



**HAL**  
open science

# Machine learning for integrating combustion chemistry in numerical simulations

Huu-Tri Nguyen, Pascale Domingo, Luc Vervisch, Phuc-Danh Nguyen

## ► To cite this version:

Huu-Tri Nguyen, Pascale Domingo, Luc Vervisch, Phuc-Danh Nguyen. Machine learning for integrating combustion chemistry in numerical simulations. *Energy and AI*, 2021, 5, pp.100082. 10.1016/j.egyai.2021.100082 . hal-03372600

**HAL Id: hal-03372600**

**<https://normandie-univ.hal.science/hal-03372600>**

Submitted on 24 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Machine learning for integrating combustion chemistry in numerical simulations

Huu-Tri Nguyen<sup>a,b</sup>, Pascale Domingo<sup>a</sup>, Luc Vervisch<sup>a,\*</sup>, Phuc-Danh Nguyen<sup>b</sup>

<sup>a</sup>*CORIA – CNRS, Normandie Université, INSA Rouen Normandie, 76801 Saint-Etienne-du-Rouvray, France*

<sup>b</sup>*ArcelorMittal Global Research and Development, 57280 Maizières-lès-Metz, France*

---

## Abstract

A strategy based on machine learning is discussed to close the gap between the detailed description of combustion chemistry and the numerical simulation of combustion systems. Indeed, the partial differential equations describing chemical kinetics are stiff and involve many degrees of freedom, making their solving in three-dimensional unsteady simulations very challenging. It is discussed in this work how a reduction of the computing cost by an order of magnitude can be achieved using a set of neural networks trained for solving chemistry. The thermochemical database used for training is composed of time evolutions of stochastic particles carrying chemical species mass fractions and temperature according to a turbulent micro-mixing problem coupled with complex chemistry. The novelty of the work lies in the decomposition of the thermochemical hyperspace into clusters to facilitate the training of neural networks. This decomposition is performed with the Kmeans algorithm, a local principal component analysis is then applied to every cluster. This new methodology for combustion chemistry reduction is tested under conditions representative of a non-premixed syngas oxy-flame.

*Keywords:* Combustion chemistry, Micro-mixing modeling, Principal component analysis, Artificial neural network, Chemistry reduction

---

## 1. Introduction

2 Despite of the continuous progress in supercomputing methodologies, all the degrees of free-  
3 dom required to fully describe the detail of combustion chemistry cannot be introduced in the nu-

---

\*Corresponding author

Email address: [luc.vervisch@insa-rouen.fr](mailto:luc.vervisch@insa-rouen.fr) (Luc Vervisch)

4 merical simulations of large-scale combustion systems, such as boilers and furnaces. Several meth-  
5 ods have been proposed in the literature to reduce and optimize large chemical kinetic schemes,  
6 such as Quasi-Steady State Approximation (QSSA) and partial equilibrium [1], Directed Rela-  
7 tion Graph (DRG) [2], Directed Relation Graph with error propagation [3], DRG-aided sensitivity  
8 analysis (DRGASA) [4] or again unimportant reaction elimination [5] (not exhaustive list).

9 Even after reducing the number of differential equations solved, because of the very stiff char-  
10 acter of differential systems associated to combustion chemistry, the CPU time devoted to their  
11 integration remains quite significant and in most cases, too large to allow for performing the mul-  
12 tiple simulations needed to complete the quest of the optimal design for a combustion system.

13 These computing limitations motivated the introduction of artificial neural networks (ANN) to  
14 deal with chemistry reduction and its time integration [6] and also to develop data driven turbulent  
15 combustion modeling [7], or to analyse experimental measurements [8] and to setup digital-twins  
16 for process control [9].

17 Among these works, ANN and convolutional neural networks (CNN) were trained on a repre-  
18 sentative dataset to replace chemistry integration [10] and the method was applied to perform direct  
19 numerical simulation (DNS) of a syngas oxy-flame. Nevertheless, the training of the full dataset  
20 using a single or two ANNs required specific treatments, such as data augmentation through man-  
21 ufactured intermediate solution points for dealing with some fast reacting species [10]. These  
22 operations requiring fine tuning by the user were found necessary to secure good accuracy over  
23 the whole domain covering from chemically frozen-flow mixing up to the equilibrium state and  
24 pollutant emissions, with fast ignition and fuel oxidation in between.

25 To avoid relying on these additional more or less ad-hoc operations, thus allowing for having  
26 a computerised and fully automatic procedure, a novel method is proposed in the present study  
27 for turbulent non-premixed flames in which techniques such as data clustering with K-means [11]  
28 and dimension reduction using LPCA [12], are combined with ANN training to secure, by con-  
29 struction, accuracy over the entire composition space domain (mixing, ignition, combustion and  
30 equilibrium state).

31 The paper is organised as follow, in the subsequent section the operating conditions considered  
32 and the novel methodology proposed are discussed. After that, the obtained ANN is tested in a

33 posteriori manner for chemistry integration, to study the robustness of the method and the CPU  
34 time reduction.

## 35 **2. Chemistry integration with machine learning**

### 36 *2.1. Thermochemical conditions*

37 The operating condition considered in this study is taken from previous works on syngas, with  
38 a chemical composition representative of those found in the steel industry in recycled exhaust  
39 gases [13]. This non-premixed syngas oxy-flame operates under a pressure of 341.3kPa, with the  
40 two inlets: (i) syngas fuel injected at 1223K for a mass flow rate of 0.74kg/s and the composition  
41 in mass fraction H<sub>2</sub>: 0.0085, CO: 0.7852, CO<sub>2</sub>: 0.0514, N<sub>2</sub>: 0.1549; (ii) pure oxygen at 298K with  
42 a mass flow rate of 0.69kg/s.

