www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

PAN ARCHIVES OF ACOUSTICS
Vol. 43, No. 3, pp. 385–395 (2018)
Copyright © 2018 by PAN – IPPT
DOI: 10.24425/123910

Research Papers

Vibroacoustic Real Time Fuel Classification in Diesel Engine

Andrzej BĄKOWSKI[(1)], Michał KEKEZ[(1)], Leszek RADZISZEWSKI[(1)], Alžbeta SAPIETOVA[(2)]

[(1)] *Faculty of Mechatronics and Mechanical Engineering, Kielce University of Technology*
Al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland; e-mail: {abakowski, m.kekez, lradzisz}@tu.kielce.pl

[(2)] *University of Žilina*
Univerzitná 8215/1, 010 26 Žilina, Slovakia; e-mail: alzbeta.sapietova@fstroj.uniza.sk

(*received September 18, 2017; accepted April 18, 2018*)

Five models and methodology are discussed in this paper for constructing classifiers capable of recognizing in real time the type of fuel injected into a diesel engine cylinder to accuracy acceptable in practical technical applications. Experimental research was carried out on the dynamic engine test facility. The signal of in-cylinder and in-injection line pressure in an internal combustion engine powered by mineral fuel, biodiesel or blends of these two fuel types was evaluated using the vibro-acoustic method. Computational intelligence methods such as classification trees, particle swarm optimization and random forest were applied.

Keywords: fuel recognition; classification trees; particle swarm optimization; random forest.

1. Introduction

Proper engine performance control is directly related to the recognition of the fuel used, as the engine power and the composition of exhaust gases depend on fuel properties (Morón-Villarreyes, 2007). Diesel fuel used in internal combustion (IC) engines is a mixture of several dozen constituents, with 7% of biodiesel as a blend component. Biofuels are produced from different plant materials, such as rapeseed, soybean, olives, corn, jatropha, coconuts, sunflowers, peanuts or fish oil (Sakthivel, 2016) depending on their availability. While the required ranges for the fuel parameters are generally available (Safety Data Sheets), accurate values of physical and chemical properties are the trade secrets of fuel manufacturers. There are a number of factors that may affect fuel properties. These are improper storage, transportation or unfair practices of intermediaries. The actual fuel composition may also be altered by the left residue in the fuel tank mixed with new fuel. Unstable properties of a combusted fuel impede the control of engine operation, alter the exhaust composition, and contribute to generating more noise.

Fuel type recognition in laboratories is conducted through very accurate and expensive devices that analyse fuel chemical composition and physical characteristics. In the field, portable analysers (relatively costly) are used, for example, ERASPEC fitted with an interferometer operating within a wavelength range 630 to 4000 cm^{-1}, which allow performing spectral analyses and determining physical and chemical parameters such as density, cetane number and aromatic content. The long measurement time, about 60 seconds, can be a limitation (Teixeira *et al.*, 2008).

The authors of this article propose the vibro-acoustic method for recognizing injected fuel types but by making use of experimentally obtained indicator diagrams of pressure in the cylinder and in the injection line (Kekez, Radziszewski, 2011). The purpose of this study is to construct models for the real time recognition of fuel grade injected into a diesel engine, with an accuracy of not less than 95%. The models will use the descriptors of pressure in the combustion chamber or injection piping and the algorithms based on artificial intelligence methods for implementing them in 8-bit microcontrollers. The fuel should be recognized after refuelling, under stationary operation of the engine and within not more than a few seconds.

2. Vibro-acoustic monitoring and diagnostics of an internal combustion engine

Reciprocating machinery, such as the internal combustion (IC) engine, generates disturbances that induce vibration and noise. The key vibration and

noise sources are classified as mechanical, aerodynamic and fuel combustion process related. The disturbances propagate through air, solids or both of these media, and tend to interact. Vibration, stress waves or noise measuring systems collect condition monitoring and fault diagnostics data for analysis. As a transducer is a part of the signal path, care must be used when mounting it so that disturbance measurands can be evaluated with the lowest possible uncertainty. Each mechanical system has its own, characteristic vibro-acoustic signature which is a signal measured under defined operating conditions of that system. The challenge is to decide which types of failure and under which conditions can use vibro-acoustic measurement as an indicator, and to select the right transducer and adequate diagnostic technique (Delvecchio *et al.*, 2018). Structural resonance-based monitoring and diagnostic methods use the information generated by a range of impulsive sources, which can include piston slap, engine knock, misfiring, chain-sprocket, injector faults. Vibration signals are measured by accelerometers fixed on the engine block. The typical frequency response range is 10 kHz. In (Szymański, Tomaszewski, 2016), an experimental diagnostic model is proposed that is based on resonance vibration frequency data collected from IC engine sub-assemblies during impact excitations. The method uses eigenfrequency band bound with the reaction forced by an impulse.

To analyse the combustion processes, the transducer (piezoelectric crystal sensor or an optical sensor) is placed inside the combustion chamber (Payri *et al.*, 2010). Piezoelectric properties of materials (e.g. SiO_2, $LiNbO_3$, or $LiTAO_3$) are strongly temperature-dependent – they decline with the rising temperature. A rapid decrease in piezoelectric properties of silicon is observed at a temperature of 523 K while at 846 K the piezoelectric properties disappear. Gallium orthophosphate is stable up to the temperature of 1206 K and its sensitivity remains constant up to 773 K. By the reason of the above, particular attention should be paid to maintaining low (below the Curie point) and stable working temperature of the piezoelectric transducer. A variety of piezoelectric transducer designs are used to indicate a diesel engine, typically cooled with water. Cooling with a liquid prevents the transducer from overheating, reduces thermal drift, prevents the reduction in insulation resistance, and enables installation of the transducer directly in the combustion chamber. In-cylinder pressure signals are indicative of the rate of pressure rise, indicated mean effective pressure, combustion phasing, heat release and ignition delay for different working conditions (Maurya *et al.*, 2013).

