International Journal of Hydrogen Energy 81 (2024) 47–55

Contents lists available at ScienceDirect

International Journal of Hydrogen Energy

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

# Probabilistic machine learning framework for chemical source term integration with Gaussian Processes: $H_2$/air auto-ignition case

Cihat Emre Üstün *, Amin Paykani

*School of Engineering and Materials Science, Queen Mary University of London, London, E1 4NS, United Kingdom*

ARTICLE INFO

*Keywords:*
Probabilistic machine learning
Gaussian Processes
Artificial neural networks
Hydrogen auto-ignition
Uncertainty quantification (UQ)

ABSTRACT

The integration of chemistry poses a major bottleneck in numerical combustion modelling, as a significant amount of simulation time is consumed in the direct integration (DI) of differential equations into thermo-chemistry modelling. In this work, a probabilistic machine learning (ML) framework, using Gaussian processes (GPs), is developed as a chemical source term integrator to replace DI for the prediction of a $H_2$/air auto-ignition case. In this context, two algorithms, namely Gaussian Process Regression (GPR) and Gaussian Process Autoregressive Regression (GPAR), are investigated. The training dataset is generated using zero-dimensional (0D) isobaric $H_2$/air auto-ignition simulations in Cantera. To address the scalability issue of the GPs, the variational inducing points method is used. This method leverages a subset of the original data for training, allowing sparse approximations. The performances of the GPR and GPAR are compared to a standard artificial neural network (ANN) model. A priori comparison with direct integration shows that both GPR ($R^2_{test} = 0.997$) and GPAR ($R^2_{test} = 0.998$) outperform the ANNs ($R^2_{test} = 0.988$) by capturing latent dynamics of the chemistry when working with small datasets. Additionally, GP models offer the capability to quantify the uncertainty of each prediction, providing deeper insights into the model's limitations. It is also shown that the inference with GP-based models is slower than ANNs with speed-up factors of 1.9-2.1 relative to the 0D reactor model, whereas the ANN speed-up factor goes up to 3.0.

## 1. Introduction

Decarbonising the power generation and long-range transportation sectors is a crucial step in addressing climate change. These hard-to-abate sectors still rely on combustion for their high-temperature processes, with fossil fuels being a cost-effective option. Zero-carbon fuels like hydrogen ($H_2$) have the potential to play a strategic role in long-term decarbonisation and reduction of greenhouse gas (GHG) emissions from these sectors. Hydrogen has a high gravimetric energy density and produces zero-carbon emissions, making it an attractive alternative fuel [1]. Its high burning velocity and wide flammability limits [2] make it ideal for various applications, including rocket propulsion, power generation, and transportation [3,4]. Hydrogen's high diffusivity allows for more uniform mixing with air, enhancing combustion efficiency and stability. Additionally, its low ignition energy enables quick and reliable ignition even under extremely lean conditions [5]. Recent studies have shown promising advancement in achieving ultra-low NOx emissions by ultra-lean combustion of $H_2$ [6]. Ultra-lean combustion, where the air-fuel mixture contains excess air, not only reduces NOx formation but also improves thermal efficiency [7]. This is particularly important for meeting stringent environmental regulations and mitigating the impact of air pollution on public health.

To better understand hydrogen combustion and achieve clean combustion, advanced computational techniques are essential. Combustion modelling with computational fluid dynamics (CFD) simulations has been vital for developing less-polluting and highly-efficient engine technologies. Simulating reacting flows in such devices with strong turbulence and fully-resolved scales and species requires quite demanding computational resources. Additionally, to carry out reliable simulations, comprehensive multi-species chemical kinetic mechanisms should be included, where the majority of the simulation time is in the direct integration (DI) of these equations into the thermo-chemistry modelling approach, making it a computationally intensive simulation [8]. One standard alternative solution is to use an operator-splitting method that solves the chemical source terms separately as an ordinary differential equation (ODE) system [9,10]. However, direct integration (DI) of chemical source term remains the main challenge of combustion simulations as it can take up to 95% of the total run time for the numerical integration of the ODE systems corresponding to the chemical step [11,12].

* Corresponding author.
*E-mail address:* c.e.ustun@qmul.ac.uk (C.E. Üstün).

https://doi.org/10.1016/j.ijhydene.2024.07.220
Received 3 April 2024; Received in revised form 10 July 2024; Accepted 14 July 2024
Available online 23 July 2024

**Nomenclature**

**Abbreviations**

| | |
|---|---|
| 0D | Zero-dimensional |
| ANN | Artificial neural network |
| CFD | Computational fluid dynamics |
| CNN | Convolutional neural network |
| DI | Direct Integration |
| DNN | Deep neural network |
| GeLu | Gaussian error linear unit activation function |
| ReLu | Rectified linear unit activation function |
| GHG | Greenhouse gas |
| GP | Gaussian process |
| GPAR | Gaussian process autoregressive regression |
| GPR | Gaussian process regression |
| IP | Inducing point |
| LBV | Laminar burning velocity |
| LUT | Look up table |
| MAE | Mean absolute error |
| ML | Machine learning |
| ODE | Ordinary differential equation |
| SciML | Scientific machine learning |
| SOM | Self organising map |

**Symbols**

| | |
|---|---|
| $k$ | Covariance function |
| $N$ | Number of observations |
| $y_i$ | Set of outputs |
| $\alpha$ | Scale mixture |
| $\boldsymbol{\mu}$ | Mean vector |
| $\boldsymbol{\theta_\mu}$ | Mean vector hyperparameter |
| $\boldsymbol{\theta_k}$ | Covariance matrix hyperparameter |
| $\boldsymbol{I}$ | Identity matrix |
| $\boldsymbol{K}$ | Covariance matrix |
| $\boldsymbol{X}$ | Input matrix |
| $\boldsymbol{x}_i$ | Set of inputs |
| $\boldsymbol{Y}$ | Output matrix |
| $\epsilon_i$ | Gaussian noise term |
| $l$ | Length scale |
| $\mathcal{D}$ | Training dataset |
| $\mathcal{N}$ | Normal distribution |
| $\mathcal{O}$ | Complexity |
| $\mu$ | Mean function |
| $\phi$ | Equivalence ratio |
| $\sigma$ | Standard deviation |
| $\sigma^2$ | Variance |
| $\sigma_f$ | Signal variance |
| $\sigma_n$ | Noise variance |
| $P_i$ | Initial pressure |
| $R^2$ | Coefficient of determination |
| $T$ | Temperature |
| $T_i$ | Initial temperature |
| $Y$ | Mass fraction |