### 43 *2.2. Database generation*

44 A database of chemical evolutions is first generated by solving for the mass fractions and the  
45 enthalpy of a set stochastic particles featuring at initial condition the concentration in the fuel or  
46 in the oxidizer inlet. Then, the species mass fraction  $Y_i^p(t)$  and sensible enthalpy  $h_s^p(t)$  of each  
47 stochastic particle evolve according to the following equations:

$$\frac{dY_i^p(t)}{dt} = \text{MIX}_i^p(\tau_T) + \dot{\omega}_i^p, \quad (1)$$

$$\frac{dh_s^p(t)}{dt} = \text{MIX}_{h_s}^p(\tau_T) + \dot{\omega}_{h_s}^p, \quad (2)$$

48 where ‘MIX’ denotes the stochastic turbulent micro-mixing closure for the diffusive budget in  
49 a non-premixed system.  $\tau_T$  denotes the micro-mixing time and  $\dot{\omega}_i^p$  and  $\dot{\omega}_{h_s}^p$  are the species and  
50 enthalpy chemical sources, respectively. In the present study, the Euclidean minimum spanning  
51 tree (EMST) micro-mixing model [14] is used to model the turbulent micro-mixing.

52 This approach to generate thermochemical conditions representative of those observed in a real  
53 system for studying its chemical response, without explicitly solving for the flow, has been shown  
54 effective for partial oxidation of natural gas [15], aircraft engines combustion chambers [16], large-

55 scale furnaces with urea DeNO<sub>x</sub> automated control [17], micro- and meso-scale combustion sys-  
56 tems [18] and also to build the database for machine learning application to sooting flames [19].

57 Here a total number of 1429 stochastic particles is employed with 740 particles (corresponding  
58 to the mass flow rate of 0.74kg/s) assigned to Fuel inlet and 689 particles (corresponding to the  
59 mass flow rate of 0.69kg/s) to O<sub>2</sub> inlet.

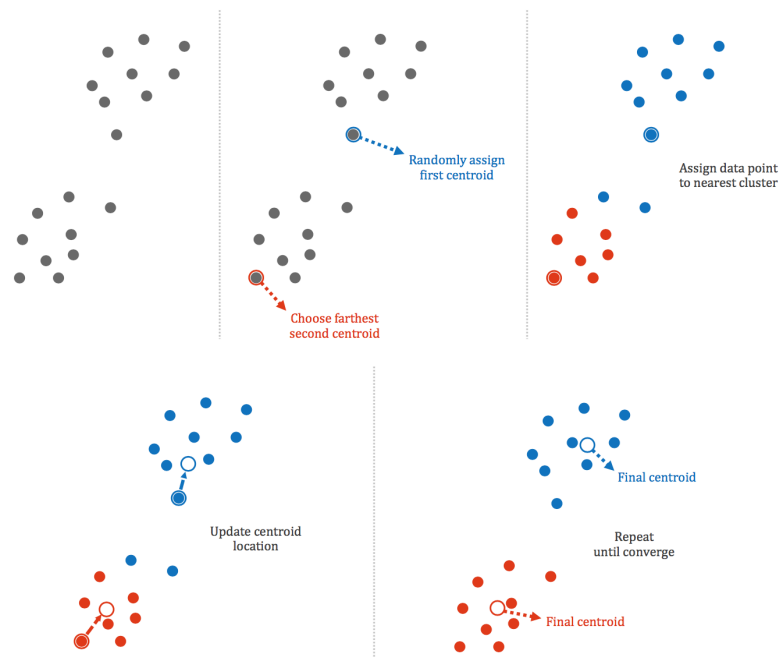


Figure 1: Clustering with K-means

60 From the evolution of these particles, the first step consists of identifying, from a detailed  
61 chemical scheme, the important species to be retained and whose evolutions shall be learned by the  
62 ANNs. This is done using the direct relation graph with error propagation method (DRGEP) [3].  
63 For the present operating conditions, this step was previously performed and reported in [10].  
64 Starting from the detailed scheme GRI3.0 [20], 11 species were identified as essential to follow  
65 the dynamics of the syngas chemical evolutions: O, O<sub>2</sub>, H, H<sub>2</sub>, OH, HO<sub>2</sub>, H<sub>2</sub>O, H<sub>2</sub>O<sub>2</sub>, CO, CO<sub>2</sub>  
66 and N<sub>2</sub>. A companion reduced chemistry involving 23 reactions was then optimised (see Table 2  
67 of [13]).

