.

As an alternative to numerical optimization methods, genetic algorithms (GAs) are less likely to get trapped in local minima, and are not restricted by continuity or differentiability requirements on the objective function. As a drawback, the computational time may be larger than that of numerical methods. Taking this into account, GAs were used as the optimization technique in the work described in this paper.

The overall objective raised in this study integrates a problem of complex modeling together with an optimization process characterized by a high number of variables to be simultaneously modified and a high number of constraints to be satisfied. The scope of this paper is to evaluate the suitability of combining both ANN and GA techniques in this difficult and important task.

II. Experimental SetUp

The engine used to conduct the experiments was a light duty (LD) four-cylinder direct injection diesel engine (turbocharged and aftercooled) that was equipped with a common rail system. The engine was fully instrumented, and connected to a data acquisition system, so that many different operating parameters could be measured and precisely controlled.

The electronic control unit (ECU) controlled all the engine electronic equipment. Its hardware (sensors, actuators, and regulators) monitored different variables and its software controlled the engine according to the registered signals and the integrated cartographies. The cartographies are a set of different lookup tables that, for each couple of measured engine speed and fuel mass injected variables (that define an *engine condition*), impose the values of each of the engine operating parameters. Each engine includes its own original cartography, which has been previously defined by the engine manufacturer and that is considered as the nominal combination of operating parameters for each engine condition. The communication between the ECU and the test cell system made it possible to modify the original cartographies and to fix the desired operating conditions to be established.

The exhaust stream of the engine was diverted to a Horiba 7100 analyzer from which NOx, CO, HC emissions were measured. Smoke emissions were evaluated with an Avl 415 Variable Sampling Smoke Meter, which estimates the opacity of a filter through which the exhaust gases have been drawn. These measurements are reported as filter smoke number (FSN). The experimental FSN values can be transformed into dry soot mass emissions by means of the correlation proposed by Christian *et al.* [24]. The emission regulations related to PM require the interaction of the exhaust gases with the ambient environment to be taken into account, as the dry soot mass will adsorb hydrocarbon compounds and small amounts of sulfates, metals and ash in this process, increasing the mass of the final PM obtained [25]. Considering the large number of experimental tests necessary for this study and that the application of the prediction and optimization method presented in this work remains valid, the direct FSN measurement was selected.

TABLE I
Range of Variation Covered by the Different Operating Parameters Within the Experimental Study

Operating parameter	Units	Range of Variation
n	rpm	[750, 4000]
Mf	kg/h	[0.38, 24.34]
Ma	kg/h	[20.2, 520.3]
EGR	%	[0, 55.4]
IP	bar	[200, 1350]
SOIm	°BTDC	[-4.85, 21]
SOIp	°BTDC	[0, 49]
Tint	°C	[27.4, 130.2]
Tw	°C	[49.4, 96.9]

III. Experimental Database

The selection of the experimental test cases to be measured for an empirical model such as ANN is of great importance, due to the fact that the database determines the applicability, the range of valid predictions, and also the quality of the predictions. In this sense, the experimental database must be broad, present low noise levels, and also be representative of the problem to be modeled.

The experimental database in this study was developed under stationary conditions. A total number of 19 engine conditions were selected in order to cover a wide range of situations in the engine working area. These 19 engine conditions were selected from different areas: full load, conditions covering the European transient cycle [26], and other areas where the emissions are less restricted. This classification has been also important in the definition of the allowed limits for emissions and fuel consumption in the optimization procedure.

As mentioned in the previous section, each engine condition is defined by the values of two operating parameters, namely, engine speed (n) and fuel mass injected (Mf). There are seven additional operating parameters, the values of which are determined, for each engine condition by the engine cartography. These additional operating parameters are: air mass (Ma), exhaust gas recirculation (EGR), injection pressure (IP), start of pilot injection (SOIp), start of main injection (SOIm), intake temperature (Tint), and water temperature (Tw). Table I shows the complete list of operating parameters.

The experimental database was generated by taking each of the 19 engine conditions, and producing different variations for values of the seven additional operating parameters. Starting from the nominal values of these parameters for each engine condition, independent variations of each parameter at three or four levels were established. In this way, a total number of

This article has been accepted for inclusion in a future issue.

440 test cases, representative of the LD engine and close to the manufacturer cartography, were obtained.

All nine operating parameters were used as the inputs to the ANN, considering that the empirical model is only able to predict trends in emissions and fuel consumption that have been integrated into the test cases used as training patterns. As predictions and optimization were integrated together in this study, the definition of the inputs in the ANN also determined the variables to be modified in the optimization process.

Another important feature of the experimental database from the point of view of ANN and optimization development was the range of variations of the operating parameters, as the validity of the predictions and the range of variation in the optimization were restricted for those values. The ranges of variation covered by the different operating parameters within this experimental study are presented in Table I.

The quality of the ANN predictions, due to the empirical nature of the model, is highly dependent on the experimental noise in the measurements. The noise in the experimental data comes from different sources: first of all, it is influenced by the measuring devices accuracy (which can be obtained from the manufacturer), secondly, it depends on the experimental measuring methodology and mainly on the engine and measuring devices stability with time.

The Horiba manufacturer ensures a precision of $\pm 2\%$ for the emission measurements. Later, experimental studies performed in single-cylinder engine showed poorer accuracy, lowering the value to $\pm 4\%$ [27]. The resolution for the particulate estimation device (Avl 415) assures 0.05 FSN of repetitivity for values within 0.5 and 6 [28]. The BSFC was derived from two measurements: fuel mass (obtained with an Avl S733 gravimetric balance with $\pm 0.2\%$ of accuracy [27]) and torque (measured with load cell Zöllner A-350/AE with an accuracy of ± 2 Nm [27]).

The measuring protocol and methodology, including the postprocessing of the measured data, was defined to reduce to the minimum the sources of error obtained during the experimental study. A global maintenance test was carried out for most of the devices employed for the variables used as inputs/outputs to the ANN. Daily calibration and purging between test cases were performed in the emission equipment. Large stabilization times were considered before measuring each stationary test, in which the variables were averaged during 30 seconds. As the measurement of the large experimental test matrix extended for months, all the emissions as well as the fuel consumption were corrected taking into account atmospheric conditions, as described in the European legislation [26]. All the experimental data were additionally put under a quality control, in which anomalous data were detected and the measurements repeated when necessary.