In recent years, scientific machine learning (SciML) applications have gained popularity due to a proven ability to speed up scientific computing tasks [13]. Machine learning (ML) offers significant time-savings, particularly in problems that encompass stiff and highly non-linear characteristics. ML in combustion has found applications in various topics ranging from chemistry integration [14], flame speed prediction [15,16], ignition delay time prediction [17], explosion limits [18] and more [19]. Early applications of ML for combustion chemistry integration focused on using artificial neural networks (ANNs). Christo et al. [20] showed one of the earliest applications of ANNs to predict chemical source term, applying them to describe turbulent diffusion flames of $H_2/CO_2$ mixtures. The implementation and the comparisons with DI and look-up table (LUT) method were presented. The advantages of ANN over DI and LUT were discussed in terms of computational speed-up and memory savings. Later, Blasco et al. [21] carried out a similar study focused on $CH_4$ combustion using a reduced mechanism with 13 species and 18 reactions. In their study, two distinct ANNs were employed to predict species mass fractions and temperature and density, respectively. The study revealed that ANN exhibited orders of magnitude faster computational speeds compared to DI with much less memory requirement compared to the tabulation approach. An et al. [22] studied the application of ANNs to model $H_2$/CO/kerosene fuelled supersonic engines where conventional combustion modelling is computationally expensive. The dataset was first clustered using self-organising maps (SOM) and particle swarm optimisation was used to optimise the training process. This approach offered up to 19 times faster simulations compared to a standard ODE solver using a 41-species mechanism. Another approach, also developed by An et al. [23], utilised deep convolutional neural networks (CNNs) for combustion simulation. This method reconstructs the turbulent flame rather than estimating the chemistry source term. Although successful reconstruction of the flame was obtained, further research is needed to test the applicability of the method to other complex combustion systems.

More recently, Sharma et al. [24] used deep feed forward ANNs to predict the source term for $H_2$/air combustion using a mechanism with 10 species and 34 reactions. The algorithm was initially applied to a single thermochemical condition and then expanded, resulting in a large dataset consisting of 7.2 million temporal vectors. Although an accurate representation of the temporal evolution of the source term was captured, the dataset size was prohibitive for a larger coverage of the thermochemical space. Later, Wan et al. [25] applied ANNs in the context of a non-adiabatic non-premixed combustion case using the GRI 3.0 mechanism. The ANNs were trained to predict the reaction rates of a reduced number of species compared to the original mechanism, achieving a 25 times speed-up compared to direct integration. Brown et al. [26] introduced novel deep neural networks (DNNs) to approximate stiff ODEs arising from combustion integration. The approach was based on parallel ResNets architecture using a $H_2/O_2$ combustion dataset. The model initially demonstrated poor generalisation when the equivalence ratio was varied, but it improved when the initial temperature was varied. This challenge was addressed by segregating the data into fuel-lean, stoichiometric, and fuel-rich regions. Zhang et al. [27] investigated the optimum sampling and transformation methods for DNN-based combustion integration. The method was applied to 0D, 1D, 2D, and 3D datasets. They used a large dataset with 5.3 million data points to cover a wide thermochemical space ($T_i$ = 800–3100 K, $P_i$ = 0.5–2.0 atm, no $\phi$ limits). They found that the multi-scale sampling method and box-cox transformation can effectively handle the multi-scale nature of the combustion data, making them well-suited for DNN training. Nikitin et al. [28] applied multilayer NNs to compute $H_2/O_2$ combustion. The number of layers and blocks were studied to efficiently model the evolving chemical system. The results showed that a relatively small network could accurately reconstruct the time–history of the species.

In recent years, uncertainty quantification (UQ) has gained significant attention in the field of combustion. This approach has been applied to various aspects such as kinetic modelling [29,30], laminar burning velocity [31], and fuel compositional uncertainty [32]. However, there is a noticeable gap in its application to chemical source term

integration. Deterministic machine learning models, including ANNs and DNNs, exhibit predictive errors that can accumulate during source term integration, leading to non-physical and inaccurate thermochemical evolution. Incorporating UQ in source term prediction is crucial to ensure that predictions are reliable and beneficial compared to the inherently deterministic nature of ANNs/DNNs.

One of the primary challenges in existing literature is the limited generalisability of models across various conditions, often constrained to a single initial state. Additionally, a very large dataset size is often required for acceptable prediction accuracy in most ML frameworks. To address these issues, we establish a Gaussian process (GP) based probabilistic machine learning framework (GP framework) and demonstrate its efficacy on chemical source term prediction of a $H_2$/air auto-ignition case. The novelty of the current work lies in the application of GPs as chemistry integrator to explore their data-efficient, uncertainty-quantified inference capabilities. Specifically, we explore and compare the performance of GP regression (GPR) and GP autoregressive regression (GPAR) algorithms, comparing them to ANNs in terms of speed and predictive accuracy. The methodology presented in this study provides an interesting approach that shows the potential for offering reduced computational costs of combustion simulations.

## 2. Methodology

### 2.1. Gaussian Process Regression (GPR)

Gaussian process regression provides a probabilistic, Bayesian-based framework for accomplishing single-output regression tasks with non-linear characteristics [33]. GPR possesses a modular structure, making it computationally manageable and partially interpretable. These qualities facilitate the creation of complex non-linear models through the design of covariance functions, i.e. kernels. These models can subsequently undergo rigorous evaluation through principled methods, such as the calculation of marginal likelihood, enabling a better understanding of their individual components.

Consider a process or a function, $f$, that takes a set of inputs $(\boldsymbol{x}_i)$ and generates set of outputs $(y_i)$ with $i = 1, \ldots, N$, with $N$ being number of observations. The function is regulated by a noise term, $\epsilon_i$, accounting for the uncertainty of the observations. The observations, $y_i$, can be expressed as follows:

$$y_i = f\left(\boldsymbol{x}_i\right) + \epsilon_i, \tag{1}$$

where $\epsilon_i$ is Gaussian distributed, $\mathcal{N}\left(0; \sigma_n^2\right)$, with a standard deviation of $\sigma_n$. The function, $f$, can be modelled as a GP.

Formally, a GP is characterised by a collection of random variables, where any finite subset of these variables follows a joint Gaussian distribution. It is defined by its mean and covariance functions, often referred to as kernels. The mean $(\mu(\boldsymbol{x}))$ and covariance functions $(k\left(\boldsymbol{x}, \boldsymbol{x}'\right))$ of a process $f(\boldsymbol{x})$ are as follows:

$$\begin{aligned} \mu(\boldsymbol{x}) &= \mathbb{E}[f(\boldsymbol{x})] \\ k\left(\boldsymbol{x}, \boldsymbol{x}'\right) &= \mathbb{E}\left[\left(f(\boldsymbol{x}) - \mu(\boldsymbol{x})\right)\left(f\left(\boldsymbol{x}'\right) - \mu\left(\boldsymbol{x}'\right)\right)\right] \end{aligned} \tag{2}$$

The GP, including the hyperparameters $\boldsymbol{\theta}_{\boldsymbol{\mu}}$ and $\boldsymbol{\theta}_{\boldsymbol{k}}$, can be written as,

$$f(\cdot) \sim \mathcal{GP}\left(\mu\left(\boldsymbol{x}; \boldsymbol{\theta}_{\mu}\right), k\left(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta}_k\right)\right) \tag{3}$$