68 To save CPU time during training and to demonstrate the integration of chemistry with ANN,

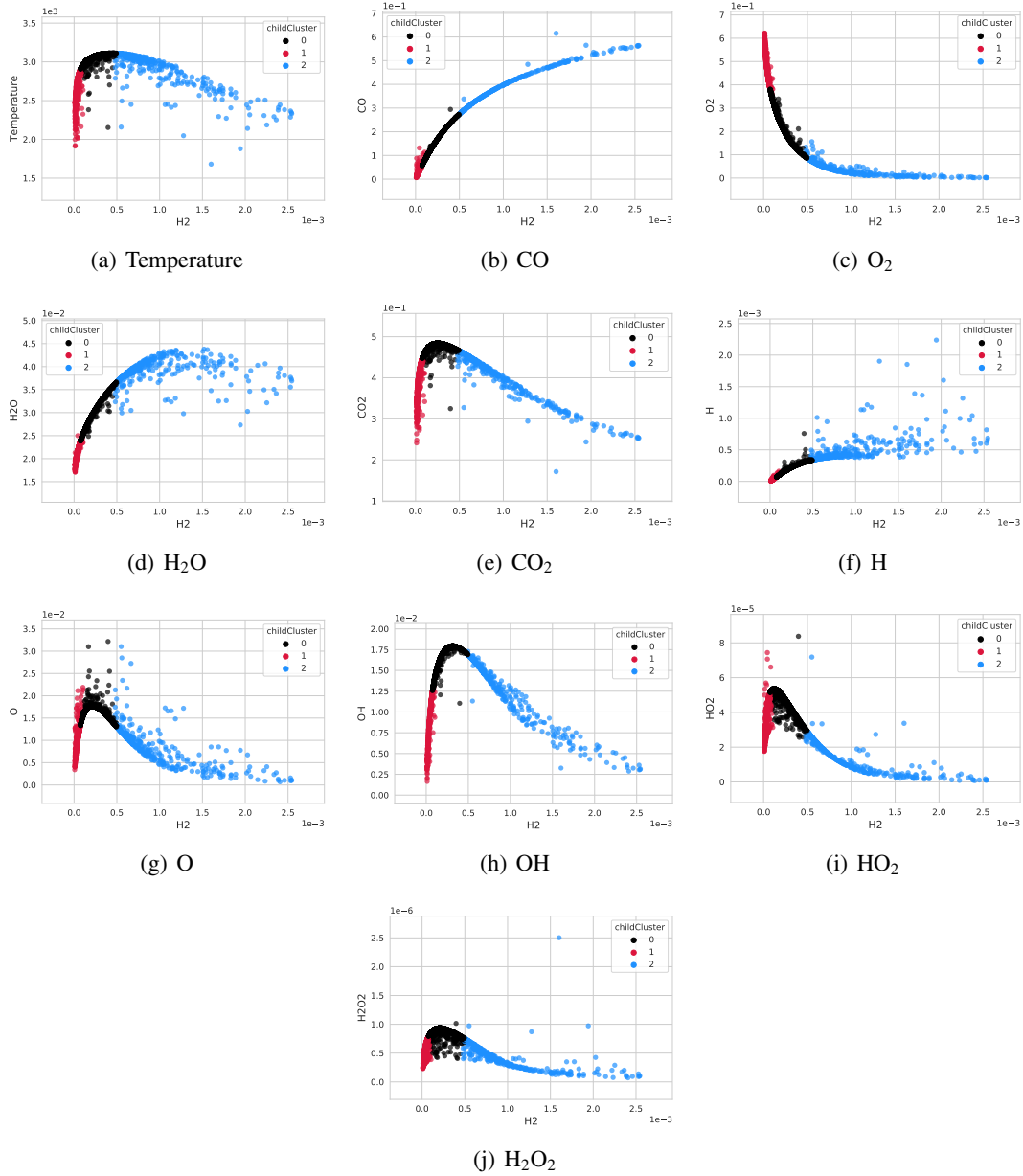


Figure 2: Data clustering for parentCluster1. Temperature [K] (a) and species mass fraction [-] (b)-(j) vs  $H_2$  mass fraction. Red: childCluster1\_0. Black: childCluster1\_1. Blue: childCluster1\_2.

69 this 11-species and 23 elementary reactions scheme is used to build the time evolution dataset of  
 70 all stochastic particles, over 4999 time-steps of  $0.3\mu s$  with a mixing time  $\tau_T = 0.3ms$  (a repre-  
 71 sentative value for a characteristic turbulent mixing-time as observed in shear layers for Reynolds

72 numbers typical of burners).  $N_2$  was removed from the database for the pre-treatment and ANN  
73 learning steps. The full database thus consists of the 11 feature columns (10 species and temper-  
74 ature) of the 1429 stochastic particles collected over 4999 time-steps, therefore the raw data is  
75 composed of  $1429 \times 4999 = 7143571$  vectors of 10 components.

### 76 2.3. *Data pre-treatment*

77 To secure an appropriate analysis of the multivariate nature of the problem, the thermochem-  
78 ical variables obtained from the solving of Eqs. 1 and 2 are going through pre-treatments, such  
79 as centering and rescaling and unsupervised machine learning, like clustering with K-means [11],  
80 and dimension reduction with LPCA [21], before the ANN training is to be applied. This strategy  
81 is expected to close the gap between the previous works [10], in which artificial data augmenta-  
82 tion was required to secure accuracy, and the introduction of complex chemistry in the numerical  
83 simulation of large scale furnaces.

84 Following the recent study by D'Alessio et al. [22] in the context of adaptive chemistry for  
85 reacting flow simulation, the original dataset is first pre-processed with centering and auto-scaling  
86 (standardizing) to insure reliable and robust results.

#### 87 2.3.1. *Data clustering*

88 K-means [11] is a widely used and well-established unsupervised machine learning algorithm,  
89 which consists in grouping similar data into different clusters. K-means is particularly suited for  
90 the classification of the micro-mixing stochastic particles, which evolve through a large variety of  
91 stages from chemically-frozen flow mixing, rapid ignition, fuel oxidation and combustion up to  
92 the chemical equilibrium state. Moreover, the particles from Fuel and  $O_2$  inlets feature different  
93 time histories, which benefit from being classified in separate clusters for chemistry ANN integra-  
94 tion. After clustering into separate groups, every individual subset of data has similar and regular  
95 distribution to ensure effectiveness of the subsequent operations.

96 The different steps in K-means used for the thermochemical quantities (species mass fractions  
97 and temperature) are illustrated in Figure 1 and summarized as follows:

- 98 • Definition of the number of clusters.

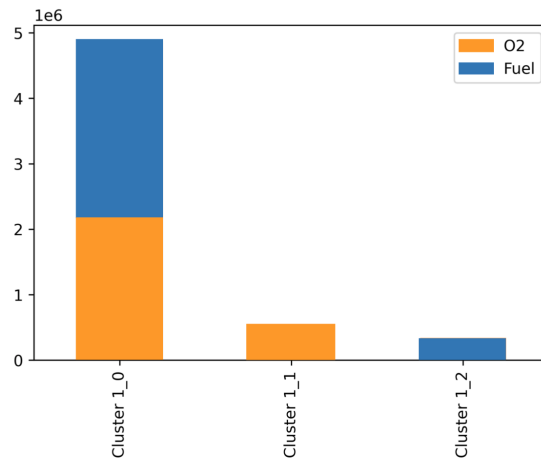


Figure 3: Repartition of data points in childCluster1\_0, childCluster1\_1 and childCluster1\_2. Abscissa is cluster numbering, ordinate is number of data points. Orange: particles from O<sub>2</sub> inlet. Blue: particles from Fuel inlet.

- 99      • Selecting randomly an initial cluster centroid: To ensure the quality of clustering and im-
- 100      prove the convergence speed, the K-means++ initialization algorithm [23] is applied. The
- 101      algorithm assigns randomly a first centroid, a second centroid is then chosen as farthest
- 102      as possible to the first centroid, a third centroid is as farthest as possible to the first two
- 103      centroids, and so on.
- 104      • Assigning data point to cluster with shortest Euclidean distance.
- 105      • Updating centroid locations by averaging all data points in each cluster.
- 106      • Repeat the process until the centroid locations are converged by evaluating the shifting tol-
- 107      erance, which is chosen at  $1e-20$ .

