A MODEL BASED DESIGN APPROACH FOR KNOCK CONTROL IN INTERNAL COMBUSTION ENGINES USING MACHINE LEARNING

Eduardo V. Falcão*, Péricles R. Barros*, Vinicius M. Mafra†
*Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brazil
†Universidade Federal da Paraíba
João Pessoa, Paraíba, Brazil

Emails: eduardo.falcao@ee.ufcg.edu.br, prbarros@dee.ufcg.edu.br, brazilvmafram@hotmail.com

Abstract— Knock is the byproduct of an abnormal combustion taking place in the combustion chamber and is associated to high vibration levels, undesirable noise, increased emissions and degraded fuel economy. Conventional control strategies address knock in a corrective fashion by tuning the spark timing in order to decrease in-cylinder pressure and temperature. Although effective in avoiding engine’s prolonged exposure to the harmful vibratory forces, it may induce excessive spark retard and is implemented through a considerable number of calibrations. Recent breakthroughs on knock investigation show that knock intensity inherits random behavior and this statement plays a crucial role on how to appropriately design knock control laws. By leveraging this assumption using machine learning concepts and treating knock occurrence as a classification problem, a logistic regression model dependent on in-cylinder pressure, compressed mixture temperature and engine RPM is defined. The Likelihood Ratio Test is performed to validate the suggested hypothesis and McFadden’s pseudo-$R^2$ is calculated to quantify its accuracy. By leveraging the hypothesis stated, a proportional gain feedback control is designed to account for steady state operation of combustion engines around a target knock rate. Simulation results show the performance of both logistic regression model and proportional gain knock control against a conventionally defined knock control.

Keywords— Engine Knock, Logistic Regression, Machine Learning, Model based Design

1 Introduction

Knock is the audible noise that results from spontaneous ignitions that take place in the unburned zone of the air-fuel mixture enclosed in the combustion chamber and mainly characterized by fast heat releasing chemical reactions (Kiencke and Nielsen, 2005). As torque demand increases, spark timing has to be adjusted in a way that part of the combustion process is proportionally shifted towards the compression stroke in order to deliver more power. This leads to an increase in both temperature and pressure inside the cylinder as well as formation of hot spots in the combustion chamber. These conditions, occasionally, provide the reactive radicals in the fuel molecules with sufficient activation energy to ignite spontaneously. This causes an insufficient pressure equalization within the combustion chamber and high vibration levels are induced in the engine due to the fast expanding flame front colliding with cylinder wall, valves, spark plug and piston head but also from the spontaneous ignition flame front collision with the spark plug ignited flame front. Besides the harmful vibration levels that can damage engine components and undesirable noise, it is also associated to NOx formation (Jones et al., 2013).

Different control strategies were developed to address engine knock. A standard approach in the automotive industry is to retard the spark timing upon knock detection and incrementally advance it. This prevents engine’s long exposure to the vibratory forces but sacrifices engine efficiency. Such a corrective approach bases itself on the statement that engine knock is a deterministic phenomenon when, in fact, knock intensity possesses strong random nature (Selmanaj et al., 2018). This baseline leads to the conclusion that for an effective control law to be designed there has to be a model that approaches the problem in accordance to the appropriate mathematical framework.

Logistic regression is a powerful predictive analysis technique that attempts to describe the relationship between one dependent categorical random variable and measurable quantities that correlate with it. By addressing knock occurrence as a classification problem and defining a convenient features vector, it becomes feasible to apply logistic regression to compute statistical models in order to infer how engine knock is unraveling during its operation and define controllers that can prevent spark timing variability upon knock occurrence by maintaining engine operation around a target knock rate.

The work described in the following sections details the development of a predictive proportional gain knock control for internal combustion engines using logistic regression. Section 2 details engine knock and how it is traditionally addressed. Section 3 describes how logistic regression applies to knock detection and Section 4 describes the naturally aspirated combustion engine model. Section 5 shows how both conventional and proposed control strategies behave under the
Figure 1: Typical in-cylinder pressure profile for a knocking cycle as a function of crankshaft angle $\varphi$. In red, the sharp pressure peak caused by the fast heat releasing combustion from spontaneous ignitions.

