Nuclear Engineering and Technology 49 (2017) 1423–1430

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Gas detonation cell width prediction model based on support vector regression

Jiyang Yu [a], Bingxu Hou [b, *], Alexander Lelyakin [c], Zhanjie Xu [c], Thomas Jordan [c]

[a] Department of Engineering Physics, Tsinghua University, Qinghuayuan 1, 100084 Beijing, China
[b] CNNC China Zhongyuan Engineering Corporation, Xizhimenwai Street 140, 100044 Beijing, China
[c] Institute for Nuclear and Energy Technologies, Karlsruhe Institute of Technology, P.O. Box 3640, 76021 Karlsruhe, Germany

ARTICLE INFO

Article history:
Received 28 November 2016
Received in revised form
26 April 2017
Accepted 19 June 2017
Available online 14 July 2017

Keywords:
Detonation Cell Width
Hydrogen Safety
Machine Learning
Support Vector Regression

ABSTRACT

Detonation cell width is an important parameter in hydrogen explosion assessments. The experimental data on gas detonation are statistically analyzed to establish a universal method to numerically predict detonation cell widths. It is commonly understood that detonation cell width, λ, is highly correlated with the characteristic reaction zone width, δ. Classical parametric regression methods were widely applied in earlier research to build an explicit semiempirical correlation for the ratio of λ/δ. The obtained correlations formulate the dependency of the ratio λ/δ on a dimensionless effective chemical activation energy and a dimensionless temperature of the gas mixture. In this paper, support vector regression (SVR), which is based on nonparametric machine learning, is applied to achieve functions with better fitness to experimental data and more accurate predictions. Furthermore, a third parameter, dimensionless pressure, is considered as an additional independent variable. It is found that three-parameter SVR can significantly improve the performance of the fitting function. Meanwhile, SVR also provides better adaptability and the model functions can be easily renewed when experimental database is updated or new regression parameters are considered.

1. Introduction

After a severe accident in a nuclear power plant, hydrogen, mainly produced from zirconium–water reaction, can leak into the inner space of the containment. The hazard of hydrogen detonation threatens the integrity of the containment, as in the case of the Fukushima Daiichi nuclear disaster.

When gas detonation takes place, regularly or irregularly distributed cell structures are observed behind the shock wave. The average width of the cells is defined as the detonation cell width, which is represented by λ [1]. The value of λ depends on the initial thermodynamic condition of the flammable gaseous mixture, including the temperature and pressure and the concentrations of the components [2].

λ is commonly used to quantify the detonation risk of flammable gaseous mixtures. For a given gaseous mixture in a given confined volume, the hazard potential caused by detonation may be evaluated with the help of the dimensionless ratio of λ to the characteristic geometric size, noted as L. According to the deflagration-to-detonation transition criterion, which is also known in the field of hydrogen safety research as the λ-criterion, detonation is likely to happen when L/λ is greater than 7 [3]. Therefore, it is of significant importance to reliably predict λ for detonable gases.

To date, there has been no sophisticated theoretical expression to compute λ because of the complexity and uncertainty of the detonation phenomenon. Present solutions to predict the value of λ are all based on empirical or semiempirical correlations established by fitting a large amount of experimental data.

At present, there are two commonly used approaches to estimate λ [4]. One offers simple and direct determination of λ using the initial conditions. In Dorofeev et al [5], a correlation between λ and the initial concentrations of hydrogen and steam, and the initial temperature and pressure is presented. It is clear that such a fit can be used only within the range of the experimental data. Extrapolation of the fitted function beyond the range of measurements does not give reliable values. Another approach is based on an analysis of the correlation of the characteristic reaction zone widths, δ, with λ. The idea that λ and δ can be correlated was first

* Corresponding author.
E-mail address: h.bx@163.com (B. Hou).

http://dx.doi.org/10.1016/j.net.2017.06.014

proposed in [6]; such an approach allows estimations of detonation cell sizes for general gas mixtures apart from hydrogen–air–steam mixtures (e.g., hydrocarbon–air mixtures). Quite a few works in the literature [7–9] suggest that two main parameters can influence the stability of the wave and cellular structure. One is the dimensionless effective activation energy, which can be interpreted as a characteristic of the sensitivity of the global reaction time to variations of the postshock temperature caused by changes in the strength of the leading shock. The other is a dimensionless temperature describing the relation between the released chemical energy and the initial thermal energy. In Gavrikov et al [10], mixture composition and initial thermal–dynamic conditions are at first preconditioned into these two parameters. Then, using analytical expressions and the least squares fitting method, a correlation is built between λ/δ and the two dimensionless parameters.

The correlation in Gavrikov et al [10] achieved success in precisely predicting λ within a wide range of thermal conditions for hydrogen–oxidizer mixtures. However, the approach is incapable of predicting λ for hydrocarbon–oxidizer mixtures when the equivalence ratio is larger than 2. Moreover, an attempt to introduce a third parameter to obtain better fitness failed because unphysical oscillations showed up on the regression curves. Because earlier studies suggested that some other parameters could be significant in determining the λ/δ value [1,11], it is necessary to apply better mathematical tools to take more than two parameters into consideration and to improve the precision of the function.

Actually, the aforementioned two approaches are both based on parametric regression. This kind of regression demands an *a priori* assumption of the correlation form, which can be very complicated and difficult to achieve. Meanwhile, parametric regression suffers from weak robustness and poor adaptability performance [12]. When new data points are added and/or additional regression parameters are considered, the regression model can become invalid and needs to be replaced.

