aerospace

MDPI

Article

Prediction of Hydrogen–Oxygen Combustion Ignition Delay Time Based on Gaussian Process Regression

Qingmiao Ma, Jiaming Fu, Qizheng Zhou *, Yang Zhao and Weige Liang *

Naval University of Engineering, Wuhan 430033, China
* Correspondence: 1512021014@nue.edu.cn (Q.Z.); 1312021010@nue.edu.cn (W.L.)

Abstract

To address the insufficient prediction accuracy of traditional Arrhenius-type empirical correlations for hydrogen–oxygen combustion ignition delay time (IDT) over wide operating conditions, this paper proposes a data-driven prediction method based on Gaussian Process Regression (GPR). Based on 413 sets of shock tube and rapid compression machine experimental data, with temperature, pressure, and equivalence ratio as input features, an IDT prediction model was constructed. The GPR model adopts the Matérn 3/2 kernel function combined with the Automatic Relevance Determination (ARD) strategy for hyperparameter optimization. As a benchmark comparison, LASSO regression was employed for feature selection and parameter estimation of the Arrhenius empirical correlation. On a stratified sampling test set, the GPR model achieved a coefficient of determination R^2 of 0.8587 (R^2_{ln} = 0.9271 on logarithmic scale), significantly outperforming the Arrhenius model's R^2 of 0.5422 (R^2_{ln} = 0.7027). In 50 random-split stability tests, GPR yielded a mean R^2 of 0.9180 with a standard deviation of only 0.0333, demonstrating excellent robustness. This study provides a valuable reference for ignition performance prediction of hydrogen–oxygen combustion systems.

Keywords: hydrogen–oxygen combustion; ignition delay time; gaussian process regression; arrhenius correlation

1. Introduction

Hydrogen–oxygen combustion, as a key technology for clean energy utilization, holds significant application value in hydrogen-fueled internal combustion engines, gas turbines, rocket propulsion, and scramjets [1–4]. Ignition Delay Time (IDT) refers to the time interval from when a combustible mixture reaches specified initial conditions (temperature, pressure, and equivalence ratio) until auto-ignition occurs. It is a core kinetic parameter characterizing fuel reactivity and directly affects ignition reliability, combustion stability, and pollutant emission control of combustion systems. Accurate prediction of IDT for hydrogen–oxygen mixtures is of great significance for combustor design, fuel formulation optimization, and combustion operation over wide operating ranges [5,6].

Traditionally, researchers have relied primarily on two approaches to obtain ignition delay times: direct measurement through experimental apparatus such as shock tubes and rapid compression machines, and zero-dimensional or three-dimensional combustion simulation using detailed chemical reaction mechanisms [7,8]. Although shock tube experiments can provide high-precision IDT data, they are costly, time-consuming, and have difficulty covering the full range of conditions spanning wide temperature, pressure, and equivalence ratio regimes. On the other hand, detailed chemical mechanisms possess a solid physical

Check for updates

Academic Editor: James 'Chris' Thomas

Received: 2 June 2026
Revised: 28 June 2026
Accepted: 22 July 2026
Published: 23 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

foundation but involve complex solution procedures with high computational resource consumption and long time costs [9]. Therefore, researchers have devoted efforts to simplified IDT prediction methods. Among these, the Arrhenius-type empirical correlation is the most classical, whose parameters carry clear physical meaning and can describe the influence laws of temperature, pressure, and equivalence ratio on ignition delay within certain operating ranges [10]. However, hydrogen–oxygen combustion reaction pathways vary dramatically with operating conditions, especially over wide equivalence ratio and pressure ranges, where the dependence of IDT on temperature and pressure exhibits strong nonlinearity and multi-scale features. The traditional linearized form of the Arrhenius correlation cannot fully capture such complex dependencies, resulting in limited prediction accuracy under wide-ranging conditions.

In recent years, machine learning methods have demonstrated great potential in the field of combustion prediction, providing new technical pathways to address the aforementioned problems [11,12]. In the area of IDT prediction, studies employing artificial neural networks (ANN), support vector machines (SVM), random forests, and other methods have achieved promising results. For instance, Cui et al. [13,14] predicted the ignition delay times of n-butane/hydrogen mixtures and multi-component surrogate fuels using back-propagation (BP) neural networks, respectively; Huang et al. [15] used ANN models to predict the IDT of Jet A-1/hydrogen blended fuels and proposed nested sub-networks to improve local prediction accuracy; and Tuan et al. [16] compared ANN and SVM in diesel/biodiesel engine IDT prediction, finding that SVM outperformed ANN under certain operating conditions. Nonlinear regression models based on machine learning offer significant advantages in capturing the complexity of combustion chemistry.