The experimental dispersion observed in the dependent variables integrated the accuracy of the devices, the measuring methodology considered, and the deviation of the engine and measuring devices equipment with time. This dispersion was evaluated comparing the same reference test case repeated every measuring day. The reference test case was a nominal operating condition from the manufacturer cartography that belonged to the area where the EGR is present, i.e., the European transient cycle area [26]. Up to 40 repetitions of the reference test case were analyzed and repetitivity of the different emissions and the fuel consumption studied.

TABLE II
VARIANCE COEFFICIENT EVALUATED FOR THE VARIABLES IN THE REFERENCE TEST CASE

Variable	Units	VarCoeff[%]
NOx	g/h	4.08
PM	FSN	6.41
CO	g/h	7.03
HC	g/h	20.56
BSFC	g/kWh	1.23

Equations (1) and (2) define the variance coefficient (VarCoeff), which was chosen as an estimation of the stability of the measurements

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (\mathrm{VAR}_{(i)} - \overline{\mathrm{VAR}})^2}{N-1}} \tag{1}$$

$$\mathrm{VarCoeff}[\%] = \frac{\sigma}{\overline{\mathrm{VAR}}} \cdot 100 \tag{2}$$

where σ is the standard deviation, N is the total number of test cases evaluated, VAR is the variable for which the repetitivity is being evaluated, and $\overline{\mathrm{VAR}}$ is the mean value of that variable. The results obtained for the reference test case are presented in Table II.

As observed in Table II, the smallest variance was obtained for BSFC, which was by far the most stable output to be predicted. Within the emissions, the NOx presented the highest stability, which was followed with similar results in variance coefficients by PM and CO. Measuring problems, such as uncontrolled temperatures influencing HC condensation, may have been responsible for the low HC emission stability.

It was important to analyze these results, as the quality of the ANN predictions has been highly influenced by the repetitivity of each of the variables to be predicted. According to Table II, and assuming that ANNs are empirical models, BSFC was expected to show the most accurate ANN predictions, while HC emissions were expected to present the poorest results.

IV. ANN DEVELOPMENT AND PREDICTIONS

One of the main ANN's advantages is their ability to model complex nonlinear relationships between multiple input variables and the required outputs [29]. ANNs are also able to self-identify these complicated relationships, so that the ANN designer does not need to make any assumption about the equations governing the process to be reproduced [30]. In the case of emission modeling in a diesel engine, where the complex and local phenomena are not yet completely understood and are difficult to model, this fact implies an enormous simplification. Another important advantage of the ANN approach is its fast response, which enables it to be included in more complex procedures, such as optimization applications [31]. More information about the general features and working principles of ANNs can be found in [29]–[31] .

The main ANN modeling objective in this study was to be able to reproduce diesel exhaust emissions (NOx, PM, CO, and HC) together with the fuel consumption (BSFC), starting from the engine setting information obtained from the experimental test database. In this work, both ANN implementation and training was developed using the neural network toolbox of Matlab [32].

Different ANNs were built (one for each output), rather than using one large ANN including all the output variables. This strategy allowed for better adjustment of the ANN for each specific problem, including the optimization of the architecture for each output. It was also pointed out by some authors [33] that including many outputs in the same ANN causes the least accurate variable to drag down the prediction of the other outputs.

The ANN architecture used in this work was based on the multilayer perceptron (MLP) with only one hidden layer, as including more layers could increase the risk of finding local minima in the error surface, thus degrading the ANN prediction accuracy [33]. The activation functions chosen were sigmoids for the hidden neurons as they allowed nonlinear relationships between inputs and outputs, and linear activation function for the output neurons. Both types of functions complied with the differentiability conditions imposed by the training algorithm.

As described next, different steps were followed in the development of the ANN-based approach: training procedure, data transformation or scaling, and ANN specific architecture definition. More details about the procedure for the selection of ANN architecture and training can be found in previous works of the authors [34], [35].

A. Training Procedure

One of the most relevant aspects of a neural network is its ability to generalize, that is, to predict cases that are not included in the training set. An important concern in ANN training is the risk of overfitting, which occurs when good predictions are achieved for the training data set, but a significantly lower success rate is achieved when predicting the test data set [31].

Due to the difficulty and the cost associated to taking measurements in a diesel engine, the available measurement database was reduced to contain only 440 patterns. In order to compensate for the possible problems of overfitting associated with the small data set available, several countermeasures were applied during the ANN training, as described in the following paragraphs.

Early stopping of the ANN training procedure can be quite effective when dealing with this problem. It has been demonstrated [36] that there is an optimal stopping time which guarantees generalization prior to convergence of the network on the training set. Nevertheless, the fact that an optimal stopping time exists does not imply that a stopped training method will be able to find it.

Taking this into account, the training methodology employed in this study combines cross-validation with early stopping, in order to avoid overfitting, while being able to use all the patterns. The method consists of three stages, as described next.

The first stage randomly divided all the patterns into a training set (90% of the patterns) and a validation set (10% remaining patterns). The neural network was trained with a traditional cross-validation scheme, and the training set error corresponding to the minimum validation error was stored. This process was repeated for 50 runs, randomly dividing the patterns into the training and validation sets in each run. Next, the training errors were averaged, obtaining a mean-square training error (MSTR) which was the stopping error to be used in the second and the third phases of the training procedure.

The second stage was used as an intermediate step to evaluate the effects of the number of neurons for the ANN predictions quality (see Section IV-C). At this second stage, the ANN was initialized five times, and for each of them, ten different iterations divided the patterns into 90% for training and 10% for ANN prediction evaluation. Each ANN was only trained until the averaged MSTR obtained in the previous stage was reached. For each of the five iterations, the quality of the ANN predictions was evaluated for a matrix which only contained data not seen by the ANN during the training. The different iterations allowed the evaluated matrix to contain as many data as the total experimental patterns, which increased the statistical validity of the results.

In the third stage, the neural network was trained, using all the patterns, until the training error reached the MSTR calculated in the first phase. The training was repeated 20 times, which resulted in 20 different ANNs due to the randomness of the initial ANN weights. As the optimization process is dependent on the specific ANN considered, the 20 trained ANN for each output were assembled in an ANN committee, in order to compensate for the differences in initialization and local minima in each ANN.