Most promising approaches to IC engine condition monitoring include the techniques relying on sound pressure (SP) and acoustic emission (AE) signals. AE measurement techniques involve the detection of elastic waves, usually within a range of 0.1 to 1 MHz (Ranachowski, Bejger, 2005). AE methods are used for detection of exhaust valve faults, misfire, injector faults, and diesel knock (Lowe *et al.*, 2011). SP signals measured remotely without contact within the broad frequency range of 20 Hz–20 kHz (Deptuła *et al.*, 2016) are mainly used to identify engine noise sources (combustion/mechanical), decompose pure airborne signals into their sources (alternator, turbocharger, ventilation fan) and detect major sources of engine noise, i.e. piston slap, advanced/delay of injection (Albarbar *et al.*, 2010). Problems can occur due to the high level of masking effect from background noise in the locations where SP signals are measured. Continuous Wavelet Transform or Discrete Wavelet Transform and independent Component Analysis can be adopted to extract the informative content from the real noise sources, but as the surrounding noise is neither white, nor harmonic, nor stationary, the separation techniques may be problematic (Figlus *et al.*, 2014). In enclosed spaces, the characteristics of the sound source are influenced by reflection or inference from other signals. Time domain analyses are more difficult to interpret than the results of frequency domain analyses. No criterion for the number and position of microphones capturing SP signals has been well defined. The most common technique for monitoring and diagnosing IC engine condition using SP signals is the time-frequency analysis, the results of which are represented as a three-dimensional graph. The use of three-dimensional graphs may be inefficient in the case of deriving simple features and complex decision algorithms need to be used in analysis (Delvecchio *et al.*, 2018).

A non-intrusive methodology based on vibration and pressure levels measured on cylinder heads is proposed in (Barelli *et al.*, 2009). Both signals are strictly related to the phenomena inside the cylinder, depending on the combustion frequency. Close relationships between the combustion process and the vibration and noise in the engine were also demonstrated in (Chiatti *et al.*, 2015).

Vibro-acoustic methods have been used for fuel recognition. Authors in (Hardenberg, Hase, 1979) demonstrated that the change in the fuel cetane number affected the angle of self- ignition delay and combustion process. In (Elghamry *et al.*, 1998), the authors measured acoustic emission, vibration and pressure in a gaseous fuel Perkins engine to find that AE signal parameters could be used to evaluate mixture quality, and the vibration and in-cylinder pressure data could be used for fuel type identification. In (Valencia, Armas, 2005), the fuel type was recognized based on the cetane number determined via the analysis of in-cylinder pressure in a marine engine. Flekiewicz (2007) used the vibro-acoustic method based on the measurement of cylinder block vibrations in a petrol- and gas-fuelled engine. The contin-

uous wavelet transform was used to analyse acceleration. The relationship between the time derivative of the wavelet coefficient and the type of fuel was demonstrated. Attempts to identify fuel in a spark ignition engine through vibration analysis were made by (Gravalos *et al.*, 2013), who used RMS values of filtered acceleration signals to identify unleaded petrol from blends of ethanol or methanol, however, the percentage of the blends could not be identified. An algorithm for identifying the type of fuel used in spark ignition engine was proposed. The object of the study in (Grajales *et al.*, 2017) was to assess the influence of different blends of petrol and ethanol (E8, E20 and E30) on the spectral composition of vibration signals collected from the engine operating under simulated misfire. Accelerometers were installed at three points, and the measurement time was 2 seconds. Normal FFT and high frequency (Envelope, Spike Energy and Peak Value) analysis of the signals indicated no significant differences between blends. Fuel blend classification was possible after evaluation of the statistical features mean, RMS and energy in the spectrum from the Peak Value, using the filters: high-pass at 5 kHz, band-pass at 5–7 kHz and 5–10 kHz. Since the vibration and AE measurements and their analyses are difficult to conduct during typical vehicle operation, the use of this method is limited.

3. Test facility

The test facility used in the experiments consisted of a medium-speed aspirated, compression ignition engine with direct injection, a water brake and a control unit for controlling the facility and displaying engine and brake operation parameters. A Perkins three cylinder AD3.152 UR engine was used in the tests. The design of the test facility and the schematic of the measurement paths are reported in (Ambrozik *et al.*, 2014). The experiments on the test bench (Kekez, Radziszewski, 2011) included the measurement of fuel injection line pressure, in-cylinder pressure and injector needle lift. The in-cylinder pressure measurements were made using the water-cooled piezoelectric sensor AVL QC34D, mounted in the cylinder head. Large changes in in-cylinder temperature and the strains on the membrane of the pressure transducer made the relative error of maximum pressure measurement amount to about 1%. The pressure in the fuel injection pipes was measured using the CL31 ZEPWN Marki transducer. The parameters of the piezoelectric transducers used for pressure measurements are compiled in Table 1.

All parameters were measured as a function of crank angle, with a resolution of 1.4°, which provided 512 measuring points per one engine duty cycle.

The results used in this paper were from the tests of the engine operating at full load and at the crankshaft speeds ranging from 1000 to 2000 rpm. Five different fuels were used: diesel, RME (rapeseed oil methyl esters) and blends of these two fuels, B10, B20 and B30, containing 10, 20 and 30% of RME, respectively. The values for the 50 full duty cycles were recorded per kind of operation conditions. Indicator diagrams, acquired in this way, were averaged thereafter. Of the measured signals, the indicator graphs of pressure in the cylinder and in the injection line are the most important when identifying the fuel type. Figure 1 shows a graph of cylinder pressure dependence on a crank angle (CA) for a diesel engine powered by diesel fuel or RME for an example of selected three crankshaft speeds. These graphs have common characteristics, independent of the type of fuel, and others that distinguish them. For the full load operation of an engine running on diesel fuel, as shown in Figure 1a, the intense pressure increase starts in the range of 3.5 MPa to 4 MPa for CA from 353° to 359°. The position of the point representing maximum pressure is different at each rotational speed and does not go beyond 8 MPa to 8.8 MPa. For the full load operation of an engine running on RME, as shown in Fig. 1b, the intense pressure increase starts in the range of 3.3 MPa to 3.9 MPa for CA from 351° to 357°. The position of the point representing maximum pressure is different for each rotational speed and does not exceed within 8 MPa to 8.5 MPa. With the engine fueled with B10, B20, B30, the above differences for individual fuels were the easiest to observe at 1200 rpm. – then the intense increase in pressure caused by spontaneous combustion starts in the range of 2.8 MPa to 3.2 MPa and at 352°CA. The position of the point representing maximum pressure is different for each fuel and stays within 8.3 MPa to 8.6 MPa. It should be noted that the parameters indicated may be different for each engine type. This necessitates experimental research for each type of engine of interest. The CA range between 337.5° and 375.5° was taken into account to find the maximum and minimum pressure points in the cylinder. The minimum value was at 337.5°CA. The CA range was chosen so that the valves were

Table 1. Selected specification data of the transducers used for measuring pressure in the cylinder and injection pipes (Bąkowski, Radziszewski, 2015).

