Chaos theory-based time series analysis of in-cylinder pressure and its application in combustion control of SI engines

Huanyu DI*, Yahui ZHANG* and Tielong SHEN*
*Department of Engineering and Applied Sciences, Sophia University
7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554, Japan
E-mail: zhangyahui@eagle.sophia.ac.jp

Abstract
Combustion control is a significant topic for achieving high efficiency and low emissions of internal combustion engines. Recently, in-cylinder pressure sensor-based closed-loop control strategies have become the preferred solution. However, their practical applications in automotive industries are limited due to the intensive acquisition of pressure series for a whole cycle and subsequent calculation of combustion indicators. This paper proposes a method for in-cylinder pressure information extraction and combustion phase estimation of spark ignition (SI) engines based on pressure measurements at several points coordinated by the crank angle. First, nonlinear dynamics analysis is introduced to analyze the system of in-cylinder pressure evolution, which is proved to be a deterministic nonlinear dynamic system with chaotic characteristics. Then, a 3-dimensional system state variable is determined to replace the pressure series during combustion. Second, with the determined system state variable, the in-cylinder pressure series during combustion and the combustion phase are learned and estimated by a machine learning method, namely, extreme learning machine (ELM). As a result, only pressure measurements at 3 points and ELM estimation models are required, instead of intensive data acquisition and calculation. The experimental validations carried out on a gasoline engine test bench have proved that the reconstruction and estimation results are accurate and that the method can perform well in real-time combustion control.

Keywords: Nonlinear dynamics analysis, State reconstruction, Combustion phase estimation, In-cylinder pressure, Extreme learning machine, Spark advance control

1. Introduction
For internal combustion engines, one of the effective factors that affect efficiency and pollution emissions is the combustion phase represented by CA50, which is defined as the crank angle when 50% of the fuel has been burned (Farag et al., 2017). In spark ignition (SI) engines, the combustion phase is determined by spark advance (SA) angle (Hepkaya et al., 2014). In most production engines, the SA parameter is controlled in an open-loop by means of maps stored in electronic control units (ECUs). Such kinds of control cannot guarantee that an engine will operate at its best performance since the optimal combustion phase is influenced by many variables, including environment conditions, fuel quality, and engine aging (Enrico and Forte, 2010). On condition that in-cylinder pressure contains sufficient information, it can be used to calculate the combustion phase (Hibi et al., 2013; Ceviz et al., 2011). Therefore, many closed-loop sparking timing control methods have been proposed based on the measurement of in-cylinder pressure (Gao et al., 2017). For example, feedback control is applied to regulate the combustion phase to a certain range (Pipitone, 2014). The extremum seeking (ES) method is used to calibrate the SA optimal point online (Hellström et al., 2013). However, the precise pressure sensor, numerous collected data, and complex calculation method hinder those methods when applied to production engines. Efforts should be made to reduce data collection and to simplify the calculation process. A possible solution is to select a few points to represent the whole pressure curve. The selection method is determined by the analysis of in-cylinder pressure data series.
Nonlinear dynamics analysis, especially chaos theory, can be used to analyze data series with internal relationships. For example, a wind power time series has been analyzed by chaos theory, and a kernel function prediction model based on the chaos analysis is introduced (Ouyang et al., 2016). Chaos theory is also applied to traffic flow information analysis, and short-term traffic speed forecasting is completed by the support vector machine model (Wang et al., 2013).

Chaos theory was introduced to engine research in the 2000s. Period doubling bifurcations, one of the typical chaos phenomena, are observed in homogeneous charge compression ignition (HCCI) engine combustion. Then based on a near chaos dynamic model, combustion timing has been predicted successfully (Lee et al., 2009). Li apply Poincaré section and return map to show the chaos dynamics of the in-cylinder pressure of a lean-burn natural gas engine (Li et al., 2008). And Matsumoto use Pyragas’ method, which is an effective method to control chaos dynamic systems, to reduce the low-frequency oscillations of power output in a motor engine (Matsumoto et al., 2008). However, experimental evidence has shown that both random and chaotic features spark can affect ignition engines. This ambiguity therefore leads to a continuing debate on the truth of state evolution in the cylinders of SI engines (Daw et al., 1998).