To overcome the disadvantages of parametric regression, a nonparametric regression method, support vector regression (SVR), is proposed and discussed in this study. It requires no assumptions about the form of the function, and thus it does not offer an explicit equation but a high-dimensional matrix that describes the relationship among the data points. The fitting process is totally driven by the data itself, leading to better fitness of data and higher stability. The adaptability of the method makes it easier to include new additional data points or new parameters.

The physical idea proposed by Gavrikov et al [10], that λ/δ is a function of dimensionless activation energy and dimensionless temperature, is used and developed in the study. A third parameter, dimensionless pressure, is taken into account to achieve better fitness of data and higher precision.

2. Definitions of parameters

Most of the definitions of the physical parameters involved in this paper utilize the ideas of Gavrikov et al [10]. Some modifications for better regression results are proposed.

2.1. Characteristic reaction zone width

There are several definitions of the characteristic reaction zone width, and in this paper the one expressed by Shepherd [13] is adopted, in which the width is defined as the distance between the leading shock and the location of the maximum rate of temperature rise. The value of the width can be approximately evaluated by the product of the characteristic reaction time, t_{ch}, and the postshock gas velocity, $\overline{D}$:

$$\delta = \overline{D} t_{ch} \tag{1}$$

The original definition of t_{ch} is a period during which the consumption of the limited component is equal to 90% of the total consumption defined by the final equilibrium state. The term "limited component" stands for fuel for lean mixtures or oxidizer for rich mixtures. With such a definition, however, discontinuity occurs for t_{ch} when the equilibration ratio is close to 1, at which point a transition of the reaction process indicator occurs between the fuel and the oxidizer. Therefore, the definition of t_{ch} used in this paper is expressed as an oxidizer consumption equal to 90% of the total consumption defined by the final equilibrium state.

2.2. Dimensionless activation energy

The dimensionless activation energy is defined as $E_a/R/T_{ps}$, where E_a is the effective activation energy, R the gas constant, and T_{ps} the postshock temperature of the gas. The definition of E_a is based on calculations of the characteristic reaction times $t_{ch,1}$ and $t_{ch,2}$ behind shocks with two different traveling speeds D_1 and D_2:

$$\frac{E_a}{R} = \frac{\ln(t_{ch,1}/t_{ch,2})}{1/T(D_1) - 1/T(D_2)} \tag{2}$$

where $T(D_1)$ and $T(D_2)$ are temperatures of the gas behind the shocks with speeds D_1 and D_2, respectively. Meanwhile, these temperatures are also the actual starting temperatures of the chemical reactions. The parameter E_a in Eq. (2) describes the mean sensitivity of the reaction time to the changes in shock strength between D_1 and D_2.

According to an analysis of the reaction conditions in actual multidimensional detonations, as well as prediction accuracy comparisons, Gavrikov et al [10] chose the pair $(1.0D_{cj}, 1.6D_{cj})$ as the values of (D_1, D_2) from the following four candidates:

$$(1.0D_{cj}, 1.6D_{cj}); (1.2D_{cj}, 1.6D_{cj}); (1.0D_{cj}, 1.4Dcj); (1.0D_{cj}, 1.4D_{cj}).$$

where D_{cj} is the Chapman–Jouguet velocity. Two more pairs, $(0.9D_{cj}, 1.6D_{cj})$ and $(0.9D_{cj}, 1.4D_{cj})$, are tested in this study. It is found empirically that, among the six candidate pairs, the pair of $(0.9D_{cj}, 1.6D_{cj})$ always offers the best predictions. According to the classic one-dimensional ZND detonation model, detonation does not occur when D_{cj} is not fully reached. However, to compute activation energy in this model, the speed of $0.9D_{cj}$ in the pair $(0.9D_{cj}, 1.6D_{cj})$ is taken merely as a reference speed. Thus, the calculation of the activation energy is based on the shock strength, regardless of the occurrence of detonation. Therefore, activation energy is computed with $D_1 = 0.9D_{cj}$ and $D_2 = 1.6D_{cj}$, while their mean value $\overline{D} = 1.25D_{cj}$ is used to compute the postshock properties.

2.3. Dimensionless temperature

The ratio of the energy released in reaction to the initial energy of the unignited mixture is expressed as $Q/(C_vT_0)$, where Q is the reaction heat, C_v the constant-volume heat capacity, and T_0 is the initial temperature of the unignited mixture. This ratio describes the relationship between the released chemical energy and the initial thermal energy. However, the value of Q is not defined for the real thermochemistry of a combustible mixture. Meanwhile, a dimensionless temperature T_{vn}/T_0 is related to $Q/(C_vT_0)$ and the variation in the ratio $[Q/(C_vT_0)]/(T_{vn}/T_0)$ is small for most detonable mixtures. It is estimated that, for strong waves of which the Mach number is larger than 2, the variation is within ±25% of an average value of 4.8. Therefore, T_{vn}/T_0 is selected to replace $Q/(C_vT_0)$. T_{vn} is

the von Neumann temperature or the temperature of the von Neumann state, which is the maximum extent to which the flammable gaseous mixture can be compressed by a shock wave before chemical reaction starts. The value of T_{vn} in the study varies between 1,300 K and 2,400 K, while the value of T_{vn}/T_0 is within the range of 2–8.

2.4. Third dimensionless variable

To improve the performance of the regression in this study, several potential dimensionless parameters are tested as a third variable for regression besides $E_a/R/T_{ps}$ and T_{vn}/T_0. The parameters include: (1) the ratio of induction time to energy release time, t_{ind}/t_{rel}; (2) the Mach number of the postshock gas, Ma; (3) the expansion ratio of the gas during the reaction, σ; (4) the heat capacity ratio of the reactant, γ; and (5) the ratio of the postshock pressure to the initial pressure, p_{ps}/p_0. It is found that, among these parameters, considering p_{ps}/p_0 as the third independent regression variable leads to the best regression result. Therefore, $E_a/R/T_{ps}$, T_{vn}/T_0, and p_{ps}/p_0 are decided on as independent variables for three-variate regression.