108      Defining an adequate number of clusters is crucial. The usual K-means++ initialization algo-

109      rithm may lead to poor clustering in our case, where the majority of clusters concentrate in the

110      ignition and combustion stages. To overcome this difficulty, an additional hierarchical clustering

111      strategy was adopted; the first clustering leads to only 2 clusters (named parentCluster0 and parent-

112      Cluster1), which separates the low and high temperature zones. The first parent containing the data

113      of mixing, ignition and combustion stages is then clustered into 14 child clusters (childCluster0\_0

114      to childCluster0\_13). The subset of data in the second parent, representing the end of combustion



115 followed by the equilibrium state, requires a special treatment to guarantee the accuracy of ANN  
116 training and prediction. For this reason, the second parent is classified into only 3 child clusters  
117 (childCluster1\_0, childCluster1\_1 and childCluster1\_2), and then these 3 child clusters will have,  
118 respectively, 20, 10, 10 grandchildren (grandChild1\_0\_0 to grandChild1\_0\_19, grandChild1\_1\_0 to  
119 grandChild1\_1\_9, grandChild1\_2\_0 to grandChild1\_2\_9). Each child cluster of the first parent and  
120 each grandchild cluster of the second parent benefits from its own ANN training.

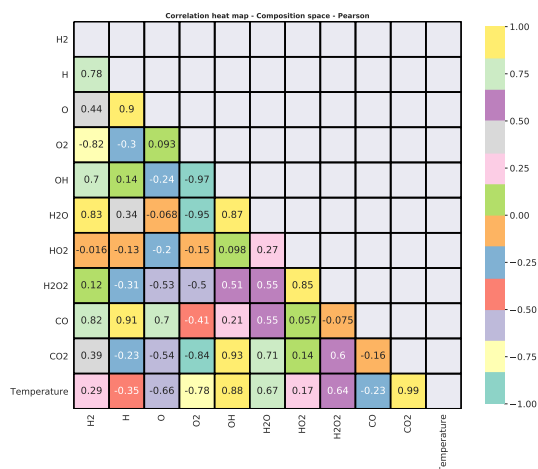
121 The effectiveness of K-means algorithm applied to the flame turbulence/chemistry interaction  
122 (Eqs. 1 and 2) is illustrated in figures 2 and 3. Figure 2 shows how K-means automatically classifies  
123 data in parentCluster1 (representing the end of combustion and the equilibrium state) into 3 distinct  
124 subsets, i.e. childCluster1\_0, childCluster1\_1 and childCluster1\_2. The childCluster1\_1 contains  
125 the particles from O<sub>2</sub> inlet, while the childCluster1\_2 gathers the particles from Fuel inlet (Figure  
126 3); both childClusters represent the end of combustion. The particles from O<sub>2</sub> and Fuel inlets  
127 finally approach the equilibrium state and their data points are classified in a single common cluster  
128 childCluster1\_0.

### 129 2.3.2. *Local principal component analysis (LPCA)*

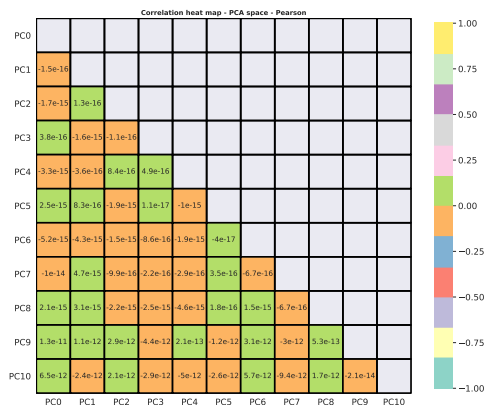
130 Principal component analysis (PCA), also known as a dimensional reduction technique, is  
131 used to reduce a large number of correlated variables (species mass fractions and temperature,  
132 in our case) to a smaller number of uncorrelated variables (principal components) [24]. Because  
133 of the strongly non-linear character of combustion chemistry, it is usually recommended to adopt  
134 a locally linear approach called ‘local PCA’ (LPCA) [21], to properly describe such dynamical  
135 system.

136 In our case, each cluster represents a piece-wise variation of species mass fractions and tem-  
137 perature during a specific combustion stage. The application of LPCA in each cluster transforms  
138 correlated variables of high variance into the most important uncorrelated principal components  
139 (PCs), which will be carefully addressed during the ANN training process. The choice of the  
140 number of PCs is essential to secure a good quality of ANN prediction, and a quite high number  
141 of PCs is usually required dealing with chemistry in reacting flows [22].

142 In practice, it was observed that reducing from 11 PCs to 6-8 PCs would still allow for keeping



(a) Composition space



(b) PCA space

Figure 4: Heat map showing correlation coefficients of data in grandChildCluster1.1.5. (a) Original variables in composition space. (b) PCs in PCA space.

143 more than 98% of the relevant information. However, here the full set of control variable is trans-  
 144 formed into PC (i.e., 10 chemical species and temperature). Therefore, this does not reduce the  
 145 size of the problem to be learned, but the parametrisation of the data from the PCs considerably  
 146 eases the ANN learning process. Indeed, species and temperature are usually well-correlated, the  
 147 PCA analysis transposes these variables to non-correlated PCs and the performances of ANNs  
 148 then become much greater with these non-correlated inputs.