In this paper, the evolution of in-cylinder pressure series is reconstructed by the mutual information (MI) method and Grassberger-Procaccia (G-P) approach. A 3-dimensional vector consisting of three moments of pressure data is applied to describe the system state.

The nonlinear dynamic analysis indicates that each reconstructed data point contains the nearby evolution information. It is reasonable to predict the transformation of system state based on the evolution information. Using a proper reconstructed system state point, the in-cylinder pressure during combustion is estimated by an extreme learning machine (ELM). This approach indicates that one proper state point contains the information of in-cylinder pressure during combustion, making it possible to use one reconstructed system state point of one cycle to estimate the combustion phase (CA50).

The estimation of CA50 is achieved by ELM. This estimation is verified under different operating conditions. The cycle-to-cycle varied CA50 is precisely estimated according to three moments of pressure data by learning results of the ELM method. In addition, a control experiment is undertaken as an application example. The estimated CA50 is applied to the extremum seeking (ES) method to control SA to a high efficiency point. The control result is contrasted with the situation when calculated CA50 is applied. This contrasting result shows that this estimation of CA50 can replace the complex calculations (Appendix A) in real time control.

The rest of this paper is organized as follows. In section two, the engine test bench and design of experiments are introduced. The in-cylinder pressure is reconstructed and analyzed in section three. The Lyapunov exponent of in-cylinder pressure is calculated, and surrogate test is introduced. In section four, ELM model is introduced and then utilized for estimation of in-cylinder pressure during combustion and combustion phase. Section five shows experimental validations of the proposed reconstruction and estimation method in SA ES control. Final conclusions are drawn in section six.

2. Experimental setup

Given the fact that the commonly used SI engines of gasoline-fueled automobiles employ the four-stroke Otto cycle, an SI engine is also applied in this study.

To analyze nonlinear dynamic of SI engines, the experiment is designed based on the test bench, in which an SI gasoline engine is coupled to a HORIBA low inertial alternative current (AC) dynamometer. The dynamometer has two operational modes, which are mode with constant load or constant rotate speed. In this paper the constant rotate speed mode is selected. The engine speed is controlled by the SPARC controller of dynamometer.

The test bench control and measurement system include an engine electronic control unit, dynamometer SPARC controller, and a dSPACE1006 multi-processor system connected to a personal computer. In addition, a dSPACE2004 highspeed data acquisition card is used to collect the submillisecond pressure data. Fig. 1 shows the test bench configuration. The engine parameters are tabulated in Table 1.

As shown in Fig. 2, the experiment is conducted under several operating conditions: the engine speed is 1200 rpm (revolutions per min); the ARF (air fuel ratio) is approximately 14.6; the ignition angle SA (spark advance) ranges from 7 to 22 degrees BTDC (before top dead center); and the TA (throttle angle) ranges from 5.7 to 8.8 degrees. During the experiment, the in-cylinder pressure is measured against the crank angle from 0 to 719 degrees in each cycle. The engine is operated under 6 different throttle angles, and under each condition, the SA is changed from 10 to 13 deg. BTDC.

According to the measured data, in-cylinder pressure varies from cycle to cycle. Those kinds of variations appear after 300 degrees and before 450 degrees of crank angle, which is almost the same as the phase of in-cylinder combustion.
It is clear that this variation is introduced by in-cylinder combustion. This range of in-cylinder pressures is analyzed.

3. Nonlinear dynamics of in-cylinder pressure

In this section, the nonlinear dynamics of in-cylinder pressure series are analyzed. In the field of internal combustion engine engineering, Wiebe function is widely used to formulate the process of in-cylinder combustion. However, the linear dynamics formulated by Wiebe function cannot totally match the experiment measurements. A brief analysis from the view of physical mechanism is introduced.

To analyze the system dynamic of in-cylinder combustion, the system state need to be ascertained at first. As an usual practice, the system is described by crank angle $\theta$ instead of time $t$. The most important dynamic state is the heat release $Q$. The heat release amount and rate can represent the combustion state directly.