Among these machine learning methods, Gaussian Process Regression (GPR) [17], as a non-parametric Bayesian probabilistic model, possesses two unique advantages beyond its ability to automatically learn complex nonlinear relationships: first, it can theoretically provide uncertainty quantification of predictions, which is particularly important for combustion system risk assessment and experimental design; second, it is relatively insensitive to hyperparameters and maintains good generalization capability even with small sample sizes [18]. GPR has been successfully applied in computational fluid dynamics and flame speed prediction [19,20]; however, research on GPR for hydrogen–oxygen combustion IDT prediction remains insufficient.

Therefore, this paper systematically investigates a GPR-based prediction method for hydrogen–oxygen combustion IDT. The main contributions include: (1) proposing a Matérn 3/2 kernel function combined with an Automatic Relevance Determination (ARD) hyperparameter optimization strategy suitable for combustion kinetics data; (2) establishing a complete GPR prediction framework and conducting systematic comparisons with the Arrhenius-type empirical correlation model in terms of prediction accuracy and model stability; and (3) evaluating the differences between the two models in prediction reliability and sensitivity to data partitioning through residual analysis and repeated random-split experiments. The research findings can provide reference for ignition performance prediction and evaluation of hydrogen–oxygen combustion systems.

2. Arrhenius Empirical Correlation Model and GPR Model Construction

2.1. Arrhenius-Type Empirical Correlation Model

2.1.1. Model Principle

The Arrhenius empirical correlation is a classical model describing the relationship between combustion reaction rate and temperature. For hydrogen–oxygen combustion IDT, its form is expressed as:

$$\tau = AP^n\phi^m \exp\left(\frac{E}{RT}\right) \tag{1}$$

where A is the pre-exponential factor, n is the pressure exponent, m is the equivalence ratio exponent, E is the activation energy, and R is the universal gas constant. This formula quantitatively describes the governing effects of pressure P, equivalence ratio ϕ, and temperature T on ignition delay time τ from a physical mechanism perspective, and each parameter carries clear physical meaning.

Taking the natural logarithm of both sides linearizes the nonlinear model to facilitate parameter estimation:

$$\ln \tau = \ln A + n\ln P + m\ln \phi + \frac{E}{R} \cdot \frac{1}{T} \tag{2}$$

2.1.2. Feature Construction

To transform Equation (2) into a solvable linear regression problem, the following features are constructed: lnP, $ln\phi$, lnT, and $1/T$, along with all pairwise interaction terms (six terms in total) to capture coupling effects between variables. The output variable is taken as $ln\tau$. This yields the extended Arrhenius linear model:

$$\ln \tau = \beta_0 + \beta_1 \ln P + \beta_2 \ln \phi + \beta_3 \ln T + \beta_4 \frac{1}{T} + \sum_{i<j} \beta_{ij} x_i x_j \tag{3}$$

where β_0 is the intercept term, β_i are the coefficients of main-effect features, and β_{ij} are the coefficients of interaction terms.

2.1.3. LASSO Parameter Estimation

The model includes multiple main-effect features and their pairwise interaction terms (10 candidate variables in total). Direct use of Ordinary Least Squares (OLS) carries a risk of overfitting. On one hand, multicollinearity may exist among features (e.g., ln T and $1/T$ are highly correlated), leading to inflated OLS coefficient variance and poor stability; on the other hand, not all interaction terms possess significant physical meaning, and retaining all variables degrades generalization capability. Therefore, LASSO (Least Absolute Shrinkage and Selection Operator) [21] is adopted to select the optimal combination of interaction terms.

LASSO regression is a linear regression method based on L_1-norm regularization. Its core idea is to impose a penalty on the sum of absolute values of regression coefficients while minimizing the sum of squared residuals. The optimization objective function is:

$$\beta = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \sum_{k=1}^{K} \beta_k x_{ik} \right)^2 + \lambda \sum_{k=1}^{K} |\beta_k| \right\} \tag{4}$$

where N is the number of training samples, K is the number of features, y_i is the target variable ($ln\tau$) for the i-th sample, x_{ik} is the k-th feature of the i-th sample, and λ is the penalty parameter whose value determines the regularization intensity. The L_1 penalty of LASSO regression has the unique property of precisely shrinking some coefficients to zero, thereby achieving automatic feature selection: when λ is sufficiently large, coefficients of unimpor-

tant features are directly set to zero, retaining only variables that contribute significantly to prediction. This property is particularly important for the present study, as it enables automatic screening of key factors affecting ignition delay time from four main effects and six interaction terms, simplifying model structure while enhancing physical interpretability.

