

Article

Combustion Regime Identification in Turbulent Non-Premixed Flames with Principal Component Analysis, Clustering and Back-Propagation Neural Network

Hanlin Zhang, Hao Lu *, Fan Xie , Tianshun Ma and Xiang Qian

School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
* Correspondence: haolu@hust.edu.cn

Abstract: Identifying combustion regimes is important for understanding combustion phenomena and the structure of flames. This study proposes a combustion regime identification (CRI) method based on rotated principal component analysis (PCA), clustering analysis and the back-propagation neural network (BPNN) method. The methodology is tested with large-eddy simulation (LES) data of two turbulent non-premixed flames. The rotated PCA computes the principal components of instantaneous multivariate data obtained in LES, including temperature, and mass fractions of chemical species. The flame front results detected using the clustering analysis do not rely on any threshold, indicating the quantitative characteristic given by the unsupervised machine learning provides a perspective towards objective and reliable CRI. The training and the subsequent application of the BPNN rely on the clustering results. Five combustion regimes, including environmental air region, co-flow region, combustion zone, preheat zone and fuel stream are well detected by the BPNN, with an accuracy of more than 98% using 5 scalars as input data. Results showed the computational cost of the trained supervised machine learning was low, and the accuracy was quite satisfactory. For instance, even using the combined data of CH₄-T, the method could achieve an accuracy of more than 95% for the entire flame. The methodology is a practical method to identify combustion regime, and can provide support for further analysis of the flame characteristics, e.g., flame lift-off height, flame thickness, etc.

Keywords: non-premixed; identification; principal component analysis; cluster; back-propagation neural network



Citation: Zhang, H.; Lu, H.; Xie, F.; Ma, T.; Qian, X. Combustion Regime Identification in Turbulent Non-Premixed Flames with Principal Component Analysis, Clustering and Back-Propagation Neural Network. *Processes* **2022**, *10*, 1653. <https://doi.org/10.3390/pr10081653>

Academic Editor: Zhihua Wang

Received: 30 June 2022

Accepted: 17 August 2022

Published: 20 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Non-premixed flames constitute a specific class of combustion processes where fuel and oxidizer are separated before burning, and have been adopted in many advanced combustion applications, such as gas turbines, direct-injection engines, and furnaces [1]. Complex multiple regimes, including fuel stream, oxidizer stream, mixing layers of fuel and oxidizer streams, etc., appear in non-premixed flames. Therefore, a global characterization parameter could not reflect all location conditions. Accurate and detailed combustion regime identification (CRI) in non-premixed flames is helpful to understand combustion phenomena, the structure of the flame and the flame characterizations, such as ignition delay time, flame lift-off height, flame thickness, flashback, etc., and further help to study the energy transfer, the transport and change of species, the interactions between flow and reaction, and combustion modeling [2–5]. In the present work, we propose a method to characterize the local regimes within turbulent non-premixed flame, and validate it in two well-known flames.

Flame index method [6] is the first kind of numerical method proposed to identify different combustion regimes. Basically, it evaluates the alignment of fuel and oxidizer gradients and, thus, gives an indication of the nature of the local combustion regime ranging between premixed flame and diffusion flame. After years of development, this

method has made great progress [2,5]. However, the main disadvantage of the flame index method is that it depends on a large number of field data, and requires detailed fine-scale information including gradients. Hartl et al. [2] and Butz et al. [7] have proposed a gradient free regime identification (GFRI) method, which combines the mixture fraction, the heat release rate (HRR), and the chemical explosive mode to detect and characterize premixed versus non-premixed reaction regimes; however, a considerable amount of computation is still required to process the data [4].

In experiments, it is not easy to apply the flame index method since high resolution combustion field data are difficult to obtain. Before the introduction of the flame index method, experimenters used planar laser-fluorescence (PLIF) imaging technology to study the flame structure, and the identification of combustion zone has been quite reliable [8,9]. Most of the efforts have concentrated on the measurement of OH distribution because of its high abundance in flames and the coincidence of OH transitions with high-power excimer laser wavelengths [8]. For this reason, OH distribution is often used to characterize the reaction regime in simulations. For instance, in direction numerical simulations of turbulent non-premixed flames, Kerkemeier [10] has adopted a commonly used OH-threshold to define the autoignition event, and the iso-surface of $Y_{OH} = 10^{-4}$ is used as a marker of flame front. Flame front identification using the OH-threshold, however, cannot give sufficient information about important chemical reaction paths, fuel consumption rate and HRR. Analyses have indicated the distribution of HCO correlates well with peak HRR. Since the PLIF measurements of HCO distribution seem not to be feasible, studies have proposed the use of the product of OH and CH_2O PLIF intensities to identify flame fronts [9]. The threshold of chemical species has become an important indicator in reaction regime identification in both experimental and numerical studies [3,8,9], because it only requires a little thermochemical information and the amount of computation is very low. However, it also has two main disadvantages: (1) it usually can only distinguish the reaction regime from other regimes by identifying the flame front; and (2) the identification results are highly sensitive to the critical value of the scalar threshold.

Machine learning is a powerful tool for multivariate analysis and can offer an alternative to traditional statistical methods for CRI and analysis of dynamic tracking of the flame. Jigjid et al. [11] have developed a predictive tool based on neural networks to identify combustion modes in MILD combustion. Wan et al. [4] have adopted a convolutional neural network (CNN) trained by the GFRI results of the experimental data from a laboratory scale burner [2,7]. This CNN method offers a pixel-wise accuracy of more than 85%, and, compared to the flame index method and the GFRI, it is ultra-fast, making it possible to envision real-time CRI for advanced flame control.

The CNN method proposed by Wan et al. [4] relies on the GFRI to provide training data to make predictions. An accurate and reliable database is very important for supervised machine learning. However, the GFRI still needs to manually determine the threshold strength of scalars to determine the classifications of combustion regime, and also it is quite computationally expensive [4]. Unsupervised learning is a fast and reliable method to analyze and classify data. It can label the combustion categories in the turbulent combustion field more objectively, rather than based on a-priori knowledge. Barwey et al. [12] have adopted the K-means, a commonly used clustering algorithm, to achieve CRI in detonation waves enabling local source term modeling. Himanshu et al. [13] have adopted another clustering algorithm to characterize MILD combustion. These studies have shown the potential of unsupervised learning techniques in characterizing flames. Thus, this study proposes the use of clustering analysis to distinguish the quantitative characteristics of the combustion field. Referring to the study by Wan et al. [11], the accurate flame classification results can be used as the input of the neural network. Unlike previous recognition methods (e.g., Wan et al. [11]), this study proposes a novel machine learning-based CRI method that requires only a small number of scalars to obtain high accuracy. The turbulent combustion flame database is briefly described in the subsequent section, and the proposed CRI method

is detailed in Section 3. Then, the method is tested over two flames to examine its capability of CRI.

2. Large-Eddy Simulation Database

The data used in this study were from the high fidelity 3-dimensional (3D) LESs [14–17] of two well-known turbulent non-premixed flames, a jet-in-hot-co-flow MILD combustion case, HM1 [18], and Sandia Flame D [19]. The chemistry was modeled using the GRI-Mech 2.11 [20], which contained 277 elementary chemical reactions of 49 species.