Parameter	QC34D AVL	CL31 ZEPWN Marki
Measuring range	0–25 MPa	0–100 MPa
Sensitivity	190 pC/MPa	126 pC/MPa
Nonlinearity	$\leq 0.2\%$	$\leq 0.5\%$
Resonance frequency	69 kHz	50 kHz
Capacitance	10 pF	8 pF
Operating temperature	293–353 K	253–323 K

closed and the amount of data for the calculations did not exceed the capacity of the 8-bit microcontroller.

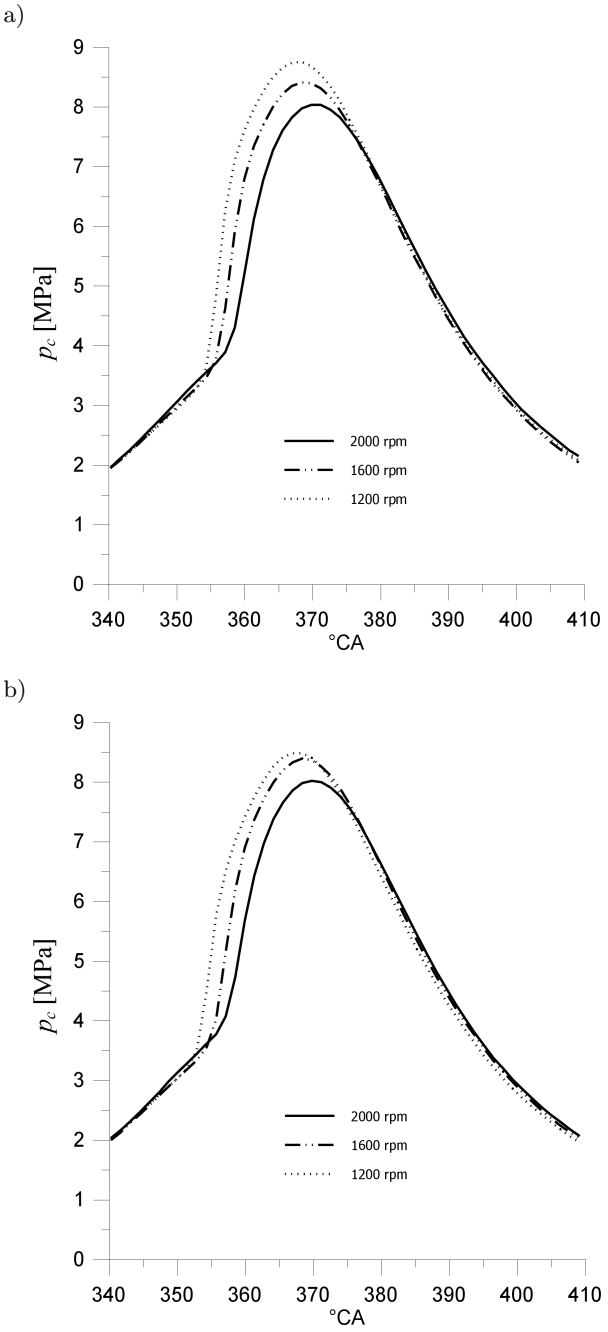

Fig. 1. In-cylinder pressure p_c waveform for selected rotational speeds of the engine operating at full load and powered with (a) diesel fuel or (b) RME.

4. Selected descriptors of pressure changes

Random disturbances in the operation of cyclic motors cause varied characteristics of the subsequent duty cycles even in the steady state. Parameters describing the IC engine unique operating characteristics can be divided into four groups related to: in-cylinder pressure distribution, the process of heat release, the development of the flame front, the removal of exhaust gas.

In the group of parameters related to the in-cylinder pressure, we can distinguish: maximum pressure, the CA at which maximum pressure or maximum pressure increase occurs, maximum pressure increase, and mean indicated pressure. Parameters containing the most information about the history of the combustion heat release are: maximum combustion heat release rate, maximum fuel mass burning rate, ignition delay angle, combustion duration angle, interval of crankshaft rotation from the ignition angle to the angle at which a portion of fuel mass is burned. The parameters defining the in-cylinder pressure are most commonly used, mainly because of the ease of performing experiments. The combustion-related parameters are determined after analyzing the pressure graphs and using the thermodynamic heat release model. The values of those signals are used to control the engine operation and for diagnostic purposes (Maurya *et al.*, 2013; Yoon *et al.*, 2007).

Known in-cylinder pressure changes and adequate thermodynamic models allow the calculation of the parameters needed to control the engine operation: maximum in-cylinder pressure $p_{c\max}$, indicated mean effective pressure, and the crank angle at which 50% combustion heat was released (α_{q50}). To determine α_{q50} with due uncertainty, adequate thermodynamic models have to be used, which makes this method difficult and time consuming. For the same reasons, other descriptors are also applied, such as the CA at which $p_{c\max}$ appears, the maximum value of the slope of the tangent line to the pressure curve. The first and second derivative of the pressure curve with respect to the crankshaft rotation angle allows determining the start of self-ignition or the coefficient of variation of the pressure signal.