149 Two heat maps are presented in Figure 4 to show the correlation of original variables in com-

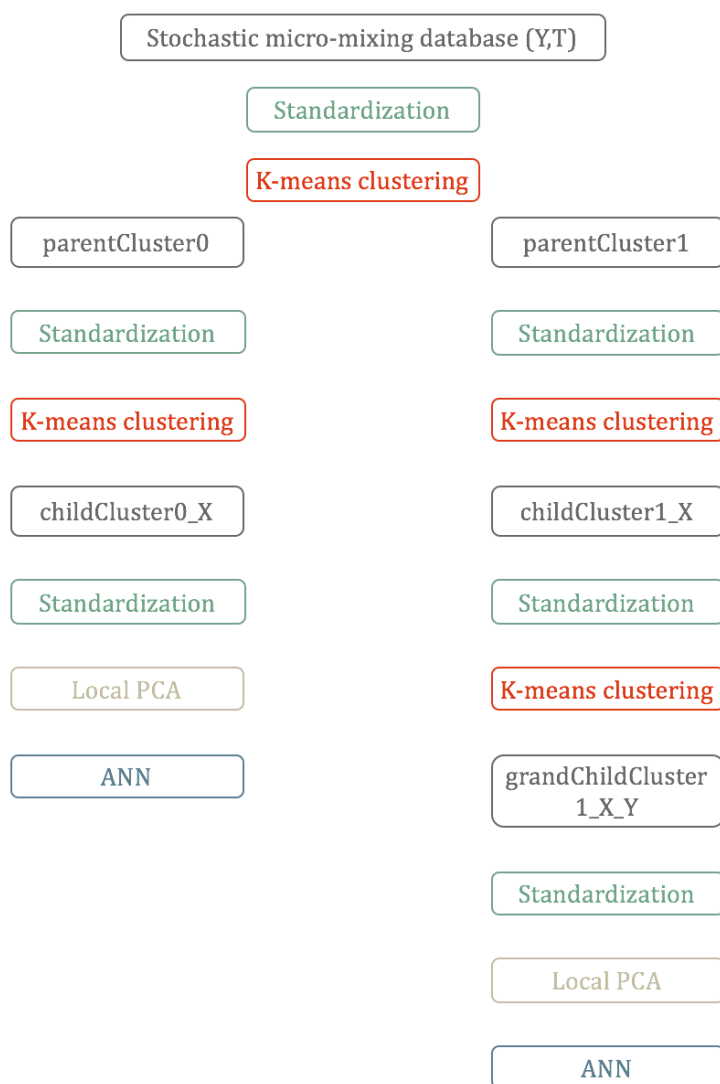


Figure 5: Chemistry reduction with trained ANN.

150 position space and of PCs in PCA space. These heat maps display the Pearson correlation coef-  
 151 ficients, negative (respectively positive) value means that two variables vary in opposite (respec-  
 152 tively same) directions, zero shows the absence of correlation. It is confirmed that the original  
 153 variables in composition space are strongly correlated (Figure 4.a) and that PCs in PCA space are  
 154 clearly uncorrelated, with practically zero correlation coefficients (Figure 4.b).

Layer	Activation function	Output shape
Input	-	(None,11)
Dense	ReLU	(None,512)
Dense	ReLU	(None,256)
Dense	ReLU	(None,128)
Dense	ReLU	(None,64)
Dense (Output)	-	(None,11)

Table 1: Structure of ANNs [10, 13]

155 *2.4. Artificial neural network*

156 The ANN regression structure (Table 1) used in this study is adopted from our previous  
 157 works [10]. The ANN consists of 1 input layer, 4 dense hidden layers with the rectified linear  
 158 activation function (ReLU) and 1 output layer. The total number of parameters (weights and bi-  
 159 ases) is 179851.

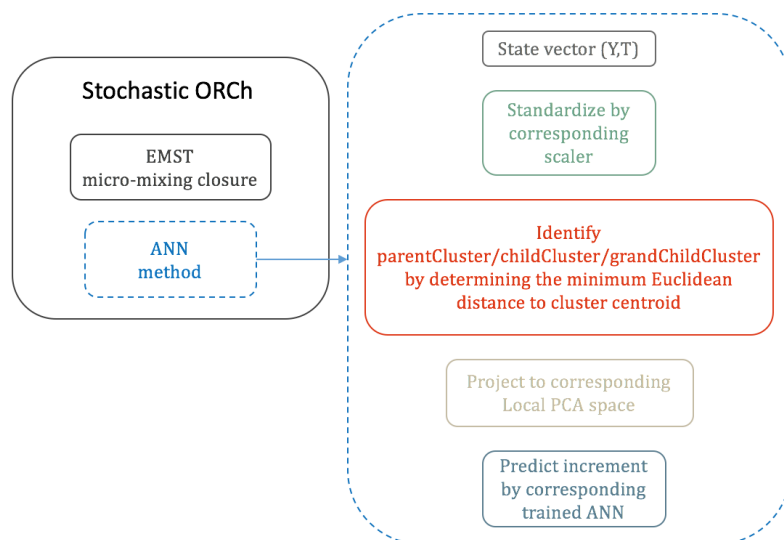


Figure 6: Flowchart of the a posteriori test. ('ORCh': Optimised and Reduced Chemistry.)

160 The ANN input is fed by the data in PCA space, composed of 11 PCs which are obtained by  
 161 the projection of the 11 original variables from composition space into PCA space. The target is  
 162 the increment of species composition and temperature (i.e., the ANN learns how to integrate the  
 163 time evolution of chemistry). The database is split into training set (81%), validation set (9%) and

164 test set (10%) for model performance evaluation.

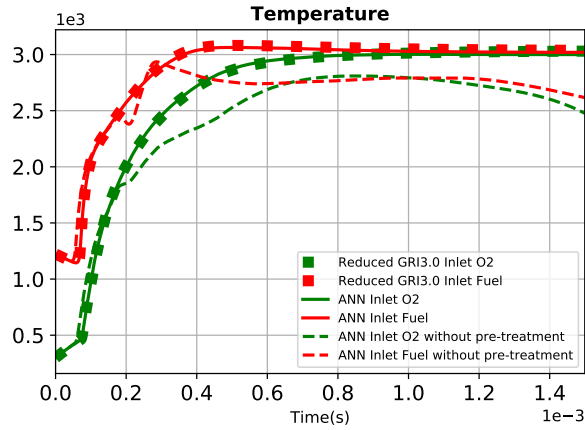


Figure 7: Time evolution of temperature [K] averaged over stochastic particles. Symbols: reference reduced scheme (Table 2 of [13]). Continuous line: Kmeans-PCA-ANN. Dashed-line: ANN without data pre-treatment. Green: O<sub>2</sub> inlet. Red: Fuel inlet.

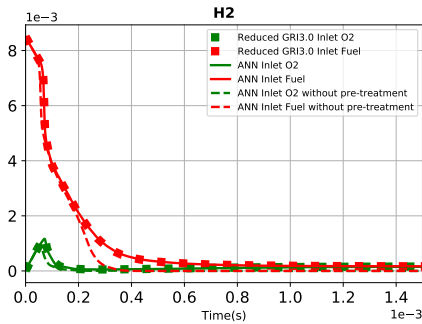
165 The training process is performed using Tensorflow 2 with GPU support (NVIDIA GeForce  
166 GTX 1080 Ti) and the Adam optimizer in default setting. To prevent the overfitting, the early  
167 stopping callback is adopted and set as 200 epochs, which means that the training will be halted  
168 after 200 epochs without improvement. Additionally, the check point callback saves the best model  
169 with the lowest mean squared error (MSE) during the training process. The total training time for  
170 all ANNs is about 11 hours.