The fuel jet mixture of the HM1 consisted of 80% CH₄ and 20% H₂ (percent-by-mass basis), and the co-flow consisted of 3% O₂, 5.5% CO, 6.5% H₂O, and 85% N₂ (percent-by-mass basis). The experimentally reported mean temperatures suggested that the fuel jet and shroud air temperature profiles were uniform with values of 305 K and 300 K, respectively. The mean co-flow temperature was approximately 1300 K. The numerical setup followed the description of the experiment, which consisted of an insulated and cooled central fuel jet with a diameter of 4.25 mm and co-flow with a diameter of 82 mm. The experimentally reported bulk flow velocity of the fuel stream was 73.5 m/s, and the corresponding Reynolds number was 10,000. The fuel jet of the Sandia Flame D contained 25% CH₄ and 75% air (percent-by-volume basis), and the equivalence ratio of the pilot/co-flow was 0.77. The experimentally reported mean temperatures suggested that the fuel jet and shroud air temperature profiles were uniform with values of 294 K and 291 K, respectively. The mean pilot/co-flow temperature was approximately 1880 K. The fuel nozzle bore a diameter of 7.2 mm and was enclosed by a pilot/co-flow nozzle with a diameter of 18.2 mm. The bulk fuel jet velocity of Flame D was 49.6 m/s, and the corresponding Reynolds number was 22,400.

3. Identification Method

3.1. Principal Component Analysis

The LES of turbulent flames can yield a large number of 3D-space data, including three velocity components, pressure, temperature, mass fractions of species in reaction, etc. Multivariate analysis methods consist of many techniques that can be used to analyze a set of data to answer complex questions involving more than two variables. In this study, temperature and chemical data were used to identify combustion regimes. However, due to the huge amount of data, if it is not processed by dimension reduction, the subsequent calculation cost is greatly increased. For this reason, the study adopted rotated principal component analysis (PCA, also known as proper orthogonal decomposition) [21] to exclude the coexistence of chemical overlapping information, and conducted a pre-sort and a dimension reduction of data.

In specific, the application of PCA decomposed the multivariate data that made up the original data into a linear combination of empirical orthogonal basis vectors, which are called principal components (PCs). These basis vectors are the eigenvectors of a covariance matrix computed from a data matrix containing multivariate data from LES data.

For LES dataset X which contains n samples and N original variables, PCs will be determined through the eigenvalue problem. The covariance of X , S , is defined as $S = 1/(n - 1)XTX$, where the superscript T indicates the transpose matrix. Through the computation of determinant of S , $\det |S - \lambda I|$, where I is an identity matrix, the eigenvalue problem is now obtained as $S \times A_i = \lambda_i \times A_i$, in which A_i represents i th eigenvector ($1 \leq i \leq N$) and λ_i is the i th eigenvalue. The eigenvectors of S in descending order of the corresponding eigenvalues are the PCs. Some investigators use the correlation matrix; however, the use of the covariance matrix reduces sensitivity to noise (redundant information) in the data [22,23]. These unrotated PCs are basic functions of the original data set of the multivariate profiles through the LES domain and then linear combinations of the unrotated PCs can be used to describe and reconstruct these data profiles. The data reconstruction can also be accomplished after applying either an orthogonal or oblique rotation to the PCs. The use of PCA ensures that the analyst's biases do not affect the

quantitative results, only their subjective physical interpretation. The rotated PCA is useful merely as a dimension reduction technique, and is ineffectual for identifying modes of variability of physical data [22–24]. For the multi-PC phenomena described later in the study, the corresponding cluster of data points spreads across a vector subspace with as many dimensions as there are PCs needed to describe the phenomena. In specific, the study adopted the rotated PCA, which assists in compressing and extracting useful information from the original matrix by removing redundant information and finally obtaining the PCs used in the subsequent analysis and calculation.

In this study, the initial thermochemical data of two cases were 50-dimensional data (including temperature and mass fractions of 49 species). We used MATLAB to perform PCA analysis of the data. Figure 1 shows can see that for both cases, the first five PCs contained more than 95% of the information of 50 scalars. In order to reduce computational cost, these five-PC datasets retained after rotated PCA were used for subsequent analysis.

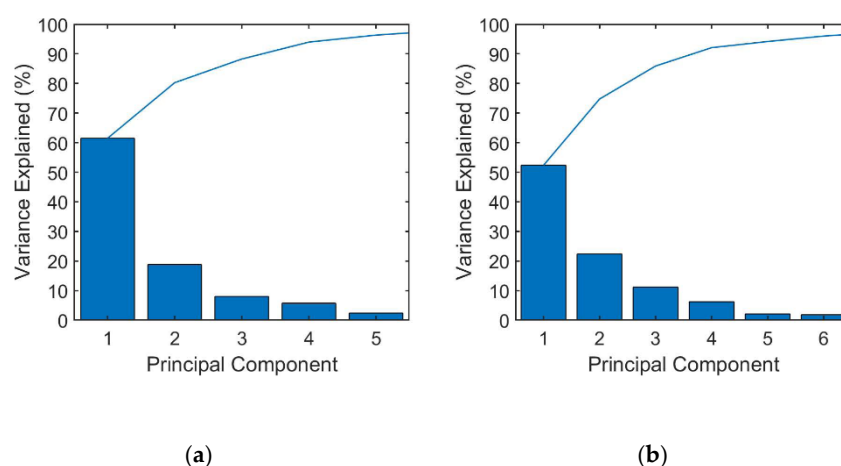


Figure 1. Pareto diagram of two cases after PCA: (a) HM1; (b) Sandia Flame D.

3.2. Clustering Analysis

The application of techniques of artificial intelligence, which can be used for analysis of complex multivariate data and prediction of nonlinearities, can potentially be useful in analyzing data, including turbulent combustion. After using rotated PCA to extract useful data from the original by removing redundant information, it was necessary to analyze the data to distinguish the characteristics of different combustion states. One of the vital means in dealing with the data was to classify or group them into a set of categories or clusters. Clustering algorithm is a branch of machine learning and belongs to unsupervised learning. The goal of clustering is to separate a finite unlabeled data set into a finite and discrete set of natural hidden data structures. Figure 2 shows a schematic diagram of combustion clustering, and scatters of temperature and fuel concentration are presented as an example. Clustering classification of scatters of low temperature and high fuel concentration represented fuel inlet; scatters of high temperature and low fuel concentration were located in the combustion zone; and the air inlet was characterized by low fuel concentration and low temperature. Connecting the fuel inlet and the combustion zone represented all reaction paths, while connecting the combustion zone and the air inlet was dominated by the heat transfer process. This diagram is only a two-dimensional diagram using fuel and temperature. If different scalars were used for analysis, the distribution of scatters would show different characteristics. For example, in the preheating area (yellow scatters in Figure 2) connecting the fuel inlet and the combustion zone, the concentration of CH_2O is very high, while it is very low in other regions. The concentrations of major products, such as H_2O , CO , and CO_2 , were high in the region represented by orange in Figure 2. Although the high-dimensional scalar distribution was difficult to represent by a two-dimensional scatter diagram, the example in Figure 2 presents the multi-dimensional

clustering analysis and can objectively classify multidimensional thermochemical data to distinguish various regimes.