The heat release rate of the combustion is determined by the reaction rate. According to the chemical kinetics, the reaction rate is related to the number of particles which get over the reaction barrier, which is called excited molecule.

\[
\frac{dQ}{d\theta} \propto N(u_i > \Delta E_b)
\]  (1)
where \( u_i \) is the energy of the unreacted particles, and \( \Delta E_b \) is its reaction barrier. In the chemical kinetics theory, when the energy of particles overcome the reaction barrier, the particle is called excited particle. The energy of each particles obeys normal distribution. The mean of particle energy \( \bar{u}_i \) is related to the internal energy of the in-cylinder gas. From the view of macroscopic, the number of the excited molecule can be represented.

\[
\frac{dQ}{dt} \propto N(u_i > \Delta E_b) = C(\text{unburned}) p(u_i > \Delta E_b) V(\text{burning})
\]

(2)

So the heat release rate is related to three important index, which is the concentration of the unburned gasoline molecule, the internal energy of the in-cylinder gas, and the volume of the burning zone.

\[
\frac{dQ}{dt} = (Q_t - Q_1) f_1(Q) f_2(S)
\]

(3)

where \( Q_t \) is the total heat release of this cycle, \( f_1 \) and \( f_2 \) are unknown function, \( S \) is the flame surface which indicate the volume of burning zone. The value of \( S \) can not be measured in the experiment, which makes the evolution of in-cylinder combustion can not easily formulated. And the complexity encourages the researchers to introduce the methods of machine learning(Kobayashi et al., 2019). According to the physical analysis of in-cylinder combustion, nonlinear dynamics analysis is introduced, and nonlinear measurements such as correlation dimension and Lyapunov exponents are calculated in this section.

3.1. Form of reconstructed data

Regarding the SI engine, the evolution of the in-cylinder environment during combustion is a multidimensional system, but the measured in-cylinder pressure is a one-dimensional series. Hence, before the analysis of the nonlinear dynamics, the phase space of the measured data needs to be reconstructed into a proper dimension coincident with its physical essence. It has been proved that any variable time series of a given system contains the dynamic information of the original system (Packard et al., 1980). According to embedding theory, the measured in-cylinder pressure data can be reconstructed into a \( d \)-dimensional series \( \{x_j\} \) as follows:

\[
x_j = \{p_j, p_{j-\tau}, \ldots, p_{j-(m-1)\tau}\}^T
\]

(4)

where \( \{p_j\} \) is the measured in-cylinder pressure at crank angle \( j \), \( m \) is the dimension of the new reconstructed state vector \( x_j \), and \( \tau \) is the delay time. Parameters \( m \) and \( \tau \) need to be selected. Based on embedding theory, the dynamic behaviors of the original and reconstructed systems are equal. However, the selection is not arbitrary because a proper selection of \( m \) and \( \tau \) will benefit extraction of the nonlinear dynamics of the system.

3.2. Determination of embedding interval

There are some mature methodologies for selecting embedding interval according to different concepts. In this paper, the mutual information (MI) method is applied. The mutual information function is defined as

\[
I(\tau) = \sum_{j=1}^{N} P(p_j, p_{j+\tau}) \log \frac{P(p_j, p_{j+\tau})}{P(p_j)P(p_{j+\tau})}
\]

(5)
Mutural information

where $P(\cdot)$ is the probability density function; $\log(\cdot)$ is the logarithmic function. In information entropy theory (Shannon, 1948), the information entropy of series $p_j$ is defined as $G(p_j) = -\sum_j P(p_j) \log P(p_j)$. Information entropy represents the amount of information contained in the distribution of the data set. The larger the information entropy, the more uncertain the distribution of the data set. If the distribution function of $p_{j+\tau}$ is a $\delta$-function when $p_j$ is ascertained, which means that the value of $p_{j+\tau}$ is definitive, the information entropy reaches its minimum, 0. Defining series $p_j$ as $s$ and series $p_{j+\tau}$ as $q$, the mutual information function can be rewritten as

$$I(\tau) = G(s) + G(q) - G(s, q) = G(q) - G(q,s),$$

where $G(q,s)$ represents the information entropy of series $q$ when the series $s$ is given. The mutual information function $I(\tau)$ is defined as the difference between the information entropy of sequence $q$ and the information entropy of sequence $q$ when sequence $s$ is determined. As a result, the function $I(\tau)$ represents information shared by series $p_j$ and $p_{j+\tau}$.