Statistical and metrological analyses showed that the maximum values of p_c signal recorded in consecutive duty cycles could be considered to be stationary under all operating conditions. The verification of the maximum values of the p_c signal with the normal distribution did not provide any grounds to reject the null hypothesis at the 5% significance level. Standard uncertainty of the in-cylinder pressure calculated for a type B evaluation was $u_B = 0.046$ MPa for uniform distribution. This analysis allowed determining pressure mean values for all CAs at which pressure had been recorded. The pressure mean values with respect to the i-th crank angle can be determined according to the following algorithm

$$\overline{p_c}(i) = \frac{1}{50}\sum_{j=1}^{50} p_c(i,j), \tag{1}$$

where j is the number of the duty cycle and takes values from 1 to 50, and i is the crank angle index,

for which the mean value is determined, and which takes values from 1 to 512. After determining $\bar{p_c}(i)$, pressure p_c deviations from the mean values can also be calculated, according to the following equation

$$\Delta p_c(i,j) = p_c(i,j) - \overline{p_c}(i). \tag{2}$$

It was demonstrated in (Bąkowski, Radziszewski, 2015) that the signal of the sum of squared deviations ($\sum \Delta p_c^2$) determined for particular crank angles can be used to find the values of the combustion onset angles (α_{ps}) and those angles for which the heat release rate reaches its maximum ($\alpha_{\dot{Q}\,\max}$). Determining the value of this descriptor does not require a lot of computing power and is even easier than determining the coefficient of standard deviation or performing a FFT analysis of the pressure signal. It should also be noted that the algorithms for determining descriptors based on variation coefficients are fairly simple to implement, do not require high-power processors or a long processing time, i.e. about 0.02 seconds. This allows their real time application.

Figure 2 shows an example of a graph for the sum of in-cylinder pressure squared deviations from its mean values ($\sum \Delta p_c^2$) with determined angles α_{ps} and $\alpha_{\dot{Q}\,\max}$ marked. Angles α_{ps} and $\alpha_{\dot{Q}\,\max}$ determined based on this graph amount to $\alpha_{ps} = 353.0°$ and $\alpha_{\dot{Q}\,\max} = 355.8°$ for the diesel-powered engine operating at full load at a speed of 1200 rpm.

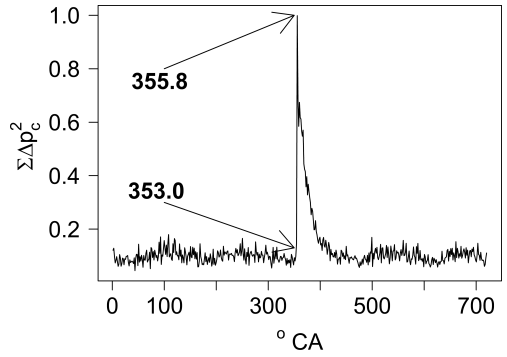

Fig. 2. Sum of squared in-cylinder pressure deviations $\Delta p_c(i,j)$ from the mean values for the first two duty cycles – for the diesel powered engine operating at full load at a speed of 1200 rpm.

Based on the analysis of the pressure signal descriptors, the authors decided to use further in the article the minimum or maximum pressure in the cylinder or in the injection lines and the CA of the maximum heat release rate for fuel type recognition. Recognition effectiveness was assumed to be the usefulness criterion for these descriptors.

5. Fuel recognition models

Computational intelligence methods (Sakthivel, 2016; Ruiza *et al.*, 2017), including artificial neural networks, which in many cases (Pietraszek *et al.*, 2014) allow building accurate classifiers, are used in the control or modelling (Brzozowski, Nowakowski, 2014) of IC engine work. Unfortunately, neural networks comprehensibility and transparency are low in comparison to that of classification trees. Moreover, the relatively long time needed to build the classifier (or a need of high computational power) is another obstacle for the user. Typically, all combinations of the network learning method, network architecture parameters (number of neurons in the hidden layer) and activation function (sigmoid or radial-basis function) must be checked. For these reasons, it was decided to test the possibility of using other artificial intelligence methods in fuel recognition models.

5.1. Application of classification trees

Decision trees (more precisely, classification trees) are applied to fuel type recognition on account of their advantages, including a short time needed to build a classifier, faster fuel recognition by the classifier built (a decision tree), an accessible form of the classifier, and the possibility of justifying the decision being made. Further in the article, terms "classification tree" and "decision tree" are used interchangeably. The C5.0/See5 system was chosen from among available software for its quick and very good results (accurate classifiers), as confirmed by the authors' experience (Kekez *et al.*, 2016). To construct a decision tree, a training dataset needs to be built. The training data was built based on the average values of in-cylinder and/or in-injection line pressure at more than ten/several dozen crank angles for which the measurements were made.

To build an accurate classifier, it is essential to select adequate descriptors of in-cylinder pressure changes. The training dataset consisted of records containing four input variables: the maximum (x_1), and minimum (x_2) pressure in the cylinder, and the maximum (x_3) and minimum (x_4) pressure in the injection line for the investigated crack angle range – Model 1. The output variable in the dataset was the fuel type (y). Model 1 contained a total of five variables. The maximum cylinder pressure varies with the engine rotational speed and type of fuel tested. The largest differences between maximum cylinder pressure values for different fuels occur at the speed of 1200 rpm, therefore the data for this rotational speed was used to prepare the training sets for this classifier and for those which will follow. For building the first classifier, the measurement data was averaged for all 40 possible consecutive work cycles taken from the 50 cycles recorded for a given fuel. The resultant dataset comprised 11 records for each fuel type (Kekez *et al.*, 2016). In less than 0.05 s, C5.0/See5 software produced a relatively simple classifier which recognized the fuel type with

100% accuracy on the training set. The classifier has a form of a decision tree with five leaves, and can be converted to an equivalent set of rules (Fig. 3). There is no variable x_4 in the derived classifier, because decision tree algorithms do not always use all input variables, (but only those that in a given node make it possible to split the dataset into groups that correspond to each decision).

y = diesel if $x_1 > 8.480765$ and $x_2 \leq 1.707941$
y = RME if $x_1 \leq 8.480765$ and $x_2 > 1.539154$
y = B10 if $x_1 \leq 8.480765$ and $x_2 \leq 1.539154$
y = B20 if $x_1 > 8.480765$ and $x_2 > 1.707941$ and $x_3 \leq 20.757$
y = B30 if $x_1 > 8.480765$ and $x_2 > 1.707941$ and $x_3 > 20.757$

Fig. 3. Model 1 classifier of fuel recognition, using the classification trees method, four input variables, C5.0/See5 software and measurement data from 40 consecutive engine cycles.