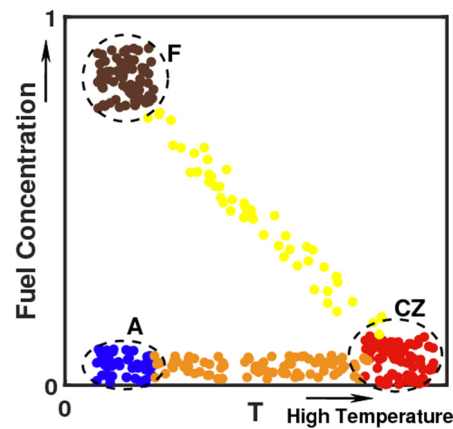


Figure 2. Schematic diagram of clustering using scatters of temperature and fuel concentration as an example. Colors of five clustering classifications will be identified later, and the three abbreviations, “CZ”, “A” and “F”, stand for combustion zone, air and fuel, respectively.

The study can use many clustering algorithms, but here we adopted two of the more commonly used algorithms. K-means [24] clustering algorithm is the earliest proposed clustering algorithm, and the K-means and its variants have been widely studied and applied in different scientific fields. The K-means requires the user to pre-specify the number of clusters present in the dataset. The K-means algorithm partitions a given set of data in a manner such that the squared-error function is minimized for a pre-specified number of clusters. The squared error function (E) is defined as:

$$E = \sum_{k=1}^K \sum_{x \in S_k} \|x - Z_k\|^2, \quad (1)$$

where K is number of specified clusters, the d -dimensional Z_k denotes the center of k th cluster and X represents a d -dimensional data vector belonging to the cluster S_k . The K-means algorithm aims to minimize the sum of squared distances between all points and the cluster centers. Compared with other clustering algorithms, the K-means clustering has an obvious advantage of fast calculation speed [25], which is also the first clustering algorithm used in this study.

Furthermore, this study also adopted the self-organizing maps (SOM) neural network method [26] for data analysis to compare with the results using K-means to see the reliability of clustering. The SOM is an unsupervised learning method to analyze various data sets, including those with missing values, and Chen et al. [27] have demonstrated that the SOM is a superior clustering technique. The SOM performs dimensionality reduction and classification and projects a high-dimensional input space onto a low dimensional topology so as to allow the number of data clusters to be visualized/determined by manual inspection. The computational steps of SOM algorithm are described below.

Step 1 (Initialization): Let X_i , $i = 1, 2, \dots, n$, be the d -dimensional vectors to be clustered. Select small random values for the initial weights, $W_{ij}(0)$, and fix the initial learning rate ($\hat{\alpha}_0$) and the neighborhood. $W_{ij}(0)$ is the d -dimensional weight vector associated with the node at location (i, j) of a 2-dimensional grid array

Step 2 (Determining the best matching unit (BMU)): Select a sample pattern, X , from the data set and determine the BMU (C_{ij}) at training iteration t , using the minimum Euclidean distance criterion.

$$\|X - W_{C_{ij}}\| = \min_{ij} \|X - W_{ij}\|; \quad i = 1, 2, \dots, L; \quad j = 1, 2, \dots, L; \quad (2)$$

where $\| \cdot \|$ is the Euclidean norm and L denotes the number of rows (and also columns) in the square 2-D SOM grid.

Step 3 (Weight updating): Update all the weights according to the Korhonen learning rule;

Step 4: Increment the iteration index, t , by unity and decrease the magnitude of the learning rate, $\hat{\alpha}_0(t)$, accordingly; shrink, then neighborhood, $NC_{ij}(t)$ of the BMU

Step 5: Repeat steps 2–4 until the change in the weight magnitudes is less than the specified threshold or the maximum number of iterations is reached.

In this study, the input layer of SOM neural network were related variables of combustion (temperature and mass fractions of species), and the output layer was a multi-dimensional spatial model (neuron). The more neurons, the more details are represented. The neuron nodes of the input variables and the output variables were trained to generate the neuron with the smallest n -dimensional distance until the clustering result was obtained. Both clustering analyses were performed using MATLAB.

3.3. Back-Propagation Neural Network

Clustering analysis is a reliable classification method. However, if it is applied to a large amount of data, it still requires a great computational cost. Nonetheless, when performing local identification of the flow field, or when only limited data were available in the experiment, clustering analysis could be inaccurate. Previous study [4] has shown that supervised learning can offer accurate CRI after sufficient training. Artificial neural network (ANN) is a machine learning algorithm that simulates the human neural structure, which can be applied to both regression and classification problems. A standard neural network consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations [28]. Back-propagation neural network (BPNN) algorithm is a simple and efficient method to correct the weight of neurons which was firstly developed by Dreyfus in 1973 [29]. The BPNN is a supervised learning algorithm. Its idea is to first calculate the gradient of the objective function to the output value of each neuron recursively through the chain rule, and then use the chain rule to calculate the gradient of the weight parameters on the edge. As shown is Figure 3, the network structure was a fully-connected neural network, which consisted of an input layer, an output layer, and a hidden layer.

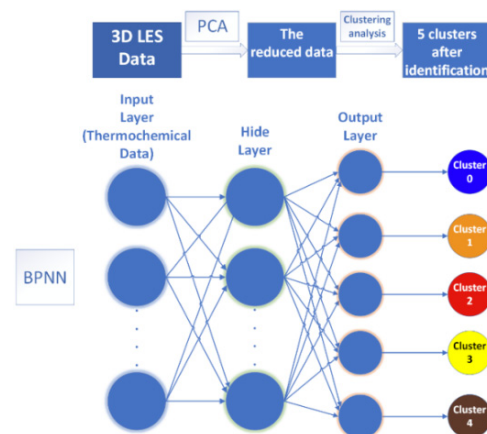


Figure 3. Process of identifying methods and the network structure.

The activation function adopted in this BPNN was the Softmax function to convert the output values of the multi-classification into relative probabilities:

$$\text{Softmax}(y_i) = \frac{e^{y_i}}{\sum_{c=1}^C e^{y_c}}, \quad (3)$$

the y_i is the output value of the i th node, and C is the number of output nodes. Loss function is the cross-entropy loss function:

$$H(y, y^*) = - \sum y \cdot \ln y^*, \quad (4)$$

the y represents the exact probability distribution and y^* represents the probability distribution of the predicted outcome.

In this study, the BPNN was trained by clustering results, and the network gradually “learns” the input/output relationship of interest by adjusting the weights to minimize the error between the actual and predicted output patterns of the training datasets. In specific, we used clustering analysis results of multiple instantaneous thermochemical data from the LESs of the HM1 and the Sandia Flame D as the training set, and then performed the BPNN predictions of two cases using the instantaneous thermochemical data at different times. The training of the BPNN was performed using the TensorFlow Python library. Once trained, returning the identification results from an input was almost instantaneous.