The original experimental data include initial temperature, pressure, concentrations of the gaseous components, and measured detonation cell width for different detonation cases. These data are mainly collected from Kaneshige and Shepherd's detonation database [14], while some are unpublished experimental data supplied by the Karlsruhe Institute of Technology. The data are preconditioned into $E_a/R/T_{ps}$, T_{vn}/T_0, p_{ps}/p_0, and $\lg(\lambda/\delta)$ using an open-source chemical dynamics code from Goodwin et al. [15]. Regressions are then performed with ($E_a/R/T_{ps}$, T_{vn}/T_0) or ($E_a/R/T_{ps}$, T_{vn}/T_0, p_{ps}/p_0) as independent variables and $\lg(\lambda/\delta)$ as dependent regression variable.

3. Support vector regression

When a parametric method such as least square regression is applied to fit the data, the fitting function attempts to go through every data point. As noise exists in experimental data, such an attempt without the consideration of the noise can lead to overfitting, resulting in heavy oscillation and bad stability of the fitting curve, as shown in Fig. 1. To overcome this disadvantage of the classical parametric regression method, SVR is applied in this study.

In the categories of machine learning, SVR belongs to supervised learning [16]. During regression, the learning program is presented with example inputs and their desired outputs. This process is called training; the goal is to summarize a rule that can map new inputs to predicted outputs. SVR is based on the theory of support vector machines. Detailed information about this method can be found in previously published studies [17,18].

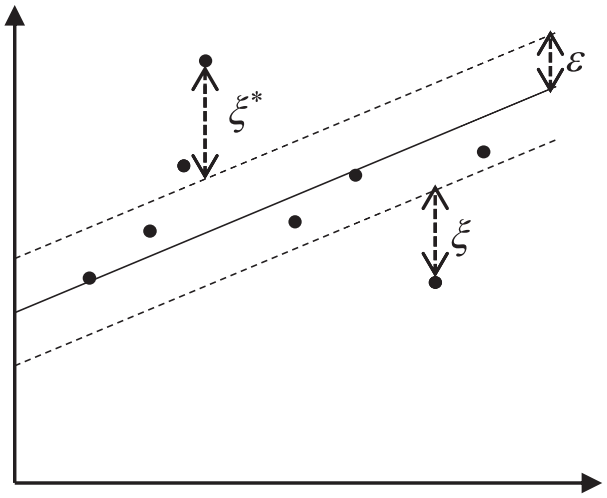

Fig. 2. One-dimensional linear SVR.

Assuming $f(x)$ is the fitting function, an error limit ε is set to $f(x)$ and data noise between $-\varepsilon$ and $+\varepsilon$ is neglected, as shown in Fig. 2. Non-negative slack variables, ξ and ξ^*, are defined to measure the deviation of the data points of which the error is less than $-\varepsilon$ or larger than $+\varepsilon$. Therefore, the task of SVR is to search for a function that covers as many as data points as possible between $f(x)-\varepsilon$ and $f(x)+\varepsilon$ and to minimize the sum of ξ and ξ^*. The final form of the fitting function is given as follows:

$$y = f\left(\vec{x}\right) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)k\left(\vec{x}_i, \vec{x}\right) + b \tag{3}$$

where $k(x_i, x)$ is the kernel function, x_i is the ith data point, x is the independent variable, and α_i and α_i^* are regression parameters determined during the regression process.

There are many kinds of kernel functions; the most commonly used one is called the Gaussian kernel:

$$k\left(\vec{x}, \vec{x}'\right) = \exp\left(-\gamma\left|\vec{x} - \vec{x}'\right|^2\right) \tag{4}$$

where γ is the kernel width representing the dispersion degree of the training data.

Before doing SVR regression, three empirical parameters must be manually decided on. These parameters are called hyperparameters. In SVR, hyperparameters include ε and γ, mentioned above, and a positive constant C determining the trade-off between the flatness of $f(x)$ and the amount up to which deviations larger

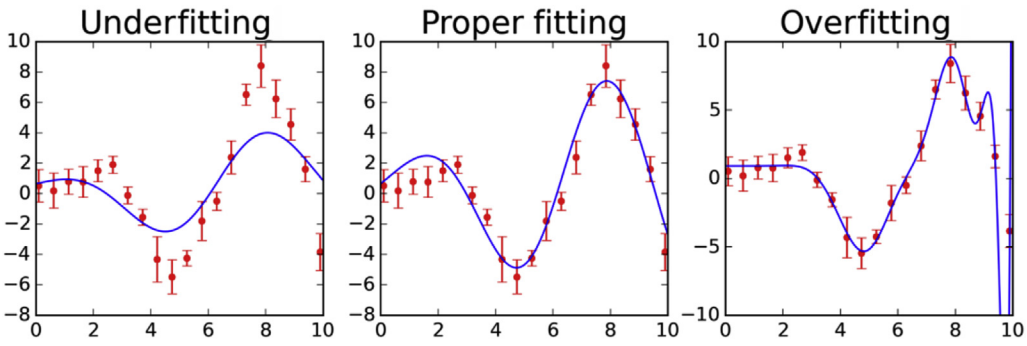

Fig. 1. Examples of underfitting and overfitting.

than ε are tolerated.