### 171 3. Results

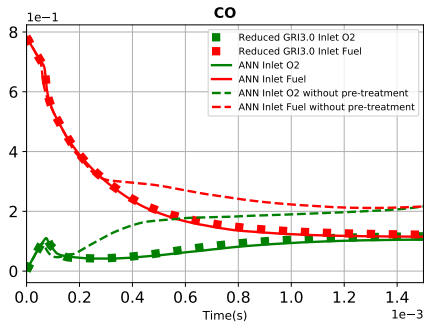
172 The methodology adopted in the study is summarised in Figure 5. To perform a posteriori tests  
173 of the model, the evolutions of the stochastic particles are recomputed replacing the integration of  
174 chemistry by the ANN, as summarised in Figure 6.

175 The time evolutions of temperature and species mass fractions averaged over the stochastic  
176 particles issued from one of the inlets and computed with and without ANNs, are shown in figures  
177 7-9.

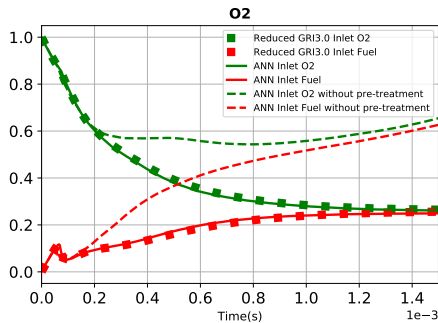
178 The case without any pre-treatment of the data (dashed-line) and a single ANN rapidly deviates  
179 from the reference solution (symbols). Without pre-treatment, the ANN captures the very first part



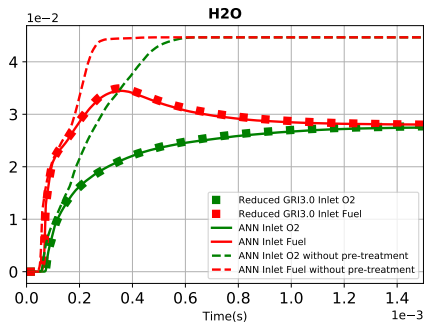
(a) H<sub>2</sub>



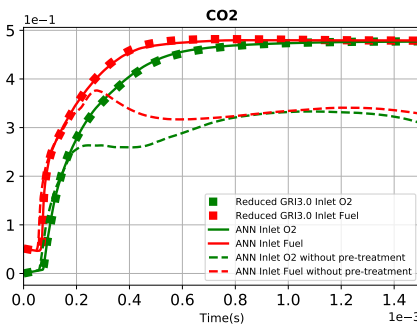
(b) CO



(c) O<sub>2</sub>



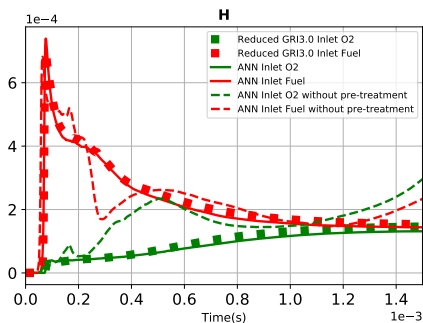
(d) H<sub>2</sub>O



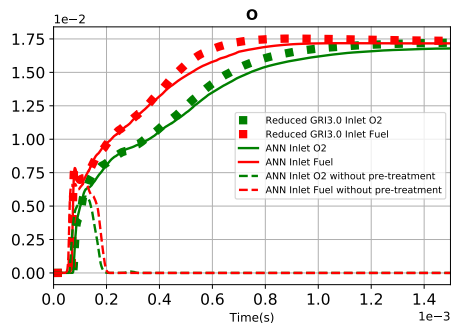
(e) CO<sub>2</sub>

Figure 8: Time evolution of major species mass fractions averaged over stochastic particles. Symbols: Reference chemical scheme (Table 2 of [13]). Continuous line: Kmeans-PCA-ANN. Dashed-line: ANN without data pre-treatment. Green: O<sub>2</sub> inlet. Red: Fuel inlet.

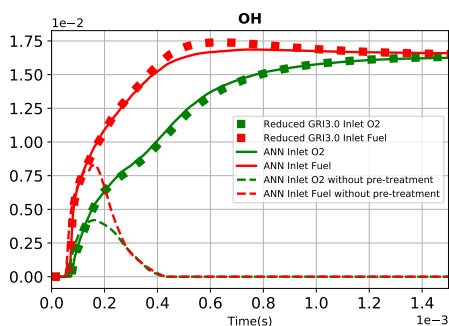
180 of ignition, but fails to reach the equilibrium condition, see for instance the temperature response in  
 181 figure 7. (Notice that one does not make use here of the introduction of manufactured intermediate  
 182 solutions for H<sub>2</sub>O<sub>2</sub>, as discussed in [10].)



(a) H



(b) O



(c) OH

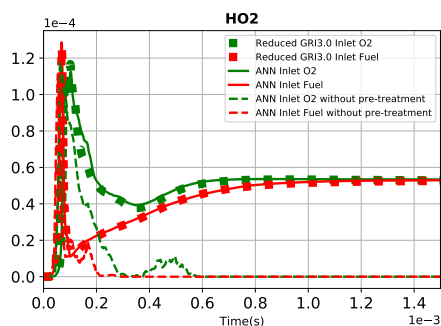
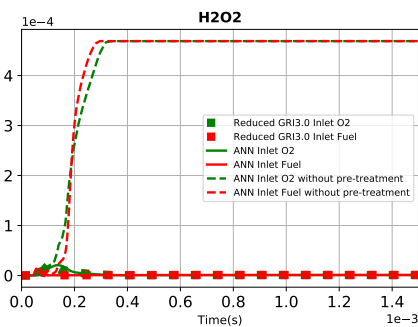
(d) HO<sub>2</sub>(e) H<sub>2</sub>O<sub>2</sub>

Figure 9: Time evolution of radicals and intermediate species mass fractions averaged over stochastic particles. Symbols: Reference chemical scheme (Table 2 of [13]). Continuous line: Kmeans-PCA-ANN. Dashed-line: ANN without data pre-treatment. Green: O<sub>2</sub> inlet. Red: Fuel inlet.

