RESEARCH ARTICLE

Optimal-robust selection of a fuel surrogate for homogeneous charge compression ignition modeling

Irene García-Camacha Gutiérrez[ID][1], **Raúl Martín Martín**[ID][1]*, **Josep Sanz Argent**[2]

1 Institute of Applied Mathematics for Science and Engineering (IMACI), University of Castilla-La Mancha, Toledo, Spain, **2** Fundación Valenciaport, Valencia, Spain

* Raul.MMartin@uclm.es

Abstract

Homogeneous Charge Compression Ignition (HCCI) combustion is a potential candidate for dealing with the stringent regulations on vehicle emissions while still providing very good energy efficiency. Despite the promising results obtained in preliminary studies, the lack of autoignition control has delayed its launch in the engine industry. In the development of the HCCI concept, the availability of reliable computer models has proved extremely valuable, due to their flexibility and lower cost compared with experiments using real engines. In order to obtain the best formulation of a fuel surrogate formulated with n-heptane, toluene and cyclohexane that efficiently estimate the autoignition behaviour, regression adjustments are made to the Root-Mean-Square Errors (RMSE) of experimental Starts of Combustion (SOC) from the modeled SOC. The canonical form of the Scheffé polynomials is widely used to fit the data from mixture experiments, however the experimenter might have only partial knowledge. In this paper we present the adaptation of the robust methodology for possibly misspecified blending model and an algorithm to obtain tailor-made optimal designs for mixture experiments, instead of using standard designs which are indiscriminately employed, to make good estimations of the parameters blending model. We maximize the determinant of the mean squared error matrix of the least square estimator over a realistic neighbourhood of the fitted regression mixture model. The maximized determinant is then minimized over the class of possible designs, yielding an optimal design. Thus, the computed desings are robust to the exact form of the true blending model. Standard mixture designs, as the simplex lattice, are around 25% efficient for estimation purposes compared with the designs obtained in this work when deviances from the considered model occur during the experiments. Once an optimal-robust design was selected (based on the level of certainty about model adequacy), we computed the optimal mixture that best reproduces the combustion property to be imitated. Optimal mixtures obtained when the considered model is inadequate agree with the results achieved in empirical studies, which validates the methodology proposed in this work.

Introduction

Governments and other institutional entities have shown their concern about the release of hazardous substances from the use of Internal Combustion Engines in transportation, and the resulting impact on human health. Consequently, they have made the regulations on pollutant emissions stricter. On the other hand, there is also global awareness regarding the efficient use of fossil fuels due to their proven impact on global warming. Thus, new methods of combustion need to be explored to generate cleaner energies achieving the best performance. Whilst other alternative methods, such as electric engines or the use of fuel cells, are being developed, internal combustion engines are still the most widely used nowadays. Engine researchers and designers look into improvements in traditional Spark Ignition (SI) and Compression Ignition (CI) engines, to reduce their emissions while maintaining good fuel economy. Homogeneous Charge Compression Ignition (HCCI) combustion is a potential candidate to meet these needs and has gathered extensive support over the last decade.

HCCI combustion is defined as a process by which a homogeneous mixture of air and fuel is compressed until autoignition occurs near the end of the compression stroke [1]. HCCI engines have the advantage of providing good fuel economy, like diesel, while producing very low particulate emissions, comparable to a gasoline engine. A comparative study of the different combustion engines can be found in [1].

One of the main features of this type of engine is that combustion is a kinetically controlled process rather than a physical phenomenon as in traditional engines [2]. Autoignition and combustion rates are strongly influenced by the chemical kinetics of the fuel. These phenomena are therefore extremely sensitive to charge composition as well as other factors involved in the combustion such as pressure and temperature. Charge composition is the main scenario of this work and different levels of pressure and temperature are considered through the parameter combinations used in the experiments. On the other hand, ignition timing is the major challenge to be addressed in these engines. This is the main drawback which has delayed its launch on the automobile market. Different solutions have been considered, such as the variable valve timing (VVT) method or the variable EGR method among others (see, for example, [3–5]). However, all these strategies have been studied experimentally. Mathematical engine models are valuable tools for predicting and analyzing these processes and allow many engine design alternatives to be considered [6]. Due to the paramount importance of the combustion kinetics of the fuel, any type of model (from zero dimensional to CFD models) will necessarily require reliable methods for accurately describing the reaction kinetics of real fuels. Because of the great complexity of commercial fuels (which include from hundred to thousands of different hydrocarbons) surrogate fuels are used instead, which are mixtures of a few hydrocarbons with well-known combustion kinetics. This paper poses the problem of fuel surrogate selection to control HCCI autoignition from an Optimal Experimental Design (OED) perspective, and incorporates the theory of robust design into the problem, to consider a more realistic scenario.

Surrogates must closely mimic market fuel properties and match engine combustion and emissions behavior [7]. A variety of selection mechanisms have been used in the literature. Hernández et al. [8] assumed that the error function matching the simulated and experimental starts of combustion had an absolute minimum, and sought this optimum using a type of bisection method. Others, however, defined an objective function representing some physical, chemical and combustion properties and optimized this complex function using commercial software optimizers [9–11] or parallel computing calculation [12]. Yu et al. [13], took a different approach, constructing surrogate fuel mixtures by directly matching the molecular structure and the key functional groups, instead of using the targets for the combustion properties

explicitly. There are as many fuel surrogates as engine properties to be imitated. In the particular case of HCCI modeling and the present research, the surrogate should adequately represent both the autoignition properties and commercial fuel compositions. Classical mixture experimental designs have been used to estimate the effects of fuel compositions to HCCI performance [14]- [15]. Their applicability depends however on the experimental factor space and on the choice of a model for the responses. The novelty of this work is that, in addition to consider the optimal design theory, considers possible departures from the assumed model. It is a common situation in this type of experimental setups due to the extreme conditions under which experiments are carried out.