4. CRI with Clustering

The data of this study were the temperature and mass fractions of 49 species of the HM1 and the Sandia Flame D. Through rotated PCA, we retained five PCs as clustering input data. According to the analysis of the Davies-Bouldin index and flame structure, the optimal number of clusters was five, and Table 1 summarizes the five clusters representing five different states in the flames. Cluster 1 described environmental air. Cluster 2 represented the co-flow region, and cluster 3 referred to a state in which reactions were active, and the concentrations of active species, such as OH, were intense. Cluster 4 represented the combustion preheating stage, where temperature was not high, and the concentrations of some typical species, such as CH₂O, were intense, and cluster 5 represented the fuel inflow stream.

Table 1. Regimes represented by different clusters.

Number of Cluster	Regime
Cluster 1	Air (A)
Cluster 2	Co-flow (CO)
Cluster 3	Combustion Zone (CZ)
Cluster 4	Preheat (P)
Cluster 5	Fuel (F)

The analysis of artificial intelligence based on clustering algorithm is a mathematical technology capable of giving the quantitative characteristic indices of thermochemical data [30]. Figure 4 shows an example of scatter plots of temperature and mass fractions of OH and CH₂O in the mixture fraction space in the HM1, and reveals the differences between the five clusters and their physical characteristics. Z is element-based Bilger mixture fraction [25], which is defined as:

$$Z = \frac{\frac{0.5(Y_H - Y_{H,2})}{W_H} + \frac{2(Y_O - Y_{O,2})}{W_C} - \frac{(Y_O - Y_{O,2})}{W_O}}{\frac{0.5(Y_{H,1} - Y_{H,2})}{W_H} + \frac{2(Y_{O,1} - Y_{O,2})}{W_C} - \frac{(Y_{O,1} - Y_{O,2})}{W_O}}, \quad (5)$$

where Y_i is the total mass fraction of element i , and W_i is the atomic mass of element i . The subscripts 1 and 2 refer to values in the fuel and air streams. The calculations of Y_i for elemental carbon, hydrogen, and oxygen should include all species in the mechanism. The characteristics of the environmental air regime included low temperature and low OH and CH₂O concentrations. In contrast, although the temperature and OH concentration in the fuel regime were also low, due to the chemical reactions gradually entering the preheating stage, it already had a certain concentration of CH₂O. The temperature in the co-flow regime was significantly higher than 800 K, but the concentrations of OH and CH₂O were low and

close to the side of the environmental air. The characteristics of the combustion zone were significantly different than others. When the combustion was intense, the temperature was high. Owing to a large amount of OH production and CH₂O consumption, the OH concentration was high and close to the fuel regime, and the CH₂O concentration was high but close to the air regime. In the preheating stage of combustion, because there was not much OH production, although the temperature and CH₂O concentration were not low, the OH concentration was low.

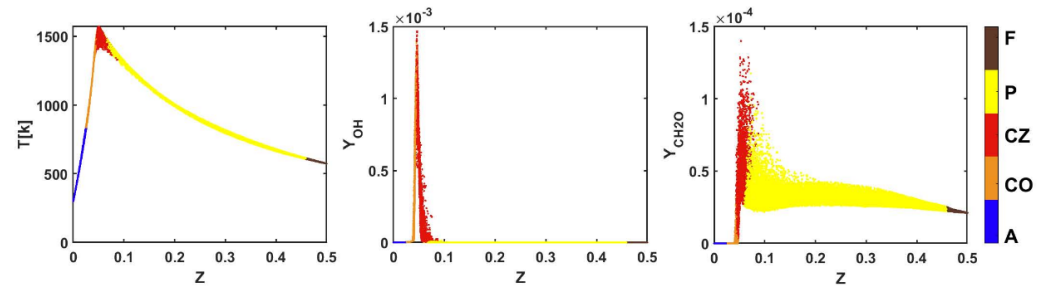


Figure 4. Scatter plot of resolved temperature and mass fractions of OH, CH₂O in the mixture fraction space in the HM1 case at the height of 60 mm. Colored with five clustering classifications.

Figure 5 compares the clustering results obtained by two clustering algorithms (the K-means and the SOM) in the HM1 and the Sandia Flame D. R is the radial distance from the center. The correlation coefficients between the K-means and SOM results were 0.999 and 1.000 for the HM1 and the Sandia Flame D, respectively. This implied that the two clustering algorithms could deliver almost the same clustering results in the two flames, and the classifications embodied in the clustering would objectively represent the different quantitative characteristics in the flame data.

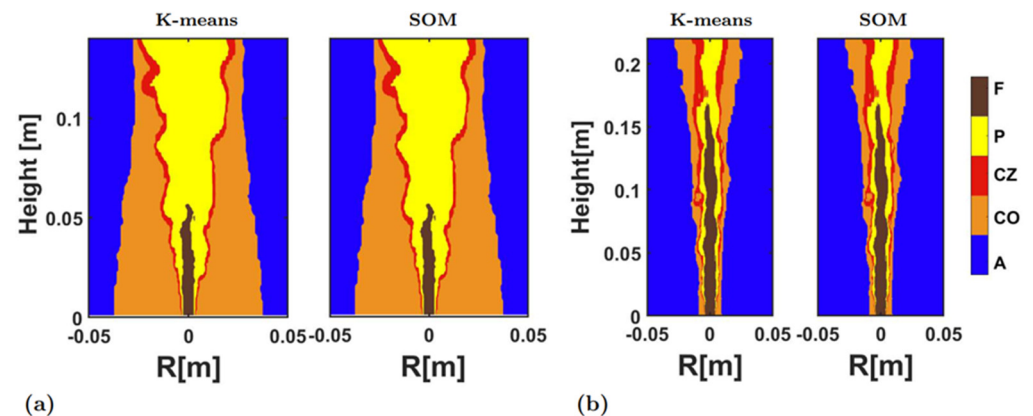


Figure 5. Distributions of two cases divided by two clustering algorithms (a) HM1; (b) Sandia Flame D.

Furthermore, Figures 6 and 7 compare the differences between the results obtained using the K-means clustering and the threshold method. Figure 6 embeds the isolines (20% of its maximum value) of HRR into the clustering results. It is a basic means to characterize the combustion zone with high HRR [3,8,9]. Since it is difficult to measure HRR directly in experiments, many methods [13] characterize flames by predicting HRR, and then measure the geometric parameters, such as flame thickness and wrinkle. It was evident that the combustion zone identified by the clustering algorithm was highly consistent with the results of the HRR threshold method. In addition, it should be noted that the clustering algorithm could not only judge the combustion zone, but also identify other characteristic regimes (such as the preheat zone), and the threshold method must use other variables, such as CH₂O, etc.

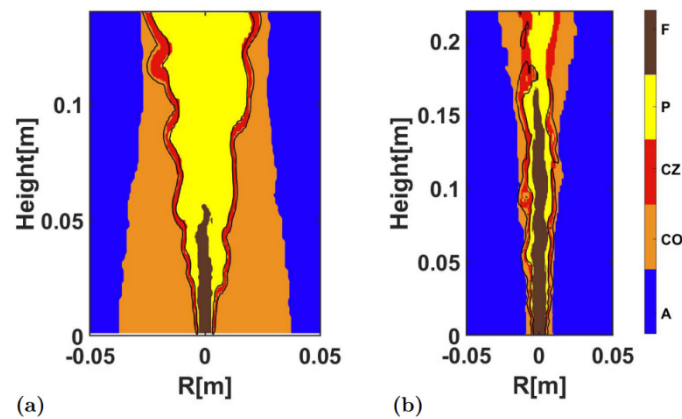


Figure 6. The clustering analysis results are compared with the HRR (20% of the maximum) isolines: (a) HM1; (b) Sandia Flame D.