Fig. 3 shows the relationship between mutual information $I(\tau)$ and time interval $\tau$. When time interval $\tau$ is too small, the mutual information of the $p_j$ and $p_{j+\tau}$ series is too large. When $p_j$ and $p_{j+\tau}$ are selected as two axes of state vector $x_j$, the information entropy of $x_j$ is too small. In addition, if the value of $p_{j+\tau}$ is close to the value of $p_j$, the system state point $x_j$ will only evolve around the diagonal of the system space. This phenomenon will lead to an unconvincing analysis of the nonlinear dynamics of the reconstructed system. When time interval $\tau$ is too large, $p_j$ and $p_{j+\tau}$ are almost independent, and they are not suitable for selection into a single system state $x_j$. Consulting previous works, in this paper, the time interval $\tau$ is selected to cause the mutual information $I(\tau)$ to reach its first minimum value. As shown in Fig. 3, the time interval $\tau$ is selected as 20 degrees of crank angle.

3.3. Determination of embedding dimension

The embedding dimension $m$ is another important parameter to be determined. In this paper, the G-P approach presented by Grassberger and Procaccia is used to select the embedding dimension. Constrained by the state function, the state point only evolves in part of the whole space; that part can be equivalent to a special dimension space, the dimension of which is called the correlation dimension $d$. The correlation dimension of a chaos dynamic system is always a noninteger. In the G-P approach, the correlation dimension $d$ and the proper embedding dimension $m$ are estimated simultaneously.

Before applying the G-P approach, the accumulative distribution function $C(r)$ is defined as

$$C(r) = \frac{1}{N(N-1)} \sum_{i,j=1 \land i\neq j}^{N} H(r - \|x_j - x_i\|),$$

where $\|x_j - x_i\|$ is the distance between the two state points $i$ and $j$ in the state space. $N$ is the total number of state points. $H(z)$ is the Heaviside function:

$$H(z) = \begin{cases} 0 & z < 0 \\ 1 & z \geq 0 \end{cases}$$

© 2020 The Japan Society of Mechanical Engineers
The accumulative distribution function $C(r)$ denotes the probability that the distance between two random points is less than $r$. If $\| x_j - x_i \| < r$, point $\{x_i\}$ is called the neighbor point of $\{x_j\}$. $C(r)$ represents the average proportion of neighbor points of each state point $\{x_j\}$.

According to topology theory, for some range of neighbor scale $r$, the state points are separated in the correlation dimension in the reconstruction space, and the average proportion of neighbor points is proportional to the measure of a neighborhood space ball in the space described by correlation dimension.

$$C(r) \propto r^d; \quad (9)$$

where $r^d$ is proportional to the measure of a neighborhood space ball in $d$-dimension space. Thus, for a different embedding dimension $m$, the corresponding correlation dimension $d(m)$ can be calculated according to logarithmic linear fitting:

$$d(m) = \frac{\ln C(r) - C_0(m)}{\ln r}. \quad (10)$$

The G-P approach is undertaken by calculating the corresponding correlation dimension $d(m_0)$ given $m_0 = 1$ and then increasing the embedding dimension $m$ until the correlation dimension $d(m)$ converges to a constant. If the correlation dimension $d(m)$ does not converge, then the given series is random.

As Fig. 4 shows, the embedding dimension $m$ is calculated from 1 to 5. It can be seen that the line slope converges when the embedding dimension $m$ is 3. Therefore, to simplify calculation, the embedding dimension $m$ is selected as 3. The corresponding correlation dimension $d(3) = 1.237$. The correlation dimension is non-integer and much lower than the embedding dimension, which is a main character of chaos dynamic system.

As the previous part of this section shown, the measured data are reconstructed into a 3-dimensional state space.

$$x_j = [p_j; p_{j-20}; p_{j-40}]^T \quad (11)$$

In this reconstructed system state space, the state points evolve, as Fig. 5 shows.

3.4. Calculation of the Lyapunov exponent

In mathematics the Lyapunov exponent of a dynamical system is a quantity that characterizes the rate of separation of infinitesimally close trajectories (Boeing, 2016). The positive Lyapunov exponent represents a phenomenon that a small disturb will lead to the state evolution diverge in this dimension, and the negative one represents convergence in corresponding dimension. If all Lyapunov exponents of the system is negative, a small disturb will disappear during the evolution, which is a character of a non-chaotic system. And if all Lyapunov exponents is positive, a small disturb will change the evolution rules even in a short term, which is a symbol of randomness system.