For all the stages, the Levenberg–Marquardt training algorithm was employed [37], [38]. The principal idea of the Levenberg–Marquardt algorithm is to maintain the second-order approach of the ANN error in order to speed up the training (like in the Newton method), but at the same time evaluating the Hessian matrix using an approximation (like the quasi-Newton methods).

B. Data Transformation

Large differences in the absolute values of the different inputs can complicate the ANN learning process. To overcome this potential problem, all the inputs and outputs were previously normalized so that only relative changes in the variables were learned by the network. In this study, the inputs and targets were normalized between their minimum and maximum values. As the ANN predictions were to be included in the performance function for the optimization algorithm, a higher accuracy for lower levels of predictions was necessary. One possibility could have been to change the error to be minimized in the training algorithm from absolute to relative. Instead of changing the training algorithm, the same objective was obtained by applying a logarithmic scale to the input and output variables [23]. Both normalization and scale transformation were applied before presenting the training patterns to the ANN, and then converted again to obtain the results in their original levels.

C. ANN Architecture

The selection of the exact number of neurons in the hidden layer for each output was based on the accuracy of results, following a trial and error procedure. This adjustment allowed the complexity of the modeled phenomenon to be taken into account. The number of neurons evaluated ranged from 1 to 30,

This article has been accepted for inclusion in a future issue.

TABLE III
OPTIMUM NUMBER OF NEURONS FOR THE ANN OF EACH OUTPUT

Output	Units	Neurons
NOx	g/h	17
PM	FSN	17
CO	g/h	24
HC	g/h	27
BSFC	g/kWh	9

which avoided large number of weights to be adjusted in the training process for the available experimental patterns. The different architectures derived for each output were trained following the first and second training stages described above. In the trial and error procedure, the accuracy of the ANN predictions results was compared and the best architecture for each output is presented in Table III.

From a mathematical point of view, a larger number of neurons implies more complex ANN relationships, which is required to model more complex processes. The differences in the optimal number of neurons for each output were in accordance with the diesel combustion knowledge. On the one hand, it is well known that the emission formation in diesel engines is much more complicated than the processes governing the fuel consumption, which is consistent with the smaller number of neurons required for the BSFC ANN predictions. On the other hand, the increasing number of neurons required for the optimized architectures for the different emissions is in accordance with the increasing variance coefficient shown in Table II. This fact may imply that the variables with less stable measurements need more complicated mathematical equations in the prediction ANN models.

D. ANN Prediction Quality

Once the different stages of the training process and the ANN architectures had been determined, and before the optimization procedure was developed, it was important to estimate the ANN prediction qualities.

Fig. 1 presents the predictions obtained for the emissions and the fuel consumption when averaging all the neural networks that integrate the ANN committee. As a result of the incorporation of the logarithmic scale, in most of the graphs, higher accuracy can be observed for low levels of predictions.

In order to quantify the quality of the predictions obtained for each output, the average relative error [as presented in (3)] is evaluated in each case

$$\mathrm{ERROR_{rel}} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{X_{\mathrm{ANN}}(i) - X_{\mathrm{expected}}(i)}{X_{\mathrm{expected}}(i)}\right| \quad (3)$$

where X_{ANN} is the ANN predicted value, X_{expected} is the expected value and N is the total number of experimental cases to be evaluated.

Fig. 2 analyzes the differences in relative error among the 20 ANN that form the committee for each output. Different conclusions can be derived from both figures. The best predictions, together with the lowest dispersion among ANN, were obtained for the BSFC case. NOx were the emissions that presented more accurate results. For the case of CO and HC, poorer predictions were achieved and the dispersion within the ANN committee increased. A notable decrease in the quality of the PM predictions could be observed.

As remarked previously, the complexity of the phenomena to be reproduced, together with the experimental problems analyzed for each output (Table II) can explain these results.

V. OPTIMIZATION PROCEDURE

It should be noted that the optimization of the whole range of engine operating conditions was not feasible due to the large number of parameters involved. Instead, this process was simplified by optimizing only a small number of particular engine conditions (defined by engine speed and total fuel mass injected). Each of these optimization subproblems was in charge of selecting the combination of operating parameters which produce the lowest fuel consumption subject to certain emission constraints.

The minimization process was subject to two different types of constraints. On the one hand, the operating parameters were bounded by a minimum and maximum value for each engine condition. On the other hand, different criteria were established for the emission constraints depending on the engine area.

With respect to the emission constraints, the main priority for the operating conditions within the European transient cycle area is the NOx emissions, so that a reduction of 20% with respect to the measured nominal value was imposed for this emission. In the case of full load operating conditions, the selected criterion was to reduce the measured FSN obtained for the nominal test case by 10%. In both cases, the maximum limits for the remaining emissions were set at their cartography nominal levels.

GAs have traditionally been employed for optimization problems with objective functions that do not present continuity or differentiability properties [39]. In these algorithms, the search for a global minimum is performed through the application of reproduction operators and "survival of the fittest" strategies. Moreover, the mutation operator introduces the possibility of exploring the whole search space, an interesting ability that reduces the risk of finding a local minimum. However, a GA does not guarantee that the global minimum is found.

A. GA Population

Considering that the optimization process was performed for each engine condition, which was indicated by a combination of engine speed (n) and fuel mass (Mf), each individual of the genetic population corresponded to a combination of values for the remaining seven operating parameters, namely: air mass (Ma), exhaust gas recirculation (EGR), injection pressure (IP), start of pilot injection (SOIp), start of main injection (SOIm), intake temperature (Tint), and water temperature (Tw).

The chromosome of each individual was represented as a vector of seven real-valued elements, where each element expressed directly the value of a parameter, using the units shown in Table I. There have been studies confirming that a real-valued representation is more efficient and produces better solutions than binary-coded GAs [40] [41]. It is also argued that a real-valued representation offers enhanced precision and more consistent results between different replications.