Ten-fold cross validation is applied to search for the optimal combination of hyperparameters. To perform the 10-fold cross validation, first, estimated values of the hyperparameters should be obtained beforehand. The whole database is then randomly divided into 10 equal-sized sets. Nine of the data sets are selected as the training set and are used for regression to obtain the fitting function. The one set left is used as a testing set to compute the correlation coefficient for the fitting function. This process is repeated 10 times with each set as a testing set and the other nine sets as training sets. The mean value of the 10 correlation coefficients computed during the above processes is called the cross-validation score. The score normally varies between 0 and 1 and is a measurement of the performance of regression. Higher cross-validation scores mean that, with the given hyperparameters, the regression result offers a better balance between the fitness to the measured data and the stability of the fitting function. Further information on cross validation can be found in Kohavi [19].

In our case, at first, possible ranges for ε, γ, and C are assumed beforehand. Then the values of the three parameters are discretized with proper distances. The values of ε, γ, and C are combined and a three-dimensional testing matrix is constructed. The 10-fold cross-validation score with each set of (ε, γ, C) then can be computed. The optimal combination of the hyperparameters is found by positioning the maximum value of the corresponding score.

4. Comparisons of regression results

The regression results of SVR and parametric regression are compared with the experimental data used as testing points. Regression work using SVR is performed with the open-source Python toolbox, scikit-learn [20]. The toolbox also provides a module to compute the cross-validation scores, which helps in finding the optimal hyperparameters. As for the parametric regressions, Gavrikov et al's [10] method is applied to construct a correlation using different simple analytical expressions and least square fitting. The coefficient of the determinant, R^2, is applied as an important indicator of the fitness to the training points. The definition of R^2 is as in Eq. (5), where y_i is the observation value of the data, f_i the value predicted by the fitting function, and $\bar{y}$ the mean of y_i.

$$R^2 \equiv 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{5}$$

4.1. Two-variate regression

The regression with the form:

$$\lg\left(\frac{\lambda}{\delta}\right) = f\left(\frac{E_a}{RT_{ps}}, \frac{T_{vn}}{T_0}\right) \tag{6}$$

is performed with two-variate SVR and the parametric method.

For the parametric method, the correlation that gives the least deviation from the data is given as follows:

$$\begin{aligned}\lg(\lambda/\delta) = {} & aX + bY + cY^2 + dXY + eXY^2 + f\ln X + g\ln Y + hX/Y \\ & + iY/X + jX^2/Y + kY^2/X + lX/Y^2 + mY/X^2 + n\end{aligned} \tag{7}$$

where a = 6.638580294; b = 9.416228276; c = −0.241216936; d = −0.903635419; e = 0.026723775; f = −30.64629363; g = 5.252928942; h = −1.84853496; i = −44.68907842; j = −0.26945211; k = 0.904164774; l = 4.701279894; m = 48.23143082; n = 20.86468644, and $X = E_a/(RT_{ps})$, $Y = T_{vn}/T_0$.

For SVR, the optimistic hyperparameters are C = 1,000, γ = 0.0021, and ε = 0.006.

4.2. Three-variate regression

The regression with the form

$$\lg\left(\frac{\lambda}{\delta}\right) = f\left(\frac{E_a}{RT_{ps}}, \frac{T_{vn}}{T_0}, \frac{p_{ps}}{p_0}\right) \tag{8}$$

is performed with three-variate SVR and the parametric method.

For the parametric method, the correlation that gives the least deviation from the data is given as

$$\begin{aligned}\lg(\lambda/\delta) = {} & aY + bZ + cX^2 + dXY + eXZ + fXYZ + g\ln X + h\ln Y \\ & + i\ln Z + jX/Z + kY/Z + lX/Y + mZ/X + nY/X \\ & + oX^2Z + pY^2Z + qXZ^2 + rX^3 + sY^3 + tX^2/Z \\ & + uY^2/X + vY^2/Z + wZ^2/Y + xX/Y^2 + yX/Z^2 \\ & + zY/Z^2 + const\end{aligned} \tag{9}$$

where a = 16.1711908; b = −1.3077105; c = 0.6118572; d = −1.6246259; e = 0.0829447; f = 0.0092600; g = 28.7740088; h = −58.7212346; i = 25.5080873; j = 61.4810312; k = 103.4928602; l = −34.1012837; m = 0.06369708; n = 5.3476021; o = −0.0032753; p = 0.0019715; q = −0.0004605; r = −0.0125591; s = −0.0241242; t = −0.7328976; u = −0.9088536; v = −7.4249634; w = 0.0240641; x = 25.3028072; y = −198.9036880; z = −546.7832086; $const$ = −60.8717020; and $X = E_a/(RT_{ps})$, $Y = T_{vn}/T_0$, $Z = p_{ps}/p_0$.

For SVR, the optimistic hyperparameters are C = 1,000, γ = 0.0035, and ε = 0.001.

The comparisons between the experimental data and the prediction results by different regression methods are presented in

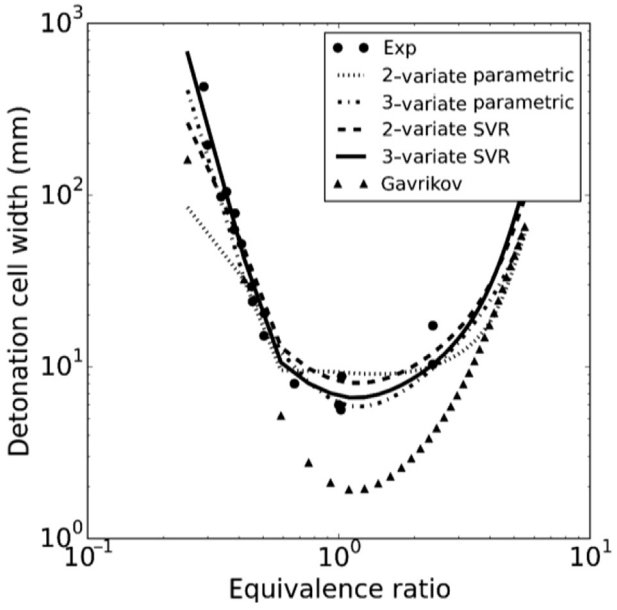

Fig. 3. Comparison among predicted values and experimental data for H_2–air mixture, initialed at 1 atm, 500 K. SVR, support vector regression.