The regularization parameter λ directly determines model complexity: $\lambda = 0$ degenerates to OLS, where the model is prone to overfitting without sparsity; as λ increases, more and more coefficients are compressed toward zero, making the model more concise but potentially prone to underfitting. To determine the optimal λ, 10-fold cross-validation combined with the 1-SE rule is employed: candidate λ values are traversed on a logarithmic scale ($\lambda \in [10^{-4}, 10^{1}]$), the cross-validation mean squared error (CV-MSE) for each λ is computed via 10-fold CV, and then the minimum λ within one standard error of the minimum CV-MSE point is selected as the optimal value, maximizing feature sparsity while guaranteeing model fitting accuracy. Through this method, $\lambda = 0.029$ was determined, and five non-zero coefficient variables were selected from the 10 candidates: main effects ln P, ln ϕ, $1/T$, and interaction terms ln P·ln T, ln ϕ·$(1/T)$. This indicates that the coupling effects of pressure with temperature and of equivalence ratio with temperature have significant influence on ignition delay time. Variables compressed to zero can be linearly represented by other features or have weak independent contributions to IDT. Substituting the selected non-zero coefficients into the linear model yields the specific expression of the Arrhenius-type empirical correlation for subsequent IDT prediction.

2.2. GPR Model

2.2.1. Model Principle

GPR is a non-parametric Bayesian probabilistic regression model based on Bayesian inference. Its core idea is to treat the regression function as a random process following a Gaussian distribution, learn the posterior distribution of the function through training data, and thereby realize prediction for unknown samples while providing the probability distribution and uncertainty quantification of prediction results.

For the input feature vector $\mathbf{x} \in \mathbb{R}^d$ ($d = 3$, corresponding to T, P, ϕ) and output variable y (corresponding to $ln\tau$), it is assumed that y follows a Gaussian process:

$$y \sim \mathcal{GP}\left(m(\mathbf{x}), k\left(\mathbf{x}, \mathbf{x}'\right)\right) \tag{5}$$

where $m(\mathbf{x})$ is the mean function (taken as zero to simplify model complexity), and $k(\mathbf{x}, \mathbf{x}')$ is the kernel function that characterizes similarity between input features and determines smoothness and flexibility of the regression function.

Given the training dataset, the predictive distribution for a test sample $\mathbf{x}_*$ follows a Gaussian distribution:

$$y_* \mid \mathcal{D}, \mathbf{x}_* \sim \mathcal{N}\left(\mu_*, \sigma_*^2\right) \tag{6}$$

where the predicted mean μ_* is the model's predicted value, and the predicted variance σ_*^2 represents prediction uncertainty, which can effectively quantify the reliability of prediction results.

2.2.2. Feature Construction and Standardization

Since GPR relies on kernel functions to automatically learn nonlinear mappings, the original physical variables T, P, and ϕ are directly adopted as inputs, with the output taken as $ln\tau$. Due to significant differences in dimensions and numerical ranges among the three features (temperature ~10^3 K, pressure ~10^0–10^1 atm, and equivalence ratio

~10^{-1}–10^{0}), inputs are standardized before training to ensure equal attention from the kernel function to each feature:

$$x'_j = \frac{x_j - \mu_j}{\sigma_j} \tag{7}$$

where μ_j and σ_j are the mean and standard deviation of the i-th feature on the training set, respectively.

2.2.3. Kernel Function and Hyperparameter Optimization

The kernel function is the core of the GPR model, and its selection directly determines the model's ability to fit complex nonlinear relationships. The Matérn 3/2 kernel function is selected. Compared with the commonly used Radial Basis Function (RBF) kernel, the Matérn kernel provides better balance between smoothness and flexibility, making it more suitable for fitting combustion kinetics data with moderate smoothness [22]. Its expression is:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \frac{\sqrt{3}}{l}\|\mathbf{x} - \mathbf{x}'\|\right) \exp\left(-\frac{\sqrt{3}}{l}\|\mathbf{x} - \mathbf{x}'\|\right) \tag{8}$$

where σ_f^2 is the signal variance, l is the length scale, and $\|\mathbf{x} - \mathbf{x}'\|$ is the Euclidean distance between input features. The Matérn 3/2 kernel combines smoothness with flexibility and can effectively fit the nonlinear variation law of hydrogen–oxygen combustion IDT while maintaining moderate computational complexity suitable for the sample size of this study.

To improve fitting accuracy, the Automatic Relevance Determination (ARD) strategy is employed for kernel hyperparameter optimization, allowing each input feature to correspond to an independent length scale l_d (d = 1, 2, 3), corresponding to temperature, pressure, and equivalence ratio, respectively. By learning different length scales for each feature, the relative importance of each feature on IDT can be automatically quantified. The goal of hyperparameter optimization is to maximize the log-likelihood function, solving for the optimal hyperparameter set $\theta^* = \sigma_f^2, l_1, l_2, l_3$ via Maximum Likelihood Estimation (MLE), with the optimization process implemented using gradient descent to ensure finding the global optimum. The optimized kernel hyperparameters are: signal variance σ_f^2 = 1.31 and length scales l_1 = 0.2165, l_2 = 2.9163, and l_3 = 2.2583. Among the three length scales, l_1 is the smallest, indicating that temperature has the most sensitive effect on ignition delay time; l_2 and l_3 are similar and relatively large, indicating that the characteristic scales of pressure and equivalence ratio are gentler, with weaker influence on model output compared to temperature.