183 The solution with data pre-treatment (solid lines in figures 7-9) is in good agreement with the  
 184 reference solution obtained using a CPU consuming stiff-solver for chemistry (symbols). This is  
 185 the case for temperature and all species considered, even radicals and intermediate ones such as O,

186 H, OH, HO<sub>2</sub> and H<sub>2</sub>O<sub>2</sub>. All the evolution stages (mixing, ignition, combustion and equilibrium)  
187 are well predicted.

188 Figure 7 shows that the ignition delay of 0.08 ms is correctly predicted by the ANN model.  
189 After the ignition, the temperature profiles feature a sharp increase to attain high values due to oxy-  
190 combustion (the highest value is 3081K for the particles from the fuel side). Then, the temperature  
191 proceed towards its equilibrium state at 3037K. The trajectories of reactants (H<sub>2</sub>, CO, O<sub>2</sub>), and  
192 main combustion products (CO<sub>2</sub>, H<sub>2</sub>O) are also perfectly captured by the ANNs-based model  
193 during their evolution in all specific stages, i.e. mixing, ignition, combustion and equilibrium  
194 (Figure 8).

195 Regarding minor and radicals species (Figure 9), the shape and peak of the trajectories are also  
196 correctly predicted except a very slight discrepancy for H<sub>2</sub>O<sub>2</sub> species departing from the oxygen  
197 inlet.

198 The computation time is drastically reduced by using the ANN model as given in Table 2. The  
199 simulation with ANNs (single processor calculation) is 8 times faster than the simulation for the  
200 reduced scheme, and 80 times faster than the simulation with the detailed GRI3.0 scheme. The  
201 CPU time reduction is measured here solving for micro-mixing and chemistry (Eq. (2)). This  
202 relative CPU reduction thus differs from our previous works [10], also relying on ANN for chem-  
203 istry, in which the speed-up was measured solving for the aerothermochemical equations in three-  
204 dimensional flows.

205 This quantification of the speed-up does not account for the preprocessing time needed to train  
206 the neural networks. However, the CPU effort required for training stays much smaller than the  
207 time spent performing the numerous operations required to generate a reduced chemical scheme  
208 (analysis of the most important chemical species and reaction paths and optimisation of the rate  
209 constants). What could slightly increase the CPU cost with ANN is the introduction of more  
210 degrees of freedom which may become mandatory to reproduce very specific combustion charac-  
211 teristics (cool flame effects, etc.).



Chemistry	Speed-up
GRI3.0 detailed scheme	1
Reduced scheme	10
ANNs (present work)	80

Table 2: Speed-up between chemistry descriptions.

## 212 4. Conclusion

213 A novel methodology based on the training of neural networks was proposed to reduce com-  
 214 bustion chemistry and applied to a non-premixed syngas oxy-flame canonical configuration. The  
 215 unsupervised learning technique K-means was used to cluster the entire dataset, obtained from  
 216 the solution of stochastic particles in a turbulent micro-mixing system, into multiple subsets. In  
 217 the case of a two-inlet problem, with one of the streams bringing the energy necessary for igni-  
 218 tion, the proposed decomposition in clusters is likely to be robust and generic. Indeed, after some  
 219 turbulent mixing, the mixture will always ignite to then evolve towards equilibrium. Some level  
 220 of finite rate chemistry may appear, but full quenching is very unlikely, especially in the case of  
 221 oxy-combustion studied here. However, the addition of heat losses and of non-adiabatic recircu-  
 222 lating burnt gases may require further examination to determine the best compromise in terms of  
 223 clustering, still following a similar approach.

224 Local Principal Component Analysis (LPCA) was then applied on each subset to reduce the  
 225 dimension of the problem and most important principal components were retained for the ANN  
 226 training process. Neural networks were finally trained with the aim to replace expensive chem-  
 227 istry integration. The simulation results showed that the proposed method is able to predict the  
 228 evolution of the thermochemical variables (temperature and species concentration) with excellent  
 229 accuracy over the whole domain covering from mixing to ignition, combustion and finally chemi-  
 230 cal equilibrium. All of those with a drastically reduced simulation time.

231 This promising result is paving the way for the introduction of reduced chemical schemes based  
 232 on ANN training in reacting flow simulations, in order to mitigate CPU times without accuracy  
 233 loss.

## 234 Acknowledgments

235 The PhD of the first author is funded by ANRT (Agence Nationale de la Recherche et de la  
236 Technology) and ArcelorMittal under the CIFRE No. 2019/0056.