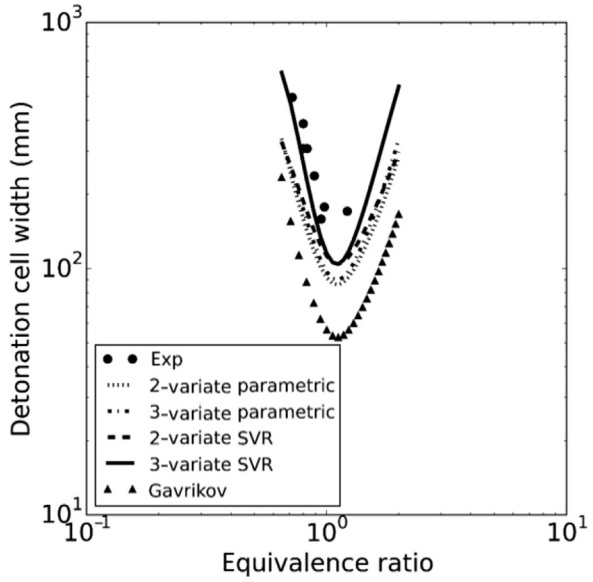

Fig. 4. Comparisons among predicted values and experimental data for H_2–air–15% CO_2 mixture, initialed at 1 atm, 293 K. SVR, support vector regression.

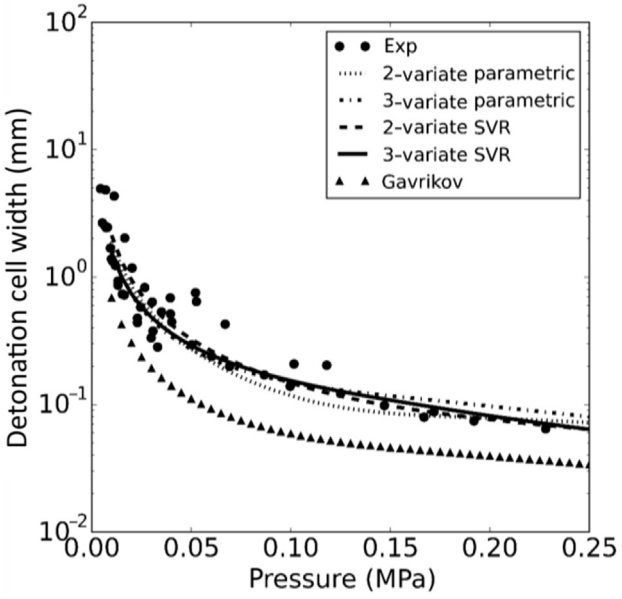

Fig. 5. Comparisons among predicted values and experimental data for 66% H_2–air initialed at 293 K, different pressures. SVR, support vector regression.

Figs. 3–7. The predictions by Gavrikov et al's [10] original correlation are also presented, though the database and the method of calculating the parameters in his study might be different.

It can be found that, apart from the hydrocarbon–oxidizer case in Fig. 7, all methods (two/three-variate SVR/parametric regressions) offer good accuracy in predicting detonation cell width for H_2–O_2–dilutant mixtures. Meanwhile, predictions by both methods with p_{ps}/p_0 as a third independent regression variable show improved agreement with the experimental data, especially in Figs. 3 and 4. Compared with those of three-variate parametric regression, three-variate SVR results show much better fitness to the measured data, especially in the cases of hydrocarbon–oxidizer mixtures, shown in Fig. 7, in which deteriorated stability of curves occurred for three-variate parametric regression.

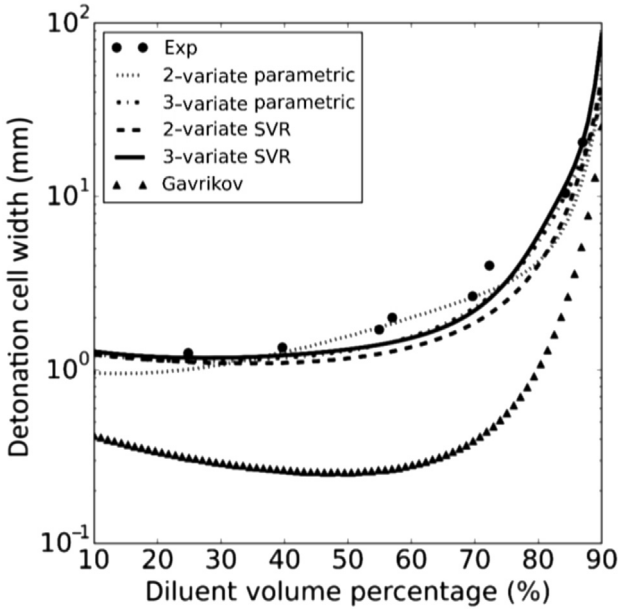

Fig. 6. Comparisons among predicted values and experimental data for H_2–O_2–Ar mixture, initialed at 1 atm, 293 K. SVR, support vector regression.