2.3. Model Comparison and Performance Evaluation

To evaluate the comprehensive performance of the two prediction models—namely the Arrhenius empirical correlation model and the GPR model—a systematic comparison is conducted from two dimensions: prediction accuracy of ignition delay time and model stability. Single-training comparisons assess prediction accuracy, while performance metrics from multiple random splits compare model stability.

To measure model prediction performance, Root Mean Square Error (RMSE) and coefficient of determination (R^2) are introduced as primary evaluation metrics. Since ignition delay time spans multiple orders of magnitude and the model output during training is log-transformed, RMSE and R^2 on the logarithmic scale are also computed. The expressions of the evaluation metrics are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\tau_i - \hat{\tau}_i)^2} \tag{9}$$

$$R^2 = \frac{\sum_{i=1}^{N}\left(\hat{\tau}_i - \frac{1}{N}\sum_{i=1}^{N}\tau_i\right)^2}{\sum_{i=1}^{N}\left(\tau_i - \frac{1}{N}\sum_{i=1}^{N}\tau_i\right)^2} \tag{10}$$

$$\mathrm{RMSE}_{\ln} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\ln\tau_i - \ln\hat{\tau}_i)^2} \tag{11}$$

$$R^2_{\ln} = 1 - \frac{\sum_{i=1}^{N}(\ln\tau_i - \ln\hat{\tau}_i)^2}{\sum_{i=1}^{N}\left(\ln\tau_i - \overline{\ln\tau_i}\right)^2} \tag{12}$$

where τ_i is the actual ignition delay time of the i-th sample in the real experimental dataset, $\hat{\tau}_i$ is the predicted ignition delay time by the model, and $\hat{\tau}_i$ is the mean of all actual values. RMSE represents the difference between predicted and actual experimental data; the closer RMSE approaches to zero, the higher the model prediction accuracy. R^2 measures the degree to which the model explains the variance of the dependent variable and can be used to compare fitting quality of different models; the closer R^2 is to 1, the better the model fitting effect.

3. Data Source and Preprocessing

3.1. Experimental Dataset Description

All experimental data used in this study are sourced from public data repositories and their associated original references [23–31], which collected and compiled 413 sets of hydrogen–oxygen combustion experimental data. Data screening criteria were: clearly documented experimental conditions including four complete variables (temperature, pressure, equivalence ratio, and ignition delay time) and acquisition methods limited to shock tube or rapid compression device experiments to ensure data reliability. Each sample in the dataset records the initial temperature T (K), initial pressure P (atm), equivalence ratio ϕ, and corresponding ignition delay time τ (μs) of a hydrogen–oxygen combustion experiment. Statistical characteristics of each variable are shown in Table 1. As seen from Table 1, the data cover a broad range from low temperature (858 K) to high temperature (2136 K), from low pressure (0.74 atm) to high pressure (34.68 atm), and from fuel-lean (ϕ = 0.10) to fuel-rich (ϕ = 4.00) conditions. The ignition delay time values span three orders of magnitude (10 μs to 13,090 μs), reflecting significant differences in combustion reaction kinetics under different operating conditions. Data cleaning confirmed no outliers, satisfying basic requirements for model construction.

Table 1. Statistical characteristics of H_2/O_2 combustion experimental data.

Variable	Min	Max	Mean	Std. Dev.
Temperature T (K)	858.0	2136.0	1136.2	173.4
Pressure P (atm)	0.74	34.68	6.73	8.23
Equivalence ratio ϕ	0.10	4.00	0.76	0.55
Ignition delay τ (μs)	10.0	13,090.0	1335.2	1946.6

3.2. Target Variable Logarithmic Transformation

The numerical range of ignition delay time τ spans three orders of magnitude. Using τ directly as model output would cause imbalanced gradient updates during training, making it difficult for the model to converge stably. Additionally, larger numerical values would dominate the loss function, severely masking prediction errors of small-magnitude samples and ultimately reducing overall prediction accuracy. Therefore, a logarithmic transformation is applied to τ to compress the numerical range, improve distribution symmetry, and reduce the impact of extreme values on the model.

3.3. Dataset Partitioning

To ensure objectivity and reproducibility of model training and evaluation, reasonable dataset partitioning is required. This study employs a stratified sampling strategy using equivalence ratio ϕ as the stratification basis. Since ϕ values in the dataset take discrete finite values (e.g., 0.1, 0.5, 1.0, 2.0, etc.) while temperature and pressure distributions are relatively continuous within their respective ranges, fully random partitioning could result in all samples of a particular ϕ value falling entirely into the training or test set, causing loss of reference for model evaluation under that specific equivalence ratio condition. Stratifying by ϕ ensures that samples within each equivalence ratio interval enter the training and test sets proportionally, avoiding evaluation bias due to missing specific operating conditions. The dataset is partitioned in an 8:2 ratio into training and test sets, with the training set containing 331 samples for model parameter estimation and training optimization, and the test set containing 82 samples for independent evaluation of generalization performance. To ensure experimental repeatability, the random seed is fixed at 75 during partitioning, ensuring identical sample subsets across runs. The sample distribution across intervals is shown in Table 2.