According to the Lyapunov exponent calculation method provided by M. Sano in (Sano, 1985), the Lyapunov exponent can be calculated based on the series of reconstructed measured data. By linearizing the evolution of distance
between two nearby reconstructed points, the transfer matrix of separation of infinitesimally close trajectories can be estimated. Then, the Lyapunov exponent can be measured based on this transfer matrix. The calculation is provided in detail in the previous work (Di et al., 2018).

The calculated Lyapunov exponents of the in-cylinder pressure data are

\[ \lambda = (0.0085, -0.0558, -0.0597) \]  

(12)

There are two negative and one positive Lyapunov exponents. According to the chaos theory, owning at least one positive Lyapunov exponent is a necessary condition of chaos dynamic system. The in-cylinder pressure evolution system satisfies this necessary condition. From view of this necessary condition, the calculated Lyapunov exponent implies that the system might have chaotic property. However, compared with other classical chaos dynamic system, the Lyapunov exponent is small in magnitude. To further study of the in-cylinder evolution system, surrogate test is introduce.

### 3.5. Surrogate Test

The method of surrogate test is applied to validation of determinacy of the in-cylinder pressure evolution. This method is widely used in the analysis of experimental time series and provides a powerful tool in the search for determinism in pseudo-periodic data, such as the in-cylinder pressure data in this paper. This method may be applied to test against the null hypothesis of a periodic orbit with uncorrelated noise in the very large number of experimental systems that exhibit pseudo-periodic behavior (Small et al., 2001).

Surrogate methods proceed by comparing the value of (nonlinear) statistics for the data and the approximate distribution for various classes of linear systems and by doing so one can test if the data has some characteristics which are distinct from stochastic linear systems. This method is widely used in the analysis of experimental time series and provides a powerful tool in the search for determinism in apparently stochastic data by comparing the processing result of the original and the surrogate data series.

According to the original measured time series \( \{x_t\}_{t=1}^N \), the following scheme is employed to construct a surrogate data series. (1) Randomly choose an initial condition \( s_1 \), where \( s_1 \in \{x_t : t = 1, \ldots, N\} \). Set \( i = 1 \). (2) For \( s_i \) choose a neighbor \( s_j \in \{x_t : t = 1, \ldots, N\} \) with probability.

\[
\text{Prob}(s_j = x_t) \propto \exp\left(-\frac{\|x_t - s_j\|}{\rho}\right).
\]  

(13)

The parameter \( \rho \) is the noise radius and will be discussed latter. (3) Set \( s_{i+1} = s_{j+1} \) and then increase \( i \). (4) Repeat this procedure from step (2) until \( i = N \). The noise radius \( \rho \) is selected to maximum the number of short segment (of length 2) that are the same for the original measured data and the surrogate data. As shown in Fig. 6, the short segment number is counted in each noise radius \( \rho \). The short segment number reaches its maximum value when \( \rho = 0.0158 \).

Now that the parameter \( \rho \) is selected, 15 groups of surrogate data series \( \{s_j\} \) are constructed. One group of this surrogate data series are illustrated in Fig. 7. Compared with Fig. 5, the surrogate and original data series have similar appearance.
In the analysis of surrogate method, the evolution of surrogate data series are contaminated by the stochastic noise. If the original data itself is stochastic, the original data series can not be distinguished form the surrogate data series in the correlation dimension map. The correlation dimension of measured data series and the 15 groups of surrogate data series are calculated. As shown in the Fig. 8, the correlation curve of original data series is far from the curves of surrogate data series, although the vector field of surrogate data series is similar with that of original data series. Therefore, the dynamic of in-cylinder combustion is deterministic nonlinear dynamics, but not stochastic. Combined with the calculation result of Lyapapov exponents and the physical mechanism analysis, we can conclude that this in-cylinder pressure evolution system is a deterministic nonlinear dynamic system with chaotic characteristics.

4. Estimation of CA50 based on chaos dynamics

According to the nonlinear dynamic analysis, there exists essential nonlinear dynamic during in-cylinder combustion. It means that the system state contains the evolution information. From another perspective, the system state can be regarded as the information extraction of the combustion process. Hence, the combustion parameters, for example CA50, can be estimated based on the state point. However, as the analysis in section 3, the physical relationship between the state points and combustion parameters can not be easily described. As a result, a machine learning method, namely, extreme learning machine (ELM), is employed to estimate the in-cylinder pressure during combustion and the combustion phase CA50.