The torque chain of a combustion engine powered vehicle comprises both slow and fast paths where the former refers to the air mass flow control and the latter relates to the spark timing. Adjustments in the torque chain are done to account for optimal vehicle driveability and safety compliance. Tuning the spark timing in the context of knock avoidance means retarding the combustion process beyond the optimal timing for maximum engine efficiency to be achieved. In general, the automotive industry addresses knock mitigation by means of the control law described in Equation 1 (Kiencke and Nielsen, 2005)

$$\phi(n) = \phi(n-1) - \Delta\phi + \beta \Delta E_\kappa(n) \quad (1)$$

where $\phi(n)$ is the spark angle for the $n$th cycle, $\Delta\phi$ the incremental spark advance and $\beta \Delta E_\kappa(n)$ the spark retard applied upon a knocking cycle based on the knock intensity above a threshold $E_0$, $\Delta E_\kappa(n) = E_\kappa(n) - E_0$. In other words, retard is applied to the spark timing upon knock occurrence and incrementally advanced in case of increasing vehicle load. This behavior is depicted in Figure 2. Knock has increased severity as load increases and engine speed decreases (Jones et al., 2013) and for higher knock rates, the control law described in Equation 1 leads to substantial variability of the actuator signal. The main downside of this approach lies on the fact that heavy knocking conditions lie in a region of engine operation of optimal fuel economy.

Considering the importance of maximizing the efficiency of combustion engines due to both environmental and driver experience reasons, a number of alternative knock control approaches were proposed throughout the recent years including standard deviation knock occurrence control (Schmilen and Rechs, 1991) and the use of fuzzy sets for spark timing adjustment (Yue and Li, 2004). In fact, the assumption of knock random nature plays a crucial role on knock control advancements and its strong correlation with peak in-cylinder pressure (Lezius and Drewelow, 2007) further supports the usage of a probabilistic framework on the definition of knock controllers. By leveraging this concept and recent studies on how machine learning techniques can be conveniently applied to forecast knocking conditions (Panzani et al., 2017), this work delivers further research on how to alternatively address the phenomenon.

### 3 Logistic Regression Applied to Knock Detection

Assuming knock occurrence as a random variable with two possible outcomes, it is possible to define the knock signal $\kappa(n)$ as (Panzani et al., 2017)

$$\kappa(n) = \begin{cases} 1, & E_\kappa(n) > E_0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Given that $\kappa$ follows a Bernoulli distribution, its
prediction defines a classification problem. One can leverage logistic regression in order to perform statistical inferences about how knocking conditions unravel by cyclically calculating log-linear models to accurately estimate the probability of ending up with a knocking cycle. Log-linear models are of the form

\[ \ln \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n = \theta \]  

(3)

where \( \theta \) is the linear predictor. The probability \( p \) modeled by the logistic function is

\[ p = P(\kappa = 1|\boldsymbol{X}) = \frac{e^\theta}{1 + e^\theta} \]  

(4)

The logistic function is defined by the unknown parameters \( \beta_0 \) and \( \beta = [\beta_1, \beta_2, \cdots, \beta_n]^T \) and the features vector \( \boldsymbol{X} = [x_1, x_2, \cdots, x_n] \). The value of \( p \) is determined based on the estimation of the unknowns that maximize the likelihood function \( \mathcal{L}(\beta_0, \beta|x) \) relative to the respective probability density function of the random variable \( \kappa \), and based on a set of \( n \) observations,

\[ \mathcal{L}(\beta_0, \beta|x) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} \]  

(5)