Table 2. Distribution of equivalence ratio intervals.

Equivalence Ratio Interval	Original	Training Set	Test Set
$\phi \leq 0.5$	101	81	20
$0.5 < \phi \leq 1.0$	95	76	19
$1.0 < \phi \leq 2.0$	160	128	32
$\phi > 2.0$	57	46	11
Total	413	331	82

It should be noted that the experimental data used in this study originate from multiple independent research facilities. Although the data sources are reliable, the influence of different facilities on the measured ignition delay time has not been taken into account, nor can the quantitative effects of facility variability be provided as additional input features.

4. Results and Discussion

4.1. Performance Comparison Under Single Partition

Using the above stratified sampling partition (80% training, 20% testing), the Arrhenius empirical correlation model and GPR model were trained separately, and various metrics were computed on the test set. The performance of both models on the test set is shown in Table 3.

Table 3. Performance comparison of the two models on the test set.

Model	RMSE	R^2	$RMSE_{ln}$	R^2_{ln}
Arrhenius	1226.22	0.5422	0.8441	0.7027
GPR	781.00	0.8587	0.3973	0.9271

As shown in Table 3, the RMSE values of the GPR model on both original and logarithmic scales are lower than those of the Arrhenius empirical correlation model, with R^2 improved to 0.8587 (original scale) and 0.9271 (logarithmic scale), indicating that GPR achieves higher prediction accuracy. Figure 1 presents scatter plots of actual versus predicted values for both models on the test and training sets, where training set samples are denoted by blue circles, test set samples by red circles, and the black dashed line represents the ideal diagonal ($y = x$). It can be visually observed that the scatter points of the GPR model cluster more closely around the diagonal line, whereas the empirical correlation model exhibits considerable deviation. Moreover, since the figure displays deviations on a logarithmic scale, some true values differ from predicted values by several to tens of times. This verifies that GPR, through automatically learning nonlinear relationships via kernel functions, can better capture the complex mapping between ignition delay time and temperature, pressure, and equivalence ratio. In contrast, although the Arrhenius model enhances nonlinear expressiveness by introducing interaction terms, it remains essentially a linear model in logarithmic space and cannot fully capture the complex nonlinear characteristics of IDT variations with temperature, pressure, and equivalence ratio.

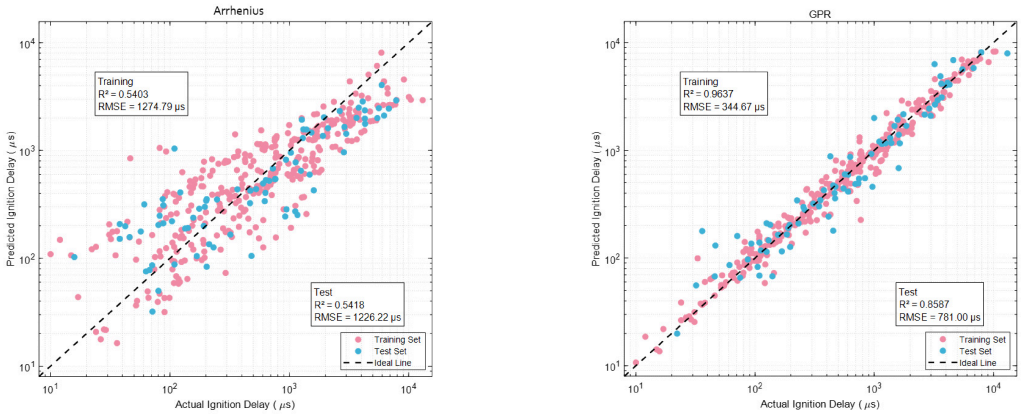


Figure 1. Scatter plots of predicted vs. actual values (blue dots: test set samples; red dots: training set samples).

Furthermore, the logarithmic-scale residual plots of both models are presented in Figure 2, with the horizontal axis representing predicted values and the vertical axis representing residuals. The residuals of the GPR model are more concentrated around the zero line with no obvious trendiness, while the residuals of the Arrhenius model appear somewhat dispersed in the low-prediction-value region, exhibiting slight systematic bias. This further confirms the advantages of GPR in uncertainty modeling and residual control.