The typical approach to learning the hyperparameters of the covariance function is to maximise the marginal likelihood, which represents the probability density of the observations based on these hyperparameters [34]. This marginal likelihood is determined by integrating over the possible values of the function $f$. If the training set, $D$, consists of input matrix $\boldsymbol{X}$ and output matrix, $\boldsymbol{Y}$, the natural logarithm of the marginal likelihood can be expressed as follows:

$$\begin{aligned} \log(p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta})) = &-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\mu})^{\top}\left(\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right)^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) \\ &-\frac{1}{2}\log\left(\left|\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right|\right) - \frac{n}{2}\log(2\pi) \end{aligned} \tag{4}$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{K}$ is the covariance matrix. Then, the GP approximation turns into a maximisation problem of Eq. (4) with respect to the mentioned hyperparameters. The posterior predictions are tractable assuming a Gaussian error, $\epsilon_i$, and they are calculated using predictive mean. The uncertainty associated with each prediction can also be quantified using the predictive variance [34].

### 2.2. Gaussian Process Autoregressive Regression (GPAR)

In most ML problems, a multi-output architecture that can capture the dependencies between the outputs is necessary. Gaussian Process Autoregressive Regression (GPAR), developed by Requeima et al. [35], accomplishes this by decomposing the output distributions into a collection of one-dimensional conditional distributions. These distributions can be seen as a decoupled group of single-output GPR problems. In chemical reactions, the reaction rates of chemical species are non-linearly dependent, and these dependencies can be captured by GPAR. A brief description of the GPAR algorithm, introduced by Requeima et al. [35], is given as follows:

Considering the problem with $M$ outputs $(y_{1:M}(\boldsymbol{x}_i) = (y_1(\boldsymbol{x}_i), \ldots, y_M(\boldsymbol{x}_i)))$ as a function of input $\boldsymbol{x}_i$. When the product rule is applied, the following can be obtained where the terms multiplied with each other,

$$\begin{aligned} &p\left(y_{1:M}(\boldsymbol{x}_i)\right) = p\left(y_1(\boldsymbol{x}_i)\right) \cdot \\ &\underbrace{p\left(y_2(\boldsymbol{x}_i) \mid y_1(\boldsymbol{x}_i)\right)}_{\substack{y_2(\boldsymbol{x}_i) \text{ as a random} \\ \text{function of } y_1(\boldsymbol{x}_i)}} \cdots \underbrace{p\left(y_M(\boldsymbol{x}_i) \mid y_{1:M-1}(\boldsymbol{x}_i)\right)}_{\substack{y_M(\boldsymbol{x}_i) \text{ as a random} \\ \text{function of } y_{1:M-1}(\boldsymbol{x}_i)}} \end{aligned} \tag{5}$$

which means that $y_1(\boldsymbol{x}_i)$ is obtained from $\boldsymbol{x}_i$, according to a random function, $f_1$; that $y_2(\boldsymbol{x}_i)$ is then obtained from $y_1(\boldsymbol{x}_i)$ and $\boldsymbol{x}_i$, and so on and so forth. This chain of relations can be shown as the following:

$$\begin{aligned} y_1(\boldsymbol{x}_i) &= f_1(\boldsymbol{x}_i), && f_1 \sim p\left(f_1\right) \\ y_2(\boldsymbol{x}_i) &= f_2\left(y_1(\boldsymbol{x}_i), \boldsymbol{x}_i\right), && f_2 \sim p\left(f_2\right) \\ &\vdots \\ y_M(\boldsymbol{x}_i) &= f_M\left(y_{1:M-1}(\boldsymbol{x}_i), \boldsymbol{x}_i\right) && f_M \sim p\left(f_M\right) \end{aligned} \tag{6}$$

As previously discussed, in a function-space view, a GP is a distribution over functions. Now, if $f_{1:M}$ is marginalised out, the conditionals in Eq. (5) are modelled with GPAR algorithm using GPs. The outputs depend not only on the local values of all other outputs but their full history. This dependency structure is also called non-local GPAR and the following generalisation can be made [35].

$$\begin{aligned} &y_m \mid y_{1:m-1} \\ &\sim \mathcal{GP}\left(0, k_m\left(\left(y_{1:m-1}, \boldsymbol{x}_i\right), \left(y_{1:m-1}, \boldsymbol{x}_i'\right)\right)\right) \end{aligned} \tag{7}$$

In both GPR and GPAR, the kernel is fixed to a squared exponential kernel, found to be the best kernel after a rigorous trial-error study. The rest of the hyperparameters, such as length scale ($l$) and signal variance ($\sigma_f$), and noise variance ($\sigma_n$) are automatically optimised using SciPy's built-in optimisation algorithm L-BFGS-B.

## 3. Results and discussion

### 3.1. Computational experiment

In this study, the capabilities of GPR and GPAR are tested as time-integrators of the thermochemical variables for $H_2$-air auto-ignition. The primary objective is to replace the ODE solver by integrating the whole thermochemical evolution given an initial condition and a fixed time-step. The auto-ignition event is chosen for two key reasons: (1) it represents stiff chemical kinetics, and (2) it shows highly non-linear characteristics. The dataset is obtained from the ignition of $H_2$-air mixture at isobaric and lean conditions using the ideal gas constant pressure reactor in Cantera [36]. The initial pressure of the mixture is

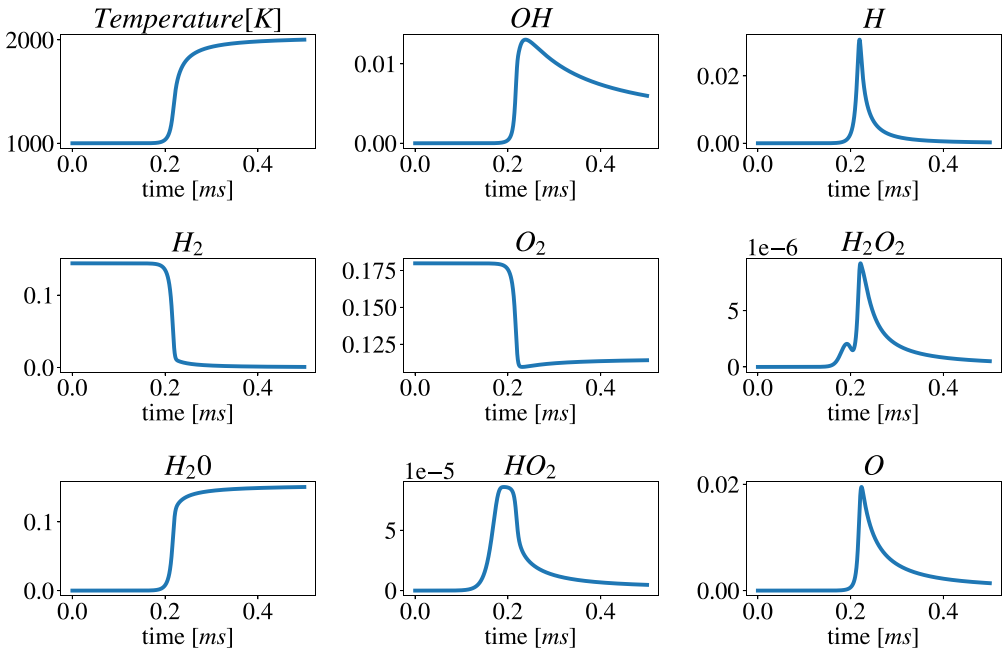


**Fig. 1.** Temperature and species mass fraction profiles of atmospheric lean auto-ignition of $H_2$-air at $T_i$ = 1000 K, $\phi$ = 0.4.