To ease the computational effort in estimating the unknown parameters it is possible to leverage the monotonic property of the logarithm function. In this case, the log-likelihood function \( \ell(\beta_0, \beta|x) \) is maximized instead, whose defines the cost function

\[ \ell(\beta_0, \beta|x) = \sum_{i=1}^{n} \ln (1-p) + \sum_{i=1}^{n} y_i \ln \frac{p}{1-p} \]  

(6)

The log-linear model proposed in this work is defined by features that correlate with knock occurrence and in order to validate it, a 1.3 liter naturally aspirated SI engine high pressure is modeled. The engine model encompasses combustion, heat transfer and piston kinematics models as well as a knock generator that draws the knock signal based on peak in-cylinder pressure values, a quantity that strongly correlates with knock intensity (Lezius and Drewelow, 2007).

4 Internal Combustion Engine Model

The 1.3 liter naturally aspirated internal combustion engine model encompasses the compression and expansion strokes of a four stroke spark ignited engine. Its parameters are described in Table 1.

In this model it is assumed that the torque generation can be taken as a lumped parameter process such that the gas states are not space dependent and the laws for ideal gases can be applied, characterizing an one-zone-model.

According to the first law of thermodynamics

\[ dU = dQ - pdW \]  

(7)

where \( dU \) and \( dQ \) are the incremental quantities for internal energy and supplied heat, respectively, and \( pdW \) the related mechanical work. For a combustion engine model dependent on the crankshaft angle \( \varphi \), the energy balance from Equation 7 can be described in terms of released and transferred heat, fuel properties and piston displacement as (Isermann, 2014)

\[ \frac{dU}{d\varphi} = \frac{dQ_f}{d\varphi} - p \frac{dV}{d\varphi} - \frac{dQ_w}{d\varphi} \]  

(8)

where \( Q_f \) is the gross released heat from combustion, \( Q_w \) the heat convectively transferred to the cylinder walls, \( m_f \) the fuel mass, \( V \) the cylinder volume and \( p \) the in-cylinder pressure. For the internal energy of the enclosed gas at constant volume, holds

\[ \frac{dU}{d\varphi} = m_q c_v \frac{dT_g}{d\varphi} \]  

(9)

with \( m_q \) and \( T_g \) denoting the mixture mass and temperature, respectively, and \( c_v \) the specific heat at constant volume. Equation 9 assumes that the air-fuel mixture is homogeneous and completely atomized. Based on the ideal gases equation and assuming stationary ideal gas constant \( R \), the derivative of Equation 9 w.r.t. the crankshaft angle is

\[ m_q R \frac{dT_g}{d\varphi} = \frac{dp}{d\varphi} \frac{V}{c_v} + p \frac{dV}{d\varphi} \]  

(10)

Plugging Equation 10 in Equation 9,

\[ \frac{dU}{d\varphi} = c_v \left( \frac{dp}{d\varphi} \frac{V}{c_v} + p \frac{dV}{d\varphi} \right) \]  

(11)

Finally, inserting Equation 11 in Equation 8 and solving for the incremental pressure changes yields

\[ \frac{dp}{d\varphi} = \frac{R}{c_v} \left( \frac{dQ_f}{d\varphi} - \frac{dQ_w}{d\varphi} - \left( 1 + \frac{c_v}{R} \right) p \frac{dV}{d\varphi} \right) \]  

(12)

Figure 3 shows different in-cylinder pressure profiles for different spark advances w.r.t. the Top
Dead Center of the compression stroke. Positive spark advances mean that the spark timing happens during the compression stroke. The more spark advance added the greater the peak in-cylinder pressure.

Since the compression stroke can be modeled as a polytropic process with sufficient accuracy (Eriksson and Andersson, 2002), the temperature of the homogeneous mixture $T_g$ can be defined as

$$T_g = T_{IVC} \left( \frac{p}{p_{IVC}} \right)^{\frac{k-1}{k}}$$

(13)

where $T_{IVC}$ and $p_{IVC}$ are the intake manifold air temperature and pressure measured at intake valve closure and $k$ is the polytropic coefficient.