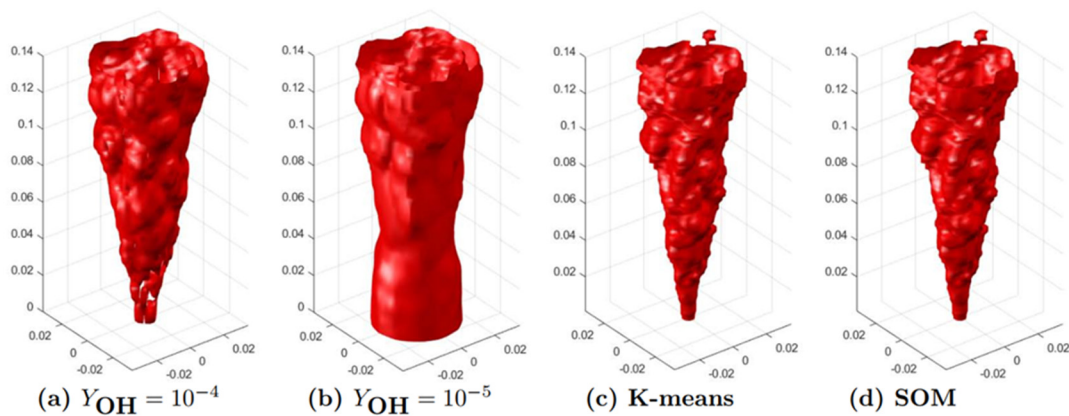


Figure 7. Flame front predicted using OH-threshold ($Y_{OH} = 10^{-4}$, and $Y_{OH} = 10^{-5}$) and two clustering algorithms (K-means and SOM) in the HM1.

As mentioned earlier, the OH-threshold method is often used to represent the combustion zone [13]. However, the flame front would appear differently in shape and thickness when different critical values of OH concentration are used. As shown in Figure 7, the flame front structure in the HM1 case would be significantly different for using $Y_{OH} = 10^{-4}$ and $Y_{OH} = 10^{-5}$. For all threshold methods, including the GFRI method mentioned above, there exists the same problem of threshold selection. In specific, when using the threshold method to determine the combustion zone, if different variables are used, such as HRR or OH concentration, or the threshold value is different, these may affect the results, thus affecting further analysis, such as obtaining the flame lift-off height, flame thickness, etc. Using multidimensional clustering analysis algorithm to carry out CRI obviously would not produce such a problem. The flame front results detected using the K-means and the SOM are identical, indicating that the quantitative characteristic index given by unsupervised machine learning is more objective and reliable.

5. CRI with BPNN Trained by Clustering Results

Combustion regime identification with rotated PCA and clustering has been shown to be robust and accurate [30]. Although PCA can cut down the number of variables and reduce CPU cost for clustering, a large amount of computations is still required to process the data and predict the corresponding combustion regime. Moreover, in experimental measurements or industrial applications, it is usually difficult to obtain multiple scalar fields as the data source for clustering analysis, and clustering analysis with limited data could be fallacious. Therefore, this study attempted to adopt the supervised machine learning to provide an ultra-fast and reliable CRI based on limited thermochemical properties, and

also tried to perform accurate CRI with fewer data, which is investigated in this section to the minimum data required for accurate CRI and the possible required data types.

For the evaluation of BPNN results, we still adopted the method of examining the flame structure, and compared it with the flame structure obtained using the cluster analysis. We also compared the BPNN result with the clustering result by three quantitative methods. Since the K-means CRI result and the SOM CRI result described above are nearly identical, we used the K-means result as the comparison object here. The accuracy rate, which is defined as the ratio of the number of correctly identified samples to the total number of samples, was used to evaluate the identification accuracy of the whole domain. The combustion zone is the focus of many studies. Thus, we can extract the data of combustion zone and binarize the flame into the flow field, and then calculate the correlation coefficient and the F_1 score between the two results. The F_1 score follows: $2/F_1 = (FP + TP)/TP + (FN + TP)/TP$, where comparison results are divided into four categories, true positive (TP, where combustion zone identified using the K-means clustering is identified as combustion zone), false positive (FP, where non-combustion zone identified using the K-means clustering is identified as combustion zone), false negative (FN, where combustion zone identified using the K-means clustering is identified as non-combustion zone) and true negative (TN, where non-combustion zone identified using the K-means clustering is identified as non-combustion zone). Note that TN was not used in the evaluation of the F_1 score.

According to the corresponding scalars represented by the first few major PCs characterized by PCA shown in the previous study [30], and taking into account the chemical species that are usually measured in experiments, we used the five scalars, including temperature and mass fractions of CH_4 , CO , OH , and CH_2O as the input data of the BPNN. Figure 8 shows the BPNN results of CRI using these five scalars. This was very consistent with the previous clustering results shown in Figure 4. In specific, the recognition accuracy in the HM1 was 98.64%, and the recognition accuracy in the Sandia Flame D was 98.88%. In both cases, selecting these five dimensions as the input data for the BPNN could accurately achieve CRI of the flame.

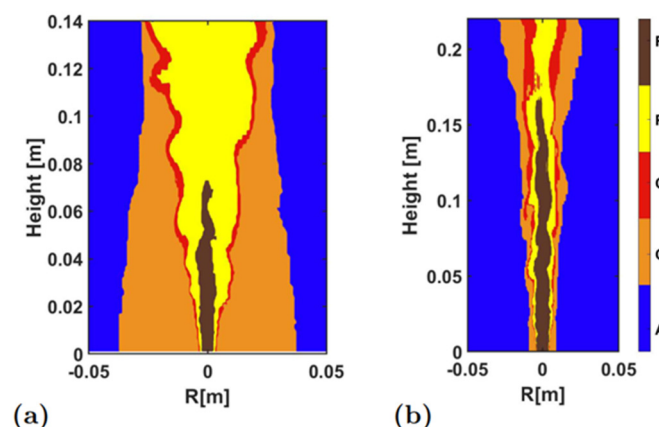


Figure 8. CRI results of BPNN with five-dimensional input: (a) HM1; (b) Sandia Flame D.

Considering the very limited data available for experimental measurements, in order to provide a simple method to determine the combustion regimes, we reduced the number of input variables step by step to see the accuracy of the BPNN results of CRI. It was also a test of the reliability of the BPNN method used in this CRI study and the minimum limit of input required for identification. Of course, reducing the dimension of the input data could also reduce the computational cost during CRI. Figure 9 shows the identification result after removing one of the five variables used in the identification shown in Figure 8, and Figure 10 presents quantitative indicators of identification accuracy. It was clear that, overall, the identification of four variables here were quite reliable. Overall, all identification accuracies were very high, remaining at a level of more than 95%. Except in the HM1, the identification accuracy dropped to almost 80% if the temperature was absent among the

variables, due to the existence of large high-temperature co-flow region. In the Sandia Flame D, however, the pilot/co-flow region was small, so the absence of temperature did not affect the identification very much. In the Sandia Flame D, the most influential factor was the absence of fuel (i.e., CH_4). Since the identified fuel region was large, the absence of fuel reduced the identification accuracy to 90%. As for the identification of the combustion zone, all identifications maintained a high accuracy. The F_1 score and the correlation coefficient were maintained above 70%, and most of them were above 80%.