Computer simulation has become a dominant tool in making HCCI a reality and in the quest for control strategies for HCCI, and has higher flexibility and lower cost than real engine experiments [6].There has been increasing interest in Optimal Experimental Design (OED) in recent year, not only for reasons of resource optimization but also due to its flexibility in handling more complex and realistic problems. Applications of this theory to engines include, for example, the modeling of the ignition angle in HCCI engines [16] or the quantification of the effects of fuel compositions on emissions [17]. On the other hand, in this type of problem, the response of the model may be easily perturbed by the extreme conditions under which experiments are carried out. Practitioners have little information about model suitability prior to running the experiments and the large number of factors involved in the combustion as well as the difficulty of considering all possible operating conditions implies that small deviations from the considered model may occur during the experiments. The robust design theory allows practitioners to establish the level of certainty about model suitability. Thus, this approach would reflect a more realistic modeling of the problem. Taking these considerations into account, the aims of this research are:

- to formulate the problem of selecting a fuel surrogate to model HCCI autoignition using optimal mixture design.

- to incorporate robustness into the formulation.

- to provide a new method to obtain an efficient $D-$optimal robust design for several levels of certainty about model suitability.

- from the selected $D-$optimal robust design, to compute the surrogate formulation that best reproduces the mechanism to be replaced as well as the combustion property to be imitated.

These objectives are aimed at selecting a fuel surrogate to estimate the autoignition time under real HCCI conditions.

Materials and methods

The strong dependency of HCCI autoignition on the chemical kinetics of the fuel implies that it is extremely sensitive to charge composition. Commercial diesel consists of hundreds of medium-high molecular weight hydrocarbons, and thus it is not feasible to consider the oxidation chemistry of all the compounds when modeling targets [8]. A strategy commonly adopted in simulation studies is to consider a reduced number of species, and then to prove that the reduced mechanism properly matches the desired combustion property through experimentation. Following Hernández et al. [8], our interest is in modeling the Root-Mean-Square Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{J} \left(\text{SOC}_i(\exp) - \text{SOC}_i(\text{mod})\right)^2}{J}}, \tag{1}$$

where $SOC_i(exp)$ and $SOC_i(mod)$ correspond to the experimental and modeled Starts Of Combustion (SOC) obtained in J engine tests. In particular, SOC reports the crank angle corresponding to 10% of the cumulative heat released during the high temperature combustion process. Eq (1) defines the response variable in the model. For notational reasons, this will be represented by $\mathrm{RMSE} \underset{Not}{\equiv} Y$. Experimental SOCs were obtained using a single-cylinder diesel engine operating under HCCI conditions while the modeled values came from the zero-dimensional single-zone kinetic model implemented in the CHEMKIN 4.0 software [18].

Single-cylinder engine tests provide a means for excellent control and reproducibility of the operating conditions [19]. The engine used for the experimental work was a four-stroke single cylinder engine, with 0.287l swept volume and four valves, second generation common rail (fuel injection system), 14:1 compression ration, 0.068m x 0.079m (bore x stroke), and 12 x 86 μm number and diameter of injector holes respectively. Different engine operating conditions were considered in the J engine tests run in (1) modifying the most important parameters affecting the two-stage oxidation kinetics of any fuel. The charge composition was quantified by the exhaust gas recirculation percentage (EGR). The engine load was quantified in terms of the indicated mean effective pressure (IMEP) by fixing the intake pressure and varying the EGR rate. The engine speed was measured in revolutions per minute (rpm), and the start of injection (SOI) was defined with respect to the top dead center (TDC) where a negative value indicates that injection starts before TDC. Operating rate data are shown in Table 1. As it can be observed, experimental HCCI conditions were achieved with a very early fuel injection. Consequently, the fuel/air mixture may assume to be homogeneous (or almost homogeneous) from the end of the injection process, since the physical delay can be considered negligible compared to the chemical one [8]. In addition, the injection pressure was kept constant for all the tests at 900 bar.

Simulations were performed using CHEMKIN code, which assumes uniform temperature, pressure and species concentration in the combustion chamber, and ignores the temperature differences near boundary zones due to heat transfer. A single zone model was used. In spite of the limitations of these models due to the assumption that the composition and temperature in the cylinder are homogeneous, they have been used to provide an estimation of the ignition

Table 1. Engine operating conditions used in the engine test.

Engine speed (rpm)	EGR rate (%)	SOI (with respect to TDC)	IMEP (bar)
1500	40	-80	5.2
			2.8
		-60	6.3
			4.1
	60	-80	4.9
			2.8
		-60	6.2
			4.1
2000	40	-80	5.9
			3.1
		-60	5.8
			4.4
	60	-80	5.6
			3.2
		-60	7.1
			4.6

timing [20–22]. As previously mentioned, another fundamental characteristic of the selected surrogate is that it must appropriately reproduce the kinetic-chemical mechanism of the original fuel. The affinity criterion means that the main hydrocarbon families must be represented in the mixture. Since original diesel consists of around 37% paraffins, 34% naphthenes and 29% aromatics by mass [8], n−heptane, cyclohexane and toluene were the selected components to represent each hydrocarbon family. n−heptane has a cetane number equal to diesel fuel and its oxidation has been widely validated under engine conditions [23–25]. Kinetic models of longer paraffins show similar autoignition results to n−heptane and, for this reason, some authors recommend the use of n−heptane for its lower computational cost [26]. The cyclohexane and toluene were selected since their oxidation chemistry is validated in the literature under similar conditions to HCCI engines [27–32]. The surrogate mechanism was built up in a step-wise fashion, using as a starting point the n−heptane detailed mechanism proposed by Curran et al. [33] which contains 550 species and 2450 reactions. Then, the reactions that describe the oxidation of toluene proposed by the Lawrence Livermore National Laboratory [27] were added (285 species and 1427 reactions). In addition, the 12 co-oxidation reactions between n−heptane and toluene proposed by Andrae et al. [34] were incorporated into the mechanism. Finally, the mechanism of cyclohexane oxidation, containing 1081 species and 4269 reactions, given by Silke et al. [32] were added. The resulting mechanism contains 1140 species and 4578 reactions.