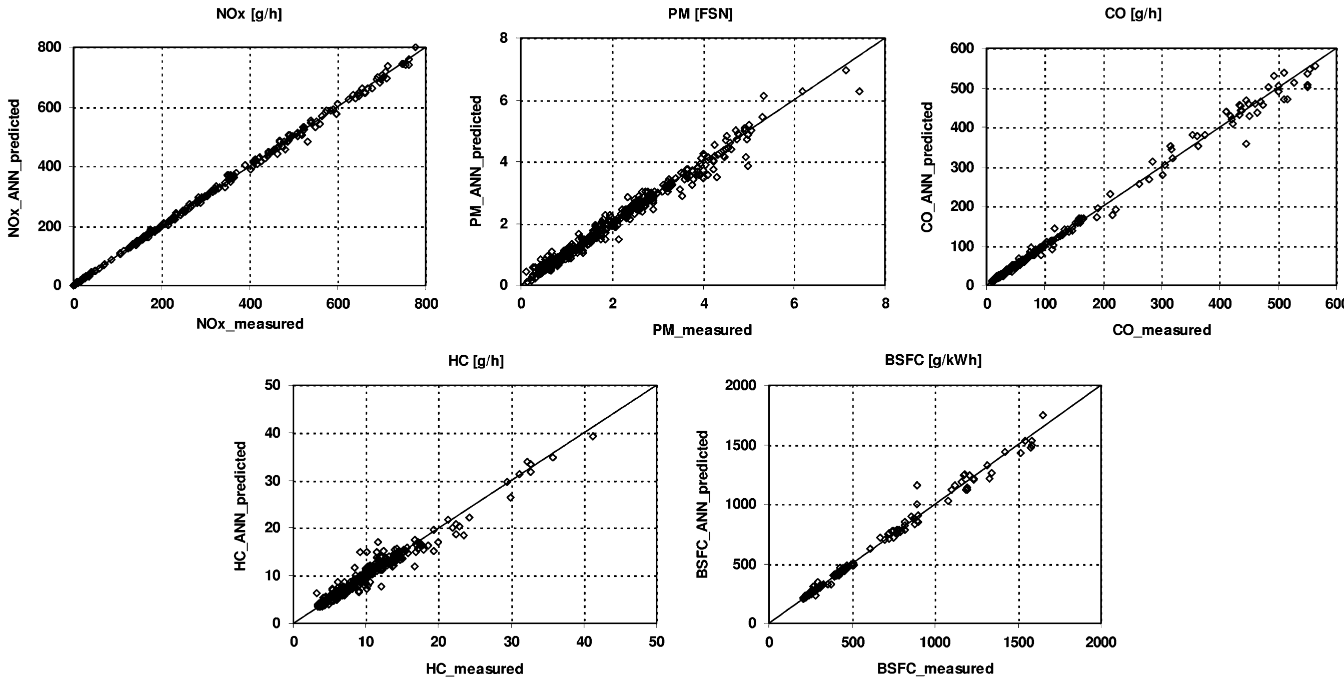


Fig. 1. Predictions for each output averaging the results of a 20 ANN committee.

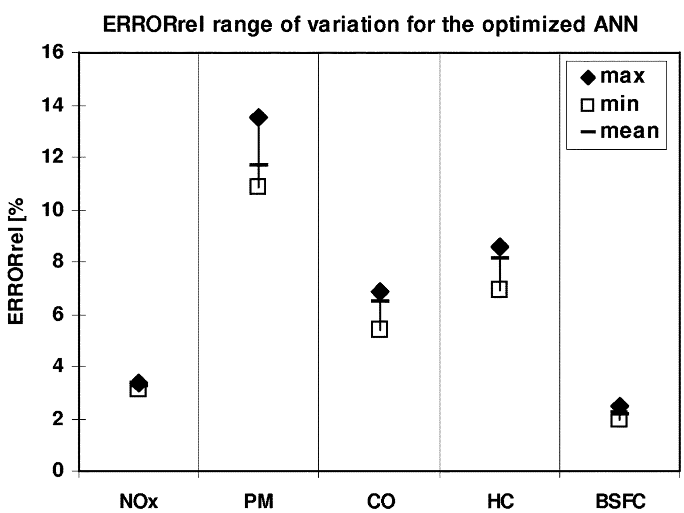


Fig. 2. Differences in the relative error among the 20 ANN committee for each output.

Therefore, the parameters of an individual were represented by real values within lower and upper bounds, these bounds being specific for each engine condition.

First of all, the genetic population was created. Even though there are approaches using microgenetic algorithms (μGA) with a very small population [42], it was found that, for the problem described in this paper, a population between 100 and 150 individuals produced higher quality solutions than those obtained with smaller populations. The usage of different subpopulations (from 2 to 8), with 150 individuals equally distributed among them, was initially considered. This technique has been shown to improve the quality of the results compared with the usage of a single population in other problems [43] [44]. However, no noticeable changes in the solution for this particular problem were obtained. Therefore, it was decided to carry out the experiments using a single population with 150 individuals.

B. GA Objective Function

The definition of the objective function in a GA is crucial for the success of the optimization process, because it assesses the fitness of an individual. The value of the objective function represents a measure of the quality of a member of the population, and thus, the searching process is guided by the results of the objective function.

Taking into account that the main target of the optimization process was to reduce the fuel consumption (BSFC), subject to the operating parameter constraints and the emission constraints, the objective function (OF) was defined as follows:

$$\mathrm{OF}(X) = \mathrm{BSFC}(X) + \sum_{i=1}^{4} \exp\left[K_i \cdot \left(\frac{E(X)_i - \mathrm{MAX}_i}{\mathrm{MAX}_i}\right)\right] \tag{4}$$

where $E(X)_i$ represents the value of emission i for the individual X, MAX_i represents the maximum value allowed for the emission, and K_i is a penalty factor. The value of K_i quantifies the penalty imposed to an individual that does not accomplish the restriction about the contaminant emission i. The larger the value of K_i, the higher the penalty, which results in a substantial increase of the objective function value, and reduces the probability of the individual to be chosen for reproduction. The emissions and the BSFC terms are obtained for each individual (combination of operating parameters) by making use of the ANN committee previously described.

The definition of the OF represented the main objectives of the optimization: first of all, to minimize fuel consumption [first term in (4)] and also to keep the emissions below certain

This article has been accepted for inclusion in a future issue.

limits, controlled by the exponential terms, which significantly increase when the upper bounds are exceeded.