The predictive accuracy of the classifier was tested using *k-fold cross validation* (Witten, 2005) for $k = 5$, $k = 10$ or $k = 11$, and the result achieved was 92.7%. With the *leave one out* method for $k = 55$, the accuracy increased to 96.4%. The strength of the proposed model is high classification accuracy and a small number of simple rules. The limitation is the large number of consecutive engine cycles needed to identify the fuel and the need to build two pressure measuring chains in the cylinder and injection lines, which complicates the measurement system and increases its cost. For these reasons, the authors decided to build a model similar to the previous one, but with only two variables.

When the training set was simplified in order to contain only the fuel type and the maximum in-cylinder pressure (one input variable instead of four) – Model 2, the classifier built by the C5.0/See5 had a simpler form, Fig. 4.

y = B10 if $x_1 \leq 8.344099$
y = RME if $8.344099 < x_1 \leq 8.480765$
y = B30 if $8.480765 < x_1 \leq 8.608824$
y = B20 if $8.608824 < x_1 \leq 8.651529$
y = diesel if $x_1 > 8.651529$

Fig. 4. Model 2 classifier of fuel recognition, using the classification trees method, one input variable, C5.0/See5 software and measurement data from 40 consecutive engine cycles.

Nevertheless, the classification accuracy for the data from 40 consecutive engine cycles remained at the level of 100%. The predictive accuracy of this classifier tested using *k-fold cross validation* for $k = 5$, $k = 10$ or $k = 11$, was found to decrease to 90.9%, 92.3%, and 92.7%, respectively, whereas for $k = 55$ (*leave one out*) it was 92.7%. This accuracy level is insufficient and 40 cycles needed to identify the fuel is too many for the real-time fuel recognition. Another model, Model 3, uses two input variables, additional decision tree optimization and a smaller number of engine cycles, as discussed in Subsec. 5.2.

5.2. Application of decision trees and particle swarm optimization

The training set in the proposed model, Model 3, was created from 10 consecutive engine cycles data, and with two input variables – maximum (x_1) and minimum (x_2) in-cylinder pressures. The Windows version of decision tree software, See5, built the classifier which had 7 leaves and 95.6% classification accuracy on the training set (Fig. 5). The accuracy of this classifier tested using *k-fold cross validation* for $k = 5$, $k = 10$ or $k = 99$ was 91.7%, 92.7% and 90.7%, respectively. The resulting *confusion matrix* from *10-fold cross validation* has the form as in Fig. 5.

y = B10 if $x_1 \leq 8.511647$ and $x_2 \leq 1.549044$
y = RME if $x_1 \leq 8.511647$ and $x_2 > 1.549044$
y = diesel if $x_1 > 8.511647$ and $x_2 \leq 1.726765$
y = B20 if $x_1 > 8.640588$ and $x_2 > 1.726765$
y = B30 if $8.511647 < x_1 \leq 8.640588$ and $x_2 > 1.768882$
y = B30 if $8.511647 < x_1 \leq 8.605294$ and $1.726765 < x_2 \leq 1.768882$
y = B20 if $8.605294 < x_1 \leq 8.640588$ and $1.726765 < x_2 \leq 1.768882$

(a)	(b)	(c)	(d)	(e)	<-classified as
40			1		(a): class diesel
	40	1			(b): class RME
		35	6		(c): class B30
		6	35		(d): class B20
	1			40	(e): class B10

Fig. 5. Model 3 classifier of fuel recognition, using the classification trees method, two input variables, C5.0/See5 software, training data from 10 consecutive engine cycles, and the confusion matrix of this classifier.

The accuracy tests showed that no fuel was classified 100% correctly, with the most common errors in the classification of B30 and B20 (14.6% of B30 cases were classified as B20 and 14.6% of B20 cases were classified as B30).

Because See5 software uses random partition of a training set into subsets in 10-*fold cross validation*, in other 50 attempts (with almost identical software, Linux counterpart C5.0) classification accuracy ranged from 88.2% to 92.7% with an average of 90.9%. The analysis showed that the recognition accuracy in Model 3 was unsatisfactory – less than 95%. For this reason, Model 3 was modified by increasing the number of engine cycles. When additional calculations were made for training sets, created as previously but containing averaged data from more engine cycles (15, 20, 25, 30, 35, 40), the accuracy after 10-*fold cross valida-*

tion (run once in each case) was 93.9%, 94.3%, 96.2%, 96.1%, 96.2%, and 94.3%, respectively. At the 50th run of 10-*fold cross validation* (with almost identical software, Linux counterpart C5.0), the averaged values amounted to 93.9%, 95.7%, 97.1%, 96.2%, 95.1%, and 93.1%, respectively. These tests found the number of successive engine cycles sufficient to achieve averaged data accuracy above 95% to be between 20 and 25. With more cycles, accuracy was lower than that from 25 cycles because the training sets contained fewer records.

To improve the classification accuracy, another modification of Model 3 was proposed, involving optimization of the values of decision borders in the tree (Fig. 4a) obtained earlier by C5.0/See5. For this purpose, *particle swarm optimization* (PSO) (Engelbrecht, 2006) was used – Model 3_v. PSO. Each of the six decision border values (Fig. 6a) occurring in the tree (8.511647, 1.549044, 1.726765, 8.640588, 1.768882, and 8.605294) was set as a PSO-optimised variable. The Standard PSO 2006 (http://www.particleswarm.info/Programs.html) with minor modification was used in the experiments – the swarm size was set at 100. In the PSO algorithm, the fit value of a particle (intended to be minimized) was the number of cases in the training set misclassified by a decision tree,

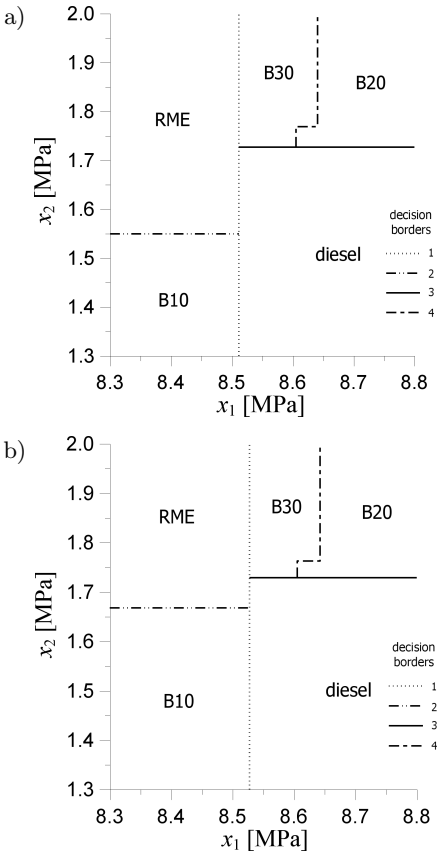

Fig. 6. Decision boundaries and decision areas in the decision tree: a) before optimization, b) after optimization with a swarm of particles.

whose decision boundaries were equal to the values of the corresponding variables describing the location of the particle in the search space. The solution, better than the classification tree in Fig. 5, was found already in the second iteration of the PSO method. This method changed the values of the six used variables to 8.527545, 1.667753, 1.728503, 8.642454, 1.762328, and 8.605295, respectively (Fig. 6b).