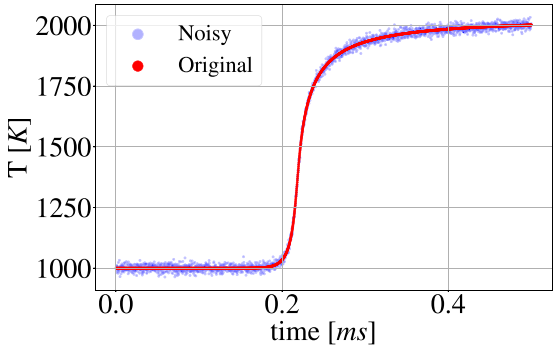


**Fig. 2.** Noise added temperature profile for the atmospheric lean auto-ignition of $H_2$-air at $T_i$ = 1000 K, $\phi$ = 0.4.

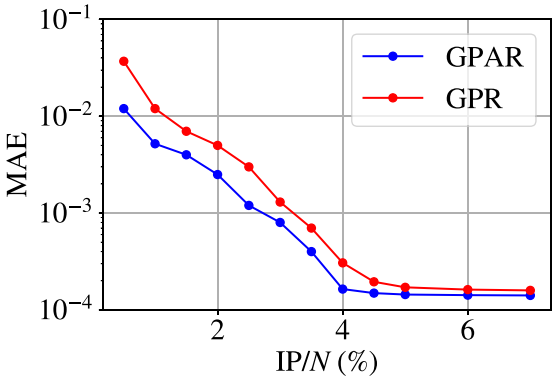


**Fig. 3.** The mean absolute error (MAE) of the normalised temperature and species mass fractions versus the percentage of IP to the total data set size, $N$.

set to $P_i$ = 1 atm, and the initial temperature and the equivalence ratio are varied between $T_i$ = 900–1100 K and $\phi$ = 0.4–1.0 with increments of 50 K and 0.1, respectively. The simulations are performed using a constant time-step of $10^{-4}$ ms for a total simulation time of 0.5 ms. The choice of time-step is based on the appropriate high-resolution representation of the evolution of chemical species and temperature.

The chemical kinetic mechanism of Conaire et al. [37] developed for $H_2/O_2$ combustion with 11 species and 19 reactions is used. After neglecting non-reacting species, the final thermochemical state vector consists of 8 reacting species and temperature ($\boldsymbol{D}\ \epsilon\ \mathbb{R}^9$). This computational experiment setup results in a total dataset size of $N$ = 85,500 observations (rows). The evolution of temperature and species profiles at $\phi$ = 0.4 is shown in Fig. 1.

$$\boldsymbol{D} = \begin{bmatrix} \vdots & \vdots & \vdots & & \vdots & \vdots \\ T & Y_1 & Y_2 & \dots & Y_{N-1} & Y_N \\ \vdots & \vdots & \vdots & & \vdots & \vdots \end{bmatrix}_{N\times 9} \tag{8}$$

### 3.2. Data preparation

The dataset is transformed using a power transformation since there are orders of magnitude differences between the feature vectors, such as temperature and $H_2O_2$. Furthermore, data transformation helps to avoid numerical stability issues with the Cholesky decomposition due to entries close to machine precision ($10^{-16}$). For new predictions, the scaled predictions are easily transformed back to the original scale without loss of information.

In real combustion systems, the profiles of thermo-chemical state variables are noisy due to perturbations originating from turbulence, diffusion, pressure fluctuations, and so on. Therefore, the data coming from homogeneous reactors only describe an idealised case without noise. Typically, the models trained using an idealised training dataset would lack robustness and be less expressive. Alternatively, datasets from 3D-CFD simulations of burners or engines could be used, but this approach is computationally intensive and would only be geometry-dependent. Therefore, to reflect the real system noise signal, a Gaussian noise term ($\epsilon_i$ in Eq. (1)) is added to each observation in the output matrix. The noise term acts as a regulariser, increasing model robustness and expressivity while reducing possible over-fitting. The noisy temperature output at $T_i$ = 1000 K, $\phi$ = 0.4 is shown in Fig. 2.

Finally, the training dataset is randomly split into three subsets: 80% for the train (68,400 observations), 10% for validation (8,550