In order to weigh the different constraints in emissions, so that a final value of the OF was representative of the global compliance of the restrictions, an iterative process for the adjustment of the K_i parameters in (4) was developed. The initial values for the K_i parameters for all exponential terms were set to one. The GA was executed until convergence and, if a certain emission did not comply with the maximum limit, its corresponding K_i value was increased. The process was repeated until all the constraints were satisfied.

C. GA Selection and Replacement

The objective function measures the quality of the individuals with respect to the problem domain. In our case, the fittest individuals had the lowest objective function values.

The *selection phase* determines the individuals that are going to produce the offspring. This includes a first step of fitness assignment, where each individual of the population receives a fitness value proportional to its probability of being selected for reproduction. As the fitness value influences the offspring for the next generation, caution should be taken so that highly fit individuals do not dominate the reproduction to avoid a rapid convergence to a possible suboptimal solution. The second step is to select individuals for reproduction, based on their fitness values. *Stochastic Universal Sampling* (SUS) [45] was employed for this selection. With this method, the individuals are mapped one-to-one into contiguous segments of a line, where the size of the segment of each individual corresponds to its fitness value. Then, as many equally spaced pointers are placed along the line as individuals are to be selected, and the number of pointers to an individual's segment indicates how many times that individual is selected for reproduction.

Crossover between individuals was performed using *Intermediate Recombination* [46] [47], a method that produces new offspring around and between the values of the parents. In this method, which is applied to real-valued representations of the individuals, the value of each gene O in a new individual is obtained following the rule:

$$O = P_1 + \alpha \cdot (P_2 - P_1) \tag{5}$$

where P_1 and P_2 are the values of the same gene in the parents, and α is a scaling factor chosen randomly for each gene within an interval (typically $[-0.25, 1.25]$). Another method considered for crossover was *Line Recombination*, which is similar to Intermediate Recombination, with the difference that the same value of α is used for all the genes in the same descendant. However, we found that the choice of any of these two methods did not affect, on average, to the quality of the solution.

Once the new offspring has been produced, a reinsertion policy specifies the insertion of the new individuals into the population, as well as the replacement of the existing individuals. Different aspects need to be taken into account here, such as which is the number of newly created individuals with respect to the population size used, which proportion of those new individuals will be actually inserted in the population, and which old individuals are to be replaced by the new ones. After several experiments, the number of individuals produced in each generation was fixed to be equal to 80% of the population size, only 90% of which were inserted in the population, the remaining 10% being discarded. The corresponding old individuals to be replaced were those found to be least fit (lowest fitness values). The size of the population remained constant through all the generations, as experimental studies showed that an increase in the population size did not provide better solutions in terms of fuel consumption. This enabled to keep the computational times low.

A high mutation rate was used for the real-valued GA implemented in this study. It was shown in [48] that high mutation rates combined together with nonbinary representation, used in complex combinatorial optimization problems, gave place to significantly better solutions than using a typical mutation rate between 0.1% and 1% [49]. Applied to this particular problem, it was found that, on average, high mutation rates allowed the GA to find individuals with lower objective function value. In particular, a mutation rate equal to the inverse of the number of variables in the individuals of the population (i.e., the number of operating parameters defining an individual) was used, as suggested in [49].

No stopping condition was imposed on the algorithm, this decision being left to the user. Thus, the user was in charge of observing periodically the evolution of the population through several graphical results that summarized the optimization process.

VI. Optimization Results

In order to assess the performance of the optimization process, two different engine conditions were optimized out of the total 19 considered. Each engine condition belonged to a different area, namely European transient cycle and full load, thus covering a wide range of engine conditions where the emissions have important restrictions. All the executions were performed on an Intel Pentium Xeon running at 2.0 GHz with 2 GBytes of RAM.

Fig. 3 shows the evolution of the objective function value during the optimization process for the two considered operating conditions. In each graph of Fig. 3, the evolution of two objective function values is shown: the mean value of the objective function averaged for all the individuals (dotted line) and the value of the objective function for the best individual in each generation (solid line).

The continuous approach along the consecutive generations of both objective function values represented in the graphs, suggests that the differences among the individuals of the population get progressively smaller. A stable objective function value is reached around the 140th generation for Fig. 3(a) and around the generation number 100 for Fig. 3(b). No further improvement in the objective function is observed in the next generations, indicating that no better individual is found in future iterations and, thus, that the algorithm has converged. The corresponding operating parameters obtained at the convergence stage were the best combination found which allows a minimized fuel consumption, while also satisfying the imposed emission constraints.