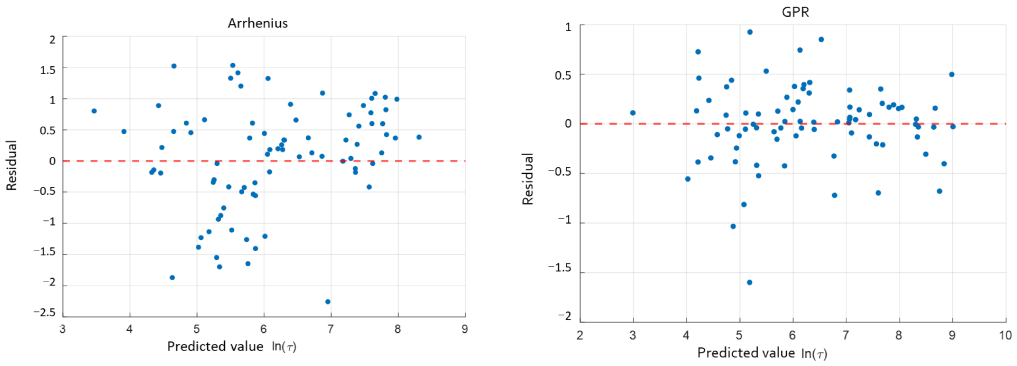


Figure 2. Logarithmic-scale residual distribution on the test set.

4.2. Model Stability Analysis

The single-partition performance evaluation described above may be influenced by the specific data partition. To comprehensively evaluate model robustness, sensitivity to data partitioning was further analyzed through 50 random stratified partitions (changing the random seed each time), recording R^2 and RMSE for each test set, and computing means and standard deviations. Figure 3 presents a density scatter plot with R^2 on the horizontal axis and RMSE on the vertical axis, where each point represents the test-set performance of one random partition, color depth indicates local point density, the red cross marks the mean point, and red dashed lines pass through the horizontal and vertical positions of the mean point. Statistical results are summarized in Table 4.

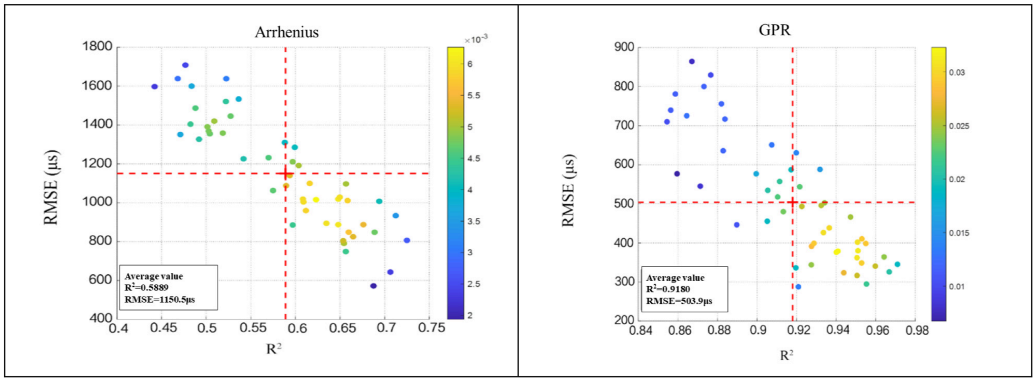


Figure 3. Scatter plot of R^2-RMSE density of the test set under multiple random divisions.

Table 4. Stability comparison of the two models (50 random partitions).

Model	R^2 Mean	R^2 Std.	RMSE Mean	RMSE Std.
Arrhenius	0.5889	0.0772	1150.52	287.56
GPR	0.9180	0.0333	503.95	157.91

As shown in Table 4, the standard deviation of R^2 for the GPR model (0.0333) is smaller than that of the Arrhenius empirical correlation model (0.0772), and the standard deviation of RMSE (157.91) is also significantly smaller than that of the Arrhenius model (287.56). In Figure 3, the points of the GPR model exhibit higher clustering density, indicating that GPR has lower sensitivity to data partitioning and higher stability, with prediction results less affected by minor changes in the training set. In contrast, due to the limited number of selectable parameters, the Arrhenius model may select different combinations of interaction terms under different partitions as the training set data varies slightly, leading to greater performance fluctuations.

In summary, the GPR model outperforms the Arrhenius empirical correlation model in both prediction accuracy and model stability.

5. Conclusions

This study develops and validates a GPR-based ignition delay time prediction model using the Matérn 3/2 kernel function with Automatic Relevance Determination (ARD) for hyperparameter optimization. Compared with the classical Arrhenius model, the GPR model significantly outperforms the Arrhenius model on the test set, and demonstrates superior stability over 50 random data splits.