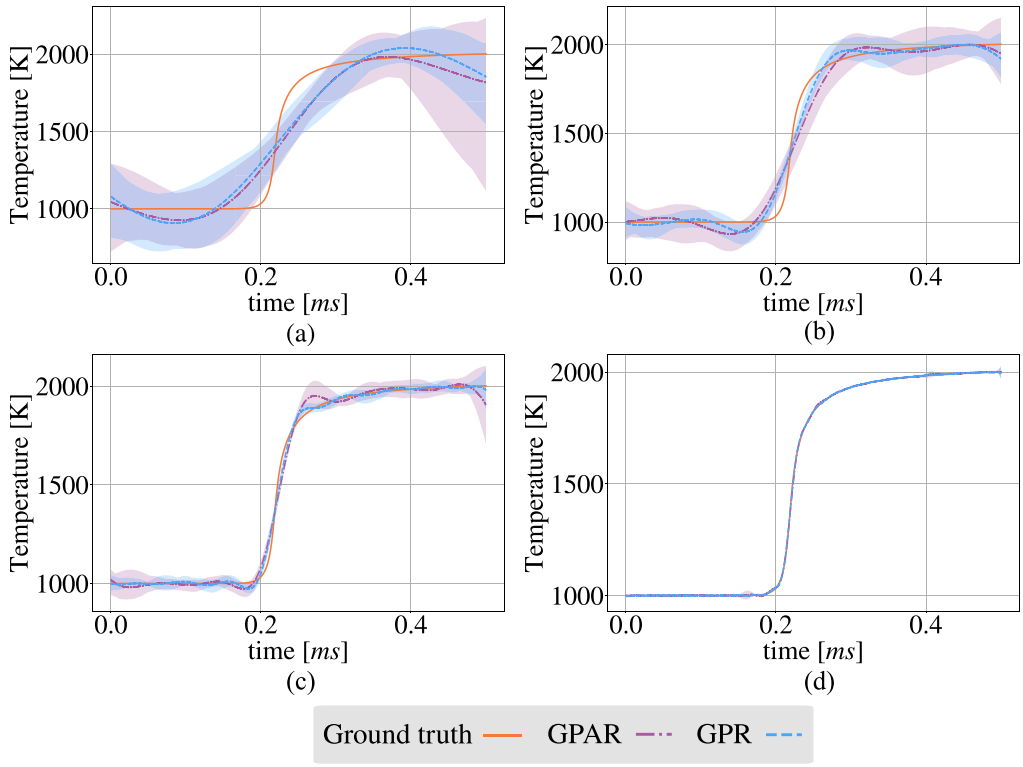


**Fig. 4.** The uncertainty (confidence interval) associated with different training dataset sizes for temperature; (a) IP/$N$ (%) = 1.0, (b) IP/$N$ (%) = 2.0, (c) IP/$N$ (%) = 3.0, (d) IP/$N$ (%) = 4.0.

observations), 10% for the test set (8,550 observations). This partitioning strategy ensures a diverse representation of the data across these subsets, enabling effective training, validation, and evaluation of the proposed framework.

### 3.2.1. Scalability

The most prominent weakness of standard GP is that it suffers from a cubic time complexity ($\mathcal{O}(N^3)$) and memory complexity of ($\mathcal{O}(N^2)$) [33]. In large datasets ($N > 10^4$), scalability becomes an issue due to the computational complexity of GPs. For general multi-output GPRs, the scaling factor is $\mathcal{O}(M^3N^3)$ while GPAR scales with $\mathcal{O}(MN^3)$ since the outputs ($M$) are essentially decoupled one-dimensional GPs.

In cases where the function is oversampled to approximate the exact solution, the variational inducing point (IP) method for sparse GPR of GPAR can be applied [38]. The variational IP method allows sparse approximation of the latent function using a subset of original data points. In this study, this technique is used to achieve faster and more scalable computation of the posterior function distributions. The number of IPs and their distribution directly affects the posterior predictions. Previous research has shown that regularly spaced IPs perform well with time series data similar to time-integration of chemistry [39].

The percentage of regularly spaced IPs of the total number of observations ($N$) in $\boldsymbol{x}_i$ required to model the latent functions for each output ($M$) with reasonable accuracy in the training dataset is shown in Fig. 3. The MAE for species mass fractions drastically drops as the number of IPs is increased, also resulting in reduced uncertainty. It is seen that the MAE reaches an asymptote at 5% for both GPR and GPAR. This corresponds to train, validation, and test sets of 3421, 427, and 427 observations each, respectively. It can also be noticed that the GPAR framework shows better overall accuracy by capturing dependencies between the outputs. The uncertainty, represented by 95% confidence intervals, for varying dataset sizes in comparison with the ground truth from the 0D auto-ignition simulation is shown in Fig. 4. It is evident that the uncertainty progressively diminishes by increasing the dataset size.

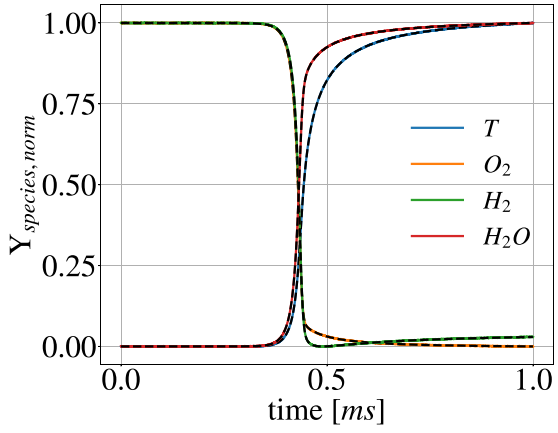


**Fig. 5.** Normalised actual (dashed lines) and predicted (solid lines) normalised species mass fractions of temperature and major species for $T_i$ = 1000 K, $P_i$ = 1 atm, and $\phi$ = 0.4 using GPAR.

The normalised predicted and actual profiles of temperature and major species at $T_i$ = 1000 and $\phi$ = 0.4 using GPAR are shown in Fig. 5. Similarly, normalised predicted and actual profiles of minor species at the same conditions are shown in Fig. 6. Since these conditions are within the initial training set, the framework accurately captures the latent dynamics using a sparse approximation.

The mean absolute error between the normalised ground truth and predictions of mass fractions for the train set is depicted in Fig. 7. It is evident that the difference between the GPR and GPAR predictions is minimal. The error magnitudes are mostly of the same order, except for the $HO_2$. This could be attributed to different local dynamics of $HO_2$ during ignition.

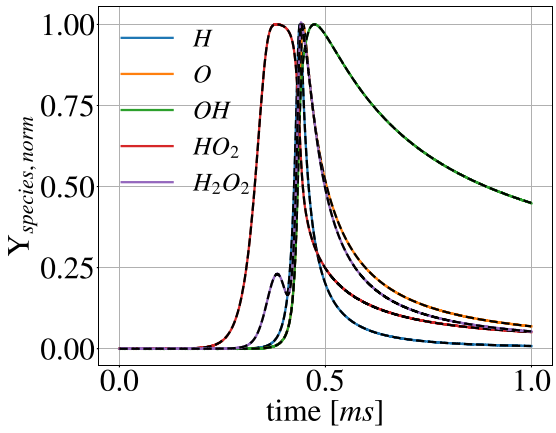


**Fig. 6.** Normalised actual (dashed lines) and predicted (solid lines) normalised species mass fractions of minor species for $T_i$ = 1000 K, $P_i$ = 1 atm and $\phi$ = 0.4 using GPAR.