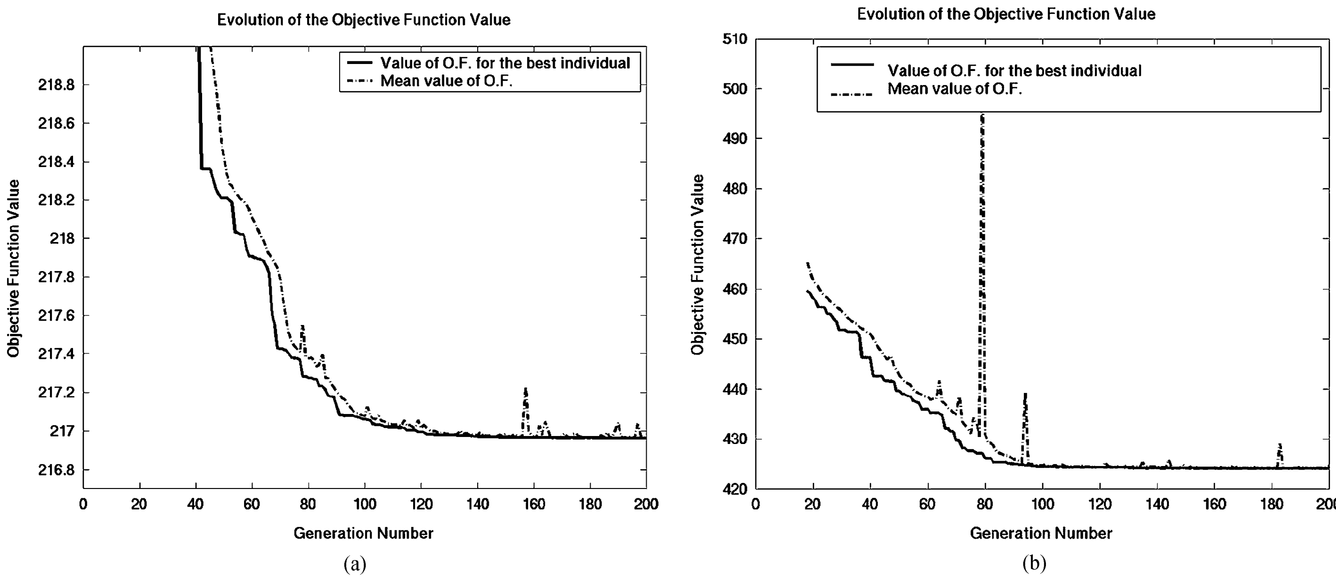


Fig. 3. (a) Evolution of the objective function value during successive generations for the full load and (b) the European transient cycle engine conditions.

TABLE IV
COMPARISON OF NOMINAL VALUES, LIMITS, AND GA OPTIMIZED VALUES FOR EMISSIONS AND BSFC IN THE FULL LOAD OPERATING CONDITION

Variable	Nominal	Limit	Best Found	% Reduction
NOx	695.36	695.36	665.11	4.4%
CO	92.67	92.67	68.72	25.8%
HC	10.14	10.14	6.47	36.2%
FSN	1.34	1.20	1.20	10.1%
BSFC	222.04	---	215.56	2.9%

TABLE V
COMPARISON OF NOMINAL VALUES, LIMITS, AND GA OPTIMIZED VALUES FOR EMISSIONS AND BSFC IN THE EUROPEAN TRANSIENT CYCLE CONDITION

Variable	Nominal	Limit	Best Found	% Reduction
NOx	5.52	4.41	4.40	20.3%
CO	47.73	47.73	44.29	7.2%
HC	10.28	10.28	10.25	0.3%
FSN	1.65	1.65	1.21	26.7%
BSFC	449.43	---	422.03	6.1%

The graphs in Fig. 3 do not show the objective function values for the first 20 generations, due to the fact that the random initialization of the individuals in the GA causes that many individuals do not satisfy the constraints during the first generations, implying an important increase in the mean objective function values. Thus, showing the value of the objective function for the first generations would have altered the scale completely, hiding the evolution of the values for later generations. The peaks in the second graph are due to mutations leading to individuals that do not satisfy a constraint, thus obtaining a large value of the objective function, and altering the mean value for all the population.

The whole optimization process required around a minute and convergence was achieved in approximately 140 generations for both the case of European transient cycle and the full load operating condition.

Once GA converged, the values for the emissions and fuel consumption (BSFC) corresponding to the obtained OF minimum were analyzed, as shown in Tables IV and V. The first of these tables presents the results for the full load operating condition, where column *Nominal* contains the values of exhaust emissions and fuel consumption obtained with the nominal values of the operating parameters, as defined by the original engine cartography. Column *Limit* shows the maximum acceptable limits imposed on each contaminant. According to what was stated in Section V, the limit of FSN is fixed to a value 10% lower than the corresponding nominal value. Column *Best Found* contains the values of contaminants and fuel consumption obtained as a result of the optimization process. Finally, the last column shows the reduction achieved with respect to the nominal values, in each emission and in fuel consumption. It can be seen that all emissions obtained after the optimization satisfied the limits imposed, particularly FSN, and reductions of up to 36.2% in the case of HC were achieved. A decrease in fuel consumption of 2.9% was also obtained.

Table V presents the same results for the case of the European transient cycle condition. In this case, the priority was to reduce the NOx emissions, as pointed out in Section V, thus the imposed limit was 20% lower than the nominal value. As in the previous operating condition, the operating parameters found by the optimization process satisfied all the emission constraints imposed, and particularly that of NOx. Additionally, important reductions were achieved both in the emission levels (up to 26.7% for FSN) and in fuel consumption (6.1%).

VII. CONCLUSION

The optimization approach described here uses ANNs as the basis for the evaluation of the objective function that is used within a GA optimization process. This is applied to solve the problem of reducing the fuel consumption, while keeping the emissions under certain maximum limits in a diesel engine under stationary conditions. An application of a similar approach to the case of transient engine conditions is currently under study.

The ANNs have been employed as predicting tools for BSFC, NOx, PM, CO, and HC. Different quality for the predictions was obtained for the outputs, which was strongly related to the complexity of the phenomena to be modeled and the experimental repetitivity of each variable. A committee of 20 ANNs for each output was produced and integrated within the optimization procedure in order to average the predictions of the different specific ANN trainings.

The GA optimization was performed for two engine operating conditions, where the limits in emissions were established according to the values in the nominal engine cartography and the main restrictions in each case. A methodology for the objective function adjustment was proposed, including special penalties for the more restrictive emission limits.

For the two operating conditions studied, the convergence of the optimization process was fast (less than a minute) and all the constraints were satisfied in the resultant solutions. Important improvements with respect to the nominal manufacturer engine cartographies were observed both in terms of emissions (up to 36% reduction for HC) and fuel consumption (between 3% and 6% reduction depending on the engine condition). This has been achieved in a problem of high dimensions (seven operating parameters) and high number of constraints (four emissions). Finally, it should be noticed that the engine emission and consumption improvement was reached without the incorporation of any new technological device, just combining in a better way the operating parameters that are currently considered by the manufacturer.

Additional efforts and future work will focus on experimental confirmation of optimized settings and also in implementation of the presented procedure for transient operation of the engine.

REFERENCES

[1] K. Mori, "Worldwide trends in heavy-duty diesel engine exhaust emission legislation and compliance technologies," SAE paper 970753, 1997.

[2] F. X. Moser, T. Sams, and W. Cartellieri, "Impact of future exhaust gas emission legislation on the heavy duty truck engine," SAE paper 2001-01-0186, 2001.

[3] C. J. Morey and J. Mark, "Diesel passenger vehicles—Can they meet air quality needs and climate change goals?," SAE paper 2000-01-1599, 2000.