Nevertheless, the following limitations should be acknowledged: (1) the current dataset is predominantly concentrated in the low-to-medium pressure range, with insufficient high-pressure data, which restricts the model's applicability across a wide pressure range. Additionally, fuel-rich conditions are sparsely sampled, and practical engineering factors such as impurity gases and turbulence intensity have not been considered. (2) The model makes predictions solely based on initial conditions and does not account for differences among experimental facilities. Consequently, its predictions are applicable to IDT evaluation under standardized constant initial conditions. (3) Systematic comparisons with detailed chemical kinetic mechanisms and other machine learning methods have not yet been conducted, and the model's generalization capability across broader regimes requires further validation. Future work may address these directions accordingly.

Author Contributions: Conceptualization, J.F., Q.M., Q.Z., Y.Z. and W.L.; methodology, J.F., Q.M., Q.Z., Y.Z. and W.L.; software, J.F.; validation, J.F. and Q.M.; formal analysis, J.F.; investigation, J.F. and Q.M.; resources, Q.Z., Y.Z. and W.L.; data curation, J.F.; writing—original draft preparation, J.F. and Q.M.; writing—review and editing, Q.Z., Y.Z. and W.L.; visualization, J.F.; supervision, Q.Z., Y.Z. and W.L.; project administration, W.L.; funding acquisition, Q.Z., Y.Z. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Alam, N.; Sharma, K.K.; Pandey, K.M. Combustion characteristics of hydrogen-air mixture in pulse detonation engines. *J. Mech. Sci. Technol.* **2019**, *33*, 2451–2457. [CrossRef]
2. Fu, H.; Luo, Q.; Bao, L.; Sun, B.; Niu, Q.; Zhang, S.; Zhang, H. Combustion characteristics and energy distribution of hydrogen engine under high oxygen concentration. *Int. J. Hydrogen Energy* **2022**, *47*, 38031–38042. [CrossRef]
3. Habib, M.A.; Abdulrahman, G.A.; Alquaity, A.B.; Qasem, N.A. Hydrogen combustion, production, and applications: A review. *Alex. Eng. J.* **2024**, *100*, 182–207. [CrossRef]
4. Gelfand, B.E.; Silnikov, M.V.; Medvedev, S.P.; Khomik, S.V. *Thermo-Gas Dynamics of Hydrogen Combustion and Explosion*; Springer: Berlin/Heidelberg, Germany, 2012.
5. Petrukhin, N.V.; Grishin, N.N.; Sergeev, S.M. Ignition delay time—An important fuel property. *Chem. Technol. Fuels Oils* **2016**, *51*, 581–584. [CrossRef]
6. Erman, A.G.; Hellier, P.; Ladommatos, N. The impact of ignition delay and further fuel properties on combustion and emissions in a compression ignition engine. *Fuel* **2020**, *262*, 116155. [CrossRef]
7. AlAbbad, M.; Javed, T.; Khaled, F.; Badra, J.; Farooq, A. Ignition delay time measurements of primary reference fuel blends. *Combust. Flame* **2017**, *178*, 205–216. [CrossRef]
8. Alabbad, M.; Li, Y.; AlJohani, K.; Kenny, G.; Hakimov, K.; Al-Lehaibi, M.; Emwas, A.-H.; Meier, P.; Badra, J.; Curran, H.; et al. Ignition delay time measurements of diesel and gasoline blends. *Combust. Flame* **2020**, *222*, 460–475. [CrossRef]
9. Wang, B.; Wei, W.; Ma, S.; Wei, G. Construction of one-step H2/O2 reaction mechanism for predicting ignition and its application in simulation of supersonic combustion. *Int. J. Hydrogen Energy* **2016**, *41*, 19191–19206. [CrossRef]