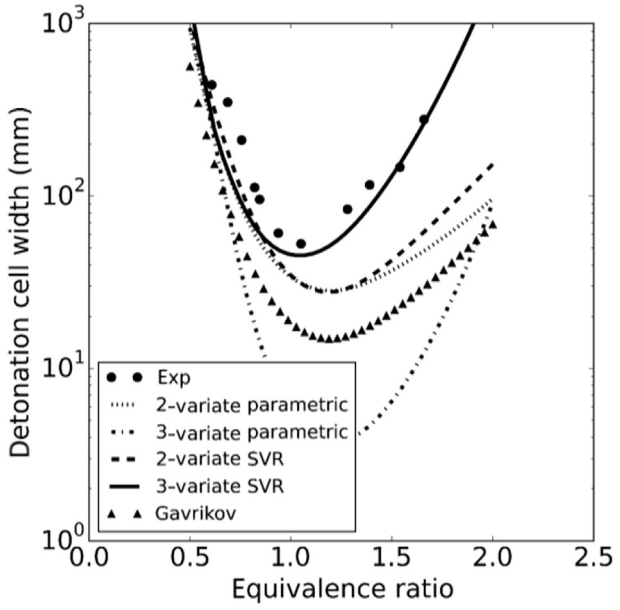

Fig. 7. Comparisons among predicted values and experimental data for C_3H_8–O_2 mixture, initialed at 1 atm, 293 K. SVR, support vector regression.

The values of R^2 for all the methods are presented in Table 1, which shows that three-variate SVR offers the best fitness to the training data. Overall, it can be concluded that three-variate SVR provides the best regression results.

4.3. Three-variate SVR predictions for all testing points

In Sections 4.1 and 4.2, only the most representative comparisons are presented. The prediction results of the best regression method, three-variate SVR, for all testing points are shown in Figs. 8–12.

Table 1
Values of R^2 for each regression method.

Regression method	Two-variate parametric	Two-variate SVR	Three-variate parametric	Three-variate SVR
R^2	0.771	0.743	0.843	0.876

SVR, support vector regression.

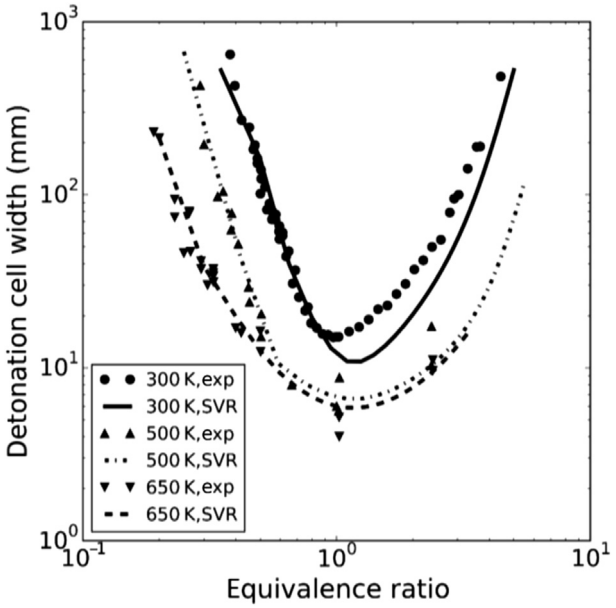

Fig. 8. Three-variate support vector regression (SVR)-predicted and measured λ versus equivalence ratio at different initial temperatures for H_2–air mixtures, 1 atm.

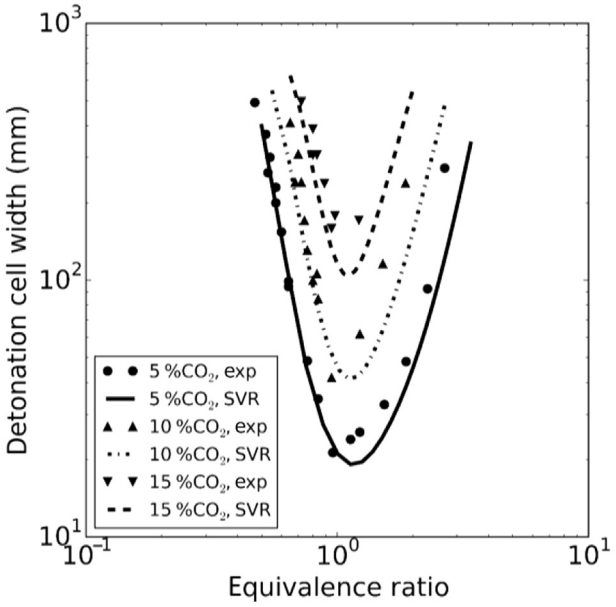

Fig. 9. Three-variate support vector regression (SVR)-predicted and measured λ versus equivalence ratio with different initial CO_2 concentrations for H_2–air–CO_2 mixtures, 1 atm, 293 K.

It can be concluded that the model is able to give reliable estimations of the cell width for a broad scope of detonable mixtures. The deviation between calculated values and experimental data is within the laboratory tolerance of the experimental data. Actually, the predictions of three-variate SVR can fit the experimental data very well within the valid range of independent variables.