The Equation 12 includes a heat exchange with the engine speed in RPM and piston bore $D_{pist}$ (H. and H., 1971). The derivation of the Wibe function yields the combustion heat release rate $dQ/d\varphi$ by

$$\frac{dQ_f}{d\varphi} = a(m + 1)H_i \Delta m_f \varphi_{\text{rel}} \frac{m_f}{\varphi_{\text{rel}}} e^{-m_f \varphi_{\text{rel}}^{m_f+1}}$$

(14)

The heat transfer to the cylinder walls as well as piston crown can be modeled through the basic equation for one-dimensional conduction (Iserrmann, 2014)

$$\frac{dQ_w}{d\varphi} = \frac{1}{\omega_{\text{eng}} A \left( T_g - T_w \right)}$$

(15)

with $\omega_{\text{eng}}$ the engine speed in rad/sec and $A$ the heat transfer effective area. The heat transfer coefficient $\zeta$ is given by (Hohenberg, 1980)

$$\zeta = 130 V_d^{-0.06} p^{-0.8} T_g^{-0.4} (\bar{v}_p + 1.3)^{0.8}$$

(16)

where $V_d$ is the displacement volume and $\bar{v}_p$ the mean piston speed. The average cylinder wall temperature $T_w$ can be estimated based on the air excess factor $\lambda$, engine speed $n_{\text{eng}}$ in RPM and piston bore $D_{pist}$ (H. and H., 1971). The incremental change in the cylinder volume relates to the crankshaft radius $r_{cs}$ and connecting rod length $l_{cr}$ by

$$\frac{dV}{d\varphi} = \frac{V_d}{2} \left( \sin \varphi + \frac{r_{cs}/l_{cr} \sin 2\varphi}{2 \sqrt{1 - (r_{cs}/l_{cr} \sin \varphi)^2}} \right)$$

(17)

### 4.1 Model Validation

In order to validate the accuracy of the combustion engine high pressure cycle model, different wide open throttle operating conditions were simulated and compared against results from dynamometer experiments. Each experiment was carried out during 30 seconds by keeping engine speed constant and slightly rich air-fuel ratios using the fuel blend advertised as E22 (22% hydrous ethanol, 78% gasoline). For each run, the peak in-cylinder pressure values were measured in each cycle and from where the average peak in-cylinder pressure $PMAX_{avg}$ and maximum peak in-cylinder pressure $PMAX$ quantities were calculated. The Figure 4 depicts how the estimated peak in-cylinder pressure by the model, $p_{max}$, compares to $PMAX_{avg}$. Since no friction losses are taken in consideration, the pressure value is the indicated mean effective pressure, or $IMEP$.

### 5 Hypothesis Assessment

The idea of correlating measurable quantities and knock occurrence with the intent of predicting the
probability of ending up with a knocking cycle was first introduced through the formulation of a features vector composed by the intake manifold temperature at intake valve closure, engine speed and a few coefficients for the in-cylinder pressure trace that represent the projection of the pressure profile in the eigenpressures identified through Principal Component Analysis (Panzani et al., 2017). In this work, the features vector is alternatively defined by the peak in-cylinder pressure, \( p_{\text{max}} \), unburned air-fuel mixture temperature at 10% of burned fuel mass fraction, \( T_{10\%} \), and engine speed \( n_{\text{eng}} \) in RPM. The log-linear model for the alternative hypothesis is, thus, given by

\[
\ln \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 p_{\text{max}} + \beta_2 T_{10\%} + \beta_3 n_{\text{eng}}
\]

The motivation behind the definition of the log-linear model in Equation 18 as being dependent on the unburned air-fuel mixture temperature lies on the fact that spontaneous ignitions take place in the end gas of the combustion chamber (Isermann, 2014). The 10% threshold indicates beginning of combustion, after flame development delay (Urlaub, 1988), and is taken as a convenient instant to better describe the temperature at which the combustion takes place.