10. Davidson, D.F.; Hanson, R.K. Interpreting shock tube ignition data. *Int. J. Chem. Kinet.* **2004**, *36*, 510–523. [CrossRef]
11. Gerard, B. Predicting the Complex Dependence of Ignition Delay Times on Pressure Using Machihne Learning. Honors Thesis, Syracuse University, Syracuse, NY, USA, 2024. Available online: https://surface.syr.edu/honors_capstone/1665 (accessed on 23 May 2026).
12. Molana, M.; Darougheh, S.; Biglar, A.; Chamkha, A.J.; Zoldak, P. Machine learning approaches for predicting ignition delay in combustion processes: A comprehensive review. *Ind. Eng. Chem. Res.* **2024**, *63*, 2509–2518. [CrossRef]
13. Cui, Y.; Wang, Q.; Liu, H.; Zheng, Z.; Wang, H.; Yue, Z.; Yao, M. Development of the ignition delay prediction model of n-butane/hydrogen mixtures based on artificial neural network. *Energy AI* **2020**, *2*, 100033. [CrossRef]
14. Cui, Y.; Liu, H.; Wang, Q.; Zheng, Z.; Wang, H.; Yue, Z.; Ming, Z.; Wen, M.; Feng, L.; Yao, M. Investigation on the ignition delay prediction model of multi-component surrogates based on back propagation (BP) neural network. *Combust. Flame* **2022**, *237*, 111852. [CrossRef]
15. Huang, Y.; Jiang, C.; Wan, K.; Gao, Z.; Vervisch, L.; Domingo, P.; He, Y.; Wang, Z.; Lee, C.-H.; Cai, Q.; et al. Prediction of ignition delay times of jet a-1/hydrogen fuel mixture using machine learning. *Aerosp. Sci. Technol.* **2022**, *127*, 107675. [CrossRef]
16. Van Tuan, N.; Minh, D.Q.; Khoa, N.X.; Lim, O. A study to predict ignition delay of an engine using diesel and biodiesel fuel based on the ANN and SVM machine learning methods. *ACS Omega* **2023**, *8*, 9995–10005. [CrossRef]
17. Isabona, J.; Ojuh, D.O. Machine learning based on kernel function controlled gaussian process regression method for in-depth extrapolative analysis of COVID-19 daily cases drift rates. *Int. J. Math. Sci. Comput.* **2021**, *7*, 14–23. [CrossRef]
18. Polke, D.; Ahle, E.; Söffker, D. Adaptive learning with gaussian process regression: A comprehensive review of methods and applications. *Mach. Learn. Knowl. Extr.* **2026**, *8*, 101. [CrossRef]
19. Morita, Y.; Rezaeiravesh, S.; Tabatabaei, N.; Vinuesa, R.; Fukagata, K.; Schlatter, P. Applying bayesian optimization with gaussian process regression to computational fluid dynamics problems. *J. Comput. Phys.* **2022**, *449*, 110788. [CrossRef]
20. Üstün, C.E.; Herfatmanesh, M.R.; Valera-Medina, A.; Paykani, A. Applying machine learning techniques to predict laminar burning velocity for ammonia/hydrogen/air mixtures. *Energy AI* **2023**, *13*, 100270. [CrossRef]
21. O'Brien, C.M. Statistical Learning with Sparsity: The Lasso and Generalizations. *Int. Stat. Rev.* **2016**, *84*, 156–157. [CrossRef]
22. Wang, J. An intuitive tutorial to gaussian process regression. *Comput. Sci. Eng.* **2023**, *25*, 4–11. [CrossRef]
23. Ren, H.; Wang, J.; Li, X. Combustion Dynamics, Sichuan, CDS, Center for Combustion Dynamics, Sichuan University. 2021. Available online: http://cds.scu.edu.cn/ (accessed on 1 April 2026).
24. Heng, Y.; Li, G.; Liang, J.; Zhang, Z. Ignition characteristics of ethanol–hydrogen mixtures with different hydrogen contents at elevated initial temperatures. *Fuel* **2020**, *281*, 118742. [CrossRef]
25. Shao, J.; Choudhary, R.; Davidson, D.F.; Hanson, R.K.; Barak, S.; Vasu, S. Ignition delay times of methane and hydrogen highly diluted in carbon dioxide at high pressures up to 300 atm. *Proc. Combust. Inst.* **2019**, *37*, 4555–4562. [CrossRef]
26. Ninnemann, E.; Koroglu, B.; Pryor, O.; Barak, S.; Nash, L.; Loparo, Z.; Sosa, J.; Ahmed, K.; Vasu, S. New insights into the shock tube ignition of H2/O2 at low to moderate temperatures using high-speed end-wall imaging. *Combust. Flame* **2018**, *187*, 11–21. [CrossRef]
27. Pan, P.; Zhang, Y.; Zhang, J.; Tian, Z.; Huang, Z. Shock tube and kinetic study of C2H6/H2/O2/Ar mixtures at elevated pressures. *Int. J. Hydrogen Energy* **2014**, *39*, 6024–6033. [CrossRef]
28. Tang, C.; Man, X.; Wei, L.; Pan, L.; Huang, Z. Further study on the ignition delay times of propane–hydrogen–oxygen–argon mixtures: Effect of equivalence ratio. *Combust. Flame* **2013**, *160*, 2283–2290. [CrossRef]
29. Kéromnès, A.; Metcalfe, W.K.; Heufer, K.A.; Donohoe, N.; Das, A.K.; Sung, C.-J.; Herzler, J.; Naumann, C.; Griebel, P.; Mathieu, O.; et al. An experimental and detailed chemical kinetic modeling study of hydrogen and syngas mixture oxidation at elevated pressures. *Combust. Flame* **2013**, *160*, 995–1011. [CrossRef]
30. Pang, G.A.; Davidson, D.F.; Hanson, R.K. Experimental study and modeling of shock tube ignition delay times for hydrogen–oxygen–argon mixtures at low temperatures. *Proc. Combust. Inst.* **2009**, *32*, 181–188. [CrossRef]
31. Herzler, J.; Naumann, C. Shock-tube study of the ignition of methane/ethane/hydrogen mixtures with hydrogen contents from 0% to 100% at different pressures. *Proc. Combust. Inst.* **2009**, *32*, 213–220. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.