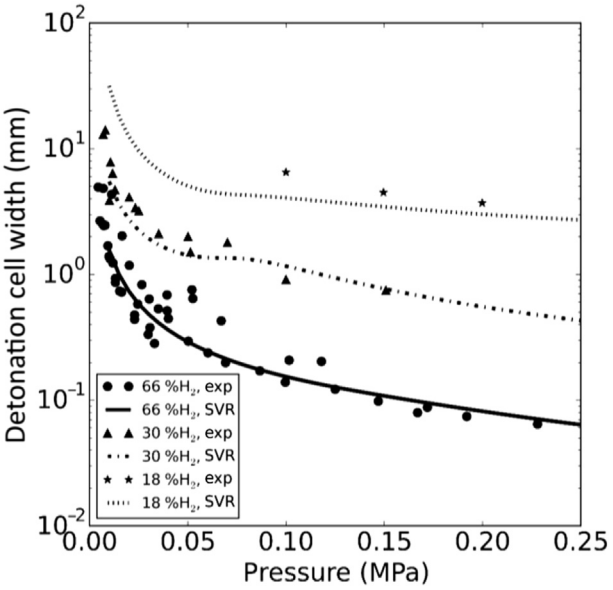

Fig. 10. Three-variate support vector regression (SVR)-predicted and measured λ versus initial pressure with different initial H_2 concentrations for H_2–air mixtures, 1 atm, 300 K.

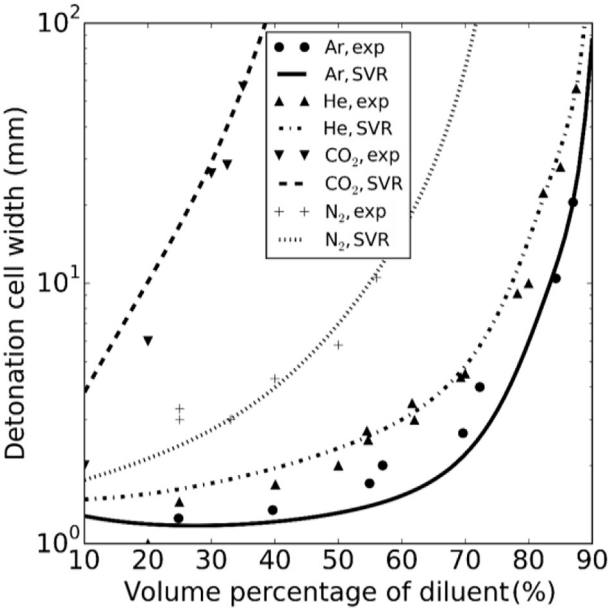

Fig. 11. Three-variate support vector regression (SVR)-predicted and measured λ versus volume percentage of different diluents for stoichiometric H_2–O_2–diluent mixtures, 1 atm, 293 K.

$$\begin{aligned} &E_a/R/T_{ps} : 4 \sim 15 \\ &T_{vn}/T_0 : 2 \sim (104 - 4x)/11 \\ &p_{ps}\Big/p_0 : 10 \sim (2020 - 120x)/11 \\ &\text{where } x = E_a/R/T_{ps} \end{aligned} \tag{10}$$

5. Discussion

One thing to be discussed is why a third regression parameter of p_{ps}/p_0 can improve the behavior of the regression to such a degree. One possible reason may be that the status change of the gaseous mixture before and after the shock wave is considered by the

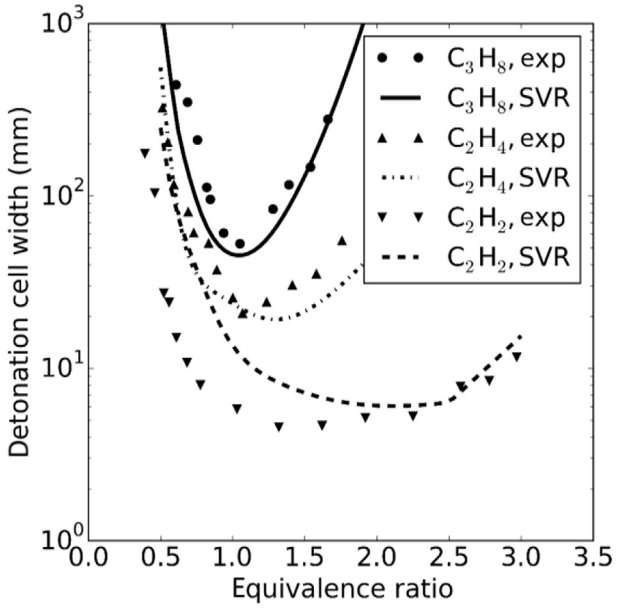

Fig. 12. Three-variate support vector regression (SVR)-predicted and measured λ versus equivalence ratio with different hydrocarbon–O_2 mixtures, 1 atm, 293 K.

introduction of the third parameter. According to earlier introductions to $E_a/R/T_{ps}$ and T_{vn}/T_0, the physical meanings of both the terms involve only the chemical reaction process itself. However, the mixture is compressed by the shock before the reaction. The pressure then rises to a value that is assumed to be constant during the reaction. The addition of p_{ps}/p_0 considers the pressure increase, and links the initial state of the gaseous mixture to the starting point of the reaction.

Another thing is the selection of an indicator to indicate the reaction process when defining t_{ch}. In Gavrikov et al [10], the concentration of the limited component is used as the indicator. However, such a definition causes unexpected discontinuity in the curves of t_{ch} when the equivalence ratio is close to 1. This problem is solved by applying the concentration of the oxidizer as the indicator, but another problem arises when more than one species of oxidizer exists. Other physical parameters are also considered. A proper indicator requires the parameter to vary monotonously and the variation trend should be the same for all detonation reactions. Temperature is abandoned because its decrease can be observed for $1.6D_{cj}$ detonation in certain conditions. The situation is similar for Gibbs free energy. Entropy, increasing in all spontaneous reactions, seems to be a perfect indicator. However, the regression results are not as good as those obtained using the concentration of oxidizer for single-oxidizer detonations. Nevertheless, entropy remains a potential indicator for future study.