4.1. Extreme learning machine model

The relationships between the inputs (the proper system state point in Eq. (11) and the output (in-cylinder pres-
sure trace or CA50) are unknown, may be highly complex, and are difficult to be expressed accurately and analytically. Machine learning methods are suited when performing modeling of such relationships. A feed-forward neural network, which is a commonly used model structure for machine learning, is shown in Fig.9. It consists of three layers: input, feature mapping, and output. Each input and each output are considered as a node or a unit. The machine learning process is used to adjust the weights between nodes. In other words, what the model learns is contained in these weights.

Extreme learning machine (ELM) is a state-of-the-art machine learning method first proposed by Huang et al. (Huang et al., 2006) to improve feedforward neural network learning speeds. The basic idea of ELM is that the input weights and feature mapping layer biases need not be tuned but can be randomly assigned instead and that the feature mapping layer can be represented by either random hidden nodes or kernels. After the input weights and the feature mapping layer are chosen randomly, the model can be simply considered as a linear system, and the output weights of the model can be analytically determined through a simple generalized inverse operation of the feature mapping layer output matrices.

In this work, the ELM model is set as a multi-dimension to one-dimension function:

$$ f(u) = h^T(u)w = \sum_{j=1}^{n_h} h_j(a_j^T u + b_j) \cdot w_j $$  \hspace{1cm} (14)

where $u$ is the input vector, $a_j$ is the input weights matrix, $b_j$ is the feature mapping layer biases vector, $w_j$ is the output weight vector. The feature mapping function in this application is the commonly used sigmoid function:

$$ h(t) = \frac{1}{1 + e^{-t}} $$  \hspace{1cm} (15)
In the ELM model, \( a_j \) and \( b_j \) are determined randomly, and \( w_j \) is optimized according to the training set. The output weight \( w \) can be solved as follows.

Consider the following training data set for ELM:

\[
D = \{(u(1), o(1)), ..., (u(i), o(i)), ..., (u(N), o(N))\},
\]

where \( N \) denotes the total number of training samples. The matrix of input set is denoted as \( U = [u(1), ..., u(N)] \), correspondingly the vector of output set is denoted as \( O = [o(1), ..., o(N)] \).

According to the ELM model:

\[
Hw = O,
\]

where the matrix \( H = \{h(u_1), ..., h(u_N)\} \), the vector \( h(\cdot) = [h_1(\cdot), ..., h_m(\cdot)]^T \). The smallest norm least squares solution of \( w \) to the above linear system can be expressed as follows:

\[
w = H^+ O,
\]

where \( H^+ \) is the Moore-Penrose pseudoinverse of matrix \( H \).

Finally, put the optimal output weight vector to the ELM model, the ELM model shown in Fig.9 is expressed as a combination of feature mapping functions.

### 4.2. ELM-based CA50 estimation

Note that the restructured system state \( x_{390} \) contains transition information of the in-cylinder system during combustion, and it is reasonable to estimate the CA50 in unit of deg. after top dead center (ATDC) by the system state \( x_{390} \). In addition, if the calculated load of that estimation is lower than that of the calculation according to the definition, the estimation has potential for application in real-time control. Based on this assumption, a model from state \( x_{390} \) to CA50 is useful.

\[
CA50 = f(x_{390})x_{390} = [p_{390}, p_{370}, p_{350}]^T.
\]

The relationship function \( f(\cdot) \) between the state point and the corresponding CA50 is learned by ELM. In this application, the ELM model uses 50 hidden nodes. The data set of 7100 combustion cycles employed in in-cylinder pressure estimation is also utilized to train and test the ELM model of CA50. In the training set, the input is \( x_{390} \) and the target output is the conventional calculated CA50 which is obtained from the heat release profile (Appendix A). Overall, 90% of the data is for training, while the remainder is used for testing. Note that the training and testing root mean square errors (RSME) are 0.202 and 0.205, respectively.

This ELM model has 3 input dimensions, which are the in-cylinder pressures at 390, 370 and 350 degrees. Using the learned ELM model, the CA50 values of the test set are estimated, as shown in Fig. 10. The estimated CA50 is accurate in comparison with the measured one, and the calculation load is much lower than for the original measurement. Therefore, this estimation can be used in real-time control.