This optimized solution had 96.6% accuracy on the training set (95.6% prior to optimization). *10-fold cross validation* on the PSO-classifier showed 93.7% accuracy, which was a better result compared with 91.2% prior to optimization, but still lower than 95%. Model 3 was modified again, but this time it was created from averaged data from 20 engine cycles, containing four decision boundaries: 8.505765, 1.540253, 8.670647, 8.62. Accuracy on the training set before and after optimization was 98.7%. In 10-*fold cross validation,* the classification accuracy without optimization was 94.8%, and 96.1% after optimisation. In another run of the test, the accuracy was 96.1%, and 97.4% after optimisation. This classifier showed that PSO of the classification tree allowed predicting the fuel type with good accuracy by using data collected from 20 engine cycles.

5.3. Application of classification trees using maximum pressure and maximum heat release data

Another in-cylinder pressure change descriptor – maximum heat release point ($Q_{\max}$) was proposed for fuel recognition (Kekez *et al.*, 2016). The method for computing $Q_{\max}$, based on instantaneous in-cylinder pressure in subsequent cycles, was discussed in (Bąkowski, Radziszewski, 2015). Two approaches of building a fuel recognition classifier were compared. The first method used only the maximum in-cylinder pressure, x_1 or $p_{c\,\max}$, whereas the second method used also the in-cylinder pressure at $Q_{\max}$, $p_{c\_Q\,\max}$. The training set consisted of records containing values of two or three variables: fuel type (y), $p_{c\,\max}$, and, in the case of the second training set, also $p_{c\_Q\,\max}$ – Model 4.

First, the averaged training items from 40 consecutive engine cycles were used. The accuracy of the classification trees built on both training sets (one with $p_{c\,\max}$ – Model 2 and the other with $p_{c\,\max}$ and $p_{c\_Q\,\max}$ – Model 4) was 100% on the training set. Time needed to recognize the fuel was below 0.05 s in both cases. The structure of the two trees was identical, and the trees varied (in two nodes) by the variable with a tested value and by the limit value in the two nodes. For the reader's convenience, let us recall that when the classifier based on 40 consecutive engine cycles, Model 2, was validated with *k-fold cross validation* for $k = 5$, $k = 10$, $k = 11$ or $k = 55$, accuracy was 90.9% 92.3%, 92.7%, and 92.7%, respectively. When the classifier based on 40 consecutive engine cycles, Model 4, was validated

with *k-fold cross validation* for $k = 5$, $k = 10$, $k = 11$ or $k = 55$, accuracy achieved in each case was higher than for the classifier in Model 2 – 92.7%, 94.3%, 94.5%, 94.5%, respectively. However, since accuracy achieved was lower than 95%, the size of the training set was increased, reducing the number of subsequent engine cycles from which the training items were averaged. The training data averaged from only 10 consecutive working cycles allowed creating the classification trees of the accuracy on the training set of 95.6% in Model 2 and 97.6% in Model 4. The *k-fold cross validation* of the classifier in Model 2 for $k = 5$, 10 or 99 showed classification accuracy of 93.2%, 92.7% and 92.6%. The accuracy of the classifier in Model 4 (Fig. 7) (containing $p_{c\_Q\,\max}$) achieved 93.7%, 94.1%, and 95.5%, respectively. The use of the $p_{c\_Q\,\max}$ variable improved the accuracy of the classification.

y = B10 if $x_1 \leq 8.365527$

y = RME if $8.365527 < x_1 \leq 8.511647$

y = B30 if $8.511647 < x_1 \leq 8.605294$

y = B20 if $8.605294 < x_1 \leq 8.619471$

y = B20 if $8.619471 < x_1 \leq 8.640588$ and $p_{c\_Q\max} \leq 4.467706$

y = B30 if $8.619471 < x_1 \leq 8.640588$ and $p_{c\_Q\max} > 4.467706$

y = B20 if $8.640588 < x_1 \leq 8.688882$

y = diesel if $x_1 > 8.688882$

Fig. 7. Classifier in Model 4 for fuel recognition, using classification trees, two input variables, C5.0/See5 software, and measurement data from 10 consecutive engine cycles.

When additional calculations were made for the training sets created in Model 2 but containing data averaged from more cycles (15, 20, 25, 30, 35, 40), classification accuracy achieved in *10-fold cross validation* (in a single run for each case) was 95.0%, 96.1%, 97.7%, 96.1%, 95.0%, 92.7%, respectively. At the 50th run of *10-fold cross validation* (Linux counterpart C5.0), the average values were 94.7%, 95.9%, 97.4%, 96.0%, 94.6%, 92.7%.

When additional calculations were made for the training sets created in Model 4 but containing data averaged from more engine cycles (15, 20, 25, 30, 35, 40), classification accuracy achieved in *10-fold cross validation* (in a single run for each case) was 95.0%, 96.1%, 97.7%, 96.3%, 96.2%, 94.7%, respectively. In all cases, accuracy achieved in Model 4 was the same or better than that in Model 2. The sufficient number of cycles for measurement data averaging was 25. At the 50th run of *10-fold cross validation* (Linux counterpart C5.0), the average values were 95.1%, 95.8%, 97.4%, 95.1%, 95.7%, 94.3%, respectively. The values obtained pass the 95% threshold for the training data averaged over each cycle number in the range of 15 to 35, unlike in Model 2, where the averaged data from 15, 35 and 40 cycles did not pass the 95% threshold.