6. Conclusions

Compared with the classical parametric methods, SVR, which is nonparametric and based on machine learning, achieves a more universal and more accurate model in predicting detonation cell width for gaseous mixtures with given compositions and thermal–dynamic conditions. Cross validation is applied to determine the optimal hyperparameters that make the best balance between the fitness to the training data and the stability of the function.

The addition of the ratio p_{ps}/p_0 as a third independent variable in the regression is found to ameliorate the regression results to a significant degree. The three-variate SVR process produces the best prediction results because of the higher R^2 value, that is, better fitness to training data and better agreement to measured data of the testing points, that is, higher prediction accuracy.

Another advantage of SVR is its strong adaptability. Once new data are added in the training set, or new parameters are considered as independent variables, the model can be modified simply by using the cross-validation tool to decide on another set of optimal hyperparameters. As for parametric regression, updating of the database or the parameters means a complete change in the correlation, of which the form is usually very complicated and hard to assume.

As for practical applications of SVR, though there is no explicit correlation that can provide a straightforward way to compute the width, the regression results can be stored in a text file that is updated only when the regression is redone. The information in this file is loaded in the computer's memory before making new predictions. The time cost in this step is tiny, so the SVR result, considering its high prediction precision, is suitable for industrial computations.

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgments

The work is financially supported by National Key Laboratory of Science and Technology on Reactor System Design Technology (HT-A100K-02-201406). Meanwhile, the authors appreciate Dr M. Kuznetsov of IKET, Karlsruhe Institute of Technology for his kind support on a part of the experimental data used in the study.

References

[1] J. Shepherd, Detonation in gases, Proc. Combust. Inst. 32 (2009) 83–98.
[2] G. Ciccarelli, T. Ginsberg, J.L. Boccio, Detonation Cell Size Measurements in High-temperature Hydrogen-air-steam Mixtures at the BNL High-temperature Combustion Facility, Rep. No. NUREG/CR–6391; BNL-NUREG–52482, U.S. Nuclear Regulatory Commission, Rockville, MD, 1997.
[3] S.B. Dorofeev, V.P. Sidorov, M.S. Kuznetsov, I.D. Matsukov, V.I. Alekseev, Effect of scale on the onset of detonations, Shock Waves 10 (2000) 137–149.
[4] W. Breitung, C. Chan, S. Dorofeev, A. Eder, B. Gelfand, M. Heitsch, R. Klein, A. Malliakos, E. Shepherd, E. Studer, P. Thibault, Flame Acceleration and Deflagration-to-detonation Transition in Nuclear Safety, Nuclear Safety Rep. No. NEA/CSNI/R(2000)7, OECD Nuclear Energy Agency, Issy-les-Moulineaux, France, 2000, pp. 1–455.
[5] S. Dorofeev, A. Bezmelnitsin, V. Sidorov, DDT Scaling for Severe Accidents. Detonation Cell Size Data as a Function of Composition and Initial Conditions, Rep. No. RRC "Kurchatov Institute" RRC KI, 80–05, 1997.
[6] K. Shchelkin, Y. Troshin, Gas Dynamics of Detonations, Mono Book Corp., Baltimore, MD, 1965.
[7] J. Shepherd, I.O. Moen, S.B. Murray, P.A. Thibalut, Analyses of the cellular structure of detonations, in: Symposium (International) on Combustion (Vol. 21, No. 1)Elsevier, Amsterdam, Netherlands, 1988, pp. 1649–1658.
[8] H. Lee, D. Stewart, Calculation of linear detonation instability: one-dimensional instability of plane detonation, J. Fluid Mech. 216 (1990) 103–132.
[9] S. Schöffel, F. Ebert, A numerical investigation of the reestablishment of a quenched gaseous detonation in a Galilei-transformed system, in: Proceedings of the 16th International Symposium on Shock Tubes and Waves, Aachen, FRG, 1987, pp. 779–786.
[10] A. Gavrikov, A.A. Efimenko, S.B. Dorofeev, A model for detonation cell size prediction from chemical kinetics, Combust. Flame 120 (2000) 19–33.
[11] R. Akbar, M. Kaneshige, E. Schultz, J. Shepherd, Detonations in H_2-N_2O-CH_4-NH_3-O_2-N_2 Mixtures, Explosion Dynamics Laboratory Rep. No. FM97–3, Graduate Aeronautical Laboratories, California Institute of Technology, Pasadena, CA, 2000.
[12] W. Hardle, E. Mammen, Comparing nonparametric versus parametric regression fits, Ann. Stat. 21 (1993) 1926–1947.
[13] J. Shepherd, AIAA Progress in Astronautics and Aeronautics, vol. 106, AIAA, New York, 1986, p. 263.
[14] M. Kaneshige, J.E. Shepherd, Detonation Database, Technical Rep. FM97–F98, GALCIT, 1997 [cited 2017 Jul 17]. Available from: http://www.galcit.caltech.edu/detn_db/html/.
[15] D. Goodwin, H.K. Moffat, R.L. Speth, Cantera: an Object-Oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes (Version 2.2.1), 2016 [cited 2017 Jul 17]. Available from: http://www.cantera.org.

[16] J.G. Carbonell, R.S. Michalski, T.M. Mitchell, An overview of machine learning, in: Machine Learning, Springer, Berlin, Germany, 1983, pp. 3–23.
[17] A. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.
[18] D. Basak, S. Pal, D.C. Patranabis, Support vector regression, Neural Inf. Process. Lett. Rev. 11 (2007) 203–224.
[19] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proc. International Joint Conference on Artificial Intelligence, vol. 14, Morgan Kaufmann, San Francisco, CA, 1995, pp. 1137–1145.
[20] F. Pedregosa, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.