### 4.3. Online validation of the ELM model

Embedding the learned estimation ELM model to dSPACE1006 multi-processor system, the CA50 of each cycle can be estimated online. The learned ELM model is aimed at the operating condition of engine speed at 1200rpm, and the throttle angle form 5.7 to 8.8. This method can be applied in all operating condition theoretically. To confirm the estimation result in the real time experiment, a verification experiment is undertaken, as Fig. 11(a) shows. The engine is operated at a fixed speed and throttle angle. The engine speed is 1200 rpm. The ARF is approximately 14.6. The throttle angle is approximately 6.5 deg.. The ignition angle SA signal is stepped from 11 to 19 deg. BTDC. The conventional CA50 calculation and the CA50 estimation results are shown in Fig. 11(b). The estimated and calculated CA50 are similar regardless of how the SA changes.

To further confirm the estimation, the engine is operated under a fix speed and varying throttle angle, as Fig. 12(a) shows. The engine speed is 1200 rpm. The ARF is approximately 14.6. Throttle angle is varied from 6 to 8 degrees. Ignition angle SA is controlled by the production ECU from the Toyota Corporation. The estimated and calculated CA50 are shown in Fig. 12(b). It is confirmed that during both the transient process and state process, the estimation result is believable. Under the combustion control of the production ECU, the CA50 is controlled at approximately 6 deg. ATDC.
Due to the turbulence of the intake gas and the fuel distribution before ignition, the CA50 varies from 2 to 10 degrees from cycle to cycle. The RMSE of this estimation is 0.362; it is accurate compared to the varying range of CA50.

When the estimation model is applied to calculate the combustion phase CA50, the data collection is reduced to 3 points. However, in the original method, the pressure at each crank angle needs to be collected. In engineering application, a sensor measuring 3 points in each cycle will be much less expensive than a sensor collecting every point in each cycle. The estimation model has the ability to replace the conventional CA50 calculation method (Appendix A) in real-time control.

5. CA50 estimation-based extremum seeking control of SA

Extremum seeking (ES) is a gradient-based iterative optimization method that has proven to be both effective and robust in extremum determinations of static maps and dynamic systems (Zhang et al., 2018). Recently, applications of the ES methods in engine calibration and control have been investigated, such as SA control to obtain maximal thermal efficiency. In this section, ES control of SA is performed based on the presented CA50 estimation, and compared with ES control based on the conventional CA50 calculation (Appendix A).

The problem formulation of SA ES control is described detailedly in (Zhang and Gao, 2018). For completeness of this paper, a brief introduction of the problem formulation is given here. As it is well known, thermal efficiency, $\eta$, is significantly affected by CA50. The following statistical model is employed to represent the relationship, which is an unknown convex function, $f_{c}(\cdot)$, coupled with a Gaussian distributed random noise, $\omega$:

$$\eta = f_{c}(CA50) + \omega, \quad \omega \sim N(0, \sigma^2).$$  \hspace{1cm} (20)
Moreover, the randomness can also be observed in the relationship between SA and CA50. The property is represented statistically as a linear function coupled with a Gaussian distributed random noise, \( d \):

\[
CA50 = -a \cdot SA + b + d, \quad d \sim N(0, \sigma^2),
\]

where \(-a \cdot SA + b\) represents the linear negative correlation between SA and CA50. From statistical point of views, for a given SA, the CA50 exhibits a Gaussian distribution with a mean \(-a \cdot SA + b\).

Optimal SA where maximal mean thermal efficiency can be achieved exists in Eq.(20) and Eq.(21), which is shown in Fig.13 for experimental data. SA ES control problem can be finally described as follow:

\[
SA^* = \arg \max_{SA} E\{\eta[CA50(SA, d), \omega]\}.
\]

where \(E\{\cdot\}\) denotes the expected value.

Several categories of ES control methods, for example, stochastic approximation-based, sinusoid-based, and natural excitation-based, can be used to solve the problem Eq.(22) (Zhang et al., 2018). In this paper, the most recently presented ES control scheme, namely, natural excitation-based ES, is performed, as shown in Fig.14. It takes advantage of the joint sample distribution of CA50 and thermal efficiency \(\eta\).