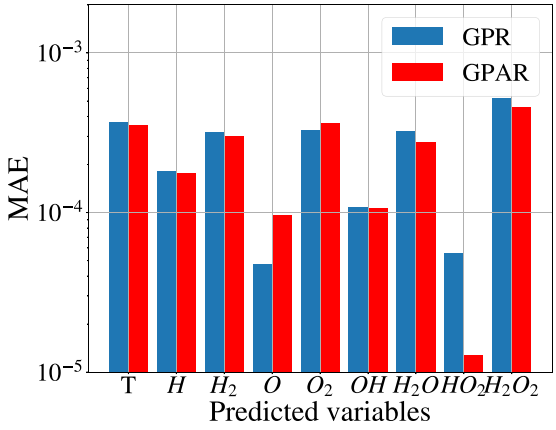


**Fig. 7.** The train set MAE of the predictions for normalised temperature and species mass fractions.

### 3.3. Artificial Neural Networks (ANN)

As previously mentioned, ANNs have been the primary choice of ML algorithm in the prediction of chemical source terms in the existing literature. Therefore, a comparison between ANNs and the GP framework in terms of dataset size requirements, accuracy, and speed is necessary. In this study, we reference the setup by Wan et al. [25] for the ANN configuration. However, the architecture of the ANN is simplified to reduce the complexity of the model while maintaining its expressivity. The architecture consists of three dense layers with 64, 32, and 16 neurons, respectively, implemented in TensorFlow as shown in Fig. 8. The batch size and epochs were set to 10 and 150, respectively. The Adam optimiser with a learning rate of $3 \times 10^{-4}$ and a decay of $10^{-4}$ recommended by Wan et al. [25] was used. The activation functions, ReLu and GeLu, were tested and it was seen that the GeLu provided better stability due to its smoothness and differentiability. Additionally, the early stopping method was used with a patience parameter of 50 to stop model training when the validation loss does not further decrease after 50 consecutive epochs. The training dataset initially consisted of the set of variational IPs used for the GP framework, totalling 3421, 427, 427 observations for each set, respectively. This dataset size was later increased to achieve the same level of accuracy as the GP framework. The change in the accuracy of the ANN with training dataset size is shown in Fig. 9. Furthermore, to ensure generality and prevent over-fitting of the model, k-fold cross-validation (10-fold) was utilised.

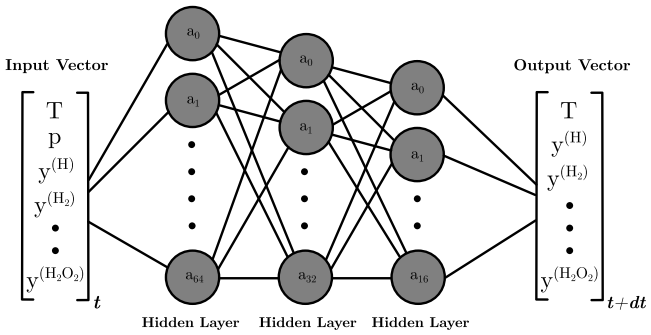


**Fig. 8.** The neural network architecture with three densely connected layers.

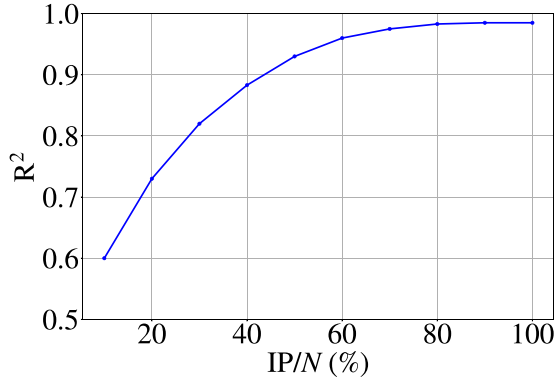


**Fig. 9.** The accuracy of ANN with increasing training dataset size.

### 3.4. Model benchmark

The performances of all models on train, validation, and test sets together with the speed-up relative to the 0D code, are given in Table 1. It is seen that both GPR and GPAR fully reconstruct the latent thermochemical space with high accuracy. Both models demonstrate good generalisation, with GPAR showing slightly better accuracy. However, the ANN model seems to suffer from insufficient dataset size, resulting in lower accuracy and slight over-fitting. This can also be observed in Fig. 10-c where the ANN model has a wider spread on the test set predictions. The accuracy of the ANN model improves with a larger dataset size as shown in Fig. 9, leading to enhanced generalisability. When the speed-up ratios relative to the 0D reactor model are compared, the ANN model outperforms the GPs with a speed-up ratio of 3.0. This highlights a limitation of GPs, where quantification of uncertainty comes at the cost of storing and processing the data points for the covariance matrix.

The models are further assessed by testing them on a randomly selected unseen condition: an initial temperature of $T_i$ = 970 K and an equivalence ratio of $\phi$ = 0.47. The ground truth and predictions from each model, including the confidence intervals from the GP models are shown in Fig. 11. It is evident from Fig. 11 that both GP models captured the latent dynamics at the given condition with minimal uncertainty. However, the uncertainty increases locally for $HO_2$, indicating challenges in capturing gradients correctly, possibly due to different dynamics compared to other species such as H, O, and $H_2O_2$ where the gradients increase sharply during ignition. In contrast, the ANN model predicts the new condition reasonably well, although deviations from the ground truth are present especially right before the ignition starts.

## 4. Conclusion

The present work proposes a probabilistic ML framework that builds upon GP algorithms, namely GPR and GPAR, for chemical source