A suitable fit of the response requires an appropriate choice of both design and model. Since the RMSE strongly depends on the considered proportions of n−heptane, toluene and cyclohexane and these are ingredients of a mixture, models for mixture experiments may be appropriate for explaining this behavior. Controlled variables in a standard mixture problem are non-negative, belonging to [0, 1] and dependent through the relationship $\mathbf{1}'_q\boldsymbol{p} = 1$ where $\mathbf{1}_q = (1,\ldots,1)' \in \mathbb{R}^q$ and $\boldsymbol{p} = (p_1,\ldots,p_q)'$ is the vector of relative proportions in a q−component mixture. These constraints define the design region as a $(q-1)$−dimensional simplex $\mathcal{S} = \{\boldsymbol{p} \in [0,1]^q : \mathbf{1}'_q\boldsymbol{p} = 1\}$. A mixture design ξ is thus an allocation rule of experimental units over $\mathcal{S}$,

$$\xi = \left\{ \begin{array}{c} \boldsymbol{p} \\ m(\boldsymbol{p}) \end{array} \right\}_{\boldsymbol{p} \in \mathcal{S}},$$

where $m(\boldsymbol{p})$ is a probability mass function. If $m(\boldsymbol{p})$ is finite-supported, it is said that ξ is a discrete design. Otherwise, it is a continuous design. As stated above, a suitable model must be selected a priori to describe the composition-response relationship. Scheffé polynomials are widely applied to fit the response surface. With regard to polynomial order, quadratic models have been used in HCCI modeling as they provide a reasonable balance between the necessary number of experiments and modeling capacity [16]. The general form of a second-degree mixture model is:

$$E[Y(\boldsymbol{p})] = \theta_0 + \sum_{j=1}^{q} \theta_j p_j + \sum_{i=1}^{q-1} \sum_{j=1}^{q} \theta_{ij} p_i p_j,$$

where the term $\sum_{j=1}^{q} \theta_j p_j$ is the linear blending portion of the model, while the term $\sum_{i=1}^{q-1} \sum_{j=1}^{q} \theta_{ij} p_i p_j$ represents the nonlinear blending (curvature) between component pairs. However, ordinary polynomials do not allow the parameters to be uniquely identified, due to co-linearity between proportions. Instead, canonical polynomials introduced by Scheffé [35], [36] are the most commonly used in mixture experiments for a large number of

practical situations, including HCCI combustion properties ([14]- [15], [37]). Thus, a canonical second-order Scheffé polynomial in the intercept form (2) was considered for modeling the RMSE

$$Y(\boldsymbol{p}) = f'(\boldsymbol{p})\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \theta_0 + \theta_1 p_1 + \theta_2 p_2 + \theta_{11} p_1^2 + \theta_{22} p_2^2 + \theta_{12} p_1 p_2 + \boldsymbol{\varepsilon},$$

(2)

where p_1, p_2 represent the proportion of n–heptane and toluene by mass respectively and the proportion of cyclohexane in mass, p_3, is implicitly determined by the relationship $p_3 = 1 - p_1 - p_2$. The vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_{11}, \theta_{22}, \theta_{12})$ corresponds to the vector of the unknown model parameters. Additive uncorrelated random errors ε with common variance will be assumed.

Classical optimal design theory establishes that the model form is completely determined, and then, the so-called optimal designs are obtained from the optimality criteria based on the inverse of the Fisher information matrix ($M^{-1}(\xi)$). J. Sanz [37] stated that there are significant discrepancies between experimental and modeled results despite the existence of validated mechanisms in the literature to reproduce HCCI engine conditions. The large number of factors involved in this process, as well as the difficulty of considering all possible operating conditions, implies that small deviations from the model may occur during the experiments. Thus the standard models could not accomodate other effects of the mixture components. Robust design theory introduced by Huber [38] and later developed by Wiens [39–41] will be suitably applied to address this situation. A more realistic framework of the problem is to consider a class of plausible responses

$$E(Y|\boldsymbol{p}) = \boldsymbol{f}'(\boldsymbol{p})\boldsymbol{\theta} + \psi(\boldsymbol{p}), \quad \boldsymbol{p} \in \mathcal{S},$$

(3)

where ψ is a *contamination function* which belongs to a certain class of functions Ψ. The choice of Ψ determines the neighborhood of possible responses. This study will consider a neighborhood orthogonal to the experimenter's assumed response, since it covers a wide range of general alternative responses (for further details see [41]). When the assumed response is (2) but the true one is (3), bias is introduced in the least square estimator (*lse*) $\hat{\theta}$ of the model parameters. Let n be the number of mixtures to perform. Then, unlike in classical optimal design theory, the mean-square-error matrix of the estimators, rather than the variance, is a good measure of design quality

$$\mathrm{mse}(\hat{\boldsymbol{\theta}}) = \mathrm{mse}(\psi, \xi) = \frac{\sigma_\varepsilon^2}{n}\mathbf{M}^{-1}(\xi) + \mathbf{M}^{-1}(\xi)\boldsymbol{b}(\psi, \xi)\boldsymbol{b}'(\psi, \xi)\mathbf{M}^{-1}(\xi)$$