[4] J. Benajes, S. Molina, and J. M. García, "Influence of pre- and post-injection on the performance and pollutant emissions in HD diesel engine," SAE paper 2001-01-0526, 2001.

[5] P. L. Herzog, L. Buerguer, E. Winklhofer, P. Zelenka, and W. Caterllieri, "NOx reduction strategies for diesel engines," SAE paper 920470, 1992.

[6] C. Havenith, J. R. Needham, A. J. Nicol, and C. H. Such, "Low emission heavy duty diesel engine for Europe," SAE paper 932959, 1993.

[7] S. P. Edwards, D. M. Grove, and H. P. Wynn, *Statistics for engine optimization*. London, U.K.: Professional Engineering Publishing Limited, 2000, ISBN 1-86058-201-X.

[8] D. T. Montgomery and R. D. Reitz, "Optimization of heavy duty diesel engine operating parameters using a response surface method," SAE paper 2000-01-1962, 2000.

[9] P. K. Senecal and D. T. Montgomery, "A methodology for engine design using multi-dimensional modeling and genetic algorithms with validation through experiments," *Int. J. Engine Research*, vol. 1, no. 3, pp. 229–248, 2000.

[10] H. Yun and R. D. Reitz, "An experimental study on emissions optimization using micro-genetic algorithms in a HSDI diesel engine," SAE paper 2003-01-0347, 2003.

[11] T. Hiroyasu, M. Miki, J. Kamiura, S. Watanabe, and H. Hiroyasu, "Multi-objective optimization of diesel engine emissions and fuel economy using genetic algorithms and phenomenological model," SAE paper 2001-01-2778, 2001.

[12] M. Guerrier and P. Cawsey, "The development of model based methodologies for gasoline IC engine calibration," SAE paper 2004-01-1466, 2004.

[13] H. Hiroyasu and T. Kadota, "Models for combustion and formation of nitric oxide and soot in direct injection engines," SAE paper 760129, 1976.

[14] J. M. Desantes, M. Lapuerta, and J. M. Salavert, "Study on independent effects of diesel engine operating conditions on nitric oxide formation and emissions trough schematical combustion simulation," in *Proc. Instn. Mech. Engineers, Part D, J. Automobile Eng.*, 1996, vol. 210, pp. 71–80.

[15] F. Payri, J. M. Desantes, and M. A. Lapuerta, "Phenomenological model for calculating D.I. diesel engine performance and emissions," in *Proc. Int. Congr. Calcul Scientifique et Automobile (SIA)*, Paris, France, 1990.

[16] S.C. Kyriatides, J. C. Dent, and P. S. Mehta, "Phenomenological diesel combustion model including smoke and NO emission," SAE paper 860330, 1986.

[17] R.D. Reitz and R. P. Hessel, "Optimization of IC engine design for reduced emissions using CFD modeling," in *Proc. 2nd Thiesel Conf. Thermo and Fluid Dynamic Processes in Diesel Engines*, 2002, pp. 399–414.

[18] R.J. Howlett, M. M. Zoysa, S. D. Walters, and P. A. Howson, "Neural network techniques for monitoring and control of internal combustion engines," in *Proc. Int. Symp. Intell. Ind. Autom.*, 1999.

[19] D.T. Montgomery and R. D. Reitz, "Evaluation of the EGR and multiple injections on particulate and NOx emissions from a D.I. diesel engine," SAE paper 960316, 1996.

[20] A. E. Klingbeil, H. Juneja, Y. Ra, and R. D. Reitz, "Premixed diesel combustion analysis in a heavy-duty diesel engine," SAE paper 2003-01-0341, 2003.

[21] M. P. Thiel, A. E. Klingbeil, and R. D. Reitz, "Experimental optimization of a heavy-duty diesel engine using automated genetic algorithms," SAE paper 2002-01-0960, 2002.

[22] P. K. Senecal and R. D. Reitz, "Simultaneous reduction of engine emissions and fuel consumption using genetic algorithms and multi-dimensional spray and combustion modelling," SAE paper 2000-01-1890, 2000.

[23] L. Hernández, "Desarrollo de una metodología para la predicción y optimización de emisiones contaminantes y consumo en motores Diesel de automoción mediante redes neuronales artificiales.," Ph.D. dissertation, Valencia, Spain, 2004, (in Spanish).

[24] R. Christian, F. Knopf, A. Jasmek, and W. Schindler, "A new method for the filter smoke number measurement with improved sensitivity," *MTZ*, vol. 54, pp. 16–22, 1993, (in German).

[25] G. G. Muntaen, "A theoretical model for the correlation of smoke number to dry particulate concentration in diesel exhaust," SAE paper 1999-01-0515, 1999.

[26] European Directive 1999/96/CE, European Official Journal.

[27] S. A. Molina, "Estudio de la influencia de los parámetros de inyección y la recirculación de gases de escape sobre el proceso de combustión, las prestaciones y las emisiones de un motor diesel de 1.8 litros de cilindrada.," Ph.D. dissertation, Valencia, Spain, 2003, (in Spanish).

[28] *"Operating Manual AVL 415. Variable Sampling Smoke Meter,"* 1996, AVL List GmbH.

[29] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice-Hall, 1994, ISBN 0-13-908385-5.

[30] J. M. Zurada, *Introduction to Artificial Neural Systems*. Boston, MA: PWS Publishing, 1992.

[31] A. J. Maren, C. T. Herston, and R. M. Pap, *Handbook of Neural Computing Applications*. New York: Academic, 1990, Ed..

[32] H. Demuth and M. Beale, *Neural Network Toolbox for Use with Matlab: User's Guide Version 3*. Natick, MA: The Mathworks, 1998.

[33] G. J. Thompson, C. M. Atkinson, N. N. Clark, T. W. Long, and E. Hanzevack, "Neural network modelling of the emissions and performance of a heavy-duty diesel engine," in *Proc. Instn. Mech. Engineers, Part D, J. Automobile Eng.*, 2000, vol. 214(D), pp. 111–126.

[34] J. M. Desantes, J. V. Benajes, S. A. Molina, and L. Hernández, "Multi-objective optimization of heavy-duty diesel engines under stationary conditions," in *Proc. Instn. Mech. Engineers, Part D, J. Automobile Eng.*, 2005, vol. 219, pp. 77–87.