The most important part of the natural excitation-based ES method is the gradient estimation. The \((CA50, \eta)\) sample distribution is utilized to approximate the gradient in SA ES control. As shown in Fig.13(b), even for a fixed SA value, the \((CA50, \eta)\) samples form a stochastic distribution, namely, natural excitation, which implies that SA should be advanced or retarded. The gradient will be estimated by using forgetting factor least squares method for the recent past samples. Based on the estimated gradient, adjustment of the previous SA value will be:

\[
\Delta SA(k) = -\Gamma \cdot \hat{g}(k) \Rightarrow SA(k) = SA(k-1) + \Delta SA(k),
\]

where \(k\) represents the combustion cycle, \(\hat{g}(k)\) and \(\Gamma\) represent the gradient estimation and the step size, respectively.

Experimental validations of the presented CA50 estimation-based and the conventional CA50 calculation-based SA ES control have been carried out respectively. The experiment is done under the same condition, in which the engine speed is 1200 rpm, the throttle angle is 6.5 deg., the ARF is approximately 14.6, and the initial ignition angle SA is 12 deg. BTDC. In this experiment, the engine speed are controlled by the Dynamometer to fix the engine operating condition. The fluctuation of engine speed and engine torque is very small, because only spark advance is controlled. Therefore, the influence of operating condition changing is neglected in this paper.

The equilibrium provided by the dynamometer fluctuate from 40 Nm to 40.5 Nm because of the optimal control of SA. This slight change will not effect the experiment result from views of engineering. As shown in Fig. 15, the SA is controlled to the optimal SA value, which is approximately 14.7 deg.BTDC. One can see that the CA50 estimation-based ES performs as well as the conventional calculation-based ES. It can be concluded that the CA50 estimation is accurate and able to used in practical combustion control.
6. Conclusions

This paper has focused on a nonlinear dynamics analysis of in-cylinder pressure series during one combustion cycle of SI engines and its applications in real-time combustion control. It is shown that the in-cylinder pressure evolution is a deterministic nonlinear dynamic system with chaotic characteristics, which makes it possible to subsequently develop a method to estimate the pressure series during combustion and combustion phase based on pressure measurements at only several points. As a result, a 3-dimensional system state variable can be determined to replace the pressure series since the state variable contains the evolutionary information of the pressure dynamics. With the determined state variable, CA50 is learned and estimated by ELM. The advantage is that only pressure measurements at 3 points are required. Hence, it is shown that the effort of pressure data acquisition and calculation of CA50 is significantly saved. The accuracy of ELM estimation are verified by experiments. Moreover, an application of SA extremum seeking control demonstrates that the
estimation-based combustion control strategy performs as well as the conventional calculation-based one. The proposed efficient pressure data acquisition and combustion phase estimation method shows great potential in practical applications on on-board engine ECUs.

Appendix A

CA50 is an important parameter of in-cylinder combustion. CA50 is defined as the crank angle when 50% of the fuel is burned, which is a indicator of combustion phase. The CA50 of the current cycle is calculated according to the definition in the real-time control, which needs sufficient calculation ability to perform the differential and integral operation. The heat release rate is calculated based on the assumption of an ideal gas and isolated process.

\[
dQ = dU + pdV = \frac{1}{1-\kappa}nRdT + pdV = \frac{1}{1-\kappa}d(pV) + pdV = \frac{\kappa}{1-\kappa}pdV + \frac{1}{1-\kappa}Vdp, \tag{24}
\]

where \(Q\) is the combustion heat release, and \(U\), \(p\), \(\kappa\), and \(V\) are the internal energy, cylinder pressure, polytropic exponent, and volume of in-cylinder gas, respectively. According to the definition,

\[
CA50 = \theta[\theta(Q(\theta) = \frac{1}{2}Q_t)], \tag{25}
\]

where \(Q_t\) is the total heat release, \(\theta\) is the crank angle, which varies from 0 to 719 degrees for each cycle, as shown in Fig.16(a). For convenience, CA50 is usually represented in unit of deg. after top dead center (deg. ATDC), which is the crank number reduced by 360 degrees, as shown in Fig.16(b)(c).

Acknowledgments

The authors would like to thank the Toyota Motor Corporation for its financial and technical support of this research.

References