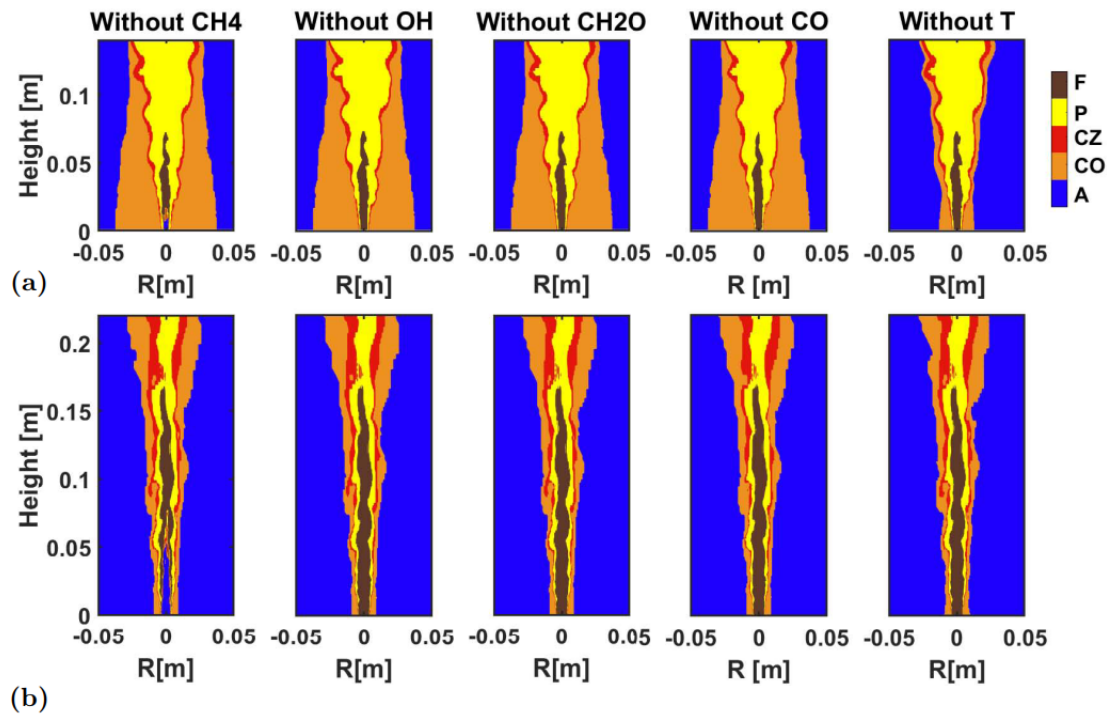


Figure 9. CRI results of BPNN with four-dimensional input: (a) HM1; (b) Sandia Flame D.

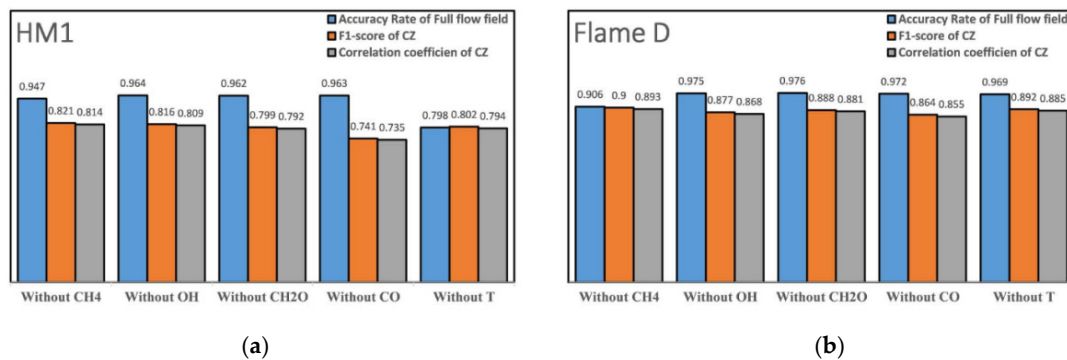


Figure 10. Identification accuracies with four-dimensional BPNN: (a) HM1; (b) Sandia Flame D.

Furthermore, Figure 11 shows the identification result using three variables, and Figure 12 presents quantitative indicators of identification accuracy. Note that the identification accuracy in the HM1 was reduced to 85% without temperature, and, thus, the results without temperature input are not shown here. Again, the overall identification was still good. Obviously, the 3-dimensional CRI results in the presence of temperature were basically above 95%; except in the Sandia Flame D, where the absence of fuel made the identification accuracy lower, which was 91.6%. For the identification of the combustion zone, the F_1 value and the correlation coefficient were both higher than 0.7 for all tests. Especially the combinations of CH_2O - OH - T and CH_4 - CO - T could better capture the combustion zone.

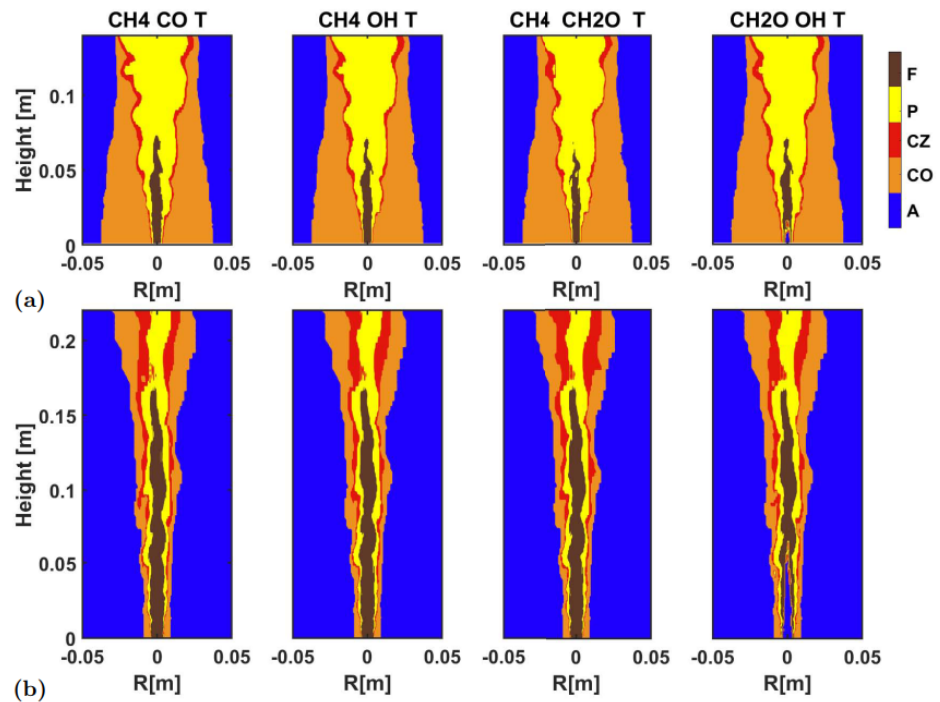


Figure 11. CRI results of BPNN with three-dimensional input: (a) HM1; (b) Sandia Flame D.