5.4. Application of Random Forests

In order to increase the recognition accuracy, Model 5 was generated in which Random Forest (Breiman, 2001) and in-cylinder pressure descriptors, $p_{c\,\max}$ and $p_{c\_Q\,\max}$ were used. The Random Forest method creates an "ensemble of classification trees" consisting of 100 trees or 500 trees, depending on the implementation. Each tree is grown on the basis of a new training set drawn by sampling with replacement. Also, a subset of input variables is drawn at random – this technique is called *random feature selection*. For a training set made up of 25 engine cycles, the Random-Forest algorithm, or its implementation in Weka software (http://www.cs.waikato.ac.nz/~ml/weka/) was run with default parameters (the number of trees was 100, no randomization of input variables). The resulting classifier achieved 100% accuracy in *10-fold cross validation*. However, in any of the trees that make up the classifier there is no $p_{c\_Q\,\max}$ variable, but only p_{cmax}. When the algorithm was started with the option of random sampling of one input variable – Model 5_v.1, the resulting classifier had both variables in most of the trees. This classifier also achieved 100% accuracy in *10-fold cross validation*. The weakness of each of the two built classifiers is their low clarity (they consist of 100 trees each), a slightly longer building time (successively 0.13 and 0.19 seconds on a PC) and the time of classifying a new case. When the algorithm with the default parameters (without random input variables) was started – Model 5 for data generated from 10, 15, 20, 30, 35, 40 engine cycles, the *10-fold cross validation* accuracy was 95.1%, 97.8%, 98.1%, 100%, 100%, 100%. When similar calculations were performed with the option of randomly drawing 1 input variable (Model 5_v.1), the accuracy levels were 95.1%, 98.3%, 100%, 99.0%, 100%, 100%, respectively.

Decision trees and other classifiers can be implemented on a PC but also in a microcontroller. The authors propose the following algorithm for the recognition of the type of fuel:

a) refuelling the car,
b) engine operating at constant speed, e.g. 1200 rpm,
c) collecting measurement data, i.e. in-cylinder pressure and injection pressure as a function of a crank angle (for 1 or 2 s),
d) sending collected data (i.e. cylinder pressure changes) to a microcontroller,
e) calculating adequate descriptors of cylinder pressure signal by the microcontroller,
f) using the classifier by the microcontroller in order to recognize the fuel type.

The 8-bit microcontroller with 2 KB of flash memory was sufficient to implement models 1, 2, 3 and 4. The classification of the fuel lasted no longer than 0.1 second. Fuel identification system using Model 5

requires the use of 8-bit microcontroller with 32 KB of flash memory.

6. Concluding remarks

All developed models, except No. 1, allow real-time fuel recognition with accuracy greater than 95% and can be implemented on an 8-bit microcontroller. Depending on the model used and its parameters, the recognition accuracy (evaluated in a single run of *10-fold cross validation*) is in the range between 92% and 100%, as shown in Table 2. Analysis of Table 2 data shows that the classification results are mostly affected by the classifiers used, in-cylinder pressure de-

Table 2. Accuracy analysis of fuel recognized in one run of *10-fold cross validation*.

Model	n	Variables in the training set: x_1, x_2, x_3, x_4, x_5	p
Model 1	40	x_1, x_2, x_3, x_4, –	92.7%
Model 2	10	x_1, –, –, –, –	92.7%
	15	x_1, –, –, –, –	95.0%
	20	x_1, –, –, –, –	96.1%
	25	x_1, –, –, –, –	97.7%
	30	x_1, –, –, –, –	96.1%
	35	x_1, –, –, –, –	95.0%
	40	x_1, –, –, –, –	92.7%
Model 3	10	x_1,x_2, –, –, –	92.7%
	15	x_1,x_2, –, –, –	93.9%
	20	x_1,x_2, –, –, –	94.3%
	25	x_1,x_2, –, –, –	96.2%
	30	x_1,x_2, –, –, –	96.1%
	35	x_1,x_2, –, –, –	96.2%
	40	x_1,x_2, –, –, –	94.3%
Model 3_v. PSO	10	x_1,x_2, –, –, –	93.7%
	20	x_1,x_2, –, –, –	96.1%
Model 4	10	x_1, –, –, –, x_5	94.1%
	15	x_1, –, –, –, x_5	95.0%
	20	x_1, –, –, –, x_5	96.1%
	25	x_1, –, –, –, x_5	97.7%
	30	x_1, –, –, –, x_5	96.3%
	35	x_1, –, –, –, x_5	96.2%
	40	x_1, –, –, –, x_5	94.7%
Model 5	10	x_1, –, –, –, x_5	95.1%
	15	x_1, –, –, –, x_5	97.8%
	20	x_1, –, –, –, x_5	98.1%
	25	x_1, –, –, –, x_5	100%
	30	x_1, –, –, –, x_5	100%
	35	x_1, –, –, –, x_5	100%
	40	x_1, –, –, –, x_5	100%
Model 5_v.1	10	x_1, –, –, –, x_5	95.1%
	15	x_1, –, –, –, x_5	98.3%
	20	x_1, –, –, –, x_5	100%
	25	x_1, –, –, –, x_5	100%
	30	x_1, –, –, –, x_5	99.0%
	35	x_1, –, –, –, x_5	100%
	40	x_1, –, –, –, x_5	100%

Denotation: n – the number of cycles from which training data were averaged,
$x_1 = p_{c\,\max}$, $x_2 = p_{c\,\min}$, $x_3 = p_{i\,\max}$, $x_4 = p_{i\,\min}$, $x_5 = p_{c\_Q\,\max}$, p – classification accuracy.

scriptors and the number of engine cycles (training data).

The highest recognition accuracy levels were achieved with variables $x_1 = p_{c\max}$ and $x_5 = p_{c\_Q\max}$ used as in-cylinder pressure discriminators and with the number of cycles $n = 25$. The 100% recognition was achieved with Model 5 using RandomForest.

The proposed procedures (methods) are universal and can be used to identify fuel in any engine. But the values of the parameters in the models (pressure descriptors, the number of cycles from which the training data was averaged) should be adapted to the engine under analysis.