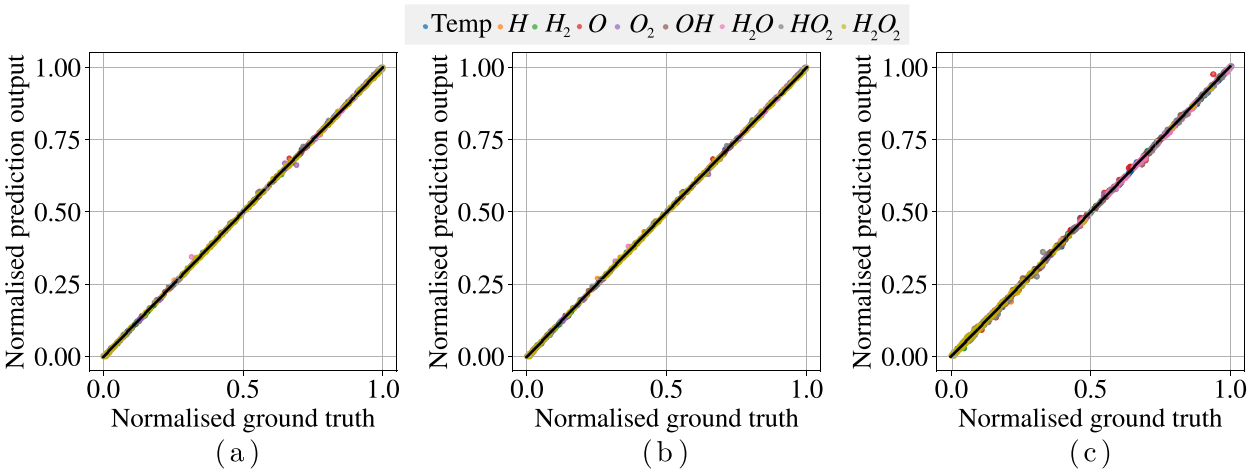


**Fig. 10.** The normalised ground truth versus predicted values for the test set for (a) GPR; (b) GPAR; (c) ANN.

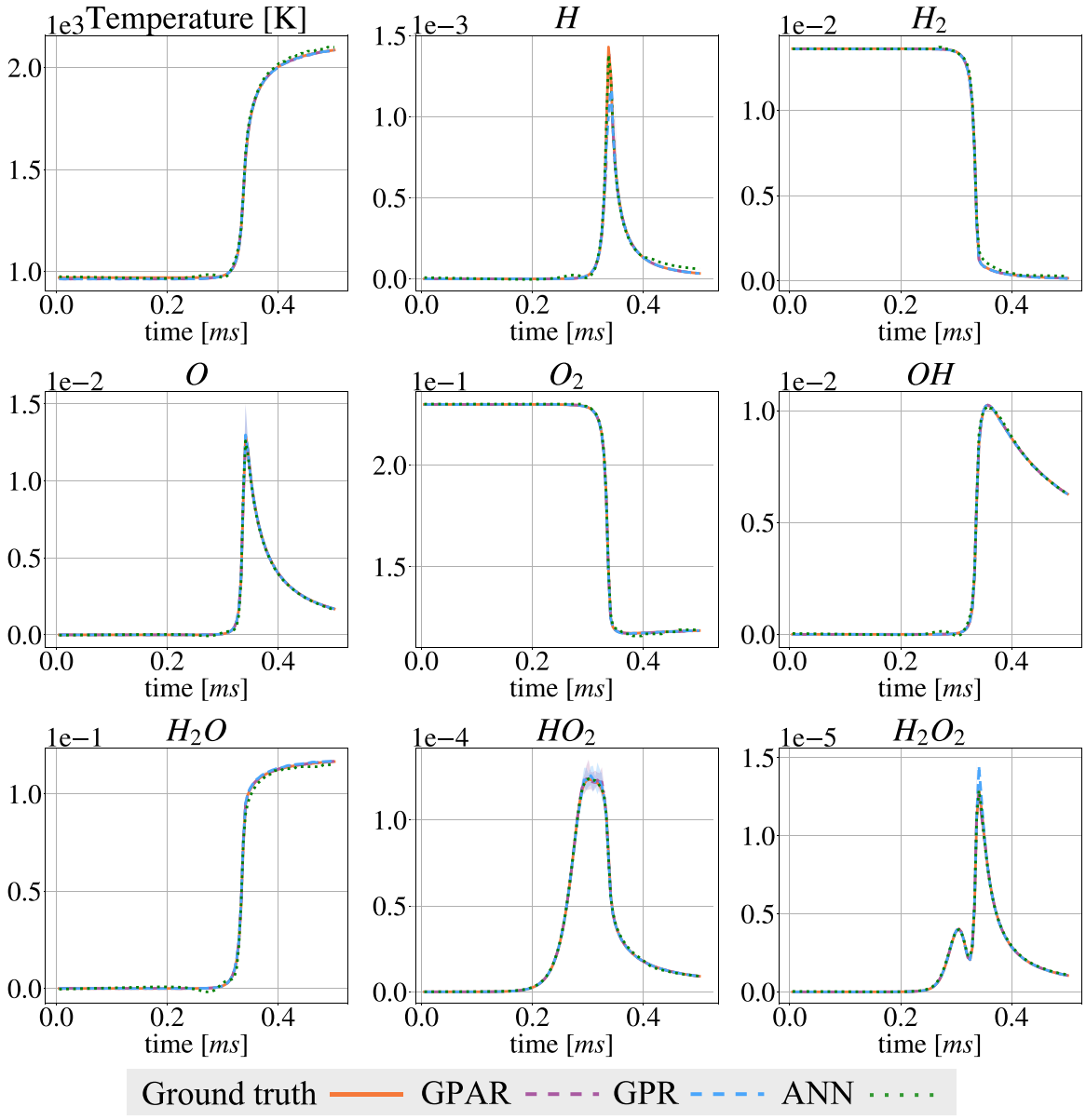


**Fig. 11.** The ground truth versus predicted values for an unseen condition at $T_i$ = 970 K and $\phi$ = 0.47 from GPR, GPAR and ANN models.

**Table 1**
Comparative performance evaluation of the models.

| ML model | Split set | $R^2$ | MAE | Relative speed-up ($t_{0D}/t_{ML}$) |
|---|---|---|---|---|
| GPR | Train | 0.998 | $1.61 \times 10^{-8}$ | 1.9 |
| | Validation | 0.998 | $1.62 \times 10^{-8}$ | |
| | Test | 0.997 | $2.12 \times 10^{-8}$ | |
| GPAR | Train | 0.999 | $1.39 \times 10^{-8}$ | 2.1 |
| | Validation | 0.999 | $1.41 \times 10^{-8}$ | |
| | Test | 0.998 | $1.53 \times 10^{-8}$ | |
| ANN | Train | 0.993 | $6.24 \times 10^{-8}$ | 3.0 |
| | Validation | 0.992 | $9.02 \times 10^{-8}$ | |
| | Test | 0.988 | $2.80 \times 10^{-7}$ | |

term integration. The probabilistic nature of GPs allows them to work with limited data and provide uncertainty-quantified predictions. The capabilities of the framework were tested on a 0D reactor-based noisy dataset for $H_2$/air auto-ignition case. Variational IPs were used to make sparse approximations of the latent dynamics of the system while keeping the training dataset size to a minimum.

The framework was benchmarked to a standard ANN model in terms of statistical error and relative speed-up factors. Although GPR and GPAR reproduced similar MAEs, GPAR slightly outperformed GPR on both accuracy and relative speed-up. In contrast, the ANN model was prone to over-fitting due to insufficient dataset size but offered a better relative speed-up factor compared to GP models. It was found that GP models have the advantage of providing confidence intervals that are especially relevant when working with small, noisy datasets derived from experiments or high-fidelity simulations.

Future work will focus on enhancing the data efficiency of the GP framework through optimal sampling methods such as active learning, aiming to speed up inference processes. Moreover, the extrapolation capabilities of GPs beyond the training domain, will be explored through the use of physics-informed kernels. Finally, we aim to investigate the performance of the GP framework with other complex fuels and mechanisms, assessing its effectiveness in higher-dimensional input and output spaces.