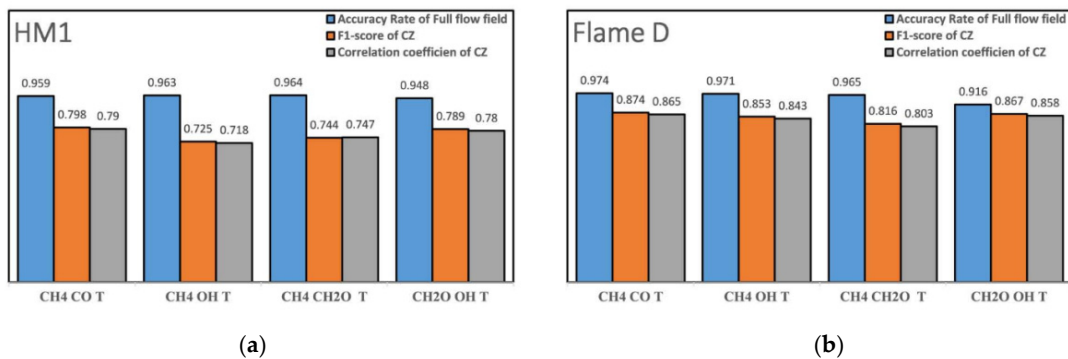


Figure 12. Identification accuracies with three-dimensional BPNN: (a) HM1; (b) Sandia Flame D.

At last, Figure 13 shows the identification result using two variables, and Figure 14 presents quantitative indicators of identification accuracy. The overall identification was acceptable, except for the identification based on temperature and OH profiles. In the HM1, the combinations of CH₄-T and CH₂O-T yielded good overall identification accuracy, while for the combustion zone, the combinations of CH₄-CO and CH₂O-OH yielded better accuracy. In the Sandia Flame D, the combinations of CH₄-T, CH₄-CO and CH₂O-T yielded good overall identification accuracy, while for the combustion zone, the combinations of CH₄-CO and CH₂O-T yielded better accuracy. Overall, the CH₄-T combination yielded the best identification accuracy, both for the entire flame and the combustion zone. These two thermochemical quantities were also a group of scalars that could be relatively easily measured in real time in experiments.

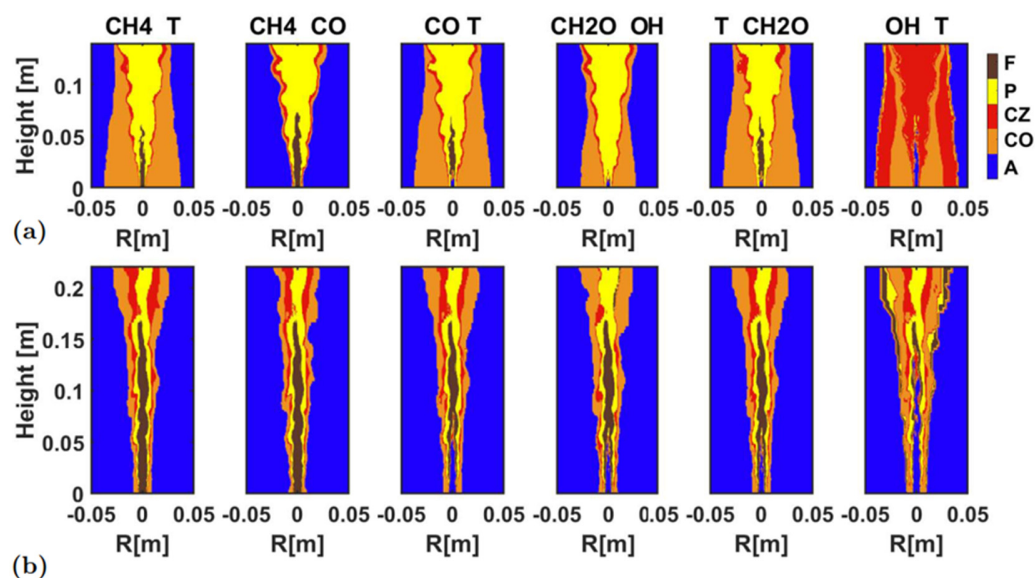


Figure 13. CRI results of BPNN with two-dimensional input: (a) HM1; (b) Sandia Flame D.

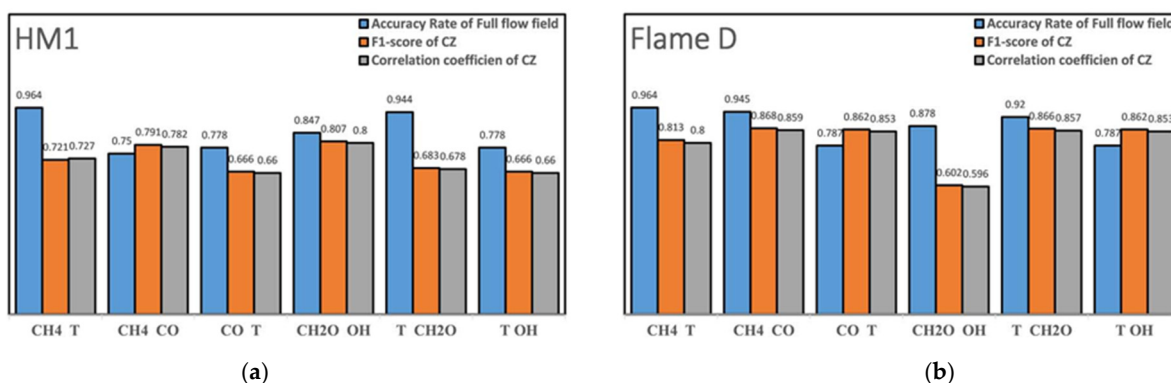


Figure 14. Identification accuracies with two-dimensional BPNN: (a) HM1; (b) Sandia Flame D.

6. Conclusions

A BPNN was trained to identify combustion regimes in two well-known turbulent non-premixed flames. The LES data was first processed with rotated PCA, and clustering analysis to generate a pixel-wise CRI database.

The application of rotated PCA can reduce 50-dimensional thermochemical data (including temperature and mass fractions of 49 species) to 5-dimensional input data, while retaining almost 95% of the information. This enables subsequent analysis to greatly reduce computation cost. Furthermore, the clustering analyses showed that these 5-dimensional data can be divided into 5 clusters, and, according to the thermochemical characteristics they represent, they are distinguished as: environmental air region, co-flow region, combustion zone, preheat zone and fuel stream. The results showed that these five regimes were well detected by the machine learning method with an accuracy of more than 98% using 5 scalars as input dataset.

Compared to other CRI approaches, such as the flame index method, the GFRI method etc., the proposed method avoids using any artificially determined thresholds for identification and does not require any scalar gradients. The computational cost of the method is very small. However, the CRI accuracy is quite satisfactory, for instance, even using the combined data of CH₄-T. The BPNN identification can achieve an accuracy of more than 95% for the entire flame. The method is a practical method to identify the combustion regime for industrial applications, and a support for further analysis of the characteristics in turbulent flames.