(4)

where

$$\mathbf{M}(\xi) = \int_{\mathcal{S}} \boldsymbol{f}(\boldsymbol{p})\boldsymbol{f}'(\boldsymbol{p})m(\boldsymbol{p})d\boldsymbol{p} \text{ and } \boldsymbol{b}(\psi, \xi) = \int_{\mathcal{S}} \boldsymbol{f}(\boldsymbol{p})\psi(\boldsymbol{p})m(\boldsymbol{p})d\boldsymbol{p}.$$

The first term in (4) corresponds to the variance-covariance matrix of $\hat{\theta}$ assuming homoscedastic errors and common variance σ_ε^2, while the second is the squared bias. Suitable functionals of (4) are defined to optimize some aspect of the model. Classical alphabetical optimality criteria [42] define the most common optimal robust criteria in terms of (4). In this context, they are known as *loss functions* $\mathcal{L}(\psi, \xi)$. The D–optimality criterion minimizes the volume of the confidence ellipsoid of the model parameter estimators. Since it has a direct interpretation and considers all model parameters, this criterion has became the most popular among practitioners and will be used in this research. Thus, our objective is to obtain the

minimax design

$$\xi^* = \arg \min_{\xi \in \Xi} \max_{\psi \in \Psi} \mathcal{L}_D(\psi, \xi),$$

where Ξ is the set of probability measures. In other words, the D−optimal robust design will be the design that minimizes the loss in the worst scenario. The maximization problem over Ψ may be equivalently formulated as a problem of computing the maximum eigenvalue $\text{ch}_{\max}$ of a certain matrix [39]. Therefore, the D−loss is defined as

$$\mathcal{L}_D(\xi) = (1 - v)\left(\frac{1 + \frac{v}{1-v} \, \text{ch}_{\max} \, \mathbf{M}^{-1}(\xi)\mathbf{G}(\xi)}{\det\mathbf{M}(\xi)}\right)^{1/k}, \tag{5}$$

where

$$\mathbf{G}(\xi) = \mathbf{K}(\xi) - \mathbf{H}(\xi), \;\; \mathbf{K}(\xi) = \int_{\mathcal{S}} \boldsymbol{f}(\boldsymbol{p})\boldsymbol{f}'(\boldsymbol{p})m^2(\boldsymbol{p})d\boldsymbol{p}, \;\; \mathbf{H}(\xi) = \mathbf{M}(\xi)\mathbf{A}^{-1}\mathbf{M}(\xi),$$

k is the number of model parameters, $A = \int_{\mathcal{S}} f(\boldsymbol{p})f'(\boldsymbol{p})d\boldsymbol{p}$ and the parameter v has been defined from the neighborhood ratio τ and the variance as

$$v = \frac{\tau^2}{\sigma_\varepsilon + \tau^2}.$$

This new parameter represents the relative importance, to the experimenter, of the errors due to bias rather than to variance. The 0 value means that the model parameters are unbiased, while 1 represents the opposite case.

To clarify the importance of the designs created by this approach, we define the efficiency of a design ξ with respect to the design $\bar{\bar{\xi}}$

$$\text{Eff}(\xi) = \left(\frac{\mathcal{L}_D(\xi)}{\mathcal{L}_D(\bar{\bar{\xi}})}\right)^{\frac{1}{k}} \times 100.$$

If a design ξ has, for instance, 70% efficiency with respect to $\bar{\bar{\xi}}$, then $\bar{\bar{\xi}}$ will produce the same precision as ξ with 30% fewer observations, thus reducing the experimental and resource cost.

Since a finite number of trials must be proposed to carry out the simulations, n−point discrete designs will be considered. This implies that prior definitions must be properly discretized. Then, for each fixed v, the optimization problem to solve is

$$\xi^* = \arg \min_{\xi \in \Xi} (1 - v)\left(\frac{1 + \frac{v}{1-v} \, \text{ch}_{\max} \, \mathbf{M}^{-1}(\xi)\mathbf{G}(\xi)}{\det\mathbf{M}(\xi)}\right)^{1/p}.$$

In most cases, this problem is analytically intractable. A genetic algorithm (GA) with specialized operators has been developed in this study for this purpose. The algorithm starts by generating a population of M random initial designs. A *generation* corresponds to a complete renovation of the population. This update is carried out through the *genetic operators*. They are classified into three groups: *selection, crossover and mutation operators*. Two *parent* designs are suitably selected and an *offspring* design is created from them by the action of the operators. A measure of probability (*fitness*) is assigned to each design. This assignment is made according to design goodness measured in terms of criterion function values. Selection operators always act until they have completed a generation while reproduction and mutation operators only act if a probability test is passed. A random $U(0, 1)$ is generated in each possible intervention and if this value is lower than a certain predetermined probability, the operator

will act. Otherwise, the design will remain unchanged. The reproduction probability chosen in this study was $P_C = 0.9$ while the mutation probability P_{Mut} was automatically updated to generate more opportunities for new solutions when the algorithm is further from the optimum, and to reduce the "biodiversity" otherwise. It is noteworthy that the appropriate selection of the operators considered in this paper, as well as the creation of new ones because of the special nature of mixture experiments, converged for all study cases. Pseudo-code for the proposed GA is given in the S1 Appendix. It was implemented in MATLAB R2018b and the code is available from the authors.