## CRediT authorship contribution statement

**Cihat Emre Üstün:** Writing – original draft, Visualization, Software, Methodology, Investigation. **Amin Paykani:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


[1] Verhelst S. Recent progress in the use of hydrogen as a fuel for internal combustion engines. Int J Hydrog Energy 2014;39(2):1071–85.
[2] Sánchez AL, Williams FA. Recent advances in understanding of flammability characteristics of hydrogen. Prog Energy Combust Sci 2014;41:1–55.
[3] Zhou H, Xue J, Gao H, Ma N. Hydrogen-fueled gas turbines in future energy system. Int J Hydrog Energy 2024;64:569–82.
[4] Abubakar S, Said MFM, Abas MA, Ismail NA, Khalid AH, Roslan MF, Kaisan MU. Hydrogen-fuelled internal combustion engines-Bibliometric analysis on research trends, hotspots, and challenges. Int J Hydrog Energy 2024;61:623–38.
[5] Momirlan M, Veziroglu TN. The properties of hydrogen as fuel tomorrow in sustainable energy system for a cleaner planet. Int J Hydrog Energy 2005;30(7):795–802.
[6] Du H, Chai WS, Wei H, Zhou L. Status and challenges for realizing low emission with hydrogen ultra-lean combustion. Int J Hydrog Energy 2024;57:1419–36.
[7] Amrouche F, Erickson P, Park J, Varnhagen S. An experimental evaluation of ultra-lean burn capability of a hydrogen-enriched ethanol-fuelled wankel engine at full load condition. Int J Hydrog Energy 2016;41(42):19231–42.
[8] Najm HN, Wyckoff PS, Knio OM. A semi-implicit numerical scheme for reacting flow: I. stiff chemistry. J Comput Phys 1998;143(2):381–402.
[9] Knio OM, Najm HN, Wyckoff PS. A semi-implicit numerical scheme for reacting flow: II. Stiff, operator-split formulation. J Comput Phys 1999;154(2):428–67.
[10] Lanser D, Verwer JG. Analysis of operator splitting for advection–diffusion–reaction problems from air pollution modelling. J Comput Appl Math 1999;111(1–2):201–16.
[11] Franke LL, Chatzopoulos AK, Rigopoulos S. Tabulation of combustion chemistry via Artificial Neural Networks (ANNs): Methodology and application to LES-PDF simulation of Sydney flame L. Combust Flame 2017;185:245–60.
[12] Liu A, Ding T, Liu R, Rigopoulos S, Luo K. Machine learning tabulation of thermochemistry for turbulent dimethyl ether (DME) flames. Fuel 2024;360:130338.
[13] Zhou L, Song Y, Ji W, Wei H. Machine learning for combustion. Energy AI 2022;7:100128.
[14] Nguyen H-T, Domingo P, Vervisch L, Nguyen P-D. Machine learning for integrating combustion chemistry in numerical simulations. Energy AI 2021;5:100082.
[15] Üstün CE, Herfatmanesh MR, Valera-Medina A, Paykani A. Applying machine learning techniques to predict laminar burning velocity for ammonia/hydrogen/air mixtures. Energy AI 2023;13:100270.
[16] Üstün CE, Eckart S, Valera-Medina A, Paykani A. Data-driven prediction of laminar burning velocity for ternary ammonia/hydrogen/methane/air premixed flames. Fuel 2024;368:131581.
[17] Han W, Sun Z, Scholtissek A, Hasse C. Machine Learning of ignition delay times under dual-fuel engine conditions. Fuel 2021;288:119650.
[18] Li J, Liang W, Han W. Predicting the explosion limits of hydrogen-oxygen-diluent mixtures using machine learning approach. Int J Hydrog Energy 2024;50:1306–13.
[19] Ihme M, Chung WT, Mishra AA. Combustion machine learning: Principles, progress and prospects. Prog Energy Combust Sci 2022;91:101010.
[20] Christo F, Masri A, Nebot E. Artificial neural network implementation of chemistry with pdf simulation of H2/CO2 flames. Combust Flame 1996;106(4):406–27.
[21] Blasco J, Fueyo N, Dopazo C, Ballester J. Modelling the temporal evolution of a reduced combustion chemical system with an artificial neural network. Combust Flame 1998;113(1–2):38–52.
[22] An J, He G, Luo K, Qin F, Liu B. Artificial neural network based chemical mechanisms for computationally efficient modeling of hydrogen/carbon monoxide/kerosene combustion. Int J Hydrog Energy 2020;45(53):29594–605.
[23] An J, Wang H, Liu B, Luo KH, Qin F, He GQ. A deep learning framework for hydrogen-fueled turbulent combustion simulation. Int J Hydrog Energy 2020;45(35):17992–8000.
[24] Sharma AJ, Johnson RF, Kessler DA, Moses A. Deep learning for scalable chemical kinetics. In: AIAA scitech 2020 forum. 2020, p. 0181.
[25] Wan K, Barnaud C, Vervisch L, Domingo P. Chemistry reduction using machine learning trained from non-premixed micro-mixing modeling: Application to DNS of a syngas turbulent oxy-flame with side-wall effects. Combust Flame 2020;220:119–29.
[26] Brown TS, Antil H, Löhner R, Togashi F, Verma D. Novel DNNs for stiff odes with applications to chemically reacting flows. In: High performance computing: ISC high performance digital 2021 international workshops, frankfurt am main, Germany, June 24–July 2, 2021, revised selected papers 36. Springer; 2021, p. 23–39.
[27] Zhang T, Yi Y, Xu Y, Chen ZX, Zhang Y, Weinan E, Xu Z-QJ. A multi-scale sampling method for accurate and robust deep neural network to predict combustion chemical kinetics. Combust Flame 2022;245:112319.
[28] Nikitin V, Karandashev I, Malsagov MY, Mikhalchenko E. Approach to combustion calculation using neural network. Acta Astronaut 2022;194:376–82.
[29] Yousefian S, Bourque G, Monaghan RF. Bayesian inference and uncertainty quantification for hydrogen-enriched and lean-premixed combustion systems. Int J Hydrog Energy 2021;46(46):23927–42.
[30] Wang H, Slavinskaya N, Haidn O. A comprehensive kinetic modeling study of hydrogen combustion with uncertainty quantification. Fuel 2022;319:123705.
[31] Soyler I, Zhang K, Duwig C, Jiang X, Karimi N. Uncertainty quantification of the premixed combustion characteristics of NH3/H2/N2 fuel blends. Int J Hydrog Energy 2023;48(38):14477–91.
[32] Soyler I, Zhang K, Jiang X, Karimi N. Effects of compositional uncertainties in cracked NH3/biosyngas fuel blends on the combustion characteristics and performance of a combined-cycle gas turbine: A numerical thermokinetic study. Int J Hydrog Energy 2024;69:504–17.
[33] Williams CK, Rasmussen CE. Gaussian processes for machine learning, Vol. 2, MIT press Cambridge, MA; 2006.
[34] Rasmussen CE. Gaussian processes in machine learning. In: Summer school on machine learning. Springer; 2003, p. 63–71.

[35] Requeima J, Tebbutt W, Bruinsma W, Turner RE. The Gaussian Process Autoregressive Regression Model (GPAR). 2019, arXiv:1802.07182.
[36] Goodwin DG, Moffat HK, Speth RL. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. 2018.
[37] Ó Conaire M, Curran HJ, Simmie JM, Pitz WJ, Westbrook CK. A comprehensive modeling study of hydrogen oxidation. Int J Chem Kinet 2004;36(11):603–22.
[38] Titsias M. Variational learning of inducing variables in sparse Gaussian processes. In: Artificial intelligence and statistics. PMLR; 2009, p. 567–74.
[39] Bui TD, Turner RE. Tree-structured Gaussian process approximations. Adv Neural Inf Process Syst 2014;27.