Author Contributions: Conceptualization, H.L. and X.Q.; Methodology, F.X. and T.M.; Project administration, H.L.; Software, H.Z. and X.Q.; Validation, H.Z.; Writing—original draft, H.Z.; Writing—review & editing, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (51776082). Computing resources were provided by the National Supercomputer Center in Guangzhou.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Poinso, T.; Veynante, D. *Theoretical and Numerical Combustion*; Edwards: Burgess Hill, UK, 2005.
2. Hartl, S.; Geyer, D.; Dreizler, A.; Magnotti, G.; Barlow, R.S.; Hasse, C. Regime identification from Raman/Rayleigh line measurements in partially premixed flames. *Combust. Flame* **2018**, *189*, 126–141. [[CrossRef](#)]
3. Doan, N.A.; Swaminathan, N. Analysis of Markers for Combustion Mode and Heat Release in MILD Combustion Using DNS Data. *Combust. Sci. Technol.* **2019**, *191*, 1059–1078. [[CrossRef](#)]
4. Wan, K.; Hartl, S.; Vervisch, L.; Domingo, P.; Barlow, R.S.; Hasse, C. Combustion regime identification from machine learning trained by Raman/Rayleigh line measurements. *Combust. Flame* **2020**, *219*, 268–274. [[CrossRef](#)]
5. Zirwes, T.; Zhang, F.; Habisreuther, P.; Hansinger, M.; Bockhorn, H.; Pfitzner, M.; Trimis, D. Identification of Flame Regimes in Partially Premixed Combustion from a Quasi-DNS Dataset. *Flow Turbul. Combust.* **2021**, *106*, 373–404. [[CrossRef](#)]
6. Yamashita, H.; Shimada, M.; Takeno, T. A numerical study on flame stability at the transition point of jet diffusion flames. *Symp. (Int.) Combust.* **1996**, *26*, 27–34. [[CrossRef](#)]
7. Butz, D.; Hartl, S.; Popp, S.; Walther, S.; Barlow, R.S.; Hasse, C.; Dreizler, A.; Geyer, D. Local flame structure analysis in turbulent CH₄/air flames with multi-regime characteristics. *Combust. Flame* **2019**, *210*, 426–438. [[CrossRef](#)]
8. Mohammadnejad, S.; Vena, P.; Yun, S.; Kheirkhah, S. Internal structure of hydrogen-enriched methane—Air turbulent premixed flames: Flamelet and non-flamelet behavior. *Combust. Flame* **2019**, *208*, 139–157. [[CrossRef](#)]
9. Böckle, S.; Kazenwadel, J.; Kunzelmann, T.; Shin, D.I.; Schulz, C.; Wolfrum, J. Simultaneous single-shot laser-based imaging of Formaldehyde, OH, and temperature in turbulent flames. *Proc. Combust. Inst.* **2000**, *28*, 279–286. [[CrossRef](#)]
10. Kerkemeier, S.G. Direct Numerical Simulation of Combustion on Petascale Platforms: Application to Turbulent Non-Premixed Hydrogen Autoignition. Ph.D. Thesis, ETH Zürich, Zürich, Switzerland, 2010.
11. Jigjid, K.; Tamaoki, C.; Minamoto, Y.; Nakazawa, R.; Inoue, N.; Tanahashi, M. Data driven analysis and prediction of MILD combustion mode. *Combust. Flame* **2021**, *223*, 474–485. [[CrossRef](#)]
12. Barwey, S.; Prakash, S.; Hassanaly, M.; Raman, V. Data-driven classification and modeling of combustion regimes in detonation waves. *Flow Turbul. Combust.* **2021**, *106*, 1065–1089. [[CrossRef](#)]
13. Dave, H.; Swaminathan, N.; Parente, A. Interpretation and characterization of MILD combustion data using unsupervised clustering informed by physics-based, domain expertise. *Combust. Flame* **2022**, *240*, 111954. [[CrossRef](#)]
14. Chen, W. Large Eddy Simulation of Sandia Flame D and F Based on Nonlinear Subgrid Model. Master's Thesis, Huazhong University of Science and Technology, Wuhan, China, 2018.
15. Lu, H.; Chen, W.; Zou, C.; Yao, H. Large-eddy simulation of Sandia Flame F using structural subgrid-scale models and partially-stirred-reactor approach. *Phys. Fluids* **2019**, *31*, 045109. [[CrossRef](#)]
16. Lu, H.; Zou, C.; Shao, S.; Yao, H. Large-eddy simulation of MILD combustion using partially stirred reactor approach. *Proc. Combust. Inst.* **2019**, *37*, 4507–4518. [[CrossRef](#)]
17. Qian, X.; Lu, H.; Zou, C.; Zhang, H.; Shao, S.; Yao, H. Numerical investigation of the effects of turbulence on the ignition process in a turbulent MILD flame. *Acta Mech. Sin.* **2021**, *37*, 1299–1317. [[CrossRef](#)]
18. Dally, B.; Karpetis, A.; Barlow, R. Structure of turbulent non-premixed jet flames in a diluted hot coflow. *Proc. Combust. Inst.* **2002**, *29*, 1147–1154. [[CrossRef](#)]
19. Barlow, R.S.; Frank, J.H. Effects of turbulence on species mass fractions in methane/air jet flames. *Symp. (Int.) Combust.* **1998**, *27*, 1087–1095. [[CrossRef](#)]
20. Bowman, C.; Hanson, R.; Davidson, D.; Lissianski, W.G.V., Jr.; Smith, G.; Golden, D.; Frenklach, M.; Goldenberg, M. GRI-Mech 2.11. Berkeley, 1995. Available online: [http://www.me.berkeley.edu/gri%backslash\\$mech/](http://www.me.berkeley.edu/gri%backslash$mech/) (accessed on 16 August 2022).
21. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [[CrossRef](#)]
22. Jolliffe, I.T. Principal components in regression analysis. In *Principal Component Analysis*; Springer: New York, NY, USA, 1986; pp. 129–155.
23. Richman, M.B. Rotation of principal components. *J. Climatol.* **1986**, *6*, 293–335. [[CrossRef](#)]
24. Jolliffe, I.T. Rotation of principal components: Some comments. *J. Climatol.* **1987**, *7*, 507–510. [[CrossRef](#)]
25. Maulik, U.; Bandyopadhyay, S. Genetic algorithm based clustering technique. *Pattern Recognit. Lett.* **2000**, *33*, 455–465. [[CrossRef](#)]

26. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [[CrossRef](#)]
27. Chen, S.K.; Mangimeli, P.; West, D. The comparative ability of Self-organizing neural networks to define cluster structure. *Omega Int. J. Manag. Sci.* **1995**, *23*, 271–279. [[CrossRef](#)]
28. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)]
29. Dreyfus, S. The computational solution of optimal control problems with time lag. *IEEE Trans. Autom. Control.* **1973**, *18*, 383–385. [[CrossRef](#)]
30. Ma, T.-S.; Lu, H.; Li, D.-G.; Zhang, H.-L.; Qian, X. Reaction zone identification in MILD combustion using multidimensional cluster analysis. *Chem. Eng. Des. Commun.* **2021**, *47*, 135–137.