References

1. Albarbar A., Gu F., Ball A.D., Starr A. (2010), *Acoustic monitoring of engine fuel injection based on adaptive filtering techniques*, Applied Acoustics, **71**, 12, 1132–1141.
2. Ambrozik A., Ambrozik T., Kurczyński D., Łagowski P. (2014), *The influence of injection advance angle on fuel spray parameters and nitrogen oxide emissions for a self-ignition engine fed with diesel oil and FAME*, Polish Journal of Environmental Studies, **23**, 6, 1917–1923.
3. Barelli L., Bidini G., Buratti C., Mariani R. (2009), *Diagnosis of internal combustion engine through vibration and acoustic pressure non-intrusive measurements*, Applied Thermal Engineering, **29**, 1707–1713.
4. Bąkowski A., Radziszewski L. (2015), *Determining selected diesel engine combustion descriptors based on the analysis of the coefficient of variation of in-chamber pressure*, Bulletin of the Polish Academy of Sciences: Technical Sciences, **63**, 2, 457–464.
5. Breiman L. (2001), *Random forests*, Machine Learning, **45**, 1, 5–32.
6. Brzozowski K., Nowakowski J. (2014), *Model for calculating compression ignition engine performance*, Maintenance and Reliability, **16**, 3, 407–414.
7. Chiatti G. *et al.* (2015), *Diagnostic methodology for internal combustion diesel engines via noise radiation*, Energy Conversion and Management, **89**, 34–42.
8. Delvecchio S., Bonfiglio P., Pompoli F. (2018), *Vibro-acoustic condition monitoring of Internal Combustion Engines: A critical review of existing techniques*, Mechanical Systems and Signal Processing, **99**, 1, 661–683.
9. Deptuła A., Kunderman D., Osiński P., Radziwanowska U., Włostowski R. (2016), *Acoustic diagnostics applications in the study of technical condition of combustion engine*, Archives of Acoustics, **41**, 2, 345–350.
10. Elghamry M.H. *et al.* (1998), *Gaseous fuel quality identification for a spark ignition gas engine using Acoustic Emission analysis*, [In:] 11th International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management COMADEM 1998, pp. 235–244.
11. Engelbrecht A.P. (2006), *Fundamentals of computational swarm intelligence*, John Wiley & Sons.
12. Figlus T. *et al.* (2014), *Condition monitoring of engine timing system by using wavelet packet decomposition of an acoustic signal*, Journal of Mechanical Science and Technology, **28**, 5, 1663–1671.
13. Flekiewicz M., Flekiewicz B., Fabiś P. (2007), *Engine block vibration level as a tool for fuel recognition*, SAE Technical Paper, No. 2007-01-2162.
14. Grajales J.A., Quintero H.F., López J.F., Romero C.A., Henao E., Cardona O. (2017), *Engine diagnosis based on vibration analysis using different fuel blends*, Diagnostyka, **18**, 4, 27–36.
15. Gravalos I. *et al.* (2013), *Detection of fuel type on a spark ignition engine from engine vibration behaviour*, Applied Thermal Engineering **54**, 171–175. doi: 10.1016/j.applthermaleng.2013.02.003.
16. Hardenberg H.O., Hase F.W. (1979), *An empirical formula for computing the pressure rise and delay of a fuel from its cetane number and from the relevant parameters of direct-injection diesel engines*, SAE Technical Paper, No. 790493.
17. Kekez M., Radziszewski L. (2011), *Modelling of pressure in the injection pipe of a diesel engine by computational intelligence*, Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, **225**, 12, 1660–1670.
18. Kekez M., Radziszewski L., Sapietova A. (2016), *Fuel type recognition by classifiers developed with computational intelligence methods using combustion pressure data and the crankshaft angle at which heat release reaches its maximum*, Procedia Engineering, **136**, 353–358.
19. Lowe D.P., Lin T.R., Wu W., Tan A.C.C. (2011), *Diesel knock combustion and its detection using acoustic emission*, Journal of Acoustic Emission, **29**, 78–88.
20. Maurya R.K. *et al.* (2013), *Digital signal processing of cylinder pressure data for combustion diagnostics of HCCI engine*, Mechanical Systems and Signal Processing, **36**, 95–109.
21. Morón-Villarreyes J.A., Soldi C., de Amorim A.M., Pizzolatti M.G., de Mendonça Jr, A.P., D'Oca M.G. (2007), *Diesel/biodiesel proportion for by-compression ignition engines*, Fuel, **86**, 12–13, 1977–1982.
22. Payri F., Luján J.M., Martín J., Abbad A. (2010), *Digital signal processing of in-cylinder pressure for combustion diagnosis of internal combustion engines*, Mechanical Systems and Signal Processing, **24**, 1767–1784.
23. Pietraszek J., Gądek-Moszczak A., Radek N. (2014), *The estimation of accuracy for the neural network approximation in the case of sintered metal properties*, Studies in Computational Intelligence, **513**, 125–134.

24. Ranachowski Z., Bejger A. (2005), *Fault diagnostics of the fuel injection system of a medium power maritime diesel engine with application of acoustic signal*, Archives of Acoustics, **30**, 4, 465–472.
25. Ruiz F.A., Isaza C.V., Agudelo A.F., Agudelo J.R. (2017), *A new criterion to validate and improve the classification process of LAMDA algorithm applied to diesel engines*, Engineering Applications of Artificial Intelligence, **60**, 117–127.
26. Sakthivel G. (2016), *Prediction of CI engine performance, emission and combustion characteristics using fish oil as a biodiesel at different injection timing using fuzzy logic*, Fuel, **183**, 214–229.
27. Szymański G.M., Tomaszewski F. (2016), *Diagnostics of automatic compensators of valve clearance in combustion engine with the use of vibration signal*, Mechanical Systems and Signal Processing, **68–69**, 479–490.
28. Teixeira L.S.G. *et al.* (2008), *Multivariate calibration in Fourier transform infrared spectrometry as a tool to detect adulterations in Brazilian gasoline*, Fuel, **87**, 346–352.
29. Valencia F.A., Armas I.P. (2005), *Ignition quality of residual fuel oils*, Journal of Maritime Research, **2**, 3, 77–96.
30. Witten I.H., Frank E. (2005), *Data mining*, Morgan Kaufmann.
31. Yoon M. *et al.* (2007), *A method for combustion phasing control using cylinder pressure measurement in a CRDI diesel engine*, Mechatronics, **17**, 9, 469–479.