Results

Several numerical examples were run covering different levels of certainty about model suitability (v). $n = 12$−point designs were considered for illustration. The D−optimal robust designs obtained are shown in Fig 1. They were graphically depicted using the *mixexp* R package. In these graphs, simplex vertices represent pure components, edges correspond to binary blends and simplex interior points contain all three-component systems.

As one might expect, quasi-zero v value design was supported on the extremes and mid-edges of the design region according to classical D−optimal design. The class of so-called *standard mixture designs* was obtained in this study case. If $m \geq 1$ is an integer, the $\{q, m\}$−simplex lattice on $\mathcal{S}$ is defined as the set of points whose coordinates are integer multiples of $1/m$, that is $\left\{ \boldsymbol{p} \in \mathcal{S}, \ p_i = \frac{i}{m}, 0 \leq j \leq m, \ 1 \leq i \leq q \right\}$. Thus, a $\{q, m\}$−simplex lattice design is a design with support points in the $\{q, m\}$−lattice. Another class of standard mixture designs are the $\{q, m\}$−simplex centroid designs ($1 \leq m \leq q$), which are defined as a collection of points in $\mathcal{S}$ with $q-j$ coordinates equal to zero and j coordinates equal to $\frac{1}{j}, j = 1, \ldots, m$. According to the previous definitions, a $\{3, 2\}$−simplex lattice design was obtained for $v = 0.01$ and quite similar for 0.3, that is support points are pure components or binary blends; whereas a quasi $\{3, 3\}$-simplex centroid design was obtained in the case of $v = 0.5$. No pattern is deduced from the spatial distribution of design points when the bias is emphasized ($v > 0.5$). As we can see from
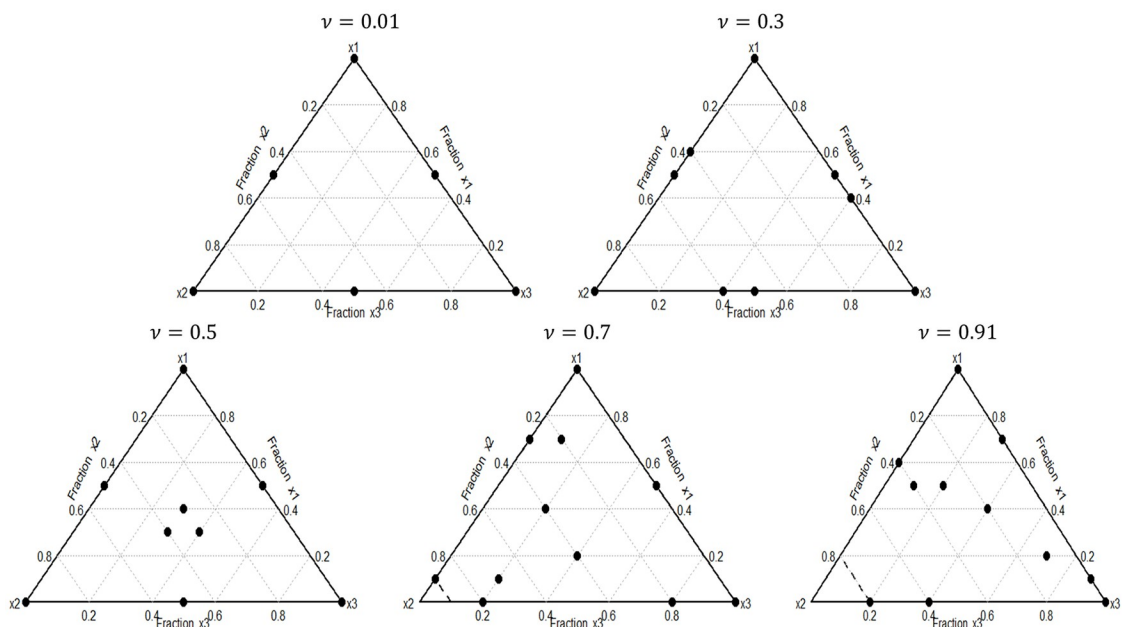

Fig 1. D−optimal robust designs. Designs obtained with the proposed GA for HCCI autoignition modeling through model (2).

Table 2. Computational costs and loss values of the D−optimal robust designs obtained from Fig 1.

v	no. it	minloss
0.01	2567	25.42
0.3	2300	17.49
0.5	1771	11.3
0.7	1803	1.99
0.91	1161	0.6782

https://doi.org/10.1371/journal.pone.0234963.t002

Fig 1, the optimal design points are spread out over the experimental region. It is a logical result since, in order to seek protection against deviations from the model, no design region should be favored over any other. This shows the importance of considering robust designs when we cannot guarantee the model choice. The number of iterations needed to reach the optimum, as well as the loss values of the designs calculated are shown in Table 2. Note that the more the bias is ignored, the higher the loss and the computational cost.

Since standard mixture designs such as simplex lattice or simplex centroid are the most widely used by practitioners to estimate the unknown parameters, we compute their efficiencies with respect to the designs obtained through the proposed methodology when deviations from the assumed model occur. The efficiency is a number between 0 and 100 which measures the goodness of a design ξ for estimating purposes. Thus, a {3, 2}−simplex lattice is 40% efficient regarding the design shown in Fig 1 when $v = 0.7$, whereas a {3, 3}−simplex centroid is 25% efficient in this case. A more extreme study case is when $v = 0.91$. In this context, the efficiencies are 28% and 20% according to a {3, 2}−simplex lattice and a {3, 3}−simplex centroid respectively with respect to the design shown in Fig 1 for $v = 0.91$.

Once a D−optimal robust design is selected, the next step consists of appropriately fitting the model 2. The response in this model explains the error between the start of combustion obtained experimentally and simulated via the surrogate (1). The optimal choice of the fuel surrogate will be one that minimizes the response in a complex feasible region.