[35] J. Galindo, J. M. Luján, J. R. Serrrano, and L. Hernández, "Combustion simulation of turbocharger HSDI Diesel engines during transient operation using neural networks," *Appl. Thermal Eng.*, vol. 25/5-6, pp. 877–898, 2005.

[36] C. Wang, S. S. Venkatesh, and J. S. Judd, "Optimal stopping and effective machine complexity in learning," in *Proc. Conf. Neural Inf. Process. Syst.*, Denver, CO, 1993, pp. 303–310.

[37] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quart. Appl. Math.*, vol. 2, pp. 164–168, 1944.

[38] D.W. Marquardt, "An algorithm for least-squares estimations of non-linear parameters," *SIAM J. Appl. Math.*, vol. 11, pp. 431–441, 1963.

[39] L. Davis, *The Handbook of Genetic Algorithms*. New York: Van Nostrand Reingold, 1992.

[40] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*. New York: Springer-Verlag, 1996, AI Series.

[41] A. H. Wright, "Genetic algorithms for real parameter optimization," in *Foundations of Genetic Algorithms*. San Mateo, CA: Morgan Kauffman, 1991, pp. 205–218.

[42] D. D. Wickman, P. K. Senecal, and R. D. Reitz, "Diesel engine combustion chamber geometry optimization using genetic algorithms and multi-dimensional spray and combustion modeling," SAE 2001-01-0547, 2001.

[43] M. J. Brauer, M. T. Holder, L. A. Dries, D. J. Zwickl, P. O. Lewis, and D. M. Hillis, "Genetic algorithms and parallel processing in maximum-likelihood phylogeny inference," *Molecular Biology and Evolution*, vol. 10, no. 19, pp. 1717–1726, 2002.

[44] T. Starkweather, D. Whitley, and K. Mathias, "Optimization using distributed genetic algorithms," *Lecture Notes in Computer Science*, vol. 496, pp. 176–185, 1991.

[45] J. E. Baker, "Reducing bias and inefficiency in the selection algorithm," in *Proc. 2nd Int. Conf. Genetic Algorithms*, 1987, pp. 14–21.

[46] H. Mühlenbein and D. Schlierkamp-Voosen, "Predictive models for the breeder genetic algorithms," *Evol. Comput.*, vol. 1, no. 1, pp. 25–49, 1993.

[47] R. Salomon, "Short notes on the schema theorem and the building bock hypothesis in genetic algorithms," in *Evolutionary Programming VII*. New York: Springer-Verlag, 1998, pp. 113–122.

[48] D. M. Tate and A. E. Smith, "Expected allele coverage and the role of mutation in genetic algorithms," in *Proc. ICGA 5*, 1993, pp. 31–37.

[49] A. Chipperfield, P. Fleming, H. Pohlheim, and C. Fonseca, Genetic algorithm toolbox user's guide 1.2 Dept. Autom. Control Syst. Eng., Univ. Sheffield. Sheffield, U.K. [Online]. Available: http://www.shef.ac.uk/content/1/c6/03/35/06/manual.pdf

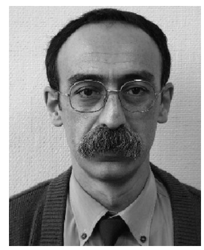
José M. Alonso received the M.Sc. degree in computer science from the Universidad Politécnica de Valencia (UPV), Valencia, Spain, in 1997. He is currently working towards the Ph.D. degree in the field of parallel and grid computing in simulation of building structures at UPV

He is a member of the High Performance Networking and Computing Group, UPV, since 1996. Since 2000, he has been an Assistant Lecturer at UPV. He has also participated in different research projects about the application of parallel and grid computing techniques to several areas of engineering. Other fields of interest include neural networks and evolutionary computation.

Fernando Alvarruiz received the M.Sc. degree in computer science from the Universidad Politécnica de Valencia (UPV), Valencia, Spain, in 1995. He is currently working towards the Ph.D. degree, related to the application of parallel computing to simulation and optimization processes in water distribution networks at UPV.

Since 2000, he has been an Assistant Lecturer at UPV. Other fields of interest include neural networks and evolutionary computation.

Mr. Alvarruiz received a scholarship for postgraduate studies in the High Performance Networking and Computing Group, UPV.



José M. Desantes received the Degree in mechanical engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 1978 and the Ph.D. degree in engineering from the Universidad Politécnica de Valencia (UPV), Valencia, Spain, in 1980.

Since 1984, he has been a Full Professor at the Department of Thermal Engines, UPV. Director of the University Research Institute CMT-Motores Térmicos, he has collaborated in 150 journal and conference papers related to ICE topics. He has supervised 15 Ph.D. thesis, has collaborated in 200 research projects funded by national programs, European union programs, or directly by the industry. His research interest is on thermofluidynamics of internal combustion engines (ICE), mainly focused in pollutant emission control and engine efficiency improvements.

Leonor Hernández received the Degree in physics from the Universidad de Valencia, Valencia, Spain, in 1998 and the Ph.D. degree from the Universidad Politécnica de Valencia (UPV), Valencia, in 2004.

Since 2005, she holds an Assistant Lecturer position at the Universitat Jaume I, Castellón, Spain. Her research interests are in emission modeling and optimization techniques.

Vicente Hernández received the Ph.D. degree in mathematics from the Universidad de Valencia, Valencia, Spain, in 1979.

Since 1991, he has been a Full Professor in the Department of Information Systems and Computation, Universidad Politécnica de Valencia (UPV), Valencia. Leader of the High Performance Networking and Computing Group, he has written over 220 journal and conference papers on numerical methods, parallel computing, grid computing and their applications on different fields. He has supervised over 20 Ph.D. thesis on computational science and applied mathematics. He has acted as coordinator or main researcher in more than 30 projects funded by national and international programs.

Germán Moltó was born in Valencia, Spain, on January 6, 1979. He received the M.Sc. degree in computer science from the Universidad Politécnica de Valencia (UPV), Valencia, Spain, in 2002. He is currently working towards the Ph.D. degree in the field of grid computing applied to biomedical applications at UPV.

Since 2002, he has been a Researcher at the High Performance Networking and Computing Group, UPV. In 2005, he received an Assistant Lecturer position.