5.1 Hypothesis Test and McFadden’s pseudo-\( R^2 \)

The hypothesis that describes Equation 18, namely \( H_\alpha \), represents an alternative approach to describe how knocking conditions unravel during engine’s dynamic operating conditions. Test procedures such as the Likelihood Ratio Test have to be carried out in order to determine the consistency of \( H_\alpha \) towards a hypothesis which, prior to the test, is taken as a true statement for the phenomenon investigated and known as null hypothesis, or \( H_0 \) (Christensen, 1997). By stating a null hypothesis model \( H_0 \) whose parameters \( \beta \) are a subset of those from \( H_\alpha \), the \( G^2 \) statistic defined as

\[
G^2 = -2(\ell_0 - \ell_1)
\]

is calculated. Based on Wilks Theorem (Wilks, 1938), Equation 19 approaches a \( \chi^2 \) distribution for a large number of samples used to calculate \( \ell_0 \) and \( \ell_1 \) (Christensen, 1997). Given the significance level \( \alpha \) and the number of degrees of freedom \( df \) the Likelihood Ratio Test states that if the alternative model holds, then \( G^2 > \chi^2(\alpha, df) \) and we can reject \( H_0 \).

To perform the hypothesis test, a 1000 cycles wide open throttle simulation undertaking a linear increase in engine speed from 2000 to 5000 RPM at slightly rich air-fuel mixture and normal temperature and pressure conditions was carried out. Figure 5 depicts how \( p_{\text{max}} \) and \( T_{10\%} \) increase during the driving cycle. In the test conducted, the null hypothesis is the intercept model

\[
\ln \left( \frac{p}{1-p} \right) = \beta_0
\]

which represents the expected mean value of \( \kappa \) for the \( n = 1000 \) cycles considered. The data set produced by the simulation leads to the calculation of \( \ell(\beta_0, \hat{\beta}|x) = -686.64 \) and \( \ell(\beta_0, \hat{\beta}|x) = -548.40 \), yielding the \( G^2 \) statistic \( G^2 = 276.47 \). Establishing a significance level of \( \alpha = 5\% \), the null hypothesis can be rejected if the value of the respective \( \chi^2 \) distribution with number of degrees of freedom \( df = 3 \), the dimension of the features vector \( X \), is less than \( G^2 \). Since \( \chi^2(df = 3) \) for \( \alpha = 5\% \) is approximately 7.81, we can reject the null hypothesis and consider \( H_\alpha \) as a valid assumption.

Coefficients of determination are calculated to quantify the performance of statistical models on describing the variance of a given random variable. For dichotomous random variables, McFadden’s pseudo-\( R^2 \), \( R^2_M \) is calculated:

\[
R^2_M = 1 - \frac{\ell_1}{\ell_0}
\]

Good statistical models will produce values in the range \( 0.2 \leq R^2_M \leq 0.4 \). For the exposed data set, \( R^2_M = 0.22 \), indicating that the model is a good fit.

5.2 Simulation

Performance evaluation for both alternative and null hypothesis on knock probability estimation was conducted in a 1000 cycles wide open throttle simulation at 2500 RPM for fuel blend E22, all the while inducing a dynamic behavior w.r.t. the spark advance and a linear increase in engine speed as in the hypothesis test. Knock events were generated based on knock occurrence probabilities drawn from a table indexed by peak in-cylinder...
pressure values. At each cycle an uniformly distributed random number within the range [0,1] is generated and compared to the knock probability at that operating condition, outputting $\kappa = 1$ whenever it doesn’t exceed the mapped probability. Figure 6 depicts the behavior of the knock generator.

The simulation results are shown in Figure 8. The hypothesis $H_a$ has better performance regarding tracking variations in the knock probability when compared to the null hypothesis, as expected. Given that the null hypothesis outputs the mean knock occurrence observed within the windowed cycles, its performance is conditioned to observations made on $\kappa$ only. On the other hand, since $H_a$ relies on the set of features $p_{\text{max}}, T_{10\%}$ and $\omega_{\text{eng}}$, it is possible to fast track how knocking conditions are evolving, where the reference value is the knock probability fed into the knock generator.