Taking into account the designs depicted in Fig (1), the least-squares method was used to fit the (2) with the data obtained after displaying the experimental setup described in the *Material and Method Section*. Data used to perform the adjustment are displayed in S1 Table. Thus, for different levels of certainty about model suitability, we obtain

$$\hat{Y}(\boldsymbol{p}) = \hat{\theta}_0 + \hat{\theta}_1 p_1 + \hat{\theta}_2 p_2 + \hat{\theta}_{11} p_1^2 + \hat{\theta}_{22} p_2^2 + \hat{\theta}_{12} p_1 p_2,$$

where the third-component proportion is implicitly determined through the relationship $p_3 = 1 - p_1 - p_2$. According to the similarity criterion, the optimal choice of the surrogate for this experimental setup will be the solution of the optimization problem

$$(p_1^*, p_2^*) = \arg \min_{\substack{\hat{Y}(\mathbf{p}) \geq 0 \\ 0 \leq p_1, p_2 \leq 1 \\ p_1 + p_2 \leq 1}} \hat{\theta}_0 + \hat{\theta}_1 p_1 + \hat{\theta}_2 p_2 + \hat{\theta}_{11} p_1^2 + \hat{\theta}_{22} p_2^2 + \hat{\theta}_{12} p_1 p_2, \tag{6}$$

where $p_3^* = 1 - p_1^* - p_2^*$. Using a nonlinear optimizer in MATLAB with linear constraints, we obtained the optimal mixture compositions of the fuel surrogate that best reproduce HCCI autoignition depending on the degree of model suitability. Optimal mixtures (given in % in mass) as well as fitted models are shown in Table 3.

Table 3. Fitted models for optimal-robust designs shown in Fig 1 and optimal mixtures obtained from solving the optimization problem (3).

v	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_{11}$	$\hat{\theta}_{22}$	$\hat{\theta}_{12}$	% n–heptane	% toluene	% cyclohexane
0.01	38.04	-80.97	-66.64	46.51	54.55	53.73	72	25	3
0.3	37.74	-83.64	-42.43	49.55	30.98	34.42	75	25	0
0.5	37.26	-71.34	-56.9	36.86	44.8	43.98	80	20	0
0.7	36.58	-74.58	-37.46	40.91	24.85	33.91	83	17	0
0.91	41.18	-80.92	-5.47	42.07	-24.82	-4.44	70	30	0

Discussion

HCCI combustion is a potential candidate for dealing with the more stringent regulations on pollutant emissions from motor vehicles, as well as the increasing concern about the emission of carbon dioxide from fossil fuels and its impact on global warming. Despite the promising results obtained in preliminary studies [1], the lack of autoignition control has delayed its launch in the engine industry. Further research is required to address this challenge. Computer simulation, however, has become a dominant tool in making HCCI a reality and in the quest for control strategies for HCCI. In addition, it has higher flexibility and lower cost compared with real engine experiments [6]. The main objective of this paper is to select a fuel surrogate capable of reproducing the autoignition characteristics of commercial fuels as well as fuel properties. To achieve this, we first provide a methodology to optimally and robustly select the simulations to obtain the best inferences of the parameters when a class of plausible responses (3) may be used to model the RMSE (2). Once the model is fitted, the optimal choice of the surrogate is obtained from solving the optimization problem (6).

Most experiments are performed in order to develop and optimize the most adequate engine control strategies do not use the tools of optimal experimental design. Consequently, data analysis will be informative only if data are themselves. Even if this methodology is applied, departures of the considered model may occur during the experimental stage invalidating, thus, the collected information in the observations. It may be due to extreme conditions under which the experiments are carried out or due to the lack of knowledge about model suitability. Practitioners have little information about model prior to running the experiments and the large number of factors involved in the combustion, as well as the difficulty of considering all possible operating conditions, implies that small deviations from the considered model may occur in a real situation. Therefore, there is a need to use a methodology which obtains robust results against possible departures from the considered model. Using this novel idea, engine researchers and designers may optimally design their experiments depending on their level of certainty about the assumed model. The main challenge faced by this research was the complex optimization design problem to be solved. A genetic algorithm was developed especially for this purpose, which is original to the authors. We should underline that this algorithm may be used regardless of the number of components or the model considered. This methodology is, therefore, general and may be applied in other engine processes where an optimal formulation is required.

Conclusions

To conclude, the experiments obtained with the proposed methodology enhance the efficiency for purposes of estimation. Thus, depending on the level of certainty, the quality of similar designs may be different by up to 80%. In particular, considering the optimal-robust designs shown in Fig (1) and the experimental setup of [8], the optimal formulations that best match

the original fuel to simulate HCCI autoignition are shown in Table 3. According to these results, optimal mixtures obtained when bias is emphasized ($v \geq 0.3$) agree with the results obtained in recent empirical studies [43]. In these cases, optimal robust designs incorporate interior points of the simplex. Therefore, this idea supports the need to consider optimal-robust designs when the model form is not completely known.

Supporting information

S1 Appendix. Genetic algorithm for computing *D*−optimal robust mixture designs.
(PDF)

S1 Table. Data used to illustrate the results with the obtained designs through the proposed methodology.
(XLSX)

S2 Table. *D*−optimal robust designs obtained from Fig 1 and values of RMSE achieved during the experiments.
(PDF)

Author Contributions

Conceptualization: Irene García-Camacha Gutiérrez, Raúl Martín Martín, Josep Sanz Argent.

Data curation: Josep Sanz Argent.

Formal analysis: Irene García-Camacha Gutiérrez, Raúl Martín Martín.