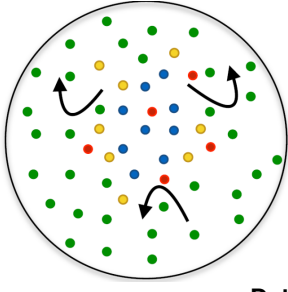
## 237 References

## 238 References

- 239 [1] N. Peters, Systematic reduction of flame kinetics: Principles and details, in: A. L. Kulh, J. R. Bowen, J.-C.  
240 Leyer, A. Boris (Eds.), Dynamics of reactive systems, No. 113, Washington, DC., 1988, pp. 67–86.
- 241 [2] T. Lu, C. K. Law, A directed relation graph method for mechanism reduction, Proc. Combust. Inst. 30 (1) (2005)  
242 1333–1341.
- 243 [3] P. Pepiot-Desjardins, H. Pitsch, An efficient error-propagation-based reduction method for large chemical kinetic  
244 mechanisms, Combust. Flame 154 (1-2) (2008) 67–81.
- 245 [4] F. M. Karalus, K. B. Fackler, I. V. Novosselov, J. C. Kramlich, P. C. Malte, A skeletal mechanism for the reactive  
246 flow simulation of methane combustion, in: Turbo Expo: Power for Land, Sea, and Air, Vol. 55119, American  
247 Society of Mechanical Engineers, 2013, p. V01BT04A065.
- 248 [5] T. Lu, C. K. Law, Strategies for mechanism reduction for large hydrocarbons: n-heptane, Combust. flame 154 (1-  
249 2) (2008) 153–163.
- 250 [6] C. Chi, G. Janiga, D. Thévenin, On-the-fly artificial neural network for chemical kinetics in direct numerical  
251 simulations of premixed combustion, Combust. Flame 226 (2021) 467–477.
- 252 [7] Z. Nikolaou, C. Chrysostomou, L. Vervisch, R. S. Cant, Progress variable variance and filtered rate modelling  
253 using convolutional neural networks and flamelet methods, Flow Turbulence Combust. 103 (2) (2019) 485–501.
- 254 [8] K. Wan, S. Hartl, L. Vervisch, P. Domingo, R. Barlow, C. Hasse, Combustion regime identification from machine  
255 learning trained by Raman/Rayleigh line measurements, Combust. Flame 219 (2020) 168–274.
- 256 [9] G. Aversano, A. Bellemans, Z. Li, A. Coussement, O. Gicquel, A. Parente, Application of reduced-order models  
257 based on PCA & Kriging for the development of digital twins of reacting flow applications, Comput. Chem.  
258 Eng. 121 (2) (2019) 422–441.
- 259 [10] K. Wan, C. Barnaud, L. Vervisch, P. Domingo, Chemistry reduction using machine learning trained from non-  
260 premixed micro-mixing modeling: Application to DNS of a syngas turbulent oxy-flame with side-wall effects,  
261 Combust. Flame 220 (2020) 119–129.
- 262 [11] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of  
263 the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1, Oakland, CA, USA, 1967, pp.  
264 281–297.

- 265 [12] S. Alqahtani, T. Echekki, A data-based hybrid model for complex fuel chemistry acceleration at high tempera-  
266 tures, *Combust. Flame* 223 (2021) 142–152.
- 267 [13] K. Wan, C. Barnaud, L. Vervisch, P. Domingo, Machine learning for detailed chemistry reduction in DNS of a  
268 syngas turbulent oxy-flame with side-wall effects, *Proc. Combust. Inst.* 38 (2021) 2825?2833.
- 269 [14] S. Subramaniam, S. B. Pope, A mixing model for turbulent reactive flows based on euclidean minimum spanning  
270 trees, *Combust. Flame* 115 (4) (1998) 487–514.
- 271 [15] N. Jaouen, L. Vervisch, P. Domingo, G. Ribert, Automatic reduction and optimisation of chemistry for turbulent  
272 combustion modelling: Impact of the canonical problem, *Combust. Flame* 175 (2017) 60–79.
- 273 [16] A. Bouaniche, N. Jaouen, P. Domingo, L. Vervisch, Vitiated high Karlovitz n-decane/air turbulent flames: Scal-  
274 ing laws and micro-mixing modeling analysis, *Flow Turbulence Combust.* 102 (1) (2019) 235–252.
- 275 [17] C. Locci, L. Vervisch, B. Farcy, N. Perret, Selective non-catalytic reduction (SNCR) of nitrogen oxide emissions:  
276 A perspective from numerical modeling, *Flow Turbulence Combust.* 100 (2) (2018) 301–340.
- 277 [18] K. Bioche, L. Vervisch, G. Ribert, Premixed flame-wall interaction in a narrow channel: Impact of wall thermal  
278 conductivity and heat losses, *J. Fluid Mech.* 856 (2018) 5–35.
- 279 [19] A. Seltz, P. Domingo, L. Vervisch, Solving the population balance equation for non-inertial particles dynamics  
280 using probability density function and neural networks: Application to a sooting flame, *Phys. Fluids* 33 (1)  
281 (2021) 013311.
- 282 [20] G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg et al. GRI mech, [http:  
283 //combustion.berkeley.edu/gri-mech/version30/text30.html](http://combustion.berkeley.edu/gri-mech/version30/text30.html).
- 284 [21] N. Kambhatla, T. K. Leen, Dimension reduction by local principal component analysis, *Neural Comput.* 9 (7)  
285 (1997) 1493–1516.
- 286 [22] G. D'Alessio, A. Parente, A. Stagni, A. Cuoci, Adaptive chemistry via pre-partitioning of composition space  
287 and mechanism reduction, *Combust. Flame* 211 (2020) 68–82.
- 288 [23] D. Arthur, S. Vassilvitskii, K-means++: The advantages of careful seeding, Tech. rep., Stanford (2006).
- 289 [24] J. Jolliffe, *International Encyclopedia of Statistical Science*, Springer (2011) 1094–1096.

**Stochastic particles undergoing turbulence/chemistry interaction**

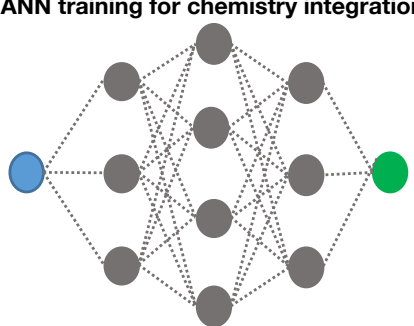


$$\frac{dY_i^p(t)}{dt} = \text{MIX}_i^p(\tau_T) + \dot{\omega}_i^p$$

$$\frac{dh_s^p(t)}{dt} = \text{MIX}_{h_s}^p(\tau_T) + \dot{\omega}_{h_s}^p$$

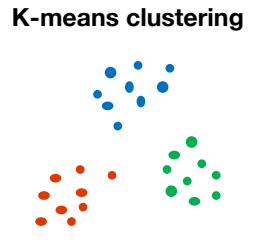
**Database generation**

**ANN training for chemistry integration**

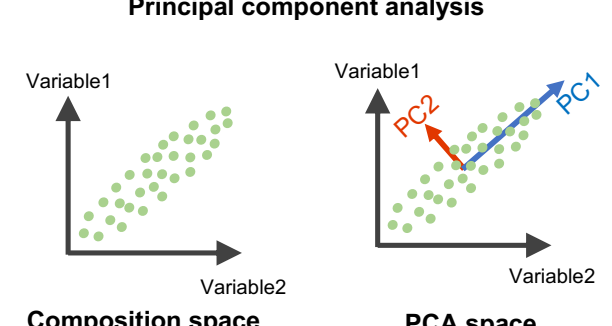


**Data pretreatment for efficient machine learning chemistry integration**

**K-means clustering**



**Principal component analysis**



**Composition space**      **PCA space**

↑

→