6 Alternative Knock Control

In order to leverage the information contained in the log-likelihood function $\ell(\beta_0, \beta_1 \mid x)$ an alternative control law is hereby proposed. Instead of approaching knock control in a corrective fashion as in Equation 1, one can accurately calibrate the knock intensity threshold $E_0$ and allow for light knock to occur sporadically without damaging engine components nor degrading driver’s experience and, thus, maximize engine efficiency.

By defining the error signal $e_t = e_{\ell(\beta_0, \beta_1)} - e_{\ell_{\text{tgt}}}$, where $\ell_{\text{tgt}}$ is given by the target log-likelihood for a given target knock rate $t_k$

$$\ell_{\text{tgt}} = \ln \left( t_k^{n_\kappa} (1 - t_k)^{n(1 - n_\kappa)} \right)$$  \hspace{1cm} (22)

it is possible to design a proportional gain closed-loop controller whose gain is tuned in a way that the spark advance adjustments follow the dynamic behavior of the log-likelihood function.

Figure 6: Knock generator behavior

Figure 7: Iterative log-linear models estimation performance for the hypothesis tested. At each cycle, the training data is updated and a new set of unknown parameters is estimated. In this case, the training data is composed by $n = 200$ cycles.

6.1 Simulation

The simulation carried out in Figure 8 depicts the steady state behavior for both conventional controller and proposed controller for a rather high target knock rate of $t_k = 10\%$. The quantities $\beta_0 E_y$ and $\Delta \phi$ in Equation 1 were tuned in a way to achieve such a target. The Table 2 describes the fixed spark retard and advance gains as well as the proportional gain $K$ for the proposed controller.

<table>
<thead>
<tr>
<th>Conventional Control</th>
<th>Proposed Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \phi = 0.2^\circ$</td>
<td>$K = 0.01$</td>
</tr>
<tr>
<td>$\Delta E_y = 2^\circ$</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Spark advance correction quantities for the conventional knock controller and proportional gain for proposed controller.

The Figure 8 shows how the proportional gain controller behaves upon request to control the target knock rate around $t_k$. The initial spark advance is $12^\circ$, the maximum spark advance for non-knocking cycles at 2000 RPM. The conventional knock control is not ideal for this purpose due to the high variability it induces in the spark advance. To meet torque demand, the control unit has to compensate the retarded spark timing by either rapidly advancing it or adjusting the induced air and fuel injected into the cylinder, and this may have a negative impact in fuel economy for extended engine operation in knocking conditions. On the other hand, the proportional gain controller acts upon changes in the log-likelihood, which is computed based on data cyclically updated during the simulation. For this particular run, the window size was $n = 200$. It is able of keeping track of how prone the engine is to run into undesirable knocking conditions and slowly
adjusts the spark advance to meet the target without overshooting the actuator signal. By keeping the air-fuel-ratio constant, an average indicated torque of 185 Nm was achieved. However, torque variability for the conventional control was 31.3 Nm whereas the proposed controller had negligible variance not considering the transient response.

7 Conclusion

This work further concludes that based on the assumption of knock occurrence behaving as a categorical random variable, and addressing engine knock as a classification problem, it is possible to accurately quantify the cyclical knock probability along a driving cycle through data analysis of quantities such as in-cylinder pressure, temperature and engine speed. The Likelihood Ratio Test evidences the relevance of the alternative hypothesis and the simulation results show how it outperforms the null hypothesis.

Further work has to be done in, first, expanding the combustion engine model to incorporate air-fuel ratio control and enable fuel economy analysis followed by validation of the proposed strategy in a test bench. Then, better control strategies will be evaluated in order to provide with a quicker response for transient operation. Next studies will also assess how it is possible to leverage artificial neural networks to accurately predict knocking cycles.

References