Funding acquisition: Raúl Martín Martín.

Investigation: Irene García-Camacha Gutiérrez, Raúl Martín Martín.

Methodology: Irene García-Camacha Gutiérrez, Raúl Martín Martín.

Resources: Raúl Martín Martín, Josep Sanz Argent.

Software: Irene García-Camacha Gutiérrez, Josep Sanz Argent.

Supervision: Raúl Martín Martín, Josep Sanz Argent.

Validation: Irene García-Camacha Gutiérrez, Raúl Martín Martín, Josep Sanz Argent.

Visualization: Irene García-Camacha Gutiérrez, Raúl Martín Martín.

Writing – original draft: Irene García-Camacha Gutiérrez, Raúl Martín Martín.

Writing – review & editing: Irene García-Camacha Gutiérrez, Raúl Martín Martín, Josep Sanz Argent.

References

1. Hasan MM, Rahman MM. Homogeneous charge compression ignition combustion: Advantages over compression ignition combustion, challenges and solutions. Renewable and Sustainable Energy Reviews. 2016; 57:282–291. https://doi.org/10.1016/j.rser.2015.12.157

2. Najt PM, Foster DE. Compression-Ignited Homogeneous Charge Combustion. In: SAE Technical Paper. SAE International; 1983. Available from: https://doi.org/10.4271/830264.

3. Caton PA, Simon AJ, Gerdes JC, Edwards CF. Residual-effected homogeneous charge compression ignition at a low compression ratio using exhaust reinduction. International Journal of Engine Research. 2003; 4(3):163–177. https://doi.org/10.1243/146808703322223306

4. Yamaoka S, Kakuya H, Nakagawa S, Okada T, Shimada A, Kihara Y. HCCI Operation Control in a Multi-Cylinder Gasoline Engine. In: SAE Technical Paper. SAE International; 2005.

5. Fathi M, Saray RK, Checkel MD. The influence of Exhaust Gas Recirculation (EGR) on combustion and emissions of n-heptane/natural gas fueled Homogeneous Charge Compression Ignition (HCCI) engines. Applied Energy. 2011; 88(12):4719–4724. https://doi.org/10.1016/j.apenergy.2011.06.017

6. Sherazi HI, Li Y. Homogeneous charge compression ignition engine: a technical review. In: The 17th International Conference on Automation and Computing. IEEE; 2011. p. 315–320.

7. Szymkowicz PG, Benajes J. Single-cylinder engine evaluation of a multi-component diesel surrogate fuel at a part-load operating condition with conventional combustion. Fuel. 2018; 226:286–297. https://doi.org/10.1016/j.fuel.2018.03.157

8. Hernandez JJ, Sanz-Argent J, Benajes J, Molina S. Selection of a diesel fuel surrogate for the prediction of auto-ignition under HCCI engine conditions. Fuel. 2008; 87(6):655–665. https://doi.org/10.1016/j.fuel.2007.05.019

9. Xu Q, Liu Z, Zhang C, Hui X, Lin Y. Surrogate formulation for a coal-based jet fuel using a mixing model based on explicit equations and artificial neural network. Fuel. 2017; 210:262–271. https://doi.org/10.1016/j.fuel.2017.08.055

10. Huang Z, Xia J, Ju D, Lu X, Han D, Qiao X, et al. A six-component surrogate of diesel from direct coal liquefaction for spray analysis. Fuel. 2018; 234:1259–1268. https://doi.org/10.1016/j.fuel.2018.07.138

11. Szymkowicz PG, Benajes J. Development of a Diesel Surrogate Fuel Library. Fuel. 2018; 222:21–34. https://doi.org/10.1016/j.fuel.2018.01.112

12. Yu W, Zhao F, Yang W, Tay K, Xu H. Development of an optimization methodology for formulating both jet fuel and diesel fuel surrogates and their associated skeletal oxidation mechanisms. Fuel. 2018; 231:361–372. https://doi.org/10.1016/j.fuel.2018.05.121

13. Yu J, Ju Y, Gou X. Surrogate fuel formulation for oxygenated and hydrocarbon fuels by using the molecular structures and functional groups. Fuel. 2016; 166:211–218. https://doi.org/10.1016/j.fuel.2015.10.085

14. Perez PL, Boehman AL. Experimental Investigation of the Autoignition Behavior of Surrogate Gasoline Fuels in a Constant-Volume Combustion Bomb Apparatus and Its Relevance to HCCI Combustion. Energy & Fuels. 2012; 26(10):6106–6117. https://doi.org/10.1021/ef300503b

15. Perez PL, Boehman AL. Effects of the Chemical Structure and Composition of Surrogate Gasoline Fuels on Homogeneous Charge Compression Ignition Combustion in a Single-Cylinder Engine. Energy & Fuels. 2014; 28(5):3377–3390. https://doi.org/10.1021/ef302106m

16. Hernández JJ, Sanz-Argent J, Carot JM, Jabaloyes JM. Modelling of the auto-ignition angle in diesel HCCI engines through D-optimal design. Fuel. 2010; 89(9):2561–2568. https://doi.org/10.1016/j.fuel.2009.11.014

17. Chen L, Zhang Z, Gong W, Liang Z. Quantifying the effects of fuel compositions on GDI-derived particle emissions using the optimal mixture design of experiments. Fuel. 2015; 154:252–260. https://doi.org/10.1016/j.fuel.2015.03.081

18. Kee RJ, Rupley FM, Miller JA, Coltrin ME, Grcar JF, Meeks Eea. CHEMKIN Release 4.0, Reaction Design, Inc. San Diego, CA; 2004.

19. Benajes JV, López JJ, Novella R, García A. Advanced methodology for improving testing efficiency in a single-cylinder research diesel engine. Experimental Techniques. 2008; 32(6):41–47. https://doi.org/10.1111/j.1747-1567.2007.00296.x

20. Komninos NP, Rakopoulos CD. Modeling HCCI combustion of biofuels: A review. Renewable and Sustainable Energy Reviews. 2012; 16(3):1588–1610. https://doi.org/10.1016/j.rser.2011.11.026

21. Aceves SM, Flowers DL, Martinez-Frias J, Smith JR, Dibble R, Au M, et al. HCCI Combustion: Analysis and Experiments. In: SAE Technical Paper. SAE International; 2001.

22. Jia M, Xie M. A chemical kinetics model of iso-octane oxidation for HCCI engines. Fuel. 2006; 85(17-18):2593–2604. https://doi.org/10.1016/j.fuel.2006.02.018

23. Curran HJ, Fisher EM, Glaude PA, Marinov NM, Pitz WJ, Westbrook CK, et al. Detailed Chemical Kinetic Modeling of Diesel Combustion with Oxygenated Fuels. In: SAE Technical Paper. SAE International; 2001.

24. Hasegawa R, Sakata I, Koyama T, Yanagihara H. Numerical Analysis of Ignition Control in HCCI Engine. In: SAE Technical Paper. SAE International; 2003.

25. Noel L, Maroteaux F, Ahmed A. Numerical Study of HCCI Combustion in Diesel Engines Using Reduced Chemical Kinetics of N-Heptane With Multidimensional CFD Code. In: SAE Technical Paper. SAE International; 2004.

26. Westbrook CK, Pitz WJ, Herbinet O, Curran HJ, Silke EJ. A comprehensive detailed chemical kinetic reaction mechanism for combustion of n-alkane hydrocarbons from n-octane to n-hexadecane. Combustion and Flame. 2009; 156(1):181–199. https://doi.org/10.1016/j.combustflame.2008.07.014

27. Pitz W, Seiser R, Bozzelli J, Seshadri K, Chen C, Da Costa I, et al. Chemical kinetic study of toluene oxidation. Lawrence Livermore National Lab., CA (US); 2001.

28. Dagaut P, Pengloan G, Ristori A. Oxidation, ignition and combustion of toluene: Experimental and detailed chemical kinetic modeling. Physical Chemistry Chemical Physics. 2002; 4(10):1846–1854. https://doi.org/10.1039/B110282F

29. Bounaceur R, Da Costa I, Fournet R, Billaud F, Battin-Leclerc F. Experimental and modeling study of the oxidation of toluene. International journal of chemical kinetics. 2005; 37(1):25–49. https://doi.org/10.1002/kin.20047

30. Granata S, Faravelli T, Ranzi E. A wide range kinetic modeling study of the pyrolysis and combustion of naphthenes. Combustion and Flame. 2003; 132(3):533–544. https://doi.org/10.1016/S0010-2180(02)00465-0

31. Buda F, Heyberger B, Fournet R, Glaude PA, Warth V, Battin-Leclerc F. Modeling of the gas-phase oxidation of cyclohexane. Energy & fuels. 2006; 20(4):1450–1459. https://doi.org/10.1021/ef060090e

32. Silke EJ, Pitz WJ, Westbrook CK, Ribaucour M. Detailed chemical kinetic modeling of cyclohexane oxidation. The Journal of Physical Chemistry A. 2007; 111(19):3761–3775. https://doi.org/10.1021/jp067592d

33. Curran HJ, Gaffuri P, Pitz WJ, Westbrook CK. A Comprehensive Modeling Study of n-Heptane Oxidation. Combustion and Flame. 1998; 114(1):149–177.

34. Andrae J, Johansson D, Björnbom P, Risberg P, Kalghatgi G. Co-oxidation in the auto-ignition of primary reference fuels and n-heptane/toluene blends. Combustion and Flame. 2005; 140(4):267–286. https://doi.org/10.1016/j.combustflame.2004.11.009

35. Scheffé H. Experiments with mixtures. Journal of the Royal Statistical Society. 1958; B(20):344–369.

36. Scheffé H. The simplex-centroid design for experiments with mixtures. Journal of the Royal Statistical Society. 1963; 2B(25):235–263.

37. Josep SA. Análisis cinético-químico del proceso de autoencendido del combustible diesel en condiciones de carga homogénea. Ediciones de la Universidad de Castilla-La Mancha; 2010. Available from: http://hdl.handle.net/10578/1442.

38. Huber PJ. Robustness and designs. In: A Survey of Statistical Design and Linear Models. J.N. (Ed.) North Holland;. p. 287–303.

39. Wiens DP. Minimax designs for approximately linear regression. Journal of Statistical Planning and Inference. 1992; 31(3):353–371. https://doi.org/10.1016/0378-3758(92)90142-F

40. Wiens DP. Designs for approximately linear regression: Maximizing the minimum coverage probability of confidence ellipsoids. Canadian Journal of Statistics; 21(1):59–70.

41. Daemi M, Wiens DP. Techniques for the construction of robust regression designs. Canadian Journal of Statistics. 2013; 41(4):679–695. https://doi.org/10.1002/cjs.11194

42. Atkinson A, Donev A, Tobias R. Optimum Experimental Designs, With SAS. Oxford Statistical Science Series. OUP Oxford; 2007.

43. Hernández JJ, Sanz-Argent J, Monedero-Villalba E. A reduced chemical kinetic mechanism of a diesel fuel surrogate (n-heptane/toluene) for HCCI combustion modelling. Fuel. 2014; 133:283–291. https://doi.org/10.1016/j.fuel.2014.05.029